PROGRAMA DE DOCTORADO EN BIOLOGÍA MOLECULAR Y BIOMEDICINA

TESIS DOCTORAL

ESTUDIO MOLECULAR DE LOS MECANISMOS RESPONSABLES DEL PAPEL DE LOS COMPLEJOS SWI/SNF EN LA PROGRESIÓN TUMORAL Y SU POSIBLE USO PARA EL DESARROLLO DE TERAPIAS CONTRA EL CÁNCER

PHD TESIS

MOLECULAR CHARACTERIZATION OF THE ROLE OF SWI/SNF COMPLEX IN TUMOR PROGRESSION AND ITS POTENTIAL APPLICATION FOR THE DEVELOPMENT OF NEW THERAPEUTIC APPROACHES AGAINST CANCER

AUTORA

Beatriz Monterde Martínez

DIRECTOR

Dr. Ignacio Varela Egocheaga

UNIVERSIDAD DE CANTABRIA

Escuela de Doctorado de la Universidad de Cantabria

Santander 2023

El Dr. IGNACIO VARELA EGOCHEAGA, Profesor Titular de la Universidad de Cantabria (UC) e Investigador Principal del grupo “Genómica Funcional de la Progresión Tumoral”, en el Departamento de Señalización Celular y Molecular del Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC)

CERTIFICA: que BEATRIZ MONTERDE MARTÍNEZ ha realizado bajo su dirección la presente tesis doctoral, titulada: “Estudio molecular de los mecanismos responsables del papel de los complejos SWI/SNF en la progresión tumoral y su posible uso para el desarrollo de terapias contra el cáncer/ Molecular characterization of the role of SWI/SNF complexes in tumor progression and its potential application for the development of new therapeutic approaches against cancer”.

Considero que este trabajo reúne los requisitos de originalidad y calidad científica necesarios para su presentación como Memoria de Doctorado al objeto de optar al grado de Doctor en Biología Molecular y Biomedicina por la Universidad de Cantabria, con mención internacional.

Y para que conste y surta los efectos oportunos, firmo el presente certificado en Santander, a 6 de marzo de 2023.

Fdo.: Ignacio Varela Egocheaga

La presente tesis doctoral titulada “Estudio molecular de los mecanismos responsables del papel de los complejos SWI/SNF en la progresión tumoral y su posible uso para el desarrollo de terapias contra el cáncer/Molecular characterization of the role of SWI/SNF complexes in tumor progression and its potential application for the development of new therapeutic approaches against cancer” ha sido realizada en el Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), perteneciente a la Universidad de Cantabria (UC) y al Consejo Superior de Investigaciones Científicas (CSIC), en el laboratorio del grupo “Genómica Funcional de la Progresión Tumoral”, en Santander (Cantabria, España). Además, durante el desarrollo de la formación predoctoral, se completó la investigación realizando una estancia internacional en el grupo “Mechanism of Repression by Polycomb Proteins”, en el Institut Curie (Paris, Francia).

La financiación para la realización de esta tesis doctoral ha sido proporcionada por las siguientes fuentes:

- Ministerio de Economía y Competitividad. Proyecto: “Estudio molecular de los mecanismos responsables del papel de los complejos SWI/SNF en la progresión tumoral y su posible uso para el desarrollo de terapias contra el cáncer” (2016). SAF2016-76758-R.
- Ministerio de Ciencia e Innovación. Proyecto: “Conexiones entre las alteraciones en el complejo remodelador SWI-SNF y las rutas tumorales canónicas en cáncer de pulmón” (2021). PID2020-117539GB-I00.

La autora de esta tesis ha disfrutado de las siguientes fuentes de financiación:

- Ayuda para la Formación del Profesorado Universitario (FPU) (referencia FPU18/00755, área de Biomedicina), concedida por el Ministerio de Ciencia, Innovación y Universidades.
- EMBO Scientific Exchange Grant (reference 9207), concedida por European Molecular Biology Organization.
- Ayuda para estancias breves dentro del programa FPU (referencia EST21/00243, área de Biomedicina), concedida por el ministerio de Universidades.

A mi familia y amigos cercanos,

por impulsarme a lograr todas mis metas.

Resumen

Los complejos SWI/SNF son remodelares de cromatina dependientes del uso de ATP, formados por varias subunidades que se pueden combinar de diferentes maneras y que se encargan de regular de manera precisa varios procesos biológicos esenciales, como son la transcripción génica, la estructura de la cromatina, la diferenciación celular o la integridad del genoma. Los proyectos de secuenciación del genoma del cáncer han revelado una elevada frecuencia mutacional de los genes codificantes del complejo SWI/SNF, la cual se distribuye en un amplio espectro de cánceres y se ha asociado a una menor tasa de supervivencia en los pacientes. Se estima que en torno a un 25% de todos los cánceres humanos contienen alteraciones en alguno de los componentes del complejo SWI/SNF, lo que lo convierte en la segunda alteración más frecuente en cáncer, justo después de las mutaciones en *TP53*.

A pesar de la extensa evidencia científica apoyando la intervención del complejo SWI/SNF en el desarrollo tumoral, los mecanismos moleculares específicamente involucrados en este proceso todavía se desconocen. En esta tesis doctoral, hemos caracterizado los mecanismos moleculares regulados por los complejos SWI/SNF para intentar descifrar la manera mediante la cual su alteración promueve la formación de tumores. Asimismo, la búsqueda de vulnerabilidades terapéuticas en tumores deficientes en SWI/SNF es de gran interés porque podría dar lugar a una mejora en el diagnóstico y el tratamiento de los pacientes de cáncer.

Para lograr estos objetivos, hemos generado una colección de líneas celulares procedentes de diferentes tipos tumorales que expresan de manera inducible shRNAs dirigidos específicamente contra las subunidades del complejo SWI/SNF más frecuentemente mutadas en cáncer (*ARID1A*, *ARID1B*, *ARID2* o *SMARCA4*). Con la finalidad de caracterizar las redes génicas directamente reguladas por los complejos SWI/SNF, hemos generado datos sobre el perfil transcripcional, la estructura de la cromatina y el perfil epigenético de estas líneas celulares tras inactivar distintas subunidades del complejo SWI/SNF, los cuales se han obtenido a través de diferentes metodologías ómicas (RNA-Seq, ATAC-Seq y CUT&RUN).

Gracias a estos estudios, hemos hallado redes génicas reguladas específicamente por diferentes subunidades del complejo en la misma línea celular, así como diferentes programas de expresión génica controlados por la misma subunidad del complejo en diferentes tejidos. Por lo tanto, los genes diana de los complejos SWI/SNF dependen de su composición, así como de su interacción con factores de transcripción específicos de tejido.

A pesar del bajo grado de solapamiento entre los genes diferencialmente expresados en cada condición, hemos encontrado un solapamiento relativamente mayor a nivel de rutas moleculares, siendo la transición epitelio-mesénquima y la reparación del DNA las vías de señalación alteradas más relevantes en nuestro contexto.

Asimismo, hemos observado que, como consecuencia de la alteración del complejo SWI/SNF, se produce un aumento de la resistencia a inhibidores de EGFR, con independencia de la presencia de mutaciones en *KRAS* o *TP53*. Además, la deficiencia de *ARID2* da lugar a un aumento de la sensibilidad a quimioterápicos e inhibidores de PARP. Por lo tanto, a partir de estas observaciones experimentales se puede deducir que los pacientes con mutaciones en el complejo SWI/SNF presentarían un resultado clínico menos favorable frente a aquellos tratamientos basados en la inhibición de EGFR. En cambio, es más probable que se beneficien de aquellas opciones terapéuticas cuyo mecanismo de acción es la generación de daño en el ADN. Estos resultados sugieren que la estratificación de los pacientes de cáncer de pulmón de acuerdo a la presencia de alteraciones en el complejo SWI/SNF podría ser de gran utilidad clínica.

La alteración del complejo SWI/SNF se asocia a cambios estructurales específicos en la cromatina, con un claro predominio de la pérdida de accesibilidad en enhancers. Estas observaciones apoyan la idea de que los complejos SWI/SNF son necesarios para mantener abierta la estructura de la cromatina en torno a estas regiones reguladoras. Sin embargo, no hemos sido capaces de demostrar una correlación general entre las alteraciones transcripcionales y los cambios de accesibilidad de los enhancers que teóricamente están asociados a estos genes, de acuerdo a la información derivada del proyecto GeneHancer. Por lo tanto, asumimos que la regulación de la expresión génica mediada por los complejos SWI/SNF es más compleja y no se puede explicar teniendo en cuenta únicamente los cambios en la accesibilidad de los enhancers.

Del mismo modo, la alteración del complejo SWI/SNF también produce cambios en la distribución de las modificaciones de histonas H3K27ac y H3K4me1 en enhancers. En particular, cada subunidad es necesaria para mantener la accesibilidad, la acetilación y la mono-metilación de diferentes enhancers, lo cual concuerda con el perfil transcripcional tan específico que hemos observado en los diferentes modelos celulares generados.

Por último, hemos identificado algunos ejemplos concretos de este mecanismo de regulación. La actividad de los complejos SWI/SNF se requiere para mantener una estructura accesible de la cromatina, así como un perfil epigenético concreto en los enhancers y promotores que regulan la expresión de genes específicos (*MTSS1, SDK1, SCGB3A2, EPGN, NRG1, KLF6* y *SOX6*). Estas observaciones podrían estar directamente implicadas en los fenotipos celulares observados.

La integración de los resultados generados por las diferentes metodologías empleadas nos lleva a proponer que la alteración en los complejos SWI/SNF promueve el desarrollo tumoral a través de dos mecanismos complementarios:

(1) En primer lugar, afecta a la capacidad de las células para reparar el daño en el ADN, incrementando su sensibilidad a tratamientos antitumorales que promueven la inestabilidad genómica.

(2) En segundo lugar, las alteraciones del complejo SWI/SNF producen una pérdida de la regulación de genes importantes para definir la identidad celular y, por lo tanto, favorecen un estado menos diferenciado en el que las células son menos dependientes de las vías de señalización iniciadas por EGFR (receptor del factor de crecimiento epidérmico). Como consecuencia, las células aumentarían la expresión de marcadores mesenquimales y factores de transcripción que promueven la transición epitelio-mesénquima, al tiempo que reducirían la expresión de ligandos específicos de receptores con actividad tirosina quinasa. Por tanto, las células deficientes en SWI/SNF experimentan la transición epitelio-mesénquima de una manera más sencilla y rápida, mostrando una mayor resistencia a aquellos tratamientos basados en la inhibición de EGFR.

Summary

SWI/SNF complexes are ATP-dependent chromatin remodelers that exist in various combinatorial assemblies and are in charge of the tight regulation of many essential biological processes, including genetic transcription, chromatin structure, cell differentiation, or genomic integrity. Cancer genome sequencing studies have evidenced a remarkably high mutation frequency of genes encoding subunits of the SWI/SNF complex, which are widely distributed across human cancers, and have been associated with worse prognoses. It is estimated that almost 25% of all human malignancies contain alterations in any of their components, which places SWI/SNF as the second most frequent alteration in cancer, just after the presence of mutations in *TP53*.

Besides the solid evidence supporting the role of SWI/SNF complex in cancer development, the specific molecular mechanisms involved in this process remain largely unknown. In this doctoral thesis, we have characterized the molecular pathways controlled by SWI/SNF complexes to increase our knowledge of how their alteration promotes tumorigenesis. In addition to this, the exploration of therapeutic vulnerabilities associated with SWI/SNF deficiency is of great interest because it could lead to an improvement in the diagnosis and treatment of cancer patients.

To achieve these objectives, we have generated a collection of cell lines obtained from different tumor types that express shRNAs in an inducible way, targeting the most commonly mutated subunits of the SWI/SNF complex (*ARID1A, ARID1B, ARID2,* or *SMARCA4*). In order to characterize the gene networks directly regulated by SWI/SNF complexes, we have integrated information of the transcriptomics, the chromatin structure, and the histone epigenetic profile of SWI/SNF-deficient cells using different omic methodologies (e.g. RNA-Seq, ATAC-Seq, and CUT&RUN).

We have found gene networks specifically regulated by different subunits of the complex in the same cell line and also different gene expression programs controlled by the same subunit of the complex in different cellular models. Therefore, target genes of SWI/SNF

complexes depend not only on their composition, but also on their interaction with tissue-specific transcription factors.

Besides the low overlap between the differentially expressed genes reported in each condition, we have observed a considerably higher overlap at the pathway level, being the epithelial to mesenchymal transition and the DNA repair the most relevant signaling pathways altered.

SWI/SNF alteration results in increased resistance to EGFR inhibitors, irrespective of the mutational status of *KRAS* or *TP53*. In addition to this, *ARID2* deficiency gives rise to an enhanced sensitivity to chemotherapeutic agents that produce DNA damage and PARP inhibitors. Therefore, it could be inferred that patients harboring mutations in genes encoding components of the SWI/SNF complex would present a worse clinical outcome to treatments based on the inhibition of EGFR, while they are more likely to benefit from therapeutic options based on the generation of DNA damage.

We have observed that SWI/SNF alteration is accompanied by specific chromatin structural changes at the complete genome, with a significant loss of accessibility affecting enhancers. These findings support the idea that SWI/SNF is required to keep an open chromatin structure around these regulatory regions. However, we did not observe a clear correlation between transcriptional alterations and changes in accessibility at the theoretical enhancers controlling the expression of those genes according to the GeneHancer project. Therefore, we assume that gene expression regulation mediated by SWI/SNF complexes is more intricate and cannot be explained just by changes in enhancer accessibility.

SWI/SNF alteration results in deep changes in the deposition of both H3K27ac and H3K4me1 histone modifications at enhancers. In addition to this, each subunit is required to maintain the accessibility, the acetylation, and the mono-methylation of different enhancers, in accordance with the subunit-specific transcriptional program reported in the different cellular models generated.

Finally, we have identified some examples of SWI/SNF-mediated gene expression regulation. Thus, SWI/SNF complex is required to ensure an open chromatin structure, as well as an active histone epigenetic profile at the enhancers and promoters in charge of the regulation of specific genes (e.g. *MTSS1, SDK1, SCGB3A2, EPGN, NRG1,*

KLF6, and *SOX6*), that are likely involved in the phenotypes observed in the cells.

Taking into consideration all the data generated in our cellular models, we propose that SWI/SNF alteration fosters tumor development by at least two parallel mechanisms:

(1) First, it affects the ability of the cells to repair DNA damage, which is translated into a higher sensitivity to antitumoral treatments based on the promotion of genomic instability.

(2) Second, SWI/SNF alteration results in the dysregulation of cell-identity genes and the promotion of a less differentiated state that is less dependent on those signaling pathways initiated by EGFR (epidermal growth factor receptor). Consequently, cells would increase the expression of mesenchymal markers and transcription factors favoring the epithelial to mesenchymal transition, while reducing the expression of specific ligands of tyrosine kinase receptors. Thus, SWI/SNF-deficient cells are more prone to the epithelial to mesenchymal transition, exhibiting a higher resistance to those treatments based on the inhibition of EGFR.

“It is not the result of scientific research
that ennobles humans and enriches their nature,
but the struggle to understand while performing
creative and open-minded intellectual work.”

– Albert Einstein

El secreto para obtener resultados que duren consiste en no dejar nunca de hacer mejoras.

Es notable lo que puedes construir si no te detienes.

Es notable el negocio que puedes construir si no dejas de trabajar.

Es notable el cuerpo que puedes llegar a tener si no dejas de entrenar.

Es notable el conocimiento que puedes adquirir si no dejas de aprender.

Es notable la fortuna que puedes llegar a amasar si no dejas de ahorrar.

Son notables las amistades que puedes construir si no dejas de preocuparte por los demás.

Los pequeños hábitos no se suman.

Los pequeños hábitos se componen.

Ese es el poder de los hábitos atómicos.

Pequeños cambios.

Resultados extraordinarios.

Hábitos atómicos – James Clear

Agradecimientos

Me gustaría expresar mi más sincero agradecimiento hacia todas aquellas personas que han formado parte de esta etapa formativa, que sin duda me ha enriquecido y me ha hecho evolucionar a nivel profesional, pero sobre todo a nivel personal.

En primer lugar, a mi familia, por transmitirme los valores que me identifican, por enseñarme el significado de la perseverancia, la dedicación y el compromiso, por ser la luz que ha guiado mis pasos desde los inicios, por brindarme la oportunidad de formarme en lo que más me apasiona, por su incondicional apoyo en todas las aventuras que he emprendido y por animarme a perseguir aquello que me ilusiona. Os quiero, mamá y papá, sois mi mayor ejemplo de superación.

A Nacho y a todo el equipo 03.05 por acogerme y guiarme durante el proceso. Me sigue pareciendo increíble todo lo que he aprendido durante estos años. Gracias por brindarme las herramientas para lograr cosas que me parecían impensables, por ayudarme a descubrir habilidades que desconocía, por despertar en mí el espíritu crítico, la reflexión y, sobre todo, por aprender a pensar y dar con las preguntas adecuadas. Gracias por permitirme conocer la ciencia fuera de España y por darme la oportunidad de asistir a congresos y eventos científicos representado nuestra investigación.

A todos los que han compartido laboratorio conmigo, desde los inicios hasta la persona en la que me he convertido hoy. Especial mención a Laurita, comenzamos clonando proteínas fluorescentes y hemos terminado siendo bioinformáticas. Gracias por enseñarme en los inicios, por ayudarme a sentir como en casa y por evolucionar a mi lado durante todos estos años.

Al Dr. Raphaël Margueron, por acogerme en su excepcional grupo de investigación y darme la oportunidad de vivir una de las mejores etapas de mi vida. Fue un auténtico placer poder aprender de la mano de expertos en un campo que desconocía hasta ese momento. A Samuel, Daniel Jr., Daniel Sr., Audrey, Laia, Sabina, Emma, Carlos, Michel, Manuela, Aurélien, Seynabou, Pierre, Aya, Yazan, Soraya, Xavi, Vedrana, Ignacio, Rami. Especial mención a Camila, por ser mi Mindy nada más llegar a Paris, gracias por tu apoyo, las experiencias y el tiempo que pasamos juntas descubriendo la ciudad. También a mi

pequeña familia española parisina, Carlota, Lyane, Pablo, María, Martina, Sinan, Alba, Vini, Amir, Mir, Víctor. Siempre nos quedará París.

Au Dr Raphaël Margueron, pour m'avoir accueilli dans son exceptionnel groupe de recherche et m'avoir donné l'opportunité de vivre l'une des plus belles étapes de ma vie. Il a été un réel plaisir de pouvoir apprendre auprès d'experts dans un domaine que j'ignorais jusqu'alors. À Samuel, Daniel Jr., Daniel Sr., Audrey, Laia, Sabina, Emma, Carlos, Michel, Manuela, Aurélien, Seynabou, Pierre, Aya, Yazan, Soraya, Xavi, Vedrana, Ignacio, Rami. Mention spéciale à Camila, pour avoir été ma Mindy depuis mon arrivée à Paris, merci pour ton soutien, les expériences et le temps que nous avons passé ensemble en découvrant la ville. Aussi à ma petite famille parisienne espagnole, Carlota, Lyan, Pablo, María, Martina, Sinan, Alba, Vini, Amir, Mir, Víctor. Nous aurons toujours Paris.

A Dolo y Ana, por seguir de cerca mi trabajo durante todos estos años, por vuestras palabras de apoyo al final de cada evaluación, por vuestro criterio y consejos, por ayudarme en cualquier momento que lo he necesitado y por la objetividad con la que habéis guiado y valorado mi trabajo.

A todas las personas que han participado en presentaciones, charlas, seminarios, meetings de los que he formado parte. Vuestra opinión y feedback me han ayudado a mejorar mi criterio, a encontrar las preguntas correctas o a dar otro enfoque a nuestra investigación. Especial mención a los PI de los Cancer Club, Javier, Dolo, Chepe, Berta, Juanqui, Fernando, vuestra opinión y críticas constructivas me han animado a continuar desde que di mi primera charla. También a los investigadores de otros campos que, desde su experiencia, me han aportado nuevas perspectivas y han seguido incentivando mi curiosidad. A Mapi, Álvaro, María Lucas, AnaVi, Ruli, Santi, Fernando.

A quienes he conocido en congresos y cursos, por la ilusión de los comienzos, por compartir nuestra pasión, por enriquecerme con distintas perspectivas. A Isaac, Pablo, Marc, Alicia, Dido, Alba, Javi, Sara, Àngela, Alex. Y también a los excepcionales ponentes y organizadores de estos eventos, por despertar en nosotros el afán de seguir adelante, por ilusionarnos y emocionarnos con sus grandes descubrimientos.

A mis compañeros de otros laboratorios del IBBTEC, por su predisposición a ayudar siempre que lo he necesitado, por las colaboraciones y la gran familia que hemos formado. A Diane, Patri, Alfonso, Alex, Víctor, María, David, Silvia, Ana, Meri, Vane, Jorge, David, Aurora, Carlos, Lolu, Marta, Teresa, Fer, Judit, Esther, Nuri, Ful, Agus, Santi, Emilio, Omar,… Somos muchos, seguro que se me olvida mencionar a alguien, no tengo palabras para agradecer lo mucho que me habéis ayudado en esta etapa, por hacer más agradable la convivencia, por compartir nuestros experimentos fallidos, pero también por celebrar las pequeñas victorias del día a día. Por las excursiones, los cafés tomando el sol en la rotonda, las barbacoas, las comidas y cenas para celebrar los pequeños logros.

Al laboratorio de apoyo, al personal técnico y administrativo, al equipo de seguridad y de limpieza, así como a todo el personal técnico del IBBTEC, por su inestimable labor para lograr que la ciencia siga su camino hacia delante y por hacer que nuestro día a día sea un poco más sencillo.

A mis compañeros del máster de Cantabria, con quienes empezó esta andadura, en la que al final tomamos caminos separados. A Eva, Ingrid, Sandra, Óscar.

También a mis compañeros del máster online en bioinformática, aunque no nos conozcamos en persona, creamos una gran comunidad. Ni una pandemia fue capaz de frenar nuestro ímpetu. Especial mención a Mapi y Santi, del team IBBTEC, por aconsejarme, convencerme para iniciar esta aventura y brindarme ayuda siempre que fue necesario. También a Elena y Pedro, mis directores de TFM, por confiar en mi desde el primer momento, guiarme durante el camino, enseñarme herramientas y enriquecerme como científica y persona. Ha sido un auténtico placer aprender a vuestro lado.

A mis primeros maestros y mentores, que alimentaron mi curiosidad y mis ganas de descubrir cómo funciona la realidad. A Manolo, Gloria, Lola y Carlos por ver aquello que destacaba en mí, por ayudarme a escoger mi camino y por animarme a lograr mis sueños. A mis profesores de la facultad por enseñarme a pensar, por despertar en mí el espíritu crítico, por hacerme ver lo increíble que es la naturaleza y las herramientas que nos aporta para diseñar prácticamente todo lo que podamos imaginar, por guiar mis primeros pasos en el laboratorio y por hacerme ver qué es lo que me apasiona.

También a mis compañeros de piso, con los que compartido tan buenos momentos durante estos años. A Rocío, mi primer compañera de vida, me acogiste como una hermana pequeña, a Diana, por enseñarme los rincones con encanto de la tierruca, a Ana Belén, por ser mi compañera de tesis y compartir nuestras experiencias, a Paula, por disfrutar de los partidos en el Sardinero, a Ray y a Kary, por ayudarme a descubrir mi esencia y a brillar de nuevo, convivimos poco tiempo, pero me llevo amigas de por vida, a Ángela, por tu apoyo, el entusiasmo con el que me siempre me escuchas hablar de ciencia y por brindarme apoyo siempre que lo he necesitado, me hiciste recuperar la ilusión.

A mis amigos de Zaragoza, aunque nos separe la distancia, siempre os he sentido cerca cuando os he necesitado. Por los años que pasamos juntos en el colegio o la facultad, disfrutando de cada etapa y soñando con lo que nos depararía el futuro. A Elena, Paula, Elisa, Elena, Pablo, Mercedes, Sergio, Edu, Mikel, Raquel,… Especial mención a ti Nat, siempre has estado ahí, hemos evolucionado juntas desde el colegio, aprendiendo a entendernos en cada etapa y dándonos apoyo siempre que lo hemos necesitado. Por nuestros paseos interminables por el centro, los cafés, las largas horas de conversaciones y por esa sensación de que no pasa el tiempo cada vez que estoy contigo. Quiero verte brillar y ser feliz en lo que haces. Por muchos más viajes aprendiendo contigo sobre historia y arte. Espero tu visita allá donde vaya.

A mi pequeña familia alocada, Tutti Frutti, el caos que aporta paz a mi vida. A Ray, Titi, Kary y Linda, las estrellas de mi universo. Por ser mi refugio, por cuidarme, por aportarme equilibrio, por ayudarme a brillar y a superar mis miedos. Por los días de playa, las excursiones, las pelis en casa, los shootings, las sesiones de belleza,… Por esa cara de fascinación que ponéis cuando os hablo de ciencia, por escucharme aunque no entendáis la mitad de lo que digo, pero sabéis lo importante que es para mí, por hacer que me sigan brillando los ojos. Nos quedan muchos viajes y aventuras que disfrutar. Uno se queda donde sabe que tiene la libertad de poder irse.

A mi familia fit, por ayudarme a crecer en todos los sentidos, a construir la fortaleza mental que me caracteriza y el espíritu de superación. A Jaime, Coke, Laura, Nando, Nieves, Aarón, Ana, Adri, Krysh, Sarai, Álvaro, Amina, Andrea. Nos quedan muchos viajes y aventuras por vivir, gyms que probar, PRs que conseguir y máquinas que desbloquear, rutas de tortillas, burgers, sushi, pizza. Especial mención a mis gym sisters, Marta y Mariajo, por hacerme ver que puedo lograr todo lo que me proponga, por escucharme con entusiasmo y admiración cada vez que os hablo de ciencia, por vuestro apoyo incondicional en todo momento y por ayudarme a sacar mi mejor versión. Vuestras palabras me han impulsado en cada etapa, me han ayudado a afrontar mis miedos y a superar mis límites. Yo también presumo con orgullo de vosotras. Cierro una etapa de mi vida en la que habéis dejado huella.

Muchas gracias a todos por formar parte de mi vida, pongo fin a un largo camino de aprendizaje, introspección y evolución que recordaré con cariño y nostalgia, pero sin el cual no me habría convertido en la persona que soy a día de hoy.

Atentamente, Bea

"Believe in yourself and all that you are. Know that there is something inside you that is greater than any obstacle." - Christian D. Larson

ABBREVIATIONS

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.”
– Sir William Lawrence Bragg

List of Abbreviations

5-mC	5-Methylcytosine
ACTB	β-Actin
ADP	Adenosine diphosphate
ARID1A	AT-rich interactive domain-containing protein 1A
ARID1B	AT-rich interactive domain-containing protein 1B
ATAC-Seq	Assay for Transposase-Accessible Chromatin with high-throughput sequencing
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BAM	Binary Alignment Map
BET	Bromodomain and extra-terminal domain
bp	Base pair
cBAF	Canonical BRG1-associated factor
cDNA	Complementary DNA
CDS	Coding DNA Sequence
CFSE	Carboxyfluorescein diacetate succinimidyl ester
CGI	CpG island
CHD	Chromodomain helicase DNA-binding
ChIP-Seq	Chromatin immunoprecipitation followed by deep sequencing
CNV	Copy number variation
COSMIC	Catalogue Of Somatic Mutations In Cancer
CPX	Ciprofloxacin
CRE	Cis-regulatory element
C_T	Cycle threshold

CUT&RUN	Cleavage Under Targets & Release Using Nuclease
ddH_2O	Double-distilled water
DAPI	4, 6’-diamino-2 phenylindole dihydrochloride
DEGs	Differentially expressed genes
DMEM	Dulbecco's Modified Eagle Medium
dNTP	Deoxynucleotide triphosphate
DNAse-Seq	DNAse I coupled to deep sequencing
DSB	Double-strand break
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetracetic acid
EGFR	Epidermal growth factor receptor
EGTA	Ethyleneglycoltetraacetic acid
EMT	Epithelial to mesenchymal transition
ENCODE	Encyclopedia of DNA Elements
FAIRE-Seq	Formaldehyde-Assisted Isolation of Regulatory Elements followed by deep sequencing
FITC	Fluorescein isothiocyanate
FBS	Fetal bovine serum
FC	Fold change
FDR	False discovery rate
FPKM	Fragments Per Kilobase Million
GATK	Genome Analysis Toolkit
GREAT	Genomic Regions Enrichment of Annotations Tool
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
H3K4me1	Histone H3 lysine 4 mono-methylation
H3K4me2	Histone H3 lysine 4 di-methylation

H3K4me3	Histone H3 lysine 4 tri-methylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 tri-methylation
H3S10P	Histone H3 serine 10 phosphorylation
H3S28P	Histone H3 serine 28 phosphorylation
H3T118P	Histone H3 threonine 118 phosphorylation
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HiC-Seq	High-throughput chromosome conformation followed by deep sequencing
HMT	Histone methyltransferase
HOMER	Hypergeometric Optimization of Motif EnRichment
HRP	Horseradish peroxidase
IARC	International Agency for Research on Cancer
IC_{50}	Half-minimum inhibitory concentration
ICI	Immune checkpoint inhibitor
ICL	Interstrand crosslink
IFN	Interferon
IGV	Integrative Genomics Viewer
Indels	Small insertions and deletions
INO80	Inositol auxotroph 80
ISWI	Imitation switch
kb	Kilobase
KD	Knock-down
kDa	Kilodalton

LCC	Large cell carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MeDIP-Seq	Methylated DNA immunoprecipitation followed by sequencing
mRNA	Messenger RNA
MRT	Malignant rhabdoid tumor
ncBAF	Non-canonical BRG1-associated factor
NES	Normalized enriched score
NER	Nucleotide excision repair
NP-40	Nonyl phenoxypolyethoxylethanol
NSCLC	Non-small cell lung cancer
NURD	Nucleosome remodeling and deacetylase
NGS	Next-generation sequencing
OCCC	Ovarian clear cell carcinoma
pA-MN	Protein A-MNase
PARP	Poly ADP ribose polymerase
PAGE	Polyacrylamide Gel Electrophoresis
PBAF	Polybromo BRG1-associated factor
PBS	Phosphate buffered saline
PBT	PBS containing 0.05% Triton X-100
PCR	Polymerase chain reaction
PE	Paired-end
PEI	Polyethylenimine
PNK	Polynucleotide kinase
PPi	Pyrophosphate
PRC1	Polycomb repressive complex 1

PRC2	Polycomb repressive complex 2
Q	Phred quality score
qRT-PCR	Reverse transcription and quantitative polymerase chain reaction
RAMSES	Realignment Assisted Minimum Evidence Spotter
RIN^e	RNA Integrity Number equivalent
RNA-Seq	Transcriptome sequencing
RPKM	Reads Per Kilobase Million
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute Medium
RQ	Relative quantification
RTK	Tyrosine kinase receptor
SAM	S-adenosyl methionine
SCCOHT	Small cell carcinoma of the ovary, hypercalcemic type
SCLC	Small cell lung cancer
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEM	Standard error of the mean
shRNA	Short hairpin RNA
shARID1A	shRNA against *ARID1A* gene
shARID1B	shRNA against *ARID1B* gene
shARID2	shRNA against *ARID2* gene
shSMARCA4	shRNA against *SMARCA4* gene
SNV	Single nucleotide variant
SOLiD	Supported Oligonucleotide Ligation and Detection
ssDNA	Single-stranded DNA

SWI/SNF	SWItch/Sucrose Non-Fermentable
TAD	Topologically Associated Domain
TBS-T	Tris Buffer Saline–Tween 20
TCGA	The Cancer Genome Atlas
TE	Tris-EDTA
TGF-β	Transforming growth factor-β
TKI	Tyrosine kinase inhibitor
TMB	Tumor Mutation Burden
tRFP	Turbo-red fluorescent protein
TSS	Transcription start site
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

INDEX

“What I love about science is that as you learn, you don’t really get answers. You just get better questions.”
– John Green

Index

LIST OF FIGURES AND TABLES

“The art and science of asking questions is the source of all knowledge.”
– Thomas Berger

List of Figures

Introduction

Experimental procedures

Results

Figure 4.1. Knock-down validation of A549 stably-transduced cell lines by qRT-PCR.

Figure 4.2. Transcriptomic landscape of SWI/SNF-deficient A549 cell lines.

Figure 4.3. Positive enrichment of EMT in SWI/SNF-deficient A549 cell lines.

Figure 4.4. Transcriptomic landscape of *ARID1A* or *ARID1B*-deficient cell lines.

Figure 4.5. Transcriptomic landscape of SWI/SNF-deficient NCI-H460 cell lines.

Figure 4.6. Positive enrichment of EMT in SWI/SNF-deficient NCI-H460 cell lines.

Figure 4.7. Transcriptomic landscape of SWI/SNF-deficient lung cancer cell lines.

Figure 4.8. *ARID2* deficiency increases proliferative and metastatic potential of cells *in vitro* and *in vivo*.

Figure 4.9. Mutual exclusivity between SWI/SNF and canonical pathways altered in lung adenocarcinoma.

Figure 4.10. Sensitivity of SWI/SNF-deficient cell lines to EGFR inhibition.

Figure 4.11. Transcriptional alterations of *ARID2*-deficient H460 cells treated with the EGFR inhibitor dacomitinib.

Figure 4.12. DNA repair mechanisms are positively enriched in the absence of *ARID2*.

Figure 4.13. ARID2 colocalizes with DNA repair proteins.

Figure 4.14. Effect of *ARID2* deficiency on DNA repair mechanisms in A549 cells.

Figure 4.15. Effect of *ARID2* deficiency on DNA repair mechanisms in NCI-H460 cells.

Discussion

List of Tables

Introduction

Experimental procedures

Results

List of Supplementary Tables

INTRODUCTION

“Science is not only a disciple of reason but also one of romance and passion.”
– Stephen Hawking

1. Lung cancer

Lung cancer remains the leading cause of cancer-related deaths worldwide, representing one out of five cancer deaths. Although incidence rates are becoming to decrease, the average 5-year survival rate for those with the disease remains below 20%, irrespective of the subtype (Lovly and Carbone, 2011). Therefore, advances in basic and translational research are required to improve the outcome of this fatal disease.

1.1. Epidemiology

According to The International Agency for Research on Cancer (IARC) (https://gco.iarc.fr/), the global incidence of lung cancer in 2020 raised to nearly 2.2 million new cases, with a mortality rate that almost reached 1.8 million deaths, mostly in men (66% of the total).

Lung cancer not only presents an alarmingly high mortality rate but also exerts the lowest survival rate compared to other types of cancer. In particular, lung cancer survival rate is highly variable depending on the stage of the disease: while stage I diagnosis is associated with an average 1-year survival of around 89%, stage IV survival rate drops drastically to 21.3%.

The major risk of developing lung cancer is cigarette smoking, which is a complex mixture containing more than 60 carcinogens (like polycyclic aromatic hydrocarbons, nitrosamines, aldehydes, inorganic compounds, and free radicals) that directly cause mutations in DNA (Pfeifer et al., 2002). Smokers present up to a 30-fold increased risk of developing lung cancer compared to non-smokers (Walser et al., 2008). It has been postulated that women are more susceptible to developing lung cancer at the same exposure, due to the higher levels of DNA adducts caused by cigarette smoking (Rivera, 2013). In addition to this, smoking leads to a prominent microbial dysbiosis in the respiratory tract to degrade its harmful compounds (Shapiro et al., 2022).

Exposure to environmental carcinogens (e.g. arsenic, asbestos, radon, nickel, or vinyl chloride), air pollution (e.g. carbon monoxide, nitrogen dioxide aldehydes, diesel, or metals), respiratory illnesses (e.g. pneumonia, asthma, or chronic obstructive pulmonary disease), genetic alterations affecting DNA repair mechanisms, or cell cycle control can

also play an important role in the development of the disease (Akhtar and Bansal, 2017).

Although smoking is associated with the vast majority of lung cancer deaths, about 10% of the cases are patients who have never smoked, which suggests additional mechanisms contributing to tumorigenesis.

1.2. Classification

Lung cancer is a heterogeneous disease that can be divided into several subtypes with different molecular, histological features, and clinical outcomes. According to these criteria, lung cancers are currently classified into non-small cell lung cancer (NSCLC), which accounts for the majority of the cases (85%), and small cell lung cancer (SCLC), which only represents 15% of all cases (Bender, 2014).

Among them, SCLC is the most clearly associated subtype with smoking and is characterized by a high proliferative rate, early metastases, and poor prognosis (Rudin et al., 2021). It is thought to be derived from neuroendocrine cells or progenitors present in the lung epithelium and it usually forms near the bronchi (Karachaliou et al., 2016). This tumor type is one of the most aggressive forms of cancer, as most patients present metastatic lesions at the time of diagnosis, and its 5-year survival rate is very low (6%) compared to NSCLC (21%) (Bender, 2014).

NSCLC is further divided into three main pathological subtypes: (1) lung adenocarcinoma (LUAD, 40% of the cases), (2) lung squamous cell carcinoma (LUSC, 30% of the cases), and (3) large cell carcinoma (LCC, 15% of the cases).

LUAD is the most common subtype of lung cancer in both genders, with women having an even higher predominance, and is also the most frequent subtype in non-smokers. It is a malignant epithelial tumor that usually forms in the glands that secrete mucus in the alveoli and presents an acinar structure (Rodriguez-Canales et al., 2016). As LUADs express markers of differentiation, patients tend to have a better prognosis compared to other subtypes of lung cancer.

LUSC is more frequent in men than in women. It arises in cells located in the airway epithelium exposed to air pollution, which could provide an explanation for its strong association with smoking, considering that 96% of patients reported being smokers ("Comprehensive genomic characterization of squamous cell lung cancers," 2012). Tumor cells lack both glandular structure and mucin production and show keratinization and expression of markers of squamous cell differentiation.

LCC is very proliferative and tends to spread to the lymph nodes and distant sites. Tumors are characterized by a poorly differentiated state, are composed of large cells with abundant cytoplasm and large nucleoli, and are usually diagnosed ruling out other subtypes of lung cancer (Rodriguez-Canales et al., 2016).

The most common sites of lung metastasis include bones, other parts of the lungs, the brain, adrenal glands, and the liver (Popper, 2016).

1.3. Genetic alterations found in lung cancer

Lung cancer presents one of the highest rates of somatic mutations (Alexandrov et al., 2013), which challenges the identification of driver gene alterations due to a large number of passenger events.

In lung adenocarcinoma (LUAD), *TP53* (46%), *KRAS* (33%), and *EGFR* (14%) are the most commonly mutated genes, followed by *BRAF* (10%), *PI3KCA* (7%), and *MET* (7%). Mutations in *KRAS* and *EGFR* are mutually exclusive. LUADs also present mutations in tumor suppressor (*STK11, KEAP1, NF1, RB1,* and *CDKN2A*), and chromatin modifying genes (*SETD2, ARID1A,* and *SMARCA4*). Genomic rearrangements are also a common event and include amplifications (*TERT, KRAS*, *EGFR,* and *MET*), deletions (*CDKN2A*), and translocations (*ALK*, *RET,* and *ROS1*) (The Cancer Genome Atlas Research Network, 2014).

Genomic analyses of lung squamous cell carcinomas (LUSCs) have identified a high rate of copy number alterations, amplifications (*SOX2, PDGFRA, NFE2L2, EGFR, MYC, CDK6,* and *MDM2*), and deletions (*CDKN2A*, *FOXP1*, *PTEN,* and *NF1*). *TP53* is the mainly mutated gene (81%), followed by *CDKN2A*, *PTEN*, *PIK3CA*, *KEAP1*, *HLA-A*, *NOTCH1,* and *RB1* ("Comprehensive genomic characterization of squamous cell lung cancers," 2012).

Small-cell lung cancer (SCLC) is mainly driven by the concomitant inactivation of the tumor suppressor genes *TP53* (89%) and *RB1* (64%). Other functionally validated alterations include *MYC* and *FGFR1* amplifications and mutations affecting *PTEN*, *NOTCH* receptors, and chromatin remodeling genes (*KMT2D*, *CREBBP,* and *ARID1A*) (Rudin et al., 2021). This tumor type is genetically characterized by a high tumor mutation burden (TMB), chromosomal rearrangements, and intratumor heterogeneity that has been linked to metastases and resistance to therapies.

The effects of smoking have been extensively characterized in different types of human cancer and normal bronchial epithelium (Alexandrov et al., 2016; Yoshida et al., 2020). These studies revealed that tobacco exposure causes DNA adducts (covalent binding of activated reactive species of carcinogens to DNA bases) that finally lead to DNA damage, increasing copy number alterations, insertions, deletions, and substitutions, that mainly consist in cytosine to adenine (C>A) nucleotide transversions. Thanks to big consortium studies and the identification of different recurrent cancer mutational signatures (Alexandrov et al., 2016, 2013), it has been described that lung cancer patients that are smokers present a pattern of mutations associated with mutational signatures 2, 4, 5, and 13, which can be attributed to mistakes during the replication of damaged DNA, overactivation of APOBEC cytidine deaminases, and defects associated with aging.

When it comes to driver events of the disease, *KRAS* mutations occur more frequently in smokers and appear to confer a worse prognosis (Barta et al., 2019). Never-smokers present a higher prevalence of driver mutations in *EGFR* and *HER2*, as well as *ALK-EML4* rearrangements, and *ROS1* translocations (Bergethon et al., 2012; Govindan et al., 2012; Ou, 2013; Paik et al., 2012).

1.4. Molecular pathways altered in lung cancer

In LUAD, the most frequently altered signaling pathways are the ones initiated by tyrosine kinase receptors (RTKs), such as the MAPK/ERK and PI3K-mTOR pathways, which are activated in 76 and 25% of the cases, respectively (The Cancer Genome Atlas Research Network, 2014). *KRAS, EGFR, BRAF,* and *MET* are the most frequently activated genes, together with inactivating mutations in *STK11, NF1,* and *PTEN*. In addition to this, p53 pathway (*TP53* and *ATM*), cell cycle

regulators (*CDKN2A* and *RB1*), and nucleosome remodeling (*ARID1A, ARID1B, ARID2,* and *SMARCA5*) disruption are also common events in this tumor type.

Alterations found in LUSC reveal a deregulation of cell cycle control (*CDKN2A* and *RB1*), response to oxidative stress (*NFE2L2, KEAP1,* and *CUL3*), and squamous cell differentiation (*SOX2, TP63,* and *NOTCH*) ("Comprehensive genomic characterization of squamous cell lung cancers," 2012). Molecular pathways downstream RTKs are again the most affected, considering that key components of the PI3K/AKT and MAPK/ERK pathways are collectively altered in 69% of the cases. Therefore, in an attempt to exploit therapeutic vulnerabilities, the development of targeted therapies against RTK/KRAS/PI3K pathways seems to be a reasonable alternative to improve lung cancer treatment.

In SCLC, cell cycle progression, genomic instability, lineage plasticity, and neuroendocrine differentiation are key molecular pathways altered as a consequence of *TP53* and *RB1* loss (Rudin et al., 2021).

1.5. Diagnosis

Lung cancer diagnosis requires the extraction of cancer cells by different methodologies (e.g. sputum cytology, thoracentesis, or needle biopsy), the evaluation of their morphology under the microscope, and the performance of specific immunohistochemical assays (p63, p40, cytokeratin 5/6, TTF-1, napsin A, chromogranin A, synaptophysin, and CD56) (Inamura, 2018) to further identify the specific subtype of lung cancer.

Immunohistochemistry and/or targeted sequencing can be further applied to detect driver genetic alterations (*EGFR*, *BRAF, ALK*, *ROS1, MET,* and *RET*), which has led to the development of molecularly targeted therapies that have dramatically improved the outcome in selected subgroups of patients.

1.6. Treatment

Treatment of lung cancer patients is stage- and subtype-specific (Detterbeck et al., 2017). Histological characterization and the presence of specific genetic alterations dictate the final treatment that the patient receives, which can include surgery, radiation therapy, chemotherapy, immunotherapy, or molecularly targeted therapy, either alone or in combined modalities.

Different lung cancer surgeries are available depending on the size of the affected regions. Thus, sleeve resection consists of the removal of a small piece of the lung where the tumor is located; in a segmentectomy, part of the lobe is resected; in a lobectomy, the complete lobe contacting the tumor is removed; and finally, in a pneumonectomy, an entire lung is extracted. Nearby lymph nodes are frequently removed to avoid possible metastases.

Radiation therapy and chemotherapy are unspecific treatments based on the generation of DNA damage through radiation or the perturbance of cell cycle, respectively. Both strategies take advantage of the higher proliferation rate of cancer cells, but are also cytotoxic to the bone marrow and other highly-proliferative tissues. Chemotherapy includes alkylating agents (e.g. carboplatin, cisplatin, lurbinectedin, and cyclophosphamide), anti-tumor antibiotics (e.g. doxorubicin and neocarzinostatin), anti-metabolites (e.g. gemcitabine and pemetrexed), topoisomerase inhibitors (e.g. irinotecan, etoposide, and topotecan), and mitotic inhibitors (e.g. docetaxel, paclitaxel, albumin-bound paclitaxel (nab-paclitaxel), and cabazitaxel), among others.

Targeted therapy and immunotherapy are strategies specifically directed against tumor cells, which results in less frequent side effects. Molecular targeted therapy consists of the use of drugs against antigens specifically altered in cancer cells, whereas immunotherapy refers to the boost of the immune system of the patient to promote an anticancer response. Among the different kinds of targeted therapies developed, tyrosine kinase receptor (RTK), signal transduction, proteasome, or angiogenesis inhibitors can be highlighted. Immunotherapy includes many different strategies, such as checkpoint inhibitors (directed against PD-1, PD-L1, CTLA-4, or LAG-3), chimeric antigen receptor (CAR) T-cell therapy, cytokines (IL-2 and IFN-alfa), or immunomodulators (e.g. thalidomide, lenalidomide, pomalidomide, and BCG), to name a few. A complete list of the treatments used in lung cancer divided by their mechanism of action is shown in **Table 1.1.**

1.6.1. Treatment of early-stage NSCLC

Surgical resection is the preferred option in early-stage disease (Duma et al., 2019), when tumors are confined to the lungs (**Figure 1.1**). Cisplatin-based adjuvant chemotherapy remains the standard of care for patients with resected tumors at high risk of recurrence (Chaft et al., 2021). Stereotactic body radiation therapy, alone or in combination with chemotherapy (cisplatin and pemetrexed), is the option of choice when the tumors are not completely resectable or the patient is not medically operable. In addition to this, the use of anti-PD-L1 antibodies (durvalumab) improves overall survival (Chaft et al., 2021).

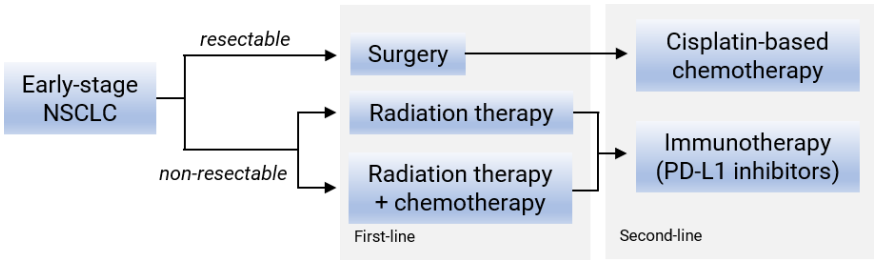


Figure 1.1. Management of early-stage NSCLC.

1.6.2. Treatment of metastatic NSCLC

Advanced-stage NSCLCs are classified according to their molecular features and genetic alterations (Alexander et al., 2020; Duma et al., 2019; Thomas et al., 2015):

- Squamous histological type is commonly treated with chemotherapeutic agents, alone or in combination with antibodies against PD-1 (platinum + paclitaxel or nab-paclitaxel + pembrolizumab). Second- and third-line treatments typically consist of chemotherapeutics, alone or in combination with RTK inhibitors (docetaxel + ramucirumab, gemcitabine, erlotinib, or necitumumab) (**Figure 1.2**).

- Non-squamous histological subtype is further classified depending on their driver alterations. Thus, specific treatments directed against *EGFR* (erlotinib, afatinib, or osimertinib), *BRAF* (dabrafenib and trametinib), or *NTRK* (entrectinib or

larotrectinib) mutations, *ALK/ROS1* rearrangements and *MET* mutations (crizotinib or alectinib), and finally *RET* rearrangements (selpercatinib or cabozantinib) can be considered.

Despite being the most frequently mutated oncogene in NSCLC, there are no available targeted treatments for *KRAS* p.G12C-mutated tumors, although sotorasib has shown promising results in phase 2 clinical trials (Skoulidis et al., 2021).

Non-squamous tumors with no actionable alterations commonly receive chemoimmunotherapy, a combination of platinum-based and PD-1/PD-L1 or VEGF antibodies (platinum/pemetrexed + pembrolizumab; or carboplatin+ taxane, atezolizumab + bevacizumab).
Second- and third-line treatments typically include chemotherapeutic agents or RTK inhibitors (docetaxel + ramucirumab, gemcitabine, or necitumumab).

One of the main issues concerning targeted therapeutics is the appearance of resistance to pharmacological inhibition of oncogenic drivers (Thomas et al., 2015). Therefore, genomic analysis of serial biopsies is essential to understand the evolution of drug-resistance mechanisms and to enable the selection of the most appropriate therapy.

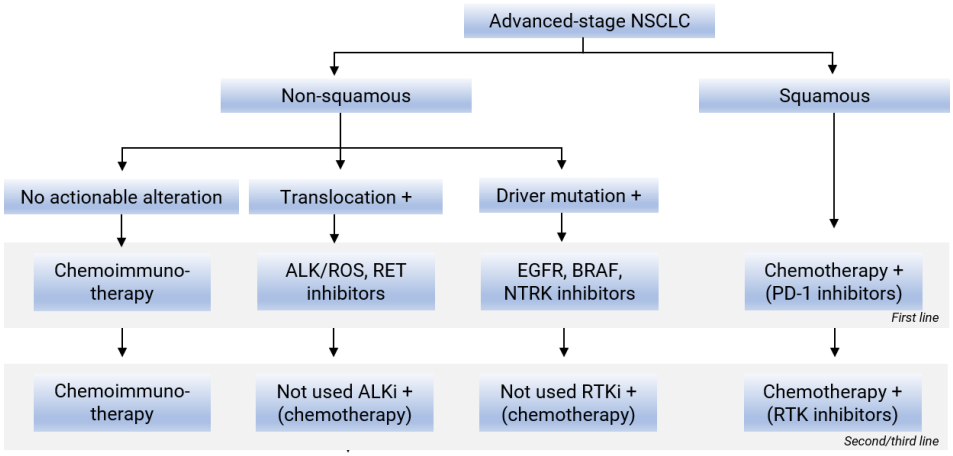


Figure 1.2. Management of advanced-stage NSCLC.

1.6.3. Treatment of SCLC

SCLCs are also treated differently depending on the stage of the disease: (1) stage I tumors are surgically resected or treated with radiotherapy with an additional adjuvant chemotherapy (cisplatin and etoposide) (**Figure 1.3**); (2) stage II and III tumors are treated with chemoradiotherapy, a combination of thoracic radiotherapy and chemotherapeutics (cisplatin and etoposide), which can be followed by irradiation in responding patients; (3) and stage IV tumors receive chemotherapy (platinum agent (cisplatin or carboplatin) together with etoposide), alone or in combination with PD-L1 inhibitors (atezolizumab or durvalumab). Second-line treatments for recurrent patients include a topoisomerase I inhibitor (topotecan) or an alkylating agent (lurbinectedin). Finally, third-line treatments consist of PD-1 inhibitors (pembrolizumab or nivolumab) (Rudin et al., 2021).

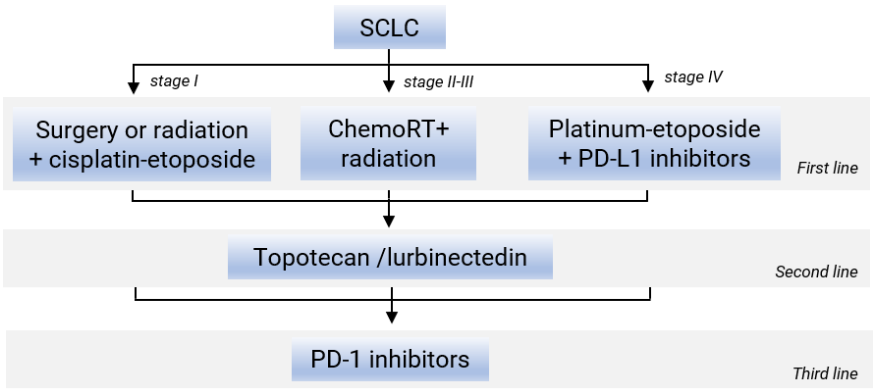


Figure 1.3. Management of SCLC. ChemoRT: chemoradiotherapy.

Treatment	Description
Chemotherapy	Platinum-based (cisplatin, carboplatin, oxaliplatin, nedaplatin, or lobaplatin), etoposide, pemetrexed, taxanes (paclitaxel, nab-paclitaxel, docetaxel), gemcitabine, topotecan, lurbinectedin.
Targeted therapy	Erlotinib, gefitnib, afatinib, dacomitinib, or osimertinib (EGFRi), necitumumab or cetuximab (α-EGFR), dabrafenib (BRAFi), trametinib (MEKi), entrectinib or larotrectinib (NTRKi), crizotinib, ceritinib or alectinib (ALKi), selpercatinib or cabozantinib (RETi).
Immunotherapy	Durvalumab or atezolizumab (α-PD-L1), pembrolizumab or nivolumab (α-PD-1), ramucirumab (α-VEGFR2), bevacizumab (α-VEGF).

Table 1.1. Description of the currently used treatments in lung cancer, classified according to their mechanism of action. Abbreviations: i, inhibitor; α, anti.

2. Gene expression regulation by chromatin structure

Gene expression is initiated by the binding of specific transcription factors and RNA polymerase to the genomic sequence immediately located around the transcription start site (core promoter). In some cases, the expression of the target gene is further activated by the physical interaction between promoters and distal sequence elements (enhancers) that are brought into proximity by chromatin looping. Nevertheless, the function of these regulatory sequences significantly relies on a permissive chromatin structure characterized by DNA accessibility, which allows the binding of specific transcription factors, cofactors, and regulators, as well as the interactions between DNA elements. Moreover, nucleosomes in the proximity of active genes and enhancers contain histones with characteristic post-translational modifications that can further activate or repress transcription. Therefore, chromatin structure plays essential roles in gene expression regulation.

Each human cell contains approximately 2 m of DNA confined in a few micrometric nuclear compartment. Thus, DNA is tightly packed into chromatin, a dynamic structure composed of basic units known as nucleosomes, whose density and position sterically determine the ability of the cellular machinery to access the genome. Histones are the central component of nucleosomal core particles, forming an octamer containing four core histone proteins (two H3-H4 dimers and one H2A-H2B tetramer) around which a fragment of 147-bp is wrapped twice. Histone linker H1 is bound to the outside of each nucleosome with a 20-90 bp of linker DNA for its stabilization and connection to the adjacent nucleosomes.

There are two types of chromatin depending on its level of compaction: heterochromatin and euchromatin. Heterochromatin is highly condensed, transcriptionally silent, and contains many repetitive DNA sequences, whereas euchromatin has a less compact structure, being consequently more accessible to transcription, and mainly comprises the coding regions of genes. Both types of chromatin are characterized by different histone modifications. Heterochromatin tends to be enriched in H3 histone molecules trimethylated on lysine 27 (H3K27me3), as well as di- or trimethylated on lysine 9 (H3K9me2/3); whereas euchromatin shows histone H3 molecules monomethylated on lysine 4 (H3K4me1), together with acetylation of several other lysine residues (van Steensel and Furlong, 2019).

DNA accessibility is mediated by two classes of enzymes: ATP-dependent nucleosome remodelers (which will be extensively described in chapter 3: “SWI/SNF chromatin remodeling complexes”) and histone modifying enzymes (which will be addressed in this section).

2.1. Histone post-translational modifications

Histones are highly conserved basic proteins with an N-terminal unstructured and flexible tail enriched in lysine (K) and arginine (R) residues that project from the nucleosome. These tails are accessible on their surface and therefore subjected to covalent post-translational modifications that coordinately determine chromatin accessibility, and hence the biological outcome. Most histone modifications are reversible but can be maintained over cellular divisions.

Histone modifications can alter the electrostatic interaction between the positively-charged residues of the histones and the negatively-charged pyrophosphate groups from the DNA, modifying chromatin compaction. In addition to this, histone modifications can alter the structure of histone tails and they can also serve as recognition modules for specific binding proteins. As a result, different cofactors, regulators, and transcription factors are able to bind DNA, allowing gene expression, DNA replication, or DNA repair, but also the maintenance of a repressive chromatin state.

The diverse set of histone modifications includes acetylation, phosphorylation, methylation, ubiquitination, SUMOylation, adenosine diphosphate (ADP) ribosylation, O-GlcNAcylation, deimination/ citrullination, N6-formylation, propionylation, succinylation, butyrylation, serotonylation, and crotonylation (Audia and Campbell, 2016; Fang and Han, 2021). Among them, methylation and acetylation epigenetic marks play an important role in transcriptional regulation.

Generally, the nomenclature of these modifications includes:

(1) The modified histone (H1, H2A, H2B, H3, or H4). Histone 3 (H3) is the most frequently modified histone.

(2) The amino acid residue affected (K, R, S, T, Y, H, or E) and its locations in the primary sequence of the protein. The majority of histone modifications occur in lysine residues (e.g. K4, K9, or K27 in histone H3).

(3) The type of post-translational modification (acetylation, methylation, phosphorylation, ubiquitination, and so on). Acetylation (ac) is typically present in a monomeric form, whereas lysine or arginine residues can be mono-, di-, or tri-methylated (me1, me2, or me3, respectively).

A growing body of evidence suggests that distinct histone modifications can influence the deposition of other epigenetic marks to regulate the biological outcome. For instance, histone H3 trimethylated on lysine 79 (H3K79me3) and lysine 4 (H3K4me3) residues are present at coding regions and co-localizes on active chromatin (Nakanishi et al., 2009); histone H3 monomethylated on lysine 79 (H3K79me1) is found on genomic regions that are also enriched in H3K4me3 and H3K27me3 (Steger et al., 2008); and histone H3 acetylated on lysine 122 (H3K122ac) co-localizes with many marks characteristics of active genes (H3K4me3) and active enhancers (H3K4me1 and H3K27ac) (Tropberger et al., 2013).

2.1.1. Writers, readers, and erasers

There is a plethora of enzymes responsible for placing, recognizing, and removing histone modifications (**Table 1.2**), and are generally described as epigenetic modifiers of the chromatin. These epigenetic players exert key roles in the intricated regulation of gene expression and can be categorized as writers (introduce chemical modification on DNA and histones), readers (contain specific domains to identify post-translational modifications), and erasers (responsible for the removal of these chemical groups) (Biswas and Rao, 2018; Schuettengruber et al., 2017).

Modification	Significance	Location	Writer	Eraser	Reader
H3K4me1	Transcriptional activation	Enhancers	MLL1-5	KDM1A/B	MLL, CHD1, BPTF, RAG2
H3K4me3	Active euchromatin	Promoters	SETD1A/D	KDM5A/B/C	ING, KDM5, TAF3
H3K27ac	Transcriptional activation	Enhancers	CBP/P300	HDACs/ SIRTs	BRD4
H3K27me3	Transcriptional silencing	Enhancers and promoters	EZH2, EZH1	KDM6A/B	EED, PC

Table 1.2. Most studied post-translational modifications of human histone H3. The biological effect of each modification, its main location in the genome, and the enzymes responsible for its deposition (writers), removal (erasers), and

binding (readers) are indicated. Adapted from *Audia and Campbell, 2016; Lawrence et al., 2016.*

Recurrent somatic mutations affecting the protein machinery in charge of writing, reading, and erasing histone marks have been reported in several types of cancer. Indeed, some of these mutations are considered oncogenic drivers that contribute to tumor progression (Nacev et al., 2019).

2.1.2. Histone methylation

Histone methylation can occur at lysine or arginine residues of histones H3 and H4 and include mono-, di- or tri-methylation (me1, me2, or me3, respectively). Histones methyltransferases (HMTs) use their SET domain to transfer methyl groups from the substrate S-adenosyl methionine (SAM) onto the lysine residues of histones. The removal of these methyl groups is catalyzed by demethylases. Chromo, Tudor, MBT, PHD, and PWWP domains are responsible for the recognition of methylated lysine residues (Kadoch et al., 2016).

Different grades of methylation of histone H3 on lysine 4 (H3K4), lysine 36 (H3K36), and lysine 79 (H3K79) residues are present in actively transcribed genes, whereas methylation of lysine 9 (H3K9), lysine 27 (H3K27), and lysine 20 (H3K20) residues are associated with transcriptional repression and heterochromatin formation (Fang and Han, 2021).

Histones that flank active enhancers are often marked by histone H3 lysine 4 mono-methylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac). Inactive or poised enhancers bear the active H3K4me1 and the repressive mark histone H3 trimethylated on lysine 27 (H3K27me3). Enhancers that are not yet active but are primed for activation at a later development point or in response to an external stimulus are pre-marked by H3K4me1 (**Figure 1.4**). If these enhancers are bound by repressive transcription factors, they are considered repressed enhancers (Bozek and Gompel, 2020).

H3K4 di-methylation (H3K4me2) marks the 5’ end of transcribed genes but is also present in enhancers (Alver et al., 2017). H3K4 tri-methylation (H3K4me3) is mainly distributed at promoters of actively transcribed genes (Fang and Han, 2021). In addition to this, some gene promoters are bivalently marked by both H3K4me3 and H3K27me3

(poised promoters). Consequently, these genes are silent, but their promoters are protected against *de novo* DNA methylation and ready to be activated in the proper developmental context (Greenberg, 2021).

When it comes to repressive marks, H3K27me3 is considered a hallmark of gene repression. The addition of di- and trimethyl groups (H3K27me2/3) is catalyzed by the EZH2 subunit of the PRC2 complex. This epigenetic modification is present in broad domains of the chromatin at the promoters of silenced genes but is also enriched at poised enhancers. H3K27me3 epigenetic modification cannot co-occur with H3K27ac on the same histone (Shlyueva et al., 2014).

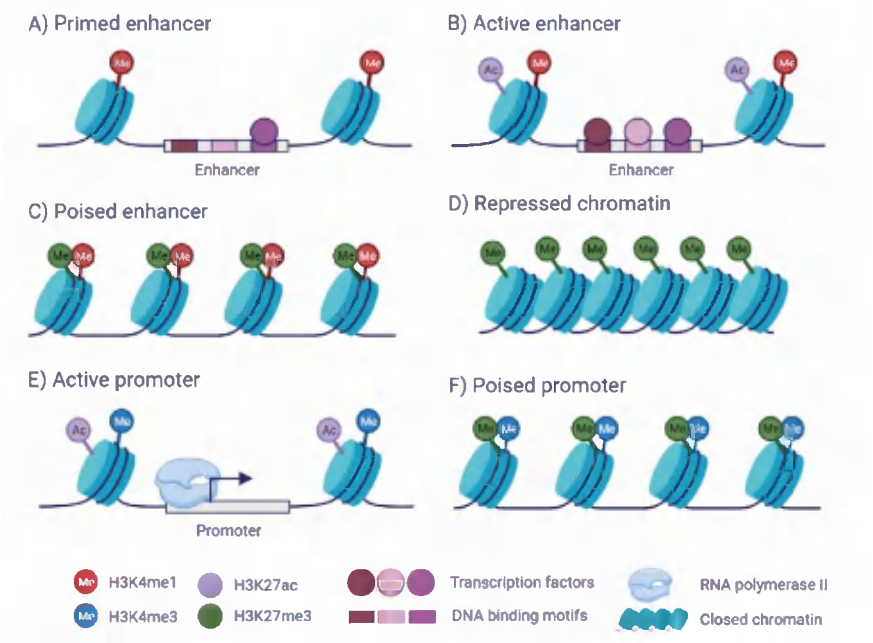


Figure 1.4. Histone post-translational modifications at regulatory elements. **A)** Enhancers primed for activation are pre-marked by H3K4me1. **B)** Histones that flank active enhancers are marked by H3K4me1 and H3K27ac. **C)** Poised enhancers bear the active H3K4me1 and the repressive H3K27me3 marks. **D)** Nucleosomes that contain the H3K27me3 modification are associated with transcriptional repression. **E)** Active promoters are transcribed by RNA Pol II and their adjacent nucleosomes are characterized by H3K4me3 and H3K27ac deposition. **F)** Poised promoters are defined by the simultaneous presence of histone modifications H3K4me3 and H3K27me3, associated with gene activation and repression, respectively.

SETD1A/COMPASS catalyzes the mono-, di- and tri-methylation of H3K4me3 at active promoters. Similarly, MLL1 is required for H3K4me3 deposition at specific genes, whereas MLL2 is responsible for H3K4me3 at bivalent promoters in ESCs. MLL3/4 are the main methyltransferases mediating H3K4me1 at enhancers. On the other hand, histone demethylases KDM1, KDM2A, and KDM5 are involved in the removal of methyl groups on H3K4, whereas KDM6A and KDM6B are able to remove the repressive H3K27me3 mark introduced by the PRC2 complex (Schuettengruber et al., 2017).

2.1.3. Histone acetylation

Histone acetylation is tightly regulated by the balance between histone acetyltransferases (HATs) and histone deacetylases (HDACs). Acetylation can influence the compaction state of chromatin by neutralizing the positive charge of the lysine residues, impairing the binding between histone tails and negatively charged DNA. Thus, histone acetylation is usually correlated with an open chromatin structure and active transcription.

The bromodomain and extra-terminal domain (BET) proteins (BRD1, BRD2, BRD3, BRD4, and BRDT) can recognize histone acetylated lysine residues through their bromodomains, acting as scaffolds to recruit transcription factors and the RNA polymerase, allowing gene expression (Audia and Campbell, 2016; Fang and Han, 2021). Other bromodomain-containing proteins with different biological functions (e.g. histone-modifying writers p300 and MLL, or chromatin remodeler subunits SMARCA2, SMARCA4, PBRM1, BRD7, and BRD9) have been described.

There are plenty of lysine residues that can be acetylated in histones (H3K4, H3K9, H3K14, H3K18, H3K23, H3K27, H3K36, H4K5, H4K8, H4K12, H4K16, or H4K20). When looking at its primary sequence, is quite striking that acetylatable lysine residues are placed regularly along the sequence at a distant coincident with the length of an α-helix (Strahl and Allis, 2000).

Histone H3 lysine 27 acetylation (H3K27ac), which is the most well-studied residue, is present at the promoters and enhancers of actively transcribed genes (Alver et al., 2017), where it colocalizes with H3K4me3 and H3K4me1, respectively. H3K27ac can form large broad

domains in the intergenic regions forming super-enhancers to further promote gene expression.

SMARCB1 and SMARCA4 subunits of the SWI/SNF complexes are able to recruit p300/CBP histone acetyltransferase to enhancers, contributing to the modulation of H3K27ac levels (Alver et al., 2017). On the other side, polycomb can inhibit the acetyltransferase activity of this enzyme, favoring H3K27 methylation and transcriptional repression (Schuettengruber et al., 2017).

2.1.4. Other histone modifications

Histone lysine residues can also be post-translational modified to take part in DNA repair mechanisms. Thus, in DNA damage conditions, PARP1 catalyzes the addition of an ADP-ribose group to lysine residues of histones. As a result of this action, chromatin becomes more accessible and the DNA repair machinery can be recruited (Fang and Han, 2021). Another example is the mono-ubiquitination at residues K119/K120 of H2AX mediated by the PRC1 complex. This modification is required for the recruitment of early sensors of DNA damage, such as the ATM kinase, which is the enzyme responsible for the phosphorylation of histone variant H2AX to form γH2AX (Panier and Durocher, 2013).

Similar to histone acetylation, histone phosphorylation affects the binding of DNA and histones by adding a negative charge, which will probably cause the decondensation of chromatin. Concerning this, phosphorylation of histone H3 on serine 10 (H3S10P) and serine 28 (H3S28P) residues is required for proper segregation and condensation of chromosomes during cell division (Hirota et al., 2005). Moreover, phosphorylation of histone H3 on threonine 118 (H3T118P) results in enhanced nucleosome mobility and accessibility, allowing nucleosome disassembly by the SWI/SNF complex (North et al., 2011).

2.2. Three-dimensional chromatin structure

Genome-wide mapping of DNA-binding proteins, histone modifications, and chromosome conformation capture technologies have shown that chromatin is hierarchically organized in higher-order structures separated by insulating boundaries (regions of low interactions). In vertebrates, insulators are typically bound by CTCF, which together with the cohesin complex can form long-range chromatin loops that delimitate the boundaries of regulatory domains and limit enhancer activity (Dixon et al., 2012) (**Figure 1.5**).

Thus, nucleosome fibers fold into chromatin loops, giving rise to topologically associated domains (TADs), which serve as functional platforms for the physical interactions between regulatory elements (Schuettengruber et al., 2017; Zhang et al., 2021). Intradomain chromatin contacts are much more frequent than contacts with regions outside these domains. TADs and their boundaries are largely conserved between cell types and species (Robson et al., 2019). During embryonic development, TADs acquire a more structured architecture as they become involved in transcriptional regulation mediating enhancer-promoter contacts (Andrey and Mundlos, 2017).

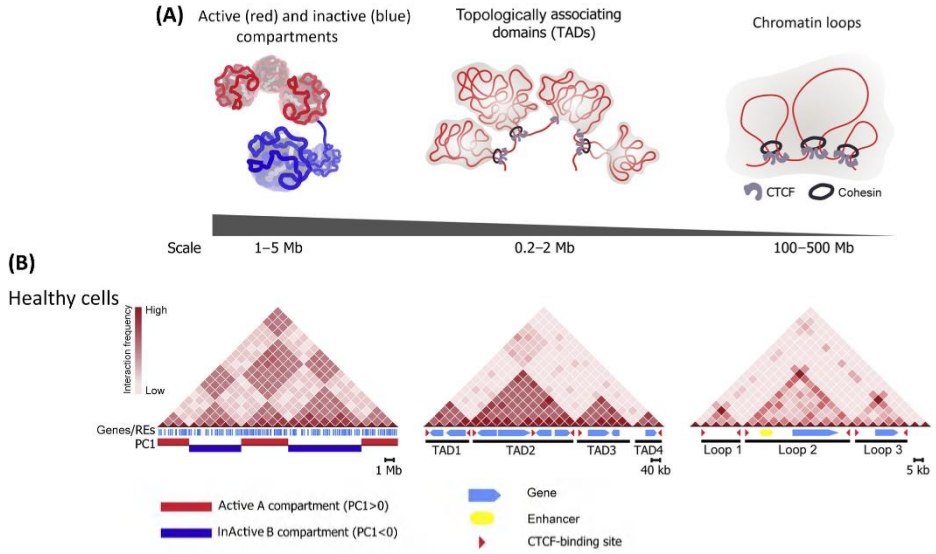


Figure 1.5. Three-dimensional organization of chromatin. A) Chromatin is organized into transcriptionally active (red) or inactive (blue) compartments. In each region, chromatin is folded into topologically associating domains (TADs), delimited by CTCF and cohesin binding. **B)** The Hi-C map reflects the average interaction frequency between regulatory elements (genes and enhancers). Figure modified from *Kantidze et al., 2020*.

Therefore, TADs allow the physical interaction among enhancers and promoters that can be located far away on the nucleotide sequence. However, enhancers do not promiscuously activate the expression of all genes present within a TAD (Ghavi-Helm et al., 2019). It is thought that this relationship is mediated by the sequence composition of core promoters and distal genetic elements (Pachano et al., 2021).

Promoters are regulatory regions located at the transcription start site of a gene, whereas enhancers are typically located at large genomic distances and can be found either upstream or downstream of the target gene promoter. Enhancers contain binding sites for regulatory proteins (such as, transcription factors, mediator complex, cohesin complex, or chromatin remodelers) and utilize chromatin folding to bypass these large distances, and facilitate the activation of their target genes by physical proximity. Three types of active enhancers have been described: (1) classical, (2) closed chromatin, and (3) chromatin dependent (Sahu et al., 2022).

This historical distinction between enhancers and promoters is not able to capture all its complexity, as both regulatory elements share chromatin and sequence architectural features (such as, the H3K27ac deposition, they are located in nucleosome-devoid chromatin, and are actively transcribed by RNA polymerase II (Sahu et al., 2022)). Regarding this, several promoters have enhancer activity, while active enhancers are also able to drive local transcription at their boundaries (Andersson and Sandelin, 2020).

One of the main differences between enhancers and promoters is their GC content. Approximately 70% of gene promoters overlap with CpG islands (CGIs), which provides them with a permissive chromatin state that facilitates transcription initiation; whereas almost no enhancer does. It is particularly remarkable that in mouse embryonic stem cells, poised enhancers that present CGIs promote the physical and functional communication with distally located genes within the same TAD, particularly those with large CGI clusters in their promoters (Pachano et al., 2021).

Super-enhancers are a subclass of enhancers that consist of clusters of highly active enhancers, which present high levels of transcriptional coactivators binding, especially Mediator, and are involved in the control of master regulators of cell identity, (Pott and Lieb, 2015; Whyte et al., 2013). They usually contain several chromatin-dependent elements associated with a classical enhancer (Sahu et al., 2022).

Finally, interactions between TADs with similar epigenetic marks further spatially segregate chromatin according to its activity (Robson et al., 2019). In line with this, the human genome can be divided into either A compartments of actively transcribed genes and active histone modifications and B compartments characterized by lower transcription and chromatin repressive marks (Beagan and Phillips-Cremins, 2020).

3. SWI/SNF chromatin remodeling complexes

Chromatin remodeling complexes are key components in the modulation of genomic architecture. They use the energy provided by ATP hydrolysis to transiently disrupt nucleosome-DNA contacts, move nucleosomes along the DNA, and catalyze its ejection, insertion, or exchange (Gonzalez-Perez et al., 2013; Wilson and Roberts, 2011). Consequently, they modify the accessibility of specific regions of DNA to the enzymatic transcriptional machinery, as well as different DNA-binding proteins, cofactors, and regulators, playing important roles in chromatin assembly, gene expression regulation, cell differentiation, and DNA repair.

According to their subunit composition and biochemical activity, chromatin remodeling complexes can be divided into four major families: SWI/SNF, INO80, ISWI, and CHD, being SWI/SNF the most clearly implicated in tumor development (Gonzalez-Perez et al., 2013; Wilson and Roberts, 2011).

3.1. SWI/SNF subfamilies

SWI/SNF was firstly described in *Saccharomyces cerevisiae* after the performance of two genetic screenings aimed at identifying mutations in genes affecting the mating-type switching (SWI) and the sucrose-fermentation (SNF) pathways (L et al., 2010; Wilson and Roberts, 2011).

SWI/SNF complexes are complicated macromolecular assemblies of approximately 1-1.5 MDa consisting of many diverse and variable subunits. They are evolutionarily conserved from yeast to mammals (Wilson and Roberts, 2011). Mammalian SWI/SNF complexes are currently divided into three broad subfamilies: canonical BRG1-associated factor (cBAF); polybromo BGR1-associated factor (PBAF), and the recently described non-canonical BRG1-associated factor (ncBAF, also known as GBAF) (Mashtalir et al., 2018) (**Figure 1.6**). Each subfamily contains over ten subunits which are combinatorially assembled from the product of twenty-nine genes (Michel et al., 2018).

All complexes contain a group of the so-called core subunits, which are highly conserved between species and include SMARCC1/BAF155, SMARCC2/BAF170, SMARCD1/2/3, ACTB, ACTL6A/B, BCL7A/B/C, and either one of the two mutually exclusive ATPase subunits (SMARCA2/BRM or SMARCA4/BRG1). Core subunits SMARCC1, SMARCC2, and SMARCD1/2/3 are essential for proper complex assembly (X. Wang et al., 2019). SWI/SNF complexes also contain numerous accessory subunits that provide each of the subcomplexes with a distinct identity. Thus, SMARCB1/BAF47/SNF5 and SMARCE1/BAF57 are common to cBAF and PBAF families, whereas ARID1A and ARID1B are specific to cBAF family and ARID2, PBRM1, and BRD7 are specific to PBAF. Finally, ncBAF complexes utilize instead BRD9, GLTSCR1, and GLTSCR1L subunits for their core assembly (Mashtalir et al., 2018; Michel et al., 2018; Mittal and Roberts, 2020; Tsuda et al., 2021; Wilson and Roberts, 2011). Accessory subunits are thought to contribute to the targeting, the assembly, and the regulation of lineage-specific gene networks (Mittal and Roberts, 2020; Wilson and Roberts, 2011).

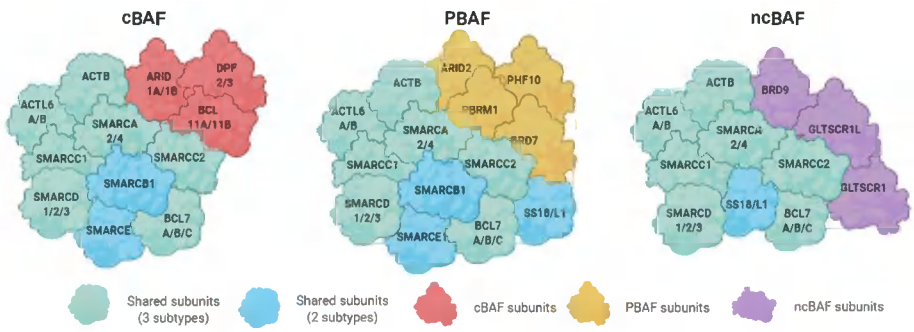


Figure 1.6. Mammalian SWI/SNF complexes. Shared subunits among all complexes are represented in green, whereas common subunits of either two of the subfamilies appear in blue. Accessory subunits specific to each subfamily are depicted in different colors: red (cBAF), yellow (PBAF), and purple (ncBAF).

Some common subunits of the SWI/SNF complexes are encoded by genes that produce different isoforms by alternative splicing. Moreover, they belong to gene families that often display differential lineage-restricted expression, such as those selectively incorporated in a certain type of neurons (Lessard et al., 2007) or embryonic stem cells (L et al., 2009). It is therefore likely that several

hundred versions of SWI/SNF complexes exist in mammals, which in turn control distinct gene expression programs in different cellular contexts.

This regulation is even more intricate taking into consideration that each subfamily presents different patterns of remodeling activity. Furthermore, alterations of nucleosome structure affect not only their recruitment but also their activity (Mashtalir et al., 2021).

The catalytic subunits of the complex contain DEXDc and HELICc domains that are responsible for its ATPase activity, but they also have bromodomains, which are also present in other subunits (BRD7, BRD9, and PBRM1) to recognize acetylated lysine residues (**Figure 1.7**). In addition to this, other domains present in the accessory subunits are thought to facilitate interactions with proteins (LXXLL, BAH, SANT, SWIRM, SWIB), DNA (ARID, HMG, Zn finger, Leucine zipper), and post-translational modifications present on histone tails (bromodomain, chromodomain, and PHD domains) (Kadoch et al., 2016).

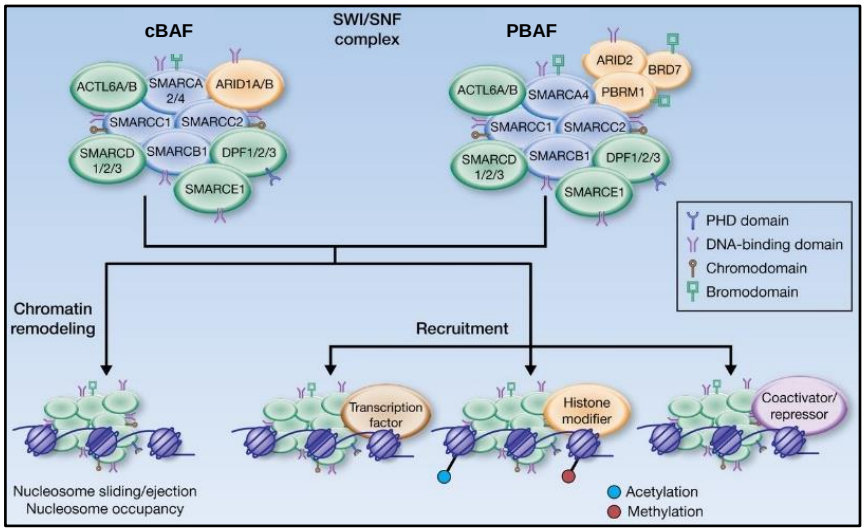


Figure 1.7. Protein domains present in cBAF and PBAF complexes. Core subunits are shown in blue, subunits encoded by multigene families appear in green, and the accessory subunits that differentiate BAF from PBAF are shown in orange. Figure modified from *Wang et al., 2014*.

3.2. Influence of SWI/SNF in transcriptional regulation and development

Mammalian SWI/SNF complexes exert general roles in transcriptional regulation. Under physiological conditions, they are involved in the differentiation of many cellular lineages. Regarding this, essential roles for these complexes have been identified during neurogenesis, myogenesis, adipogenesis, osteogenesis, and hematopoiesis (Mittal and Roberts, 2020). Thus, it is likely that they cooperate with tissue-specific transcription factors to coordinately balance the activation of lineage-specific genes, the promotion of differentiation, and the suppression of proliferative programs in such different cellular contexts (Wilson and Roberts, 2011).

The precise mechanism behind the role of SWI/SNF complexes in transcriptional regulation is not fully understood. Recent reports indicate that they are highly enriched at enhancers, where they cooperate to establish an open chromatin state required for transcription and oppose to the inactivation mediated by polycomb repressive complexes (Nakayama et al., 2017; Wang et al., 2017). It is speculated that cBAF activity occurs mainly at enhancers, whereas PBAF and ncBAF tend to be more enriched at promoters (L et al., 2009; Lessard et al., 2007; Mashtalir et al., 2021; Wang et al., 2017).

Enhancer dysregulation is a common feature in SWI/SNF-mutated tumors and has been linked with the promotion of stemness and self-renewal and the blockage of development and differentiation programs (Jones et al., 2022).

About this, in a rhabdoid tumor cell model, loss of *SMARCB1* alters SWI/SNF complex integrity and its targeting to enhancers implied in cell differentiation, while keeping SWI/SNF binding to super-enhancers required for tumor survival and cell plasticity (Wang et al., 2017). In agreement with this observation, as a result of *SMARCB1* re-expression in *SMARCB1*-deficient sarcoma cell lines, there is an increase in genome-wide BAF complex occupancy, which mediates enhancer activation and opposes to the polycomb-mediated repression at bivalent promoters (Nakayama et al., 2017).

Similar studies have been done in colorectal and liver cancer, where *ARID1A* disruption leads to the dysregulation of enhancers that govern cell identity (Guo et al., 2022; Mathur et al., 2017). In addition to this, increased activity at super-enhancers responsible for invasion properties has been reported in endometrial malignancies (Wilson et al., 2020).

3.3. Antagonistic role between SWI/SNF and PRC complexes

SWI/SNF complexes have a direct role in antagonizing the transcriptional silencing mediated by polycomb repressive complexes 1 and 2 (PRC1 and PRC2). Whereas proteins from the polycomb group are frequently overexpressed in cancer, SWI/SNF subunits are usually inactivated (Wilson et al., 2010). In addition to this, tumors harboring SWI/SNF inactivating mutations are particularly vulnerable to genetic and pharmacological inhibition of the EZH2 histone methyltransferase subunit of the PRC2 complex (Januario et al., 2017; Lu and Allis, 2017; Wilson et al., 2010).

SWI/SNF activity can occur at gene promoters, distal lineage-specific enhancers, and super-enhancers, where it leads to an open state of the chromatin associated with active transcription (Alver et al., 2017) (**Figure 1.8**). Alterations in its composition can cause defective complex assembly, failure to oppose PRC1 and PRC2 at promoters and typical enhancers of genes involved in lineage specification, while maintaining its attachment to super-enhancers of genes important to maintain cell identity (Alver et al., 2017; Katherine C Helming et al., 2014; Stanton et al., 2017; Wang et al., 2017; Wilson et al., 2010). The resulting imbalance between differentiation and self-renewal promotes tumorigenesis.

In addition to this, oncogenic SWI/SNF complexes containing SS18-SSX1 fusion protein (a common alteration found in human synovial sarcoma) are mistargeted genome-wide to polycomb target sites (Kadoch and Crabtree, 2013), and directly evict PRC1 complex from the chromatin (Kadoch et al., 2017). Finally, it has been proved by coimmunoprecipitation that SWI/SNF and PRC1 complexes interact in a chromatin-independent manner (Stanton et al., 2017).

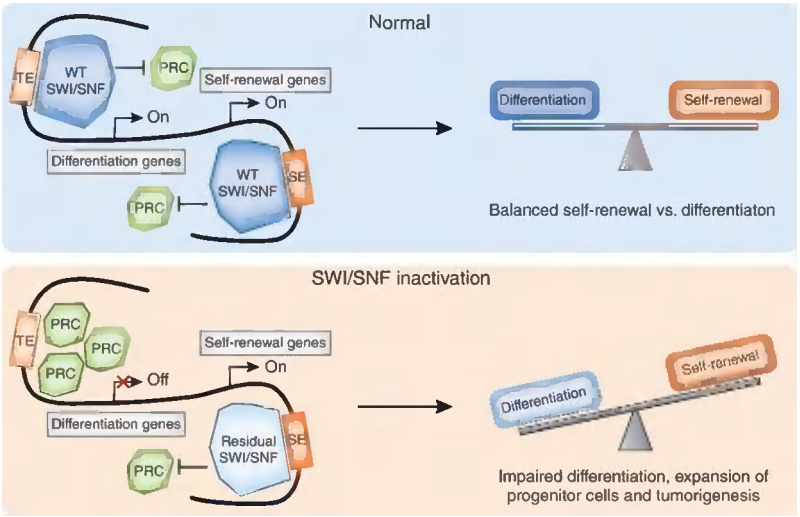


Figure 1.8. The antagonistic role between SWI/SNF and PRC complexes. Abbreviations: TE, typical enhancers; SE, super-enhancers. Figure source *Lu and Allis, 2017*.

3.4. Role of SWI/SNF in human cancer

Large genome sequencing studies have evidenced a prominent role of chromatin structure in cancer development. Interestingly, mutations in specific SWI/SNF genes are enriched in particular cancer types, whereas other subunits are broadly mutated (**Figure 1.9**).

The catalytic and accessory subunits of SWI/SNF complexes have been found recurrently mutated in several tumor types and their alteration has been associated with tumor progression (Kadoch et al., 2013; Mittal and Roberts, 2020; Moreno et al., 2021; Nakayama et al., 2017; Reisman et al., 2003; Wang et al., 2017), whereas core components in general present a very low mutation rate. One significant exception to this is *SMARCB1*, which is inactivated via biallelic mutations in 98% of malignant rhabdoid tumors (MRTs) (rare and aggressive childhood cancers that appear in the soft tissue of the brain) (Kadoch and Crabtree, 2015; Versteege et al., 1998).

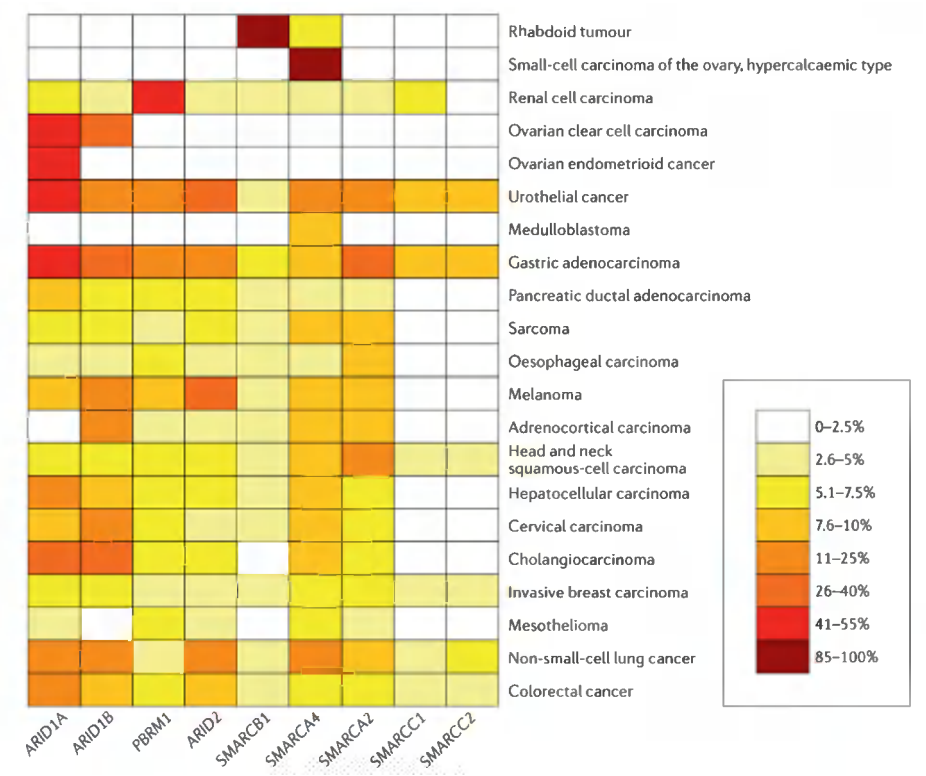


Figure 1.9. Distribution of SWI/SNF subunit mutations across human cancers. Frequency and distribution of non-synonymous mutations and deletions affecting the nine subunits of the complex most commonly mutated in cancer. Figure source *Mittal and Roberts, 2020*.

Regarding accessory and catalytic subunits, *PBRM1* is inactivated in 41% of clear cell renal cell carcinoma patients (Gerlinger et al., 2014; Varela et al., 2011), where mutations in *KDM6A* and *KDM5C* (histone demethylases), as well as *SETD2* (histone methylase) genes have also been identified but less frequently (Shain and Pollack, 2013; Varela et al., 2011). These findings suggest key roles of chromatin structure in the development of this tumor type.

ARID1A is altered in 50% of ovarian clear cell carcinomas (OCCCs, one of the most lethal subtypes of ovarian cancer), in 35% of endometroid carcinomas (Jones et al., 2010; Wiegand et al., 2010), in 9.4% of colorectal and in 8.2% of lung cancers (Kadoch et al., 2013). Other mechanisms that lead to *ARID1A* loss apart from loss-of-function mutations have been reported, such as copy number alterations (in 47%

of pancreatic cancer (Shain et al., 2012)) and epigenetic silencing of its promoter (in almost half of breast cancer patients (Zhang et al., 2013)).

It is estimated that more than 90% of the cases of an ovarian cancer subtype called small-cell carcinoma of the ovary hypercalcemic type (SCCOHT) harbor biallelic inactivating mutations in *SMARCA4* (Lu and Shi, 2019). *SMARCA2* is epigenetically silenced or transcriptionally inactivated in rhabdoid tumors (Kahali et al., 2014) and SCCOHTs (Karnezis et al., 2016). In non-small cell lung cancer (NSCLC) SMARCA2 and SMARCA4 expression is frequently lost, which is in turn associated with a worse prognosis (Herpel et al., 2017; Reisman et al., 2003). SMARCA2 expression is as well reduced in prostate cancer, where its absence correlates with advanced stages of disease progression and poor prognosis (Reisman et al., 2009).

Another accessory subunit of cBAF complexes, *ARID1B*, has been found mutated in a variety of tumor types, including melanoma, neuroblastoma, ovarian clear cell carcinoma, gastric cancer, and non-small cell lung cancer (Kadoch et al., 2013; Shain and Pollack, 2013).

Lastly, *ARID2* has been found inactivated in 18.2% of hepatocellular carcinomas (Li et al., 2011), and is also a significantly mutated gene in melanoma (Hodis et al., 2012).

Overall, genes encoding subunits of the SWI/SNF chromatin remodeling complexes are collectively mutated in almost 25% of all human cancers, which places SWI/SNF as the second most frequent alteration in cancer, just after *TP53* mutations (Kadoch et al., 2013; Mittal and Roberts, 2020; Shain and Pollack, 2013).

Intriguingly, although the tumor-suppressive mechanism of the complex might be general, the alteration of each subunit tends to be manifested in an apparently tissue-specific manner (Kadoch and Crabtree, 2015), which might indicate tissue-specific functions of each subunit.

Among the different subunits of SWI/SNF complexes, *ARID1A* is by far the most broadly and frequently mutated gene across human cancers and is often inactivated by frameshift or truncating mutations that are distributed along the length of the gene, resulting in a lack of protein expression (Kadoch and Crabtree, 2015; Otto and Kadoch, 2017).

As the majority of *ARID1A* mutations are loss-of-function, this gene is hypothesized to be a tumor suppressor. What is more, nearly all human samples tested present heterozygous mutations. This finding implies that the inactivation of a single *ARID1A* allele might be enough to drive tumorigenesis and suggests that it might function as a negative-dominant tumor suppressor (Kadoch and Crabtree, 2015).

However, the inactivation of both *Arid1a* alleles in mice is necessary for the development of tumors in the colon (Mathur et al., 2017), which probes that *Arid1a* functions as a recessive tumor suppressor in that particular animal model.

In contrast, in some hepatocellular carcinoma patients, *ARID1A* is overexpressed in primary tumors, but not in metastatic lesions (Sun et al., 2017). Concerning this, it has been observed that conditional deletion of *Arid1a* in mouse liver protects against tumor initiation, whereas its suppression accelerates progression and metastases in established tumors.

Similar findings have been observed in other types of cancer. Thus, *Arid1a* inactivation blocks tumorigenesis in *Apc*-mutant colorectal cancer but drives cancer in microsatellite instability context (Mathur et al., 2017). What is more, *Arid1a* inactivation leads to differentiation in *Apc/Pten*-mutant ovarian carcinomas (Zhai et al., 2016), whereas it drives tumor formation in *Pik3ca*-mutant context (Chandler et al., 2015). Consequently, *Arid1a* has context-dependent oncogenic and tumor-suppressive functions in the development of some kinds of tumors, depending on the mutational background.

3.4.1. Mutual exclusivity between SWI/SNF subunits

A substantial body of scientific evidence supports that mutations in genes encoding components of SWI/SNF complexes tend to be mutually exclusive in the same patient.

For instance, in colorectal cancer *ARID1A* is usually mutated in samples where neither *SMARCA4*, *ARID2*, *SMARCA2* nor *ARID1B* are mutated (Gonzalez-Perez et al., 2013). These experimental observations could be explained either by a synthetic lethality

relationship among different subunits or by a loss of selective pressure to inactivate genes encoding additional SWI/SNF subunits after the first mutation.

In agreement with the first hypothesis, it has been postulated that *ARID1A* and *ARID1B* present a synthetic lethal relationship in the colorectal cancer cell line HCT116, after realizing that *ARID1A*$^{-/-}$ cells present a prominent decrease in cell proliferation when downregulating the expression of *ARID1B* (Kelso et al., 2017). Moreover, NSCLC cell lines lacking the expression of *SMARCA4* form fewer colonies and are more prone to cellular senescence after *SMARC2* knock-down (Oike et al., 2013), which suggests that mutations in both catalytic subunits are lethal. Similar synthetic lethality relationships have been described among *SMARCA4* and *ARID2*, *SMARCA4* and *ACTB*, as well as *SMARCC1* and *SMARCC2* (Schick et al., 2019).

3.4.2. SWI/SNF complex disruption in lung cancer

In the last two decades, several studies provided much evidence of the important role of SWI/SNF complexes disruption in lung cancer development. Thus, SMARCA2 and SMARCA4 expression is lost in 30% of NSCLC cell lines. In addition to this, patients that maintain the expression of the catalytic subunits of the complex showed a significantly higher survival rate compared to the low-expression group (Reisman et al., 2003), highlighting the role of these genes as tumor suppressors. The prognostic role of SMARCA2/SMARCA4 expression has also been confirmed in larger cohorts of lung cancer patients (Bell et al., 2016; Matsubara et al., 2013), strengthening the fact that decreased expression of these subunits correlates with a worse prognosis.

Similar observations have been done in other accessory subunits of the complex. For instance, ARID1A expression is lost in almost 35% of NSCLCs, which significantly correlates with a poorly differentiated stage, a higher TNM score, and nodal metastases (Zhang et al., 2014). Additionally, *ARID2* has been found mutated in a loss-of-function manner in 5% of the cases of lung adenocarcinomas (Manceau et al., 2013).

According to the publicly available information deposited on the Catalogue Of Somatic Mutations In Cancer (COSMIC) at the time of the elaboration of this doctoral thesis, the most affected subunits of SWI/SNF complexes in lung cancer are *SMARCA4* (7.7%), *ARID1A* (7.1%), *ARID1B* (4.6%), and *ARID2* (4.2%) (**Figure 1.10**).

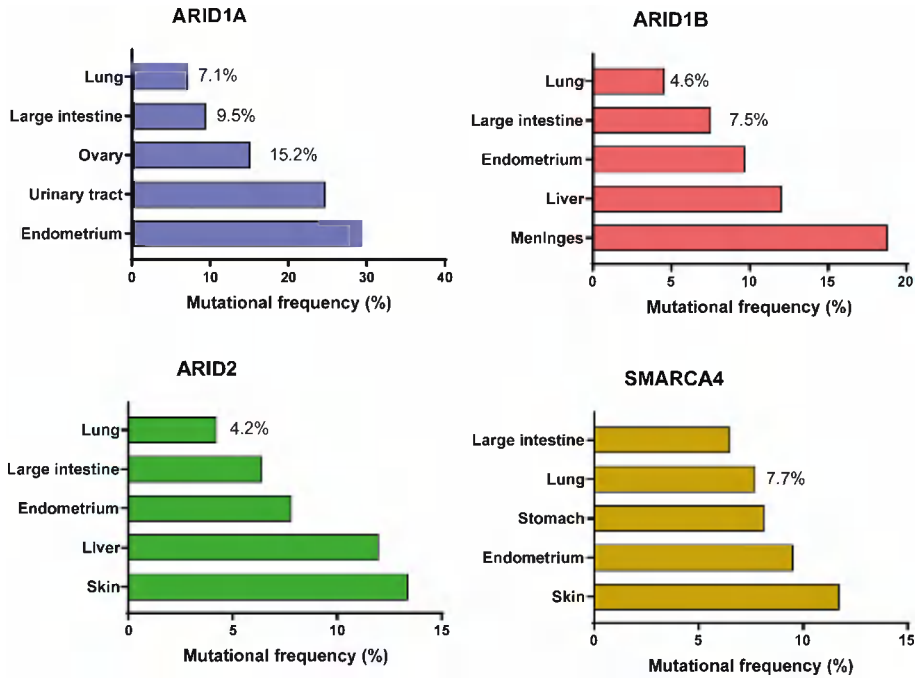


Figure 1.10. Mutation rate of accessory and catalytic subunits of SWI/SNF complexes across different tumor types. For each subunit (ARID1A, ARID1B, ARID2, or SMARCA4), the most commonly altered tissues according to the public database COSMIC (nov-2022, release 29) are indicated. In addition to this, mutation rates of those tissues that are specifically relevant in the context of this doctoral thesis are depicted.

Consistent with these findings, in our laboratory we sequenced a cohort of 225 lung cancer patients and found that *ARID2* mutations occurred at an even higher frequency (7.5%) (Moreno et al., 2021).

Finally, TRACERx Consortium has published that in NSCLC, whereas driver mutations in *EGFR*, *MET*, *BRAF*, and *TP53* were almost always clonal, mutations in SWI/SNF genes were systematically subclonal, indicating a late appearance during disease progression (Jamal-Hanjani et al., 2017).

3.5. Molecular pathways altered in SWI/SNF-mutant cancers

Despite the clear role of SWI/SNF perturbations in the development of different types of cancer, the molecular pathways behind its contribution to tumorigenesis still remain elusive (Gonzalez-Perez et al., 2013; Mashtalir et al., 2018; Mittal and Roberts, 2020; Wilson and Roberts, 2011). Among the postulated hypotheses that have more acceptance in the scientific community, we can stand out: (1) DNA repair alteration, (2) activation of canonical cancer-promoting pathways, and (3) promotion of epithelial to mesenchymal transition (EMT).

3.5.1. DNA repair alteration

There is a considerable amount of scientific evidence linking different subunits of SWI/SNF complexes with DNA repair mechanisms, which suggests that a potential mechanism of tumorigenesis in SWI/SNF-deficient cells is through the promotion of genomic instability.

The catalytic subunits of the complex (SMARCA2 and SMARCA4) are recruited to double-strand breaks (DSBs) through their bromodomains, where they increase chromatin accessibility to allow the proper assembly of protein complexes required for this DNA damage repair mechanism (Rother and van Attikum, 2017). In particular, SMARCA4 is able to bind to γH2AX nucleosomes (Lee et al., 2010) and interact with PARP1 (Chen et al., 2019), but it also takes part in the interstrand crosslinks (ICLs) damage repair through its interaction with BRCA1 (Wang et al., 2016).

SMARCA4 is as well involved in the decatenation of newly replicated sister chromatids, exerting a key role in the precise chromosome segregation accomplished after each mitosis. Consequently, *SMARCA4* loss gives rise to spontaneous anaphase bridges (Dykhuizen et al., 2013), which may result in an inequitable distribution of chromosomes, providing an additional source of genomic instability.

ARID1A recruitment to DSBs is mediated by its interaction with ATR (a kinase that controls cellular response to DNA damage), where it facilitates DNA DSB end resection (Shen et al., 2015). This accessory subunit is also required for the G_2-M DNA damage checkpoint (Shen et

al., 2015). Consequently, *ARID1A* loss might allow cells to progress through the cell cycle, even in case they had accumulated DNA damage.

In a renal cancer model, the absence of *PBRM1* results in a delay of p53-dependent response to DNA damage, characterized by a less efficient repression of genes that promote cell cycle progression (Feng et al., 2022). Consequently, an increasing number of cells progress through mitosis with unrepaired damage.

SMARCB1 promotes nucleotide excision repair (NER) by interacting with UV damage recognition factor XPC at DNA damage sites. In addition to this, SMARCB1 is needed for the recruitment of ATM to DNA damage foci and its activation by phosphorylation (Ray et al., 2009).

Supporting the role of SWI/SNF complexes in DNA repair, *ARID1A*-deficient colorectal cancer cells present a significantly higher sensitivity to PARP inhibitors (Shen et al., 2015). Similarly, *SMARCA4*-mutated lung cancer cell lines show an increase in DNA damage foci, together with an activation of the ATR pathway (Gupta et al., 2020).

3.5.2. Activation of canonical cancer-promoting pathways

SWI/SNF complex is able to interact directly with genes belonging to signaling pathways already known to be dysregulated in cancer, such as *TP53, RB1,* and *MYC*, suggesting that defects in these chromatin remodeling complexes could promote cancer through the activation of canonical cancer pathways.

Regarding this, it has been demonstrated that *SWI/SNF* and *TP53* or *PTEN* mutations do not usually concur in the same patient nor in colorectal or ovarian cancers (Kadoch et al., 2013; Wu et al., 2017). Subsequent studies have confirmed that *ARID1A* and *TP53* mutations are mutually exclusive in gynecological malignancies (Wu et al., 2017). Similarly, *BRD7* has been found frequently deleted in a subset of breast tumors that express wild-type *TP53,* which suggests that there might be exclusivity between mutations in both genes (Drost et al., 2010). Although it remains to be characterized in further detail, it has been postulated that BRD7 may act as a cofactor for TP53 in the oncogene-induced senescence process (Burrows et al., 2010; Drost et al., 2010).

Rhabdoid tumors harboring *SMARCB1* inactivating mutations are characterized by the overexpression of E2F target genes and the downregulation of $p16^{INK4A}$ (a cyclin-dependent kinase inhibitor) (Isakoff et al., 2005). Additionally, it has been observed that SWI/SNF complexes bind RB, facilitating the repression of its target genes, *E2F* and *CCND1* (Trouche et al., 1997). Collectively, both observations suggest that SWI/SNF alterations might result in an impaired RB tumor-suppressor pathway and consequently, the control upon cell cycle might be less restrictive.

Another potential tumor suppressor role for SMARCB1 has been proposed regarding its interaction with the oncogene *MYC*. In particular, SMARCB1 prevents MYC binding to DNA and represses its binding to SMARCC1. When SWI/SNF is altered, residual complexes are able to interact with MYC, increasing the transcription of its target genes (Jones et al., 2022).

Important relationships between the catalytic subunit of the complex SMARCA4 and critical molecular pathways have also been reported, which might provide molecular clues for its role in tumorigenesis. Thus, in NSCLC *SMARCA4* loss is synthetic lethal with CDK4/6 inhibition both *in vitro* and *in vivo* (Xue et al., 2019). Similarly, in the same tumor type *SMARCA4* alterations show a mutually exclusive pattern with mutations in *EGFR* (Fernando et al., 2020) and *MAX* (Romero et al., 2014) genes, together with *MYC* amplification (Medina et al., 2008; Romero et al., 2012). In line with this, SMARCA4 regulates the expression of MAX, and is required to activate neuroendocrine transcriptional programs and to upregulate MYC target genes, such as the ones implicated in glycolysis. Furthermore, in a murine model of LUAD where *SMARCA4* is ablated, there is enhanced oxygen consumption (Lissanu Deribe et al., 2018), which suggests that *SMARCA4*-deficient cell lines rely on this metabolic process to sustain their high metabolic rate. Supporting this idea, derived tumor cells are more sensitive to OXPHOS inhibition.

Mutual exclusivity relationships among SWI/SNF subunits and EGFR/MAPK-pathway have also been reported in other tumor types. Thus, *ARID1A* and *EGFR* mutations (including short variant activating mutations and gene amplification) are mutually exclusive in lung cancer (Johnson et al., 2022). In colorectal cancer, *ARID1A* mutations do not concur with mutations in genes encoding EGFR downstream effectors, such as *BRAF* or *KRAS* (Johnson et al., 2022).

Another layer of complexity appears when considering the distinct biochemical forms of SWI/SNF complexes and their interaction with transcriptional factors and coregulators. Regarding this, it has been proposed that ARID1B and ARID2 function coordinately to modulate the expression of hundreds of genes, depending on their interactions with factors that activate (BRCA1 or MAX) or repress (TEAD4 or TCF7L2) transcription (Raab et al., 2015). By contrast, ARID1A seems to have the opposite effect of ARID2 in transcriptional regulation (Raab et al., 2015).

Similarly, in an osteoblast murine cell line SWI/SNF complexes are able to both promote or repress the expression of the oncogene *MYC,* depending on their subunit composition. During cell differentiation, ARID1A-containing SWI/SNF complexes can repress directly *MYC* transcription (Nagl et al., 2006), whereas ARID1B-containing complexes have the opposite effect (Nagl Jr et al., 2007). Therefore, *ARID1A* inactivation may result in the aberrant activation of MYC programs, which could provide an explanation for tumor formation.

SWI/SNF disruption has been linked with the activation of other cancer-promoting pathways, such as PI3K-AKT (Kadoch and Crabtree, 2015) and β-catenin (Pugh et al., 2012). Thus, *ARID1A* or *ARID1B* loss-of-function mutations are frequently accompanied by an activating mutation in the PI3K-AKT pathway. What is more, both *Arid1a* inactivation and *Pik3ca* activation are required to give rise to ovarian clear cell carcinoma in mouse models (Chandler et al., 2015). It has also been reported that β-catenin interacts directly with SMARCA4, regulating its ability to promote the expression of its target genes (Barker et al., 2001; Kadoch and Crabtree, 2015).

3.5.3. Epithelial to mesenchymal transition

In the context of lung cancer, loss of any of the catalytic subunits of SWI/SNF (SMARCA2 or SMARCA4) results in a switch from an epithelial to mesenchymal state, evidenced by morphological changes, low E-cadherin and high vimentin expressions (Matsubara et al., 2013). Similar observations have been made in the absence of other subunits of the complex, as is the case of *ARID1A* in renal (Somsuan et al., 2019), pancreatic (Tomihara et al., 2021; W. Wang et al., 2019), and breast cancer (T. Wang et al., 2020), or *ARID2* in liver cancer (Jiang et al., 2020).

It is thought that this phenotypic switch is mediated by the inactivation of the TGF-β/Smad4 pathway, leading to the lack of occupancy at enhancers regulating the expression of genes associated with the maintenance of an epithelial cell phenotype (such as *TGFBR2*, *KLF4*, *FOXQ1*, and *BMP6*) in breast and liver cancer (Guo et al., 2022; Jdeed et al., 2022). In addition to this, the transcription factor KLF4 is able to interact directly with SWI/SNF complexes (through SMARCA4 and SMARCC2 subunits) to increase accessibility at enhancer sites that favor the expression of endothelial genes (Moonen et al., 2022).

3.6. Impact of SWI/SNF mutations in current cancer treatments

Due to the recurrence of SWI/SNF disruption in different tumor types, any new knowledge about how to exploit these alterations for therapy could improve the treatment of many cancer patients. According to the currently available treatments for NSCLC, several studies have provided some promising results.

Firstly, due to its role in DNA repair, many authors reported that the lack of some accessory subunits of the SWI/SNF complex sensitizes tumor cells to therapeutic strategies aimed at promoting genomic instability. For instance, *SMARCA4* (Bell et al., 2016), *ARID1A* (Shen et al., 2015), and *ARID2* (Moreno et al., 2021) alterations have been linked to a higher sensitivity to platinum-based chemotherapy or treatments based on the inhibition of PARP. In line with this, it has been proved that the combined use of BRD4 and PARP inhibitors is lethal in several cancer lineages (Sun et al., 2018). Moreover, *SMARCA4* (Gupta et al., 2020) and *ARID1A* (Karachaliou et al., 2018)-mutated lung cancer cell lines are more sensitive to ATR inhibition. Additionally, *SMARCA4*-

mutant NSCLC cell lines present a higher sensitivity to aurora kinase inhibition (Lissanu Deribe et al., 2018). Therefore, the loss of any of the above-mentioned accessory subunits might be used as biomarkers in lung cancer patients that could benefit from therapeutic strategies that exploit defects in the DNA repair machinery of tumor cells (**Figure 1.11**).

Secondly, it has been described that loss of the *SMARCE1* subunit gives rise to enhanced resistance to MET and ALK inhibitors in NSCLC cell lines, which is restored by the use of EGFR inhibitors (Papadakis et al., 2015). It is thought that this resistance relies on AKT and ERK activation. If these results are consistent in human patients, *SMARCE1* expression might emerge as a potential predictive marker for drug response to MET and ALK inhibitors. In addition to this, *SMARCE1*-mutated tumors may benefit from treatments based on the pharmacological inhibition of EGFR.

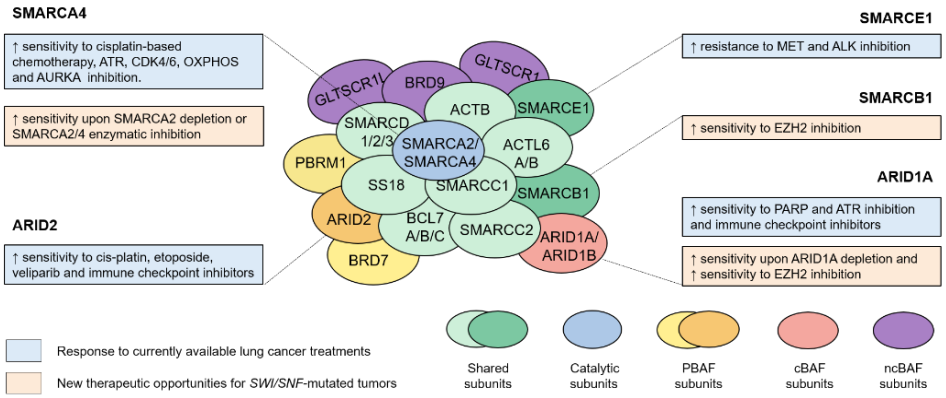


Figure 1.11. Therapeutic implications of SWI/SNF-altered lung tumors. The most commonly mutated subunits of the complex and their association with different sensitivities to lung cancer treatments are depicted in different colors, depending on whether they are currently available (blue) or under development (orange). Shared subunits of SWI/SNF complexes are represented in green. The catalytic subunits of the complex (SMARCA2 and SMARCA4) appear in blue. The accessory subunits that define PBAF, cBAF, and ncBAF subfamilies are represented in yellow, red, and purple, respectively.

Another strategy that has been proven successful in many cancer types, including lung cancer, is the boost of the proper immune response of the patient against their tumor cells. Regarding this, it has been reported that the murine melanoma cell line B16F10 increased its sensitivity to T cell-mediated cytotoxicity after mutating different subunits of the PBAF complex (in particular, *Arid2*, *Pbrm1,* and *Brd7*) (Pan et al., 2018). These results seem to indicate that tumors harboring mutations in the SWI/SNF complex might respond better to immunotherapy.

In agreement with this hypothesis, mice bearing tumors induced with *Pbrm1*-deficient B16F10 cells were more strongly infiltrated by cytotoxic T cells, developed smaller tumors, and had improved survival in comparison with control cells. Furthermore, whole exome sequencing of human metastatic clear cell renal carcinomas revealed that *PBRM1* inactivation is associated with a better clinical outcome from immune checkpoint inhibitor (ICI) therapy (Miao et al., 2018). It is thought that this response is mediated by increased levels of cytosolic DNA that in turn activates inflammatory signaling pathways (Feng et al., 2022).

The potential relationship between SWI/SNF and response to immunotherapy has been extended to other accessory subunits of the complex, such as *ARID1A* and *ARID1B* (Zhu et al., 2021). However, the precise mechanism that explains why *ARID1A*, *ARID1B,* or *ARID2*-mutated tumors have more chances to benefit from ICI therapy has not been elucidated yet.

Finally, a recent whole-genome sequencing study performed on patients treated with immune checkpoint inhibitors has revealed that, collectively, SWI/SNF alterations are linked with better clinical outcome in non-small cell lung cancer (Wang et al., 2023).

3.7. New therapeutic opportunities for SWI/SNF-mutated patients

SWI/SNF complexes exist in multiple compositions since several subunit positions can be occupied alternatively by proteins encoded from different genes. According to that, many authors have reported a dependency on the remaining functional alternative subunit in cells with alterations in specific SWI/SNF components. That is the case of SMARCA2 in *SMARCA4*-deficient cells (Hoffman et al., 2014; Oike et al., 2013) or ARID1B in *ARID1A*-deficient cells (Katherine C Helming et al., 2014; Kelso et al., 2017). Consequently, it has been postulated that this dependency could be exploited therapeutically. According to that, SMARCA2/4 specific protein degraders or ATP inhibitors impair celullar proliferation in *SMARCA4*-deficient lung cancer cellular models (Farnaby et al., 2019; Papillon et al., 2018). In line with this idea, as other SWI/SNF subunits contain bromodomains, a specific sensitivity to PFI-3 (Gerstenberger et al., 2016; Schiaffino-Ortega et al., 2014), BI-7273 (Hohmann et al., 2016) or I-BRD9 (Theodoulou et al., 2016) inhibitors could be considered.

However, it is also worth keeping in mind that these synthetic lethality relationships have been reported in certain cellular contexts, and simultaneous mutations in paralog subunits are prevalent in cancer. For instance, over 94% of SCCOHT tumors present mutations in the *SMARCA4* gene and its paralog subunit, SMARCA2, is epigenetically silenced (Karnezis et al., 2016), which probes the existence of viable cells besides the dual loss of both catalytic subunits. In addition to this, *ARID1A* and *ARID1B* mutations co-occur in a reduced subgroup of gastric, endometrial, and liver malignancies (1-5%) (Z. Wang et al., 2020), more than a third of *ARID1A*-mutant cell lines harbor *ARID1B*-inactivating mutations (Katherine C Helming et al., 2014), and dual loss of these accessory subunits promotes liver carcinogenesis in animal models (Z. Wang et al., 2020).

Other therapeutic strategies take advantage of the deficient activity of histone modifying enzymes in certain subgroups of tumors. Thus, *SMARCA4*-deficient cells express reduced levels of histone demethylases KDM6A and are strongly dependent on its activity. Therefore, the use of the histone lysine demethylase inhibitor GSK-J4 has shown a strong anti-tumor effect on lung and ovarian cancers orthotopically transplanted in mouse models (Romero et al., 2021).

Lastly, another therapeutic approach is based on the opposing role between polycomb repressive complex 2 (PRC2) and SWI/SNF (Bitler et al., 2015; Knutson et al., 2013). According to this dependency, the inhibition of the catalytic subunit of the complex (EZH2) has shown very promising anti-tumoral activity in a variety of SWI/SNF-deficient cell lines (highlighting *SMARCB1*, *ARID1A,* and *SMARCA4*) in MRTs (Knutson et al., 2013), OCCCs (Bitler et al., 2015) and in a subset of lung carcinomas (Januario et al., 2017). Therefore, EZH2 inhibition could suppose a future treatment opportunity for SWI/SNF-deficient tumors. In agreement with these promising results, a specific inhibitor of EZH2, tazemetostat, has been approved for the treatment of epithelioid sarcoma and MRT (Centore et al., 2020), constituting the first targeted therapy described for SWI/SNF-mutant tumors.

4. Next-generation sequencing technologies

DNA sequencing is a process based on different biochemical reactions and detection methods that allows the determination of the correct order of nucleotides in a DNA fragment.

The completion of the Human Genome Project in 2003 using the Sanger chain termination method (Sanger et al., 1977) represented a milestone in genomic research, improving our knowledge of human biology and disease. However, the limited throughput and the high cost of sequencing remained major barriers to the study of big patient cohorts, revealing the need for more advanced technologies (Goodwin et al., 2016).

The emergence of next-generation sequencing (NGS) technologies, also known as massively parallel or high-throughput sequencing, in the first decade of the 21st century supposed another big jump forward in the field of genomic research, due to the progressive decrease in the amount of time and the economic investment involved in this process. Thus, the first human genome required 15 years to be sequenced and its cost almost reached 3 billion dollars. With the new NGS technologies, the same amount of information can be determined within a week by less than 600 dollars (https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data).

4.1. Next-generation sequencing overview

The first commercial NGS platform developed was the 454 pyrosequencing from Roche in 2005 (Ronaghi et al., 1996). In this strategy DNA was amplified through emulsion PCR, using oligonucleotides immobilized on beads that were in turn captured inside emulsion droplets. In this way, thousands of identical copies of a DNA fragment were produced, which ensured that the generated signal could be distinguished from the background noise (Goodwin et al., 2016).

In each sequencing cycle, one of the four nucleotides (dATP, dCTP, dGTP, or dTTP) was added and incorporated by complementarity in some of the growing chains by DNA polymerase. The by-product of this reaction was a pyrophosphate group (PPi), that was afterward converted into ATP by ATP sulfurylase. In the presence of ATP, luciferase catalyzed the conversion of luciferin to oxyluciferin,

generating light. Each burst of light was detected by a camera, which monitored the incorporation of a given nucleotide in a particular bead in each of the cycles.

As a consequence of its detection method, each of the four deoxynucleotides (dNTPs) had to be added separately to make sure that a single dNTP was responsible for the detected signal. It did not require the use of terminators because the absence of other nucleotides prevented the elongation reaction from happening.

This platform was able to generate long reads (up to 700 bp), which were easier to map to a reference genome. The main problem of pyrosequencing technology was the resolution of homopolymer regions (Kchouk et al., 2017) because it had to be determined by the intensity of the produced light (too-high or too-low intensity signals lead to under- or overestimation of the number of nucleotides of a homopolymer). Regardless of its advantages, the 454 platform rapidly became non-competitive in terms of yield and cost, leading to its discontinuation in 2016.

Following the same sequencing principle, Life Technologies developed the Ion Torrent system (Rothberg et al., 2011) in 2010, which was the first NGS platform that detected the incorporation of nucleotides without requiring an optical system detection. This sequencing method takes advantage of the fact that when a nucleotide is incorporated into a DNA growing chain, a single proton is released as a by-product of this reaction, producing a change in the pH solution. This change can be detected by a sensor and converted into a voltage signal, which is proportional to the number of nucleotides incorporated (Kchouk et al., 2017).

Among the major advantages of Ion Torrent sequencers, it can be highlighted the high read length (up to 600 bp) and the fast sequencing time (2-8 hours). It shares some drawbacks with 454 pyrosequencing, such as the lack of accuracy in sequencing homopolymers larger than 6-8 bp (Goodwin et al., 2016).

Supported Oligonucleotide Ligation and Detection (SOLiD) was an NGS sequencer acquired by Applied Biosystems in 2007. This system relied on a sequencing by ligation method based on the use of fluorophore-labeled probes that were composed of two known nucleotides followed by degenerate bases (Valouev et al., 2008). During the process of sequencing, these probes bound to a DNA template by complementarity, allowing the ligase to incorporate the oligonucleotide into the growing chain, and the corresponding fluorescent signal was captured. Degenerate bases and the fluorophore were cleaved off, leaving a 5-bp extended fragment. As this process occurred, two out of every five bases were identified.

There were no spectrally resolvable fluorophores to differentiate the possible sixteen dinucleotide combinations. Therefore, four fluorophores corresponding to four different dinucleotide combinations were used. To get the complete sequence of a given fragment, the process of binding, ligation, imaging, and cleavage was repeated, moving the original position to n + 1, n + 2, n + 3, and n +4.

The main strength of the SOLiD platform was the high accuracy, as each base was read twice. However, one of its main limitations was the very short read length (maximum of 75 bp), which limited its use for the identification of structural variants or the assembly of a genome (Goodwin et al., 2016). These technical limitations and the amount of time involved in the reaction (up to several days) had limited its use to a small niche within the industry.

In 2006 Solexa developed a new method of sequencing, which was then acquired by Illumina in 2007. Among the different platforms developed in that decade, Illumina is by far the most widely used system for short-read sequencing.

Illumina platform is a sequencing by synthesis method based on the use of fluorescently labeled nucleotides with a reversible terminator (Bentley et al., 2008). In particular, the ribose 3’-OH group of each nucleotide is blocked by an O-azidomethyl group, preventing its elongation (Goodwin et al., 2016; Metzker, 2010).

During the first steps of library preparation, total DNA is fragmented into small sequences, and adapters are ligated to both ends of each fragment. In this technology, a library of DNA fragments is loaded into a flow cell, where each fragment is captured on a solid surface completely covered by oligonucleotides complementary to the

ligated adapters that are common to all fragments, and is then amplified into a clonal cluster through bridge amplification. A highly variable number of millions of clusters can be generated depending on the specific Illumina platform, which impacts in the total sequence production. It is also important to control precisely the concentration of the template, so that the amplification takes place locally, generating no overlapping clonal clusters.

In each cycle of sequencing, a mixture of four individually labeled and 3'-blocked dNTPs is added. After the incorporation of a single nucleotide into the growing chain, unbound dNTPs are removed and the fluorescent emission from each cluster is recorded. A reducing agent (tris(2-carboxyethyl)phosphine) is then added to remove the dye and regenerate the 3'-OH group, allowing the beginning of a new sequencing cycle. As all nucleotides are present in each sequencing cycle, natural competition reduces sequence context-specific errors, reaching a highly accurate performance compared to other technologies (Bentley et al., 2008).

For a given cluster, all identical strands are read simultaneously and hundreds of millions of clusters are sequenced in a massively parallel process. Each strand of DNA is read separately by selectively cleaving off those fragments captured by 3' or 5' ends, respectively.

Due to the possibility of using two different adapters in 3' and 5' ends, the same DNA fragment can be sequenced from both sides (paired-end sequencing), which favors a more accurate read alignment compared to single-end sequencing, where a DNA template is sequenced in just one direction.

In addition to this, paired-end (PE) sequencing allows the possibility of detecting insertion and deletions and the removal of PCR duplicates (a common artifact generated during the PCR amplification).

The number of cycles determines the length of the read, which can vary between 50-300 bp according to the sequencing platform used (**Figure 1.12**). Additionally, Illumina technology has an overall accuracy rate greater than 99.5% (Bentley et al., 2008).



	iSeq	MiniSeq	MiSeq
	Benchtop sequencers		
Applications	Small WGS, targeted gene sequencing	Small WGS, targeted gene sequencing, targeted gene expression profiling, 16S metagenomic sequencing	Small WGS, targeted gene sequencing, 16S metagenomic sequencing
Run Time	9.5-19 h	4-24 h	4-55 h
Maximum Output	1.2 Gb	7.5 Gb	15 Gb
Maximum Reads Per Run	4×10^6	25×10^6	25×10^6
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp



	NextSeq	HiSeq	NovaSeq
	Production-scale sequencers		
Applications	Small WGS, WES, targeted gene sequencing, transcriptome sequencing,, single-cell profiling, miRNA and small RNA analysis	Large WGS, small WGS, WES, transcriptome sequencing, chromatin analysis, methylation sequencing, metagenomic sequencing, miRNA and small RNA analysis	Large WGS, WES, single-cell profiling, transcriptome sequencing, chromatin analysis, methylation sequencing, metagenomic sequencing, metagenomic profiling, cell-free sequencing and liquid biopsy analysis
Run Time	11-48 h	7 h -6 days	13-44 h
Maximum Output	120-330 Gb	9-9000 Gb	6000 Gb
Maximum Reads Per Run	$0.4\text{-}1.1 \times 10^9$	$0.3\text{-}4 \times 10^9$	20×10^9
Maximum Read Length	2 × 150 bp	2 × 250 bp	2 × 250 bp

Figure 1.12. Illumina sequencing platforms. Main applications and specifications of different sequencers available at Illumina. Benchtop and production scale sequencers' specifications summarize the properties of the different systems developed (run time, maximum output, reads per run, and read length).

Establishing the complete sequence of a genome cannot be fully achieved with the previously described methodologies because there are repetitive and complex regions difficult to assemble and resolve with such short reads (Goodwin et al., 2016). This technical issue explains why the complete sequence of the human genome was not deciphered until 2022 (Nurk et al., 2022), when the latest long-read sequencing methodologies were applied.

Unlike the previous methods, in these new technologies, DNA amplification is not required. Instead, they detect directly the composition of a single-stranded DNA (ssDNA) molecule measuring light (PacBio) or electrical current fluctuations (Nanopore). The average read length obtained using these approaches varies between 10-15 kb, which makes them ideal for *de novo* genome assembly or the resolution of complex genomic structures.

Pacific Biosciences developed the first single-molecule real-time sequencing approach (Eid et al., 2009). Template fragments are ligated to hairpin adapters at each end, resulting in a circular DNA molecule with ssDNA at the ends and the double-stranded DNA (dsDNA) template in the middle (Goodwin et al., 2016). This platform makes use of nanometric wells where a DNA polymerase and the target DNA fragment are attached. During the sequencing reaction, a new DNA strand is synthesized by a DNA polymerase using fluorescently labeled nucleotides. When a dNTP is incorporated, it releases a luminous signal recorded by a sensor (Kchouk et al., 2017).

This technology presents a high error rate (about 13%) (Kchouk et al., 2017), with insertions and deletions being the most common mistakes. These errors are randomly distributed along the length of the read. As this platform uses a unique circular template, it can be sequenced multiple times, reaching a high coverage and improving the resulting accuracy (Goodwin et al., 2016).

Oxford Nanopore technology emerged in 2009 (Clarke et al., 2009). Unlike other methods, it detects directly DNA composition without any synthesis process. DNA is fragmented and two different adapters (a leader and a hairpin) are ligated to facilitate the interaction of the DNA with a protein nanopore, which can be an α-hemolysin or a Mycobacterium smegmatis porin A (Kircher and Kelso, 2010; Wang et al., 2015). Once the individual nucleotides translocate through the pore, due to the establishment of an electrical voltage, a current blockade

occurs. Each nucleotide provides a characteristic current fluctuation that is recorded to unravel the DNA composition of the fragment (Slatko et al., 2018).

The distinguishing features of this methodology include: (1) the sequencing of a single molecule, which avoids the need for PCR amplification and the consequent incorporation of errors in each cycle; (2) the sequencing of the direct strand without the potential errors associated with DNA synthesis, (3) the fast sequencing in real time, instead of being paused after each incorporation (Schadt et al., 2010), and (4) the generation of long sequencing reads, resolving problems with repetitive regions (Kchouk et al., 2017).

In 2014 Oxford Nanopore Technologies released the MinION device, which was the first commercially available and portable size sequencer based on this methodology that can be plugged into a USB port. This platform combines the potential for long read lengths (average read length of 10 kb) with high speed (70 bp/s) (Goodwin et al., 2016). Despite its high sequencing error rate (12%), it can be run off a personal computer, turning into a reasonable alternative for clinical diagnostic applications and hard-to-reach field locations.

4.2. Next-generation sequencing applications

As technology continues evolving, an increasing number of NGS library preparation methodologies appear in an attempt to address more complex biological questions related to the genome, the transcriptome, and the epigenome of any organism. The final information extracted from each of them is determined by a unique combination of the starting material, the specific protocols used during library preparation, and the specific bioinformatic tools used for its analysis (Metzker, 2010). Accordingly, a new generation of bioinformatic tools has emerged as a requisite to analyze sequencing data.

Regardless of the biological origin of the sample, most NGS library methodologies are based on three basic steps: generation of a collection of DNA fragments, adapter ligation at 3' and 5' ends, and PCR amplification (**Figure 1.13**).

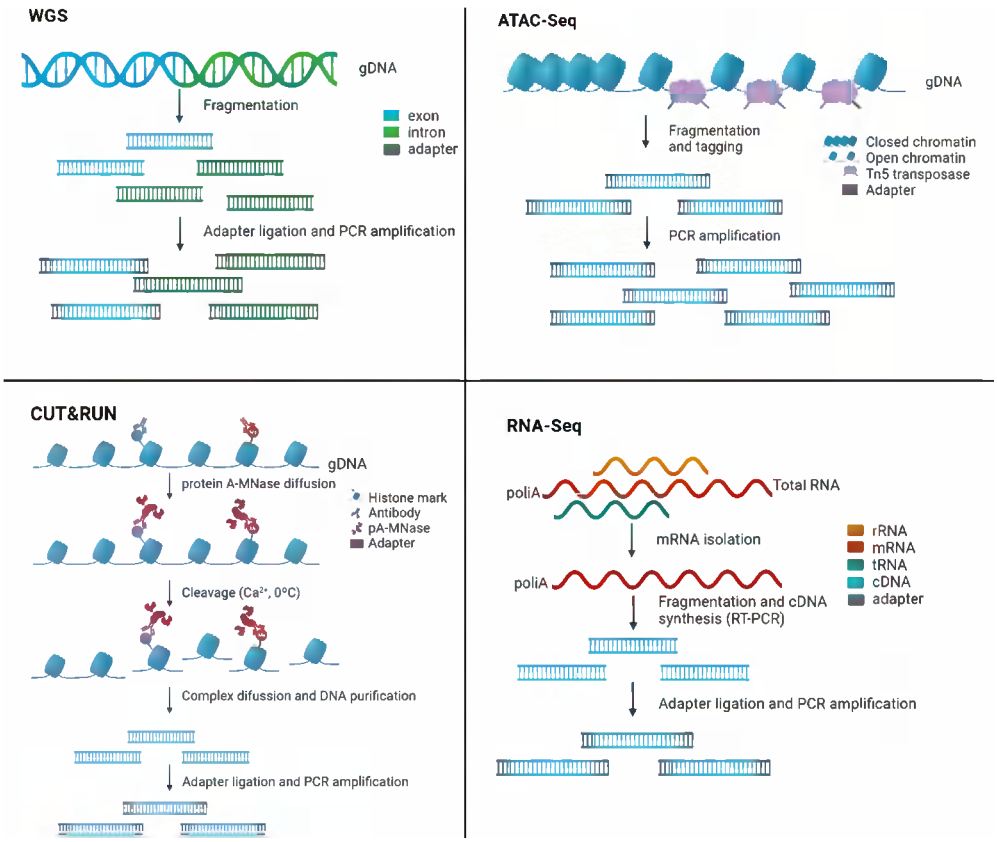


Figure 1.13. Overview of different NGS library preparation methods. Schematic representation of the required steps of a selection of genomic and transcriptomic library preparation methods. Abbreviations: WGS, whole

genome sequencing; ATAC-Seq, Assay for Transposase-Accessible Chromatin with high-throughput sequencing; CUT&RUN, Cleavage Under Targets & Release Using Nuclease; RNA-Seq, transcriptome sequencing.

In particular, Illumina adapters include at least two parts: (1) a fragment required for sequencing, which is the primer binding site that allows the binding of DNA polymerase, and (2) a capture sequence, which allows the DNA sample to be bound into the flow cell sequencing. Furthermore, depending on the application, additional sequences (such as single-molecule or sample-specific barcodes) can be incorporated into the adapters.

Before preparing the sequencing reaction, it is recommended to perform a pre-quality control to check the ligation efficiency and the concentration of the library. The identification of a narrow peak typically around 200-400 bp (depending on the application) in a capillary electrophoresis analysis, without smaller or larger fragments (adapter dimers or genomic DNA fragment contamination, respectively) is indicative of a good quality library.

Despite the great diversity of NGS methodologies, once the sequence of the fragments generated has been identified, data analysis frequently involves a sequencing quality control to remove repetitive or low-quality sequences followed by the alignment to a reference genome. In the following chapters, I will focus on those applications that are especially relevant in the context of this doctoral thesis.

4.2.1. DNA-Seq applications

Whole-genome sequencing (WGS) is the most commonly used application to identify the full sequence of a previously unknown genome (*de novo* sequencing), but it can also be used to identify differences versus an already available reference genome (re-sequencing).

WGS has already been applied in thousands of samples to identify sequence variants present in the entire genome, including single nucleotide variants (SNVs) or substitutions, small insertions and deletions (indels), copy number variations (CNVs), and large-scale reorganizations, such as translocations, deletions, insertions, or inversions.

Despite the reduction in sequencing costs, the study of the whole genome at high coverage still involves a large economic investment. Considering that the exome (the collection of all exons in the genome) represents only 1 % of the human genome (Teer and Mullikin, 2010), but contains the majority of variants responsible for diseases, whole exome sequencing (WES) represents a cheaper alternative to WGS. Additionally, targeted sequencing to an even smaller and specific set of genomic regions is commonly used, especially when the group of genes responsible for a specific disease that wants to be diagnosed has been clearly identified. By limiting the amount of required sequence to cover the region of interest, it is possible to increase the depth of the study. Regarding this, validation of very low-frequency sequence variants requires an extremely high coverage (10,000x) that cannot be achieved with WGS (Goodwin et al., 2016).

4.2.1.1. Sequence quality control

Once the libraries have been sequenced, the quality of the reaction is determined by the Phred quality score (Q). In particular, a Q30 is considered a benchmark for quality in NGS, which means that the error rate is 1/1000, ensuring an accuracy rate of 99.9%. Low-quality reads are often filtered off to avoid troubles with downstream analysis.

Tools such as FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) allow the extraction of some basic statistics, like the sequence quality through the read length, the GC content, the distribution of insert sizes, or the abundance of specific repetitive sequences. The diversity of the library (indicated by the amount of PCR duplicates) can also be determined using other programs, such as SAMTools (Li et al., 2009) or PICARD (http://broadinstitute.github.io/picard/).

4.2.1.2. Read alignment

Most NGS applications require a resource-consuming step in which the generated short sequences (reads) are aligned to a reference genome. Almost all aligners developed for this purpose try to identify all possible locations of these short sequences on the complete genome for a specific part of the read (seed). Subsequently, this seed is extended and the number of mismatches or gaps required for the alignment of the complete read is computed by the aligner in order to choose the most probable location of the read. Additionally, a quality score representing the certainty of the alignment is calculated.

SOAP (R. Li et al., 2008) and MAQ (H. Li et al., 2008) were the first aligners developed using the previously explained strategy. These tools were afterward improved by BWA (Li and Durbin, 2009) and Bowtie (Langmead et al., 2009) applying the Burrows-Wheeler transformation to index the reference genome. In line with this, Hisat2 (Kim et al., 2019) developed the graph FM indexing approach to enable a rapid and accurate alignment of the sequencing read with considerably lower memory requirements.

The above-mentioned tools generate a SAM format file (or the compressed versions, BAM or CRAM) that can be further edited by sorting, cleaning, or indexing using SAMtools, PICARD, or GATK (McKenna et al., 2010). The resulting files of the alignment can be graphically visualized using other tools, like Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013).

4.2.1.3. Identification of genetic variants

DNA-Seq applications allow the identification of different types of genetic variants, which can be grouped according to the length of nucleotides affected in the following categories:

(1) Point mutations: nucleotide substitutions and small insertions or deletions (indels).
(2) Structural variations: inversions, deletions, duplications, translocations, and copy number variations (CNVs).

Among the different computational strategies designed for variant calling, the identification of nucleotide substitutions is by far the most developed. Many tools are based on the Bayesian model initially described in MAQ (H. Li et al., 2008), and compute the different

probabilities associated with each potential genotype in a specific genomic position, according to the sequence reported for all the reads that cover that specific position. This approximation is suitable for germline substitutions. However, in cancer somatic mutations, the presence of contaminating normal DNA, copy number alterations, and intratumor heterogeneity prevent the proper construction of a prior model of expected frequency.

To overcome this issue, new tools have been specifically developed for the detection of somatic SNVs, such as Strelka (Saunders et al., 2012), VarScan 2 (Koboldt et al., 2012), Seurat (Christoforides et al., 2013), MuTect (Cibulskis et al., 2013), or RAMSES (Martínez et al., 2014).

The detection of indels is a more difficult task, as very few tools display good sensitivity and specificity. The main issue of concern relies on the alignment of reads containing this type of mutation. For this reason, some tools like Pindel (Ye et al., 2009) and Dindel (Albers et al., 2011) use a second alignment step on singleton reads to improve the identification of insertions and deletions. However, these tools are not completely specific and generally require additional filtering steps.

Illumina sequencers offer the possibility of performing paired-end sequencing, which can be applied to identify structural variations, such as insertions, deletions, duplications, inversions, and translocations. Regarding this, genomic breakpoints can be identified considering the alignment orientation of the reads, as well as the alignment distance between them. In high-coverage sequencing, this observation can also be complemented by the presence of split-apart reads, which contain segments that align in different genomic locations. Among the specific software designed to detect genomic breakpoints, BreakDancer (Chen et al., 2009) and DELLY (Rausch et al., 2012) are two of the most widely used.

The quantitative nature of NGS technologies can also be used to detect copy number variations. Thus, the number of reads coming from a specific region depends on the number of DNA copies for that specific region in the starting genome. There is a variety of tools that allow the identification of deletions and amplifications taking advantage of this fact, such as CODEX (Jiang et al., 2015), CNV-seq (Xie and Tammi, 2009), Control-FREEC (Boeva et al., 2012), or ExomeCNV (Sathirapongsasuti et al., 2011).

4.2.1.4. Identification of epigenetic alterations

Epigenetic alterations refer to variations in the molecular structure of the DNA or the associated proteins that do not result in a nucleotide change in its sequence. Among these modifications, DNA methylation and histone covalent modifications are some of the most deeply studied.

Many of these alterations can be detected using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) (Johnson et al., 2007). In this technique, there is an initial step of DNA fragmentation, followed by enrichment through immunoprecipitation with specific antibodies. Once purified and sequenced, these fragments provide information about the genome-wide location of the tested modification. ChIP-seq is the most frequently used method for assaying protein-DNA interaction *in vivo*.

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (Skene et al., 2018) is a chromatin profiling strategy with similar applications to ChIP-Seq. This method is performed with fresh cells and uses a fusion protein that is able to cut the specific regions where the antibody is attached, releasing those protein-DNA complexes into the supernatant for sequencing. Among the main advantages of this technique, it can be highlighted that it does not require cross-linking or DNA fragmentation, it is quicker than ChIP-Seq protocols, it can be performed with low quantities of starting material, and has shown good results in the case of histone modifications and transcription factors.

ChIP-Seq can be applied as well to detect DNA methylation with specific antibodies targeting 5-methylcytosine (MeDIP-Seq) (Down et al., 2008). Alternatively, bisulfite sequencing (Li and Tollefsbol, 2011), where DNA is treated with bisulfite to convert unmethylated cytosines to uracils before sequencing, has also been widely used to determine the pattern of methylation of a given sequence.

4.2.1.5. Identification of changes in chromatin structure

Another group of DNA-Seq applications offers the possibility of studying chromatin conformation in terms of accessibility. Thus, DNAse-seq (sensitive regions to DNAse I digestion) (Boyle et al., 2008), FAIRE-seq (sensitive regions to cross-link with paraformaldehyde) (Giresi et al., 2007), or ATAC-seq (accessible regions by Tn5 transposase) (Buenrostro et al., 2015) provide information about the genome-wide location of open or closed regions (Meyer and Liu, 2014). These regions are unprotected by proteins or nucleosomes and therefore are accessible to the action of a variety of enzymes.

HiC-Seq (Lieberman-Aiden et al., 2009) provides another layer of complexity enabling the analysis of spatial genome organization and the identification of topologically associated domains (TADs). This technique quantifies the number of interactions between genomic elements that are closely located in the three-dimensional structure of DNA but may be separated by many nucleotides in the linear genome.

4.2.1.6. Peak calling and enrichment analysis

Regardless of the methodological differences among the techniques previously exposed for epigenetic studies, once raw sequences have been aligned to a reference genome, the next step is the identification of specific regions with significant enrichment in sequence coverage compared to the background noise of the genome (enriched region identification). These regions or peaks where counts are accumulated can be identified using different tools. The most popular one is MACS3 (Zhang et al., 2008), but FindPeaks (Fejes et al., 2008), F-seq (Boyle et al., 2008), QuEST (Valouev et al., 2008) or SICER (Xu et al., 2014) can also be used for the same purpose. Subsequently, significant differences in the coverage of each of these peaks among samples can be identified using DESeq2 (Love et al., 2014).

Genome annotation and functional interpretation of the identified peaks can be achieved with different software, such as ChIPseeker (Yu et al., 2015), ChIPpeakAnno (Zhu et al., 2010), HOMER (Heinz et al., 2010) or GREAT (McLean et al., 2010). HOMER can also be used for motif discovery, offering the possibility of analyzing sequence patterns of already known transcription factors, as well as *de novo* motif discovery. Finally, ChromHMM (Ernst and Kellis, 2017) is a software designed for characterizing chromatin states using a multivariate

hidden Markow model. Thus, multiple histone modification datasets can be integrated to identify recurrent spatial patterns of marks that are then computed and annotated to calculate the most probable state for each genomic segment.

4.2.2. RNA-Seq applications

Transcriptome characterization can also be addressed using NGS, taking advantage of the fact that the number of generated reads corresponding to a specific genes is proportional to the initial number of RNA molecules in the sample. A commonly used option is the sequencing of the whole transcriptome to get a complete view of a cellular transcriptional profile at a given moment, but targeted sequencing of a specific subset of genes represents another possibility.

Most experimental protocols begin with an isolation step of messenger RNA (mRNA) before its conversion to cDNA, although there are specific protocols developed to capture small RNAs or non-coding RNAs.

One of the main advantages of studying the complete transcriptome is the possibility of detecting RNA isoforms, novel transcripts, gene fusions, splice variants, and other post-transcriptional modifications that frequently occur during tumorigenesis (Z. Wang et al., 2009).

4.2.2.1. Identification of transcriptional alterations

Similar to DNA-Seq applications, once the quality of the raw sequences has been assessed, they are aligned to a reference genome. TopHat2 (Kim et al., 2013) and Hisat2 (Kim et al., 2019) accomplish this task in two stages. First, all the exons of the genome are identified by a stringent direct alignment of the reads to the reference genome. Unaligned reads are subsequently used to find the junctions (splice events) between exons. STAR (Dobin et al., 2013) presents a higher mapping rate and speed compared to other aligners, maintaining high sensitivity and precision. It is also suitable for the detection of events of non-canonical splicing or fusion transcripts.

Gene expression alterations can be determined by taking advantage of the quantitative nature of NGS technologies. The level of expression of a given transcript is usually normalized as RPKM or FPKM (reads or fragments per transcript kilobase and per million total reads) to compare gene expression levels across different sequencing reactions. The most commonly used tool to determine the number of counts per transcript is HTSeq (Anders et al., 2015).

Two of the most broadly used software to detect changes in gene expression are DESeq2 (Love et al., 2014) and Cufflinks (Trapnell et al., 2012). Other tools developed for the same purpose are EdgeR (Robinson et al., 2010), limma (Ritchie et al., 2015), and NOISeq (Tarazona et al., 2015). Among them, the Cufflinks suite counts with specific tools to identify new genes, transcript-specific expressions, or new splice variants. Furthermore, if both alleles of a specific gene present different nucleotide sequences, RNA-Seq analysis enables the identification of specific expression patterns for each of the different alleles.

HYPOTHESIS AND OBJECTIVES

“Speculation and the exploration of ideas beyond what we know with certainty are what lead to progress.”
– Lisa Randall

Despite the recurrence of SWI/SNF alterations in multiple types of cancer, the molecular mechanisms behind its contribution to tumorigenesis remain largely unknown. Another point of active research is studying whether these alterations could confer vulnerabilities therapeutically actionable.

Consequently, we hypothesize that identifying direct targets of SWI/SNF complexes can increase our knowledge of the molecular mechanisms involved in tumor development. Additionally, this knowledge might be translated into clinical practice by ameliorating the prognosis and the treatment of lung cancer patients.

Along this line, the general objective of this doctoral thesis is to improve our understanding of the molecular mechanisms by which SWI/SNF alteration promotes tumor progression, with the final aim of improving the clinical outcome of cancer patients.

This general aim is implemented with the following specific objectives:

1. To characterize the molecular mechanisms behind the role of SWI/SNF complexes in lung cancer development.

2. To identify potential therapeutic opportunities in SWI/SNF-mutated lung cancer patients.

3. To study the molecular mechanisms involved in SWI/SNF-mediated gene expression regulation.

EXPERIMENTAL PROCEDURES

“Every brilliant experiment, like every great work of art, starts with an act of imagination.”

– Jonah Lehrer

---

1. Cell biology methods

1.1. Cell lines and culture conditions

Human cancer cell lines A549 (ATCC® CRM-CCL-185™), NCI-H460 (ATCC® HTB-177™), NCI-H226 (ATCC® CRL-5826™), NCI-H1568 (ATCC® CRL-5876™), Caco-2 (ATCC® HTB-37™), SK-OV-3 (ATCC® HTB-77™), and human embryonic kidney 293T (HEK 293T/17) (ATCC® CRL-11268™) were obtained from The Francis Crick Institute and the American Type Culture Collection (ATCC) repositories, authenticated by STR profiling and tested for mycoplasma (as detailed in section 1.2. "Mycoplasma PCR").

SK-OV-3, A549, and HEK 293T/17 cell lines were maintained in DMEM (Corning, 10-017-CV) supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco, 10270-106), 30 μg/ml of gentamycin sulfate (Biowest, L011-010) and 2 μg/ml of ciprofloxacin (CPX) (Acros Organics, 449620050) at 37 ºC in a humidified atmosphere containing 5% of CO_2. The Caco-2 cell line was grown in the same culture conditions but its growth medium was supplemented up to 20% FBS, according to the ATCC's culture specifications. Accordingly, NCI-H460, NCI-H226, and NCI-H1568 cell lines were maintained in RPMI-1640 (Corning, 10-040-CV) supplemented with 10% FBS (**Table 3.1**).

Cell densities were maintained below 80% confluence by subculturing them at 1:2-1:8 ratios. Considering that all cell lines used in this project are adherent, after aspiring the growth medium, cells were washed with 1x phosphate buffered saline (PBS) (Gibco, 10010023) and then detached using 0.25 % Trypsin-EDTA (Gibco, 25200-072). Finally, cells were collected by the addition of complete culture cell media to neutralize the activity of trypsin.

A panel of lung cancer cell lines was selected to have a representation of the different mutational backgrounds found in patients. Thus, A549 and NCI-H460 cell lines harbor activating *KRAS* mutations (G12S and Q61H, respectively), NCI-H1568 has an inactivating mutation in *TP53* (H179R), whereas NCI-H226 is wild-type for *KRAS* and *TP53* genes (**Table 3.1**). Regarding its biological origin, most cell lines were obtained from primary tumors, with the exception of NCI-H1568 and NCI-H226 cell lines, which came from metastatic sites.

Cell line	Tissue	Disease	Growth Medium	Mutations
A549	Lung	Carcinoma	DMEM + 10% FBS	*KRAS* pG12S, *SMARCA4* p.Q729Cfs*4
NCI-H460	Lung	Large cell carcinoma	RPMI-1640 + 10% FBS	*KRAS* pQ61H, *PIK3CA* pE545K *ARID1A* p.I2135_L2136del
NCI-H226	Lung	Squamous cell carcinoma	RPMI-1640 + 10% FBS	*PBRM1* p.Q1452*
NCI-H1568	Lung	Adenocarcinoma	RPMI-1640 + 10% FBS	*TP53* p.H179R, *EGFR* p.D1127N, *SMARCA4* p.E1449*, *ARID1B* p.R1808L
Caco-2	Colon	Colorectal adenocarcinoma	DMEM + 20% FBS	*CTNNB1* pG245A
SK-OV-3	Ovary	Adenocarcinoma	DMEM + 10% FBS	*TP53* p.S90Pfs*33, *PIK3CA* p.H1047R, *EP300* p.Y1414C

Table 3.1. Characterization of cell lines used in this doctoral thesis. The tissue of origin, the disease, the growth medium conditions, and the mutations present in each cell line are indicated.

1.2. Mycoplasma PCR

Mycoplasmas are small prokaryotic organisms that lack a cell wall, which makes them resistant to common antibiotics (e.g. penicillin or streptomycin), and their contamination is not directly visible under the microscope or by turbidity of the growth medium. For testing mycoplasma contamination, cells were cultured for at least 48 hours in the corresponding complete media of each cell line. 1 ml of cell culture supernatant was harvested and centrifuged to discard cell detritus. Next, 1 μl of the remaining supernatant was added to a PCR reaction mixture (KAPA2G Fast HotStart ReadyMix with dye (2X), Peqlab, KK5609) with specific oligonucleotides to amplify a 480-bp sequence of the gene encoding ribosomal RNA (**Table 3.2**).

Primer	Sequence
Myco_Fwd	GGCGAATGGGTGAGTAACACG
Myco_Rev	CGGATAACGCTTGCGACCTATG

Table 3.2. Nucleotide sequence of primers used to detect mycoplasma contamination.

PCR mixture was prepared in a final volume of 25 μl per reaction with 12.5 μl of a master mix (2x), 0.8125 μl of 10 μM oligonucleotides forward and reverse (reaching a final concentration of 0.325 μM), 9.875 μl of double-distilled water (ddH_2O) and 1 μl of the centrifuged supernatant of each cell culture to be tested. PCR reaction was performed as detailed in **Table 3.3**.

Temperature	Time	Cycles	Step
94 ºC	5 min	Hold	Initial denaturation
94 ºC 60 ºC 72 ºC	1 min 1 min 1 min	30	Denaturation Annealing Extension
72 ºC	5 min	Hold	Final extension
10 ºC	∞	Hold	Storage

Table 3.3. PCR reaction conditions to detect mycoplasma contamination using KAPA2G DNA polymerase.

PCR products were separated on a 2% agarose gel and revealed under a UV transilluminator (Bio-Rad, Gel Doc EZ Imager). The presence of a PCR product at around 500 bp was indicative of mycoplasma contamination.

In case cell lines were affected by mycoplasma contamination, treatment with 5 μg/ml of CPX and 10 μg/ml of tetracycline was initiated for at least 7 days. After this period of time, a new aliquot of the supernatant was taken to verify by PCR amplification if mycoplasma contamination had been efficiently removed.

1.3. Generation of stably-transduced cell lines

Stably-transduced cell lines were generated using doxycycline-inducible pTRIPZ constructs (Dharmacon, Horizon Discovery), which are lentiviral vectors that contain a short hairpin RNA (shRNA) targeting different subunits of the SWI/SNF complexes (**Figure 3.1**). In particular, constructs V3THS_343789, V3THS_343791, and V3THS_41001 were used for *ARID1A* knock-down, constructs V3THS_306688, V3THS_306689, and V3THS_306691 were selected to downregulate *ARID1B,* constructs V2THS_74399 and V3THS_347660 were used to downregulate *ARID2,* and finally constructs V3THS_317179,

V3THS_317182, and V3THS_317183 targeting *SMARCA4* were selected. The empty vector (RHS4750) was used to generate a control to which refer the results obtained in each cell line.

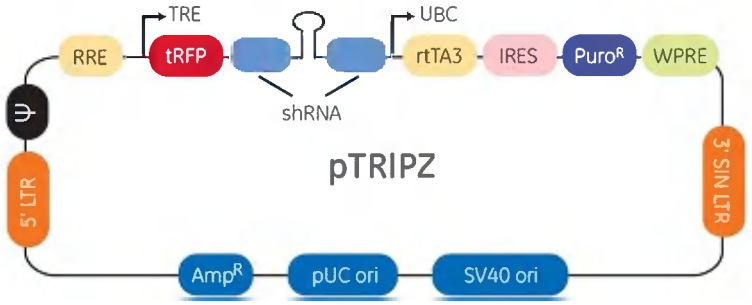


Figure 3.1. Detailed vector map of pTRIPZ. Schematic representation of the elements of the pTRIPZ lentiviral vector. This vector contains the sequence of an shRNA that interferes with the expression of a target gene. shRNA expression is controlled by an inducible promoter (tetracycline-inducible promoter, represented as TRE), activated by the addition of doxycycline. Additionally, it contains a reporter gene, which is a red fluorescent protein (turboRFP, represented as tRFP) that is transcribed in the same transcription unit as the shRNA, and a eukaryotic selection marker, which is a gene that confers resistance to the antibiotic puromycin (Puro^R). Figure source: http://dharmacon.horizondiscovery.com/uploadedFiles/Resources/ptripz-inducible-lentiviral-manual.pdf

Lentiviruses harboring the sequence of shRNAs were generated in the HEK 293T clone 17 (293T/17) packaging cell line. 293T/17 cells were seeded at 20% confluence and transfected with 2 μg of a given pTRIPZ construct, in addition to 1.5 μg of psPAX2 (Addgene, 12660) and 0.5 μg of pMD2.G (Addgene, 12259) plasmids, which are required for a correct packaging of the lentiviruses. Polyethylenimine (PEI) (1 mg/ml, Sigma-Aldrich, 408719l) was used as the transfection reagent in a 4:1 proportion (4 μl of PEI/1 μg of DNA).

The next day, the supernatant containing recombinant lentiviral particles was collected, filtered through a 0.45 μm pore size filter (FP 30/0.45 CA-S, Whatman™, 10462100), and mixed in 1:1 proportion with the corresponding growth medium used to maintain the cells that were going to be infected, adding polybrene (Santa Cruz Biotechnology, sc-134220) up to a final concentration of 8 μg/ml. In some cases, an additional centrifugation step at 134g for 45-60 min was carried out to

force the process of infection. Additionally, in order to achieve a higher infection titre, the same procedure was repeated the following day.

Transduced cells were selected adding puromycin to the medium, reaching a final concentration that ranged between 0,25-5 μg/ml, depending on the sensitivity of each cell line. The minimum amount of antibiotic required to efficiently kill all non-transduced cells had been previously assessed in each cell line. During the process of selection, which lasted at least 8 days, puromycin was refreshed whenever the medium was renewed or cells were subcultured. In addition to this, a negative control, consisting of non-infected cells treated with puromycin, was parallelly used to determine the end of the selection process.

Once the stable cell lines had been successfully selected, the amount of puromycin was reduced to a maintenance dose, which ranged from one-half to one-tenth of the selection dose, to allow antibiotic-selective pressure.

1.4. Induction of shRNA and turboRFP expression

With the aim of reaching an effective knock-down of *ARID1A*, *ARID1B*, *ARID2,* or *SMARCA4* the expression of the corresponding shRNA was induced by the addition of doxycycline (Sigma-Aldrich, D9891) to the medium for at least 5 days, reaching a final concentration of 1 μg/ml. Considering that the half-life of this inductor in cell culture is 24 hours, it was refreshed daily in order to avoid fluctuations in the level of expression of the shRNAs. Additionally, the induction process was followed in a fluorescence inverted microscope (Eclipse TS100, Nikon) checking the expression of turboRFP, which could be detected between 8-16 hours after induction.

1.5. FACS sorting of stably-transduced cell lines

In order to enrich the proportion of transduced cells, a cell sorting step was performed using a FACS-Aria II cell sorter (Becton Dickinson) and the PE channel. Once the cells transduced with shRNAs or empty vectors had been selected, the expression of turboRFP was induced by adding 1 μg/ml of doxycycline for at least 16 hours. Cells

were harvested, twice washed in phosphate-buffered saline (PBS), counted, and, resuspended in 0.2 µM-filtered sorting buffer containing 15 mM HEPES, 1% BSA (w/v), 1% penicillin/ streptomycin (v/v) (100 U/ml of each), and 2 mM EDTA in D-PBS (Ca^{2+} and Mg^{2+} free). Finally, cell suspensions of 5×10^6 cells/ml were prepared and filtered using a nylon mesh with a pore size of 70 µm (MACS® SmartStrainers (70 µm), Miltenyi Biotec, 130-098-462).

FACS sorting was used to purify the top 5% of the cells expressing the highest levels of turboRFP. For proper cell recovery after the sorting process, the cells were collected in polypropylene tubes containing their corresponding growth medium supplemented with 50% FBS to prevent dry out and cell death. Finally, the surviving cells were seeded and expanded in complete growth medium.

1.6. Proliferation assays

Growth inhibition assays were performed to determine the half-minimum inhibitory concentration (IC_{50}) values of different chemotherapeutic agents (etoposide: Sigma-Aldrich, E1383; and cisplatin: Sigma-Aldrich, 23210), PARP (veliparib (ABT-888): Selleckchem, S1004) and EGFR inhibitors (dacomitinib (PF-00299804): Selleckchem, S2727; and erlotinib: Santa Cruz Biotechnology, sc-202154).

In order to have the correct confluence for each experiment, cells were harvested and automatically counted using the TC20™ Automated Cell Counter (Bio-Rad, 1450102) following the manufacturer's instructions. In addition, the Neubauer chamber (Marienfeld, 0640111) was used for some experiments, together with Trypan Blue (Sigma-Aldrich, T8154) for checking cell viability.

Cells were seeded in 384-well plates at 1.6×10^4 cells/ml density. Once they were attached to the surface, drug-containing medium was added. Cell viability was determined after 48 hours of exposure to the chemical agent using CellTiter-Glo® Luminescent Cell Viability Assay (Promega, G7570), which is a luminescence-based method that measures intracellular ATP concentration as a signal of metabolically active cells.

IC_{50} values for each drug were calculated with Prism Software (GraphPad), using a non-linear regression model with dose-response inhibition that represents the logarithmic values of each concentration point against the normalized response.

1.7. Immunofluorescence

Before performing the immunofluorescence assays, cells were grown exponentially to 50-70% confluence on 10 mm diameter sterile cover glasses (VWR, ECN 631-1576). Additionally, poly-L-lysine (Sigma-Aldrich, P4707) was used to increase cell attachment to the surface of the coverslips.

Subsequently, cells were rinsed three times with PBS and fixed with fresh 4% paraformaldehyde in PBS for 15 min at room temperature. The cells were permeabilized using 0.5% Triton X-100 in PBS for 5 min at room temperature, and blocked with 3% BSA in PBT (PBS containing 0.05% Triton X-100) for 30 min. Finally, they were subjected to immunofluorescence staining with primary antibodies at the optimized dilution for at least 30 min, at room temperature, in a humidity chamber.

In particular, ARID2 (E3, Santa Cruz Biotechnology, sc-166117), γH2AX (anti-phospho-Histone H2A.X Ser 319, clone JBW301, Merck Millipore, 05-636-I), and 53BP1 antibodies (H-300, Santa Cruz Biotechnology, sc-22760) were used. After three washing steps with PBS for 5 min, cover glasses were incubated with a FITC-labeled secondary antibody (Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, FITC, Invitrogen, F-2761) for 30 min, at room temperature, in a humidity chamber protected from the light. Cells were three times washed with PBS for 5 min and finally cover glasses were mounted in ProLong™ Gold antifade reagent with DAPI (Invitrogen, P36935) to preserve fluorescence.

The cells were finally examined by fluorescence microscopy (Axio Imager M1, Zeiss). Colocalization of ARID2 with 53BP1 or γH2AX was performed by measuring the variation in the intensity across the lines drawn using the linescan tool from MetaMorph® (Molecular Devices). Automatic foci quantification was performed in randomly acquired images using a modified version of the speckle counting pipeline from CellProfiler open-source software (McQuin et al., 2018).

For DNA repair assays, cells were treated with 10 μM of etoposide for 1 hour. After that time, the medium was refreshed for up to 72 hours to allow cell recovery. In order to assess the effect of the chemotherapeutic agents on genomic stability, γH2AX immunostaining was used as a marker of DNA damage. DAPI staining was used to verify that γH2AX was specifically located in the nucleus. DNA-damage foci were quantified in randomly selected regions by computing the γH2AX signal using a Macro of Image J software.

2. Molecular biology methods

2.1. Nucleic acid extraction and purification

Total RNA was isolated and purified using NZY Total RNA isolation kit (Nzytech, MB13402) according to the manufacturer's instructions. β-mercaptoethanol was used as a reducing agent and cell lysates were further disrupted through a disposable 20G syringe (Henke Sass Wolf, 4710009025) to reduce viscosity. Additionally, to increase the yield and reduce contamination by the presence of genomic DNA, an enzymatic digestion with DNase I was performed. RNA concentrations were measured at 260 nm in a microvolume spectrophotometer (Thermo Fisher Scientific, ND-2000), and their spectres were checked to verify the absence of contaminating chemical compounds.

Genomic DNA was extracted from fresh cell lines using the Agencourt DNAdvance Beckman Coulter kit (Beckman Coulter, A48705) following the manufacturer's instructions. DNA preparations were quantified using the Qubit® dsDNA BR Assay (Life Technologies, Q32851).

2.2. Reverse transcription and quantitative polymerase chain reaction (qRT-PCR)

To accomplish the reverse transcription reaction, PrimeScript™ RT Reagent Kit (Takara, RR037B) was used, following the manufacturer's instructions. The reaction mixture was prepared in a final volume of 20 µl, containing 4 µl of 5x PrimeScript™ buffer, 1 µl of 50 µM oligo dT primer (25 pmol), 1 µl of 50 µM random 6-mers (50 pmol), and total RNA prepared in ddH_2O, which allows the generation of complementary DNA (cDNA) from up to 1 µg of total RNA per reaction. The final mixture was incubated under the conditions indicated in **Table 3.4.**

Temperature	Time	Step
37 ºC	15 min	Reverse transcription
85 ºC	5 s	Inactivation of reverse transcriptase with heat treatment
4 ºC	∞	Storage

Table 3.4. Reverse transcription program.

mRNA expression of the corresponding genes was measured by qRT-PCR using PowerUp™ SYBR Green Master Mix (Applied Biosystems, A25777) with the StepOnePlus™ Real-Time PCR system (Applied Biosystems). The reaction mixture was prepared in a final volume of 20 μl, containing 10 μl of PowerUp™ SYBER™ Green Master Mix (2x), 1.8 μl of 10 μM forward and reverse oligos (800 nM), 5.4 μl of ddH_2O and 1 μl of DNA template (diluted if necessary to ensure between 1-10 ng per reaction as starting material). PCR conditions are detailed in **Table 3.5**. A melting curve analysis was also included to verify that a single PCR product was obtained in each reaction.

PCR conditions				Melting curve
Temp	**Time**	**Cycles**	**Step**	
50 ºC	5 min	Hold	UDG activation	40x 95 ºC 15 s (1.6 ºC/s) 60 ºC 1 min (1.6 ºC/s) 95 ºC 15s (0.15 ºC/s)
95 ºC	2 min	Hold	Dual-lock DNA polymerase	
95 ºC 60 ºC	15 s 1 min	40	Denaturation Annealing and extension	

Table 3.5. qPCR conditions for primers with a Tm > 60 ºC and dissociation curve conditions.

As an internal control, gene expression was normalized to the housekeeping gene β-actin. The $\Delta\Delta C_T$ method (Livak and Schmittgen, 2001) was used for quantification and comparison.

$$\Delta C_T = C_{T(\text{gene of interest})} - C_{T(\text{housekeeping gene})}$$

$$\Delta\Delta C_T = \Delta C_{T(\text{SWI/SNF-deficient})} - \Delta C_{T(\text{shEmpty})}$$

$$RQ = 2^{-\Delta\Delta Ct}$$

In order to assess significant gene expression differences between cells transduced with shRNAs or empty vectors, a one-tailed unpaired *t*-test was performed, assuming equal variances and a 95% confidence interval.

The complete list of the primers used for the qRT-PCR reactions can be found in **Table 3.6**. qPCR primers were designed using the Primer 3 online software tool (https://primer3.ut.ee/), selecting parameters in line with the general PCR standards: length from 20 to 24 bp, CG content between 40-65 %, no secondary structures (confirmed using the online tool OligoAnalyzer, https://eu.idtdna.com/calc/analyzer), Tm around 60 ºC, and amplicon size between 100-150 bp. The specificity and efficiency of all primers were also tested.

Primer	Sequence
HsRT-ACTB-Fwd	CCCAGCACAATGAAGATCAA
HsRT-ACTB-Rev	CGATCCACACGGAGTACTTG
HsRT-ARID1A-Fwd	CCAGTAAGGGAGGGCAAGAA
HsRT-ARID1A-Rev	AGAGCTCCTTCTGTCCCCAT
HsRT-ARID1B-Fwd	GCGCAACAAAGGAGTCACC
HsRT-ARID1B-Rev	GCCCATGCCATACAACTGAG
HsRT-ARID2-Fwd	CAGCCCATAACTTTGACGCA
HsRT-ARID2-Rev	TGGTGCAATTCCATCTTCCT
HsRT-SMARCA4-Fwd	CGGCGGCTTCTTTGTTTCGT
HsRT-SMARCA4-Rev	GACATCTTCACGGGAGCTGC
HsRT_NRG1-Fwd	ACTCCTCTCTCAGATTCAAGTGGT
HsRT_NRG1-Rev	GGCACTGTCATTTCCTAATTTGCT
Hs_RT_SDK1_Fwd	CAGCTACCCCAGACCTCAAG
Hs_RT_SDK1_Rev	CCAGCTGATTCTCCAATGTG
Hs_RT_MTSS1_Fwd	TAAGAAAGCCCGCCAAGAGA
Hs_RT_MTSS1_Rev	TGGAGAGCACTGTCCAACTGA

Table 3.6. Nucleotide sequence of primers used for gene expression analysis.

2.3. Western blot

Total protein extracts were obtained by resuspending cellular pellets on kinase assay buffer containing 20 mM HEPES pH 7.5, 10 mM EGTA pH 8, 40 mM glycerol phosphate, 1% NP-40 (v/v), 2.5 mM $MgCl_2$, 2 mM orthovanadate, and 1 mM DTT. Protease (Halt™ protease inhibitor Cocktail (100x), Thermo Fisher Scientific, 78429) and phosphatase inhibitors (Phosphatase Inhibitor Cocktail 2, Sigma-Aldrich, P5726) were freshly added to keep the phosphorylated version of the proteins. Cell lysis was achieved by mechanical disruption of samples, which were kept on ice for 30 min, pipetting cell suspensions up and down every 10 min. In case the desired protein had a nuclear location or was firmly attached to the chromatin, a sonication step consisting of 10 cycles of 30s/30s (ON/OFF cycles) at high power was performed using the Bioruptor® (Dia-genode). Finally, a centrifugation step at 12,000 g for 15 min at 4 ºC was included to isolate the protein supernatants.

The protein concentration of each extract was determined using Qubit™ Protein Assay Kit (Life Technologies, Q33212). 50 µg of total protein lysate were mixed with 5x Laemmli buffer (250 mM Tris pH 6.8, 20% β-mercaptoethanol, 10% SDS, 0.25% bromophenol blue, and 50% glycerol) and heated at 95 ºC for 5 min. Then, samples were separated by SDS-PAGE in 8-10% polyacrylamide gels (depending on the molecular weight of the analyzed proteins) in a Mini-PROTEAN® Tetra System cuvette (Bio-Rad) with running buffer (1x Tris-Glycine pH 8.4, 1% SDS) at 130-180V (stacking and running gels, respectively).

Once the electrophoresis finished, proteins were transferred to nitrocellulose membranes (Amersham™ Protran™ Premium 0.45 mM NC, GE Healthcare Life Sciences, 10600003) by wet electroblotting in constant current conditions (400 mA for the desired time, assuming that 1 kDa of protein requires 1 min to move out the gel and attach to the membrane) using the transfer buffer (1x Tris-Glycine pH 8.4, 10% methanol).

Subsequently, membranes were washed with TBS-T buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% Tween 20) and blocked for 1-2 hours at room temperature, using 5 % BSA (Sigma-Aldrich, A9418) (w/v) solution in TBS-T as blocking agent. Membranes were then incubated with primary antibodies against pEGFR (9H2, Santa Cruz

Biotechnology, sc-57545), EGFR (A-10, Santa Cruz Biotechnology, sc-373746), pAKT (phospho-Akt(Ser473), Cell Signaling, #9271), AKT (H-136, Santa Cruz Biotechnology, sc-8312), ARID2 (E3, Santa Cruz Biotechnology, sc-166117), ARID1A (Abcam, ab176395), SMARCA4 (G-7, Santa Cruz Biotechnology, sc-17796), and Actin (I-19, sc-1616, Santa Cruz), at the recommended manufactured dilution in TBS-T containing 1 % BSA, in an orbital shaker (Biogen, DRS-12), at 4°C overnight.

Then, membranes were three times carefully washed in TBS-T. After that, membranes were incubated with HRP-conjugated secondary antibodies (goat anti-rabbit: Immun-Star Goat Anti-Rabbit (GAR)-HRP Conjugate, BioRad, 1705046; goat anti-mouse: Immun-Star Goat Anti-Mouse (GAM)-HRP Conjugate, BioRad, 1705047; and mouse anti-goat: mouse anti-goat IgG-HRP, Santa Cruz Biotechnology, sc-2354) at 1:10,000 dilutions and incubated for 1 hour at room temperature. Finally, membranes were again three times washed in TBS-T before visualizing chemiluminescence signals using the Amersham ImageQuant 800 system (Cytiva).

2.4. RNA-Seq library preparation

For RNA-Seq library preparation, 10 µg of total RNA were treated with DNase (TURBO DNA-free™ Kit, Ambion, AM1907) to remove any genomic DNA contamination. Additionally, profiles of eukaryotic ribosomal RNAs (18 and 28S) and RNA Integrity Number equivalent (RIN^e) (0-10) were assessed as a quality score, requiring a minimum value of 8, using an RNA ScreenTape (Agilent, 5067-5576). Then, between 1-5 µg of total RNA treated with DNase were subjected to an enrichment step in order to increase the proportion of mRNA in each sample, using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, E7490S).

mRNA was randomly fragmented by heating the samples at 94 ºC for 15 min. The first strand of cDNA was generated using the PrimeScript™ RT Reagent Kit (Takara, RR037B), incubating the samples at 37 ºC for 15 min, followed by a 5 s step at 85ºC to inactivate the reverse transcriptase. The synthesis of the second strand of cDNA was initiated by the addition of DNA polymerase I (Thermo Fisher Scientific,

EP0042), and the remaining molecules of RNA were degraded by adding RNase HI (Thermo Fisher Scientific, EN0201) to the previous mixture, which was incubated at 15 ºC for 2 hours. T4 DNA polymerase (Thermo Fisher Scientific, EP0062) was added to each sample to extend and complete the synthesis of the second strand, and an incubation step at 15 ºC for 5 min was performed before adding 5 µl of 0.5 M EDTA pH 8.0 to each sample.

The double-stranded cDNA fragments obtained were purified using Agencourt AMPure XP (Beckman Coulter, A63881). Different proportions of magnetic beads ranging from 0.8 to 1.8x (v/v) were used to isolate DNA fragments according to their sizes.

Sequencing libraries were prepared through a series of enzymatic steps. Firstly, DNA fragments were blunted and phosphorylated using an end-repair mix (Fast DNA End Repair Kit, Thermo Fisher Scientific, K0771), which contains an optimized mixture of T4 DNA polymerase (for the 5'→ 3' extension), Klenow fragment (for the removal of 3' overhangs using its 3'→5' exonuclease activity to form blunt ends), and T4 Polynucleotide Kinase (PNK) (for efficient phosphorylation of 5' ends). The resulting DNA fragments were adenylated on their 3' ends (Klenow Fragment exo-, Thermo Fisher Scientific, EP0422) before their ligation to paired-end adapters (T4 DNA Ligase, Thermo Fisher Scientific, EL0014) that had been previously obtained through the hybridization of two phosphorylated and complementary synthetic oligonucleotides (whose sequences are detailed in **Table 3.7**).

Adapter	Sequence
Adapt_TruSeq_P5a	CTACACGACGCTCTTCCGATCT
Adapt_TruSeq_P5b	GATCGGAAGAGCGTCGTGTAG
Adapt_TruSeq_P7a	GATCGGAAGAGCACACGTCTG
Adapt_TruSeq_P7b	CAGACGTGTGCTCTTCCGATCT

Table 3.7. Nucleotide sequence of P5 and P7 primers used to generate paired-end adapters compatible with Illumina sequencing platforms.

Finally, a PCR indexing amplification with a high-fidelity DNA polymerase (Phusion high fidelity DNA polymerase, Thermo Fisher Scientific, F530L) was performed to increase the amount of starting material and to add specific sequences in order to identify the reads from each sample. The specific PCR conditions used can be found in **Table 3.8**.

Temperature	Time	Cycles	Step
98 ºC	3 min	Hold	Initial denaturation
98 ºC 62 ºC 72 ºC	10 s 30 s 30 s	15	Denaturation Annealing Extension
72 ºC	5 min	Hold	Final extension
10 ºC	∞	Hold	Storage

Table 3.8. PCR indexing reaction conditions.

Different combinations of P5 and P7 indexes, shown in **Table 3.9** and **Table 3.10**, were used in order to pool and sequence simultaneously different samples.

Primer	Sequence
D501	AATGATACGGCGACCACCGAGATCTACAC**TATAGCCT**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D502	AATGATACGGCGACCACCGAGATCTACAC**ATAGAGGC**ACACTCT TTCCCTACACGACGCTCTTCCGATCT
D503	AATGATACGGCGACCACCGAGATCTACAC**CCTATCCT**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D504	AATGATACGGCGACCACCGAGATCTACAC**GGCTCTGA**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D505	AATGATACGGCGACCACCGAGATCTACAC**AGGCGAAG**ACACTCT TTCCCTACACGACGCTCTTCCGATCT
D506	AATGATACGGCGACCACCGAGATCTACAC**TAATCTTA**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D507	AATGATACGGCGACCACCGAGATCTACAC**CAGGACGT**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D508	AATGATACGGCGACCACCGAGATCTACAC**GTACTGAC**ACACTCTT TCCCTACACGACGCTCTTCCGATCT

Table 3.9. Nucleotide sequence of P5 oligonucleotides used during RNA-Seq library preparation. Specific indexing 8-bp sequences used to identify each sample are remarked in bold.

Primer	Sequence
D701	CAAGCAGAAGACGGCATACGAGAT**CGAGTAAT**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D702	CAAGCAGAAGACGGCATACGAGAT**TCTCCGGA**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D703	CAAGCAGAAGACGGCATACGAGAT**AATGAGCG**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D704	CAAGCAGAAGACGGCATACGAGAT**GGAATCTC**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D705	CAAGCAGAAGACGGCATACGAGAT**TTCTGAAT**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D706	CAAGCAGAAGACGGCATACGAGAT**ACGAATTC**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D707	CAAGCAGAAGACGGCATACGAGAT**AGCTTCAG**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D708	CAAGCAGAAGACGGCATACGAGAT**GCGCATTA**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D709	CAAGCAGAAGACGGCATACGAGAT**CATAGCCG**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D710	CAAGCAGAAGACGGCATACGAGAT**TTCGCGGA**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D711	CAAGCAGAAGACGGCATACGAGAT**GCGCGAGA**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D712	CAAGCAGAAGACGGCATACGAGAT**CTATCGCT**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT

Table 3.10. Nucleotide sequence of P7 oligonucleotides used during RNA-Seq library preparation. Specific indexing 8-bp sequences used to identify each sample are remarked in bold.

The resulting PCR products were purified using Agencourt AMPure XP (Beckman Coulter, A63881) and eluted in 20 µl of low TE buffer (Thermo Fisher Scientific, 12090015). Size distribution of the RNA-Seq libraries obtained was analyzed using D1000 (Agilent, 5067-5582) or High Sensitivity D1000 ScreenTape Assays (Agilent, 5067-5584), depending on their concentration, which was quantified using the Qubit® dsDNA BR Assay Kit (Life Technologies, Q32851).

Before sequencing the RNA-Seq libraries, a quality control qPCR reaction was performed to determine the efficiency of the adaptor ligation in the preparations, using the Illumina universal sequencing primers indicated in **Table 3.11**. A fragment that contained the complete length of the PE Illumina adapters was used as a positive control.

Primer	Sequence
QC_Fwd	AATGATACGGCGACCACCGAGA
QC_Rev	CAAGCAGAAGACGGCATACGAG

Table 3.11. Nucleotide sequence of primers used to amplify a common region of PE Illumina adapters.

Finally, equimolecular amounts of each sample were mixed to obtain a minimum of 60-80×10^6 of paired-end reads per sample. The sequencing of the resultant pool was carried out in a HiSeq, NextSeq, or NovaSeq platforms from Illumina, depending on the required number of reads per reaction, with a 75-150 bp paired-end protocol. Three biological replicates independently generated were prepared for each sample group (shEmpty or shRNAs targeting different subunits of SWI/SNF in different cell lines) to assess the statistical significance of all differences.

2.5. ATAC-Seq library preparation

ATAC-Seq libraries were generated using the previously described method by the Greenleaf lab (Buenrostro et al., 2015). Stably-transduced cell lines were harvested, twice washed in PBS, and counted.

For each sample, 50×10^4 cells were resuspended in 50 µl of cold lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% IGEPAL, CA-630). Cell suspensions were centrifuged immediately at 500g for 10 min at 4 °C. Nuclei pellets were resuspended in transposase mixture (25 µl of Tagment DNA Buffer, 2.5 µl of Tagment DNA Enzyme I (Nextera DNA Library Preparation Kit, Illumina, 15028220) and 22.5 µl of ddH_2O). Next, cell nuclei were incubated at 37 ºC for 30 minutes to allow DNA cleavage and tagged (tagmentation). During this reaction, Tn5 transposase is able to bind to accessible chromatin regions, which become flanked by the addition of Illumina adapters.

Transposed DNA was purified using the MinElute Reaction Cleanup Kit (QIAGEN, 28004) and eluted in 10 µl of elution buffer. Finally, a PCR indexing amplification under the same conditions described in "RNA-Seq library preparation" was followed (**Table 3.8**), using Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific, F530L), and the customized Nextera PCR primers designed by the Greenleaf lab (Buenrostro et al., 2015). Different combinations of these indexes allow the recognition of samples during the sequencing reaction (**Table 3.12**).

Size distribution, concentration, and quality of ATAC-Seq libraries were determined as previously described in the "RNA-Seq libraries preparation" section.

Primer	Sequence
Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCG GCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGAT**TCGCCTTA**GT CTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGAT**CTAGTACG**GT CTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGAT**TTCTGCCT**GTC TCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGAT**GCTCAGGA**GT CTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGAT**AGGAGTCC**GT CTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGAT**CATGCCTA**GT CTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGAT**GTAGAGAG**GT CTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGAT**CCTCTCTG**GTC TCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGAT**AGCGTAGC**GT CTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGAT**CAGCCTCG**GT CTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGAT**TGCCTCTT**GTC TCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGAT**TCCTCTAC**GTC TCGTGGGCTCGGAGATGT

Table 3.12. Nucleotide sequence of Nextera oligonucleotides used during ATAC-Seq library preparation. Specific indexing 8-bp sequences used to identify each sample are remarked in bold.

2.6. CUT&RUN library preparation

CUT&RUN experiments were prepared as previously described by the Henikoff lab (Skene et al., 2018), starting from 0.1-1×10^6 cells per condition. As this protocol is performed with fresh cells, it is advisable to follow the steps prior to the addition of the antibodies at room temperature to minimize the stress on the cells and to avoid DNA breakage.

Fresh cultured cells were harvested and counted. Two centrifugation steps with wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and one Roche Complete Protease Inhibitor (EDTA-free) tablet (Sigma-Aldrich, 5056489001)) at 600g for 3 min and at room temperature were done.

Concanavalin-A coated beads (BioMag®Plus Concanavalin A, Quimigen, 86057-10) were activated with binding buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, and 1 mM $MnCl_2$). 10 µl of bead suspension was added to each sample, which was then gently rotated (C1 program, 20 rpm, 10 min) at room temperature using an Intelli-Mixer RM-2 Rotator (Vortex Shaker, FV-51202).

Antibodies solutions were prepared by mixing 2 ml of digitonin buffer (wash buffer containing 5% of digitonin) with 8 µl of 0.5 M EDTA and the corresponding volume of the antibody to obtain a final concentration of 1:50-1:100, according to the manufacturer's recommendations. In particular, H3K27me3 (Tri-Methyl-Histone H3(Lys27) (C36B11), Cell Signaling, 9733), H3K4me1 (Mono-Methyl-Histone H3(Lys4) (D1A9), Cell Signaling, 5326), H3K4me3 (histone H3K4me3, Active Motif, 39159), and H3K27ac (Anti-Histone H3 (acetyl K27) antibody, Abcam, ab4729) antibodies were used to characterize the epigenetic landscape of the stably-transduced cell lines.

Cellular pellets were resuspended in the antibody solution and placed on the rotator (60 mode, 34 rpm, 10 min) at room temperature. Subsequently, they were permeabilized by resuspension on digitonin buffer to allow the diffusion of the primary antibody.

Purified and tested protein A-MNase (pA-MN) fusion protein (obtained from pK19pA-MN plasmid, Addgene, 86973) was added to each sample. To increase the binding of the nuclease to the primary antibody, samples were again inverted by rotation. Target digestion of

pA-MN was reached by adding 3 µl of 100 mM $CaCl_2$ and incubating the samples for at least 30 min at 0 ºC. Cleavage was finished by adding 100 µl of 2x stop solution (333.3 mM NaCl, 19.6 mM EDTA, 3.92 mM EGTA, 5% digitonin, 0.05 mg/ml RNase A, and 0.05 mg/ml glycogen).

Specific protein-DNA complexes were released by incubating the samples at 37 ºC for 10 min and 500 rpm, followed by an additional centrifugation step at 16,000g for 5 min at 4 ºC.

Finally, CUT&RUN fragments were purified using a magnetic rack (DynaMag Magnet, Thermo Fisher Scientific, 12321D).

Eluted DNA was purified using the NucleoSpin kit (Macherey-Nagel, 740609.50), following the manufacturer's instructions. CUT&RUN libraries were prepared using the Accel-NGS 2S Plus DNA library kit (Swift Biosciences, 21024), according to the manufacturer's instructions, and sequenced on an Illumina NovaSeq 6000, with a 100 bp paired-end protocol, at the Institut Curie Next Generation Sequencing Core Facility, requiring a minimum of 20-40×10^6 of paired-end reads per condition.

A summary of the NGS libraries prepared in the different stably-transduced cell lines generated in this doctoral thesis can be found in **Table 3.13**.

Cell line	NGS libraries
A549 shARID1A	RNA-Seq, ATAC-Seq
A549 shARID1B	RNA-Seq, ATAC-Seq
A549 shARID2	RNA-Seq, ATAC-Seq
H460 shARID1A	RNA-Seq, ATAC-Seq, CUT&RUN-Seq (H3K27ac, H3K4me1, H3K4me3, H3K27ac)
H460 shARID2	RNA-Seq, ATAC-Seq, CUT&RUN-Seq (H3K27ac, H3K4me1, H3K4me3, H3K27ac)
H460 shSMARCA4	RNA-Seq, ATAC-Seq, CUT&RUN-Seq (H3K27ac, H3K4me1, H3K4me3, H3K27ac)
H460 shARID2 daco	RNA-Seq after 24 and 48 hours of exposure to EGFR inhibitor dacomitinib
Caco-2 shARID1A	RNA-Seq, ATAC-Seq
Caco-2 shARID1B	RNA-Seq, ATAC-Seq
SK-OV-3 shARID1A	RNA-Seq

Table 3.13. NGS libraries prepared on SWI/SNF-deficient cell lines.

3. Bioinformatic methods

3.1. DNA sequencing data analysis

Raw sequence data were subjected to quality control using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), removing those reads that did not accomplish the quality criteria. Then, reads were mapped to the reference human genome (GRCh38 release 84) using BWA-MEM (Li and Durbin, 2009). Format transformation, sorting, and indexing of the bam files were done with SAMTools (Li et al., 2009). The alignment was fixed and cleaned, and PCR duplicate reads were marked and excluded from the analysis with Picard (http://broadinstitute.github.io/picard/). Finally, GATK (McKenna et al., 2010) was used to perform local realignment around indels. Bedtools (Quinlan and Hall, 2010) was used to calculate the enrichment statistics and the target coverage.

Exome sequencing results from the TCGA-LUAD project available on March 2022 were downloaded from the public portal (http://www.tcgaportal.org/TCGA/Lung_TCGA_LUAD/index.html), non-synonymous mutations were found with Mutect2 (Cibulskis et al., 2013), and maftools R library was used to create the maf object and to plot the somatic variants.

3.2. RNA sequencing data analysis

Paired-end reads from RNA-Seq experiments were aligned against the human reference genome (GRCh38 release 84) using Hisat2 (Kim et al., 2019). Predict transcripts from the Ensembl database were analyzed and transcripts that would lack a CDS start or stop site were filtered out. The number of counts of each gene was quantified using HTSeq (Anders et al., 2015) and the GRCh38 human genome assembly (annotation GENCODE v96). Differentially expressed genes (DEGs) between SWI/SNF-deficient cells and their respective control were identified using DESeq2 (Love et al., 2014). These R packages for transcriptome expression profile analysis were used according to the software developer's indications, requiring a minimum of three counts per gene in more than two independent samples to be considered as expressed. The results obtained were filtered requiring a minimum

absolute logarithmic value of fold change greater than 0.5 and an adjusted *p*-value of less than 0.05 ($|\log_2 FC| > 0.5$ & adjusted *p*-value < 0.05). The set of genes that accomplish both conditions was manually reviewed and the final list of DEGs was created. Heatmaps showing a selection of differentially expressed genes were obtained using the pheatmap package developed by R (https://cran.r-project.org/web/packages/pheatmap/index.html).

Gene Set Enrichment analysis (GSEA) (Subramanian et al., 2005) was performed to identify significant enrichments at the molecular pathway level using a permutation type based on phenotype distribution, the normalized counts of all genes expressed in the different conditions tested, and a hallmark gene set (h.all.v2022.1.Hs.symbols.gmt). Those pathways with a nominal *p*-value below 0.05 and a false discovery rate (FDR) q-value cutoff of less than 0.25 were considered significantly altered. The FDR is an estimation of the probability that a given gene set with a particular enrichment score (normalized for gene set size) constitutes a false positive. An FDR of 25% means that the result is likely to be valid three out of four times. Considering the different sizes of the gene sets, the small number of samples in each biological group, the variability of the inactivation of the target gene, and the type of permutation, we selected the top 20 positively enriched pathways to not overlook potential hypotheses that in any case require further validation in future research.

In addition to this, GOrilla (http://cbl-gorilla.cs.technion.ac.il/) (Eden et al., 2009) and AmiGO2 (http://amigo.geneontology.org/amigo) were used to identify enriched Gene Ontology (GO) terms (biological process, molecular function, or cellular component) using a target list of genes compared to a background list of genes.

3.3. ATAC and CUT&RUN sequencing data analysis

ATAC-Seq and CUT&RUN reads were aligned against the reference human genome (GRCh38 release 84) using BWA-MEM (Li and Durbin, 2009). Accessible regions were identified using MACS3 (Zhang et al., 2008). In the case of the histone epigenetic marks, some parameters of the analysis (such as the expected size of the peak and the statistical cut-off) had to be adjusted due to the distribution of the epigenetic mark (e.g. H3K27me3 has broad chromatin peaks, whereas H3K4me1 peaks tend to be sharper). A similar strategy was followed to identify general changes in chromatin or at specific regulatory regions (enhancers and promoters):

(1) To assess global changes in the complete genome, a combined list of all accessible or positive regions for the corresponding histone modification mark identified in any of the tested samples was created.

(2) The genomic positions of promoters were obtained assuming that they are located close to the transcription start site (TSS) of their corresponding gene, selecting a region of ±1 kb around the TSS.

(3) The complete list of enhancers identified in the GeneHancer project (Fishilevich et al., 2017), which includes regulatory elements identified in all types of tissues, was used as an input in case no histone epigenetic marks datasets were available for the selected cell line.

(4) In case histone epigenetic marks datasets had been generated, ChromHMM software (Ernst and Kellis, 2017) was used to characterize chromatin states corresponding to enhancers. For that purpose, histone modifications datasets (H3K27ac, H3K27me3, H3K4me1, and H3K4me3) from all tested samples (SWI/SNF-deficient and their corresponding controls) were included, and the segmentation was based in 12 states. According to the epigenetic pattern and the candidate state annotation described by the Kellis lab (Ernst and Kellis, 2017),

four different regions were identified as possible enhancers with different biological outcomes (**Table 3.14**).

Region	Description	Chromatin characterization
Enhancer_1	Active enhancer 1 (state #9)	↑ H3K27ac, ↑ H3K4me1
Enhancer_2	Active enhancer 2 (state #10)	↓ H3K27ac, ↓↓H3K4me1
Enhancer_3	Weak/primed enhancer (state #11)	↓ H3K27ac, ↑ H3K4me1
Enhancer_4	Bivalent/poised enhancer (state #15)	↑ H3K27me3, ↑ H3K4me1

Table 3.14. Types of enhancers identified using ChromHMM. The name, description (candidate state annotation and the corresponding chromatin state identified by the Kellis lab), and histone epigenetic pattern of each enhancer are indicated.

Once the complete list of genomic positions identified in all samples had been characterized, a significant region enrichment analysis was performed in SWI/SNF-deficient cells against control cells using DESeq2 (Love et al., 2014).

Region annotation and visualization of their genomic distribution according to their distance to the nearest transcription start site were performed using the ChIPSeeker software (Yu et al., 2015). The overlapping of genomic regions with altered accessibility or histone modifications was assessed using BEDTools (Quinlan and Hall, 2010). Motif enrichment analyses were performed using HOMER (Heinz et al., 2010). In case information of specific histone epigenetic marks had not been generated in our laboratory, ENCODE publicly available ChIP-seq datasets (https://www.encodeproject.org/data-standards/chip-seq/) were downloaded and compared with our ATAC-Seq data, assuming that accessible regions are likely to carry active histone modifications. The overlapping results were plotted using deepTools2 (Ramírez et al., 2016). TADs genomic positions were downloaded from a publicly available database (http://3dgenome.fsm.northwestern.edu/view.php). Finally, read alignments were visualized using the IGV genome browser (Robinson et al., 2017).

3.4. Statistical analysis

In all cases, at least three independent experiments were performed in order to assess the statistical significance of all differences. In figure legends, the specific statistical test performed in each case is indicated. Results are represented as the mean of independent biological replicates with error bars representing the standard error of the mean (SEM). In general, for quantitative variables, a two-tailed *t*-test assuming equal variance was used to identify significant differences between control and experimental groups. For qualitative variables, a two-tailed Fisher's exact test was used to identify significant differences between groups. Generally, a *p*-value of less than 0.05 was considered significant and represented in the graphs with an asterisk (*), with the following criteria: *$*p < 0.05$, $**p < 0.01$,* and *$***p < 0.001$*.

RESULTS

“All outstanding work, in art as well as in science, results from immense zeal applied to a great idea.”
– Santiago Ramón y Cajal

1. Characterization of the molecular mechanisms behind the role of SWI/SNF alteration in lung cancer development

1.1. Transcriptional alterations found in SWI/SNF-deficient A549 cell lines

Loss of different subunits of SWI/SNF complexes has been linked with tumor progression and worse prognosis in animal models (Moreno et al., 2021) and lung cancer patients (Bell et al., 2016; Matsubara et al., 2013; Zhang et al., 2014). In addition to this, accessory subunits of the complex have been proposed to mediate the targeting, the assembly, and the regulation of specific gene networks (Mittal and Roberts, 2020).

Therefore, in order to unravel the molecular mechanisms behind the contribution of SWI/SNF alteration to the development of lung cancer, we knocked down the expression of the most commonly mutated accessory subunits (*ARID1A*, *ARID1B,* or *ARID2*) (**Figure 4.1**) in the lung cancer cell line A549, and we performed RNA-Seq experiments.

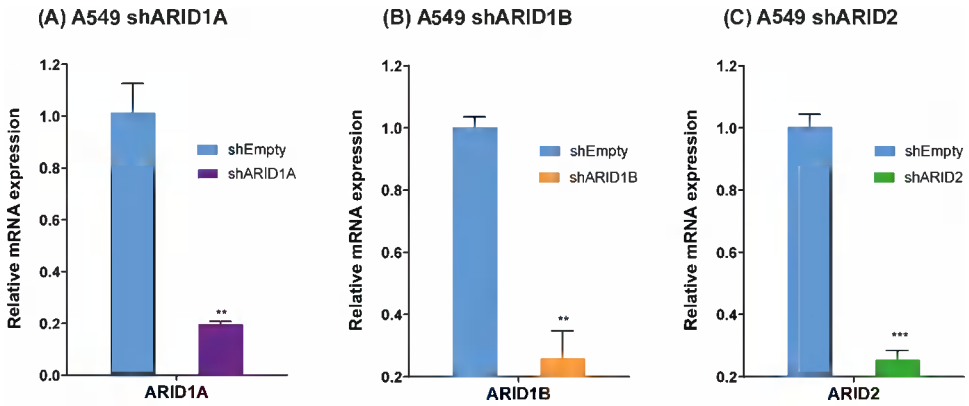


Figure 4.1. Knock-down validation of A549 stably-transduced cell lines by qRT-PCR. Representative figures of the relative mRNA expression for the inactivated genes in A549 shARID1A (**A**), shARID1B (**B**), and shARID2 (**C**) cellular models. The results are represented as mean ± SEM of at least three independent experiments (one-tailed *t*-test **p < 0.05, **p < 0.01,* and ****p < 0.001*).

The analysis of the results showed that in general, SWI/SNF deficiency was accompanied by profound transcriptional changes in the cells (**Figure 4.2.A**). In particular, cBAF alteration (*ARID1A* or *ARID1B* deficiency) resulted in a higher proportion of overexpressed genes (shARID1A: 1393 upregulated and 540 downregulated; shARID1B: 1719

upregulated and 1135 downregulated), whereas PBAF disruption (*ARID2* deficiency) was associated with a greater proportion of underexpressed genes (366 upregulated and 789 downregulated).

Consequently, cBAF alteration resulted in a higher number of differentially expressed genes (DEGs) compared to the PBAF-altered model (DEGs shARID1A: 1933, shARID1B: 2854, shARID2:1155). These findings suggest that in the lung cancer cell line A549, cBAF complexes are essential to keep the expression of a higher number of genes, whereas PBAF complexes might in turn regulate the transcription of a more reduced set of genes.

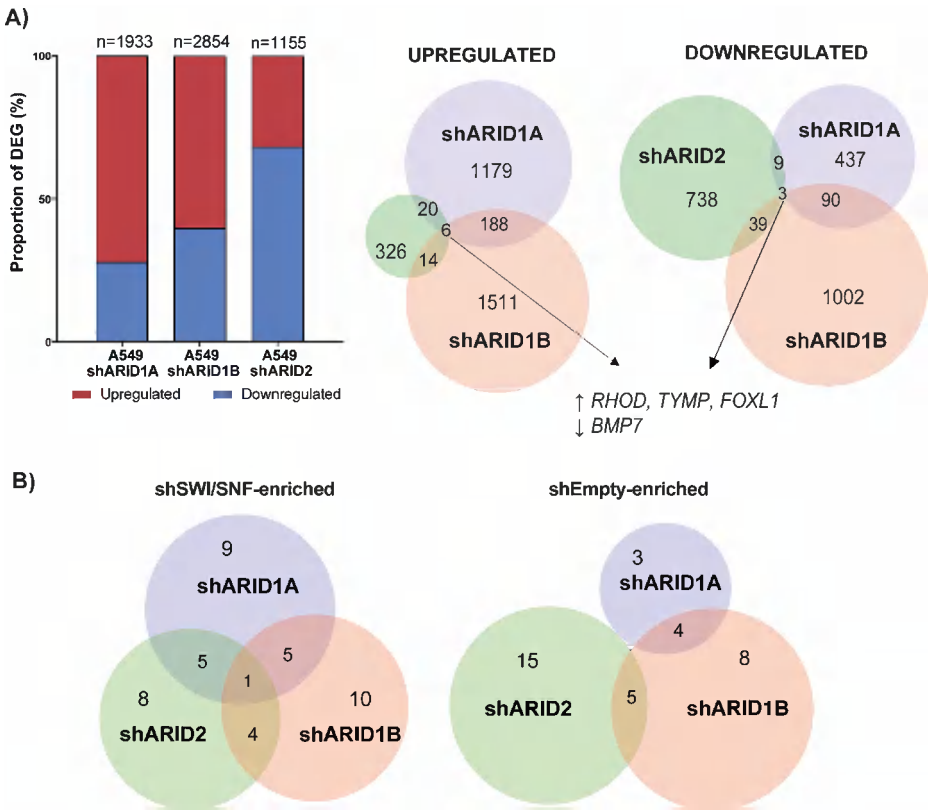


Figure 4.2. Transcriptomic landscape of SWI/SNF-deficient A549 cell lines. A) Proportion of DEGs reported in *ARID1A*, *ARID1B,* or *ARID2*-KD A549 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each A549 deficient condition, shARID1A (purple), shARID1B (orange), or shARID2 (green), separated according to their fold change. **B)** Venn diagram showing the results of Gene Set Enrichment Analysis (GSEA) in SWI/SNF-deficient A549 cell lines. On the left, there is a representation of the overlap between the top 20 gene sets positively enriched

in *ARID1A* (purple), *ARID1B* (orange), or *ARID2* (green)-deficient A549 stably-transduced cells, whereas those pathways enriched in the control cells are represented on the right.

A complete list of the normalized counts and the results of the differential gene expression analysis performed in each SWI/SNF-deficient cell line can be found in **Supplementary Table 1**. To allow comparisons between the different models, only the commonly expressed genes in all models have been compiled.

In addition to this, the comparison of the list of DEGs obtained in each condition revealed specific gene set regulation networks controlled by ARID1A, ARID1B, or ARID2-containing SWI/SNF complexes, evidenced by the presence of a very low number of shared genes (6 upregulated and 3 downregulated) (**Figure 4.2.A**). This finding suggests that cBAF and PBAF complexes are required to regulate the expression of different gene expression programs in the context of lung cancer.

It is quite remarkable that *ARID1A* and *ARID1B*-deficient cells share a higher proportion of DEGs (194 upregulated and 93 downregulated) than other comparisons. Regarding that both accessory subunits are considered to be mutually exclusive in cBAF complexes, the expression of those genes that are specifically controlled by *ARID1A* or *ARID1B* would be controlled by the remaining paralog. Therefore, having found commonly dysregulated genes in the absence of either *ARID1A* or *ARID1B* suggests the existence of cBAF-mediated gene networks whose expression requires the presence of both ARID1A and ARID1B-containing complexes.

In order to check if the altered genes could be organized in similar molecular functions, we performed functional analysis with Gene Set Enrichment Analysis (GSEA) software in each knock-down (KD) cell line using the hallmark gene collection (Liberzon et al., 2015) (**Figure 4.2.B**). Due to the small number of samples in each biological condition, the small size of the gene sets and/or the variability in SWI/SNF knock-down, we considered the top 20 positively enriched pathways of the GSEA analysis, although in most cases, this enrichment was not statistically significant. Nevertheless, in order to remove false positives, we only considered the commonly enriched pathways in the different cellular models for further studies.

Beginning with the cBAF complexes, we observed an enrichment of angiogenesis, upstream *KRAS*, apical junction, and response to interferon signaling pathways in *ARID1A* or *ARID1B*-KD cells compared to their respective controls. In some cases, the upregulated genes belonging to each functional category were conserved between subunits, whereas in others we found different transcriptional alterations that could be grouped in the same molecular pathway.

Thus, among the shared alterations reported, genes encoding proteins involved in cell adhesion and motility (*TSPAN1, NRP1, CD37, SLIT2,* and *CNTN1*), together with extracellular matrix remodeling (*COL16A1, ADAM23, MMP11, ITGB1, LAMC2, LAMA3, LAMB3, SDC3, CDH3, FBN1,* and *SKAP2*) stood out. However, we found specific gene expression changes in *ARID1A*-deficient cells related to KRAS signaling (*VEGFA* and *IGF2*) and response to interferon pathways (*TNFSF10, CISH, LTB, IFITM3, FLT3LG, NFKBIZ, LIF, CSF1,* and *IER3*). Similarly, *ARID1B*-deficient cells presented other transcriptional alterations of components belonging to the KRAS signaling (*PDGFA, BTC, EGFR,* and *SLIT2*) and immune response pathways (*GBP4, IFIT3, IFITM2, IFITM3, IFI44, IRF7, IL7,* and *IL15*).

We also looked for shared DEGs and potential signaling pathways regulated by cBAF and PBAF complexes in this particular lung cancer cell model. Between those genes commonly upregulated in *ARID1A*, *ARID1B,* or *ARID2*-deficient cells, it can be highlighted *RHOD,* a Ras homolog involved in cytoskeleton reorganization, *TYMP,* an angiogenic factor, and *FOXL1,* a transcription factor involved in cell proliferation. Regarding the shared repressed genes, *BMP7,* which is a ligand of TGF-β known to reverse the epithelial to mesenchymal transition, stands out (Zeisberg et al., 2003) (**Figure 4.2.A**).

Concerning the positively enriched signaling pathways in all SWI/SNF-deficient cells, the epithelial to mesenchymal transition (EMT) was among the main potentially altered pathways. In particular, we found commonly upregulated genes in all cellular models (*TWIST2, ECM1, LAMC2, LAMA3,* and *VEGFC*), but also specific transcriptional alterations for the deficiency of *ARID1A* (*SNAI3, MMP15,* and *MMP25*), *ARID1B* (*TGFB1, CDH2, SNAI2, FOXC2, JUN, MMP2, SPOCK1, COL5A1,* and *FN1*) or *ARID2* (*AREG, VIM, SDC1, COL5A3,* and *RHOB*) (**Figure 4.3**).

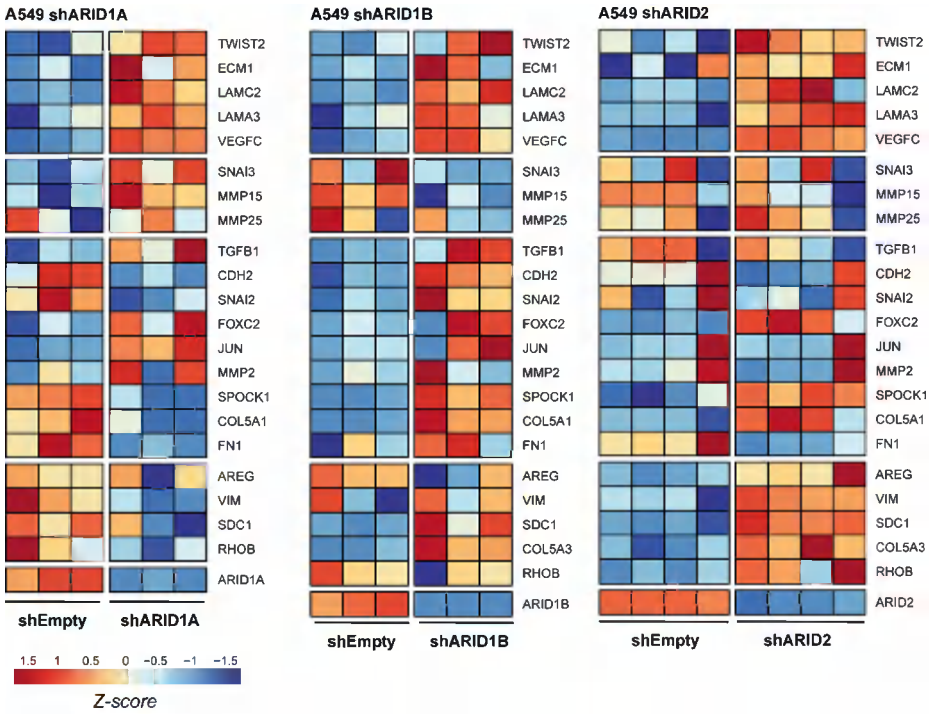


Figure 4.3. Positive enrichment of EMT in SWI/SNF-deficient A549 cell lines. Heatmap representation of a selection of upregulated genes related to the epithelial to mesenchymal transition in SWI/SNF-deficient A549 cells. Gene expression differences between the absence of different subunits of the complex (shARID1A, shARID1B, or shARID2) and shEmpty cells (n=3 in each condition) have been represented according to the log2 of the fold change and have been indicated by colors, which range from red (upregulation) to blue (downregulation).

Taken together our results indicate that in the same cellular context, SWI/SNF complexes mediate the expression of different transcriptional programs depending on their composition. However, we have observed some similarities in terms of signaling pathways controlled by each subunit, as is the case of EMT, which might indicate conserved mechanisms mediating gene expression regulation by the distinct compositions of SWI/SNF complexes.

1.2. Transcriptional alterations across different cellular contexts

Considering the low amount of shared DEGs obtained after removing the expression of different subunits of the complex in the same cell line, we hypothesize that SWI/SNF-mediated gene networks might be subunit-specific. Therefore, in order to figure out if those genes were maintained in different cellular contexts, we decided to investigate the transcriptional alterations found after the deficiency of the same accessory subunit of the SWI/SNF complex in cellular models obtained from different tumor types.

To that end, we extended our collection of stably-transduced cell lines to other cell types with recurrent SWI/SNF mutations. We selected an ovarian (SK-OV-3) and a colorectal (Caco-2) cancer cell lines, where we knocked down the expression of *ARID1A*. Moreover, we also downregulated *ARID1B* in the colorectal cancer cell line Caco-2. We prepared RNA-Seq experiments on these additional models in order to compare the results with the ones generated in the A549 cell line.

ARID1A deficiency in different cellular contexts resulted in a general transcriptional activation (**Figure 4.4.A**), as we reported a higher proportion of upregulated genes compared to the downregulated ones (SK-OV-3: 55 upregulated and 39 downregulated; Caco-2: 1249 upregulated and 1044 downregulated; A549: 1393 upregulated and 540 downregulated).

Similar results were obtained following *ARID1B* knock-down in different tumor types (A549: 1719 upregulated and 1135 downregulated; Caco-2: 419 upregulated and 396 downregulated) (**Figure 4.4.B**).

The comparison of the list of DEGs obtained in each SWI/SNF-deficient cell line showed again a very low overlap, as only 3 genes were commonly dysregulated in *ARID1A*-deficient cells (2 upregulated and 1 downregulated) and 53 in the case of *ARID1B*-deficient cells (30 upregulated and 23 downregulated). Altogether these findings indicate that ARID1A and ARID1B-containing complexes regulate the expression of different gene networks, depending also on the cellular context.

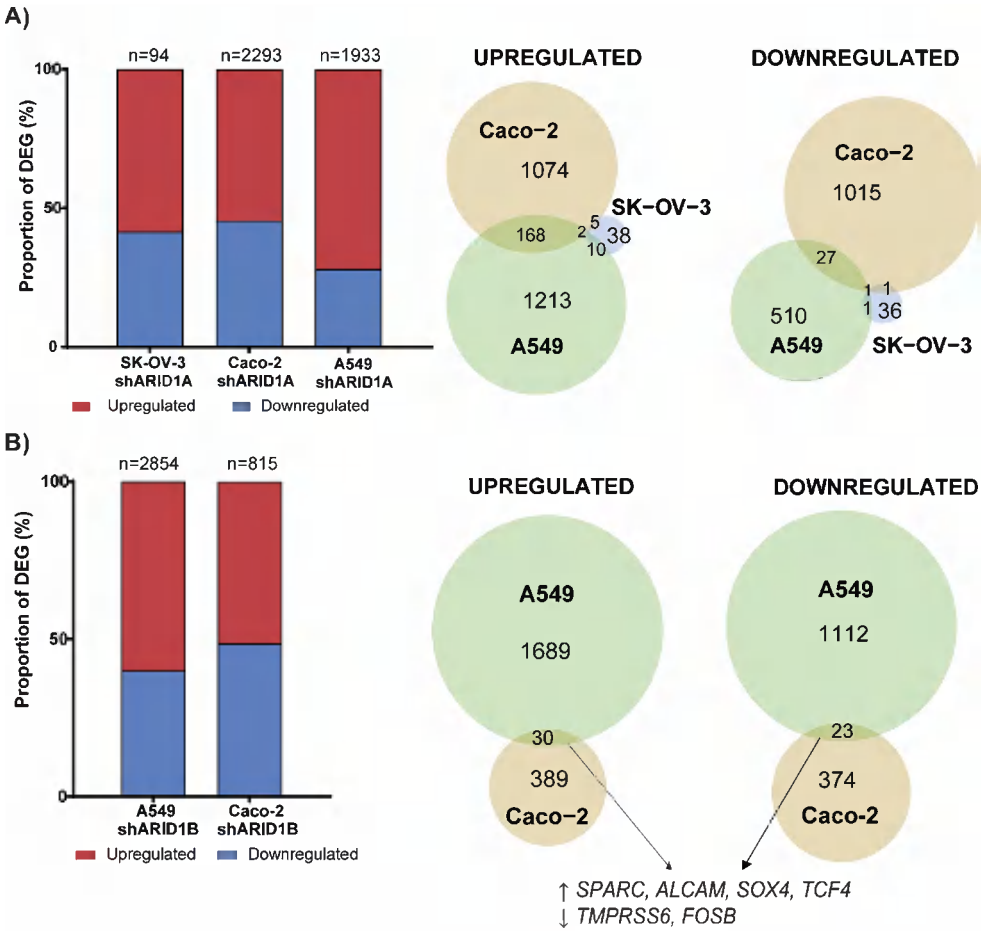


Figure 4.4. Transcriptomic landscape of *ARID1A* or *ARID1B*-deficient cell lines. A) Proportion of DEGs reported in *ARID1A*-KD SK-OV-3, Caco-2, and A549 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each *ARID1A*-deficient condition, SK-OV-3 (blue), Caco-2 (yellow), or A549 (green), separated according to their fold change. **B)** Proportion of DEG reported in *ARID1B*-KD A549 and Caco-2 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each *ARID1B*-deficient condition, A549 (green) or Caco-2 (yellow), separated according to their fold change.

Although the vast majority of transcriptional alterations tend to be tissue-specific, among the commonly dysregulated genes in *ARID1B*-deficient cells, we can highlight the upregulation of *SPARC* and *ALCAM*, which are molecules involved in cell migration, together with the repression of the adhesion protein *TMPRSS6,* and the dysregulation of transcription factors involved in development *SOX4*, *FOSB*, and *TCF4*.

Besides the very low overlap achieved at the gene level, we wondered if the altered genes could be grouped in similar molecular pathways, as previously occurred in the lung cancer cell model. Consistent with this idea, functional analysis performed in SWI/SNF-deficient cells revealed a potential alteration of some signaling pathways similar to the ones reported in the prior experiments.

Thus, we observed a positive enrichment in genes belonging to EMT and angiogenesis. Some transcriptional alterations were conserved in at least two *ARID1A*-deficient cells (A549 and Caco-2: *S100A4, JAG2,* and *NRP1*; A549 and SK-OV-3: *VCAN* and *LUM*; Caco-2 and SK-OV-3: *THBS1, INHBA, TIMP3, LAMC1, COL12A1,* and *COL4A2*). Similarly, we found common transcriptional alterations related to angiogenesis and EMT in the case of *ARID1B*-deficient cells (A549 and Caco-2: *ITGAV, STC1, FSTL1*, *FN1, TGFBI, SNAI2, SPARC, SPOCK1, IGFBP4, SLIT2, MSX1, NRP1,* and *LUM*). The fact that we have been able to find commonly altered molecular pathways in different SWI/SNF-deficient cell lines supports the existence of general mechanisms of gene expression regulation exerted by SWI/SNF complexes which are conserved across different tumor types.

In addition to this, in *ARID1A*-deficient cells we observed a general upregulation of genes related to the immune response in all cellular contexts compared, some of which were shared in at least two of the cell lines (Caco-2 and SK-OV-3: *IRF1* and *IFR7;* Caco-2 and A549: *IFIT1, IFI6, IFITM1, IFITM2, IFITM3, OAS1, NRP1,* and *GATA3;* SK-OV-3 and A549: *LTB, TNFSF10,* and *NFKBIZ*).

In light of the exposed results in this and the previous section, it can be collectively concluded the existence of signaling pathways specifically controlled by each subunit in the same cellular context and also by the same subunit of the complex in different tissues. Together, these observations mean that target genes of SWI/SNF complexes depend not only on their composition, but also on their interaction with tissue-specific transcription factors.

1.3. Transcriptional alterations found in SWI/SNF-deficient NCI-H460 cell lines

Considering the high variability of the gene networks regulated by SWI/SNF complexes in different cellular conditions, we decided to focus our research on those mechanisms that could be conserved in a specific tumor type. Consequently, we reduced the tissue-specific complexity by extending our analysis to additional lung cancer cellular models because in this particular tumor type up to three different SWI/SNF subunits (*ARID1A, ARID2,* and *SMARCA4*) present alterations in around 20% of the cases (Monterde and Varela, 2022). Accordingly, we selected the lung cancer cell line NCI-H460, which presents a similar mutational background to A549 (*KRAS*-mutant), and we generated stably-transduced cell lines by knocking down the expression of either *ARID1A*, *ARID2,* or *SMARCA4*.

RNA-Seq experiments revealed that SWI/SNF deficiency in the NCI-H460 cell line was again accompanied by a significant dysregulation in the expression of thousands of genes (**Figure 4.5**). Similar to what we had observed in the A549 cell model, cBAF alteration (*ARID1A* deficiency) resulted in a higher number of dysregulated genes compared to the PBAF-altered model (*ARID2* deficiency). This finding reinforces the requirement of cBAF complexes to keep the expression of deeper gene networks than PBAF complexes in the context of lung cancer.

Nevertheless, as opposed to what we had observed in the A549 model, in the NCI-H460 cell line *ARID2* alteration resulted in a higher proportion of overexpressed genes (shARID2: 471 upregulated and 233 downregulated), while *ARID1A* or *SMARCA4* deficiency was generally associated with transcriptional repression (shARID1A: 1153 upregulated and 1460 downregulated; shSMARCA4: 2289 upregulated and 2819 downregulated).

It is quite remarkable that *SMARCA4* knock-down resulted in the highest number of differentially expressed genes (DEGs shARID1A: 2613; shARID2: 704; shSMARCA4: 5108), which could be explained regarding that catalytic subunits are present in all complexes. Hence, its downregulation is more likely to have a greater impact compared to other accessory subunits that are only present in certain subcomplexes. Interestingly, this result indicates that SMARCA2, the other catalytic subunit of the complex, is not able to fully compensate the absence of SMARCA4, as it has been suggested in previous studies (Katherine C. Helming et al., 2014).

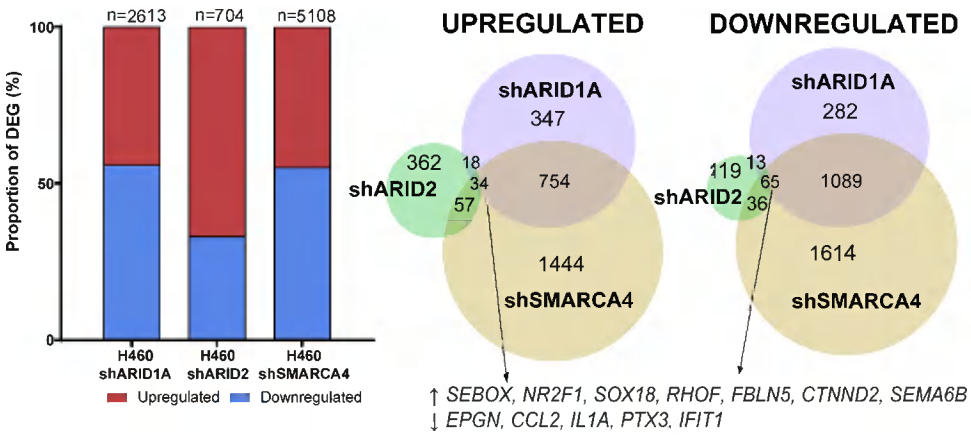


Figure 4.5. Transcriptomic landscape of SWI/SNF-deficient NCI-H460 cell lines. Proportion of DEGs reported in *ARID1A*, *ARID2,* or *SMARCA4*-KD NCI-H460 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each NCI-H460 deficient condition, shARID1A (purple), shARID2 (green), or shSMARCA4 (yellow), separated according to their fold change.

The comparison of the list of DEGs obtained in each stably-transduced cell line revealed a considerably higher overlap than in A549, although we also reported subunit-specific transcriptional alterations. This effect was particularly noticeable between shARID1A and shSMARCA4 cells, which could be explained considering that cBAF complexes are the most stoichiometrically abundant subtype (Mashtalir et al., 2018). In line with this, around 36% of dysregulated genes as a result of *SMARCA4* deficiency were commonly altered in *ARID1A*-deficient cells (1843/5108), while only 1.8% of DEGs were shared with *ARID2*-deficient cells (93/5108).

Similar to what we had noticed in the A549 cell line, *ARID1A* or *ARID2*-deficient NCI-H460 cells share a low amount of DEGs (52 upregulated and 78 downregulated), which strengthens the hypothesis that cBAF and PBAF complexes are required to regulate the expression of specific gene expression programs.

SWI/SNF deficiency was accompanied by the common dysregulation in the expression of 99 genes (34 upregulated and 65 downregulated), among which we can highlight the overexpression of transcriptional regulators implicated in development (*SEBOX, NR2F1,* and *SOX18*), molecules involved in cell adhesion and extracellular matrix remodeling (*RHOF, FBLN5, CTNND2,* and *SEMA6B*), together with the downregulation of *EPGN,* which is one of the ligands of the epidermal growth factor receptor, as well as genes related to the immune response (*CCL2, IL1A, PTX3,* and *IFIT1*).

Regarding the low proportion of shared transcriptional alterations reported in the previous experiments, we wondered if the overlap could be observed at the pathway level instead. Therefore, we made use of GSEA software to identify potential signaling pathways commonly altered as a result of the deficiency of any of the SWI/SNF subunits.

As it can be distinguished in **Figure 4.6.A**, we obtained a very high overlap at the pathway level, considering that almost half of the signaling pathways (8/20) were commonly altered in the context of SWI/SNF deficiency. Among the top 20 positively enriched molecular pathways in SWI/SNF-deficient cells, interestingly we observed again an enrichment in the epithelial to mesenchymal transition (EMT). Similarly, we observed a negative enrichment of genes related to the immune response in SWI/SNF-deficient cells.

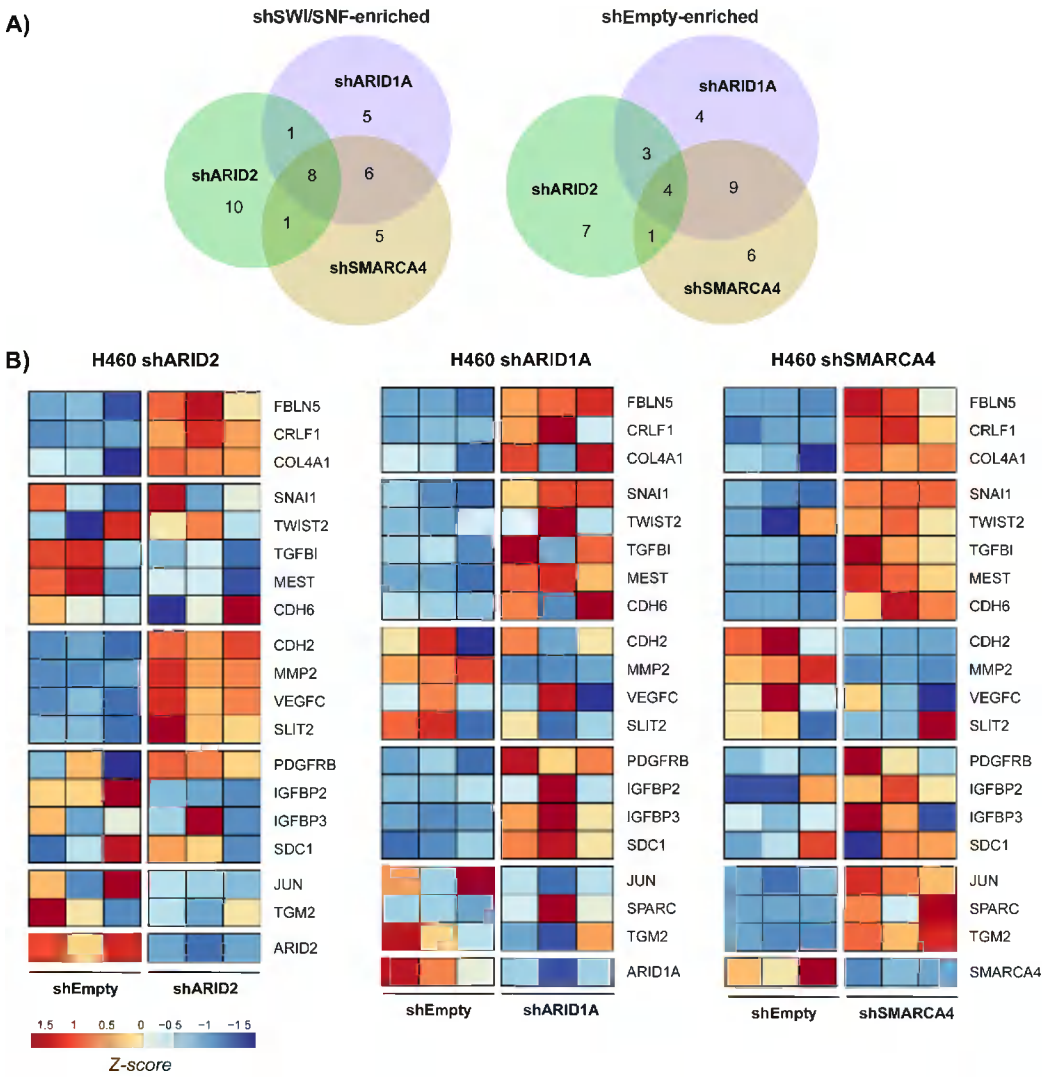


Figure 4.6. Positive enrichment of EMT in SWI/SNF-deficient NCI-H460 cell lines. A) Venn diagram representation of the overlap between the top 20 gene sets positively enriched in *ARID1A* (purple), *ARID2* (green), or *SMARCA4* (yellow)-KD NCI-H460 stably-transduced cells (left) and in control cells (right). **B)** Heatmap representation of a selection of upregulated genes related to the epithelial to mesenchymal transition in SWI/SNF-deficient NCI-H460 cells. Gene expression differences between the absence of different subunits of the complex (shARID2, shARID1A or shSMARCA4) and shEmpty cells (n=3 in each condition) have been represented according to the log2 of the fold change and have been indicated by colors, which range from red (upregulation) to blue (downregulation).

In particular, we observed that 3 genes related to the EMT were commonly upregulated in the absence of either *ARID1A*, *ARID2,* or *SMARCA4: FBLN5* (molecule involved in cell adhesion and extracellular matrix remodeling), *CRLF1* (receptor of cytokines type 1), and *COL4A1* (type of collagen that constitutes part of the basement membrane). Other transcriptional alterations were maintained in shARID1A or shSMARCA4 cells, as is the case of the upregulation of genes involved in cell migration (*ID1, PLXND1, SEMA4C, MESP1, TBX1, CORO1B, ANXA6, RADIL, LAMA5,* and *NRTN*), extracellular matrix organization (*COL6A2, COL9A2, LAMC3, LAMA5, COL18A1, COL4A2, COL14A1, MMP11, ITGB5, PLOD3, ADAM8,* and *LTBP3*), and mesenchymal markers (*SNAI1, TWIST2, TGFBI, MEST,* and *CDH6*). Moreover, we also reported specific transcriptional alterations for each subunit related to the EMT (shARID1A: *PDGFRB, IGFBP2, IGFBP3,* and *SDC1,* shARID2: *CDH2, MMP2, VEGFC*, and *SLIT2*; shSMARCA4: *JUN, SPARC,* and *TGM2*) (**Figure 4.6.B**).

On the other hand, SWI/SNF-deficient cells presented a repression of some genes belonging to the functional categories TNF-α signaling via NFKβ and response to IFN-γ. Some of these alterations were conserved in all cases (*IFIT1, IFIT3, PTGER4, PTX3, CXCL3, CCL2,* and *CXCL2*), others were shared among *ARID1A* and *SMARCA4*-deficient cells (*IL13RA2, IL13, CD74, TLR4, IL1B, IL6, IL8, IL7R, IFIH1, CD22, IGHM,* and *IGKC*), and we also observed subunit-specific transcriptional changes (shARID1A: *IFIT1, SOX4, KLF6,* and *CCL2*; shARID2: *IFIT2, IFITM2, IRF4,* and *GBP4;* shSMARCA4: *GBP6, CFB, MX2, IFIH1, CD74,* and *CXCL1*).

1.4. Common transcriptional alterations found in SWI/SNF-deficient lung cancer models

SWI/SNF deficiency resulted in an intense tissue-specific transcriptional reprogramming of cells characterized by thousands of changes in gene expression, which was highly variable depending on both the eliminated subunit and the specific cellular model. In order to discover commonly dysregulated mechanisms that are more likely to play an important role in lung cancer progression, we compared the transcriptomic landscape of either *ARID1A* or *ARID2* knock-down in two lung cancer cellular models, A549 and NCI-H460, both of which harbor *KRAS* activating mutations.

ARID1A or *ARID2* deficiency was accompanied by a profound transcriptional rewiring of cells that did not show a common pattern in any of the two models, in terms of the proportion of up- and downregulated genes (**Figure 4.7**).

However, in both cases, *ARID1A*-deficient cells present a higher number of DEGs compared to *ARID2* knock-down (DEGs A549 shARID1A: 1933, NCI-H460 shARID1A: 2613; A549 shARID2: 1155, NCI-H460 shARID2: 704). This finding is in line with our previous observations and reinforces the requirement of cBAF complexes to ensure the expression of a higher number of genes than PBAF complexes.

Accordingly, if we only consider the differentially expressed genes commonly reported in both cellular models, there were more dysregulated genes in the absence of *ARID1A* (225 genes) compared to *ARID2*-deficient cells (52 genes). What is more, the proportion of shared DEGs between the lung cancer cell models is higher in the case of *ARID1A* deficiency (8.6%) compared to *ARID2*-deficient cells (4.5%), which suggests a more conservative regulatory network between models in the case of cBAF complexes (these percentages have been calculated as the proportion of common DEGs with respect to the largest dataset compared, shARID1A: 225/2613; shARID2: 52/1155).

It is quite remarkable that less than 10% of DEGs were commonly altered between the two lung cancer models (A549 and NCI-H460) in the absence of two different accessory subunits of the complex (*ARID1A* or *ARID2*), revealing that SWI/SNF-mediated gene expression regulation is even more intricate and involves additional mechanisms specific to each cellular model.

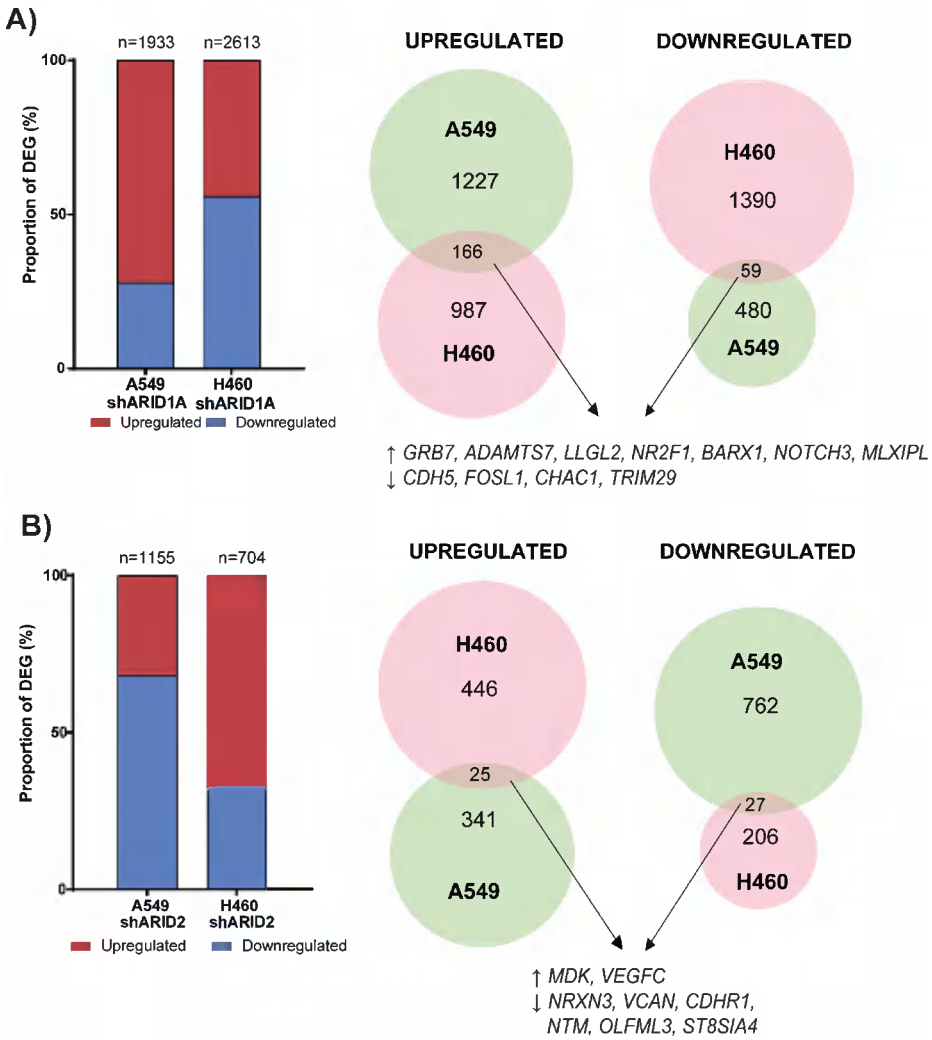


Figure 4.7. Transcriptomic landscape of SWI/SNF-deficient lung cancer cell lines. A) Proportion of DEGs reported in *ARID1A*-KD NCI-H460 and A549 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each *ARID1A*-deficient condition, NCI-H460 (red) or A549 (green), separated according to their fold change. **B).** Proportion of DEGs reported in *ARID2*-KD NCI-H460 and A549 stably-transduced cells, colored according to their fold change (red: upregulation; blue: downregulation). On the right, a Venn diagram representation of the overlap of DEGs reported in each *ARID2*-deficient condition, NCI-H460 (red) or A549 (green), separated according to their fold change.

Among the shared transcriptional alterations reported in *ARID1A*-deficient lung cancer cell lines (166 upregulated and 59 downregulated), we can highlight the overexpression of molecules involved in cell migration (*GRB7, ADAMTS7,* and *LLGL2*), the downregulation of the tumor suppressor gene *CDH5,* and a general dysregulation of developmental genes (*NR2F1, BARX1, NOTCH3, MLXIPL, FOSL1, CHAC1,* and *TRIM29*).

Functional analysis performed in both *ARID1A* knock-down cell lines revealed a positive enrichment in angiogenesis and the PI3K-AKT-mTOR signaling pathways. While there were commonly upregulated genes in both cell lines (*PLCG1, MKNK2, TNFRSF1A, TRIB3, PAK4, VAV3, PITX2, CALR,* and *JAG2*), we could also find genes specifically dysregulated in each model (A549: *VEGFA, LUM, VCAN,* and *NRP1*; NCI-H460: *FGFR1* and *VAVA2*).

In the case of *ARID2*-deficient cells, among the commonly altered genes (25 upregulated and 27 downregulated), we can highlight the upregulation of molecules that promote angiogenesis and cell growth (*MDK* and *VEGFC*), together with the downregulation of genes involved in cell adhesion (*NRXN3, VCAN, CDHR1, NTM, OLFML3,* and *ST8SIA4*).

In addition to this, functional analysis performed in *ARID2*-deficient lung cancer cell lines revealed potential alterations in several molecular pathways, highlighting a higher progression through cell cycle (*POLE, PRIM2, CDC25A, MCM2, MCM3, MCM6, BARD1, TMPO, PLK4, AURKA, XPO1, MKI67, ESPL1,* and *RAD21*) and the epithelial to mesenchymal transition (*VEGFC, LAMA3, COL1A1, ECM2,* and *DPYSL3*). We also found an upregulation of genes involved in DNA repair mechanisms, which were not conserved among cell lines (A549: *RAD51, RPA3, FEN1, POLD1, POLD3,* and *TYMS;* NCI-H460: *BRCA1, BRCA2, POLQ, POLE,* and *RAD50*).

A summary of the molecular pathways positively enriched in each SWI/SNF-deficient stably-transduced cell line can be found in **Table 4.1**. In this table we have represented a selection of the top 20 positively enriched signaling pathways in each condition and we have colored them according to their frequency across different cellular models. Among the recurrently altered events, we have found an enrichment of the epithelial to mesenchymal transition in almost all SWI/SNF-deficient models generated. However, although we found some commonly overexpressed genes after removing different SWI/SNF subunits in at least one lung cancer cellular model (*TWIST2, CDH2, MMP2, VEGFC, SDC1,* and *JUN*), the majority of upregulated genes tended to be subunit-specific. This finding might imply an activation of this biological process in the context of SWI/SNF deficiency, but how cells reach this state is probably different.

Finally, it is also worth mentioning that this transcriptomic analysis has enable us to identify potentially activated signaling pathways that could contribute to tumorigenesis. Apart from the positive enrichment of the epithelial to mesenchymal transition, in certain cases we have reported a potential activation of cell cycle, DNA repair, MAPK, and PI3K-AKT molecular pathways. Nevertheless, experimental validation is still required to prove that these alterations are certainly driving phenotypic changes in the cells.

Consistent with this idea, during the composition of this doctoral thesis, in our laboratory we described that *ARID2* deficiency in lung cancer cellular models is accompanied by a higher proliferative and metastatic potential of cells *in vivo* and *in vitro* (**Figure 4.8**). These findings are reflecting an activation of cell cycle progression that could be explained by the upregulation of MYC targets and genes involved in the G2M checkpoint.

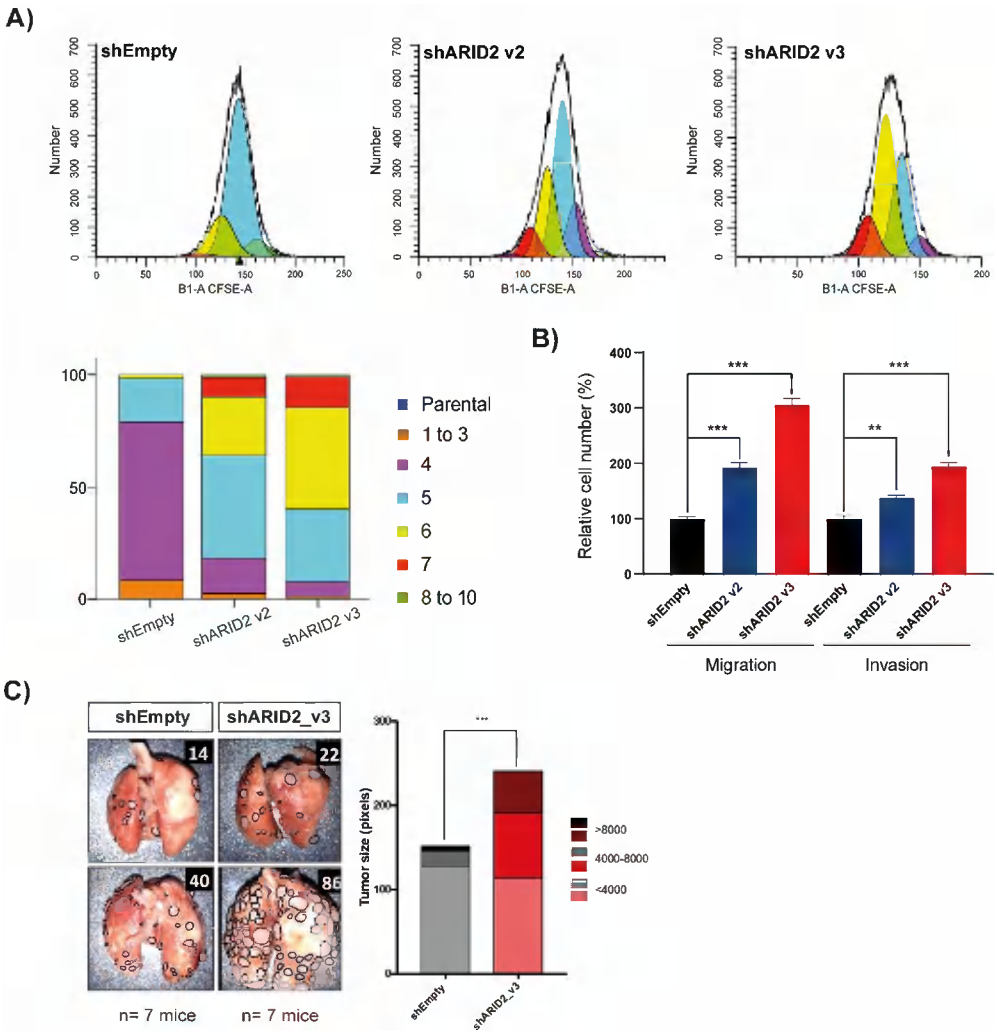


Figure 4.8. *ARID2* deficiency increases proliferative and metastatic potential of cells *in vitro* and *in vivo*. **A)** Histograms representing the proportion of A549 cells in each cell division at 48h, estimated by CFSE labeling and measured by flow cytometry. Bellow, bar representation of the percentage of cells that have suffered each number of cell divisions, indicated by different colors. **B)** Bar representation of quantified cells in destination chamber on migration and invasion assays of A549 cells transduced with two different shRNAs targeting *ARID2* (blue and red bars). Data are shown as mean ± SEM of three independent experiments, relative to control cells, A549 Empty vector (black bars), (two-tailed *t*-test **p* < 0.05, ***p* < 0.01, and ****p* < 0.001). **C)** Representative images of lung metastases generated in intravenously injected immunocompromised mice with A549 cells transduced either with shEmpty or shARID2_v3 vectors. On the right, a quantification of the number and size of tumors generated in both groups is shown (one-tailed Fisher's exact test **p* < 0.05, ***p* < 0.01, and ****p* < 0.001).

Cell line	SWI/SNF KD	Positively enriched functional terms	
A549	shARID1A	Angiogenesis	PI3K AKT mTOR signaling
		IL2 STAT5 signaling	Apical junction
		Upstream KRAS signaling	Response to IFNγ
	shARID1B	Angiogenesis	Epithelial to mesenchymal transition
		Response to IFNα	G2M checkpoint
		Upstream KRAS signaling	Response to IFNγ
	shARID2	G2M checkpoint	MYC targets
		DNA repair	mTORC1 signaling
		PI3K AKT mTOR signaling	Epithelial to mesenchymal transition
NCI-H460	shARID1A	E2F targets	Apical junction
		DNA repair	Epithelial to mesenchymal transition
		Mitotic spindle	PI3K AKT mTOR signaling
	shARID2	Apical junction	Upstream KRAS signaling
		IL2 STAT5 signaling	Upstream response to UV
		G2M checkpoint	Epithelial to mesenchymal transition
	shSMARCA4	NOTCH signaling	Downstream KRAS signaling
		Wnt β catenin signaling	Apical junction
		Angiogenesis	Epithelial to mesenchymal transition
Caco-2	shARID1A	Response to IFNγ	IL2 STAT5 signaling
		Response to IFNα	Angiogenesis
		Downstream response to UV	Epithelial to mesenchymal transition
	shARID1B	G2M checkpoint	Upstream KRAS signaling
		Angiogenesis	E2F targets
		Epithelial to mesenchymal transition	TGF-β signaling
SK-OV-3	shARID1A	IL2 STAT5 signaling	IL6 JAK STAT3 signaling
		Mitotic spindle	NOTCH signaling
		Epithelial to mesenchymal transition	Upstream KRAS signaling
Legend	1	2-3	4-5 > 5

Table 4.1. Functional categories enriched in SWI/SNF-deficient cell lines. Each column contains a selection of the most relevant signaling pathways positively enriched in SWI/SNF-deficient models from the top 20 results obtained using the GSEA hallmark gene set (Liberzon et al., 2015). Functional terms are depicted in different tones of blue depending on the number of cell lines that share the same alteration: from light blue (unique events) to dark blue (events found in more than five different cell lines).

2. Therapeutic implications of SWI/SNF-altered lung cancer patients

Once we had seen the potential molecular pathways frequently altered as a result of SWI/SNF deficiency, we wondered if this knowledge could have implications for the treatment of SWI/SNF-mutated patients. Among them, we were particularly interested in the study of the epithelial to mesenchymal transition and DNA repair, due to their potential implications in the context of lung cancer treatment.

2.1. Mutual exclusivity between SWI/SNF and *EGFR* mutations

One of the molecular pathways that we found recurrently enriched in SWI/SNF-deficient models is the epithelial to mesenchymal transition (**Table 4.1**). It has been reported that SWI/SNF complexes play an important role during the process of cellular differentiation (Jones et al., 2022). Therefore, loss of SWI/SNF activity could be accompanied by an increased cellular plasticity. Epithelial to mesenchymal transition is frequently observed in different tumors types, is characterized by the altered expression of adhesion molecules that allows cells to adopt a migratory and invasive behavior (Nieto et al., 2016), and has been associated with treatment resistance (Byers et al., 2013; Engl et al., 2015; Fischer et al., 2015), especially to targeted therapies against molecular pathways that fuel cellular proliferation.

In the particular case of lung cancer, many patients present an overactivation of the epidermal growth factor receptor (EGFR) pathway (Skoulidis and Heymach, 2019), which is typically combated with the use of specific inhibitors. In order to check if there were further evidence of a relationship between SWI/SNF alterations and EGFR signaling pathway, we analyzed the publicly available data from lung adenocarcinoma patients generated during the TCGA consortium (http://www.tcgaportal.org/TCGA/Lung_TCGA_LUAD/index.html).

According to the information deposited at the time of the elaboration of this doctoral thesis, SWI/SNF genes are collectively mutated in around 20% of the cases, with *SMARCA4* (8%), *ARID1A* (6%), *ARID2* (6%), and *SMARCA2* (3%) being the most commonly altered

subunits (**Figure 4.9.A** and **Supplementary Table 4**) (Monterde and Varela, 2022).

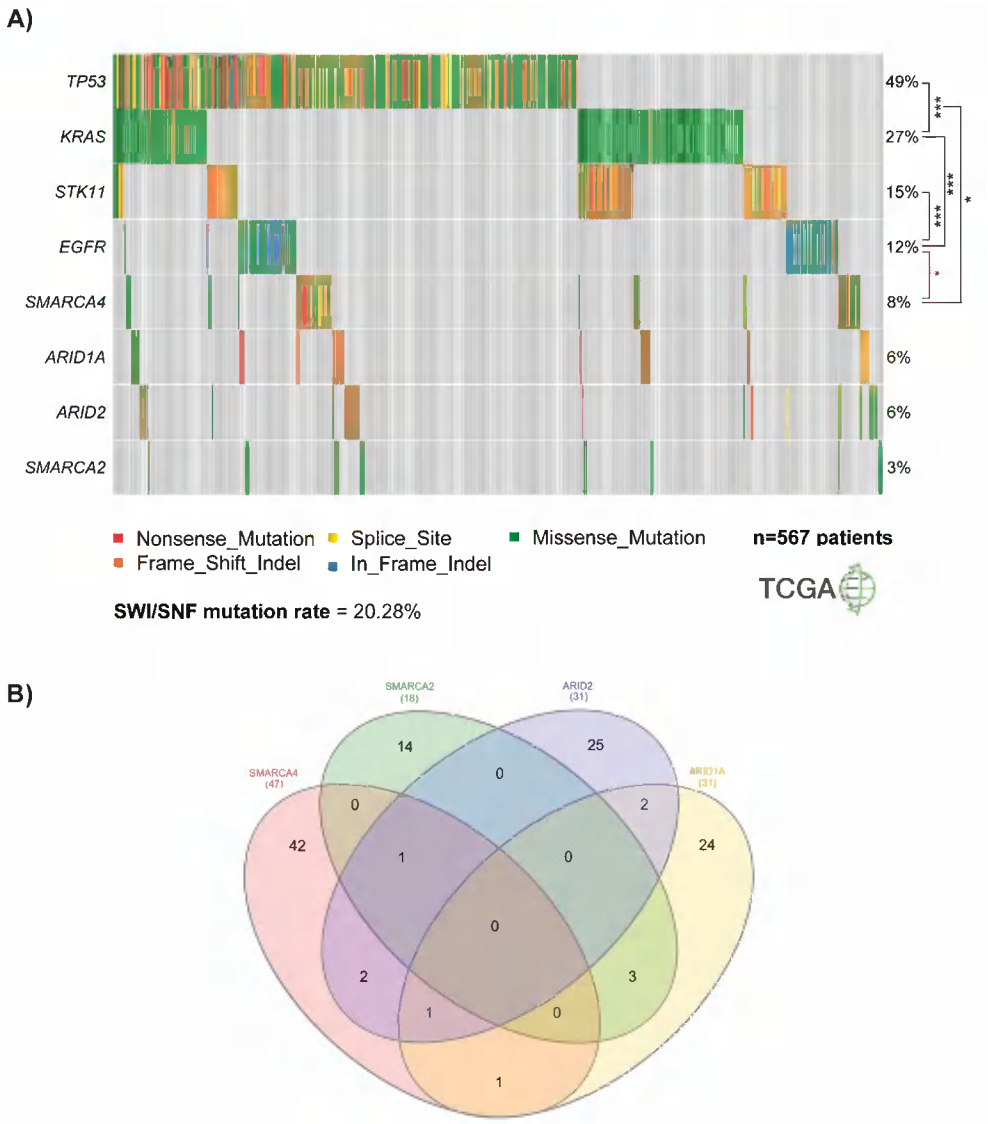


Figure 4.9. Mutual exclusivity between SWI/SNF and canonical pathways altered in lung adenocarcinoma. A) Oncostrip representation of the non-synonymous mutations found by MuTect2 (Cibulskis et al., 2013) in patients with lung adenocarcinoma (LUAD) extracted from the TCGA-LUAD project dated March 2022 . For each gene (rows), the presence of mutations in each of the 567 patients (columns) is represented in different colors, according to the mutation predicted functional consequence. SWI/SNF mutation rate has been calculated as the proportion of patients that harbor at least one mutation in one

of the components of SWI/SNF complexes in relation to the total number of patients. Silent mutations were excluded to calculate the mutation frequency of each subunit. *KRAS-EGFR, KRAS-TP53, STK11-EGFR, TP53*-SWI/SNF*, TP53-SMARCA4,* and *EGFR-SMARCA4* mutations are mutually exclusive. **B)** Venn diagram representation comparing the number of LUAD patients with mutations in the most frequently altered subunits of SWI/SNF complexes.

The fact that more than one subunit of the complex is mutated in patients of this particular tumor type suggests that SWI/SNF alteration involves common molecular mechanisms that finally lead to tumor development.

The number of SWI/SNF-mutated samples is not large enough to prove significant mutual exclusivity relationships between the individual subunits. However, from the total cohort of 115 patients with SWI/SNF-mutated tumors, only 10 patients harbored alterations in more than one subunit (8.7%), which indicates that collectively SWI/SNF mutations are mutually exclusive (*p-value*<0.0001, one-tailed Fisher's exact test) (**Figure 4.9.B**). This finding supports the idea that SWI/SNF mutations might provide selective advantages to tumor cells that once acquired, there is no further need to inactivate other components by secondary mutations.

In addition to this, SWI/SNF-mutated patients show *TP53* and *KRAS* mutations in similar ratios to no-mutated patients. Nevertheless, SWI/SNF-mutant patients show a lower *EGFR* mutation ratio (9/115 ~ 7.8%) than the 12% (68/567) observed in the complete cohort, although this difference is not statically significant (*p-value*=0.0797, one-tailed Fisher's exact test). This bias is mainly the result of *SMARCA4* mutations showing significant mutual exclusivity with *EGFR* mutations in this dataset, as only 2% (1/47) of *SMARCA4*-mutated patients show *EGFR* mutations versus the 12% that would be expected by chance (*p-value*=0.0151, one-tailed Fisher's exact test).

This observation extends previous results (Matsubara et al., 2013) and suggests the existence of either synthetic lethality relationships or redundant tumor progression-promoting activities.

2.2. SWI/SNF deficiency increases cell resistance to EGFR inhibitors

SWI/SNF-deficient cells presented an upregulation of certain genes involved in the epithelial to mesenchymal transition. Moreover, we reported a pattern of mutually exclusive mutations between SWI/SNF and *EGFR* in lung adenocarcinoma patients. Thus, we hypothesized that the presence of SWI/SNF alterations could have an impact on the sensitivity of the cells to treatments based on the inhibition of EGFR.

To figure out if that was the case, we treated the knock-down cells with different concentrations of dacomitinib and we assessed the sensitivity of each cell line by computing the half-maximal inhibitory concentration (IC_{50}). Dacomitinib is a second-generation tyrosine kinase inhibitor used in the treatment of lung cancer, which has been specially developed to overcome the *EGFR* T790M resistance mutation.

Considering that the positive enrichment in the epithelial to mesenchymal transition was specially significant in the case of *ARID2*-KD NCI-H460 cells, we first tested the sensitivity of this cellular model to EGFR inhibitors. As it can be distinguished in **Figure 4.10.A**, *ARID2*-deficient NCI-H460 cells presented an enhanced resistance to the inhibition of EGFR, evidenced by a higher drug concentration required to obtain a 50% of cell death (IC_{50}).

NCI-H460 cell line contains an activating *KRAS* mutation. Hence, in order to decipher whether this effect could rely on the activation of the KRAS signaling pathway, we performed equivalent experiments in the NCI-H1568 cell line, which is wild-type for *KRAS,* but presents an inactivating mutation in *TP53.* This assay showed that *ARID2*-deficient NCI-H1568 cells also exhibited higher resistance to EGFR inhibitors (**Figure 4.10.B**).

Finally, we conducted the same experiments in the NCI-H226 cell line, which is wild-type for both genes, and we confirmed that this increased resistance was conserved, regardless of the mutational background of *KRAS* and *TP53* genes (**Figure 4.10.C**).

Thus, our findings collectively indicate that, in the context of *ARID2* deficiency, resistance to EGFR inhibition is independent of the mutational status of *KRAS* and *TP53*.

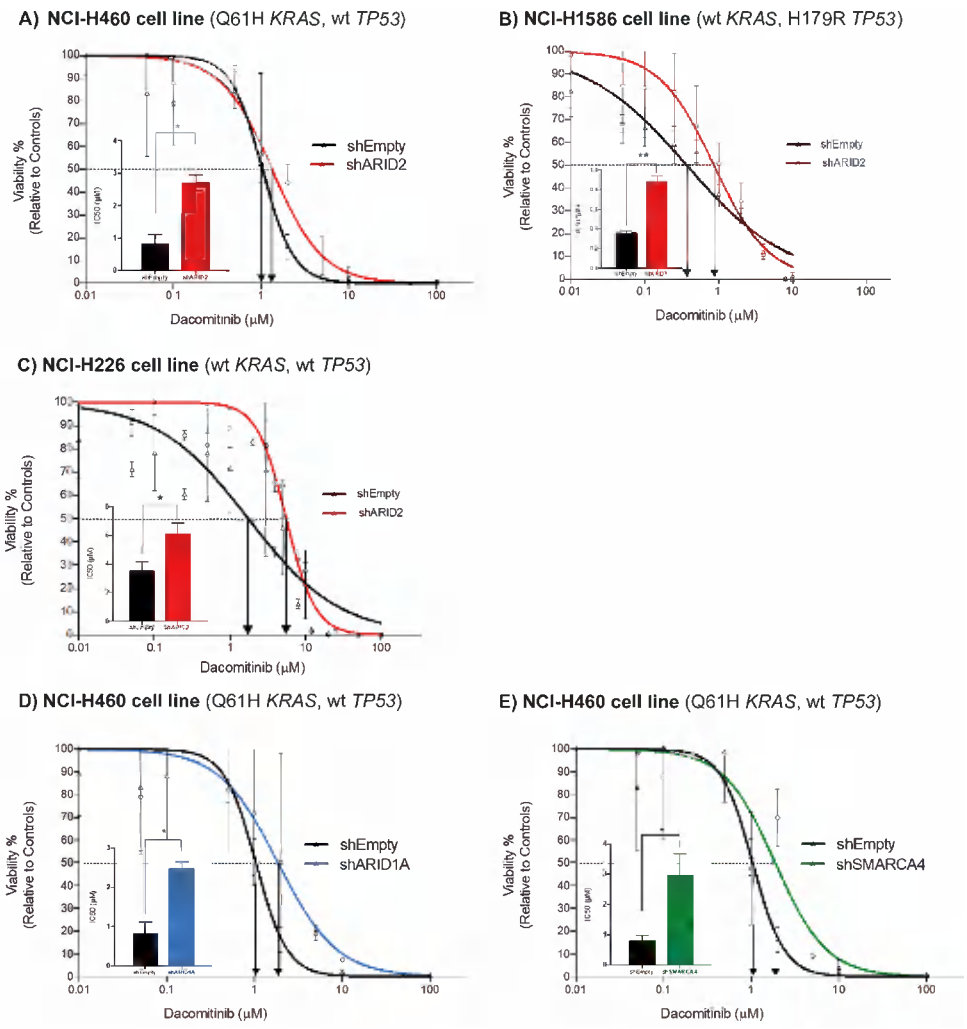


Figure 4.10. Sensitivity of SWI/SNF-deficient cell lines to EGFR inhibition. Representative experiments measuring cell survival to increasing concentrations of the EGFR inhibitor dacomitinib (PF-00299804) on NCI-H460 (**A**, **D**, and **E**), NCI-H1568 (**B**), and NCI-H226 (**C**) cell lines transduced with shEmpty (black), shARID2 (red), shARID1A (blue), or shSMARCA4 (green) vectors. Bar graphs represent the calculated IC_{50} value for each experiment. In all cases, the results are represented as mean ± SEM of at least three independent experiments (two-tailed *t*-test **p < 0.05, **p < 0.01,* and ****p < 0.001*).

Next, we wondered if this effect would be maintained after inactivating other subunits of the complex, taking into account that we had observed a positive enrichment in the epithelial to mesenchymal transition pathway in other SWI/SNF-deficient NCI-H460 cell lines. As shown in **Figure 4.10.A**, **Figure 4.10.D** and **Figure 4.10.E**, *ARID2*, *ARID1A* or *SMARCA4* deficiency resulted in higher resistance to EGFR inhibition. These findings reinforce the idea that the alteration of a particular subunit of the SWI/SNF complex is enough to achieve an enhanced resistance to EGFR inhibitors in the context of lung cancer.

Therefore, SWI/SNF mutational status appears to be very relevant in lung cancer treatment decisions, as our data suggest that SWI/SNF-mutated patients might present a worse response to treatments based on the inhibition of the EGFR signaling.

In order to unravel the molecular mechanisms contributing to the increased resistance to EGFR inhibition, we performed additional RNA-Seq experiments on *ARID2*-deficient NCI-H460 and control cells after adding dacomitinib to the medium. A complete list of the normalized counts and the results of differential expression analysis of *ARID2*-deficient and control NCI-H460 cells treated with EGFR inhibitor dacomitinib for 0, 24 and 48 hours can be found in **Supplementary Table 3.**

Although we have already observed an upregulation of several mesenchymal genes in *ARID2*-deficient cells compared to controls, after the treatment with dacomitinib these differences deeply increased (**Figure 4.11**). The role of some of these genes (e.g. *MYL9, EMP3, FN1, TGBI,* and *ECM1*) in the epithelial to mesenchymal transition has been extensively characterized before (Dong et al., 2022; Gan et al., 2018; Kahm et al., 2021). Even though the treatment with dacomitinib also triggered an epithelial to mesenchymal transition in control cells, this process takes longer and is less intense compared to *ARID2*-deficient cells. The epithelial to mesenchymal transition is a reported effect of EGFR inhibition and constitutes one of the main mechanisms associated with resistance to epidermal growth factor receptor inhibition (Li et al., 2017; Weng et al., 2019; Zhu et al., 2019).

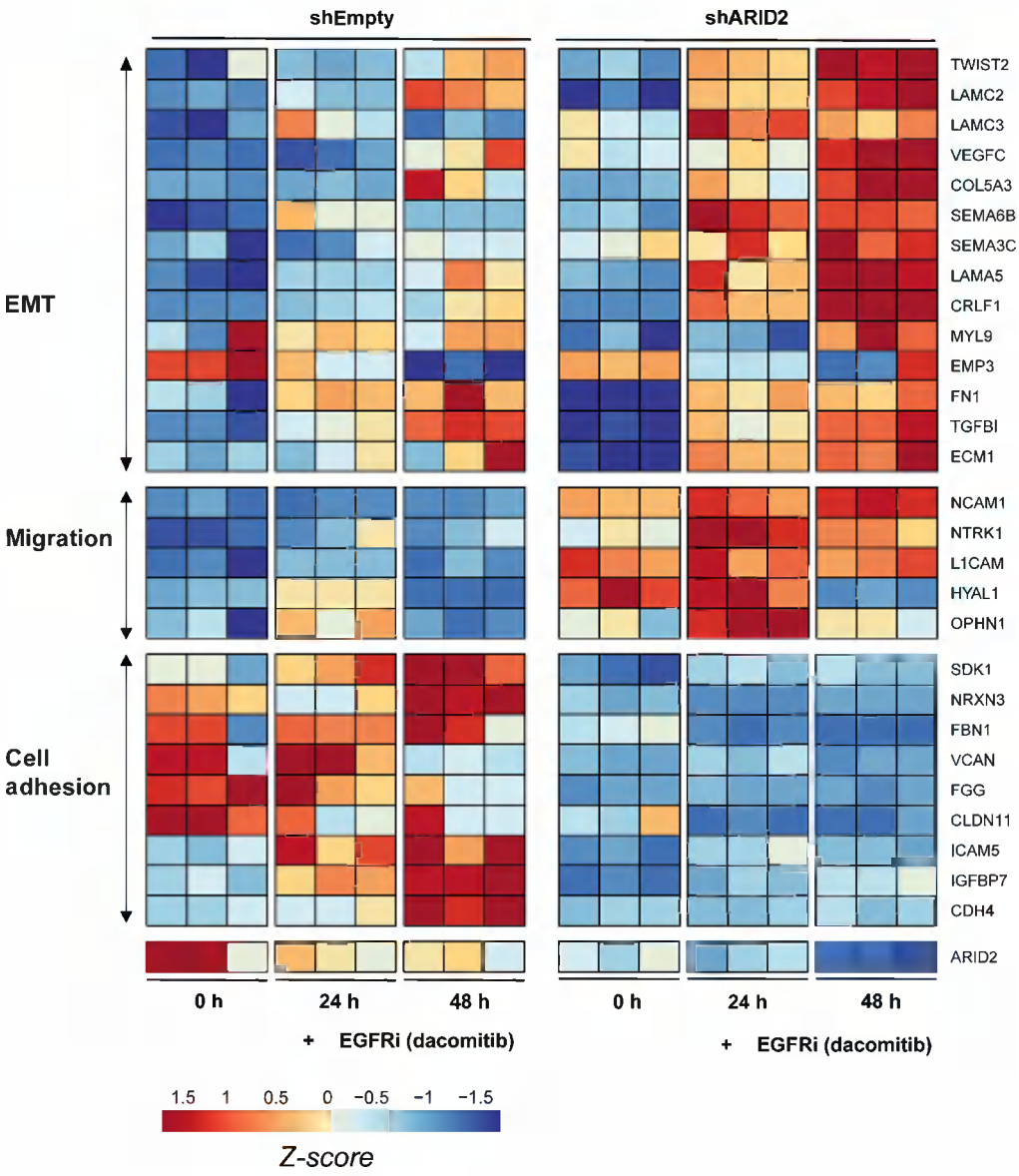


Figure 4.11. Transcriptional alterations found in *ARID2*-deficient NCI-H460 cells treated with the EGFR inhibitor dacomitinib. Heatmap representation of a selection of differently expressed genes in *ARID2*-deficient NCI-H460 cells (n = 3) at different times of exposure to EGFR inhibitor dacomitinib (0, 24 and 48 h) and grouped according to their biological function. Gene expression differences between shARID2 and shEmpty cells have been represented according to the log2 of the fold change and have been indicated by colors, which range from red (upregulation) to blue (downregulation).

Thus, our results are in agreement with the general model in which SWI/SNF-deficient cells are in a less robust epithelial differentiation state and, therefore, undergo this transition easier and faster than control cells. This behaviour likely explains the higher resistance to dacomitinib treatment observed in *ARID2*-deficient cells. Accordingly, these changes were also accompanied with an upregulation of genes involved in cell migration (*NCAM1, NTRK1, L1CAM, HYAL1,* and *OPHN1*), together with a downregulation of genes associated with cell adhesion (*SDK1, NRXN3, FBN1, VCAN, FGG, CLDN11, ICAM5, IGFBP7,* and *CDH4*).

2.3. DNA repair alteration in *ARID2*-deficient cells

Functional analysis of the transcriptional alterations found in SWI/SNF-deficient cells, and particularly *ARID2* knock-down models, showed a significant upregulation of genes involved in DNA damage detection and repair, suggesting a defective DNA damage response (**Figure 4.12**). We selected the GSEA gene sets sorted by biological function to figure out if DNA repair was a recurrently altered pathway in other databases. As a result, we observed that Gene Ontology terms “DNA repair”, “Damaged DNA binding”, “DNA synthesis involved in DNA repair”, and “Transcription coupled to Nucleotide Excision Repair" were particularly overrepresented in *ARID2*-deficient A549 cells.

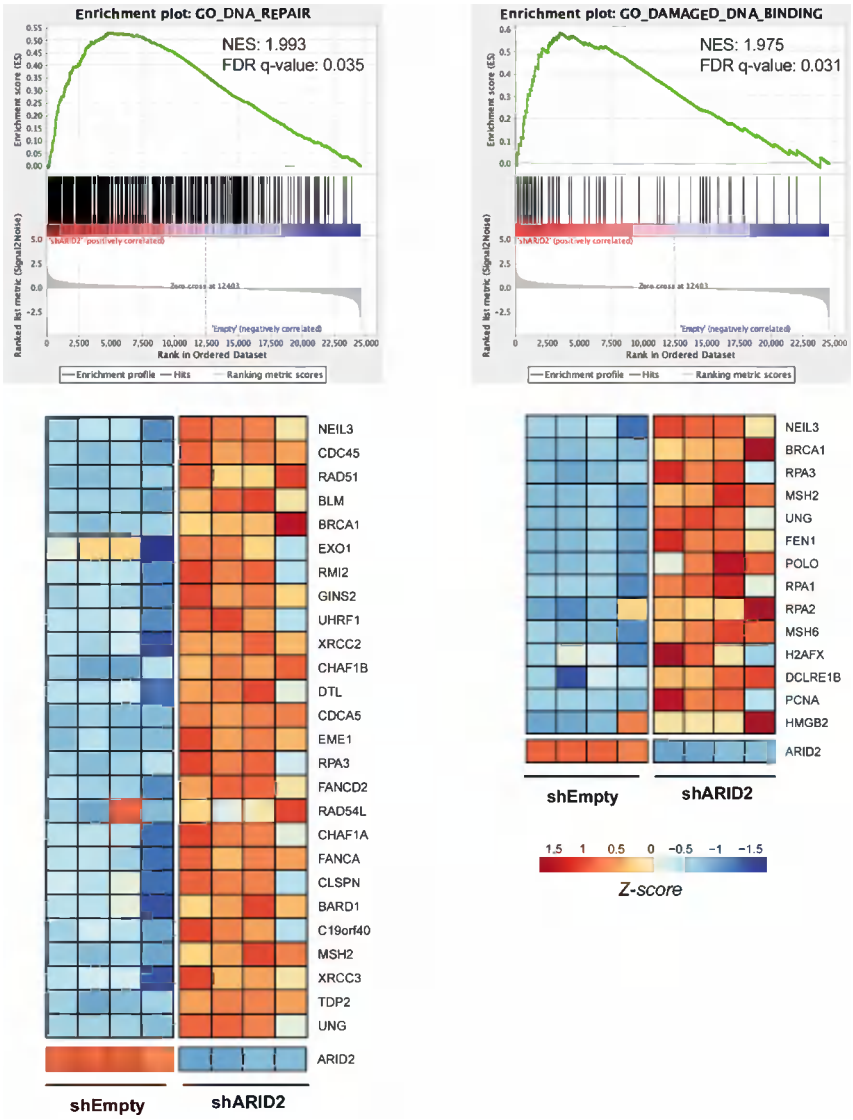


Figure 4.12. **DNA repair mechanisms are positively enriched in the absence of *ARID2*.** Representative signaling pathways overrepresented in *ARID2*-deficient A549 cells using Gene Set Enrichment Analysis (GSEA) from RNA-Seq data. Bellow a heatmap representation showing the expression of a selection of upregulated genes belonging to these ontologies in the different replicates is included. Gene expression differences between shARID2 and shEmpty cells have been represented according to the log2 of the fold change and have been indicated by colors, which range from red (upregulation) to blue (downregulation).

As DNA repair alteration leads to increased genomic instability, one of the recurrently observed cancer hallmarks (Hanahan and Weinberg, 2011), and can be exploited to improve the treatment of cancer patients, we decided to analyze the potential role of ARID2 in DNA repair. To that end, we performed immunofluorescence assays in DNA-damaging conditions to check the location of ARID2. In particular, we treated wild-type cells with etoposide, a chemotherapeutic agent that forms a ternary complex with DNA and topoisomerase II, inhibiting the ability of the enzyme to ligate cleaved DNA fragments and producing DNA strand breaks (van Maanen et al., 1988).

As shown in **Figure 4.13**, ARID2 colocalizes with markers of DNA damage (e.g. 53BP1 and γH2AX) at the DNA-repair foci in untransduced A549 cells. This finding strengthens the idea that ARID2 is recruited to DNA damage foci and exerts an active role in DNA repair mechanisms.

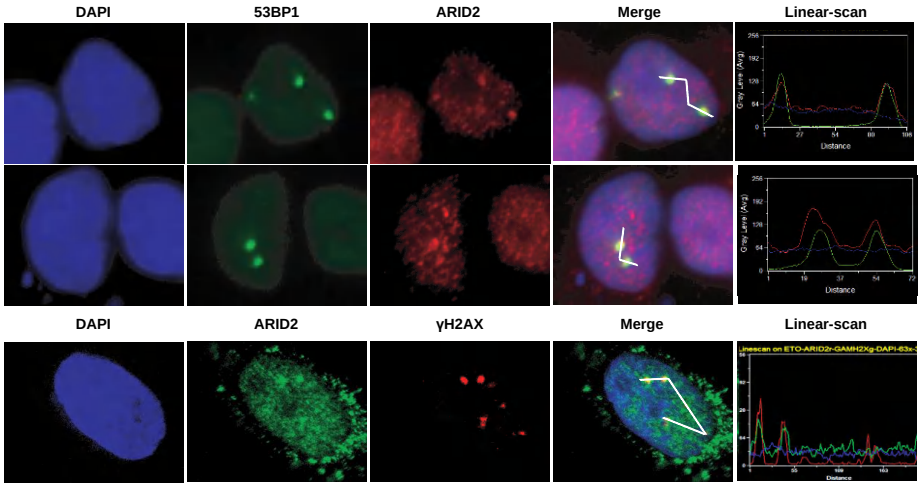


Figure 4.13. ARID2 colocalizes with DNA repair proteins. Representative images of immunofluorescence experiments performed in wild-type A549 cells demonstrating colocalization of 53BP1 (green) and ARID2 (red) (upper panel) and γH2AX (red) and ARID2 (green) (lower panel) at DNA damage foci. Colocalization was confirmed on the right through parallel quantifying the red and green signals on a manually selected path (white line) across the image using the LineScan tools from Methamorph software.

Subsequently, we decided to test whether *ARID2* deficiency could have an impact on the accumulation of DNA damage. Thus, we performed time-lapse experiments consisting of inducing the DNA damage response by a 1-hour treatment with etoposide, followed by cell recovery at different time points. We measured DNA damage

accumulation by γH2AX immunostaining. In order to avoid possible fluctuations between the basal γH2AX signal in each cell line, a control consisting of cells without treatment with etoposide was included to normalize the results measured at each time point.

As it can be observed in **Figure 4.14**, although there were no statistically significant differences in the amount of DNA damage foci between *ARID2*-deficient and control cells just after the chemotherapeutic treatment, *ARID2*-deficient cells showed a clear delay in DNA damage resolution evidenced by a higher number of γH2AX foci at 24, 48 and 72 hours after removing the drug from the medium.

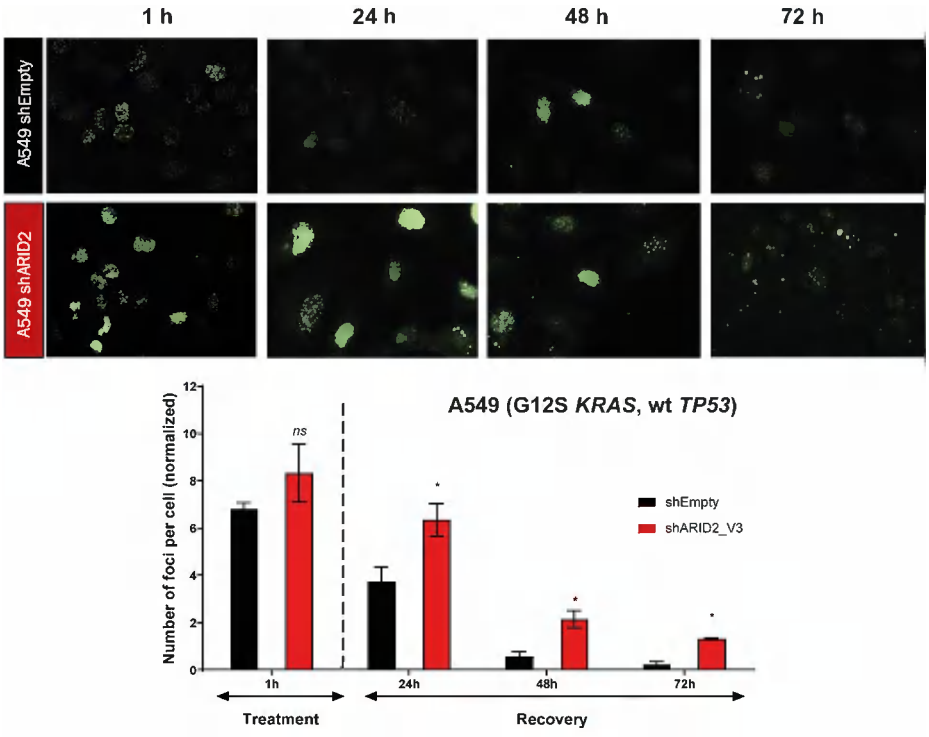


Figure 4.14. Effect of *ARID2* deficiency on DNA repair mechanisms in A549 cells. Representative images of DNA-repair foci visualized by γH2AX immunofluorescence (green) in transduced nuclei stained with DAPI at different recovery times after the treatment with etoposide in the A549 cell line. On the bottom, bar representation of the number of foci in each transduced cell line (shARID2_V3 or shEmpty) at different times of recovery after the induction of DNA damage with etoposide. For each cell line, a control consisting of the basal γH2AX signal in cells grown without etoposide was included for data normalization. The results are represented as mean ± SEM of at least three independent experiments (two-tailed t-*test* *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

In order to validate these observations in another cellular model, we performed equivalent experiments on the NCI-H460 cell line, which presents a similar mutational background to A549 cells (*KRAS*-mutated). *ARID2*-deficient NCI-H460 cells also showed a significant delay in the resolution of DNA damage upon treatment with etoposide at 24 hours of recovery, but these differences were lost at higher times of recovery (**Figure 4.15**). This finding suggests that NCI-H460 cells require less time than A549 cells to resolve the accumulated DNA damage induced by the treatment with etoposide.

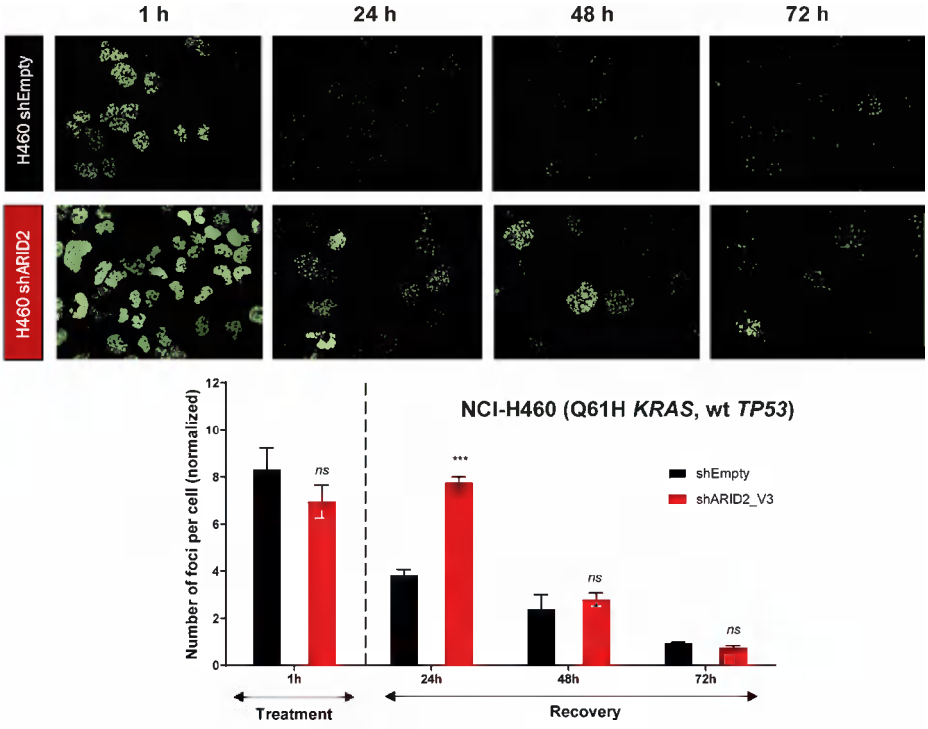


Figure 4.15. Effect of *ARID2* deficiency on DNA repair mechanisms in NCI-H460 cells. Representative images of DNA-repair foci visualized by γH2AX immunofluorescence (green) in transduced nuclei stained with DAPI at different recovery times after the treatment with etoposide in NCI-H460 cell line. On the bottom, bar representation of the number of foci in each transduced cell line (shARID2_V3 or shEmpty) at different times of recovery after the induction of DNA damage with etoposide. For each cell line, a control consisting of the basal γH2AX signal in cells grown without etoposide was included for data normalization. The results are represented as mean ± SEM of at least three independent experiments (two-tailed t-*test* $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$).

Subsequently, considering that neither A549 nor NCI-H460 cell lines contain *TP53* mutations, we decided to check the contribution of the *TP53* pathway to the observed effect on DNA damage accumulation. Accordingly, we initiated the same experiments in the NCI-H1568 cell line, which contains an inactivating mutation in *TP53* (H179R), but is wild-type for *KRAS*.

In this particular cell line, loss of *ARID2* is accompanied by a significantly higher amount of DNA damage at 48 and 72 hours of recovery (**Figure 4.16**). It is quite remarkable that the number of DNA damage foci measured per cell is the highest one compared to the previous experiments, which could be explained due to the lack of proper *TP53* function.

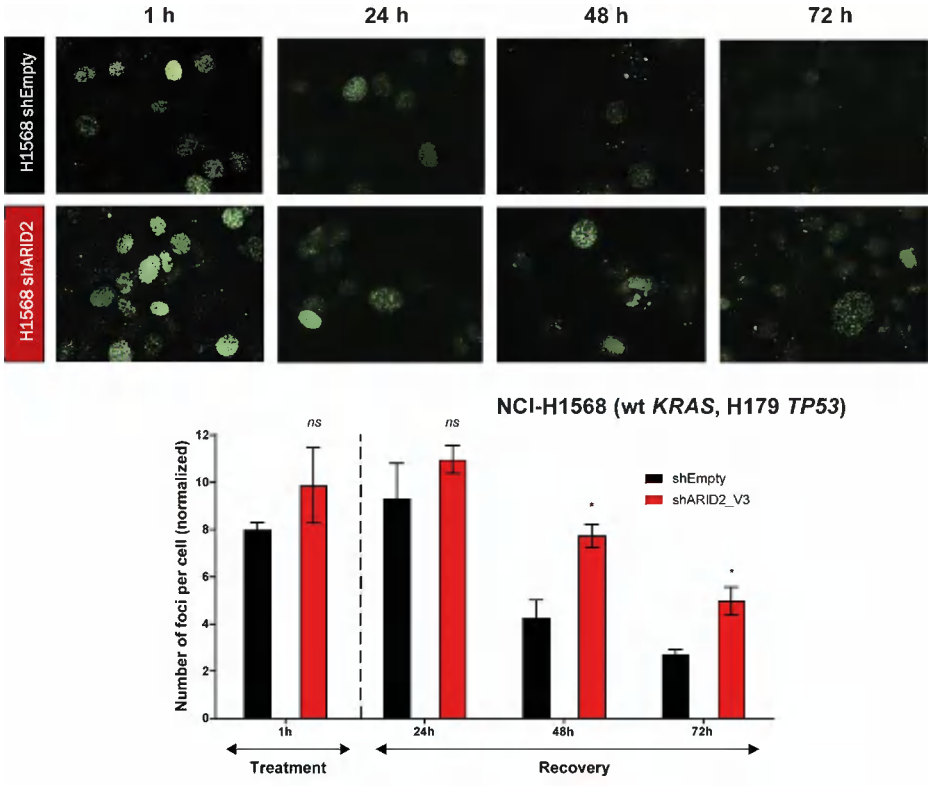


Figure 4.16. Effect of *ARID2* deficiency on DNA repair mechanisms in NCI-H1568 cells. Representative images of DNA-repair foci visualized by γH2AX immunofluorescence (green) in transduced nuclei stained with DAPI at different recovery times after the treatment with etoposide in NCI-H1568 cell line. On the bottom, bar representation of the number of foci in each transduced cell line (shARID2_V3 or shEmpty) at different times of recovery after the induction of DNA damage with etoposide. For each cell line, a control

consisting of the basal γH2AX signal in cells grown without etoposide was included for data normalization. The results are represented as mean ± SEM of at least three independent experiments (two-tailed t-*test* *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

Altogether our findings support the idea that the impaired DNA damage response observed in the absence of *ARID2* is not dependent on *KRAS* nor on *TP53* functions.

Finally, we conducted the same experiments in the NCI-H226 cell line, which is wild-type for *KRAS* and *TP53* genes. As it can be seen in **Figure 4.17**, there were no statistically significant differences between *ARID2*-deficient and control cells. Interestingly, this cell line presents the lowest amount of DNA damage foci measured per cell. This resistance to DNA damage accumulation might explain our failure to identify significant differences.

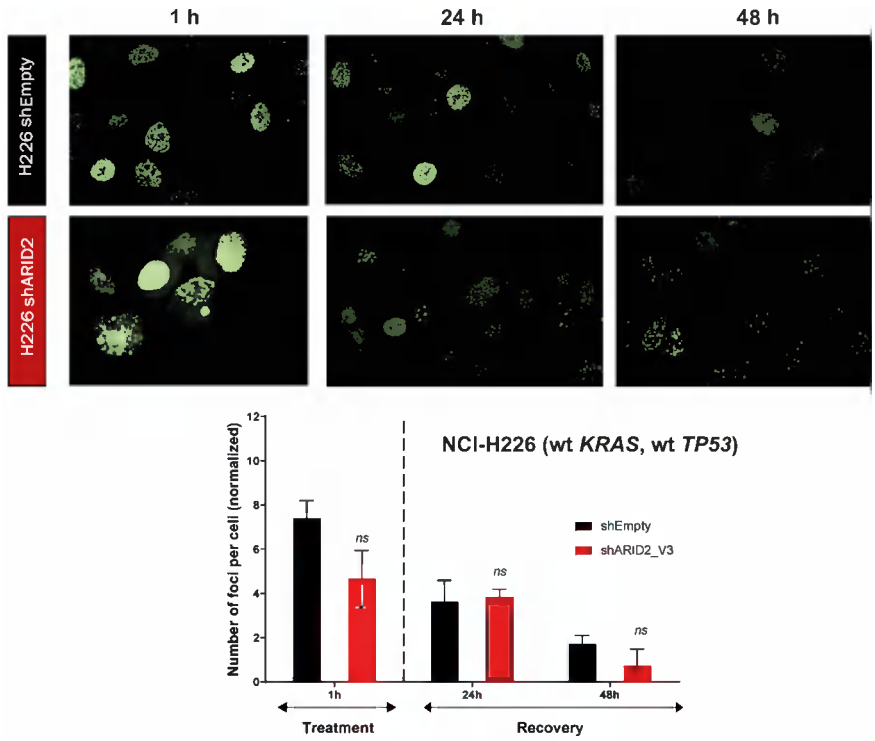


Figure 4.17. Effect of *ARID2* deficiency on DNA repair mechanisms in NCI-H226 cells. Representative images of DNA-repair foci visualized by γH2AX immunofluorescence (green) in transduced nuclei stained with DAPI at different recovery times after the treatment with etoposide in NCI-H226 cell line. On the bottom, bar representation of the number of foci in each transduced cell line (shARID2_V3 or shEmpty) at different times of recovery after the

induction of DNA damage with etoposide. For each cell line, a control consisting of the basal γH2AX signal in cells grown without etoposide was included for the normalization of the data. The results are represented as mean ± SEM of at least three independent experiments (two-tailed t-*test *p < 0.05, **p < 0.01*, and ****p < 0.001*).

2.4. *ARID2* deficiency increases cell sensitivity to chemotherapy and PARP inhibition

As platinum-based chemotherapy is widely used for the treatment of lung cancer patients (Chen et al., 2014), and we have seen that *ARID2*-deficient cells present deficiencies in DNA repair mechanisms, we hypothesized that *ARID2*-deficient cells might show different sensitivities to DNA-damaging agents.

Consequently, we first examined the sensitivity of *ARID2*-deficient cell lines to the chemotherapeutic agents cisplatin and etoposide. As it can be seen in **Figure 4.18.A**, in concordance with a defective DNA repair machinery, *ARID2*-deficient A549 cells exhibited a higher sensitivity to both compounds compared to control cells.

A similar effect was observed in NCI-H460 transduced cells (**Figure 4.18.B**). Both observations suggest that SWI/SNF-mutated lung cancer patients might show higher response rates to DNA-damaging treatments.

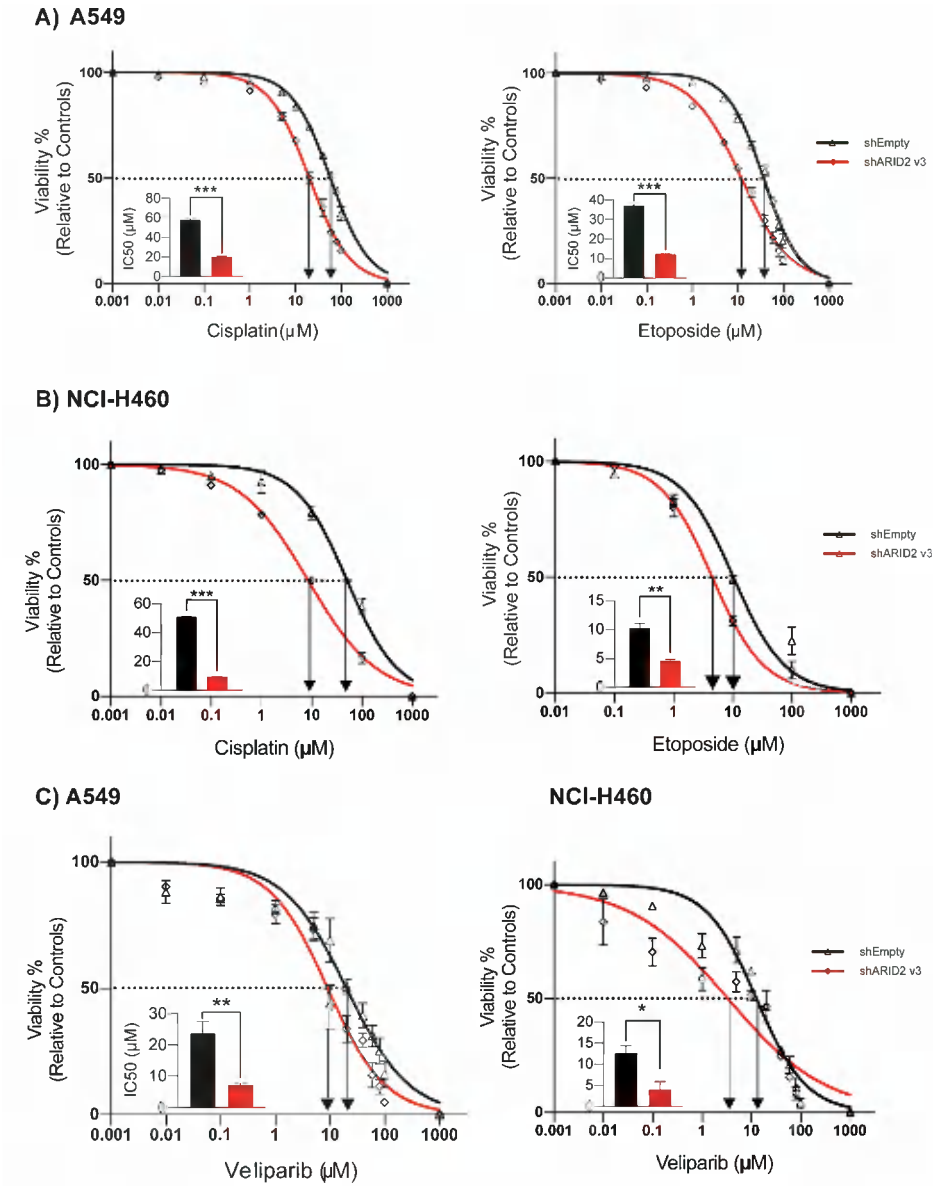


Figure 4.18. *ARID2* deficiency increases sensitivity to cisplatin, etoposide, and veliparib in A549 and NCI-H460 cell lines. Representative experiments measuring cell survival to increasing concentrations of the chemotherapeutic agents cisplatin and etoposide on A549 (**A**) and NCI-H460 cells (**B**) transduced with shEmpty (black), or shARID2_v3 (red) vectors. **C)** Sensitivity to PARP inhibitor veliparib in A549 (left) and NCI-H460 cells transduced with shEmpty (black), or shARID2_v3 (red) vectors. Bar graphs represent the calculated IC_{50} value for each experiment. In all cases, the results are represented as mean ± SEM of at least three independent experiments (two-tailed *t*-test **p < 0.05, **p < 0.01*, and ****p < 0.001*).

In addition to this, in the last decade, many researchers have described a higher sensitivity of PARP inhibitors in tumors harboring defects in DNA-repair mechanisms due to synthetic lethality (Lord and Ashworth, 2017). Consequently, we checked if this might apply as well to *ARID2*-deficient cells. As it can be seen in **Figure 4.18.C**, in both A549 and NCI-H460 cell lines ARID2 loss led to a higher sensitivity to veliparib, a well-described PARP inhibitor that is under research in several clinical trials in breast, ovarian, and most importantly, lung cancer.

Thus, our findings reinforce the importance of SWI/SNF mutational status in the assignment of the most appropriate treatment for cancer patients. In particular, *ARID2* deficiency could be used as a marker for the stratification of lung cancer patients that may benefit from treatments based on the promotion of genomic instability.

3. Molecular mechanisms involved in SWI/SNF-mediated gene expression regulation

3.1. Chromatin structure alteration of SWI/SNF-deficient cell lines

Considering that the main biological function of the SWI/SNF complex is the modulation of chromatin structure, we hypothesized that the transcriptional alterations observed in SWI/SNF-deficient cells might be the result of changes in its chromatin remodeling activity. Thus, we performed ATAC-Seq experiments on the knock-down cell lines.

The deficiency of each SWI/SNF subunit was accompanied by hundreds of specific chromatin structural changes at the complete genome, without a clear pattern of gain or loss of accessibility. This pattern was highly variable depending on both the cellular context and the eliminated subunit (**Figure 4.19**).

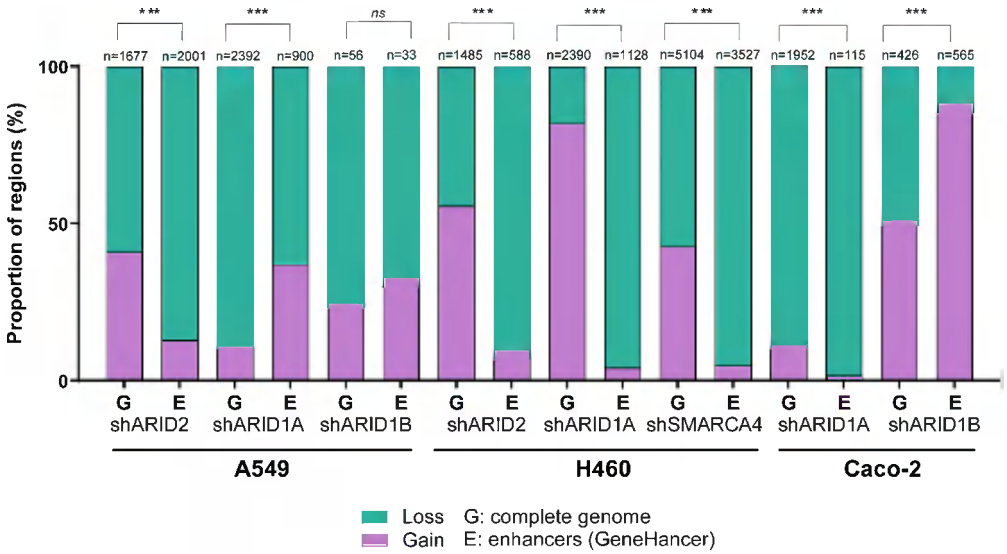


Figure 4.19. Chromatin structure alterations found in SWI/SNF-deficient cell lines. Barplot representing the proportion of regions that gain (purple) or lose (green) accessibility following *ARID1A*, *ARID1B*, *ARID2*, or *SMARCA4* knock-down in A549, NCI-H460, and Caco-2 cell lines, at the complete genome (G), and at enhancers (E) identified in the GeneHancer project (Fishilevich et al., 2017).

It is particularly noticeable that *ARID1B* deficiency barely altered chromatin accessibility neither in lung nor in colorectal cancer, as only 56 and 426 regions, respectively, showed significant changes compared to the many hundreds observed in the case of other subunits.

Taking into account the proposed role of SWI/SNF complexes at regulatory regions, we checked the global effect on the regions described as enhancers as part of the GeneHancer project (Fishilevich et al., 2017). In most cases, we observed a clear loss of accessibility affecting enhancers, which is statistically significant compared to the alterations found on the complete genome (**Figure 4.19**).

Those genomic regions with altered chromatin accessibility presented a low overlap, meaning that each subunit is required to modulate the structure of different sites. In addition to this, cBAF alteration (*ARID1A* deficiency) generally mediated the accessibility of more regions than PBAF complexes (*ARID2* deficiency), supporting its broader role in gene expression regulation. In all cellular contexts, we observed a higher overlap between those regions with a significant loss of accessibility (**Figure 4.20.A**).

Similar to what was observed in the transcriptomic analysis of SWI/SNF-deficient cell lines (previously addressed in chapter 1: “Characterization of the molecular mechanisms behind the role of SWI/SNF alteration in lung cancer development”), changes in chromatin accessibility showed high variability depending on the subunit composition of the complex.

Interestingly, in the context of SWI/SNF deficiency, A549 cell lines presented a predominant loss of accessibility at distal regions to gene transcription start sites, which could be the result of a specific effect on enhancers (**Figure 4.20.B**). Similar findings were reported in the colorectal cancer cell line Caco-2. However, in the case of NCI-H460 cell lines, we also observed regions located near the transcription start site especially affected by changes in accessibility, which might probably be affecting promoters. This effect is particularly remarkable in the case of *ARID1A*-deficient cells (shARID1A: 74.1%, 558/753, shARID2: 27.3% 180/660, shSMARCA4: 38.2%, 1117/2924).

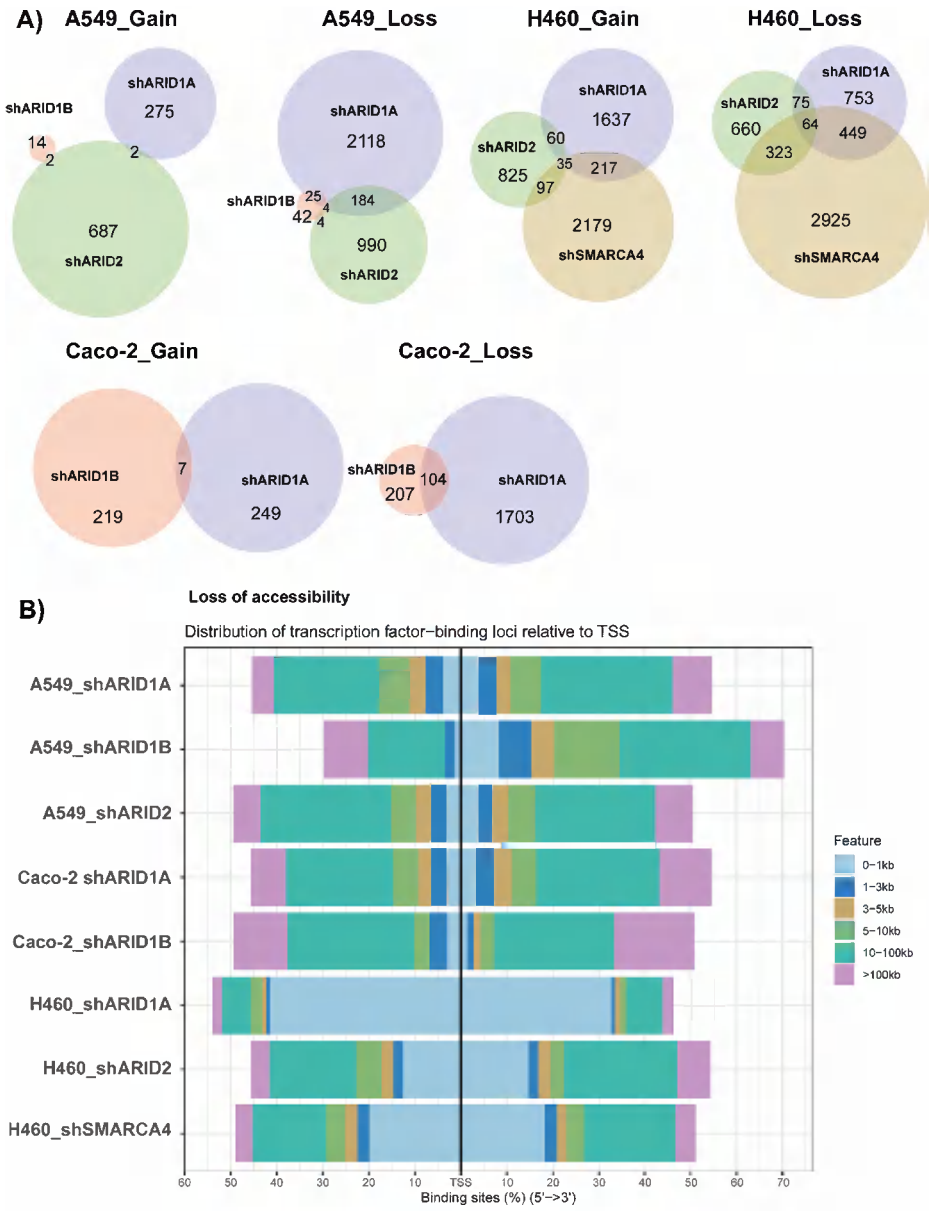


Figure 4.20. Alteration of chromatin accessibility of SWI/SNF-deficient cell lines. A) Venn diagram representation of the overlap between regions that gain or lose accessibility in each A549, NCI-H460, and Caco-2 deficient condition, shARID1A (purple), shARID2 (green), shARID1B (orange), shSMARCA4 (yellow). **B)** Barplot distribution of the proportion of genomic regions that significantly lose accessibility as a result of SWI/SNF deficiency according to their distance to the nearest transcription start site (TSS) and their relative orientation (upstream or downstream). Categories representing several distances to TSS are plotted in different colors.

Subsequently, we performed sequence-based motif analysis on the affected regions using HOMER to identify DNA-binding sites indicative of specific transcription factors. The results of this analysis showed that, in most cases, binding motifs of the AP-1 transcription factors were particularly enriched in those genomic regions that showed less accessibility on SWI/SNF-deficient lung cancer cell lines (**Figure 4.21**).

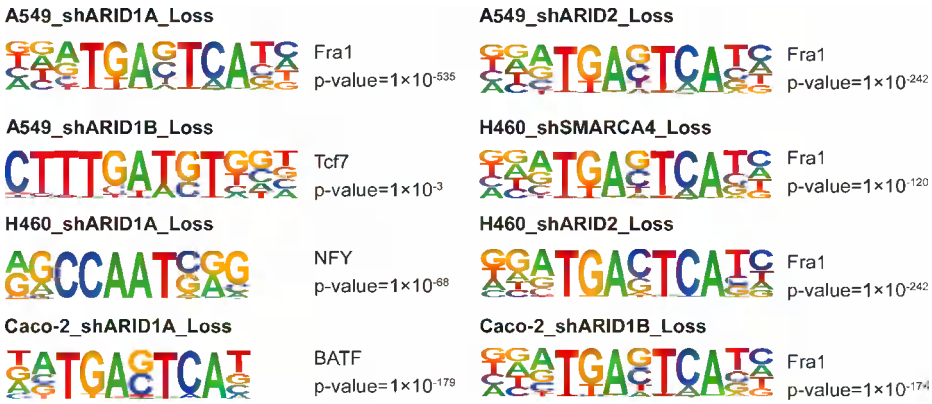


Figure 4.21. Enrichment of sequence motifs identified by HOMER in those regions that lost chromatin accessibility after SWI/SNF deficiency. The most significant motifs identified in each SWI/SNF-deficient cell model are shown. In most cases, a recurrent sequence, TGA(C/G)TCA, recognized by the AP-1 transcription factors was found.

AP-1 transcription factor family has been previously described to be abundantly present in enhancers (Sheffield *et al.*, 2013; Bejjani *et al.*, 2019). Consequently, our findings are in agreement with the proposed specific role of SWI/SNF complexes in establishing an open chromatin conformation at enhancers.

The identified recurrent sequence, TGA(C/G)TCA (TRE site) (Wu et al., 2021), was also detected after knocking down some accessory subunits of the complex in the colorectal cancer cell line Caco-2 (**Figure 4.21**). In this particular cellular context, Sen and coworkers identified a direct binding of ARID1A in AP-1 occupied enhancers (Sen et al., 2019). Moreover, the Greenberg lab demonstrated that AP-1 transcription factors mediate the targeting of SWI/SNF complexes to specific enhancers across the genome during the process of cell differentiation (Vierbuchen et al., 2017). Altogether, these results strengthen the idea that SWI/SNF alteration might affect the ability of the complex to regulate enhancer accessibility and developmental processes.

In the particular case of *ARID1A*-deficient NCI-H460 cells, we selected the genomic regions that were located near the transcription start site (± 1kb), as they were the most affected by changes in accessibility (74.1%, 558/753). The main finding of sequenced-based motif analysis was a sequence recognized by the NF-Y transcription factor, which is known to bind promoter proximal sites to ensure the maintenance of nucleosome-depleted regions (Oldfield et al., 2019) (**Figure 4.21**). Therefore, this result supports the fact that ARID1A cooperates with other nuclear components to keep an open chromatin structure at promoters.

Our findings collectively point to a major effect of SWI/SNF complexes in ensuring chromatin accessibility at regulatory regions responsible for gene expression regulation, such as enhancers and promoters.

Although we have seen a general loss of accessibility affecting enhancers as a result of SWI/SNF deficiency, the affected regions presented a low overlap (**Figure 4.22**). This result suggests that each subunit modulates the accessibility of different enhancers. Similar to what was reported in the alterations of the complete genome, there was a higher overlap between those enhancers with a significant loss of accessibility.

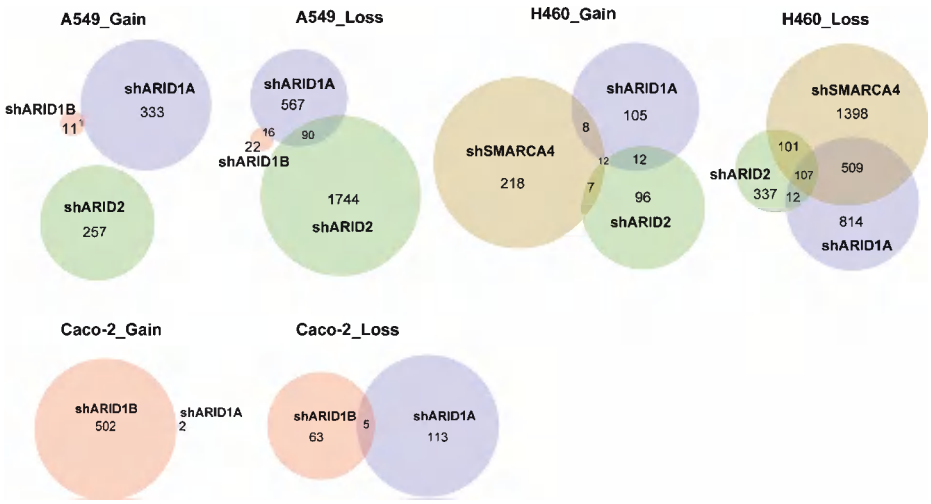


Figure 4.22. Alteration of enhancer accessibility of SWI/SNF-deficient cell lines. Venn diagram representation of the overlap between enhancers identified with the GeneHancer database that gain or lose accessibility in each A549, NCI-H460, and Caco-2 deficient condition, shARID1A (purple), shARID2 (green), shARID1B (orange), shSMARCA4 (yellow).

3.2. Correlation between alterations in enhancer accessibility and gene expression in SWI/SNF-deficient cell lines

Considering that SWI/SNF deficiency altered the gene expression program of the cells and its chromatin remodeling effect was mainly affecting enhancers, we hypothesize that dysregulated genes might be the result of changes in the accessibility of their theoretical enhancers.

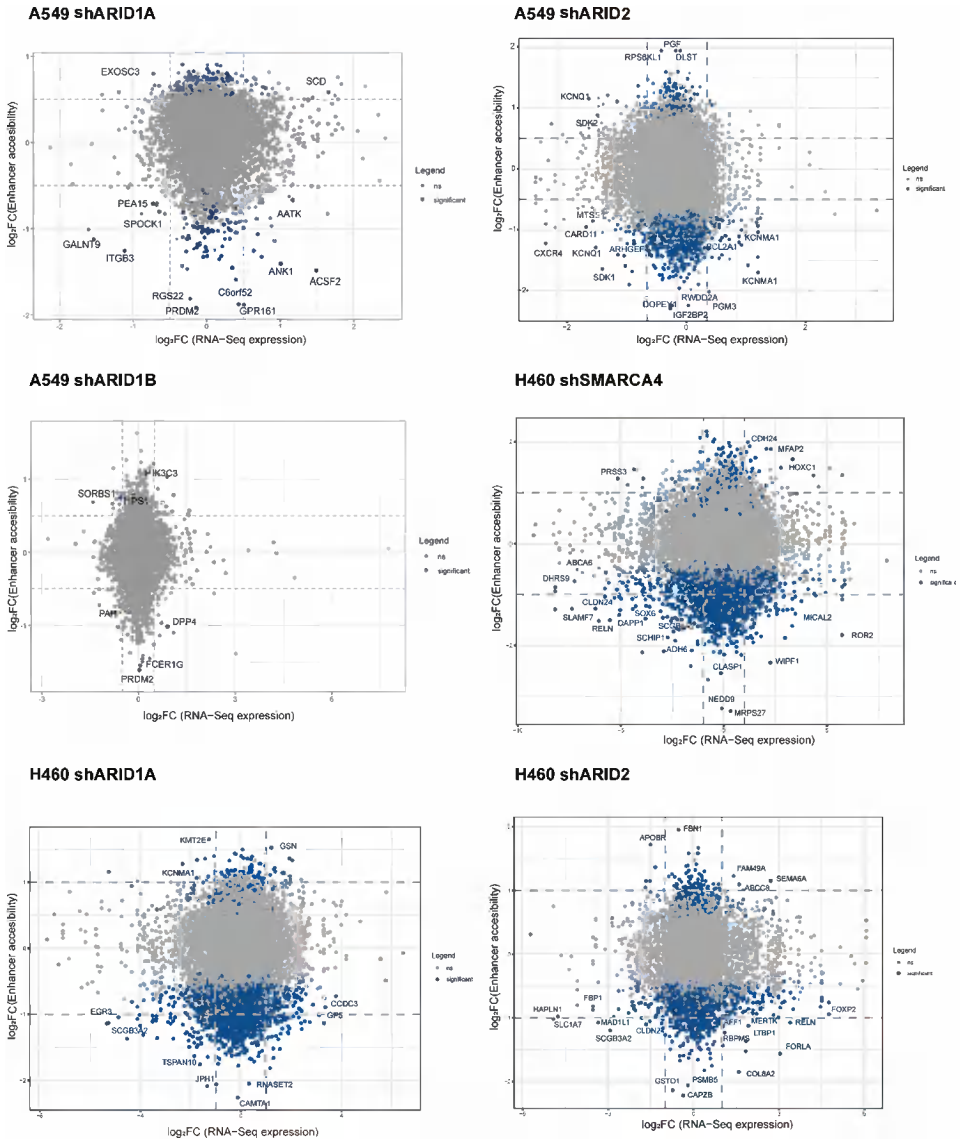


Figure 4.23. Correlation between gene expression changes and alterations in chromatin accessibility of enhancers in SWI/SNF-deficient lung cancer cell lines. Dot plots representing the correlation between the accessibility changes

in enhancers obtained by ATAC-Seq (Y-axis) according to the GeneHancer project and expression changes on their target genes obtained by RNA-Seq (X-axis) for the different stably-transduced cell lines are shown. Blue dots represent enhancers that showed significant accessibility changes in SWI/SNF-deficient lung cancer cell lines.

Therefore, in order to identify potential targets of the complex, we looked for a general correlation between transcriptional alterations and changes in the accessibility of their associated enhancers according to the GeneHancer project.

However, as shown in **Figure 4.23** and **Figure 4.24**, we did not observe a clear correlation between alterations in gene expression and enhancer accessibility in any of the SWI/SNF-deficient lung or colorectal cancer models.

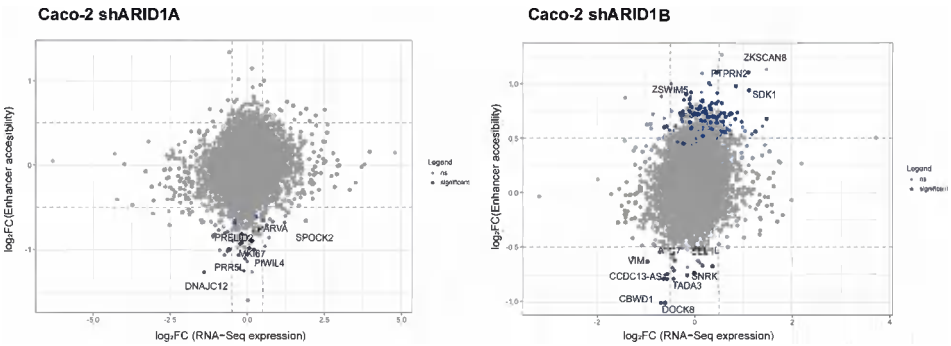


Figure 4.24. Correlation between gene expression changes and alterations in chromatin accessibility of enhancers in SWI/SNF-deficient colorectal cancer cell lines. Dot plots representing the correlation between the accessibility changes in enhancers obtained by ATAC-Seq (Y-axis) according to the GeneHancer project and expression changes on their target genes obtained by RNA-Seq (X-axis) for the different stably-transduced cell lines are shown. Blue dots represent enhancers that showed significant accessibility changes in SWI/SNF-deficient colorectal cancer cell lines.

Consequently, we hypothesize that the mechanisms of gene expression regulation mediated by SWI/SNF complexes are more intricate and cannot be explained just by changes in the accessibility of enhancers.

3.3. Alteration of enhancer activity in SWI/SNF-deficient lung cancer cell lines

In order to get further insight into how SWI/SNF regulates enhancer activity, we characterized the genome-wide distribution of histone epigenetic marks (H3K27ac, H3K4me1, H3K4me3, and H3K27me3) in SWI/SNF-deficient NCI-H460 cell lines using the CUT&RUN methodology (a detail description of this technique is addressed in chapter 2.6).

Subsequently, we used ChromHMM software, which integrates epigenetic data from different histone modifications and predicts the presence of different chromatin states that can be manually annotated. As a result, we obtained four different types of enhancers with different patterns of histone marks (**Figure 4.25**), that correspond to those regulatory regions previously identified as active (E1 and E2), primed (E3), and bivalent enhancers (E4) by the developers of ChromHMM (Ernst and Kellis, 2017).

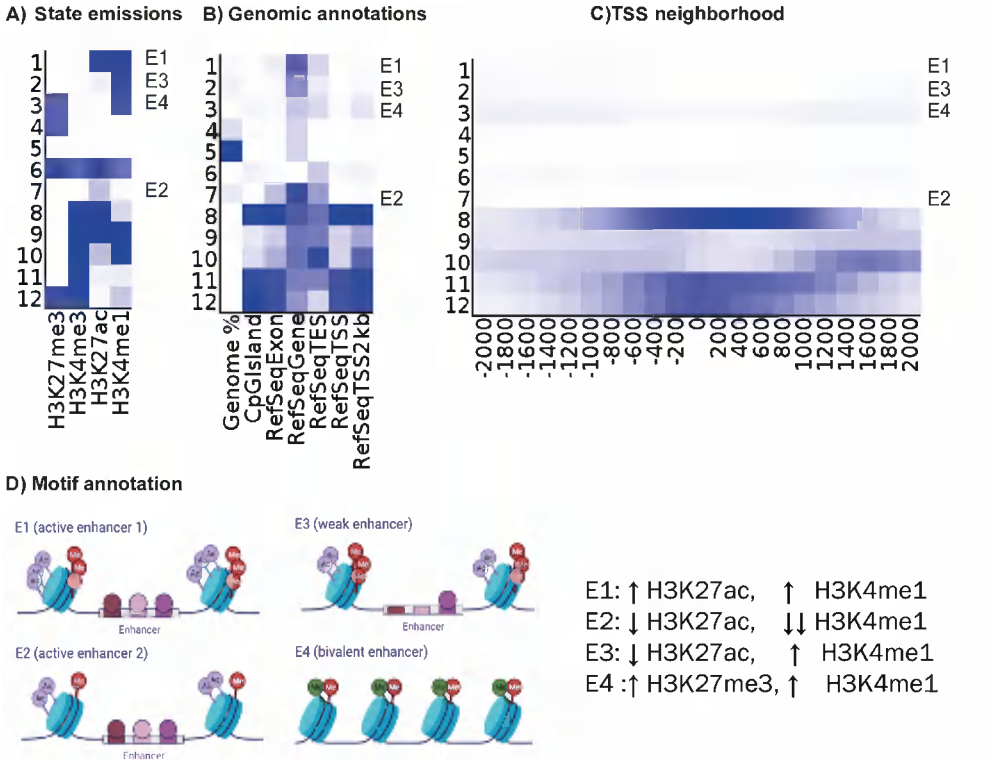


Figure 4.25. Types of enhancers identified with ChromHMM software. Epigenetic modifications H3K27ac, H3K4me1, H3K4me3, and H3K27me3 from SWI/SNF-deficient NCI-H460 cell lines were integrated to find a total of 12

chromatin states. Plots showing the distribution of histone modifications (**A**), the genomic annotation (**B**), and the distance to the nearest TSS (**C**) for each chromatin state are depicted. In these heatmaps each row corresponds to a different chromatin state and each column a different histone modification (**A**), genomic annotation (**B**) or distance to TSS (**C**), coloured in different tones of blue according to their probability (darker tones represent higher probabilities). **D)** Diagram showing the pattern of histone marks and its description in the different types of enhancers identified (the correspondence to chromatin states identified by the Kellis lab can be found in **Table 3.13**).

We hypothesized that gene expression changes resulted from SWI/SNF alteration might be caused by changes in the activity of their corresponding enhancers. Thus, among the epigenetic marks characterized, we focused on the study of H3K27ac and H3K4me1 due to their implication in enhancer functionality.

H3K27ac is a histone modification located at promoters and enhancers of actively transcribed genes. SWI/SNF deficiency resulted in a highly variable alteration in the deposition of this mark, depending on both the eliminated subunit and the type of enhancer affected (**Figure 4.26.A**).

Considering the total number of enhancers with altered H3K27ac deposition, *SMARCA4*-deficient cells present the highest number of affected regions (shARID1A: 896; shARID2: 2362, shSMARCA4: 3983). In addition to this, *ARID2* deficiency resulted in almost a three times higher number of affected enhancers compared to *ARID1A*-deficient cells, which suggests that PBAF complexes might control the acetylation or a larger number of regions than cBAF complexes.

Similar to what was concluded in the transcriptional and the accessibility studies, the overlap of enhancers with altered acetylation in the context of SWI/SNF deficiency was quite low, which indicates that each subunit is required to modulate the acetylation levels of specific regions (**Figure 4.26.B**).

In most cases, we reported a higher overlap between *ARID1A* and *SMARCA4*-deficient cells, which means that the acetylation of some enhancers relies on more than one subunit of the complex. In addition to this, we reported some overlap between *ARID1A* and *ARID2*-deficient cells, besides being part of different SWI/SNF subfamilies (cBAF and PBAF, respectively).

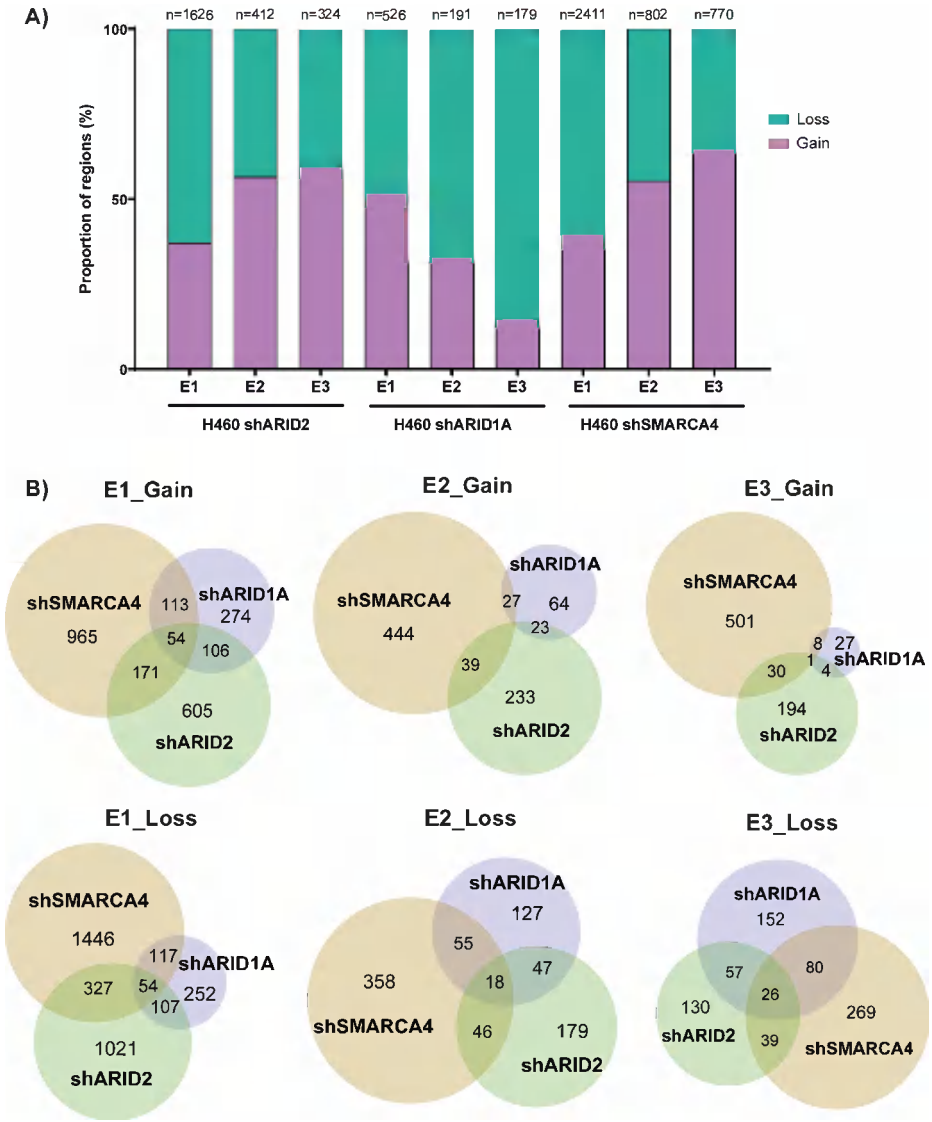


Figure 4.26. H3K27ac alteration at enhancers identified in SWI/SNF-deficient lung cancer cell lines. A) Barplot representing the proportion of regions that gain (purple) or lose (green) acetylation following *ARID1A*, *ARID1B*, or *SMARCA4* knock-down in NCI-H460 cell lines in different types of enhancers identified with ChromHMM. E1 and 2: active enhancer; E3: primed enhancer. E4 type enhancers are not included in this comparison as they do not present H3K27ac signal. **B)** For each type of enhancer, Venn diagrams showing the overlap of enhancers that gain or lose H3K27ac in the absence of different subunits of the SWI/SNF complex, shARID1A (purple), shARID2 (green), shSMARCA4 (yellow), requiring a minimum of 20% of overlap between two regulatory regions.

Therefore, although there might be enhancers whose acetylation relies on both subfamilies, H3K27ac regulation of most enhancers tends to be subunit-specific. The complete analysis of H3K27ac deposition on the regions identified as enhancers using ChromHMM in all SWI/SNF-deficient NCI-H460 cell lines is available in **Supplementary Table 6**.

H3K4me1 is another epigenetic mark associated with enhancer activity. SWI/SNF deficiency resulted in a general loss of H3K4me1 deposition at all regulatory regions and in all cell lines (**Figure 4.27.A**). Similar to what we had observed in the case of H3K27ac modification, *SMARCA4* deficiency resulted in the highest number of affected regions, and *ARID2* knock-down impacted a significantly higher number of enhancers compared to *ARID1A*-deficient cells (shARID1A: 2780; shARID2: 12943, shSMARCA4: 21958).

In addition to this, we observed a low overlap between those enhancers with changes in the deposition of H3K4me1. This finding reinforces the idea that each subunit is necessary to ensure the activity of specific enhancers, in this case indicated by the mono-methylation of lysine 4 on histone H3 (**Figure 4.27.B**). However, we reported some cases in which H3K4me1 regulation depended on more than one subunit of the complex. In line with this, we obtained a quite high overlap between *ARID2* and *SMARCA4*-deficient cells, but we also found some examples of enhancers whose mono-methylation relies on either ARID1A or ARID2.

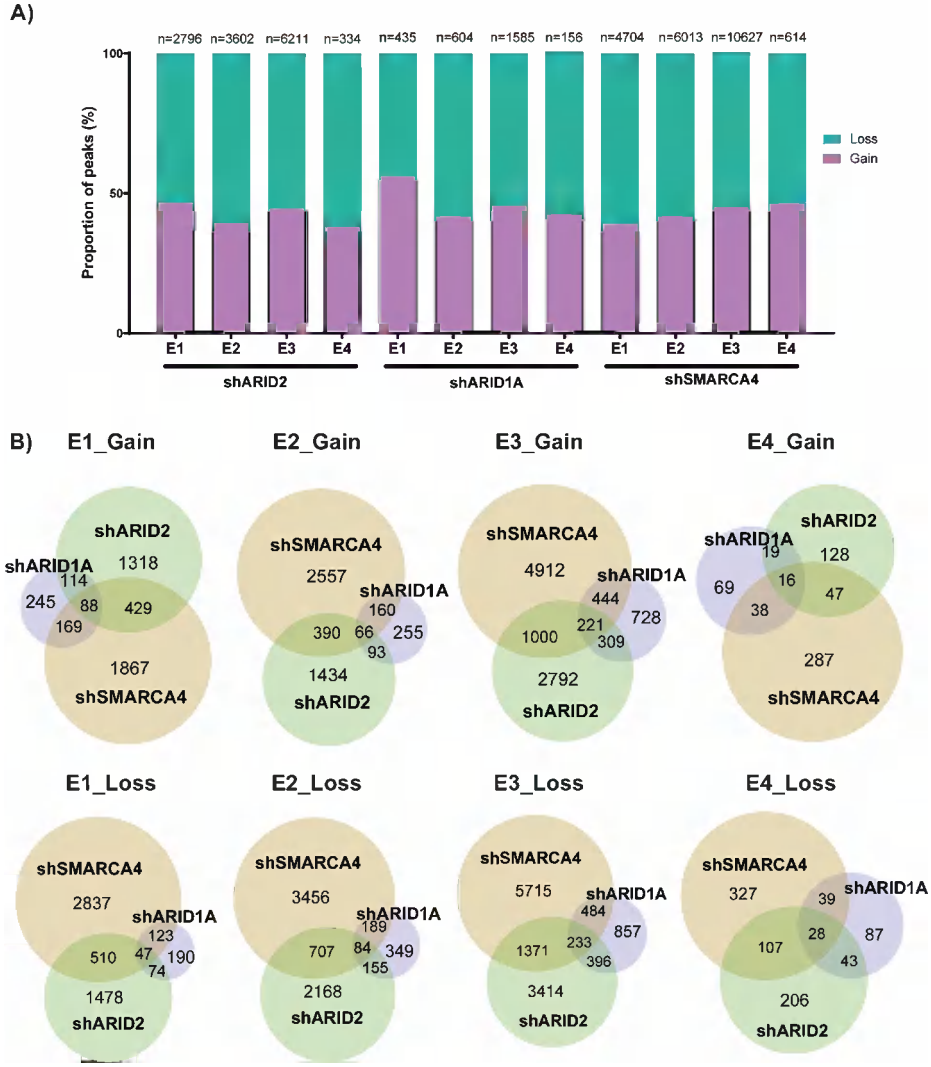


Figure 4.27. H3K4me1 alteration at enhancers identified in SWI/SNF-deficient lung cancer cell lines. A) Barplot representing the proportion of regions that gain (purple) or lose (green) acetylation following *ARID1A*, *ARID1B*, or *SMARCA4* knock-down in NCI-H460 cell lines in different types of enhancers identified with ChromHMM. E1 and 2: active enhancer; E3: primed enhancer; E4: bivalent enhancers. **B)** For each type of enhancer, Venn diagrams showing the overlap of enhancers that gain or lose H3K4me1 in the absence of different subunits of the SWI/SNF complex, shARID1A (purple), shARID2 (green), shSMARCA4 (yellow), requiring a minimum of 20% of overlap between two regulatory regions.

Finally, when we compared the number of affected regions by acetylation or mono-methylation, it seems that SWI/SNF complexes exert a more predominant role in the deposition of H3K4me1 at enhancers (H3K27ac shARID1A: 896, shARID2: 2362, shSMARCA4: 3983; H3K4me1 shARID1A: 2780; shARID2: 12943, shSMARCA4: 21958).

In summary, SWI/SNF alteration is associated with deep changes in the deposition of histone marks H3K27ac and H3K4me1 at enhancers, which suggests that their function might be compromised.

3.4. Alteration of developmental gene network in SWI/SNF-deficient lung cancer cell lines

Considering the high variability of the transcriptional programs and the chromatin structure changes observed in the different SWI/SNF-deficient cell lines, we wondered if that might be the result of the interaction between SWI/SNF and specific transcription factors that could have different target genes in each cellular context.

Accordingly, we selected the DEGs reported in the context of SWI/SNF deficiency, obtained their genomic coordinates, and performed a sequence-based motif analysis against the complete list of human promoters, in order to identify if the former presented an enrichment in a specific sequence pattern. The results of this analysis indicated that upregulated genes in *ARID1A* or *SMARCA4*-deficient NCI-H460 cells presented binding sites of Hox genes (Hoxa11, Hoxa13, Hoxd11, Hoxd13, HOXB13, and Hoxa9), suggesting a dysregulation of developmental gene networks (**Figure 4.28**). Among them, *HOXA9* and *HOXB13* are the most commonly altered genes in solid malignancies (Bhatlekar et al., 2014), and particularly, *HOXA13* is upregulated in lung adenocarcinomas compared to normal tissues (Deng et al., 2018).

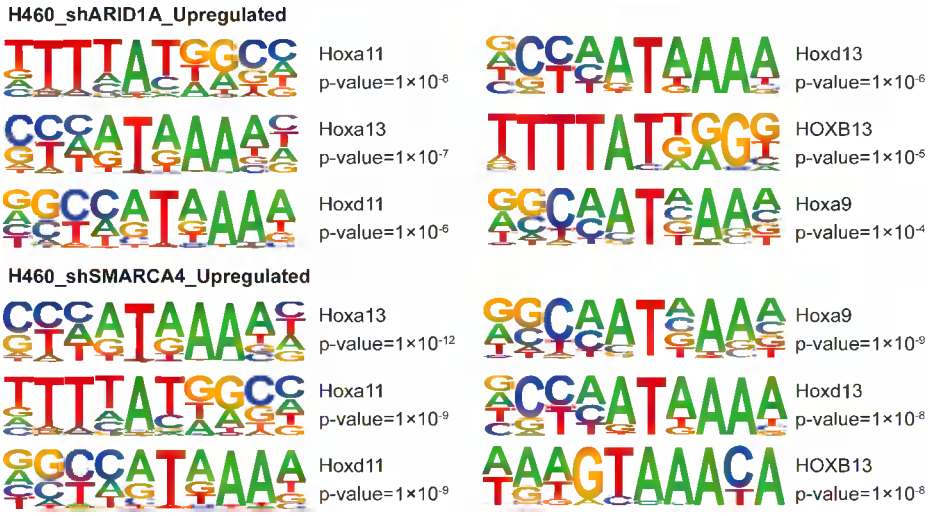


Figure 4.28. Enrichment of sequence motifs identified by HOMER in the promotors of DEGs in the context of SWI/SNF deficiency. The most significant motifs identified in each SWI/SNF-deficient lung cancer cell lines are shown.

Due to the potential alteration of the developmental gene expression program, we next examined if dysregulated genes as a result of SWI/SNF deficiency were associated with cell identity. To that end, we used a list of genes proposed to be specially involved in cellular differentiation, kindly provided by the Rada lab, which are defined by the presence of CpG islands on their promoter regions and H3K27me3 broad domains while they are inactive in pluripotency (the complete list of developmental genes can be found in **Supplementary Table 7**). For each SWI/SNF-deficient condition, we checked whether there was a significant enrichment in developmental genes among those identified as differentially expressed compared to the whole set of expressed genes.

Upregulated genes as a consequence of *ARID2* deficiency in the NCI-H460 cell line were significantly enriched in the developmental category (125/471 ~ 27%, *p-value* = 0.0006, one-tailed Fisher's exact test) (**Figure 4.29**).

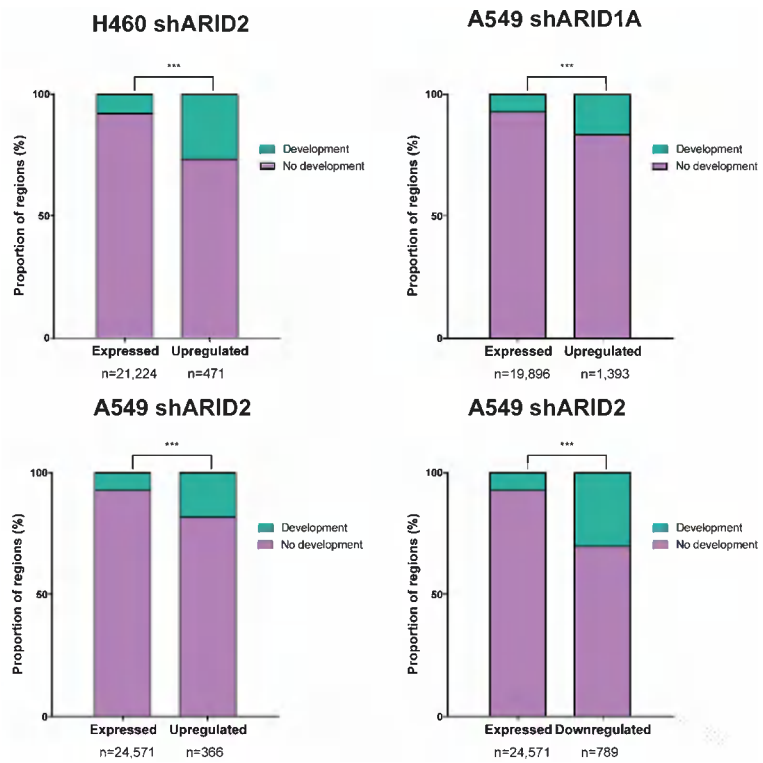


Figure 4.29. Enrichment of developmental category in DEGs as a result of SWI/SNF deficiency. Barplot representing the proportion of genes that overlap with developmental signature among those expressed and those dysregulated (up- or downregulated) in different SWI/SNF-deficient lung cancer models.

To validate the consistency of this alteration, we performed an equivalent analysis in the lung cancer cell model A549, where we observed that upregulated genes in the context of *ARID1A*-deficiency and dysregulated genes in *ARID2*-deficient cells presented a significantly higher proportion of developmental genes compared to the expected proportion by chance (**Figure 4.29**) (shARID1A upregulated 232/1393 ~ 17%, *p-value* = 0.0011, two-tailed Fisher's exact test; shARID2 upregulated 67/366 ~ 18%, *p-value* = 0.0309, two-tailed Fisher's exact test; shARID2 downregulated 239/789 ~ 30%, *p-value* = 0.0001, two-tailed Fisher's exact test).

Thus, our data indicate that SWI/SNF alteration might result in the perturbation of developmental gene networks. The fact that we have identified a significantly higher proportion of developmental genes in both upregulated and downregulated categories is probably revealing that SWI/SNF is required to keep a balance expression of this gene expression program.

Consequently, alterations in SWI/SNF composition likely result in a more unstable differentiation, promoting a cellular state which could explain the faster ability of cells to undergo the epithelial to mesenchymal transition in the context of SWI/SNF deficiency.

3.5. Identification of potential targets of SWI/SNF complexes in lung cancer

Once we had characterized the transcriptomic profile, the chromatin accessibility, and the epigenetic landscape of SWI/SNF-deficient lung cancer cell lines, we decided to propose candidate genes as direct SWI/SNF targets that might be directly involved in the final behavior of our cellular models. To that end, we looked for specific correlations between the data extracted from the different experiments performed in SWI/SNF-deficient cells.

Although we were not able to prove a direct correlation between gene expression changes and alterations in enhancer accessibility (**Figure 4.23**), we found some particular cases that accomplished both conditions. Thus, in the particular case of the A549 cell line, *ARID2* deficiency resulted in the downregulation of *MTSS1*, a well-described metastasis inhibitor (Kayser et al., 2015), as well as *SDK1*, a molecule involved in cellular adhesion (**Figure 4.30.A**).

The decreased expression of both genes correlated with a significant reduction of chromatin accessibility on their corresponding enhancers, identified as part of the GeneHancer project (**Figure 4.30.B**). We additionally downloaded from the ENCODE database ChIP-Seq experiments performed in the same cellular model using antibodies against histone modifications indicative of enhancer activity (H3K27ac and H3K4me1). Thus, we observed that those regions identified as theoretical enhancers of *MTSS1* and *SDK1* genes presented a strong signal of both histone modifications in the ENCODE data, confirming that both regions are indeed active enhancers on normal conditions.

Therefore, we propose that *MTTS1* and *SDK1* expressions are modulated by ARID2-containing SWI/SNF complexes, which are essential to keep an open chromatin structure at their corresponding enhancers.

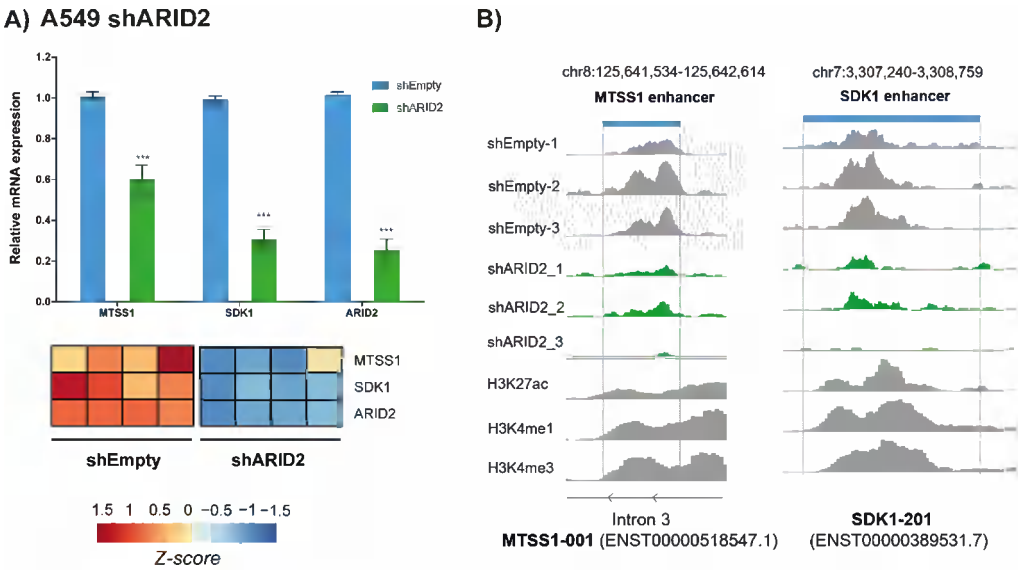


Figure 4.30. Correlation between *MTSS1* and *SDK1* expression and accessibility of their theoretical enhancers. A) Bar graph representation of the relative mRNA expression of *MTSS1*, *SDK1*, and *ARID2* genes in *ARID2*-deficient A549 cells obtained by qRT-PCR. The relative mRNA expression is plotted as mean ± SEM of at least three biological replicates (two-tailed *t*-test **p < 0.05, **p < 0.01*, and ****p < 0.001*). Bellow a heatmap representing the expression of these genes obtained by RNA-Seq experiments is shown. **B)** Visualization of read alignments for the different replicates of our ATAC-Seq experiments in *ARID2*-deficient A549 cells for the regions identified as enhancers of *MTSS1* (chr8:125,641,534-125,642,614) and *SDK1* (chr7:3,307,240-3,308,759) genes according to the GeneHancer database (Fishilevich et al., 2017). Read alignments of ChIP-Seq experiments performed in wild-type cells against different histone modification marks extracted from the ENCODE project are also represented (H3K27ac: ENCSR000AUI; H3K4me1: ENCSR000AUM; H3K4me3: ENCSR000ASH).

Following the same approach on SWI/SNF-deficient NCI-H460 cells lines, we found a significant decrease in the expression of the *SCGB3A2* gene that correlated with reduced accessibility of its enhancer in the absence of either *ARID1A*, *ARID2*, or *SMARCA4* (**Figure 4.31**). *SGB3A2* is a secretoglobin produced in the airway epithelial cells that has anti-inflammatory, anti-fibrotic, and anti-cancer activity (Kimura et al., 2022). Thus, regarding its tumor suppressor role, *SGB3A2* loss could partially explain why SWI/SNF-mutated lung tumors present worse clinical outcome (Peinado et al., 2020).

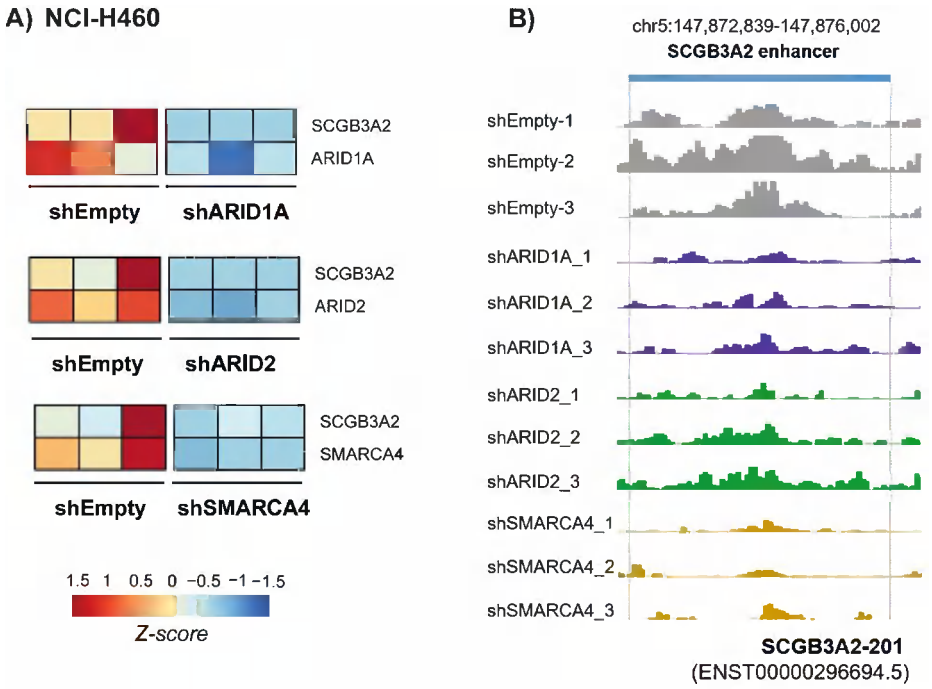


Figure 4.31. Correlation between *SCGB3A2* expression and accessibility of its theoretical enhancer. A) Heatmap representing the expression of *SCGB3A2* gene in SWI/SNF-deficient NCI-H460 and control cells obtained by RNA-Seq experiments. **B)** Visualization of read alignments for the different replicates of our ATAC-Seq experiments in *ARID1A, ARID2,* or *SMARCA4*-deficient NCI-H460 cells for the region identified as an enhancer of *SCGB3A2* (chr5:147,872,839-147,876,002).

We next focused on the establishment of associations between genes and enhancers identified using ChromHMM in order to identify which are the genes under the influence of each enhancer. To that end, we studied the correlation between gene expression alterations and H3K27ac levels of enhancers belonging to the same TAD, as a higher

probability of interaction is expected considering that both elements are physically proximal in the three-dimensional conformation of DNA. A complete list indicating the genomic coordinates of TADs identified in the NCI-H460 cell line, the corresponding TAD of human genes and enhancers identified with ChromHMM can be found in **Supplementary Table 5**.

As an example, we found that in the absence of *ARID2*, there is a downregulation of *NRG1*, which is a ligand of tyrosine kinase receptors ErbB3 and ErbB4, together with a decrease in the acetylation of an enhancer belonging to the same TAD (TAD_0840). Therefore, we hypothesize that ARID2 regulates the expression of the *NRG1* gene by maintaining the acetylation of its enhancer (**Figure 4.32**).

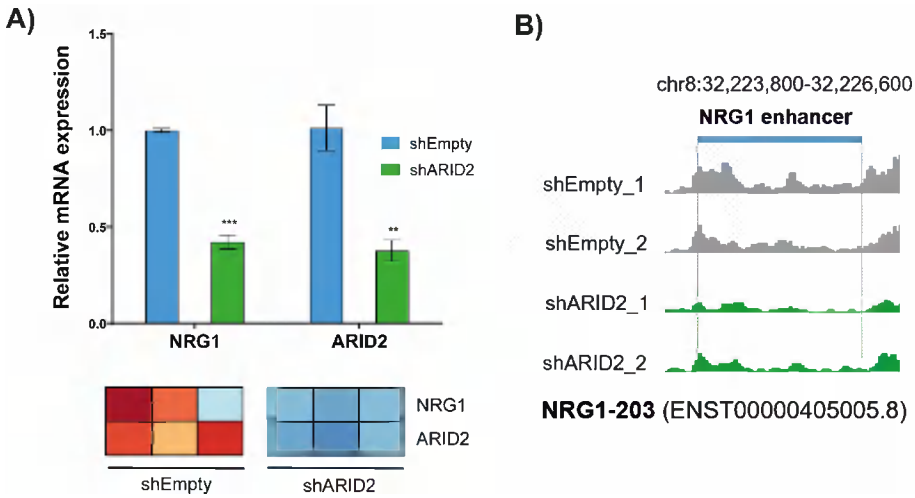


Figure 4.32. Correlation between *NRG1* expression and acetylation of its potential enhancer. A) Bar graph representation of the relative mRNA expression of *NRG1* and *ARID2* genes in *ARID2*-deficient NCI-H460 cells obtained by qRT-PCR. The relative mRNA expression is plotted as mean ± SEM of at least three biological replicates (two-tailed *t*-test **p < 0.05, **p < 0.01,* and ****p < 0.001*). Bellow a heatmap representation of the expression of both genes obtained by RNA-Seq experiments is shown. **B)** Visualization of read alignments for the different replicates of our CUT&RUN experiments against H3K27ac performed in *ARID2*-deficient NCI-H460 cells in a genomic region identified as a possible enhancer of *NRG1* gene (chr8:32,223,800-32,226,600).

In addition to this, we prepared the **Supplementary Table 2** containing the normalized counts and the differential gene expression analysis of SWI/SNF-deficient NCI-H460 cell lines, indicating also the TAD where each gene belongs.

Subsequently, we searched for alterations that were conserved among cell lines. We found that, in the context of SWI/SNF deficiency, there was a consistent decrease in the expression of *EPGN*, one of the ligands of EGFR (**Figure 4.33.A**).

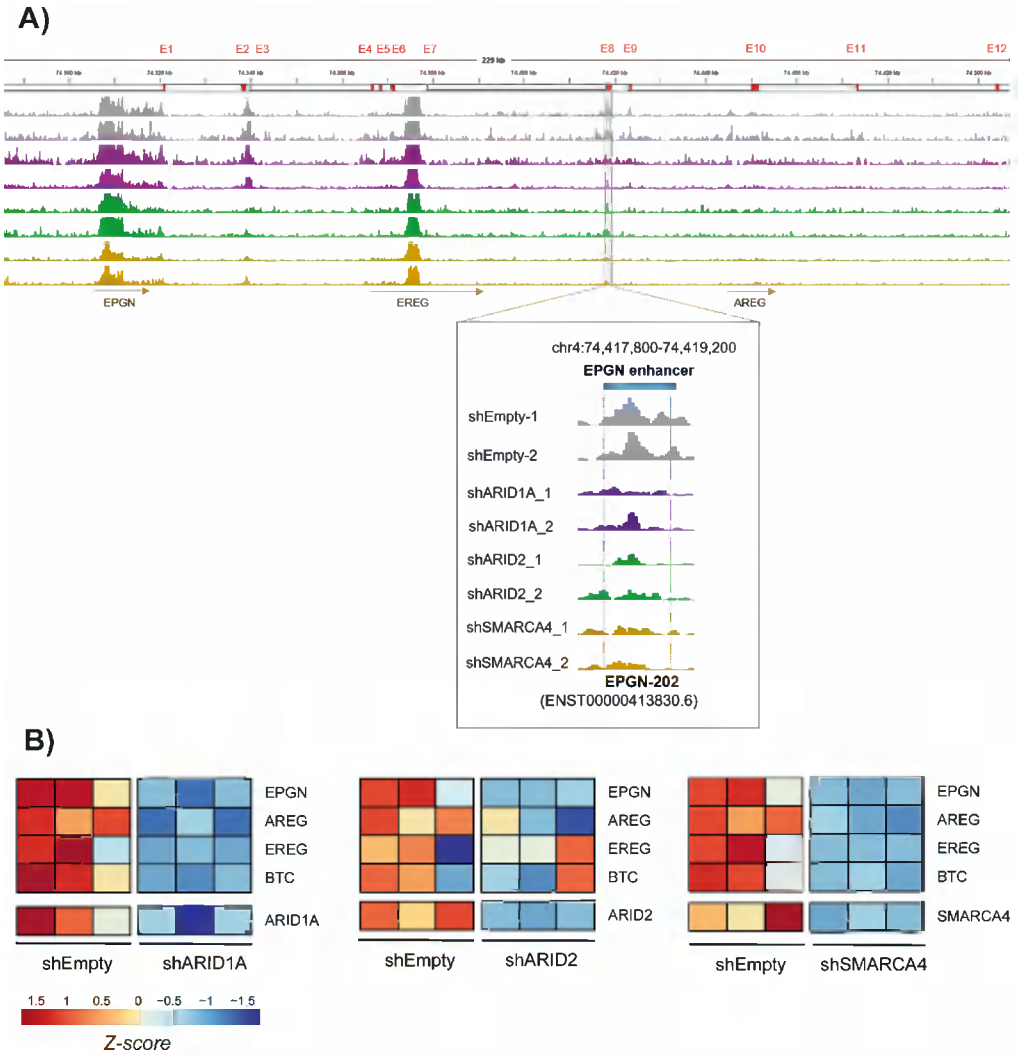


Figure 4.33. Correlation between the expression of EGFR ligands and the acetylation of their potential enhancer. A) Visualization of read alignments for the different replicates of our CUT&RUN experiments against H3K27ac performed in SWI/SNF-deficient NCI-H460 cells (shARID1A: purple, shARID2: green, shSMARCA4: yellow) in a genomic region identified as a possible enhancer of genes encoding EGFR ligands (chr4:74,417,800-74,419,200). **B)** Heatmap representing the relative mRNA expression of some ligands of EGFR (*EPGN*, *AREG*, *EREG*, and *BTC*) in *ARID1A*, *ARID2*, or *SMARCA4*-deficient cells obtained by RNA-Seq experiments.

In addition to this, we reported a significant reduction in the acetylation of an enhancer that is located within the same TAD (TAD_0444) (**Figure 4.33.B**). Interestingly, this TAD contains the coding genes of other ligands of EGFR, some of which were significantly downregulated in *ARID1A* or *SMARCA4*-deficient cells (*AREG, EREG*, and *BTC*). Therefore, we hypothesize that SWI/SNF mediates the expression of some ligands of EGFR by maintaining the activity of their potential enhancer.

Moreover, we observed additional mechanisms contributing to SWI/SNF-mediated gene expression regulation, as is the case of a loss of accessibility at the *EPGN* promoter in *ARID1A* or *ARID2*-deficient cells, which was likewise accompanied but the downregulation of that gene in both cellular models (**Figure 4.34**).

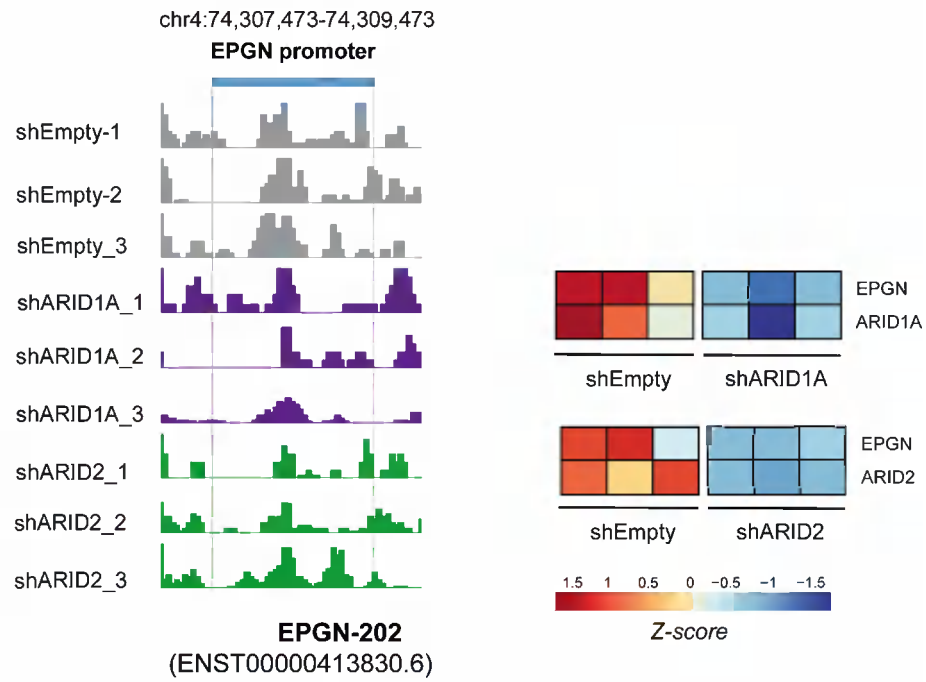


Figure 4.34. Correlation between the accessibility of *EPGN* promoter and its gene expression. Visualization of read alignments for the different replicates of our ATAC-Seq experiments performed in SWI/SNF-deficient NCI-H460 cells (shARID1A: purple, shARID2: green) in the promoter of *EPGN* gene (chr4:74,307,473-74,309,473). On the right, a heatmap representing the decreased expression of *EPGN* in *ARID1A* or *ARID2*-deficient cells obtained by RNA-Seq experiments.

With the final aim of understanding the molecular mechanisms that explained the final transcriptomic profiles of SWI/SNF-deficient cells, we looked for changes in the expression of genes and in the activity of enhancers relevant to the epithelial to mesenchymal transition, as it was the main recurrently altered pathway.

Thus, in *SMARCA4*-deficient NCI-H460 cells we observed a decrease in the expression of transcription factors *KLF6* and *SOX6*, both of which have been reported to impair the epithelial to mesenchymal transition (Jiang et al., 2018; Wang et al., 2021). In order to know how SWI/SNF could mediate the expression of both genes, we looked for changes in the activity of the enhancers belonging to the same TAD. As it can be distinguished in **Figure 4.35.A** and **Figure 4.35.B**, we reported a loss of H3K27ac deposition at their theoretical enhancers.

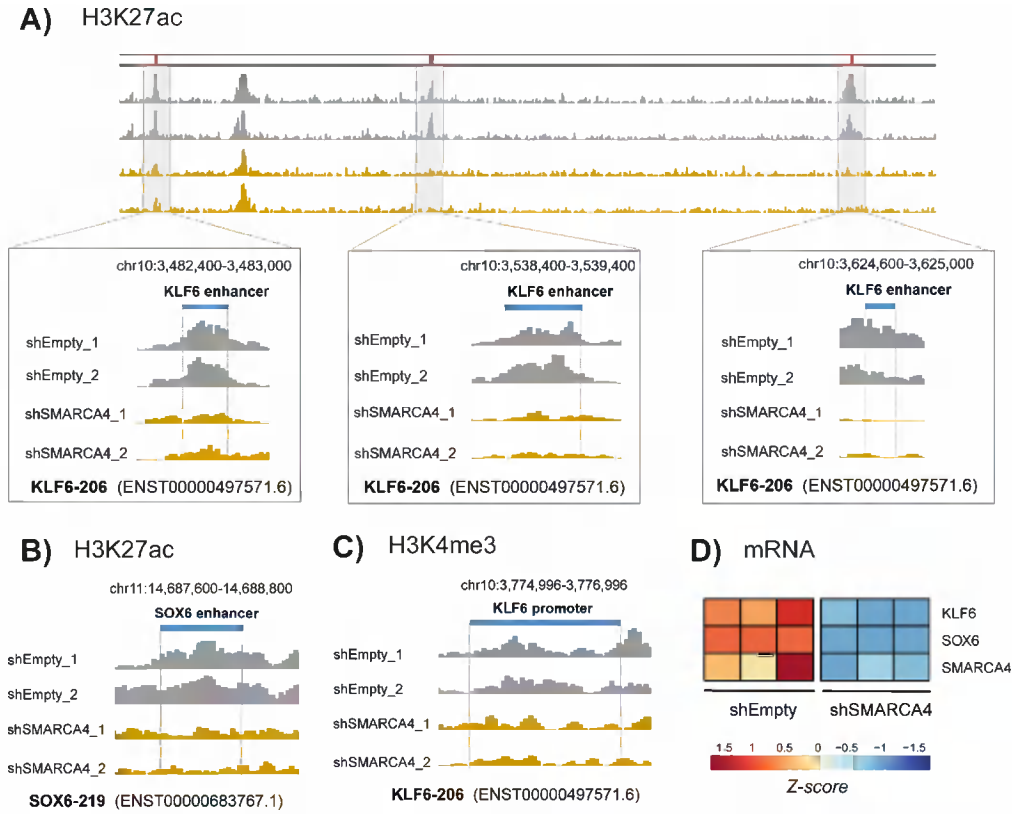


Figure 4.35. Correlation between *KLF6* and *SOX6* expression and histone epigenetic profile. Visualization of read alignments for the different replicates of our CUT&RUN experiments against H3K27ac performed in *SMARCA4*-deficient NCI-H460 cells in (**A**) three genomic regions identified as potential enhancers of *KLF6* (chr10:3,482,400-3,483,000; chr10:3,538,400-3,539,400;

chr10:3,624,600-3,625,000) and (**B**) a genomic region identified as a possible enhancer of *SOX6* gene (chr11:14,687,600-14,688,800). **C)** Visualization of read alignments for the different replicates of our CUT&RUN experiments against H3K4me3 performed in *SMARCA4*-deficient H460 cells in the promoter of *KLF6 gene* (chr10:3,774,996-3,776,996). **D)** Heatmap representing the relative mRNA expression of *KLF6* and *SOX6* genes in *SMARCA4*-deficient cells obtained by RNA-Seq experiments.

In the particular case of the *KLF6* gene, we found three putative enhancers at different locations upstream the TSS (150-300 kb) that presented a significant decrease in H3K27ac deposition. Besides the large genomic distance, potential enhancers could become close to their target gene in the three-dimensional structure of DNA due to the formation of chromatin loops. In addition to this, we also noticed a significant decrease in H3K4me3 deposition at the *KLF6* promoter (**Figure 4.35.C**).

Together, our results seem to indicate that SWI/SNF complex mediates the expression of specific genes by ensuring a proper chromatin conformation and an active histone epigenetic pattern at regulatory regions. However, further experiments are required to prove if SWI/SNF complex binds directly to these genomic regions, as well as the implication of all these genes and pathways in lung cancer progression.

DISCUSSION

“Research is to see what everybody else has seen, and to think what nobody else has thought.”

– Albert Szent-Györgyi

1. Characterization of the molecular mechanisms behind the role of SWI/SNF alteration in lung cancer development

Lung cancer remains the leading cause of cancer-related deaths globally and shows one of the lowest survival rates. Therefore, there is a clear need to better understand the molecular mechanisms driving tumorigenesis in this tumor type in order to develop more effective therapies. In this context, the presence of frequent alterations in genes encoding subunits of the SWI/SNF complex represents a great research opportunity to decipher its role in tumor progression and to exploit these deficiencies for the development of antitumoral treatments.

SWI/SNF complexes mobilize nucleosomes at target promoters and enhancers but are also able to interact with transcription factors, coactivators, and corepressors to modulate gene expression (Liao et al., 2017). In addition to this, SWI/SNF complexes exert key roles in the activation of differentiation and the suppression of proliferative programs of many cellular lineages (Mittal and Roberts, 2020).

In this doctoral thesis, we have characterized the transcriptomic, the chromatin conformation, and the epigenetic landscape of SWI/SNF-deficient lung cancer cell lines with the final aim of improving our knowledge of the molecular mechanisms involved in tumor progression.

Accordingly, we have observed that SWI/SNF alteration results in an intense tissue-specific transcriptional reprogramming of cells characterized by thousands of alterations, which is highly variable depending on both the eliminated subunit and the specific tissue. In addition to this, SWI/SNF deficiency was associated with the alteration of both shared and tissue-specific molecular pathways.

In particular, our transcriptomic analysis showed a very low overlap between DEGs after removing different subunits of the complex in the same cell line or after downregulating the expression of the same subunit of the complex in different cellular contexts. Both results revealed that the majority of transcriptional alterations tend to be subunit- or tissue-specific, respectively.

Consequently, we hypothesize that target genes of SWI/SNF complexes rely not only on their subunit composition, but also on their interaction with tissue-specific transcription factors. Another important point to keep in mind is that the final effect of SWI/SNF activity might depend on its interaction with other cellular components, such as coactivators, and corepressors. With this respect, alteration of specific subunits of the complex might affect its interaction with transcription factors or coregulators, impairing SWI/SNF recruitment to specific genomic loci.

In most cases, cBAF alteration resulted in a higher number of dysregulated genes. Considering that these subcomplexes are the most abundant in cells (Mashtalir et al., 2018), they are more likely to mediate the expression of more highly connected gene expression networks than PBAF complexes. In addition to this, it has been postulated that ARID1A and ARID2 have an opposite effect on transcriptional regulation, whereas ARID1B and ARID2 are required for the repression of hundreds of genes in the context of liver cancer (Raab et al., 2015). The low overlap among DEGs reported in our lung cancer cell models suggests that SWI/SNF subcomplexes regulate different gene networks in that particular cellular context, but we cannot completely rule out the possibility of cooperative or competitive relationships among subunits.

Concerning the mutually exclusive subunits of the complex, it could be inferred that ARID1A and ARID1B functions are not completely equivalent, as we have reported the dysregulation of thousands of genes in each knock-down cell line. This implies that the expression of the remaining paralog is not enough to cover the lack of expression of the inactivated subunit. In agreement with our results, it has been recently reported that there is a subset of ARID1A-dependent genes (p53 and estrogen receptor targets) in ovarian clear carcinoma whose expression cannot be compensated by the overexpression of ARID1B (Trizzino et al., 2018). Moreover, dual loss of ARID1A and ARID1B in the murine liver resulted in a significantly higher number of dysregulated genes compared to the individual knockouts (Z. Wang et al., 2020), suggesting synergistic effects between both subunits in gene expression regulation.

Similarly, as a result of *SMARCA4* deficiency, we obtained thousands of dysregulated genes, which might imply that SMARCA2 is not enough to cover the functions exerted by its paralog subunit, at least in our cellular models. In addition to this, it has been reported that SMARCA2 and SMARCA4 are able to interact with different transcription factors (Kadam and Emerson, 2003). This finding adds another layer of complexity to the transcriptional regulation exerted by SWI/SNF complexes.

Despite all this complexity, we have been able to identify a common enrichment of several signaling pathways in the context of SWI/SNF deficiency. In most cases, the specific dysregulated genes belonging to these functional categories were not conserved, which seems to indicate that perhaps SWI/SNF complexes are important for the regulation of specific mechanisms or biological functions rather than particular genes. In line with this, there is considerable evidence supporting the role of SWI/SNF complexes in maintaining the cellular identity, but how cells reach this final stage of differentiation varies in each cell type. It is therefore likely that SWI/SNF complexes are in charge of the regulation of different set of genes in each cellular context that may have similar molecular functions.

Another possibility to consider is whether all these transcriptional alterations are directly or indirectly caused by SWI/SNF alteration, as we have conducted the experiments for several days to ensure a proper downregulation of the target gene. In any of the possible scenarios, to our knowledge, this is the first study that characterizes whether SWI/SNF-regulated gene networks are conserved or specific across human tumor types.

Concerning the commonly altered signaling pathways, most SWI/SNF-deficient cells showed evidence of apparently undergoing the epithelial to mesenchymal transition and in some cases, we reported a perturbation of DNA damage response (**Table 4.1**). Both molecular processes are especially relevant in this context, not only because they provide a possible explanation for the increased oncogenic properties of SWI/SNF-mutant tumors, but also because they open up the opportunity to explore potential therapeutic vulnerabilities.

The epithelial to mesenchymal transition appeared among the positively enriched pathways in SWI/SNF-deficient cells, but in the majority of the cases, this enrichment was not statistically significant. This result might be explained by the low number of samples in each biological condition, the small size of the gene sets and/or the high variability among samples. However, the prominent role of SWI/SNF in maintaining a differentiated state prompted us to study this process in further detail. In line with this, dual loss of ARID1A and ARID1B in a murine liver cancer model led to a prominent decrease of genes involved in differentiation (Z. Wang et al., 2020). More interestingly, and in agreement with our findings, ARID1A loss in a pancreatic cellular model resulted in the activation of the epithelial to mesenchymal transition and a stem-like transcriptional program, together with increased migratory and invasive properties (W. Wang et al., 2019).

In spite of all the reported evidence, further experimental validations are required to decisively prove the activation of the epithelial to mesenchymal transition in our SWI/SNF-deficient cellular models and its potential involvement in lung cancer progression.

It is also worth noting that SWI/SNF complexes are frequently mutated across a wide spectrum of cancers, but some subunits tend to be specifically altered in certain tumor types (for instance, *SNF5* in malignant rhabdoid tumors, *SMARCA4* in small cell ovarian carcinomas, or *PBRM1* in renal cell carcinomas). Therefore, there might be conserved mechanisms mediating gene expression regulation by the distinct subunits of SWI/SNF complexes that, as a result of their disruption, lead to tumorigenesis. It may seem controversial that the loss of a single subunit of the complex could promote cancer development. However, it is important to keep in mind that it has been reported that the alteration in just one subunit results in the loss of activity at enhancers mediating the expression of genes linked to differentiation and development, together with an activation of super-enhancers important for cell survival and invasion (Guo et al., 2022; Mathur et al., 2017; Nakayama et al., 2017; Wang et al., 2017; Wilson et al., 2020).

In addition to this, residual SWI/SNF subcomplexes resulted from the loss of different subunits are aberrantly targeted to other genomic loci (Nakayama et al., 2017; Wang et al., 2017), perturbating the homeostasis of SWI/SNF-mediated gene networks and the assembly of other subcomplexes. In relation to this, Wang and collaborators proved a prominent loss of ARID2 genome-wide binding sites upon dual loss of the accessory subunits ARID1A and ARID1B (Z. Wang et al., 2020).

Another possibility to contemplate is the potential assembly of new complexes besides the canonically described ones. In relation to this, ARID2 has been proposed to participate in DNA repair as part of complexes that do not contain the catalytic subunit SMARCA4 (de Castro et al., 2017; Oba et al., 2017). It has also been suggested that *ARID2* inactivation impairs the assembly of the PBAF complex beyond its core components (Sadek et al., 2022). However, this would not affect those functions that do not rely on the whole complex but are performed but the individual subunits instead. Moreover, SWI/SNF complexes can assemble without SMARCB1, SMARCA4, or ARID1A subunits (Doan et al., 2004; Katherine C Helming et al., 2014; Hoffman et al., 2014; Wilson et al., 2014), but the remaining complexes are aberrantly targeted to their corresponding genomic regions (Schick et al., 2019). This is particularly relevant in the context of tumorigenesis, as *ARID1A*, *SMARCA4*, or *SMARCB1* are frequently mutated in a loss-of-function manner and these subunits are crucial for targeting cell-specific enhancers (X. Wang et al., 2009).

Considering all this complexity, we propose a model on which SWI/SNF is required to regulate the expression of cell type-specific gene networks that are essential to keep the cellular identity thanks to its interaction with different transcription factors and regulators. As a result of SWI/SNF deficiency, the lineage identity losses its robustness, facilitating the transition to a mesenchymal phenotype.

2. Therapeutic implications of SWI/SNF-altered lung cancer patients

Any advance in the possibility of therapeutically exploiting vulnerabilities associated with SWI/SNF deficiency is of great interest considering that almost 25% of all human cancers harbor alterations in any of their components. Another point of active research is the study of the molecular mechanisms involved in resistance to treatments, which are becoming more frequent in the last years.

In this doctoral thesis, we have provided evidence about the importance of SWI/SNF mutational status in the assignment of the most appropriate treatment for lung cancer patients. Thus, we have reported that SWI/SNF-deficient cells presented an increased resistance to treatments based on the inhibition of EGFR, regardless of the mutational background of *KRAS* and *TP53* genes (**Figure 4.10**).

The transcriptomic analysis performed in SWI/SNF-deficient models revealed a recurrent alteration in the epithelial to mesenchymal transition (EMT). We hypothesize that this molecular process might explain the enhanced resistance to EGFR inhibition, as the promotion of a less differentiated state would allow cell survival without requiring the strong proliferative signals provided by the receptor of EGF.

The role of the EMT in the acquirement of resistance to epidermal growth factor tyrosine kinase inhibitors (EGFR-TKIs) in non-small cell lung cancer has been extensively characterized (Li et al., 2017; Weng et al., 2019; Zhu et al., 2019). In addition to this, alterations of SWI/SNF complex have also been linked with EMT in other studies. Thus, it has been suggested that *ARID1A* alterations are associated with an EMT expression signature (Wilson et al., 2019). Other researches have proved that ARID1A loss results in the upregulation of mesenchymal markers (fibronectin, vimentin, N-cadherin), while downregulating the expression of the epithelial marker E-cadherin, accompanied by changes in cell morphology and loss of cell polarity in different cellular contexts (renal, pancreatic, and breast cancer) (Somsuan et al., 2019; Tomihara et al., 2021; T. Wang et al., 2020; W. Wang et al., 2019).

Supporting the idea of a less robust epithelial identity of SWI/SNF-deficient cells, we have observed that *ARID2*-deficient cells treated with EGFR inhibitor dacomitinib upregulated faster many genes involved in mesenchymal development and cell migration compared to control cells, while decreasing the expression of molecules related to cellular adhesion (**Figure 4.11**). This finding is indicating a potential response triggered by the treatment with the EGFR inhibitor that is happening faster in the context of *ARID2* deficiency. In addition to this, untreated *ARID2*-deficient cells presented an upregulation of some indicators of the epithelial to mesenchymal transition (*FBLN5, CRLF1, COL4A1, CDH2, MMP2, VEGFC,* and *SLIT2*) compared to control cells, suggesting that the starting point of these cells might be a more mesenchymal phenotype. Accordingly, we propose that SWI/SNF-deficient cells are more prone to the activation of the epithelial to mesenchymal transition in response to the treatment, enabling them to tolerate higher concentrations of EGFR inhibitors.

Our findings could provide new insights to decipher the molecular mechanisms involved in resistance to treatments, which is especially relevant in this context, as EGFR inhibitors are one of the preferred therapeutic options for NSCLCs harboring EGFR-activating mutations (Thomas et al., 2015). In addition to this, the majority of patients that initially respond to EGFR-TKIs treatment inevitably develop resistance after less than a year (Weng et al., 2019).

Another therapeutic approach that we have addressed in this work is the sensitivity of SWI/SNF-deficient cells to treatments that cause DNA damage, due to positive enrichment in the DNA repair pathway revealed by our transcriptomic analysis. In particular, we have proved that *ARID2*-deficient cells present a higher sensitivity to the chemotherapeutic compounds cisplatin and etoposide (**Figure 4.18**).

In line with this, we observed an active role of *ARID2* in the detection and repair of *in vitro*-produced DNA damage in lung cancer cell lines with different mutational backgrounds. In particular, we demonstrated that *ARID2*-deficient cells present an important delay in the resolution of DNA damage foci, independently of *TP53* or *KRAS* activities (**Figures 4.14-4.16**).

In accordance with our results, other members of SWI/SNF complexes have been associated with different steps of DNA damage repair mechanisms (Lee et al., 2010; Niimi et al., 2012; Ray et al., 2009; Shen et al., 2015). Thus, as a result of SWI/SNF alteration, knock-down cells would present an impaired response to DNA damage, increasing their genomic instability and consequently, their sensitivity to treatments based on the generation of DNA damage. We, therefore, propose that the higher sensitivity of *ARID2*-deficient cells to chemotherapeutic agents and PARP inhibitors is the result of its involvement in DNA repair. In line with this, during the time of the elaboration of this doctoral thesis, it has been published that SWI/SNF-mutated lung adenocarcinomas present a higher tumor mutation burden compared to wild-type tumors (Peinado et al., 2022).

Considering that platinum-based chemotherapy is still widely used in the treatment of lung cancer patients with highly variable success (Chen et al., 2014), our results suggest that ARID2 expression might be explored as a stratification marker for the efficacy of chemotherapy. However, the toxicity of this drug limits its use, highlighting the need for other compounds with less side effects.

In relation to this, we showed that *ARID2* deficiency shows synthetic lethality with PARP inhibition using veliparib, an inhibitor that has shown favorable results in the treatment of breast cancer (Rugo et al., 2016) and is becoming included in several clinical trials, such as breast, ovarian, and, most importantly, lung cancer. Therefore, our findings seem to indicate that the stratification of lung cancer patients according to ARID2 expression could improve as well the efficiency of PARP inhibitors in NSCLC.

Our experimental observations collectively indicated that SWI/SNF-mutant lung cancer patients are more likely to benefit from treatments based on the promotion of genomic instability.

Functional analysis performed on SWI/SNF-deficient cellular models also revealed alterations of genes involved in immune response (**Table 4.1**), which might imply different sensitivities to immunotherapy approaches. In particular, we have reported an upregulation of genes expressed in response to interferon in A549 shARID1A, A549 shARID1B, and Caco-2 shARID1A cells. This seems to be a promising preliminary result because there is a vast body of literature supporting a positive correlation between the expression of IFN-stimulated genes and a favorable disease outcome in several tumor types (Zitvogel et al., 2015). In line with this, a recent study has shown that *ARID2*-deficient melanoma cells are particularly sensitive to immunotherapy through alterations in mTORC1 and IFNγ pathways (Pan et al., 2018).

Additionally, the higher tumor mutation burden reported in SWI/SNF-mutated tumors could result in a higher generation of neoantigens, and therefore tumor cells could be better targets for immunotherapy (Turajlic et al., 2017). All these findings support the potential use of SWI/SNF mutational status as a new stratification marker for personalized treatment in lung cancer patients.

Other therapeutic opportunities of SWI/SNF-mutant tumors that have not been fully covered in this doctoral thesis are the synthetic lethality approach among different subunits (Farnaby et al., 2019; Papillon et al., 2018), the use of EZH2 inhibitors (Januario et al., 2017), and bromodomain inhibitors (Gerstenberger et al., 2016; Schiaffino-Ortega et al., 2014; Theodoulou et al., 2016).

3. Molecular mechanisms involved in SWI/SNF-mediated gene expression regulation

We have demonstrated that SWI/SNF alteration is accompanied by specific chromatin structural changes at the complete genome, without a clear pattern of gain or loss of accessibility, which is highly variable depending on both the cellular context and the eliminated subunit (**Figure 4.19**). In addition to this, we did not observe a clear association between the proportion of transcriptional alterations (up- and downregulated genes) and the proportion of regions with altered chromatin accessibility, which is probably indicating that not all these genomic changes are affecting regions directly implicated in gene expression.

In most cases cBAF alteration affected the accessibility of more genomic regions (including enhancers) than PBAF complexes, supporting its broader role in gene expression regulation. In all cellular contexts, we reported a higher overlap between those regions with a significant loss of accessibility. What is more, these regions presented binding sites of AP-1 transcription factors, indicating that they were probably affecting enhancers.

In line with this, we proved that SWI/SNF is required to keep an open chromatin conformation around enhancers, in agreement with previous evidence (Alver et al., 2017; Kelso et al., 2017; Nakayama et al., 2017). Enhancer activity seems to have a more prominent role in maintaining the cell identity than promoter accessibility (Thurman et al., 2012), which could explain why the effect of SWI/SNF deficiency was mainly affecting enhancers.

Considering the global effect of SWI/SNF in ensuring enhancer accessibility, we hypothesized that gene expression alterations might correlate with changes in the accessibility of the enhancers linked to those genes according to the GeneHancer project. However, we were not able to see a clear correlation between the alterations of both variables in any of the lung or colorectal cancer models generated (**Figure 4.23** and **Figure 4.24**). This result indicated two aspects. First, that gene-enhancer relationships are difficult to establish and can be highly dependent on the cellular model, and, second, that gene expression regulation exerted by SWI/SNF complex is more intricate

and does not only rely on the accessibility of enhancers. Accordingly, we decided to study additional mechanisms mediating gene expression regulation, such as the epigenetic landscape of enhancers.

To our knowledge, there is no evidence of identified enhancers in the lung cancer cell line NCI-H460. Considering the tissue-specificity of these regulatory regions, we characterized the epigenetic landscape of our SWI/SNF-deficient cell lines integrating the genome-wide occupancy of several histone modifications (H3K27ac, H3K4me3, H3K4me1, and H3K27me3) with the ChromHMM software. As a result, we identified the genomic coordinates of the four types of described enhancers by the Kellis lab (Ernst and Kellis, 2017), characterized by different patterns of histone modifications, and hence potentially distinct biological functions.

SWI/SNF alteration affected the activity of enhancers, indicated by the deposition of histone modifications H3K27ac and H3K4me1 in their adjacent nucleosomes. The majority of the altered enhancers by changes in accessibility, acetylation, or mono-methylation tended to be subunit-specific, indicating that each subunit is required to ensure the activity of different regulatory regions.

It is important to highlight that, due to our experimental approach, we are not able to detect rapid changes in chromatin accessibility or epigenetic features, as several days of induction of the shRNA are required to achieve an effective knock-down of the target gene. This technical issue greatly limits our ability to detect direct SWI/SNF targets.

We hypothesize that SWI/SNF-mediated gene expression regulation might be affected by the interplay with specific transcription factors, which could explain the high variability of the transcriptional program and the chromatin structure of SWI/SNF-deficient cell lines. In line with this, the promoters of dysregulated genes in certain contexts presented binding sites of the Hox genes (**Figure 4.28**). We have additionally proved that in some models dysregulated genes were significantly associated with development compared to the whole set of expressed genes (**Figure 4.29**). Altogether, our findings seem to indicate that SWI/SNF complexes are required to ensure the homeostasis of developmental networks.

Consistently, its alteration would affect the balanced expression of genes responsible for keeping the cell linage identity and the differentiation state. As we have not observed a consistent alteration in developmental networks in all the cellular models, SWI/SNF deficiency might be necessary but not sufficient to disrupt the tight regulation of this transcriptional program.

We next integrated the information extracted from different omic approaches to identify potential targets of SWI/SNF complex in the context of lung cancer.

Thus, in our lung cancer cellular models *ARID2* is required to keep an open chromatin structure at *MTSS1* and *SDK1* enhancers, evidenced by a loss of accessibility correlated with decreased expression of their target genes (**Figure 4.30**). *MTSS1* is a well-described migration and invasion inhibitor, which has been associated with worse prognosis in several tumor types, and *SDK1* plays important roles in cell to cell adhesion (Giacobbe et al., 2016; Kayser et al., 2015; Taylor et al., 2018). The deficiency of both genes provides a plausible explanation for the higher migration and invasion capabilities of *ARID2*-deficient cells, together with the increased oncogenic potential in animal models (**Figure 4.8**).

Similarly, in SWI/SNF-deficient NCI-H460 cells we described a downregulation of some ligands of tyrosine kinase receptors (*NRG1* and *EPGN*) that correlated with a decrease in the acetylation of a potential enhancer belonging to the same TAD (**Figure 4.32** and **Figure 4.33**).

Thus, we propose that SWI/SNF complex is able to mediate the expression of specific genes by maintaining an accessible conformation or an active state of their corresponding enhancers. Our results should be carefully interpreted as they suggest potential targets of SWI/SNF complexes. Nevertheless, further research is required to effectively confirm if SWI/SNF complexes are in charge of the regulation of these genes by directly binding to their regulatory regions. Other methodological approaches to map the three-dimensional structure of chromatin (e.g. live-cell imaging, chromosome conformation capture, or genome architecture mapping) would be of great interest to identify enhancer-promoter contacts (Kempfer and Pombo, 2020).

In addition to these examples of SWI/SNF-mediated gene expression regulation, some of our findings collectively provided some evidence about the molecular mechanisms explaining the higher resistance to EGFR inhibition reported in SWI/SNF-deficient cells.

In this way, we found a consistent decrease in the expression of *EPGN*, which was accompanied by the downregulation of other ligands of tyrosine kinase receptors in specific cell lines, such as *AREG*, *EREG*, *NRG1,* or *BTC* (**Figure 4.33**). In addition to this, in *SMARCA4*-deficient cells, we reported a specific downregulation of the transcription factors *KLF6* and *SOX6* (**Figure 4.35**), both of which impair the epithelial to mesenchymal transition (Jiang et al., 2018; Wang et al., 2021). These findings were accompanied by a decrease in the deposition of H3K27ac of potential enhancers belonging to the same TAD.

It has been previously demonstrated that SWI/SNF alteration results in the dysregulation of enhancers important for cell differentiation (Jones et al., 2020; Nakayama et al., 2017; Schick et al., 2019; Wang et al., 2017; X. Wang et al., 2009). In line with this, we observed an upregulation of specific mesenchymal markers (different in each SWI/SNF-deficient cell line) that collectively turned the epithelial to mesenchymal transition into one of the main recurrently altered signaling pathways. Taken as a whole, our experimental observations suggest that, in the context of SWI/SNF deficiency, cells are able to transient more easily to a less differentiated state that is less dependent on the epidermal signaling pathways, such as those initiated by EGFR. Consequently, cells might decrease the expression of specific ligands of these receptors, as they do not longer rely on these molecular pathways. We, therefore, propose that SWI/SNF-deficient cells undergo faster the epithelial to mesenchymal transition and increase their resistance to those treatments based on the inhibition of EGFR (**Figure 5.1**).

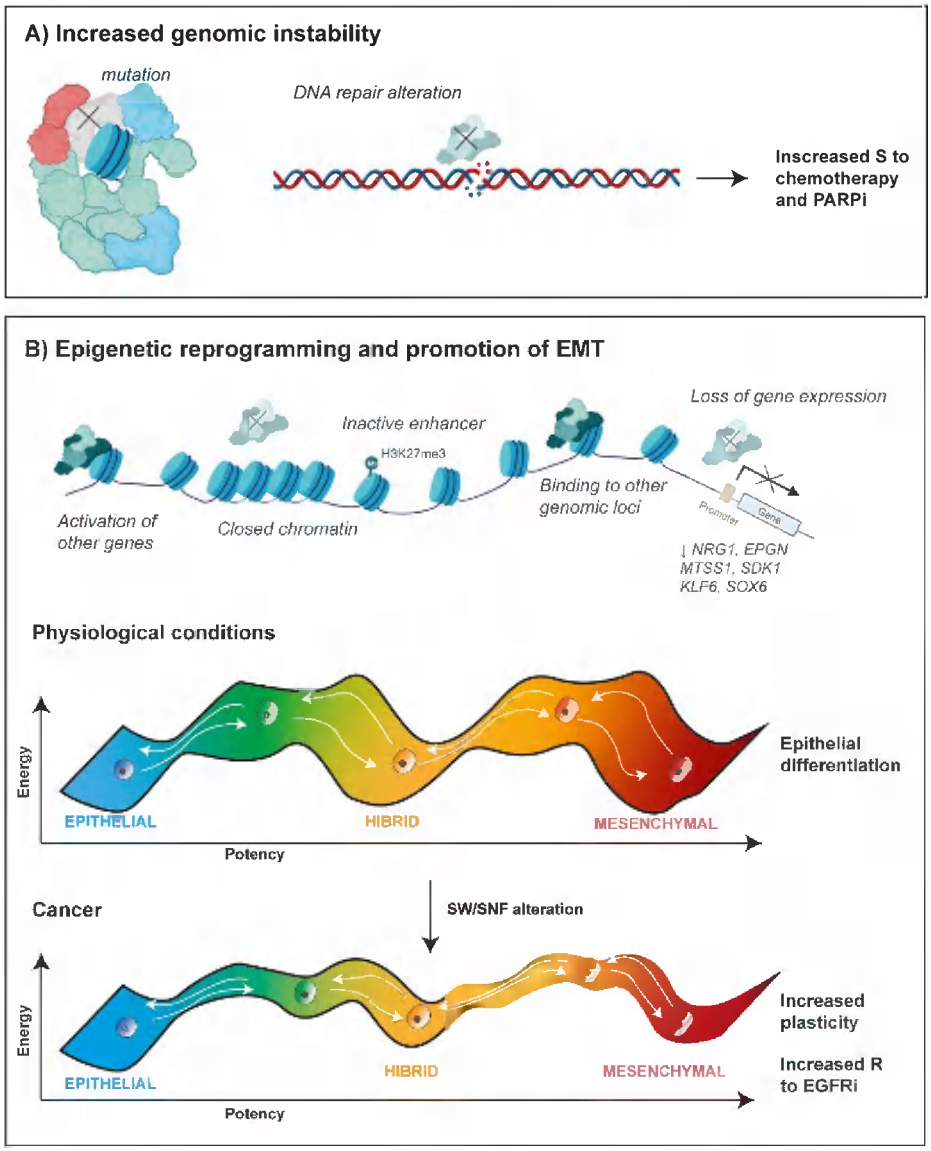


Figure 5.1. Summary of the pro-tumoral mechanisms mediated by SWI/SNF alterations. A) Alteration of DNA repair mechanisms increases cell sensitivity to chemotherapeutic compounds and PARP inhibitors. **B)** Epigenetic reprogramming of cells favors the epithelial to mesenchymal transition. From a mechanistic point of view, SWI/SNF alteration affects its genome-wide location, the epigenetic pattern and the chromatin accessibility of cells, resulting in an imbalanced gene expression, specially affecting lineage determinant genes. As a consequence, SWI/SNF-deficient cells undergo easier and faster the epithelial to mesenchymal transition and become resistant to the inhibition of EGFR. Abbreviations: S: sensitivity; R, resistance.

Complementary to our findings, during the time of the elaboration of this doctoral thesis, the Weissman lab demonstrated that SMARC4 re-expression in an ovarian cellular model induced an epithelial signature and the promoters of the highly upregulated genes contained binding sites of transcription factor members of the AP-1 family (Orlando et al., 2020). Thus, the alteration of SWI/SNF complexes might affect the expression of genes important for epithelial cell identity, resulting in a less differentiated state.

Regarding the role of SWI/SNF complexes in ensuring the accessibility and the activity of enhancers, and the presence of AP-1 binding motifs on the affected regions, it seems that both elements cooperate with other mediators inside the cell to regulate gene expression. In agreement with this idea, SWI/SNF complexes interact with the AP-1 family of transcription factors (some members present a variable pattern of expression in response to extracellular stimuli) and cell type-specific transcription factors to select cell type-specific enhancers during the process of differentiation (Vierbuchen et al., 2017). Accordingly, all this complexity could explain the low overlap between the dysregulated genes in the different cellular models generated.

CONCLUSIONS

“Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world.”

– Louis Pasteur

1. SWI/SNF alteration results in an intense tissue-specific transcriptional reprogramming of cells characterized by thousands of changes in gene expression, without a clear pattern of up- or downregulation.

2. SWI/SNF complexes mediate the expression of different gene networks, depending on their composition and the cellular context.

3. cBAF complexes are required for the regulation of deeper gene networks than PBAF complexes.

4. SWI/SNF deficiency is accompanied by the potential activation of an epithelial to mesenchymal transition program in almost all analysed cellular models.

5. *EGFR* and SWI/SNF mutations are mutually exclusive in LUAD, which suggests the existence of either synthetic lethality relationships or redundant tumor progression-promoting activities.

6. SWI/SNF alteration results in an increased resistance to EGFR inhibitors, irrespective of the mutational status of *KRAS* or *TP53*.

7. *ARID2*-deficient lung cancer cell lines present a delay in the resolution of DNA damage foci upon treatment with etoposide, which is independent of the presence of *KRAS* or *TP53* mutations.

8. In the context of lung cancer, *ARID2* deficiency gives rise to an enhanced sensitivity to chemotherapeutic agents and PARP inhibitors.

9. SWI/SNF alteration is accompanied by specific chromatin structural changes at the complete genome, with a general loss of accessibility affecting enhancers.

10. Changes in chromatin accessibility show high variability depending on the subcomplex composition and the cellular context.

11. In general terms, gene expression changes resulted from SWI/SNF alteration do not correlate with changes in chromatin accessibility at their associated enhancers according to the GeneHancer project. This suggests additional mechanisms mediating gene expression regulation.

12. SWI/SNF alteration results in deep changes in the deposition of H3K27ac and H3K4me1 at enhancers.

13. We have generated a list of potential direct targets of SWI/SNF on which the activity of the complex is required to ensure a proper chromatin conformation and an active histone epigenetic profile at enhancers and promoters in charge of the regulation of specific genes (*MTSS1*, *SDK1*, *SCGB3A2*, *EPGN*, *NRG1*, *KLF6*, and *SOX6*).

REFERENCES

“Everything is theoretically impossible, until it is done.”

– Robert A. Heinlein.

Akhtar, N., Bansal, J.G., 2017. Risk factors of Lung Cancer in nonsmoker. Current Problems in Cancer 41, 328–339. https://doi.org/10.1016/j.currproblcancer.2017.07.002

Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., Durbin, R., 2011. Dindel: Accurate indel calls from short-read data. Genome Res 21, 961–973. https://doi.org/10.1101/gr.112326.110

Alexander, M., Kim, S.Y., Cheng, H., 2020. Update 2020: Management of Non-Small Cell Lung Cancer. Lung 198, 897–907. https://doi.org/10.1007/s00408-020-00407-5

Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., Campbell, P.J., Vineis, P., Phillips, D.H., Stratton, M.R., 2016. Mutational signatures associated with tobacco smoking in human cancer. Science 354, 618–622. https://doi.org/10.1126/science.aag0299

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D.T.W., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., López-Otín, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N.J., Valdés-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., Stratton, M.R., 2013. Signatures of mutational processes in human cancer. Nature 500, 415–421. https://doi.org/10.1038/nature12477

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., Roberts, C.W.M., 2017. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nat Commun 8, 14648. https://doi.org/10.1038/ncomms14648

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andersson, R., Sandelin, A., 2020. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet 21, 71–87. https://doi.org/10.1038/s41576-019-0173-8

Andrey, G., Mundlos, S., 2017. The three-dimensional genome: regulating gene expression during pluripotency and development. Development 144, 3646–3658. https://doi.org/10.1242/dev.148304

Audia, J.E., Campbell, R.M., 2016. Histone Modifications and Cancer. Cold Spring Harb Perspect Biol 8, a019521. https://doi.org/10.1101/cshperspect.a019521

Barker, N., Hurlstone, A., Musisi, H., Miles, A., Bienz, M., Clevers, H., 2001. The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation. EMBO J 20, 4935–4943. https://doi.org/10.1093/emboj/20.17.4935

Barta, J.A., Powell, C.A., Wisnivesky, J.P., 2019. Global Epidemiology of Lung Cancer. Ann Glob Health 85, 8. https://doi.org/10.5334/aogh.2419

Beagan, J.A., Phillips-Cremins, J.E., 2020. On the existence and functionality of topologically associating domains. Nat Genet 52, 8–16. https://doi.org/10.1038/s41588-019-0561-1

Bejjani, F., Evanno, E., Zibara, K., Piechaczyk, M., Jariel-Encontre, I., 2019. The AP-1 transcriptional complex: Local switch or remote command? Biochim Biophys Acta Rev Cancer 1872, 11–23. https://doi.org/10.1016/j.bbcan.2019.04.003

Bell, E.H., Chakraborty, A.R., Mo, X., Liu, Z., Shilo, K., Kirste, S., Stegmaier, P., McNulty, M., Karachaliou, N., Rosell, R., Bepler, G., Carbone, D.P., Chakravarti, A., 2016. SMARCA4/BRG1 Is a Novel Prognostic Biomarker Predictive of Cisplatin-Based Chemotherapy Outcomes in Resected Non–Small Cell Lung Cancer. Clinical Cancer Research 22, 2396–2404. https://doi.org/10.1158/1078-0432.CCR-15-1468

Bender, E., 2014. Epidemiology: The dominant malignancy. Nature 513, S2-3. https://doi.org/10.1038/513S2a

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R.,

Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara E Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, Andrew C., Pike, Alger C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, John, Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, Jane, Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59. https://doi.org/10.1038/nature07517

Bergethon, K., Shaw, A.T., Ou, S.-H.I., Katayama, R., Lovly, C.M., McDonald, N.T., Massion, P.P., Siwak-Tapp, C., Gonzalez, A., Fang, R., Mark, E.J., Batten, J.M., Chen, H., Wilner, K.D., Kwak, E.L., Clark, J.W., Carbone, D.P., Ji, H., Engelman, J.A., Mino-Kenudson, M., Pao, W., Iafrate, A.J., 2012. ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol 30, 863–870. https://doi.org/10.1200/JCO.2011.35.6345

Bhatlekar, S., Fields, J.Z., Boman, B.M., 2014. HOX genes and their role in the development of human cancers. J Mol Med (Berl) 92, 811–823. https://doi.org/10.1007/s00109-014-1181-y

Biswas, S., Rao, C.M., 2018. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. Eur J Pharmacol 837, 8–24. https://doi.org/10.1016/j.ejphar.2018.08.021

Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A.V., Schultz, D.C., Liu, Q., Shih, I.-M., Conejo-Garcia, J.R., Speicher, D.W., Zhang, R., 2015. Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. Nat Med 21, 231–238. https://doi.org/10.1038/nm.3799

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., Barillot, E., 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28, 423–425. https://doi.org/10.1093/bioinformatics/btr670

Boyle, A.P., Guinney, J., Crawford, G.E., Furey, T.S., 2008. F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics 24, 2537–2538. https://doi.org/10.1093/bioinformatics/btn480

Bozek, M., Gompel, N., 2020. Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity: Considering Quantitative Accessibility Changes between Different Regulatory States of an Enhancer Deconvolutes the Complex Relationship between Accessibility and Activity. Bioessays 42, e1900188. https://doi.org/10.1002/bies.201900188

Buenrostro, J., Wu, B., Chang, H., Greenleaf, W., 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Burrows, A.E., Smogorzewska, A., Elledge, S.J., 2010. Polybromo-associated BRG1-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. Proc Natl Acad Sci U S A 107, 14280–14285. https://doi.org/10.1073/pnas.1009559107

Byers, L.A., Diao, L., Wang, J., Saintigny, P., Girard, L., Peyton, M., Shen, L., Fan, Y., Giri, U., Tumula, P.K., Nilsson, M.B., Gudikote, J., Tran, H., Cardnell, R.J.G., Bearss, D.J., Warner, S.L., Foulks, J.M., Kanner, S.B., Gandhi, V., Krett, N., Rosen, S.T., Kim, E.S., Herbst, R.S., Blumenschein, G.R., Lee, J.J., Lippman, S.M., Ang, K.K., Mills, G.B., Hong, W.K., Weinstein, J.N., Wistuba, I.I., Coombes, K.R., Minna, J.D., Heymach, J.V., 2013. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and

PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin Cancer Res 19, 279–290. https://doi.org/10.1158/1078-0432.CCR-12-1558

Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C., Chan, H.M., 2020. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends in Genetics 36, 936–950. https://doi.org/10.1016/j.tig.2020.07.011

Chaft, J.E., Rimner, A., Weder, W., Azzoli, C.G., Kris, M.G., Cascone, T., 2021. Evolution of systemic therapy for stages I-III non-metastatic non-small-cell lung cancer. Nat Rev Clin Oncol 18, 547–557. https://doi.org/10.1038/s41571-021-00501-4

Chandler, R.L., Damrauer, J.S., Raab, J.R., Schisler, J.C., Wilkerson, M.D., Didion, J.P., Starmer, J., Serber, D., Yee, D., Xiong, J., Darr, D.B., Pardo-Manuel de Villena, F., Kim, W.Y., Magnuson, T., 2015. Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. Nat Commun 6, 6118. https://doi.org/10.1038/ncomms7118

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., Shi, X., Fulton, R.S., Ley, T.J., Wilson, R.K., Ding, L., Mardis, E.R., 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6, 677–681. https://doi.org/10.1038/nmeth.1363

Chen, Y., Zhang, H., Xu, Z., Tang, H., Geng, A., Cai, B., Su, T., Shi, J., Jiang, C., Tian, X., Seluanov, A., Huang, J., Wan, X., Jiang, Y., Gorbunova, V., Mao, Z., 2019. A PARP1-BRG1-SIRT1 axis promotes HR repair by reducing nucleosome density at DNA damage sites. Nucleic Acids Res 47, 8563–8580. https://doi.org/10.1093/nar/gkz592

Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., Wong, K.-K., 2014. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer 14, 535–546. https://doi.org/10.1038/nrc3775

Christoforides, A., Carpten, J.D., Weiss, G.J., Demeure, M.J., Von Hoff, D.D., Craig, D.W., 2013. Identification of somatic mutations in cancer through Bayesian-based analysis of sequenced genome pairs. BMC Genomics 14, 302. https://doi.org/10.1186/1471-2164-14-302

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., Getz, G., 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213–219. https://doi.org/10.1038/nbt.2514

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., Bayley, H., 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nature Nanotech 4, 265–270. https://doi.org/10.1038/nnano.2009.12

Comprehensive genomic characterization of squamous cell lung cancers, 2012. . Nature 489, 519–525. https://doi.org/10.1038/nature11404

de Castro, R.O., Previato, L., Goitea, V., Felberg, A., Guiraldelli, M.F., Filiberti, A., Pezza, R.J., 2017. The chromatin-remodeling subunit Baf200 promotes homology-directed DNA repair and regulates distinct chromatin-remodeling complexes. J Biol Chem 292, 8459–8471. https://doi.org/10.1074/jbc.M117.778183

Deng, Y., He, R., Zhang, R., Gan, B., Zhang, Y., Chen, G., Hu, X., 2018. The expression of HOXA13 in lung adenocarcinoma and its clinical significance: A study based on The Cancer Genome Atlas, Oncomine and reverse transcription-quantitative polymerase chain reaction. Oncol Lett 15, 8556–8572. https://doi.org/10.3892/ol.2018.8381

Detterbeck, F.C., Boffa, D.J., Kim, A.W., Tanoue, L.T., 2017. The Eighth Edition Lung Cancer Stage Classification. Chest 151, 193–203. https://doi.org/10.1016/j.chest.2016.10.010

Di Giammartino, D.C., Kloetgen, A., Polyzos, A., Liu, Y., Kim, D., Murphy, D., Abuhashem, A., Cavaliere, P., Aronson, B., Shah, V., Dephoure, N., Stadtfeld, M., Tsirigos, A., Apostolou, E., 2019. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. Nat Cell Biol 21, 1179–1190. https://doi.org/10.1038/s41556-019-0390-6

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. https://doi.org/10.1038/nature11082

Doan, D.N., Veal, T.M., Yan, Z., Wang, W., Jones, S.N., Imbalzano, A.N., 2004. Loss of the INI1 tumor suppressor does not impair the expression of multiple BRG1-dependent genes or the assembly of SWI/SNF enzymes. Oncogene 23, 3462–3473. https://doi.org/10.1038/sj.onc.1207472

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, N., Ma, X., Shen, J., Zheng, Y., Li, G., Zheng, S., Huang, X., 2022. Identification and validation of critical genes with prognostic value in gastric cancer. Front Cell Dev Biol 10, 1072062. https://doi.org/10.3389/fcell.2022.1072062

Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E.M., Thorne, N.P., Bäckdahl, L., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Marioni, J.C., Birney, E., Hubbard, T.J.P., Durbin, R., Tavaré, S., Beck, S., 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26, 779–785. https://doi.org/10.1038/nbt1414

Drost, J., Mantovani, F., Tocco, F., Elkon, R., Comel, A., Holstege, H., Kerkhoven, R., Jonkers, J., Voorhoeve, P.M., Agami, R., Del Sal, G., 2010. BRD7 is a candidate tumour suppressor gene required for p53 function. Nat Cell Biol 12, 380–389. https://doi.org/10.1038/ncb2038

Duma, N., Santana-Davila, R., Molina, J.R., 2019. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc 94, 1623–1640. https://doi.org/10.1016/j.mayocp.2019.01.013

Dykhuizen, E.C., Hargreaves, D.C., Miller, E.L., Cui, K., Korshunov, A., Kool, M., Pfister, S., Cho, Y.-J., Zhao, K., Crabtree, G.R., 2013. BAF complexes facilitate decatenation of DNA by topoisomerase IIα. Nature 497, 624–627. https://doi.org/10.1038/nature12146

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48. https://doi.org/10.1186/1471-2105-10-48

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133–138. https://doi.org/10.1126/science.1162986

Engl, W., Viasnoff, V., Thiery, J.P., 2015. Epithelial Mesenchymal Transition Influence on CTL Activity, in: Bonavida, B., Chouaib, S. (Eds.), Resistance of Cancer Cells to CTL-Mediated Immunotherapy, Resistance to Targeted Anti-Cancer Therapeutics. Springer International Publishing, Cham, pp. 267–284. https://doi.org/10.1007/978-3-319-17807-3_12

Ernst, J., Kellis, M., 2017. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc 12, 2478–2492. https://doi.org/10.1038/nprot.2017.124

Fang, D., Han, J. (Eds.), 2021. Histone Mutations and Cancer, Advances in Experimental Medicine and Biology. Springer, Singapore. https://doi.org/10.1007/978-981-15-8104-5

Farnaby, W., Koegl, M., Roy, M.J., Whitworth, C., Diers, E., Trainor, N., Zollman, D., Steurer, S., Karolyi-Oezguer, J., Riedmueller, C., Gmaschitz, T., Wachter, J., Dank, C., Galant, M., Sharps, B., Rumpel, K., Traxler, E., Gerstberger, T., Schnitzer, R., Petermann, O., Greb, P., Weinstabl, H., Bader, G., Zoephel, A., Weiss-Puxbaum, A., Ehrenhöfer-Wölfer, K., Wöhrle, S., Boehmelt, G., Rinnenthal, J., Arnhof, H., Wiechens, N., Wu, M.-Y., Owen-Hughes, T., Ettmayer, P., Pearson, M., McConnell, D.B., Ciulli, A., 2019. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. Nat Chem Biol 15, 672–680. https://doi.org/10.1038/s41589-019-0294-6

Fejes, A.P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., Jones, S.J.M., 2008. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics 24, 1729–1730. https://doi.org/10.1093/bioinformatics/btn305

Feng, H., Lane, K.A., Roumeliotis, T.I., Jeggo, P.A., Somaiah, N., Choudhary, J.S., Downs, J.A., 2022. PBAF loss leads to DNA damage-induced inflammatory signaling through defective G2/M checkpoint maintenance. Genes Dev. https://doi.org/10.1101/gad.349249.121

Fernando, T.M., Piskol, R., Bainer, R., Sokol, E.S., Trabucco, S.E., Zhang, Q., Trinh, H., Maund, S., Kschonsak, M., Chaudhuri, S., Modrusan, Z., Januario, T., Yauch, R.L., 2020. Functional characterization of SMARCA4 variants identified by targeted exome-sequencing of 131,668 cancer patients. Nat Commun 11, 5551. https://doi.org/10.1038/s41467-020-19402-8

Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T.C., Choi, H., El Rayes, T., Ryu, S., Troeger, J., Schwabe, R.F., Vahdat, L.T., Altorki, N.K., Mittal, V., Gao, D., 2015. Epithelial-to-mesenchymal transition is not required

for lung metastasis but contributes to chemoresistance. Nature 527, 472–476. https://doi.org/10.1038/nature15748

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., Cohen, D., 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017, bax028. https://doi.org/10.1093/database/bax028

Gan, L., Meng, J., Xu, M., Liu, M., Qi, Y., Tan, C., Wang, Y., Zhang, P., Weng, W., Sheng, W., Huang, M., Wang, Z., 2018. Extracellular matrix protein 1 promotes cell metastasis and glucose metabolism by inducing integrin β4/FAK/SOX2/HIF-1α signaling pathway in gastric cancer. Oncogene 37, 744–755. https://doi.org/10.1038/onc.2017.363

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., Martinez, P., Phillimore, B., Begum, S., Rabinowitz, A., Spencer-Dene, B., Gulati, S., Bates, P.A., Stamp, G., Pickering, L., Gore, M., Nicol, D.L., Hazell, S., Futreal, P.A., Stewart, A., Swanton, C., 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet 46, 225–233. https://doi.org/10.1038/ng.2891

Gerstenberger, B.S., Trzupek, J.D., Tallant, C., Fedorov, O., Filippakopoulos, P., Brennan, P.E., Fedele, V., Martin, S., Picaud, S., Rogers, C., Parikh, M., Taylor, A., Samas, B., O'Mahony, A., Berg, E., Pallares, G., Torrey, A.D., Treiber, D.K., Samardjiev, I.J., Nasipak, B.T., Padilla-Benavides, T., Wu, Q., Imbalzano, A.N., Nickerson, J.A., Bunnage, M.E., Müller, S., Knapp, S., Owen, D.R., 2016. Identification of a Chemical Probe for Family VIII Bromodomains through Optimization of a Fragment Hit. J Med Chem 59, 4800–4811. https://doi.org/10.1021/acs.jmedchem.6b00012

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korbel, J.O., Furlong, E.E.M., 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet 51, 1272–1282. https://doi.org/10.1038/s41588-019-0462-3

Giacobbe, A., Compagnone, M., Bongiorno-Borbone, L., Antonov, A., Markert, E.K., Zhou, J.H., Annicchiarico-Petruzzelli, M., Melino, G., Peschiaroli, A., 2016. p63 controls cell migration and invasion by transcriptional regulation of MTSS1. Oncogene 35, 1602–1608. https://doi.org/10.1038/onc.2015.230

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17, 877–885. https://doi.org/10.1101/gr.5533506

Gonzalez-Perez, A., Jene-Sanz, A., Lopez-Bigas, N., 2013. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. Genome Biol 14, r106. https://doi.org/10.1186/gb-2013-14-9-r106

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17, 333–351. https://doi.org/10.1038/nrg.2016.49

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., Chen, K., Walker, J., McDonald, S., Bose, R., Ornitz, D., Xiong, D., You, M., Dooling, D.J., Watson, M., Mardis, E.R., Wilson, R.K., 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 150, 1121–1134. https://doi.org/10.1016/j.cell.2012.08.024

Greenberg, M.V.C., 2021. Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Frontiers in Cell and Developmental Biology 8.

Guo, B., Friedland, S.C., Alexander, W., Myers, J.A., Wang, W., O'Dell, M.R., Getman, M., Whitney-Miller, C.L., Agostini-Vulaj, D., Huber, A.R., Mello, S.S., Vertino, P.M., Land, H.K., Steiner, L.A., Hezel, A.F., 2022. Arid1a mutation suppresses TGF-β signaling and induces cholangiocarcinoma. Cell Rep 40, 111253. https://doi.org/10.1016/j.celrep.2022.111253

Gupta, M., Concepcion, C.P., Fahey, C.G., Keshishian, H., Bhutkar, A., Brainson, C.F., Sanchez-Rivera, F.J., Pessina, P., Kim, J.Y., Simoneau, A., Paschini, M., Beytagh, M.C., Stanclift, C.R., Schenone, M., Mani, D.R., Li, C., Oh, A., Li, F., Hu, H., Karatza, A., Bronson, R.T., Shaw, A.T., Hata, A.N., Wong, K.-K., Zou, L., Carr, S.A., Jacks, T., Kim, C.F., 2020. BRG1 Loss Predisposes Lung Cancers to Replicative Stress and ATR Dependency. Cancer Res 80, 3841–3854. https://doi.org/10.1158/0008-5472.CAN-20-1744

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of Cancer: The Next Generation. Cell 144, 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements

Required for Macrophage and B Cell Identities. Molecular Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Helming, Katherine C., Wang, X., Roberts, C.W.M., 2014. Vulnerabilities of mutant SWI/SNF complexes in cancer. Cancer Cell 26, 309–317. https://doi.org/10.1016/j.ccr.2014.07.018

Helming, Katherine C, Wang, X., Wilson, B.G., Vazquez, F., Haswell, J.R., Manchester, H.E., Kim, Y., Kryukov, G.V., Ghandi, M., Aguirre, A.J., Jagani, Z., Wang, Z., Garraway, L.A., Hahn, W.C., Roberts, C.W.M., 2014. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat Med 20, 251–254. https://doi.org/10.1038/nm.3480

Herpel, E., Rieker, R.J., Dienemann, H., Muley, T., Meister, M., Hartmann, A., Warth, A., Agaimy, A., 2017. SMARCA4 and SMARCA2 deficiency in non–small cell lung cancer: immunohistochemical survey of 316 consecutive specimens. Annals of Diagnostic Pathology 26, 47–51. https://doi.org/10.1016/j.anndiagpath.2016.10.006

Hirota, T., Lipp, J.J., Toh, B.-H., Peters, J.-M., 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature 438, 1176–1180. https://doi.org/10.1038/nature04254

Hodis, E., Watson, I.R., Kryukov, G.V., Arold, S.T., Imielinski, M., Theurillat, J.-P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A.H., Lawrence, M.S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R.C., Winckler, W., Ardlie, K., Wagle, N., Wargo, J., Chong, K., Morton, D.L., Stemke-Hale, K., Chen, G., Noble, M., Meyerson, M., Ladbury, J.E., Davies, M.A., Gershenwald, J.E., Wagner, S.N., Hoon, D.S.B., Schadendorf, D., Lander, E.S., Gabriel, S.B., Getz, G., Garraway, L.A., Chin, L., 2012. A landscape of driver mutations in melanoma. Cell 150, 251–263. https://doi.org/10.1016/j.cell.2012.06.024

Hoffman, G.R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., Bagdasarian, L., Huber, J., Lindeman, A., Chen, D., Romero, R., Ramadan, N., Phadke, T., Haas, K., Jaskelioff, M., Wilson, B.G., Meyer, M.J., Saenz-Vash, V., Zhai, H., Myer, V.E., Porter, J.A., Keen, N., McLaughlin, M.E., Mickanin, C., Roberts, C.W.M., Stegmeier, F., Jagani, Z., 2014. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc. Natl. Acad. Sci. U.S.A. 111, 3128–3133. https://doi.org/10.1073/pnas.1316793111

Hohmann, A.F., Martin, L.J., Minder, J.L., Roe, J.-S., Shi, J., Steurer, S., Bader, G., McConnell, D., Pearson, M., Gerstberger, T., Gottschamel, T., Thompson, D., Suzuki, Y., Koegl, M., Vakoc, C.R., 2016. Sensitivity and

engineered resistance of myeloid leukemia cells to BRD9 inhibition. Nat Chem Biol 12, 672–679. https://doi.org/10.1038/nchembio.2115

Inamura, K., 2018. Update on Immunohistochemistry for the Diagnosis of Lung Cancer. Cancers (Basel) 10, 72. https://doi.org/10.3390/cancers10030072

Isakoff, M.S., Sansam, C.G., Tamayo, P., Subramanian, A., Evans, J.A., Fillmore, C.M., Wang, X., Biegel, J.A., Pomeroy, S.L., Mesirov, J.P., Roberts, C.W.M., 2005. Inactivation of the Snf5 tumor suppressor stimulates cell cycle progression and cooperates with p53 loss in oncogenic transformation. Proc Natl Acad Sci U S A 102, 17745–17750. https://doi.org/10.1073/pnas.0509014102

Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., Salm, M., Horswell, S., Escudero, M., Matthews, N., Rowan, A., Chambers, T., Moore, D.A., Turajlic, S., Xu, H., Lee, S.-M., Forster, M.D., Ahmad, T., Hiley, C.T., Abbosh, C., Falzon, M., Borg, E., Marafioti, T., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S.M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Shah, R., Joseph, L., Quinn, A.M., Crosbie, P.A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D.A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J.F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Dentro, S., Taniere, P., O'Sullivan, B., Lowe, H.L., Hartley, J.A., Iles, N., Bell, H., Ngai, Y., Shaw, J.A., Herrero, J., Szallasi, Z., Schwarz, R.F., Stewart, A., Quezada, S.A., Le Quesne, J., Van Loo, P., Dive, C., Hackshaw, A., Swanton, C., TRACERx Consortium, 2017. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med 376, 2109–2121. https://doi.org/10.1056/NEJMoa1616288

Januario, T., Ye, X., Bainer, R., Alicke, B., Smith, T., Haley, B., Modrusan, Z., Gould, S., Yauch, R.L., 2017. PRC2-mediated repression of SMARCA2 predicts EZH2 inhibitor activity in SWI/SNF mutant tumors. Proceedings of the National Academy of Sciences 114, 12249–12254. https://doi.org/10.1073/pnas.1703966114

Jdeed, S., Lengyel, M., Uray, I.P., 2022. Redistribution of the SWI/SNF Complex Dictates Coordinated Transcriptional Control over Epithelial–Mesenchymal Transition of Normal Breast Cells through TGF-β Signaling. Cells 11, 2633. https://doi.org/10.3390/cells11172633

Jiang, H., Cao, H.-J., Ma, N., Bao, W.-D., Wang, J.-J., Chen, T.-W., Zhang, E.-B., Yuan, Y.-M., Ni, Q.-Z., Zhang, F.-K., Ding, X.-F., Zheng, Q.-W., Wang, Y.-K., Zhu, M., Wang, X., Feng, J., Zhang, X.-L., Cheng, S.-Q., Ma, D.-J.,

Qiu, L., Li, J.-J., Xie, D., 2020. Chromatin remodeling factor ARID2 suppresses hepatocellular carcinoma metastasis via DNMT1-Snail axis. Proc Natl Acad Sci U S A 117, 4770–4780. https://doi.org/10.1073/pnas.1914937117

Jiang, W., Yuan, Q., Jiang, Y., Huang, L., Chen, C., Hu, G., Wan, R., Wang, X., Yang, L., 2018. Identification of Sox6 as a regulator of pancreatic cancer development. J Cell Mol Med 22, 1864–1872. https://doi.org/10.1111/jcmm.13470

Jiang, Y., Oldridge, D.A., Diskin, S.J., Zhang, N.R., 2015. CODEX: a normalization and copy number variation detection method for whole exome sequencing. Nucleic Acids Res 43, e39. https://doi.org/10.1093/nar/gku1363

Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B., 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science 316, 1497–1502. https://doi.org/10.1126/science.1141319

Johnson, R.M., Qu, X., Lin, C.-F., Huw, L.-Y., Venkatanarayan, A., Sokol, E., Ou, F.-S., Ihuegbu, N., Zill, O.A., Kabbarah, O., Wang, L., Bourgon, R., de Sousa E Melo, F., Bolen, C., Daemen, A., Venook, A.P., Innocenti, F., Lenz, H.-J., Bais, C., 2022. ARID1A mutations confer intrinsic and acquired resistance to cetuximab treatment in colorectal cancer. Nat Commun 13, 5478. https://doi.org/10.1038/s41467-022-33172-5

Jones, C.A., Tansey, W.P., Weissmiller, A.M., 2022. Emerging Themes in Mechanisms of Tumorigenesis by SWI/SNF Subunit Mutation. Epigenet Insights 15, 25168657221115656. https://doi.org/10.1177/25168657221115656

Jones, J., Chen, Y., Tiwari, M., Li, J., Ling, J., Sen, G.L., 2020. KLF3 Mediates Epidermal Differentiation through the Epigenomic Writer CBP. iScience 23, 101320. https://doi.org/10.1016/j.isci.2020.101320

Jones, S., Wang, T.-L., Shih, I.-M., Mao, T.-L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L.A., Vogelstein, B., Kinzler, K.W., Velculescu, V.E., Papadopoulos, N., 2010. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 330, 228–231. https://doi.org/10.1126/science.1196333

Kadam, S., Emerson, B.M., 2003. Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. Mol Cell 11, 377–389. https://doi.org/10.1016/s1097-2765(03)00034-0

Kadoch, C., Copeland, R.A., Keilhack, H., 2016. PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. Biochemistry 55, 1600–1614. https://doi.org/10.1021/acs.biochem.5b01191

Kadoch, C., Crabtree, G.R., 2015. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv 1, e1500447. https://doi.org/10.1126/sciadv.1500447

Kadoch, C., Crabtree, G.R., 2013. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. Cell 153, 71–85. https://doi.org/10.1016/j.cell.2013.02.036

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., Crabtree, G.R., 2013. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet 45, 592–601. https://doi.org/10.1038/ng.2628

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., Crabtree, G.R., 2017. Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat Genet 49, 213–222. https://doi.org/10.1038/ng.3734

Kahali, B., Yu, J., Marquez, S.B., Thompson, K.W., Liang, S.Y., Lu, L., Reisman, D., 2014. The silencing of the SWI/SNF subunit and anticancer gene BRM in Rhabdoid tumors. Oncotarget 5, 3316–3332. https://doi.org/10.18632/oncotarget.1945

Kahm, Y.-J., Kim, R.-K., Jung, U., Kim, I.-G., 2021. Epithelial membrane protein 3 regulates lung cancer stem cells via the TGF-β signaling pathway. Int J Oncol 59, 80. https://doi.org/10.3892/ijo.2021.5261

Kantidze, O.L., Gurova, K.V., Studitsky, V.M., Razin, S.V., 2020. The 3D Genome as a Target for Anticancer Therapy. Trends in Molecular Medicine 26, 141–149. https://doi.org/10.1016/j.molmed.2019.09.011

Karachaliou, N., Paulina Bracht, J.W., Rosell, R., 2018. ARID1A Gene Driver Mutations in Lung Adenocarcinomas. Journal of Thoracic Oncology 13, e255–e257. https://doi.org/10.1016/j.jtho.2018.07.099

Karachaliou, N., Pilotto, S., Lazzari, C., Bria, E., Marinis, F. de, Rosell, R., 2016. Cellular and molecular biology of small cell lung cancer: an overview. Translational Lung Cancer Research 5. https://doi.org/10.3978/j.issn.2218-6751.2016.01.02

Karnezis, A.N., Wang, Y., Ramos, P., Hendricks, W.P., Oliva, E., D'Angelo, E., Prat, J., Nucci, M.R., Nielsen, T.O., Chow, C., Leung, S., Kommoss, F., Kommoss, S., Silva, A., Ronnett, B.M., Rabban, J.T., Bowtell, D.D., Weissman, B.E., Trent, J.M., Gilks, C.B., Huntsman, D.G., 2016. Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. J Pathol 238, 389–400. https://doi.org/10.1002/path.4633

Kayser, G., Csanadi, A., Kakanou, S., Prasse, A., Kassem, A., Stickeler, E., Passlick, B., Zur Hausen, A., 2015. Downregulation of MTSS1 expression is an independent prognosticator in squamous cell carcinoma of the lung. Br J Cancer 112, 866–873. https://doi.org/10.1038/bjc.2015.2

Kchouk, M., Gibrat, J.-F., Elloumi, M., n.d. Generations of Sequencing Technologies: From First to Next Generation. Biology and Medicine 9, 1–8. https://doi.org/10.4172/0974-8369.1000395

Kelso, T.W.R., Porter, D.K., Amaral, M.L., Shokhirev, M.N., Benner, C., Hargreaves, D.C., 2017. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. eLife 6, e30506. https://doi.org/10.7554/eLife.30506

Kempfer, R., Pombo, A., 2020. Methods for mapping 3D chromosome architecture. Nat Rev Genet 21, 207–226. https://doi.org/10.1038/s41576-019-0195-2

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kimura, S., Yokoyama, S., Pilon, A.L., Kurotani, R., 2022. Emerging role of an immunomodulatory protein secretoglobin 3A2 in human diseases. Pharmacol Ther 236, 108112. https://doi.org/10.1016/j.pharmthera.2022.108112

Kircher, M., Kelso, J., 2010. High-throughput DNA sequencing--concepts and limitations. Bioessays 32, 524–536. https://doi.org/10.1002/bies.200900181

Knutson, S.K., Warholic, N.M., Wigle, T.J., Klaus, C.R., Allain, C.J., Raimondi, A., Porter Scott, M., Chesworth, R., Moyer, M.P., Copeland, R.A., Richon, V.M., Pollock, R.M., Kuntz, K.W., Keilhack, H., 2013. Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. Proceedings of the National Academy of Sciences 110, 7922–7927. https://doi.org/10.1073/pnas.1303800110

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22, 568–576. https://doi.org/10.1101/gr.129684.111

L, H., Jl, R., J, W., Bt, S., L, C., A, K., J, L., Ai, N., J, R., Gr, C., 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proceedings of the National Academy of Sciences of the United States of America 106. https://doi.org/10.1073/pnas.0812889106

L, T., E, N., C, C., 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Progress in biophysics and molecular biology 102. https://doi.org/10.1016/j.pbiomolbio.2010.05.001

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lawrence, M., Daujat, S., Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet 32, 42–56. https://doi.org/10.1016/j.tig.2015.10.007

Lee, H.-S., Park, J.-H., Kim, S.-J., Kwon, S.-J., Kwon, J., 2010. A cooperative activation loop among SWI/SNF, γ-H2AX and H3 acetylation for DNA double-strand break repair. EMBO J 29, 1434–1445. https://doi.org/10.1038/emboj.2010.27

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., Crabtree, G.R., 2007. An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron 55, 201–215. https://doi.org/10.1016/j.neuron.2007.06.019

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18, 1851–1858. https://doi.org/10.1101/gr.078212.108

Li, L., Gu, X., Yue, J., Zhao, Q., Lv, D., Chen, H., Xu, L., 2017. Acquisition of EGFR TKI resistance and EMT phenotype is linked with activation of IGF1R/NF-κB pathway in EGFR-mutant NSCLC. Oncotarget 8, 92240–92253. https://doi.org/10.18632/oncotarget.21170

Li, M., Zhao, H., Zhang, X., Wood, L.D., Anders, R.A., Choti, M.A., Pawlik, T.M., Daniel, H.D., Kannangai, R., Offerhaus, G.J.A., Velculescu, V.E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R.H., Papadopoulos, N., Cai, J., Torbenson, M.S., Kinzler, K.W., 2011. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. Nat Genet 43, 828–829. https://doi.org/10.1038/ng.903

Li, R., Li, Y., Kristiansen, K., Wang, J., 2008. SOAP: short oligonucleotide alignment program. Bioinformatics 24, 713–714. https://doi.org/10.1093/bioinformatics/btn025

Li, Y., Tollefsbol, T.O., 2011. DNA methylation detection: bisulfite genomic sequencing analysis. Methods Mol Biol 791, 11–21. https://doi.org/10.1007/978-1-61779-316-5_2

Liao, S., Davoli, T., Leng, Y., Li, M.Z., Xu, Q., Elledge, S.J., 2017. A genetic interaction analysis identifies cancer drivers that modify EGFR dependency. Genes Dev 31, 184–196. https://doi.org/10.1101/gad.291948.116

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293. https://doi.org/10.1126/science.1181369

Lissanu Deribe, Y., Sun, Y., Terranova, C., Khan, F., Martinez-Ledesma, J., Gay, J., Gao, G., Mullinax, R.A., Khor, T., Feng, N., Lin, Y.-H., Wu, C.-C., Reyes, C., Peng, Q., Robinson, F., Inoue, A., Kochat, V., Liu, C.-G., Asara, J.M., Moran, C., Muller, F., Wang, J., Fang, B., Papadimitrakopoulou, V., Wistuba, I.I., Rai, K., Marszalek, J., Futreal, P.A., 2018. Mutations in the SWI/SNF complex induce a targetable dependence on oxidative phosphorylation in lung cancer. Nat Med 24, 1047–1057. https://doi.org/10.1038/s41591-018-0019-5

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262

Lord, C.J., Ashworth, A., 2017. PARP inhibitors: Synthetic lethality in the clinic. Science 355, 1152–1158. https://doi.org/10.1126/science.aam7344

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lovly, C.M., Carbone, D.P., 2011. Lung cancer in 2010: One size does not fit all. Nat Rev Clin Oncol 8, 68–70. https://doi.org/10.1038/nrclinonc.2010.224

Lu, B., Shi, H., 2019. An In-Depth Look at Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT): Clinical Implications from Recent Molecular Findings. J Cancer 10, 223–237. https://doi.org/10.7150/jca.26978

Lu, C., Allis, C.D., 2017. SWI/SNF complex in cancer. Nat Genet 49, 178–179. https://doi.org/10.1038/ng.3779

Manceau, G., Letouzé, E., Guichard, C., Didelot, A., Cazes, A., Corté, H., Fabre, E., Pallier, K., Imbeaud, S., Le Pimpec-Barthes, F., Zucman-Rossi, J., Laurent-Puig, P., Blons, H., 2013. Recurrent inactivating mutations of

ARID2 in non-small cell lung carcinoma. Int. J. Cancer 132, 2217–2221. https://doi.org/10.1002/ijc.27900

Martínez, N., Almaraz, C., Vaqué, J.P., Varela, I., Derdak, S., Beltran, S., Mollejo, M., Campos-Martin, Y., Agueda, L., Rinaldi, A., Kwee, I., Gut, M., Blanc, J., Oscier, D., Strefford, J.C., Martinez-Lopez, J., Salar, A., Sole, F., Rodriguez-Peralto, J.L., Diez-Tascón, C., García, J.F., Fraga, M., Sebastián, E., Alvés, J., Menárguez, J., González-Carreró, J., Casado, L.F., Bayes, M., Bertoni, F., Gut, I., Piris, M.A., 2014. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. Leukemia 28, 1334–1340. https://doi.org/10.1038/leu.2013.365

Mashtalir, N., Dao, H.T., Sankar, A., Liu, H., Corin, A.J., Bagert, J.D., Ge, E.J., D'Avino, A.R., Filipovski, M., Michel, B.C., Dann, G.P., Muir, T.W., Kadoch, C., 2021. Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. Science 373, 306–315. https://doi.org/10.1126/science.abf8705

Mashtalir, N., D'Avino, A.R., Michel, B.C., Luo, J., Pan, J., Otto, J.E., Zullow, H.J., McKenzie, Z.M., Kubiak, R.L., St Pierre, R., Valencia, A.M., Poynter, S.J., Cassel, S.H., Ranish, J.A., Kadoch, C., 2018. Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell 175, 1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., Roberts, C.W.M., 2017. ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat Genet 49, 296–302. https://doi.org/10.1038/ng.3744

Matsubara, D., Kishaba, Y., Ishikawa, S., Sakatani, T., Oguni, S., Tamura, T., Hoshino, H., Sugiyama, Y., Endo, S., Murakami, Y., Aburatani, H., Fukayama, M., Niki, T., 2013. Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. Cancer Science 104, 266–273. https://doi.org/10.1111/cas.12065

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., Bejerano, G., 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495–501. https://doi.org/10.1038/nbt.1630

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., Wiegraebe, W., Singh, S., Becker, T., Caicedo, J.C., Carpenter, A.E., 2018. CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol 16, e2005970. https://doi.org/10.1371/journal.pbio.2005970

Medina, P.P., Romero, O.A., Kohno, T., Montuenga, L.M., Pio, R., Yokota, J., Sanchez-Cespedes, M., 2008. Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. Hum Mutat 29, 617–622. https://doi.org/10.1002/humu.20730

Metzker, M.L., 2010. Sequencing technologies — the next generation. Nat Rev Genet 11, 31–46. https://doi.org/10.1038/nrg2626

Meyer, C.A., Liu, X.S., 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet 15, 709–721. https://doi.org/10.1038/nrg3788

Miao, D., Margolis, C.A., Gao, W., Voss, M.H., Li, W., Martini, D.J., Norton, C., Bossé, D., Wankowicz, S.M., Cullen, D., Horak, C., Wind-Rotolo, M., Tracy, A., Giannakis, M., Hodi, F.S., Drake, C.G., Ball, M.W., Allaf, M.E., Snyder, A., Hellmann, M.D., Ho, T., Motzer, R.J., Signoretti, S., Kaelin, W.G., Choueiri, T.K., Van Allen, E.M., 2018. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science 359, 801–806. https://doi.org/10.1126/science.aan5951

Michel, B.C., D'Avino, A.R., Cassel, S.H., Mashtalir, N., McKenzie, Z.M., McBride, M.J., Valencia, A.M., Zhou, Q., Bocker, M., Soares, L.M.M., Pan, J., Remillard, D.I., Lareau, C.A., Zullow, H.J., Fortoul, N., Gray, N.S., Bradner, J.E., Chan, H.M., Kadoch, C., 2018. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. Nat Cell Biol 20, 1410–1420. https://doi.org/10.1038/s41556-018-0221-1

Mittal, P., Roberts, C.W.M., 2020. The SWI/SNF complex in cancer — biology, biomarkers and therapy. Nat Rev Clin Oncol 17, 435–448. https://doi.org/10.1038/s41571-020-0357-3

Monterde, B., Varela, I., 2022. Role of SWI/SNF chromatin remodeling genes in lung cancer development. Biochemical Society Transactions BST20211084. https://doi.org/10.1042/BST20211084

Moonen, J.-R., Chappell, J., Shi, M., Shinohara, T., Li, D., Mumbach, M.R., Zhang, F., Nair, R.V., Nasser, J., Mai, D.H., Taylor, S., Wang, L., Metzger, R.J., Chang, H.Y., Engreitz, J.M., Snyder, M.P., Rabinovitch, M., 2022. KLF4 recruits SWI/SNF to increase chromatin accessibility and reprogram the endothelial enhancer landscape under laminar shear stress. Nat Commun 13, 4941. https://doi.org/10.1038/s41467-022-32566-9

Moreno, T., Monterde, B., González-Silva, L., Betancor-Fernández, I., Revilla, C., Agraz-Doblas, A., Freire, J., Isidro, P., Quevedo, L., Blanco, R., Montes-Moreno, S., Cereceda, L., Astudillo, A., Casar, B., Crespo, P., Morales Torres, C., Scaffidi, P., Gómez-Román, J., Salido, E., Varela, I., 2021. ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer. Oncogene 40, 2923–2935. https://doi.org/10.1038/s41388-021-01748-y

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., Allis, C.D., 2019. The expanding landscape of "oncohistone" mutations in human cancers. Nature 567, 473–478. https://doi.org/10.1038/s41586-019-1038-1

Nagl Jr, N.G., Wang, X., Patsialou, A., Van Scoy, M., Moran, E., 2007. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. The EMBO Journal 26, 752–763. https://doi.org/10.1038/sj.emboj.7601541

Nagl, N.G., Jr., Zweitzig, D.R., Thimmapaya, B., Beck, G.R., Jr., Moran, E., 2006. The c-myc Gene Is a Direct Target of Mammalian SWI/SNF–Related Complexes during Differentiation-Associated Cell Cycle Arrest. Cancer Research 66, 1289–1293. https://doi.org/10.1158/0008-5472.CAN-05-3427

Nakanishi, S., Lee, J.S., Gardner, K.E., Gardner, J.M., Takahashi, Y., Chandrasekharan, M.B., Sun, Z.-W., Osley, M.A., Strahl, B.D., Jaspersen, S.L., Shilatifard, A., 2009. Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. J Cell Biol 186, 371–377. https://doi.org/10.1083/jcb.200906005

Nakayama, R.T., Pulice, J.L., Valencia, A.M., McBride, M.J., McKenzie, Z.M., Gillespie, M.A., Ku, W.L., Teng, M., Cui, K., Williams, R.T., Cassel, S.H., Qing, H., Widmer, C.J., Demetri, G.D., Irizarry, R.A., Zhao, K., Ranish, J.A., Kadoch, C., 2017. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. Nat Genet 49, 1613–1623. https://doi.org/10.1038/ng.3958

Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., Thiery, J.P., 2016. EMT: 2016. Cell 166, 21–45. https://doi.org/10.1016/j.cell.2016.06.028

Niimi, A., Chambers, A.L., Downs, J.A., Lehmann, A.R., 2012. A role for chromatin remodellers in replication of damaged DNA. Nucleic Acids Res 40, 7393–7403. https://doi.org/10.1093/nar/gks453

North, J.A., Javaid, S., Ferdinand, M.B., Chatterjee, N., Picking, J.W., Shoffner, M., Nakkula, R.J., Bartholomew, B., Ottesen, J.J., Fishel, R., Poirier, M.G., 2011. Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. Nucleic Acids Res 39, 6465–6474. https://doi.org/10.1093/nar/gkr304

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. Science 376, 44–53. https://doi.org/10.1126/science.abj6987

Oba, A., Shimada, S., Akiyama, Y., Nishikawaji, T., Mogushi, K., Ito, H., Matsumura, S., Aihara, A., Mitsunori, Y., Ban, D., Ochiai, T., Kudo, A., Asahara, H., Kaida, A., Miura, M., Tanabe, M., Tanaka, S., 2017. ARID2 modulates DNA damage response in human hepatocellular carcinoma cells. J Hepatol 66, 942–951. https://doi.org/10.1016/j.jhep.2016.12.026

Oike, T., Ogiwara, H., Tominaga, Y., Ito, K., Ando, O., Tsuta, K., Mizukami, T., Shimada, Y., Isomura, H., Komachi, M., Furuta, K., Watanabe, S.-I., Nakano, T., Yokota, J., Kohno, T., 2013. A Synthetic Lethality–Based Strategy to Treat Cancers Harboring a Genetic Deficiency in the Chromatin Remodeling Factor BRG1. Cancer Res 73, 5508–5518. https://doi.org/10.1158/0008-5472.CAN-12-4593

Oldfield, A.J., Henriques, T., Kumar, D., Burkholder, A.B., Cinghu, S., Paulet, D., Bennett, B.D., Yang, P., Scruggs, B.S., Lavender, C.A., Rivals, E., Adelman, K., Jothi, R., 2019. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. Nat Commun 10, 3072. https://doi.org/10.1038/s41467-019-10905-7

Orlando, K.A., Douglas, A.K., Abudu, A., Wang, Y., Tessier-Cloutier, B., Su, W., Peters, A., Sherman, L.S., Moore, R., Nguyen, V., Negri, G.L., Colborne, S., Morin, G.B., Kommoss, F., Lang, J.D., Hendricks, W.P., Raupach, E.A., Pirrotte, P., Huntsman, D.G., Trent, J.M., Parker, J.S., Raab, J.R., Weissman, B.E., 2020. Re-expression of SMARCA4/BRG1 in small cell carcinoma of ovary, hypercalcemic type (SCCOHT) promotes an epithelial-like gene signature through an AP-1-dependent mechanism. Elife 9, e59073. https://doi.org/10.7554/eLife.59073

Otto, J.E., Kadoch, C., 2017. A Two-Faced mSWI/SNF Subunit: Dual Roles for ARID1A in Tumor Suppression and Oncogenicity in the Liver. Cancer Cell 32, 542–543. https://doi.org/10.1016/j.ccell.2017.10.014

Ou, S.-H.I., 2013. Lung cancer in never-smokers. Does smoking history matter in the era of molecular diagnostics and targeted therapy? J Clin Pathol 66, 839–846. https://doi.org/10.1136/jclinpath-2012-201296

Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H.G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., van IJcken, W.F.J., Landeira, D., Rada-Iglesias, A., 2021. Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. Nat Genet 53, 1036–1049. https://doi.org/10.1038/s41588-021-00888-x

Paik, P.K., Johnson, M.L., D'Angelo, S.P., Sima, C.S., Ang, D., Dogan, S., Miller, V.A., Ladanyi, M., Kris, M.G., Riely, G.J., 2012. Driver mutations determine survival in smokers and never-smokers with stage IIIB/IV lung adenocarcinomas. Cancer 118, 5840–5847. https://doi.org/10.1002/cncr.27637

Pan, D., Kobayashi, A., Jiang, P., Ferrari de Andrade, L., Tay, R.E., Luoma, A.M., Tsoucas, D., Qiu, X., Lim, K., Rao, P., Long, H.W., Yuan, G.-C., Doench, J., Brown, M., Liu, X.S., Wucherpfennig, K.W., 2018. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science 359, 770–775. https://doi.org/10.1126/science.aao1710

Panier, S., Durocher, D., 2013. Push back to respond better: regulatory inhibition of the DNA double-strand break response. Nat Rev Mol Cell Biol 14, 661–672. https://doi.org/10.1038/nrm3659

Papadakis, A.I., Sun, C., Knijnenburg, T.A., Xue, Y., Grernrum, W., Hölzel, M., Nijkamp, W., Wessels, L.F., Beijersbergen, R.L., Bernards, R., Huang, S., 2015. SMARCE1 suppresses EGFR expression and controls responses to MET and ALK inhibitors in lung cancer. Cell Res 25, 445–458. https://doi.org/10.1038/cr.2015.16

Papillon, J.P.N., Nakajima, K., Adair, C.D., Hempel, J., Jouk, A.O., Karki, R.G., Mathieu, S., Möbitz, H., Ntaganda, R., Smith, T., Visser, M., Hill, S.E., Hurtado, F.K., Chenail, G., Bhang, H.-E.C., Bric, A., Xiang, K., Bushold, G., Gilbert, T., Vattay, A., Dooley, J., Costa, E.A., Park, I., Li, A., Farley, D., Lounkine, E., Yue, Q.K., Xie, X., Zhu, X., Kulathila, R., King, D., Hu, T., Vulic, K., Cantwell, J., Luu, C., Jagani, Z., 2018. Discovery of Orally Active Inhibitors of Brahma Homolog (BRM)/SMARCA2 ATPase Activity for the Treatment of Brahma Related Gene 1 (BRG1)/SMARCA4-Mutant Cancers. J. Med. Chem. 61, 10155–10172. https://doi.org/10.1021/acs.jmedchem.8b01318

Peinado, P., Andrades, A., Cuadros, M., Rodriguez, M.I., Coira, I.F., Garcia, D.J., Álvarez-Perez, J.C., Baliñas-Gavira, C., Arenas, A.M., Patiño-Mercau, J.R., Sanjuan-Hidalgo, J., Romero, O.A., Montuenga, L.M., Carretero, J., Sanchez-Cespedes, M., Medina, P.P., 2020. Comprehensive Analysis of SWI/SNF Inactivation in Lung Adenocarcinoma Cell Models. Cancers 12, 3712. https://doi.org/10.3390/cancers12123712

Peinado, P., Andrades, A., Cuadros, M., Rodriguez, M.I., Coira, I.F., Garcia, D.J., Benitez-Cantos, M.S., Cano, C., Zarzuela, E., Muñoz, J., Loidi, C., Saiz, M., Medina, P.P., 2022. Multi-omic alterations of the SWI/SNF complex define a clinical subgroup in lung adenocarcinoma. Clin Epigenetics 14, 42. https://doi.org/10.1186/s13148-022-01261-3

Pfeifer, G.P., Denissenko, M.F., Olivier, M., Tretyakova, N., Hecht, S.S., Hainaut, P., 2002. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 21, 7435–7451. https://doi.org/10.1038/sj.onc.1205803

Popper, H.H., 2016. Progression and metastasis of lung cancer. Cancer Metastasis Rev 35, 75–91. https://doi.org/10.1007/s10555-016-9618-0

Pott, S., Lieb, J.D., 2015. What are super-enhancers? Nat Genet 47, 8–12. https://doi.org/10.1038/ng.3167

Pugh, T.J., Weeraratne, S.D., Archer, T.C., Pomeranz Krummel, D.A., Auclair, D., Bochicchio, J., Carneiro, M.O., Carter, S.L., Cibulskis, K., Erlich, R.L., Greulich, H., Lawrence, M.S., Lennon, N.J., McKenna, A., Meldrim, J., Ramos, A.H., Ross, M.G., Russ, C., Shefler, E., Sivachenko, A., Sogoloff, B., Stojanov, P., Tamayo, P., Mesirov, J.P., Amani, V., Teider, N., Sengupta, S., Francois, J.P., Northcott, P.A., Taylor, M.D., Yu, F., Crabtree, G.R., Kautzman, A.G., Gabriel, S.B., Getz, G., Jäger, N., Jones, D.T.W., Lichter, P., Pfister, S.M., Roberts, T.M., Meyerson, M., Pomeroy, S.L., Cho, Y.-J., 2012. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 488, 106–110. https://doi.org/10.1038/nature11329

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Raab, J.R., Resnick, S., Magnuson, T., 2015. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. PLOS Genetics 11, e1005748. https://doi.org/10.1371/journal.pgen.1005748

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160–W165. https://doi.org/10.1093/nar/gkw257

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28, i333–i339. https://doi.org/10.1093/bioinformatics/bts378

Ray, A., Mir, S.N., Wani, G., Zhao, Q., Battu, A., Zhu, Q., Wang, Q.-E., Wani, A.A., 2009. Human SNF5/INI1, a Component of the Human SWI/SNF Chromatin Remodeling Complex, Promotes Nucleotide Excision Repair by Influencing ATM Recruitment and Downstream H2AX Phosphorylation. Mol Cell Biol 29, 6206–6219. https://doi.org/10.1128/MCB.00503-09

Reisman, D., Glaros, S., Thompson, E.A., 2009. The SWI/SNF complex and cancer. Oncogene 28, 1653–1668. https://doi.org/10.1038/onc.2009.4

Reisman, D.N., Sciarrotta, J., Wang, W., Funkhouser, W.K., Weissman, B.E., 2003. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. Cancer Res 63, 560–566.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47. https://doi.org/10.1093/nar/gkv007

Rivera, M.P., 2013. Lung cancer in women: differences in epidemiology, biology, histology, and treatment outcomes. Semin Respir Crit Care Med 34, 792–801. https://doi.org/10.1055/s-0033-1358550

Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A., Mesirov, J.P., 2017. Variant Review with the Integrative Genomics Viewer. Cancer Res 77, e31–e34. https://doi.org/10.1158/0008-5472.CAN-17-0337

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robson, M.I., Ringel, A.R., Mundlos, S., 2019. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. Mol Cell 74, 1110–1122. https://doi.org/10.1016/j.molcel.2019.05.032

Rodriguez-Canales, J., Parra-Cuentas, E., Wistuba, I.I., 2016. Diagnosis and Molecular Classification of Lung Cancer. Cancer Treat Res 170, 25–46. https://doi.org/10.1007/978-3-319-40389-2_2

Romero, O.A., Setien, F., John, S., Gimenez-Xavier, P., Gómez-López, G., Pisano, D., Condom, E., Villanueva, A., Hager, G.L., Sanchez-Cespedes, M., 2012. The tumour suppressor and chromatin-remodelling factor BRG1 antagonizes Myc activity and promotes cell differentiation in human cancer. EMBO Mol Med 4, 603–616. https://doi.org/10.1002/emmm.201200236

Romero, O.A., Torres-Diz, M., Pros, E., Savola, S., Gomez, A., Moran, S., Saez, C., Iwakawa, R., Villanueva, A., Montuenga, L.M., Kohno, T., Yokota, J., Sanchez-Cespedes, M., 2014. *MAX* Inactivation in Small Cell Lung Cancer Disrupts MYC–SWI/SNF Programs and Is Synthetic Lethal with BRG1. Cancer Discovery 4, 292–303. https://doi.org/10.1158/2159-8290.CD-13-0799

Romero, O.A., Vilarrubi, A., Alburquerque-Bejar, J.J., Gomez, A., Andrades, A., Trastulli, D., Pros, E., Setien, F., Verdura, S., Farré, L., Martín-Tejera, J.F., Llabata, P., Oaknin, A., Saigi, M., Piulats, J.M., Matias-Guiu, X., Medina, P.P., Vidal, A., Villanueva, A., Sanchez-Cespedes, M., 2021. SMARCA4 deficient tumours are vulnerable to KDM6A/UTX and

KDM6B/JMJD3 blockade. Nat Commun 12, 4319. https://doi.org/10.1038/s41467-021-24618-3

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., Nyrén, P., 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242, 84–89. https://doi.org/10.1006/abio.1996.0432

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–352. https://doi.org/10.1038/nature10242

Rother, M.B., van Attikum, H., 2017. DNA repair goes hip-hop: SMARCA and CHD chromatin remodellers join the break dance. Philos Trans R Soc Lond B Biol Sci 372, 20160285. https://doi.org/10.1098/rstb.2016.0285

Rudin, C.M., Brambilla, E., Faivre-Finn, C., Sage, J., 2021. Small-cell lung cancer. Nat Rev Dis Primers 7, 3. https://doi.org/10.1038/s41572-020-00235-0

Rugo, H.S., Olopade, O.I., DeMichele, A., Yau, C., van 't Veer, L.J., Buxton, M.B., Hogarth, M., Hylton, N.M., Paoloni, M., Perlmutter, J., Symmans, W.F., Yee, D., Chien, A.J., Wallace, A.M., Kaplan, H.G., Boughey, J.C., Haddad, T.C., Albain, K.S., Liu, M.C., Isaacs, C., Khan, Q.J., Lang, J.E., Viscusi, R.K., Pusztai, L., Moulder, S.L., Chui, S.Y., Kemmer, K.A., Elias, A.D., Edmiston, K.K., Euhus, D.M., Haley, B.B., Nanda, R., Northfelt, D.W., Tripathy, D., Wood, W.C., Ewing, C., Schwab, R., Lyandres, J., Davis, S.E., Hirst, G.L., Sanil, A., Berry, D.A., Esserman, L.J., I-SPY 2 Investigators, 2016. Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. N Engl J Med 375, 23–34. https://doi.org/10.1056/NEJMoa1513749

Sadek, M., Sheth, A., Zimmerman, G., Hays, E., Vélez-Cruz, R., 2022. The role of SWI/SNF chromatin remodelers in the repair of DNA double strand breaks and cancer therapy. Front Cell Dev Biol 10, 1071786. https://doi.org/10.3389/fcell.2022.1071786

Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., Cramer, P., Kivioja, T., Taipale, J., 2022. Sequence determinants of human gene regulatory elements. Nat Genet 54, 283–294. https://doi.org/10.1038/s41588-021-01009-4

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sathirapongsasuti, J.F., Lee, H., Horst, B.A.J., Brunner, G., Cochran, A.J., Binder, S., Quackenbush, J., Nelson, S.F., 2011. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics 27, 2648–2654. https://doi.org/10.1093/bioinformatics/btr462

Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., Cheetham, R.K., 2012. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817. https://doi.org/10.1093/bioinformatics/bts271

Schadt, E.E., Turner, S., Kasarskis, A., 2010. A window into third-generation sequencing. Hum Mol Genet 19, R227-240. https://doi.org/10.1093/hmg/ddq416

Schiaffino-Ortega, S., Balinas, C., Cuadros, M., Medina, P.P., 2014. SWI/SNF proteins as targets in cancer therapy. Journal of Hematology & Oncology 7, 81. https://doi.org/10.1186/s13045-014-0081-5

Schick, S., Rendeiro, A.F., Runggatscher, K., Ringler, A., Boidol, B., Hinkel, M., Májek, P., Vulliard, L., Penz, T., Parapatics, K., Schmidl, C., Menche, J., Boehmelt, G., Petronczki, M., Müller, A.C., Bock, C., Kubicek, S., 2019. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. Nat Genet 51, 1399–1410. https://doi.org/10.1038/s41588-019-0477-9

Schuettengruber, B., Bourbon, H.-M., Di Croce, L., Cavalli, G., 2017. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell 171, 34–57. https://doi.org/10.1016/j.cell.2017.08.002

Sen, M., Wang, X., Hamdan, F.H., Rapp, J., Eggert, J., Kosinsky, R.L., Wegwitz, F., Kutschat, A.P., Younesi, F.S., Gaedcke, J., Grade, M., Hessmann, E., Papantonis, A., Ströbel, P., Johnsen, S.A., 2019. ARID1A facilitates KRAS signaling-regulated enhancer activity in an AP1-dependent manner in colorectal cancer cells. Clin Epigenetics 11, 92. https://doi.org/10.1186/s13148-019-0690-5

Shain, A.H., Giacomini, C.P., Matsukuma, K., Karikari, C.A., Bashyam, M.D., Hidalgo, M., Maitra, A., Pollack, J.R., 2012. Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. Proc Natl Acad Sci U S A 109, E252-259. https://doi.org/10.1073/pnas.1114817109

Shain, A.H., Pollack, J.R., 2013. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. PLoS One 8, e55119. https://doi.org/10.1371/journal.pone.0055119

Shapiro, H., Goldenberg, K., Ratiner, K., Elinav, E., 2022. Smoking-induced microbial dysbiosis in health and disease. Clin Sci (Lond) 136, 1371–1387. https://doi.org/10.1042/CS20220175

Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E., Furey, T.S., 2013. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res 23, 777–788. https://doi.org/10.1101/gr.152140.112

Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., Kapoor, P., Ju, Z., Mo, Q., Shih, I.-M., Uray, I.P., Wu, X., Brown, P.H., Shen, X., Mills, G.B., Peng, G., 2015. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov 5, 752–767. https://doi.org/10.1158/2159-8290.CD-14-0849

Shlyueva, D., Stampfel, G., Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15, 272–286. https://doi.org/10.1038/nrg3682

Skene, P.J., Henikoff, J.G., Henikoff, S., 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc 13, 1006–1019. https://doi.org/10.1038/nprot.2018.015

Skoulidis, F., Heymach, J.V., 2019. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. Nat Rev Cancer 19, 495–509. https://doi.org/10.1038/s41568-019-0179-8

Skoulidis, F., Li, B.T., Dy, G.K., Price, T.J., Falchook, G.S., Wolf, J., Italiano, A., Schuler, M., Borghaei, H., Barlesi, F., Kato, T., Curioni-Fontecedro, A., Sacher, A., Spira, A., Ramalingam, S.S., Takahashi, T., Besse, B., Anderson, A., Ang, A., Tran, Q., Mather, O., Henary, H., Ngarmchamnanrith, G., Friberg, G., Velcheti, V., Govindan, R., 2021. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. New England Journal of Medicine 384, 2371–2381. https://doi.org/10.1056/NEJMoa2103695

Slatko, B.E., Gardner, A.F., Ausubel, F.M., 2018. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol 122, e59. https://doi.org/10.1002/cpmb.59

Somsuan, K., Peerapen, P., Boonmark, W., Plumworasawat, S., Samol, R., Sakulsak, N., Thongboonkerd, V., 2019. ARID1A knockdown triggers epithelial-mesenchymal transition and carcinogenesis features of renal cells: role in renal cell carcinoma. FASEB J 33, 12226–12239. https://doi.org/10.1096/fj.201802720RR

Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M.G., Ku, W.L., Kadoch, C., Zhao, K., Crabtree, G.R., 2017. Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat Genet 49, 282–288. https://doi.org/10.1038/ng.3735

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., Blobel, G.A., Vakoc, C.R., 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol 28, 2825–2839. https://doi.org/10.1128/MCB.02076-07

Strahl, B.D., Allis, C.D., 2000. The language of covalent histone modifications. Nature 403, 41–45. https://doi.org/10.1038/47412

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sun, C., Yin, J., Fang, Y., Chen, J., Jeong, K.J., Chen, X., Vellano, C.P., Ju, Z., Zhao, W., Zhang, D., Lu, Y., Meric-Bernstam, F., Yap, T.A., Hattersley, M., O'Connor, M.J., Chen, H., Fawell, S., Lin, S.-Y., Peng, G., Mills, G.B., 2018. BRD4 Inhibition Is Synthetic Lethal with PARP Inhibitors through the Induction of Homologous Recombination Deficiency. Cancer Cell 33, 401-416.e8. https://doi.org/10.1016/j.ccell.2018.01.019

Sun, X., Wang, S.C., Wei, Y., Luo, X., Jia, Y., Li, L., Gopal, P., Zhu, M., Nassour, I., Chuang, J.-C., Maples, T., Celen, C., Nguyen, L.H., Wu, L., Fu, S., Li, W., Hui, L., Tian, F., Ji, Y., Zhang, S., Sorouri, M., Hwang, T.H., Letzig, L., James, L., Wang, Z., Yopp, A.C., Singal, A.G., Zhu, H., 2017. Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. Cancer Cell 32, 574-589.e6. https://doi.org/10.1016/j.ccell.2017.10.007

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A.D., Nueda, M.J., Ferrer, A., Conesa, A., 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res 43, e140. https://doi.org/10.1093/nar/gkv711

Taylor, M.D., Bollt, O., Iyer, S.C., Robertson, G.P., 2018. Metastasis suppressor 1 (MTSS1) expression is associated with reduced in-vivo metastasis and enhanced patient survival in lung adenocarcinoma. Clin Exp Metastasis 35, 15–23. https://doi.org/10.1007/s10585-017-9869-3

Teer, J.K., Mullikin, J.C., 2010. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 19, R145-151. https://doi.org/10.1093/hmg/ddq333

The Cancer Genome Atlas Research Network, 2014. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550. https://doi.org/10.1038/nature13385

Theodoulou, N.H., Bamborough, P., Bannister, A.J., Becher, I., Bit, R.A., Che, K.H., Chung, C., Dittmann, A., Drewes, G., Drewry, D.H., Gordon, L., Grandi, P., Leveridge, M., Lindon, M., Michon, A.-M., Molnar, J., Robson, S.C., Tomkinson, N.C.O., Kouzarides, T., Prinjha, R.K., Humphreys, P.G., 2016. Discovery of I-BRD9, a Selective Cell Active Chemical Probe for Bromodomain Containing Protein 9 Inhibition. J Med Chem 59, 1425–1439. https://doi.org/10.1021/acs.jmedchem.5b00256

Thomas, A., Liu, S.V., Subramaniam, D.S., Giaccone, G., 2015. Refining the treatment of NSCLC according to histological and molecular subtypes. Nat Rev Clin Oncol 12, 511–526. https://doi.org/10.1038/nrclinonc.2015.90

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14, 178–192. https://doi.org/10.1093/bib/bbs017

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutyavin, T., Lajoie, B., Lee, B.-K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Zhancheng, Zhang, Zhuzhu, Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J.,

Crawford, G.E., Stamatoyannopoulos, J.A., 2012. The accessible chromatin landscape of the human genome. Nature 489, 75–82. https://doi.org/10.1038/nature11232

Tomihara, H., Carbone, F., Perelli, L., Huang, J.K., Soeung, M., Rose, J.L., Robinson, F.S., Lissanu Deribe, Y., Feng, N., Takeda, M., Inoue, A., Poggetto, E.D., Deem, A.K., Maitra, A., Msaouel, P., Tannir, N.M., Draetta, G.F., Viale, A., Heffernan, T.P., Bristow, C.A., Carugo, A., Genovese, G., 2021. Loss of ARID1A Promotes Epithelial-Mesenchymal Transition and Sensitizes Pancreatic Tumors to Proteotoxic Stress. Cancer Res 81, 332–343. https://doi.org/10.1158/0008-5472.CAN-19-3922

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562–578. https://doi.org/10.1038/nprot.2012.016

Trizzino, M., Barbieri, E., Petracovici, A., Wu, S., Welsh, S.A., Owens, T.A., Licciulli, S., Zhang, R., Gardini, A., 2018. The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II. Cell Rep 23, 3933–3945. https://doi.org/10.1016/j.celrep.2018.05.097

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., Margueron, R., Schneider, R., 2013. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell 152, 859–872. https://doi.org/10.1016/j.cell.2013.01.032

Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M., Kouzarides, T., 1997. RB and hbrm cooperate to repress the activation functions of E2F1. Proceedings of the National Academy of Sciences 94, 11268–11273. https://doi.org/10.1073/pnas.94.21.11268

Tsuda, M., Fukuda, A., Kawai, M., Araki, O., Seno, H., 2021. The role of the SWI/SNF chromatin remodeling complex in pancreatic ductal adenocarcinoma. Cancer Sci 112, 490–497. https://doi.org/10.1111/cas.14768

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., Chambers, T., Salgado, R., Savas, P., Loi, S., Birkbak, N.J., Sansregret, L., Gore, M., Larkin, J., Quezada, S.A., Swanton, C., 2017. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol 18, 1009–1021. https://doi.org/10.1016/S1470-2045(17)30516-8

Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M., Sidow, A., 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5, 829–834. https://doi.org/10.1038/nmeth.1246

van Maanen, J.M., Retèl, J., de Vries, J., Pinedo, H.M., 1988. Mechanism of action of antitumor drug etoposide: a review. J Natl Cancer Inst 80, 1526–1533. https://doi.org/10.1093/jnci/80.19.1526

van Steensel, B., Furlong, E.E.M., 2019. The role of transcription in shaping the spatial organization of the genome. Nat Rev Mol Cell Biol 20, 327–337. https://doi.org/10.1038/s41580-019-0114-6

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., Davies, H., Jones, D., Lin, M.-L., Teague, J., Bignell, G., Butler, A., Cho, J., Dalgliesh, G.L., Galappaththige, D., Greenman, C., Hardy, C., Jia, M., Latimer, C., Lau, K.W., Marshall, J., McLaren, S., Menzies, A., Mudie, L., Stebbings, L., Largaespada, D.A., Wessels, L.F.A., Richard, S., Kahnoski, R.J., Anema, J., Tuveson, D.A., Perez-Mancera, P.A., Mustonen, V., Fischer, A., Adams, D.J., Rust, A., Chan-on, W., Subimerb, C., Dykema, K., Furge, K., Campbell, P.J., Teh, B.T., Stratton, M.R., Futreal, P.A., 2011. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 469, 539–542. https://doi.org/10.1038/nature09639

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck, M.F., Ambros, P., Handgretinger, R., Aurias, A., Delattre, O., 1998. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature 394, 203–206. https://doi.org/10.1038/28212

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., Greenberg, M.E., 2017. AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol Cell 68, 1067-1082.e12. https://doi.org/10.1016/j.molcel.2017.11.026

Walser, T., Cui, X., Yanagawa, J., Lee, J.M., Heinrich, E., Lee, G., Sharma, S., Dubinett, S.M., 2008. Smoking and Lung Cancer. Proc Am Thorac Soc 5, 811–815. https://doi.org/10.1513/pats.200809-100TH

Wang, D., Wang, J., Zhou, D., Wu, Z., Liu, W., Chen, Y., Chen, G., Zhang, J., 2023. SWI/SNF complex genomic alterations as a predictive biomarker for response to immune checkpoint inhibitors in multiple cancers. Cancer Immunol Res CIR-22-0813. https://doi.org/10.1158/2326-6066.CIR-22-0813

Wang, H., Bierie, B., Li, A.G., Pathania, S., Toomire, K., Dimitrov, S.D., Liu, B., Gelman, R., Giobbie-Hurder, A., Feunteun, J., Polyak, K., Livingston, D.M., 2016. BRCA1/FANCD2/BRG1-Driven DNA Repair Stabilizes the Differentiation State of Human Mammary Epithelial Cells. Mol Cell 63, 277–292. https://doi.org/10.1016/j.molcel.2016.05.038

Wang, S., Tong, X., Li, C., Jin, E., Su, Z., Sun, Z., Zhang, W., Lei, Z., Zhang, H.-T., 2021. Quaking 5 suppresses TGF-β-induced EMT and cell invasion in lung adenocarcinoma. EMBO Rep 22, e52079. https://doi.org/10.15252/embr.202052079

Wang, T., Gao, X., Zhou, K., Jiang, T., Gao, S., Liu, P., Zuo, X., Shi, X., 2020. Role of ARID1A in epithelial-mesenchymal transition in breast cancer and its effect on cell sensitivity to 5FU. Int J Mol Med 46, 1683–1694. https://doi.org/10.3892/ijmm.2020.4727

Wang, W., Friedland, S.C., Guo, B., O'Dell, M.R., Alexander, W.B., Whitney-Miller, C.L., Agostini-Vulaj, D., Huber, A.R., Myers, J.R., Ashton, J.M., Dunne, R.F., Steiner, L.A., Hezel, A.F., 2019. ARID1A, a SWI/SNF subunit, is critical to acinar cell homeostasis and regeneration and is a barrier to transformation and epithelial-mesenchymal transition in the pancreas. Gut 68, 1245–1258. https://doi.org/10.1136/gutjnl-2017-315541

Wang, X., Haswell, J.R., Roberts, C.W.M., 2014. Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer--mechanisms and potential therapeutic insights. Clin Cancer Res 20, 21–27. https://doi.org/10.1158/1078-0432.CCR-13-0280

Wang, X., Lee, R.S., Alver, B.H., Haswell, J.R., Wang, S., Mieczkowski, J., Drier, Y., Gillespie, S.M., Archer, T.C., Wu, J.N., Tzvetkov, E.P., Troisi, E.C., Pomeroy, S.L., Biegel, J.A., Tolstorukov, M.Y., Bernstein, B.E., Park, P.J., Roberts, C.W.M., 2017. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat Genet 49, 289–295. https://doi.org/10.1038/ng.3746

Wang, X., Sansam, C.G., Thom, C.S., Metzger, D., Evans, J.A., Nguyen, P.T.L., Roberts, C.W.M., 2009. Oncogenesis caused by loss of the SNF5 tumor suppressor is dependent on activity of BRG1, the ATPase of the SWI/SNF chromatin remodeling complex. Cancer Res 69, 8094–8101. https://doi.org/10.1158/0008-5472.CAN-09-0733

Wang, X., Wang, S., Troisi, E.C., Howard, T.P., Haswell, J.R., Wolf, B.K., Hawk, W.H., Ramos, P., Oberlick, E.M., Tzvetkov, E.P., Ross, A., Vazquez, F., Hahn, W.C., Park, P.J., Roberts, C.W.M., 2019. BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant

rhabdoid tumors. Nat Commun 10, 1881. https://doi.org/10.1038/s41467-019-09891-7

Wang, Y., Yang, Q., Wang, Z., 2015. The evolution of nanopore sequencing. Front Genet 5, 449. https://doi.org/10.3389/fgene.2014.00449

Wang, Z., Chen, K., Jia, Y., Chuang, J.-C., Sun, X., Lin, Y.-H., Celen, C., Li, L., Huang, F., Liu, X., Castrillon, D.H., Wang, T., Zhu, H., 2020. Dual ARID1A/ARID1B loss leads to rapid carcinogenesis and disruptive redistribution of BAF complexes. Nat Cancer 1, 909–922. https://doi.org/10.1038/s43018-020-00109-0

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63. https://doi.org/10.1038/nrg2484

Weng, C.-H., Chen, L.-Y., Lin, Yu-Chin, Shih, J.-Y., Lin, Yun-Chieh, Tseng, R.-Y., Chiu, A.-C., Yeh, Y.-H., Liu, C., Lin, Y.-T., Fang, J.-M., Chen, C.-C., 2019. Epithelial-mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. Oncogene 38, 455–468. https://doi.org/10.1038/s41388-018-0454-2

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., Young, R.A., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Wiegand, K.C., Shah, S.P., Al-Agha, O.M., Zhao, Y., Tse, K., Zeng, T., Senz, J., McConechy, M.K., Anglesio, M.S., Kalloger, S.E., Yang, W., Heravi-Moussavi, A., Giuliany, R., Chow, C., Fee, J., Zayed, A., Prentice, L., Melnyk, N., Turashvili, G., Delaney, A.D., Madore, J., Yip, S., McPherson, A.W., Ha, G., Bell, L., Fereday, S., Tam, A., Galletta, L., Tonin, P.N., Provencher, D., Miller, D., Jones, S.J.M., Moore, R.A., Morin, G.B., Oloumi, A., Boyd, N., Aparicio, S.A., Shih, I.-M., Mes-Masson, A.-M., Bowtell, D.D., Hirst, M., Gilks, B., Marra, M.A., Huntsman, D.G., 2010. ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med 363, 1532–1543. https://doi.org/10.1056/NEJMoa1008433

Wilson, B.G., Helming, K.C., Wang, X., Kim, Y., Vazquez, F., Jagani, Z., Hahn, W.C., Roberts, C.W.M., 2014. Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. Mol Cell Biol 34, 1136–1144. https://doi.org/10.1128/MCB.01372-13

Wilson, B.G., Roberts, C.W.M., 2011. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer 11, 481–492. https://doi.org/10.1038/nrc3068

Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.-J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H., Roberts, C.W.M., 2010. Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. Cancer Cell 18, 316–328. https://doi.org/10.1016/j.ccr.2010.09.006

Wilson, M.R., Reske, J.J., Holladay, J., Neupane, S., Ngo, J., Cuthrell, N., Wegener, M., Rhodes, M., Adams, M., Sheridan, R., Hostetter, G., Alotaibi, F.T., Yong, P.J., Anglesio, M.S., Lessey, B.A., Leach, R.E., Teixeira, J.M., Missmer, S.A., Fazleabas, A.T., Chandler, R.L., 2020. ARID1A Mutations Promote P300-Dependent Endometrial Invasion through Super-Enhancer Hyperacetylation. Cell Rep 33, 108366. https://doi.org/10.1016/j.celrep.2020.108366

Wilson, M.R., Reske, J.J., Holladay, J., Wilber, G.E., Rhodes, M., Koeman, J., Adams, M., Johnson, B., Su, R.-W., Joshi, N.R., Patterson, A.L., Shen, H., Leach, R.E., Teixeira, J.M., Fazleabas, A.T., Chandler, R.L., 2019. ARID1A and PI3-kinase pathway mutations in the endometrium drive epithelial transdifferentiation and collective invasion. Nat Commun 10, 3554. https://doi.org/10.1038/s41467-019-11403-6

Wu, S., Fatkhutdinov, N., Zhang, R., 2017. Harnessing mutual exclusivity between TP53 and ARID1 A mutations. Cell Cycle 16, 2313–2314. https://doi.org/10.1080/15384101.2017.1377503

Wu, Z., Nicoll, M., Ingham, R.J., 2021. AP-1 family transcription factors: a diverse family of proteins that regulate varied cellular activities in classical hodgkin lymphoma and ALK+ ALCL. Exp Hematol Oncol 10, 4. https://doi.org/10.1186/s40164-020-00197-9

Xie, C., Tammi, M.T., 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10, 80. https://doi.org/10.1186/1471-2105-10-80

Xu, S., Grullon, S., Ge, K., Peng, W., 2014. Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells. Methods Mol Biol 1150, 97–111. https://doi.org/10.1007/978-1-4939-0512-6_5

Xue, Y., Meehan, B., Fu, Z., Wang, X.Q.D., Fiset, P.O., Rieker, R., Levins, C., Kong, T., Zhu, X., Morin, G., Skerritt, L., Herpel, E., Venneti, S., Martinez, D., Judkins, A.R., Jung, S., Camilleri-Broet, S., Gonzalez, A.V., Guiot,

M.-C., Lockwood, W.W., Spicer, J.D., Agaimy, A., Pastor, W.A., Dostie, J., Rak, J., Foulkes, W.D., Huang, S., 2019. SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. Nat Commun 10, 557. https://doi.org/10.1038/s41467-019-08380-1

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., Ning, Z., 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25, 2865–2871. https://doi.org/10.1093/bioinformatics/btp394

Yoshida, K., Gowers, K.H.C., Lee-Six, H., Chandrasekharan, D.P., Coorens, T., Maughan, E.F., Beal, K., Menzies, A., Millar, F.R., Anderson, E., Clarke, S.E., Pennycuick, A., Thakrar, R.M., Butler, C.R., Kakiuchi, N., Hirano, T., Hynds, R.E., Stratton, M.R., Martincorena, I., Janes, S.M., Campbell, P.J., 2020. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature 578, 266–272. https://doi.org/10.1038/s41586-020-1961-1

Yu, G., Wang, L.-G., He, Q.-Y., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Zeisberg, M., Hanai, J., Sugimoto, H., Mammoto, T., Charytan, D., Strutz, F., Kalluri, R., 2003. BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition and reverses chronic renal injury. Nat Med 9, 964–968. https://doi.org/10.1038/nm888

Zhai, Y., Kuick, R., Tipton, C., Wu, R., Sessine, M., Wang, Z., Baker, S.J., Fearon, E.R., Cho, K.R., 2016. Arid1a inactivation in an Apc- and Pten-defective mouse ovarian cancer model enhances epithelial differentiation and prolongs survival. J Pathol 238, 21–30. https://doi.org/10.1002/path.4599

Zhang, X., Sun, Q., Shan, M., Niu, M., Liu, T., Xia, B., Liang, X., Wei, W., Sun, S., Zhang, Y., Liu, X.S., Song, Q., Yang, Y., Ma, Y., Liu, Y., Yang, L., Ren, Y., Zhang, G., Pang, D., 2013. Promoter Hypermethylation of ARID1A Gene Is Responsible for Its Low mRNA Expression in Many Invasive Breast Cancers. PLOS ONE 8, e53931. https://doi.org/10.1371/journal.pone.0053931

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Y., Sun, Z., Jia, J., Du, T., Zhang, N., Tang, Y., Fang, Y., Fang, D., 2021. Overview of Histone Modification. Adv Exp Med Biol 1283, 1–16. https://doi.org/10.1007/978-981-15-8104-5_1

Zhang, Y., Xu, X., Zhang, M., Bai, X., Li, H., Kan, L., Niu, H., He, P., 2014. ARID1A is downregulated in non-small cell lung cancer and regulates cell proliferation and apoptosis. Tumor Biol. 35, 5701–5707. https://doi.org/10.1007/s13277-014-1755-x

Zhu, G., Shi, R., Li, Y., Zhang, Z., Xu, S., Chen, C., Cao, P., Zhang, H., Liu, M., Pan, Z., Liu, H., Chen, J., 2021. ARID1A, ARID1B, and ARID2 Mutations Serve as Potential Biomarkers for Immune Checkpoint Blockade in Patients With Non-Small Cell Lung Cancer. Front. Immunol. 12, 670040. https://doi.org/10.3389/fimmu.2021.670040

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., Green, M.R., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237. https://doi.org/10.1186/1471-2105-11-237

Zhu, X., Chen, L., Liu, L., Niu, X., 2019. EMT-Mediated Acquired EGFR-TKI Resistance in NSCLC: Mechanisms and Strategies. Front Oncol 9, 1044. https://doi.org/10.3389/fonc.2019.01044

Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M.J., Kroemer, G., 2015. Type I interferons in anticancer immunity. Nat Rev Immunol 15, 405–414. https://doi.org/10.1038/nri3845

PUBLICATIONS

“The important thing is not to stop questioning. Curiosity has its own reason for existing.”

– Albert Einstein

Published in final edited form as:
Oncogene. 2021 April 01; 40(16): 2923–2935. doi:10.1038/s41388-021-01748-y.

ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer

Thaidy Moreno[1], **Beatriz Monterde**[1], **Laura González-Silva**[1], **Isabel Betancor-Fernández**[2], **Carlos Revilla**[1,†], **Antonio Agraz-Doblas**[1], **Javier Freire**[3], **Pablo Isidro**[4], **Laura Quevedo**[1], **Rosa Blanco**[1], **Santiago Montes-Moreno**[3], **Laura Cereceda**[3], **Aurora Astudillo**[4], **Berta Casar**[1,6], **Piero Crespo**[1,6], **Cristina Morales Torres**[5], **Paola Scaffidi**[5], **Javier Gomez-Roman**[3], **Eduardo Salido**[2], **Ignacio Varela**[1,*]

[1]Instituto de Biomedicina y Biotecnología de Cantabria. Universidad de Cantabria-CSIC. Santander, Spain

[2]Departamento de Patología, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER). Tenerife, Spain

[3]Servicio de Anatomía Patológica y Biobanco Valdecilla. HUMV/IDIVAL. Santander, Spain

[4]Biobanco del Principado de Asturias (BBPA). Hospital Universitario Central de Asturias. Oviedo, Spain

[5]Cancer Epigenetics Laboratory, The Francis Crick Institute. London, UK

[6]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC)

Abstract

The survival rate in lung cancer remains stubbornly low and there is an urgent need for the identification of new therapeutic targets. In the last decade, several members of the SWI/SNF chromatin remodeling complexes have been described altered in different tumor types. Nevertheless, the precise mechanisms of their impact on cancer progression, as well as the application of this knowledge to cancer patient management are largely unknown.

In this study, we performed targeted sequencing of a cohort of lung cancer patients on genes involved in chromatin structure. Additionally, we studied at the protein level the expression of these genes in cancer samples and performed functional experiments to identify the molecular mechanisms linking alterations of chromatin remodeling genes and tumor development.

Remarkably, we found that 20% of lung cancer patients show ARID2 protein loss, partially explained by the presence of *ARID2* mutations. Additionally, we showed that ARID2-deficiency provokes profound chromatin structural changes altering cell transcriptional programmes which

[*]***To whom correspondence should be addressed:*** Ignacio Varela, Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria, Albert Einstein 22, 39011 Santander, Spain., Ph. +34 942203931, ignacio.varela@unican.es.
[†]Deceased during the review of the manuscript

Conflict of Interest
The authors disclose no potential conflicts of interest.

Europe PMC Funders Author Manuscripts

bolsters the proliferative and metastatic potential of the cells both *in vitro* and *in vivo*. Moreover, we demonstrated that ARID2 deficiency impairs DNA repair, enhancing the sensitivity of the cells to DNA damaging agents.

Our findings support that *ARID2* is a *bona-fide* tumor suppressor gene in lung cancer that may be exploited therapeutically.

Keywords

ARID2; Lung Cancer; Next-generation sequencing technologies; SWI/SNF

Introduction

Lung cancer is the major cause of cancer-related deaths worldwide with an average 5-year survival rate below 20% irrespective of the subtype[1]. Consequently, any new knowledge about the molecular mechanisms that drive this disease could have a great impact on the treatment of patients. Recently, large genomic projects have facilitated the identification of major players in this tumor type. Thus, small cell lung cancer (SCLC) which constitutes around 15% of all cases, is mainly driven by mutations in *TP53* and *RB1,* but the role of other genes like *PTEN*, *SLIT2* or *CREBBP* has been also described[2]. Among non-small cell lung cancer (NSCLC), more than half of the cases are adenocarcinomas, where *TP53*, *KRAS*, *EGFR*, *ALK*, *ROS1* and *BRAF* are the main recurrently altered genes [3], while squamous cell carcinomas (SCC) are genetically more heterogeneous and poor in actionable mutations so far.

Lately, several members of the SWI/SNF family of chromatin remodeling complexes have been identified recurrently altered in different tumor types adding to the accumulated compelling evidence on the role of chromatin structure in cancer development. It is estimated that approximately 20% of all tumors contain alterations in these complexes, a frequency that is only exceeded by mutations in *TP53* [4]. In the case of non-small cell lung cancer, the expression of any of the two mutually exclusive catalytic ATPase subunits (SMARCA2 or SMARCA4) is lost in 30% of the cases where it is associated with worse prognosis[5]. Additionally, *ARID1A,* which encodes for one of the auxiliary subunits of the complex, frequently appears mutated in lung adenocarcinoma[6].

Materials and Methods

Detailed protocols can be found in Supplementary Methods.

Next-generation sequencing

Cancer patient primary tumor samples and, when available, matched corresponding normal samples, were obtained from different tumor biobanks after the corresponding approval of the hospital ethics committees and patient informed consent. A detailed list of the origin and characteristics of each sample can be found in Supplementary Table 1. DNA was extracted using the Agencourt DNAdvance Beckman Coulter kit (Beckman Coulter, USA), fragmented and submitted to end-repair and adenylation, adaptor ligation and PCR indexing

Europe PMC Funders Author Manuscripts

amplification. Target capture was performed using a Sure Select® user-defined probe kit (Agilent Technologies, USA).

For ATAC-Seq libraries, cell nuclei were extracted using cold lysis buffer and submitted to tagmentation (Nextera DNA Library Preparation Kit, Illumina, USA). After purification, adapter sequences were used to complete Illumina sequencing adapters by PCR with Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific, UK).

Total RNA was purified using Extract Me Total RNA Kit (Blirt, USA). Reverse transcription was performed using the Takara PrimeScript cDNA Synthesis kit (Takara Bio Europe, France). Poly-A mRNA was enriched, fragmented, and submitted to cDNA generation using PrimeScript Enzyme for first strand and RNAse HI, DNA polymerase I and T4 DNA Polymerase (Thermo Fisher Scientific, UK) for the second strand. Afterwards, genomic libraries were generated as above. Individual mRNA expression was measured by qRT-PCR using Luminaris Color HiGreen qPCR Master Mix (Thermo Fisher Scientific, UK). β-actin was used as housekeeping gene and the ΔΔCt method was used for quantification and comparison.

Sequencing Data Analysis

DNA sequence data was mapped to the human genome (hg19) using BWA 0.7.3[7]. Additionally, Samtools 0.1.18[8], Picard 1.61 (http://broadinstitute.github.io/picard/) and GATK 2.2.8[9] were used for format transformation, cleaning, sorting and indexing of the bam files, marking PCR duplicates and performing indel local realignment. RAMSES[10] and PINDEL 0.2.4d[11] were used for substitutions and small insertion and deletion identification respectively. All *ARID2* mutations were validated by PCR amplification coupled with ultrasequencing at 10,000x coverage. Additionally, a similar orthogonal validation of more than 180 mutations randomly picked showed a near 80% of specificity in the mutation calling. OncodriveFML software was run to detect genes with evidence of positive selective pressure[12].

ATAC-Seq reads were aligned against the human genome (hg19) using BWA 0.7.3[7]. Accessible regions were identified using MACS 2.1.2[13]. A combined list of all the regions identified in all the samples by MACS, as well as the list of enhancers annotated in the GeneHancer project[14], were used to identify significant changes in region accessibility in ARID2-deficient cells versus control using DESeq2[15]. Region annotation was performed using ChIPSeeker software[16]. BEDTools[17] was used to estimate the overlapping of the identified regions with ENCODE publicly available histone marks ChIP-Seq A549 data and the results were plotted using deeptools v3.3.1[18]. Motif enrichment analysis was performed using HOMER[19]. Finally, alignments were visualized using IGV genome browser[20].

RNA-Seq data was aligned using Tophat[21] to the human genome (hg19). Differentially expressed genes (DEG) were identified using HTSeq + DESeq2[22,23]. Gene-set Enrichment analysis on Gene Ontology terms was performed using GSEA software[24]. For the analysis of the TCGA database lung adenocarcinoma patients, raw counts for each patient (n=524) were downloaded and normalized using DESeq2 software. Low-ARID2 expressing lung adenocarcinoma patients (n = 64) were defined as those with a normalized expression less

than the mean minus two standard deviations calculated from the whole cohort. Similarly, high ARID2-expressing lung adenocarcinoma patients (n=78) were defined as those with a normalized expression higher than the mean plus two standard deviations. DEG between two groups were identified using DESeq2.

Cell Culture and *in vitro* assays

A549, NCI-H1568 and NCI-H460 lung cancer cell lines were obtained from The Francis Crick Institute common and ATCC repositories, authenticated by STR profiling, and tested for mycoplasma. Tetracycline-inducible pTRIPZ constructs V2THS_74399 (v2), V3THS_347660 (v3) (*ARID2*), V2THS-283735 (*ARID1A*) and V2THS-11753 (*ARID1B*) were used for stable cell line generation (Dharmacon/GE Healthcare, USA). The empty vector (RHS4750) was used as control. Virus production was performed by transfecting HEK293T/17clone cells with the pTRIPZ and packaging constructs. Infected cells were selected with 1 μg/ml puromycin and isolated by FACS using a FACS-Aria II cell sorter (Becton Dickinson, USA) based on TurboRFP expression after the induction with 1 μg/ml of Doxycycline

Extrapolated growth curves were constructed over a period of fourteen days by serial passaging and cell counting with a hemocytometer or by PrestoBlue® assay (Thermo Fisher Scientific, UK). Cell proliferation was also analyzed using the CellTrace™ CFSE Cell Proliferation Kit (Invitrogen, USA) in cells synchronized by gradual serum deprivation following published protocols[25]. Cells were harvested at 48 hours and subjected to division peak resolution by flow cytometry. The cell proliferation index was analyzed using MODFIT software (Verity, USA). Proliferation index was the sum of the cells in all generations divided by the calculated number of original parent cells.

In vitro cell migration assays were performed by using 8-μm pore size transwell chambers (Corning™ Transwell™ Multiple Well Plate) in 24-well plates using 10% FBS in the lower chamber as chemo-attractant. For invasion assays, cells were plated on growth factor-reduced Matrigel (BD Biosciences) pre-coated 8 μm pore transwell chambers. Filters or invasive cells were quantified by fixing chambers in 4% paraformaldehyde for 10 min and staining with crystal violet.

Growth inhibition assays were performed to determine the half maximal inhibitory concentration (IC50) values for different antitumoral drugs. Viability after 48 h was determined by PrestoBlue® reagent (Thermo Fisher Scientific, UK)). IC50 value for each drug were determined with Prism software (GraphPad, USA).

In vivo tumorigenesis assays

Animal studies were conducted in compliance with guidelines for the care and use of laboratory animals and were approved by the Ethics and Animal Care Committee of Universidad de Cantabria. For proliferation assays, five million cells in 500 μl of PBS were subcutaneously injected into the flanks of 6-8-week-old female nude mice (Athymic Nude-Foxn1nu, Envigo, UK). 24 days after the injection, mice were euthanizing and tumor tissues were harvested for analyses. For metastasis assays, 2.5 million of cells in 500 μl of PBS with 0,1% BSA, were tail injected into 6-8-week-old female nude mice. Hairpin expression in the

Europe PMC Funders Author Manuscripts

cells was induced with 1 μg/ml of Doxycycline 7 days before injection. After two months, mice were euthanizing and tumor tissues were harvested for analyses. In both cases, to keep the hairpin expression, the animals were treated from the day of injection with 2 mg/mL of Doxycycline in the drinking water supplemented with 1% sucrose refreshed every 2-3 days.

Western blot analysis

Total protein lysates were prepared in RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 % NP-40, 1 mM Sodium Orthovanadate, 1 mM NaF), separated by SDS-PAGE in 8% polyacrylamide gels and transferred to nitrocellulose membranes. Subsequently, membranes were washed with TBST (50 mM TRIS + 150 mM Sodium chloride + 0,1% Tween 20, pH 7,4) and blocked using 5% non-fat milk solution in TBS for 1 h at room temperature. Membranes were then incubated with primary antibodies anti-ARID2 (E-3, Santa Cruz) and anti-Actin (I-19, Santa Cruz), diluted 1:200 and 1: 1,000 in TBST with 5% (w/v) BSA at 4°C overnight, respectively. Donkey anti-mouse or donkey anti-goat secondary antibodies (LI-COR Biotechnology, USA) conjugated to IRDye 800CW (926-32212) or IRDye 680RD (926-68074) respectively were used as secondary antibodies.

Immunohistochemistry analysis

For ARID2 detection on paraffin sections, antigen retrieval was performed for 32 minutes at 97 ºC in citrate buffer pH 6, incubated with 1:300-1:500 anti-ARID2 antibody (abcam ab113283) and developed with HRP-polymer secondary antibodies (Optiview, Roche). ARID2 expression was evaluated by two pathologists on coded tissue sections, without information about the *ARID2* mutation status. Only surgical pathology cases with enough material, both tumor and non-neoplastic surrounding tissue, were considered for ARID2 immunohistochemistry. A consensus score was reached viewing the slides by two pathologists at a multiheaded scope.

Immunofluorescence was performed in cells fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. The cells were permeabilized with 0.5% Triton X-100 in PBS and blocked with 3% BSA in PBT (PBS containing 0.05% Triton X-100). Finally there were subjected to immunofluorescence staining with ARID2 antibody (E-3, sc-166117 Santa Cruz, USA or A302-230A, Bethyl Laboratories, USA), anti-phospho-Histone H2A. X Ser139 (γH2AX, clone JBW301, Merck Millipore, USA) or anti-53BP1 antibody (H-300, sc-22760, Santa Cruz, USA). Cover slides were incubated with Alexa labeled secondary antibodies and mounted in VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs, USA). Colocalization of ARID2 with 53BP1 or γH2AX was performed measuring the variation in intensity across the lines drawn using the linescan tool from MetaMorph® (Molecular Devices, USA).· For DNA repair assays, cells were treated with 10 μM Etoposide for 1 h. Subsequently, γH2AX foci were quantified using ImageJ software at different recovery times after removing the drug from the media.

Statistical analysis

In all cases, at least three independent experiments were performed in order to assess the statistical significance of all differences. In figure legends, the specific statistical test performed in each case is indicated. In general, for quantitative variables, a one-tailed t-test

with equal variance was used to identify significant differences between groups. For qualitative variables, a Fisher exact test was used in order to identify significant differences between groups of patients. The different software used for the identification of mutations, gene expression differences, enrichment transcription factor binding sites and gene ontology terms have their own statistical models explained in detailed in the references. When multiple tests were performed the significance is shown corrected for multiple testing.

Results

Loss of ARID2 protein expression in 20% of lung cancer patients is partially explained by the presence of *ARID2* mutations

In order to understand better the role of chromatin remodeling complexes in lung cancer development, we performed a genetic screening on the coding sequences of known cancer genes as well as members of the main chromatin remodeling complexes (Suppl. Table 2). We applied targeted next-generation sequencing technologies in a collection of 81 lung cancer cases (40 lung adenocarcinomas, 12 squamous cell carcinomas and 29 small cell carcinomas) (Suppl. Table 1). Interestingly, we found mutations in *ARID2* in 5 of the patients, 3 classified as lung adenocarcinoma and 2 as small cell carcinoma (Figure 1a and Suppl. Table 3). Additionally, to identify lung cancer driver genes, we run OncodriveFML[27] with our data. This software identifies genes with a number and distribution of predicted deleterious mutations higher than expected by chance, evidencing positive selection. *ARID2* ranked second in the list of genes showing significant positive selection after multi-test correction, just below *TP53* (Figure 1b and Suppl. Table 4). To validate these results, we sequenced *ARID2* coding sequences in a second cohort of 144 lung adenocarcinoma cases and found mutations in 12 patients. If we consider all analyzed lung cancer patients, irrespective of the subtype, *ARID2* mutations occur at a frequency of 7.5 % (17/225) (Suppl. Table 3). In the case of lung adenocarcinomas (40 and 144 patients from first and second cohorts respectively), we found *ARID2* non-synonymous mutations in 7% of the samples (13/184), which is near twice the frequency reported in COSMIC database for this tumor type (3.7 %, 83/2241)[28] and ranks *ARID2* among the ten genes most commonly mutated in lung cancer. In concordance with a potential role of *ARID2* as tumor suppressor, many of the identified mutations, clustered at the beginning of the protein sequence, are predicted to generate a premature truncation of the protein (Figure 1c and Suppl. Table 3). Subsequently, to check if the loss of ARID2 function is a common feature in lung cancer, we performed immunohistochemistry analyses in 139 of the studied samples finding loss or low/heterogeneous ARID2 production in approximately 20% of the cases (28/139) (Figure 1d and Suppl. Figure 1). Additionally, loss of *ARID2* expression was significantly more frequent in *ARID2*-mutated patients (6/10 Fisher exact test p=0.0098). Interestingly, this was also true for some samples with missense mutations, which suggests that these mutations might interfere with the correct folding or processing of the protein. Indeed, many of the mutations found are predicted to produce deleterious effects in the protein according to SHIFT or Polyphen algorithms (Suppl. Table 3). We observed a complete loss of ARID2 signal in many lung cancer samples suggesting a selective pressure to inactivate both *ARID2* alleles. In addition, some non-mutated samples showed also loss of ARID2 production, suggesting the existence of non-genetic mechanisms that interfere with *ARID2* expression.

Since the presence of normal tissue contamination in the tumor samples as well as the sequencing strategy followed for normal tissue, prevented an estimation of the cellularity or the zygosity of the mutations, we could not determine whether full loss of ARID2 protein is due to genetic loss or silencing of the wild-type allele.

ARID2-deficiency increases proliferative and metastatic potential *in vitro* and *in vivo*

In order to check if alterations in *ARID2* could promote lung cancer development, we knocked down the protein in different ARID2-proficient NSCLC cell lines. As it can be observed in Figure 2 a-b, *ARID2* mRNA and protein production was efficiently reduced by two different shRNAs. This reduction was accompanied by an increase in the proliferation of A549 cells, as well as their invasion and migration capacities compared to those cells transduced with the empty vector. Similar results were obtained in NCI-H460 cell line (Figure 2 c-e and Suppl. Figure 2). Moreover, when these cells were injected into immunocompromised mice, they showed a greater capacity to produce tumors *in vivo* (Figure 2 f and Suppl. Figure 3).

RNA-Seq experiments in A549 transduced cell lines showed that loss of ARID2 was accompanied by significant changes in the expression of 1155 genes (366 upregulated and 789 downregulated), that supported the observed phenotypes in the cells (Figure 2g and Suppl. Table 5). Thus, we observed a downregulation of genes involved in cellular adhesion and cell differentiation such as *NPNT*, *CDH6*, *FAT3*, *FN1, SOX2* or *SDC2* as well as an upregulation of genes associated with a higher cell-cycle progression such as *CDC45*, *MCM2* or *HIST1H1E,* which could be associated with the increased proliferation, migration and invasion capacities of ARID2-deficient cells. Additionally, we observed downregulation of other tumor suppression genes like *RPS6K2*, *TNFSF10*, *ISM1* or *LDLRAD4* together with upregulation of protumoral and anti-apoptotic genes like *HOXB1*, *BCL2A1* or *RCVRN*. Most of these alterations were not observed when we knocked down ARID1A or ARID1B subunits in the same cells (Suppl. Figure 4a) which indicates a specific gene set regulation by ARID2-containing SWI-SNF complexes. Transcriptional changes in selected genes were further validated by qRT-PCR in independently generated ARID2 knock-down cell lines (Suppl. Figure 4b).

Altogether, these results prove that *ARID2* plays a tumor suppressor function in lung cancer.

ARID2-deficiency is accompanied by widespread chromatin changes, specially affecting enhancers

We hypothesized that gene expression changes observed in ARID2-deficient cells might by the result of changes in SWI/SNF chromatin remodeling activity. To investigate this, we performed ATAC-Seq experiments in ARID2 knocked-down A549 cells. ARID2 loss was accompanied by a general loss of chromatin accessibility with 990 regions that showed a significant loss of chromatin accessibility versus 687 regions that showed increased accessibility (Suppl. Table 6). Interestingly, those regions that lost accessibility in the absence of ARID2 were located distal to gene transcription start sites (Figure 3a) and showed enrichment of enhancer specific H3K4me1 and H3k27ac histone marks according to ENCODE project data. An opposite behavior is observed on those regions that gained

accessibility after ARID2 loss (Suppl. Figure 5). Additionally, AP-1 family transcription factor binding motif, described as abundantly present in enhancers[29], is highly enriched on those regions that showed less accessibility on ARID2-deficient cells (Figure 3b). In order to explore a special impact of ARID2 loss on enhancers, we analyzed the accessibility of those regions annotated as enhancers in the GeneHancer project[14]. As it can be seen in Figure 3c, the loss of accessibility is significantly more profound in enhancers than in the rest of the genome, as 87 % (1744 of 2001) of the enhancers that showed significant accessibility changes, lose accessibility in ARID2-deficient cells. Additionally, many of the target genes of these enhancers showed significant downregulation in the RNA-Seq data (Figure 3d). All this support that *ARID2* is essential to keep an open chromatin conformation around enhancers which significantly impacts on the transcriptional regulation of specific gene networks (Suppl. Table 6).

ARID2 is essential to maintain the expression of the metastasis inhibitor MTSS1 and the adhesion molecule SDK1

In order to identify ARID2 target genes of broad relevance to lung cancer patients, we compared our RNA-Seq results with differential expression analysis performed on human lung adenocarcinoma patients from TCGA database. Eighteen genes were found upregulated in both ARID2-deficient cells, and in low-ARID2 expressing lung adenocarcinoma patients (Suppl. Table 7). Among them, we found *AREG*, *ERG* or *NGF* growth factors that might explain the higher proliferation capabilities of ARID2-deficient cells (Figure 3e and Suppl. Table 7). In addition, we found 133 genes downregulated in both datasets indicating a main gene expression activating role of ARID2 in this cellular context.

Interestingly, among those genes whose expression rely on ARID2, we found *MTSS1*, a well described metastasis inhibitor[29,29,30]., as well as *SDK1*, involved in cell-cell adhesion. The observed significant reduction of *MTSS1* and *SDK1* expression likely explain the higher invasion capabilities of ARID2-deficient cells. Additionally, we found a significant reduction of chromatin accessibility on two enhancers regulating these genes in GeneHancer database after ARID2 loss[14]. This observation is concordant with the hypothesis that *MTSS1* and *SDK1* expressions are positively regulated by ARID2 by keeping an open chromatin structure at their enhancers (Figure 3f).

ARID2 loss impairs DNA damage repair

The RNA-Seq analysis also revealed deleterious consequences suffered by ARID2-deficient cells that could be exploited therapeutically. Gene-set enrichment analyses (GSEA) on the transcriptional alterations observed in ARID2-deficient cells also showed a significant upregulation of genes involved in DNA damage detection and repair, suggesting a defective DNA damage response (Figure 4a). Supporting a role of ARID2 in DNA repair, analysis of its localization in untransduced A549 cells showed a co-localization with γH2AX and 53BP1 at the DNA repair foci (Figure 4b and Suppl. Figure 6). In addition to that, ARID2-deficient cells showed a delay in the resolution of DNA damage foci compared to wildtype cells in A549 and NCI-H460 NSCLC cell lines upon treatment with etoposide. Interestingly, this delay was not dependent on TP53 function as we could see a similar delay in ARID2-deficient NCI-H1568 cell line which is TP53 deficient (Figure 4c-d). These observations

indicate that ARID2 deficiency inhibits efficient DNA repair and suggest that its loss may sensitize cells to DNA damaging agents.

ARID2-deficiency increases cell sensitivity to chemotherapy and veliparib

As platinum-based chemotherapy is widely used for the treatment of lung cancer patients[3], we first examined ARID2-deficient cell lines sensitivity to cisplatin, as well as etoposide. As it can be seen in Figure 4e andSuppl. Figure 7, in concordance with defective DNA repair in the absence of ARID2, ARID2-deficient A549 and NCI-H460 cells exhibited a higher sensitivity to both compounds compared to control cells.

Additionally, in the last decade, many researchers have described a higher sensitivity of PARP inhibitors in tumors harboring defects in DNA repair mechanisms due to synthetic lethality[32]. Consequently, we checked if this might apply as well to ARID2-deficient cells. As it can be seen in Figure 4e, ARID2 loss led to a higher sensibility to veliparib, a well described PARP inhibitor that is under research in several clinical trials in breast, ovarian and, most importantly, lung cancer. This observation suggests ARID2 deficiency as a useful marker for the stratification of lung cancer patients that may benefit for PARP inhibitor treatment.

Discussion

Although some evidence of the presence of *ARID2* alterations in lung cancer have been reported previously[33], the relevance of these alterations for oncogenesis has not been clearly proved. Our results showed an *ARID2* mutation recurrency higher than the one reported in COSMIC database. Additionally, the distribution and predicted impact of the mutations found and our *in-vitro* and *in-vivo* experiments provided compelling evidence of the role of *ARID2* as *bona-fide* tumor suppressor in lung cancer. Supporting this, *ARID2* has been already proposed as cancer driver gene in melanoma and hepatocellular carcinoma[34,35]. We don't have enough data to definitely prove that complete ARID2 activity loss is necessary for tumor progression but several evidences support this idea. First, in many cases we observed a complete loss of ARID2 expression in the human adenocarcinoma samples which suggests a selective pressure to inactivate both *ARID2* alleles. Additionally, we observed that V2THS_74399 shRNA construct was less efficient in abrogating ARID2 expression than V3THS_347660 and cells transduced with the former typically showed less pronounced changes than cells transduced with the latter.

The precise molecular mechanisms by which alterations in chromatin remodeling complexes promote cancer development are not sufficiently understood. Interactions with well-described cancer genes like *TP53*, *RB* or *MYC* have been described[36–38]. In addition to this, they play essential roles in the activation of differentiation and the suppression of proliferative programs of many cellular lineages[39]. In this study we described a list of near 200 genes that are specifically deregulated after ARID2 loss in both our cellular model and in lung adenocarcinoma patients from the TCGA database. Some of these genes, like *AREG*, *EREG* or *NGF* growth factors might account for the higher proliferation capabilities of ARID2-deficient cells. In terms of the molecular mechanisms behind this regulation, we show that ARID2-deficiency is associated with widespread chromatin structural changes.

Europe PMC Funders Author Manuscripts

Our results prove that ARID2 is essential to keep an open chromatin structure in enhancer regions in agreement with an important role of different SWI/SNF members in regulating enhancer activity[40,41]. Two of these ARID2-dependent enhancers regulate *MTSS1* and *SKD1* expression that, consequently, showed a significant downregulation in ARID2-deficient cells. This suggests that ARID2 might regulate directly the expression of *MTSS1* and *SDK1*, although further work is necessary to finally confirm this. MTSS1 is a well described migration and invasion inhibitor, associated with worse prognosis in several tumor types[30,31,42], and SDK1 plays important roles in cell-cell adhesion. Their deficiency might well explain the higher migration and invasion capabilities of ARID2-deficient cells.

In addition, we observed an active role of ARID2 in the detection and repair of DNA damage *in vitro* in lung cancer cell lines, as ARID2-deficient cells present important delays in the resolution of DNA damage foci that were not dependent on TP53 activity. In accordance with this view, other members of the SWI/SNF complex have been shown to be involved in different steps of DNA damage repair[43–45].

Finally, any advance in the possibility of exploiting therapeutically any vulnerability associated to deficiency in SWI/SNF complex genes is of great interest, as approximately 20% of all human cancers are reported to have alterations in this complex. In this study ARID2-deficient cells showed a higher sensitivity to different DNA-damaging therapies, likely as a result of the ARID2 involvement in DNA repair. Considering that platinum-based chemotherapy is still widely used in lung cancer patients with high variable success[3], our results suggest that ARID2 expression might be explored as a stratification marker for these therapies. Moreover, we show that ARID2 deficiency shows synthetic lethality with PARP inhibition using veliparib, an inhibitor that has shown good results in the treatment of breast cancer[46] and is included in several clinical trials on breast, ovarian and, most importantly, lung cancer. Our results suggest that the stratification of lung cancer patients according to ARID2 expression might improve the efficiency of PARP inhibitors in non-small cell lung cancer. Additionally, a very recent study has shown that ARID2-deficient melanoma cells are particularly sensitive to immunotherapy through alterations in mTORC1 and IFNγ pathways[47]. Interestingly, we observed that some downstream response genes in these pathways such as *GBP2*, *GBP3* and *SCD5* are significantly downregulated in ARID2-deficient cells (Suppl. Table 5). All these results support the potential use of *ARID2* expression as a new stratification marker for personalized treatment in lung cancer patients.

In summary, here we present compelling evidence for the role of *ARID2* as tumor suppressor in lung cancer. Although *ARID2* has been proposed as a driver gene in other tumor types[35,48], little has been reported about the molecular mechanisms underlying this involvement. In this work, we propose that its role in lung cancer is exerted by fostering a specific pro-oncogenic transcriptomic program as a result of changes in chromatin structure around enhancers. Importantly, our results indicate that ARID2-deficiency could be exploited for lung cancer patient treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Europe PMC Funders Author Manuscripts

Acknowledgements

In memoriam of Carlos Revilla, we will miss you forever. We would like to thank the support of the funding agencies Spanish Ministerio de Economía y Competitividad, Fundación Ramón Areces, European Research Council, Asociación Española contra el cáncer, Cancer Research UK, UK Medical Research Council, Wellcome Trust, Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III and Fundación Bancaria Cajastur (specific grant references are included in the funding support section). We will like to thank as well the technical support of the different institutions and common services as well as the patients that agreed to participate in this study. We finally want to thank Dr. Francisco Real, Dr. Roland Rad, Dr. Jose Pedro Vaqué and Dr. Javier Leon for providing critical reagents and advice as well as to all the patients that agreed to participate in this study. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Financial Support

I. V. is supported by SAF2012-31627 and SAF2016-76758-R grants from the Spanish Ministerio de Economía y Competitividad (MINECO), by a Fundación Ramón Areces grant and by ERC2014-StG637904 grant from the European Research Council. I. V has been awardee of the Programa Ramón y Cajal (MINECO, Spain). T. M has been awardee of the Ayudas para la contratación de investigadores predoctorales (MINECO, Spain). B. M is awardee of the Ayudas para la formación de profesorado universitario (FPU, Ministerio de Educación y Formación Profesional, Spain). PC laboratory is supported by grant SAF-2015-63638R (MINECO/FEDER, UE); by Centro de Investigación Biomédica en Red de Cáncer (CIBERONC) and by Asociación Española Contra el Cáncer (AECC), grant GCB141423113. BC has been supported by a Retos Jóvenes Investigadores grant SAF2015-73364-JIN (AEI/FEDER, UE) and a grant from Fundación Francisco Cobos. P. S. is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001152), the UK Medical Research Council (FC001152). HUCA/IUOPA which is jointly financed by Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III and Fundación Bancaria Cajastur. This research was funded in part by the Wellcome Trust [FC001152]. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Data availability

DNA-Seq, ATAC-Seq and RNA-Seq data: https://www.ebi.ac.uk/ena/data/search?query=PRJEB26936. All computer code is available upon request.

References

1. Lovly CM, Carbone DP. Lung cancer in 2010: One size does not fit all. Nat Rev Clin Oncol. 2011; 8:68–70. [PubMed: 21278771]
2. Gelsomino F, Rossi G, Tiseo M. MET and Small-Cell Lung Cancer. Cancers. 2014; 6:2100–2115. [PubMed: 25314153]
3. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014; 14:535–546. [PubMed: 25056707]
4. Masliah-Planchon J, Bièche I, Guinebretière J-M, Bourdeaut F, Delattre O. SWI/SNF Chromatin Remodeling and Human Malignancies. Annu Rev Pathol Mech Dis. 2015; 10:145–171.
5. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. Cancer Res. 2003; 63:560–566. [PubMed: 12566296]
6. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. Cell. 2012; 150:1107–1120. [PubMed: 22980975]
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009; 25:2078–2079.
9. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]

10. Martínez N, Almaraz C, Vaqué JP, Varela I, Derdak S, Beltran S, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. Leukemia. 2014; 28:1334–1340. [PubMed: 24296945]
11. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25:2865–2871. [PubMed: 19561018]
12. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016; 17 doi: 10.1186/s13059-016-0994-0
13. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
14. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database. 2017; 2017 doi: 10.1093/database/bax028
15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. [PubMed: 25516281]
16. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015; 31:2382–2383. [PubMed: 25765347]
17. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
18. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016; 44:W160–W165. [PubMed: 27079975]
19. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
20. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Res. 2017; 77:e31–e34. [PubMed: 29092934]
21. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
22. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]
23. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–169. [PubMed: 25260700]
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005; 102:15545–15550. [PubMed: 16199517]
25. Lauand C, Niero EL, Dias VM, Machado-Santelli GM. Cell cycle synchronization and BrdU incorporation as a tool to study the possible selective elimination of ErbB1 gene in the micronuclei in A549 cells. Braz J Med Biol Res Rev Bras Pesqui Medicas E Biol. 2015; 48:382–391.
26. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. PLOS Biol. 2018; 16:e2005970. [PubMed: 29969450]
27. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016; 17 doi: 10.1186/s13059-016-0994-0
28. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017; 45:D777–D783. [PubMed: 27899578]
29. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. 2013; 23:777–788. [PubMed: 23482648]

Europe PMC Funders Author Manuscripts

30. Kayser G, Csanadi A, Kakanou S, Prasse A, Kassem A, Stickeler E, et al. Downregulation of MTSS1 expression is an independent prognosticator in squamous cell carcinoma of the lung. Br J Cancer. 2015; 112:866–873. [PubMed: 25625275]
31. Giacobbe A, Compagnone M, Bongiorno-Borbone L, Antonov A, Markert EK, Zhou JH, et al. p63 controls cell migration and invasion by transcriptional regulation of MTSS1. Oncogene. 2016; 35:1602–1608. [PubMed: 26119942]
32. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017; 355:1152–1158. [PubMed: 28302823]
33. Manceau G, Letouzé E, Guichard C, Didelot A, Cazes A, Corté H, et al. Recurrent inactivating mutations of *ARID2* in non-small cell lung carcinoma. Int J Cancer. 2013; 132:2217–2221. [PubMed: 23047306]
34. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A Landscape of Driver Mutations in Melanoma. Cell. 2012; 150:251–263. [PubMed: 22817889]
35. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, et al. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. Nat Genet. 2011; 43:828–829. [PubMed: 21822264]
36. Flowers S, Beck GR, Moran E. Transcriptional Activation by pRB and Its Coordination with SWI/SNF Recruitment. Cancer Res. 2010; 70:8282–8287. [PubMed: 20851996]
37. Tordella L, Khan S, Hohmeyer A, Banito A, Klotz S, Raguz S, et al. SWI/SNF regulates a transcriptional program that induces senescence to prevent liver cancer. Genes Dev. 2016; 30:2187–2198. [PubMed: 27737960]
38. Romero OA, Torres-Diz M, Pros E, Savola S, Gomez A, Moran S, et al. MAX Inactivation in Small Cell Lung Cancer Disrupts MYC-SWI/SNF Programs and Is Synthetic Lethal with BRG1. Cancer Discov. 2014; 4:292–303. [PubMed: 24362264]
39. Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011; 11:481–492. [PubMed: 21654818]
40. Alver BH, Kim KH, Lu P, Wang X, Manchester HE, Wang W, et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nat Commun. 2017; 8:14648. [PubMed: 28262751]
41. Nakayama RT, Pulice JL, Valencia AM, McBride MJ, McKenzie ZM, Gillespie MA, et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. Nat Genet. 2017; doi: 10.1038/ng.3958
42. Taylor MD, Bollt O, Iyer SC, Robertson GP. Metastasis suppressor 1 (MTSS1) expression is associated with reduced in-vivo metastasis and enhanced patient survival in lung adenocarcinoma. Clin Exp Metastasis. 2018; 35:15–23. [PubMed: 29218652]
43. Lee H-S, Park J-H, Kim S-J, Kwon S-J, Kwon J. A cooperative activation loop among SWI/SNF, γ-H2AX and H3 acetylation for DNA double-strand break repair. EMBO J. 2010; 29:1434–1445. [PubMed: 20224553]
44. Niimi A, Chambers AL, Downs JA, Lehmann AR. A role for chromatin remodellers in replication of damaged DNA. Nucleic Acids Res. 2012; 40:7393–7403. [PubMed: 22638582]
45. Ray A, Mir SN, Wani G, Zhao Q, Battu A, Zhu Q, et al. Human SNF5/INI1, a Component of the Human SWI/SNF Chromatin Remodeling Complex, Promotes Nucleotide Excision Repair by Influencing ATM Recruitment and Downstream H2AX Phosphorylation. Mol Cell Biol. 2009; 29:6206–6219. [PubMed: 19805520]
46. Rugo HS, Olopade OI, DeMichele A, Yau C, van 't Veer LJ, Buxton MB, et al. Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. N Engl J Med. 2016; 375:23–34. [PubMed: 27406347] Pan D, Kobayashi A, Jiang P, Ferrari de Andrade L, Tay RE, Luoma A, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science. 2018:eaao1710.
48. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A Landscape of Driver Mutations in Melanoma. Cell. 2012; 150:251–263. [PubMed: 22817889]
49. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018; 28:1747–1756. [PubMed: 30341162]

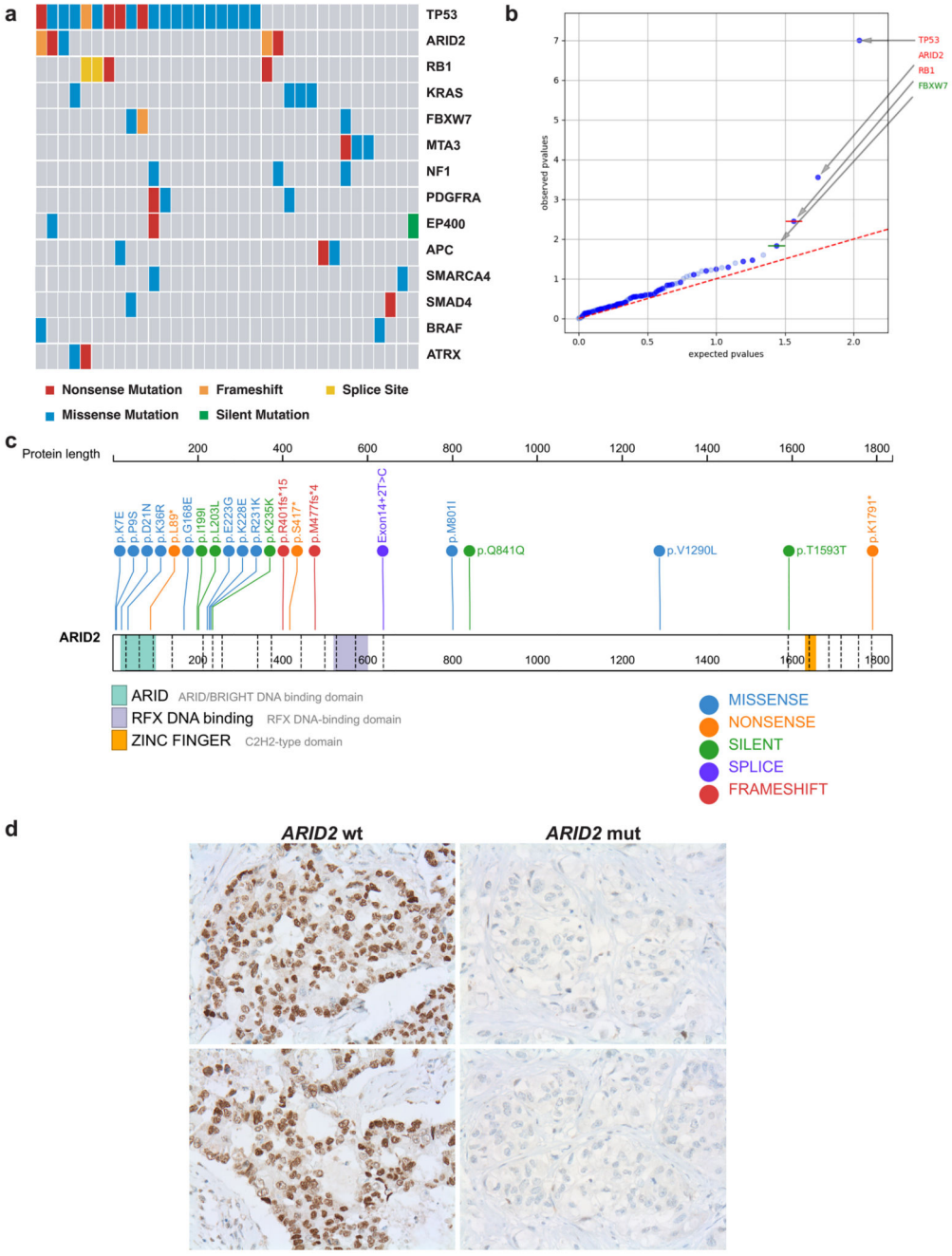


Figure 1. Frequent *ARID2* mutations associated with protein loss.

(a) Box representation of the mutated patients for the most significantly mutated genes according to OncodriveFML in the lung cancer cohort generated with Maftools [49]. Each box in the central matrix represents an independent patient. Colored boxes represent mutated patients for the corresponding gene in a color code indicating the type of mutation. (b) Representation of the significance analysis of the functional impact of the mutations found in each gene performed by OncodriveFML. Genes in read showed a q-value < 0.1 after multi-test correction. (c) Visual representation of the location of all identified *ARID2*

Europe PMC Funders Author Manuscripts

mutations in our discovery and validation lung cancer cohort in relation to the functional protein domains. (d) Representative images of ARID2 immunohistochemistry experiments in two *ARID2*-mutated (right) and two *ARID2*-wildtype (left) lung adenocarcinomas.

Europe PMC Funders Author Manuscripts

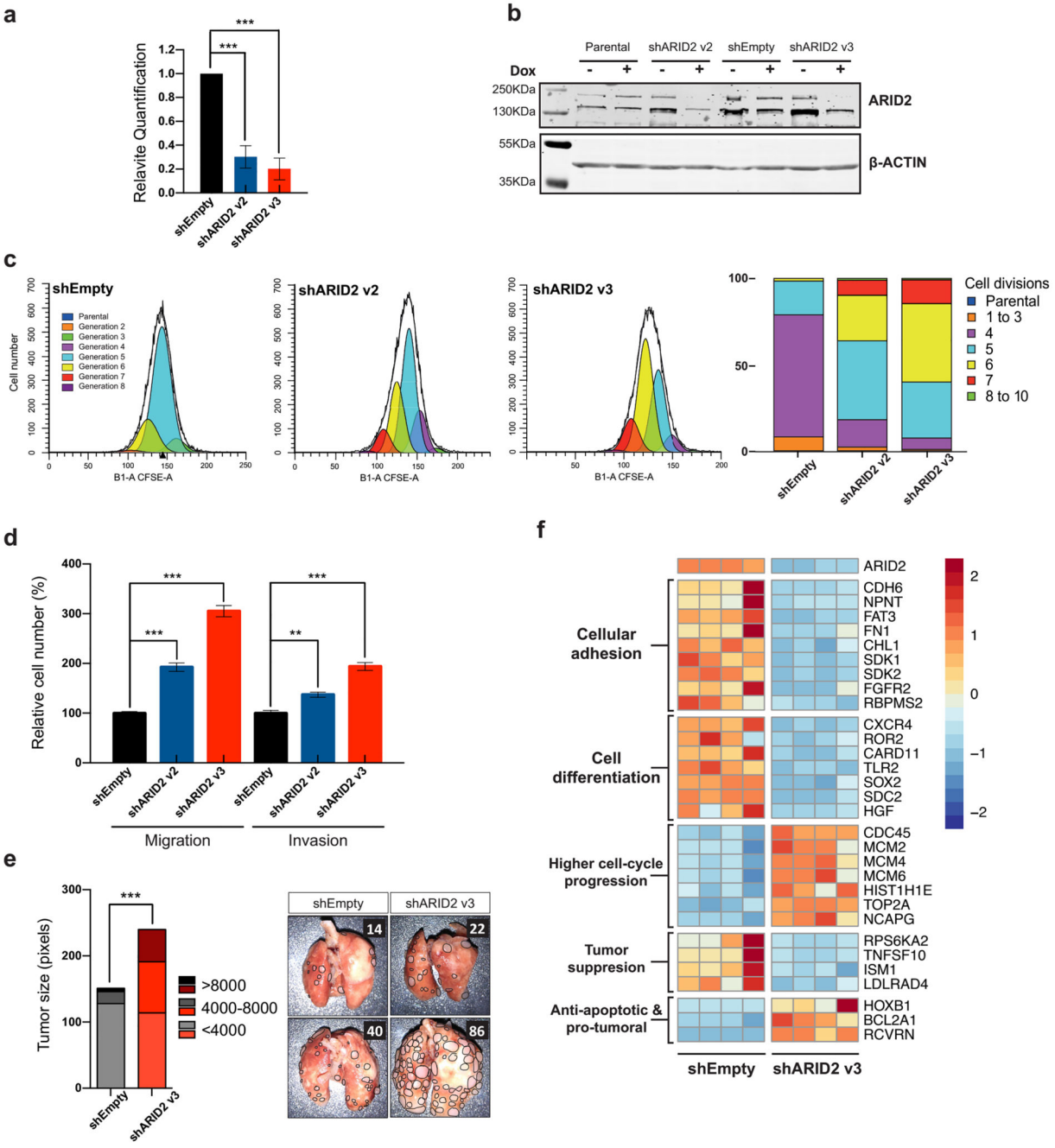


Figure 2. ARID2 deficiency is associated with an increase in oncogenesis *in vitro* and *in vivo*.

(a) Bar representation of ARID2 expression level fold changes measured by qRT-PCR in A549 cells transduced with shARID2 v2 and v3 as well as the empty vector which is used as control. Data is shown as mean ± SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed t-test *p < 0.05, ** p < 0.01 and *** p <0.001). (b) Representative image of a western blot analysis measuring ARID2 protein levels in A549 parental cells as well as those cell lines transduced with ARID2 shRNAs and the empty vector. In all the cases, the results are shown with and without induction of the

Europe PMC Funders Author Manuscripts

shRNA expression by doxycycline (Dox) treatment. (c) Representative experiment of the number of cell divisions suffered by the cells in 48h estimated by CFSE labelling in A549 cells by flow cytometry. Bar quantification on the number of cells that have suffered each number of cell divisions is represented on the right. (d) Bar representation of quantified cells in destination chamber on migration and invasion assays of A549 cells transduced with two different ARID2 shRNAs (blue and red bars). Data is shown as mean ± SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed t-test *p < 0.05, ** p < 0.01 and *** p <0.001). (e) Representative images of lung metastasis generated in intravenously injected mice with A549 cells transduced either with shEmpty, or shARID2 v3 vectors (n=7 per group). Individual metastasis are delineated in the image and counted (upper right corner numbers). On the left, a quantification of the number and size of the tumors generated in the two groups is shown. (Fisher exact test *p < 0.05, ** p < 0.01 and *** p <0.001). (f) Heatmap representation of a selection of differently expressed genes in ARID2-deficient A549 cells (n=4) and grouped according to their biological function. Expression differences go from red (upregulation) to blue (downregulation) according to the log2 of the fold change.

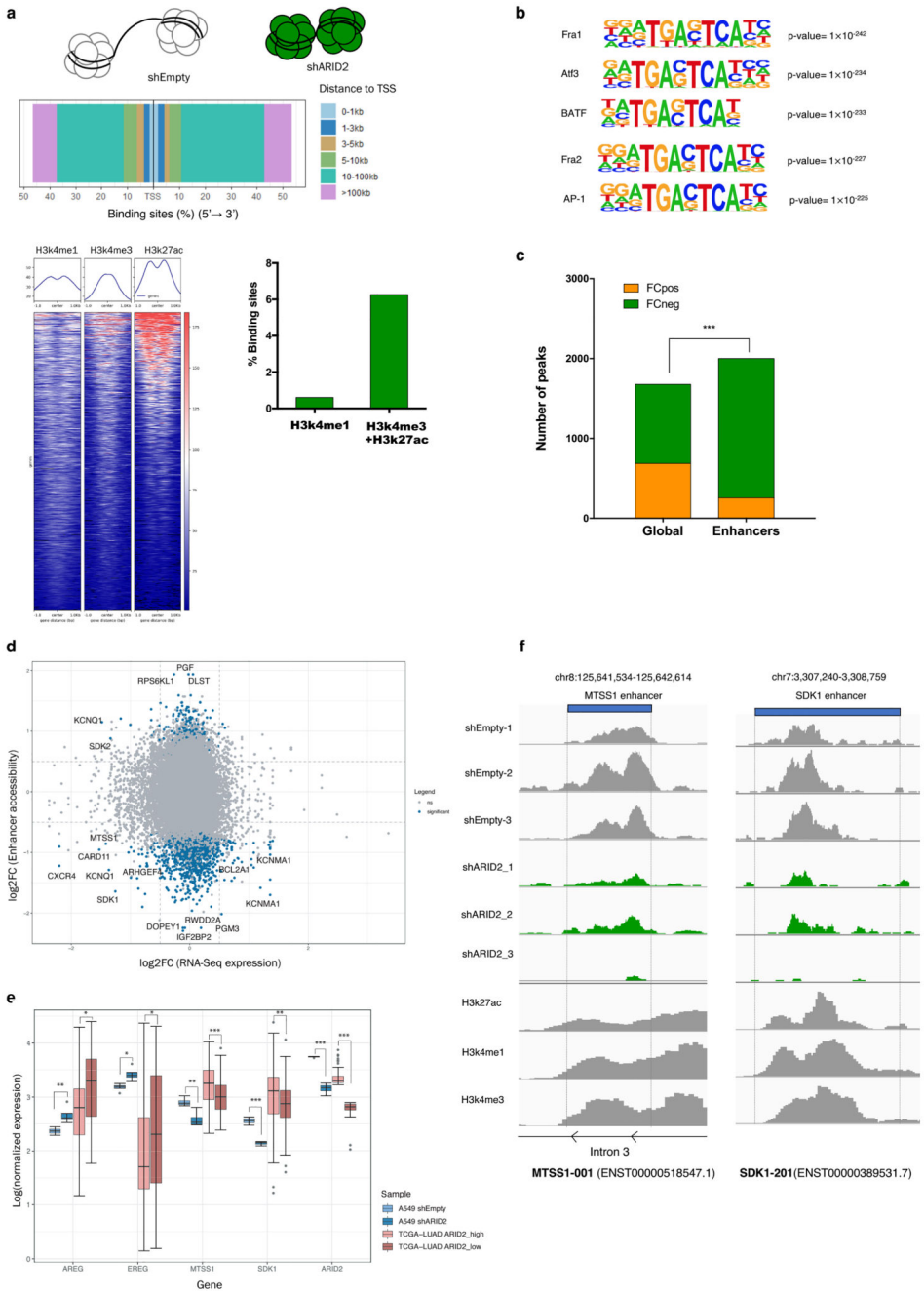


Figure 3. Profound chromatin structural changes on enhancers affect gene expression after ARID2 loss.

(a) Analysis of the genomic regions that significantly lost chromatin accessibility after ARID2 loss in A549 cells. In the upper pannel the regions are grouped according to their distance to nearest gene transcription start site (TSS). Below, the intensity of H3K4me1, H3K4me3 and H3k27ac histone marks in each identified region is represented by heatmaps (left). Additionally, the percentage of identified regions that overlap with regions with histone modification marks are represented in a bar graph (right). (b) Enrichment of sequence motifs identified by HOMER in those regions that lost chromatin accessibility

after ARID2-loss in A549 cells. (c) Bar representation of the number of genome-wide (global) regions or enhancer regions which lost (FCneg green bars) or gained (FCpos yellow bars) after ARID2 loss in A549 cells. (Fisher exact test *** p <0.001) (d) Dot plot representing the correlation between the accessibility changes in enhancer regions and expression changes on the target genes for each enhancer. Blue dots represent enhancers that showed significant accessibility changes in ARID2-deficient A549 cells. (e) Boxplot graph of gene expression differences identified in both our ARID2-deficient A549 cells and in lowly *ARID2* expressing lung adenocarcinoma patients (ARID2_low) versus highly *ARID2* expressing patients (ARID2_high) from TCGA database, (DEseq2 statistical test *p < 0.05, ** p < 0.01 and *** p <0.001). (f) Visualization, in two described *MTSS1* and *SDK1* enhancers, of read alignments for the different replicates of our ATAC-Seq experiments in ARID2-deficient A549 cells. Additionally, read alignments of ChIP-Seq experiments performed against different histone marks during ENCODE project are also represented.

Europe PMC Funders Author Manuscripts

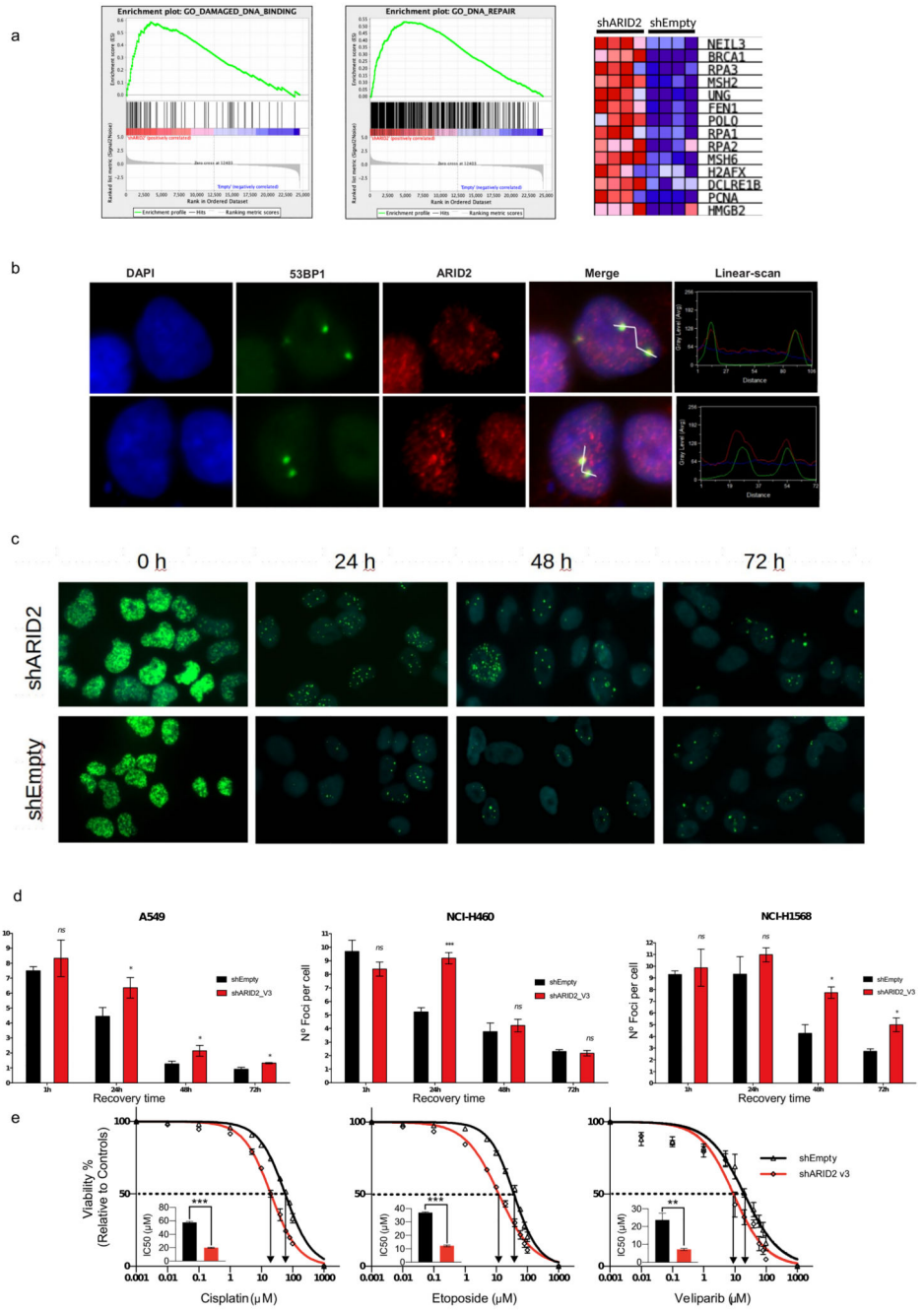


Figure 4. ARID2 deficiency affects DNA repair and affects sensitivity to anti-tumor therapies

(a) Results of the Gene-Set Enrichment Analysis (GSEA) from RNA-Seq experiments in ARID2-deficient A549 cells showing enrichment of genes involved in different DNA repair ontologies. On the right, a heatmap representation of the expression of different genes belonging to these ontologies in the different replicates is included. (b) Representative images of immunofluorescence experiments demonstrating colocalization of 53BP1 (green) and ARID2 (red) in A549 cells in DNA damage foci. Colocalization was confirmed on the right through the parallel quantification of red and green signals on a manually selected path

through the image using the LineScan tools from Methamorph software (c) Representative images of DNA-repair foci visualized by H2AX immunofluorescence (green) in transduced nuclei stained with DAPI (blue) at different recovery times after the treatment with etoposide in NCI-H460 cell lines. (d) Bar representation of the foci quantification in each transduced cell line. (d) Bar quantification of the number of foci per cell at different recovery times after DNA damage induced by etoposide, in different ARID2-deficient cell lines, the results are represented as mean +/- SEM of at least three independent experiments, (two-tailed t-test *p < 0.05, ** p < 0.01 and *** p <0.001). (e) Representative experiments measuring cell survival to increasing concentrations of cisplatin, etoposide and veliparib on A549 cells transduced with shEmpty (black), or shARID2v3 (red) vectors. Bar graphs represent the calculated IC50 value for each experiment. In all cases, the results are represented as mean +/- SEM of at least three independent experiments, (two-tailed t-test *p < 0.05, ** p < 0.01 and *** p <0.001).

Europe PMC Funders Author Manuscripts

Role of SWI/SNF chromatin remodelling genes in lung cancer development

Beatriz Monterde & Ignacio Varela

Instituto de Biomedicina y Biotecnología de Cantabria. Universidad de Cantabria-CSIC. Santander, Spain

Abstract

SWI/SNF family of chromatin remodelling complexes uses the energy of ATP to change the structure of DNA, playing key roles in DNA regulation and repair. It is estimated that up to 25% of all human cancers contain alterations in SWI/SNF, although the precise molecular mechanisms for their involvement in tumor progression are largely unknown. Despite the improvements achieved in the last decades on our knowledge of lung cancer molecular biology, it remains the major cause of cancer-related deaths worldwide and it is in urgent need for new therapeutic alternatives. We and others have described recurrent alterations in different SWI/SNF genes in nearly 20% of lung cancer patients, some of them with a significant association with worse prognosis, indicating an important role of SWI/SNF in this fatal disease. These alterations might be therapeutically exploited, as it has been shown in cellular and animal models with the use of EGFR inhibitors, DNA-damaging agents and several immunotherapy approaches. Therefore, a better knowledge of the molecular mechanisms regulated by SWI/SNF alterations in lung cancer might be translated into a therapeutic improvement of this frequently lethal disease. In this review, we summarize all the evidence of SWI/SNF alterations in lung cancer, the current knowledge about the potential mechanisms involved in their tumorigenic role, as well as the results that support a potential exploitation of these alterations to improve the treatment of lung cancer patients.

1. Introduction

Chromatin remodelling complexes are key components that modulate chromatin landscape. They use the energy provided by ATP hydrolysis to disrupt nucleosome-DNA contacts, move nucleosomes along DNA and catalyze its ejection, insertion or exchange [1,2]. Consequently, they modify the accessibility of specific regions of DNA to the enzymatic transcriptional machinery, as well as different DNA-binding proteins, cofactors and regulators, playing important roles in gene expression regulation and DNA repair.

According to their subunit composition and biochemical activity, chromatin remodelling complexes can be divided in four major families: SWI/SNF, INO80/SWR1, ISWI and NURD/CHD, being SWI/SNF complex the most clearly implied in tumor development [1].

SWI/SNF complexes are complicated macromolecular assemblies consisting of many diverse and variable subunits. They are evolutionary conserved from yeast to mammals [1]. Mammalian SWI/SNF complexes are currently divided in three broad subfamilies: canonical BAF (cBAF); polybromo-associated BAF (PBAF) and the recently described non-canonical BAF (ncBAF) [3].

All complexes contain a group of the so-called core subunits, which are highly conserved and always include SMARCC1/BAF155, SMARCC2/BAF170 and either one of the two

mutually exclusive ATPases subunits (SMARCA2/BRM or SMARCA4/BRG1), but they also contain numerous accessory subunits that provide each of the complexes with a distinct identity. Thus, SMARCB1/BAF47/SNF5 and SMARCE1/BAF57 are common to cBAF and PBAF families, whereas ARID1A and ARID1B subunits are specific for cBAF family and ARID2, PBRM1 and BRD7 are specific for PBAF. Finally, ncBAF complexes utilize instead BRD9, GLTSCR1 and GLTSCR1L subunits for their core assembly [1,3,4]. Accessory subunits are thought to contribute to the targeting, assembly and regulation of lineage-specific gene networks [1,4].

Some common subunits of the SWI/SNF complex are encoded by genes that produce different isoforms by alternative splicing. Moreover, they belong to gene families that often display differential lineage-restricted expression, what means that some subunits are only expressed in some specific tissues [4]. It is therefore likely that a large number of different SWI/SNF complexes exist in mammals, which in turn control distinct set of genes and signalling pathways in different cellular contexts.

Mammalian SWI/SNF complexes have broad roles in transcriptional regulation. Essential roles for these complexes have been identified during neurogenesis, myogenesis, adipogenesis, osteogenesis and hematopoiesis [1]. The precise mechanism behind the role of SWI/SNF complexes in transcriptional regulation is not fully understood, but recent reports indicate that they are highly enriched in enhancers, modulating their accessibility to different transcription factors and opposing to the inactivation mediated by polycomb repressive complexes [5,6]. In particular, in a rhabdoid tumor cell model, loss of *SMARCB1* alters SWI/SNF complex integrity and its targeting to enhancers implied in cell differentiation, while keeping SWI/SNF binding to super-enhancers required for tumor survival [5]. In agreement with this observation, as a result of *SMARCB1* re-expression in *SMARCB1*-deficient sarcoma cell lines, there is an increase in genome-wide BAF complex occupancy, which mediates enhancer activation and opposes to the polycomb-mediated repression at bivalent promoters [6].

Large genome sequencing studies have evidenced a prominent role of chromatin structure in cancer development. The catalytic and accessory subunits of the complex have been found recurrently mutated in several tumor types and their alteration has been associated with tumor progression [4–9], whereas core components in general present a very low mutation rate. One significant exception to this is *SMARCB1,* which is inactivated via biallelic mutations in 98% of malignant rhabdoid tumors (MRT) (rare and aggressive childhood cancers) [10,11].

Regarding accessory and catalytic subunits, *PBRM1* is inactivated in 41% of clear cell renal cell carcinoma patients [12,13]. *ARID1A*, which is probably the most broadly mutated subunit, is altered in 50% of ovarian clear cell carcinomas (one of the most lethal subtypes of ovarian cancer), in 35% of endometroid carcinomas [14,15], in 9.4% of colorectal and in 8.2% of lung cancers [7]. It is estimated that more than 90% of ovarian small-cell carcinoma hypercalcaemic type patients harbor biallelic inactivating mutations in *SMARCA4* [16]. *SMARCA2* is epigenetically silenced or transcriptionally inactivated in rhabdoid tumors [17] and ovarian small-cell carcinoma hypercalcaemic type [18]. In non-small cell lung cancer (NSCLC) SMARCA2 and/or SMARCA4 expressions are frequently lost and they are associated with worse prognosis [8,19]. Lastly, *ARID2* expression is lost in approximately 20% of NSCLC [9], has been found inactivated in 18.2% of hepatocellular carcinomas [20] and it is also a significantly mutated gene in melanoma [21].

Overall, genes encoding subunits of the SWI/SNF chromatin remodelling complex are collectively mutated in almost 25% of all human cancers, what places SWI/SNF as the second most frequent alteration in cancer, just after *TP53* [4–7].

2. SWI/SNF complex disruption in lung cancer

Lung cancer remains the major cause of cancer-related deaths worldwide, with an average 5-year survival rate bellow 20%, irrespective of the subtype [22]. Large genomic projects have facilitated the identification of major players in this tumor type: small cell lung cancer (SCLC, 15% of cases) is mainly driven by mutations in *TP53* and *RB1*, in lung adenocarcinoma (LUAD, 40% of cases) *KRAS, EGFR, ALK, ROS1* and *BRAF* are the main recurrently altered genes, while squamous cell carcinoma (LUSC) is genetically more heterogenous [23].

In the last two decades, several studies provided many evidences of an important role of SWI/SNF complexes alterations in lung cancer development. Thus, in 2003 Reisman and coworkers sequenced a total of 20 non-small cell lung cancer (NSCLC) cell lines and observed loss of SMARCA2 and SMARCA4 expression in 30% of the cases. In addition to this, when they compared the survival of SMARCA2/SMARCA4-negative against SMARCA4-positive patients, they show a significant higher survival in the second group, highlighting the role of these genes as tumor suppressors [8]. Ten years afterwards, Matsubara and coworkers confirmed the prognostic role of SMARCA2/SMARCA4 expression by comparing the survival of 442 primary lung adenocarcinomas, which were divided in three groups according to their expression of SMARCA2 and SMARCA4. In this study, they observed that the high expression of both catalytic subunits correlated with better prognosis [24]. Soon afterwards, Bell and coworkers validated in a large cohort of 440 human NSCLC that decreased expression of SMARCA4 was associated with worse prognosis [25].

Similar observations have been done in other accessory subunits of the complex. For instance, Zhang and coworkers analyzed the expression of ARID1A in 106 NSCLC and reported almost a 35% decreased staining, which significantly correlated with a poor differentiated stage, a higher TNM score and nodal metastasis [26]. Additionally, Manceau and coworkers found *ARID2* loss-of-function mutations in 5% of NSCLC [27].

In line with this, our group has recently analyzed ARID2 expression in a cohort of 139 lung cancer patients and has shown loss or low/heterogeneous ARID2 production in nearly 20% of the cases [9]. What is more, we have proved that *ARID2* deficiency in lung cancer cellular models is accompanied by a decrease in chromatin accessibility around enhancers, which is associated with a pro-tumoral transcriptional program in these cells. These results are in agreement with an increased proliferative and metastatic potential both *in vitro* and *in vivo* [9].

Finally, among NSCLC, SWI/SNF mutations occur more frequently in lung adenocarcinoma (LUAD) than in squamous cell carcinoma. According to the currently available LUAD data deposited on TCGA portal, SWI/SNF genes are collectively mutated in around 20% of the cases, with *SMARCA4* (8%), *ARID1A* (6%), *ARID2* (6%) and *SMARCA2* (3%) being the most commonly altered subunits (Figure 1). Consequently, SWI/SNF alteration might constitute a new molecular mechanism of vital importance in lung cancer progression. A better understanding of the gene networks regulated by each SWI/SNF subcomplex can unravel new opportunities for the management of lung cancer patients.

SWI/SNF-mutated patients show *TP53* and *KRAS* mutations in similar ratios to no-mutated patients. Nevertheless, *SWI/SNF*-mutant patients show a lower *EGFR* mutation ratio (9/115 ~ 7.8%) than the 12% observed in the complete cohort, although this difference is not statically significant (P value=0.0797, one-tailed Fisher's exact test). This bias is mainly the result of *SMARCA4* mutations showing significant mutual exclusivity with *EGFR* mutations in this dataset, as only 2% of *SMARCA4*-mutated patients show *EGFR* mutations versus the 12% that would be expected by chance (P value=0.0151, one-tailed Fisher's exact test). This observation extends previous results [24] and suggests the existence of either synthetic lethality relationships or redundant tumor progression promoting activities.

3. Potential pro-tumoral mechanisms associated with SWI/SNF alterations

Despite the clear role of SWI/SNF alterations in the development of different types of cancer, the molecular pathways behind its contribution to tumorigenesis still remain elusive [1–4].

There is a considerable amount of scientific evidence that links SWI/SNF with different DNA repair mechanisms, which suggests that a potential mechanism of tumorigenesis in *SWI/SNF*-deficient cells is through the promotion of genomic instability.

In particular, SMARCA4 is able to bind to γ-H2AX nucleosomes through its bromodomain, facilitating double-strand break (DSB) repair [27]. Similarly, SMARCB1 promotes nucleotide excision repair (NER) by interacting with UV damage recognition factor XPC at DNA damage sites [28]. Our laboratory described that ARID2 colocalizes with γ-H2AX and 53BP1 at the DNA-repair foci and its loss in human lung cancer cell lines is associated with a delay in the resolution of DNA damage foci, which in turn sensitizes cells to DNA-damaging agents (such as, cis-platin, etoposide and PARP inhibitors) [9]. Supporting the role of SWI/SNF complexes in DNA repair, Shen and coworkers reported a higher sensitivity to DNA-damaging agents in *ARID1A*-deficient cells [29]. Similarly, *SMARCA4*-mutated cell lines showed an increase in DNA damage foci and an activation of the ATR pathway [30]. Finally, it has been described that mutations in *ARID1A* are mutually exclusive with the ones in *TP53*, that could suggest redundant functions [31].

SWI/SNF complex is able to interact with canonical cancer associated genes, such as *RB1*, *TP53* and *MYC* [32,33], suggesting that defects in these chromatin remodelling complexes could promote cancer through the activation of canonical cancer pathways. Regarding this, it has been demonstrated that *SWI/SNF* and *TP53* or *PTEN* mutations do not usually concur in the same colorectal tumor neither in the same ovarian clear-cell carcinoma [7,30]. In the context of lung cancer, *SMARCA4* downregulation does not occur simultaneously with *EGFR* mutations [32,33], neither with *MYC* amplification [34].

Important relationships between the catalytic subunit of the complex SMARCA4 and critical molecular pathways have also been reported, which might provide molecular clues for its role in tumorigenesis. Thus, in NSCLC *SMARCA4* loss is synthetic lethal with CDK4/6 inhibition both *in vitro* and *in vivo* [35]. Additionally, in a murine mouse model of LUAD where *SMARCA4* is ablated, there is enhanced oxygen consumption [36], what suggests that *SMARCA4*-deficient cell lines rely on this metabolic process to sustain their high metabolic rate. Supporting this idea, derived tumor cells are more sensitive to OXPHOS inhibition. Moreover, it is known that *SMARCA4* regulates the expression of *MAX* and it is also required to activate neuroendocrine transcriptional programs and to upregulate *MYC* target genes, such

as the ones implicated in glycolysis. In agreement with this observation, *SMARCA4* alterations do not concur with *MAX* inactivation in SCLC cell lines [32].

Additionally, we reported that the accessory subunit ARID2 is necessary to keep an open chromatin structure around the enhancers of the metastasis inhibitor *MTSS1* and the adhesion molecule *SDK1*, allowing its expression [9]. This might explain the observed increased migration and invasion capabilities of these cells *in vitro*.

Finally, alterations in any of the catalytic subunits of the complex (SMARCA2 or SMARCA4) correlates with a loss of bronchial epithelial phenotype, low E-cadherin and high vimentin expression, what supports an enhanced epithelial to mesenchymal transition as a result of the loss of SWI/SNF function [24].

4. Impact of *SWI/SNF* mutations in current lung cancer treatments

Due to the recurrency of SWI/SNF alterations in different tumor types, any new knowledge about how to exploit these alterations for therapy could improve the treatment of many cancer patients. According to current available NSCLC treatments, several studies have provided very interesting data.

Firstly, due to its role in DNA repair, many authors reported that lack of some accessory subunits of the SWI/SNF complex sensitizes tumor cells to therapeutic strategies aimed at promoting genomic instability. For instance, *SMARCA4* [25], *ARID1A* [29] and *ARID2* [9] alterations have been linked to a higher sensitivity to platinum-based chemotherapy or treatments based on the inhibition of PARP. Moreover, *SMARCA4* [30] and *ARID1A* [37]-mutated lung cancer cell lines are more sensitive to ATR inhibition. Additionally, *SMARCA4*-mutant NSCLC cell lines are also more sensitive to aurora kinase inhibition [36]. Therefore, loss of any of the above-mentioned accessory subunits might be used as biomarkers in lung cancer patients that could benefit from therapeutic strategies that exploit defects in the DNA repair machinery of tumor cells (Figure 2).

Secondly, it has been described that loss of *SMARCE1* subunit gives rise to an enhanced resistance to MET and ALK inhibitors in NSCLC cell lines, which is restored by the use of EGFR inhibitors [38]. It is thought that this resistance relies on AKT and ERK activation. If these results are consistent in human patients, *SMARCE1* expression might emerge as a potential predictive marker for drug response to MET and ALK inhibitors and *SMARCE1*-mutated tumors may benefit from treatments based on EGFR inhibition (Figure 2).

Finally, another strategy that has been proved successful in many cancer types, including lung cancer, is the boost of the proper immune response of the patient against its tumor cells. Regarding this, in 2018 Pan and coworkers reported that the murine melanoma cell line B16F10 increased its sensitivity to T cell-mediated cytotoxicity after mutating different subunits of the PBAF complex (in particular, *Arid2, Pbrm1* and *Brd7* [39]). These results seem to indicate that tumors harboring mutations in the SWI/SNF complex might respond better to immunotherapy. In agreement with this hypothesis, mice bearing tumors induced with *Pbrm1*-deficient B16F10 cells were more strongly infiltrated by cytotoxic T cells, developed smaller tumors and had an improved survival in comparison with control cells. In accordance with this result, Miao and coworkers performed a whole exome sequencing of human metastatic clear cell renal carcinomas and verified that *PBRM1* inactivation is associated with better clinical outcome from immune checkpoint inhibitor (ICI) therapy [40]. Finally, a recent study published by Zhu and coworkers in 2021 performed on NSCLC

patients has extended this observation to other accessory subunits of the complex, such as *ARID1A* and *ARID1B* [41] (Figure 2). However, the precise mechanism that explains why *ARID1A*, *ARID1B* or *ARID2*-mutated tumors are more likely to benefit from ICI therapy has not been elucidated.

5. New therapeutic opportunities for *SWI/SNF*-mutated patients

SWI/SNF complexes exist in multiple compositions since several subunit positions can be occupied alternatively by proteins encoded from different genes. According to that, many authors have reported a dependency on the remaining functional alternative subunit in cells with alterations in specific SWI/SNF components. That is the case of SMARCA2 in *SMARCA4*-deficient cells [42,43] or ARID1B in *ARID1A*-deficient cells [44,45]. Similar synthetic lethality relationships have been described among SMARCA4 and ARID2, SMARCA4 and ACTB, as well as SMARCC1 and SMARCC2 [46]. Consequently, it has been postulated that this dependency could be therapeutically exploited. According to that, it has been demonstrated that SMARCA2/4 specific protein degraders or ATP inhibitors impair the proliferation in *SMARCA4*-deficient lung cancer cellular models [47,48].

Lastly, another proposed alternative is based on the previously reported antagonistic role between polycomb repressive complex 2 (PRC2) and SWI/SNF [49,50]. According to this dependency, EZH2 (catalytic subunit of PRC2) inhibition has shown very promising anti-tumoral activity in a variety of *SWI/SNF*-deficient cell lines (highlighting *SMARCB1* and *ARID1A*) of different tumor types and could suppose a future treatment opportunity for *SWI/SNF*-deficient lung tumors [49,50]. In agreement with these promising results, an specific inhibitor of EZH2, tazemetostat, has been approved for the treatment of epithelioid sarcoma and MRT [51], constituting the first target therapy described for *SWI/SNF*-mutant tumors.

6. Concluding remarks

In the last decade, an overwhelming amount of data supporting an important role of SWI/SNF alterations in human cancer has been generated. This is especially interesting in the case of lung cancer, the first cause of cancer-related deaths worldwide, in urgent need for the discovery of new molecular pathways to be therapeutically exploited. Thus, in this tumor type at least four different SWI/SNF genes, *SMARCA4*, *ARID1A*, *ARID2* and *SMARCA2*, are found recurrently mutated with what it seems an exclusion pattern with other highly recurrently mutated cancer genes, such as *EGFR*. This opens at least two potential alternatives: either *SWI/SNF*-mutated patients constitute a biologically different group that deserve special consideration; or SWI/SNF alterations constitute an alternative way of activating EGFR pathway. In any case, further research on the molecular pathways affected by SWI/SNF alterations could have a great impact in the treatment of lung cancer patients.

It is estimated that around 25% of all cancer patients contain SWI/SNF alterations. Thus, any new knowledge of how to exploit these alterations for therapy could have a huge impact in the treatment of multiple cancer types. In this context, it is especially remarkable the last reports linking SWI/SNF alterations with a higher sensitivity to immunotherapy. This strategy has been incorporated into the therapeutic portfolio of many malignancies and besides its great efficacy in many cases, it is still not fully understood why many patients are refractory

to this treatment. Therefore, understanding the molecular pathways by which SWI/SNF alterations modulate the immune response is key to improve the efficiency of immunotherapy strategies in cancer patients.

7. Perspectives

- SWI/SNF is collectively mutated in almost 20% of lung cancer patients, which offers a new opportunity to improve our molecular knowledge, as well as the therapeutic opportunities in the first cause of cancer-related deaths worldwide.
- SWI/SNF deficiency can be therapeutically exploited with currently available treatments used in lung cancer, such as those based on the promotion of genomic instability or immunotherapy; as well as with new strategies exploiting synthetic lethal relationships among SWI/SNF components and with other epigenetic systems.
- A better knowledge of the molecular mechanisms regulated by SWI/SNF alterations could be translated into better treatments for lung cancer patients.

8. Figures and tables

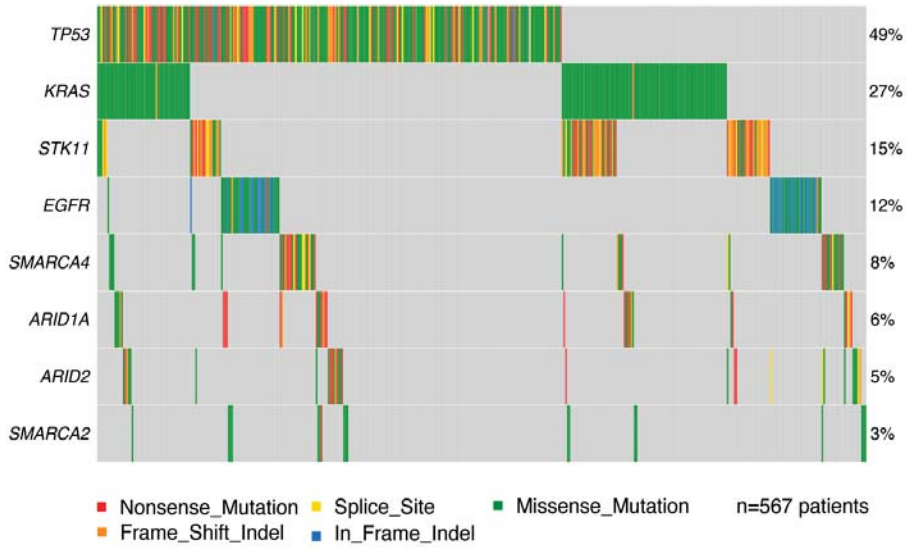


Figure 1. Mutual exclusivity between SWI/SNF and canonical pathways altered in lung adenocarcinoma. Oncostrip representation of the non-synonymous mutations found by MuTect2 [52] in patients with lung adenocarcinoma (LUAD) extracted from the TCGA-LUAD project dated March 2022 (*maftools* R library was used to create the maf object and to plot the somatic variants). For each gene (rows), the presence of mutations in each of the 567 patients (columns) is represented in different colors according to the mutation predicted functional consequence.

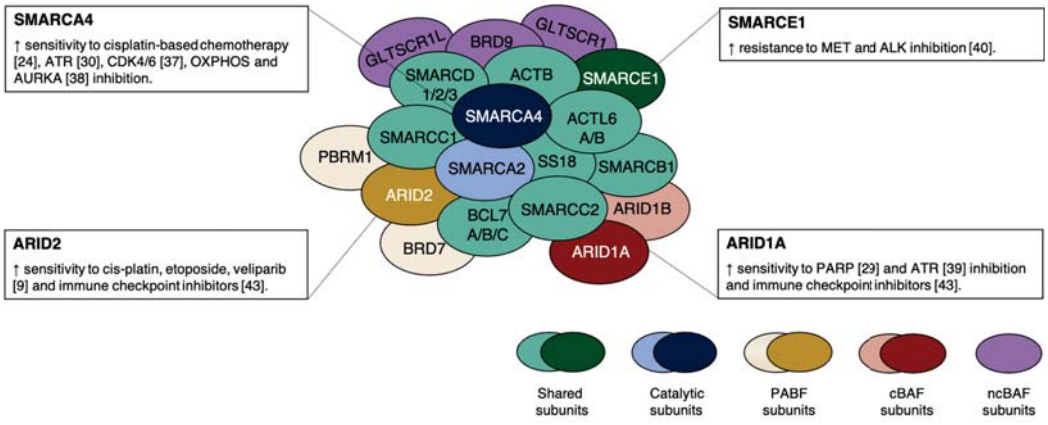


Figure 2. Therapeutic implications of SWI/SNF altered lung tumors. The most commonly mutated subunits of the complex and its association with different sensitivities to lung cancer treatments are depicted in blue or orange, depending on whether they are currently available or under development. Shared subunits of SWI/SNF complexes are represented in green. The catalytic subunits of the complex (SMARCA2 and SMARCA4) appear in blue. The accessory subunits that define PBAF, cBAF and ncBAF families are represented in yellow, red and purple, respectively. The bibliographic references supporting SWI/SNF alteration and its response to different treatments are indicated in brackets.

9. References

1. Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011 Jun 9;11(7):481–92.

2. Gonzalez-Perez A, Jene-Sanz A, Lopez-Bigas N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. Genome Biol. 2013;14(9):r106.

3. Mashtalir N, D'Avino AR, Michel BC, Luo J, Pan J, Otto JE, et al. Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell. 2018 Nov 15;175(5):1272-1288.e20.

4. Mittal P, Roberts CWM. The SWI/SNF complex in cancer — biology, biomarkers and therapy. Nat Rev Clin Oncol. 2020 Jul;17(7):435–48.

5. Wang X, Lee RS, Alver BH, Haswell JR, Wang S, Mieczkowski J, et al. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat Genet. 2017 Feb;49(2):289–95.

6. Nakayama RT, Pulice JL, Valencia AM, McBride MJ, McKenzie ZM, Gillespie MA, et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. Nat Genet. 2017 Nov;49(11):1613–23.

7. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet. 2013 Jun;45(6):592–601.

8. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in Human Lung Cancer Cell Lines and Primary Lung Cancers: Correlation with Poor Prognosis. :8.

9. Moreno T, Monterde B, González-Silva L, Betancor-Fernández I, Revilla C, Agraz-Doblas A, et al. ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to

chemotherapy in lung cancer. Oncogene. 2021 Apr;40(16):2923–35.

10. Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv. 2015 Jun;1(5):e1500447.

11. Versteege I, Sévenet N, Lange J, Rousseau-Merck MF, Ambros P, Handgretinger R, et al. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature. 1998 Jul 9;394(6689):203–6.

12. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014 Mar;46(3):225–33.

13. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011 Jan 27;469(7331):539–42.

14. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med. 2010 Oct 14;363(16):1532–43.

15. Jones S, Wang T-L, Shih I-M, Mao T-L, Nakayama K, Roden R, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science. 2010 Oct 8;330(6001):228–31.

16. Lu B, Shi H. An In-Depth Look at Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT): Clinical Implications from Recent Molecular Findings. J Cancer. 2019;10(1):223–37.

17. Kahali B, Yu J, Marquez SB, Thompson KW, Liang SY, Lu L, et al. The silencing of the SWI/SNF subunit and anticancer gene BRM in Rhabdoid tumors. Oncotarget. 2014 May 30;5(10):3316–32.

18. Karnezis AN, Wang Y, Ramos P, Hendricks WP, Oliva E, D'Angelo E, et al. Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. J Pathol. 2016 Feb;238(3):389–400.

19. Herpel E, Rieker RJ, Dienemann H, Muley T, Meister M, Hartmann A, et al. SMARCA4 and SMARCA2 deficiency in non–small cell lung cancer: immunohistochemical survey of 316 consecutive specimens. Ann Diagn Pathol. 2017 Feb 1;26:47–51.

20. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, et al. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. Nat Genet. 2011 Aug 7;43(9):828–9.

21. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A landscape of driver mutations in melanoma. Cell. 2012 Jul 20;150(2):251–63.

22. Lovly CM, Carbone DP. Lung cancer in 2010: One size does not fit all. Nat Rev Clin Oncol. 2011 Feb;8(2):68–70.

23. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 31;511(7511):543–50.

24. Matsubara D, Kishaba Y, Ishikawa S, Sakatani T, Oguni S, Tamura T, et al. Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. Cancer Sci. 2013 Feb;104(2):266–73.

25. Bell EH, Chakraborty AR, Mo X, Liu Z, Shilo K, Kirste S, et al. SMARCA4/BRG1 Is a Novel Prognostic Biomarker Predictive of Cisplatin-Based Chemotherapy Outcomes in Resected Non–Small Cell Lung Cancer. Clin Cancer Res. 2016 May 12;22(10):2396–404.

26. Zhang Y, Xu X, Zhang M, Bai X, Li H, Kan L, et al. ARID1A is downregulated in non-small

cell lung cancer and regulates cell proliferation and apoptosis. Tumor Biol. 2014 Jun;35(6):5701–7.

27. Lee H-S, Park J-H, Kim S-J, Kwon S-J, Kwon J. A cooperative activation loop among SWI/SNF, γ-H2AX and H3 acetylation for DNA double-strand break repair. EMBO J. 2010 Apr 21;29(8):1434–45.

28. Ray A, Mir SN, Wani G, Zhao Q, Battu A, Zhu Q, et al. Human SNF5/INI1, a Component of the Human SWI/SNF Chromatin Remodeling Complex, Promotes Nucleotide Excision Repair by Influencing ATM Recruitment and Downstream H2AX Phosphorylation. Mol Cell Biol. 2009 Dec;29(23):6206–19.

29. Shen J, Peng Y, Wei L, Zhang W, Yang L, Lan L, et al. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov. 2015 Jul;5(7):752–67.

30. Gupta M, Concepcion CP, Fahey CG, Keshishian H, Bhutkar A, Brainson CF, et al. BRG1 Loss Predisposes Lung Cancers to Replicative Stress and ATR Dependency. Cancer Res. 2020 Sep 15;80(18):3841–54.

31. Wu S, Fatkhutdinov N, Zhang R. Harnessing mutual exclusivity between TP53 and ARID1 A mutations. Cell Cycle Georget Tex. 2017;16(24):2313–4.

32. Romero OA, Torres-Diz M, Pros E, Savola S, Gomez A, Moran S, et al. *MAX* Inactivation in Small Cell Lung Cancer Disrupts MYC–SWI/SNF Programs and Is Synthetic Lethal with BRG1. Cancer Discov. 2014 Mar;4(3):292–303.

33. Burrows AE, Smogorzewska A, Elledge SJ. Polybromo-associated BRG1-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. Proc Natl Acad Sci U S A. 2010 Aug 10;107(32):14280–5.

34. Romero OA, Setien F, John S, Gimenez-Xavier P, Gómez-López G, Pisano D, et al. The tumour suppressor and chromatin-remodelling factor BRG1 antagonizes Myc activity and promotes cell differentiation in human cancer. EMBO Mol Med. 2012 Jul;4(7):603–16.

35. Xue Y, Meehan B, Fu Z, Wang XQD, Fiset PO, Rieker R, et al. SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. Nat Commun. 2019 Dec;10(1):557.

36. Lissanu Deribe Y, Sun Y, Terranova C, Khan F, Martinez-Ledesma J, Gay J, et al. Mutations in the SWI/SNF complex induce a targetable dependence on oxidative phosphorylation in lung cancer. Nat Med. 2018 Jul;24(7):1047–57.

37. Karachaliou N, Paulina Bracht JW, Rosell R. ARID1A Gene Driver Mutations in Lung Adenocarcinomas. J Thorac Oncol. 2018 Dec;13(12):e255–7.

38. Papadakis AI, Sun C, Knijnenburg TA, Xue Y, Grernrum W, Hölzel M, et al. SMARCE1 suppresses EGFR expression and controls responses to MET and ALK inhibitors in lung cancer. Cell Res. 2015 Apr;25(4):445–58.

39. Pan D, Kobayashi A, Jiang P, Ferrari de Andrade L, Tay RE, Luoma AM, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science. 2018 Feb 16;359(6377):770–5.

40. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science. 2018 Feb 16;359(6377):801–6.

41. Zhu G, Shi R, Li Y, Zhang Z, Xu S, Chen C, et al. ARID1A, ARID1B, and ARID2 Mutations Serve as Potential Biomarkers for Immune Checkpoint Blockade in Patients With Non-Small Cell Lung Cancer. Front Immunol. 2021 Aug 26;12:670040.

42. Oike T, Ogiwara H, Tominaga Y, Ito K, Ando O, Tsuta K, et al. A Synthetic Lethality–Based

Strategy to Treat Cancers Harboring a Genetic Deficiency in the Chromatin Remodeling Factor BRG1. Cancer Res. 2013 Sep 1;73(17):5508–18.

43. Hoffman GR, Rahal R, Buxton F, Xiang K, McAllister G, Frias E, et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc Natl Acad Sci. 2014 Feb 25;111(8):3128–33.

44. Helming KC, Wang X, Wilson BG, Vazquez F, Haswell JR, Manchester HE, et al. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat Med. 2014 Mar;20(3):251–4.

45. Kelso TWR, Porter DK, Amaral ML, Shokhirev MN, Benner C, Hargreaves DC. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. eLife. 2017 Oct 2;6:e30506.

46. Schick S, Rendeiro AF, Runggatscher K, Ringler A, Boidol B, Hinkel M, et al. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. Nat Genet. 2019 Sep;51(9):1399–410.

47. Papillon JPN, Nakajima K, Adair CD, Hempel J, Jouk AO, Karki RG, et al. Discovery of Orally Active Inhibitors of Brahma Homolog (BRM)/SMARCA2 ATPase Activity for the Treatment of Brahma Related Gene 1 (BRG1)/SMARCA4-Mutant Cancers. J Med Chem. 2018 Nov 21;61(22):10155–72.

48. Farnaby W, Koegl M, Roy MJ, Whitworth C, Diers E, Trainor N, et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. Nat Chem Biol. 2019 Jul;15(7):672–80.

49. Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A, et al. Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. Proc Natl Acad Sci. 2013 May 7;110(19):7922–7.

50. Bitler BG, Aird KM, Garipov A, Li H, Amatangelo M, Kossenkov AV, et al. Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. Nat Med. 2015 Mar;21(3):231–8.

51. Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. 2020 Dec;36(12):936–50.

52. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. bioRxiv. 2019 Jan 1;861054.

10. Author contributions, conflicts of interest, acknowledgments and funding

Author contributions

B. M. performed the literature review, wrote the first draft of the manuscript and performed suggested modifications in further versions. I. V. performed a literature review, proposed the first review schema, proposed modifications to early versions and performed the final modifications to the manuscript.

Declaration of interest

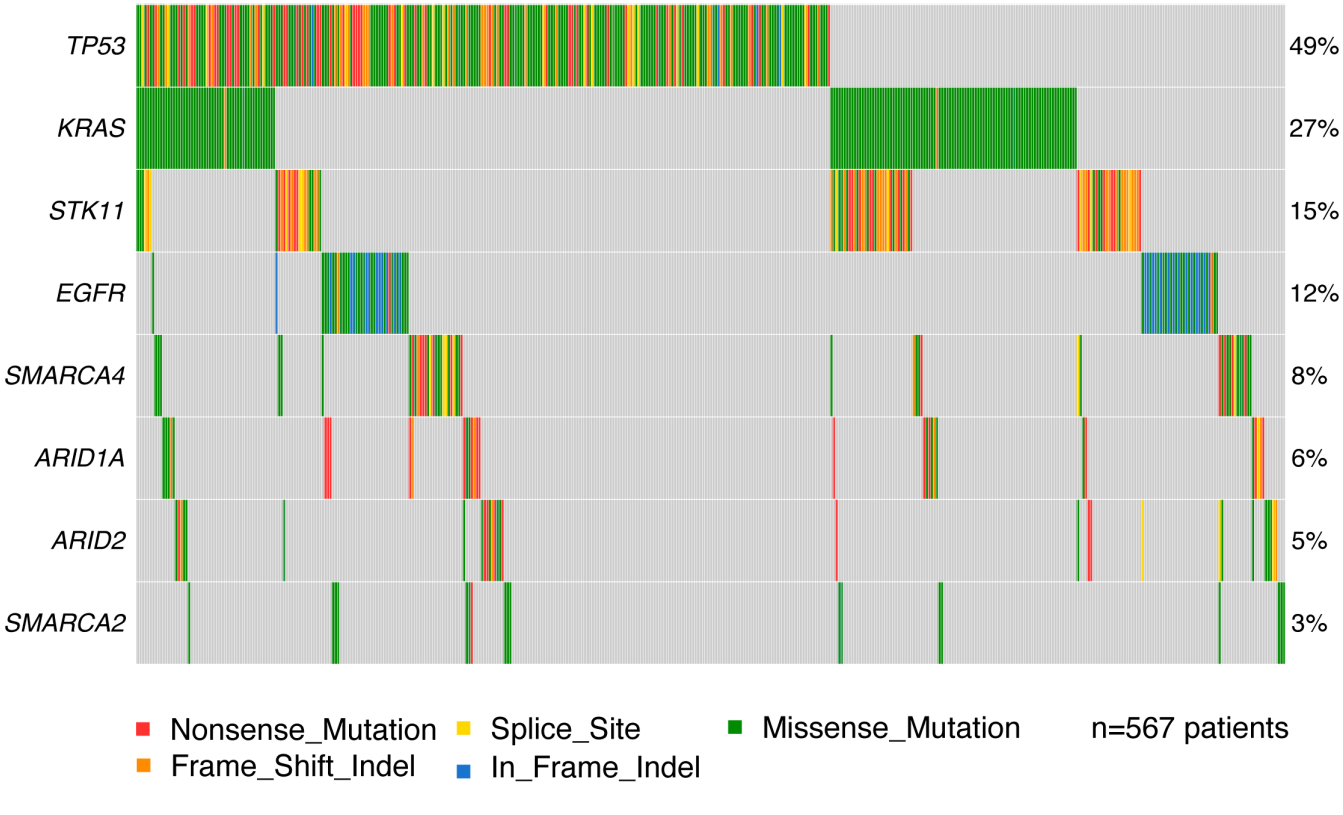
The authors declare that there are no competing interests associated with the manuscript.

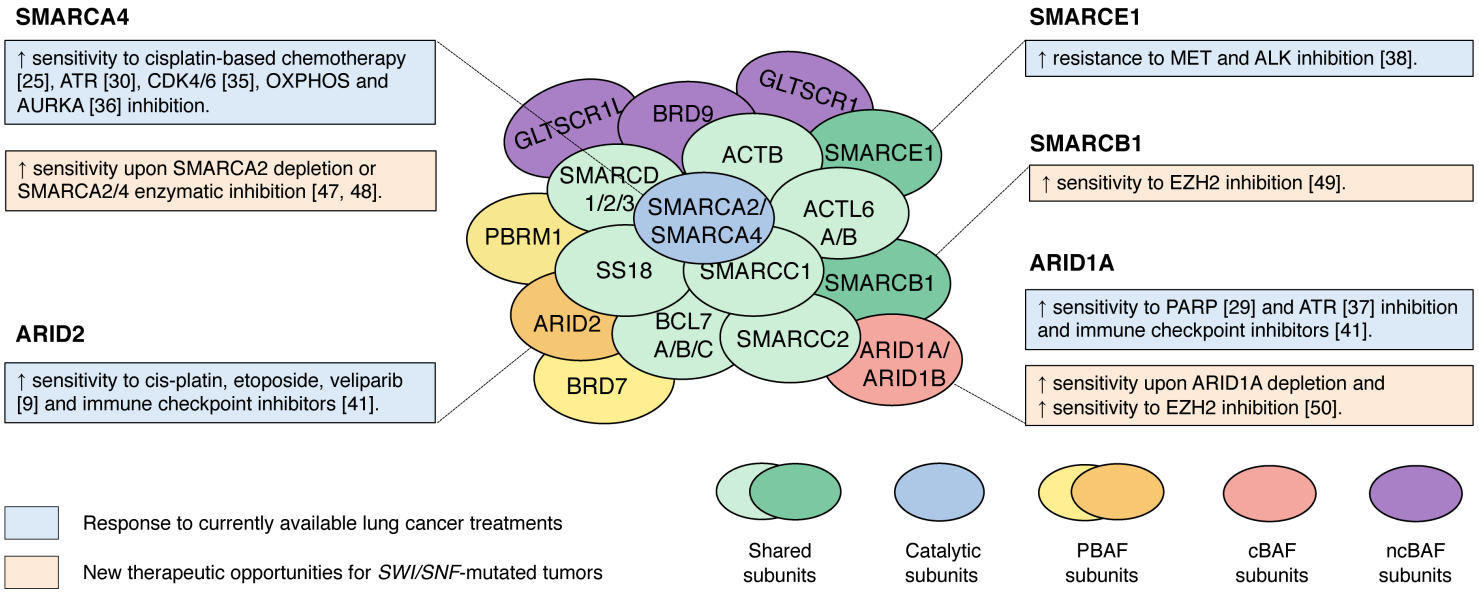
Acknowledgements

We would like to thank the staff members of the Biobanco Valdecilla (Tissue Node, PT13/0010/0024), as well as to Dr. Javier Freire, Dr. Javier Gómez-Román and Dr. Eduardo Salido for their exceptional advice in the management of lung cancer patients.

Funding

This work is supported by the Spanish Ministerio de Ciencia e Innovación (MCIN/ AEI /10.13039/501100011033. ref PID2020-117539GB-I00). B.M is awardee of the Ayudas para la formación de profesorado universitario (FPU, Ministerio de Ciencia, Innovación y Universidades, Spain).


TP53
KRAS
STK11
EGFR
SMARCA4
ARID1A
ARID2
SMARCA2
49%
27%
15%
12%
8%
6%
5%
3%
Nonsense_Mutation
Frame_Shift_Indel
Splice_Site
In_Frame_Indel
Missense_Mutation
n=567 patients


SMARCA4
↑ sensitivity to cisplatin-based chemotherapy [25], ATR [30], CDK4/6 [35], OXPHOS and AURKA [36] inhibition.
↑ sensitivity upon SMARCA2 depletion or SMARCA2/4 enzymatic inhibition [47, 48].
ARID2
↑ sensitivity to cis-platin, etoposide, veliparib [9] and immune checkpoint inhibitors [41].
GLTSCR1L
BRD9
GLTSCR1
ACTB
SMARCE1
SMARCD 1/2/3
SMARCA2/ SMARCA4
ACTL6 A/B
PBRM1
SS18
SMARCC1
SMARCB1
ARID2
BCL7 A/B/C
SMARCC2
ARID1A/ ARID1B
BRD7
SMARCE1
↑ resistance to MET and ALK inhibition [38].
SMARCB1
↑ sensitivity to EZH2 inhibition [49].
ARID1A
↑ sensitivity to PARP [29] and ATR [37] inhibition and immune checkpoint inhibitors [41].
↑ sensitivity upon ARID1A depletion and ↑ sensitivity to EZH2 inhibition [50].
Response to currently available lung cancer treatments
New therapeutic opportunities for SWI/SNF-mutated tumors
Shared subunits
Catalytic subunits
PBAF subunits
cBAF subunits
ncBAF subunits

https://doi.org/10.1093/nar/gkac619

The HDAC7–TET2 epigenetic axis is essential during early B lymphocyte development

Alba Azagra[1,2,†], Ainara Meler[1,2,†], Oriol de Barrios[1,2], Laureano Tomás-Daza[3,4], Olga Collazo[1,2], Beatriz Monterde[5], Mireia Obiols[6], Llorenç Rovirosa[3], Maria Vila-Casadesús[7], Mónica Cabrera-Pasadas[3,4], Mar Gusi-Vives[1], Thomas Graf[7,8], Ignacio Varela[5], José Luis Sardina[6], Biola M. Javierre[3] and Maribel Parra[1,2,*]

[1]Lymphocyte Development and Disease Group, Josep Carreras Leukaemia Research Institute, 08916 Badalona, Spain, [2]Cellular Differentiation Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Av. Gran Via 199, 08908 L'Hospitalet, Barcelona, Spain, [3]3D Chromatin Organization Group, Josep Carreras Leukaemia Research Institute, 08916 Badalona, Spain, [4]Barcelona Supercomputing Center (BSC), Barcelona, Spain, [5]Instituto de Biomedicina y Biotecnología de Cantabria. Universidad de Cantabria-CSIC. 39011 Santander, Spain, [6]Epigenetic Control of Haematopoiesis Group, Josep Carreras Leukaemia Research Institute, 08916 Badalona, Spain, [7]Centre for Genomic Regulation (CRG), PRBB Building, Dr. Aiguader 88, 08003 Barcelona, Spain and [8]Universitat Pompeu Fabra, Barcelona, Spain

Received July 29, 2021; Revised June 23, 2022; Editorial Decision June 25, 2022; Accepted July 05, 2022

ABSTRACT

Correct B cell identity at each stage of cellular differentiation during B lymphocyte development is critically dependent on a tightly controlled epigenomic landscape. We previously identified HDAC7 as an essential regulator of early B cell development and its absence leads to a drastic block at the pro-B to pre-B cell transition. More recently, we demonstrated that HDAC7 loss in pro-B-ALL in infants associates with a worse prognosis. Here we delineate the molecular mechanisms by which HDAC7 modulates early B cell development. We find that HDAC7 deficiency drives global chromatin de-condensation, histone marks deposition and deregulates other epigenetic regulators and mobile elements. Specifically, the absence of HDAC7 induces TET2 expression, which promotes DNA 5-hydroxymethylation and chromatin de-condensation. HDAC7 deficiency also results in the aberrant expression of microRNAs and LINE-1 transposable elements. These findings shed light on the mechanisms by which HDAC7 loss or misregulation may lead to B cell–based hematological malignancies.

INTRODUCTION

A longstanding fundamental question in the field of cell development has been: how do cells decide at a molecular level to acquire a specific cell fate during tissue and organ generation? The mammalian hematopoietic system is considered a paradigm model for answering this question. For instance, B cell lymphopoiesis is a complex developmental process that comprises several cellular transitions, including cell commitment and early and late cellular differentiation. Proper transcriptional control at each cellular transition is essential for the correct generation of B lymphocytes. Of note, aberrant establishment of specific transcriptional programs may lead to the development of B cell malignancies.

Lineage-specific networks of transcription factors (TFs) have a central role in positively regulating the transition and maintenance of the distinct B cell developmental stages. In the bone marrow, lymphoid-primed multipotent progenitors (LMPPs) have the capacity and plasticity to become either common lymphoid progenitors (CLPs) or common myeloid progenitors (CMPs) (1). At that stage, the TFs IKAROS, MEF2C, and PU.1 are crucial for the early cellular choice towards the lymphoid lineage (2,3). Later on, commitment to the B cell lineage from CLPs to B cell progenitors (pro-B) and B cell precursors (pre-B) depends on the hierarchical and coordinated actions of the TFs PAX5, E2A, EBF and MEF2C (4–6). Classically thought as genuine activators for specific gene expression, B cell TFs are now believed to be involved in the repression of inappropriate lineage or of functionally undesirable gene transcrip-

*To whom correspondence should be addressed. Tel: +34 935572800; Fax: +34 934651472; Email: mparra@carrerasresearch.org
†The authors contributed equally to this article.
Present address: Maria Vila-Casadesús, Bioinformatics Platform, CIBER Hepatic and Digestive Diseases (CIBERehd), 08036 Barcelona, Catalonia, Spain.

tion, thereby ensuring that the proper B cell identity and differentiation are maintained. In addition to positively promoting B lymphocyte gene-specific programs, the PAX5, E2A, EBF1 and MEF2C TFs induce the repression of inappropriate genes of alternative lineages, thereby ensuring maintenance of proper B cell identity and differentiation (3,7–11). These findings support the concept of transcriptional repression as an essential mechanism for proper B lymphocyte generation. However, the identity of master transcriptional repressors essential for establishing and maintaining the genetic identity of B lymphocytes has remained elusive for many years.

Besides specific TF networks, B cell differentiation also requires that epigenetic regulators and architectural proteins establish the correct permissive/non-permissive chromatin structure (euchromatin/heterochromatin) (12). There is a close relationship between transcriptional regulators and dynamic changes in the DNA epigenetic landscape during B cell development, with DNA methylation levels and the chromatin conformation state dynamically changing at every differentiation cell stage, giving them a specific epigenetic signature (13–17). Also, TFs such as PAX5 interact with architectural proteins that mediate long-range chromatin interactions (18). In mammals, DNA demethylation depends on the action of the Ten-Eleven Translocation (TET) enzyme family of TET1, TET2 and TET3, which convert 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), leading to DNA demethylation and consequent gene expression (19). Notably, TET2 has been shown to play crucial roles during hematopoiesis (20–22). Although broadly expressed within the hematopoietic system, myeloid cells express higher levels of TET2 compared with lymphoid cell populations (23–25). To date, the molecular mechanisms controlling different TET2 physiological levels within the hematopoietic systems are largely unknown.

We have previously reported that HDAC7 is a master transcriptional repressor in early B cell development, controlling the expression of lineage-inappropriate genes and thus the identity of pro-B cells. A lack of HDAC7 in pro-B cells leads to a block in B cell differentiation, aberrant activation of alternative lineage genes, a reduction of proliferation, and an increase in apoptosis (26). More recently, we have identified HDAC7 as a prognostic factor and biomarker of survival in infants with pro-B acute lymphoblastic leukemia (pro-B-ALL) and MLL-AF4 rearrangement, who display a general loss in HDAC7 expression; notably, the lowest levels of HDAC7 are associated with the poorest outcome for the infants (27). We hypothesized that these findings could be indicative of a yet-unknown HDAC7-mediated molecular mechanisms that allows proper acquisition of cell identity during early B cell development in the bone marrow.

Using a combination of transcriptomic and epigenetic genome-wide analysis, we now shed light into the molecular mechanisms that are governed by HDAC7 during early B cell development. We identified HDAC7 as a regulator of proper chromatin compaction in different stages of B cell development (pro-B and pre-B cells). Importantly, we demonstrated that HDAC7 represses TET2 expression in pro-B and pre-B cells, and that its deficiency leads to TET2 up-regulation and subsequent alteration in global and specific 5-hmC patterns. In fact, HDAC7-deficient pro-B cells showed enhanced 5-hmC global levels, resulting not only in the activation of inappropriate lineage genes, but also in the aberrant expression of non-coding elements (such as active transposon LINE-1 elements and miRNAs). Thus, our findings unveil novel molecular mechanisms that govern the maintenance of correct B cell development and identity, working through the HDAC7–TET2 axis.

MATERIALS AND METHODS

Study Design

The study aimed to define unprecedented molecular mechanisms by which class IIa HDAC7 preserves B cell identity in mice. Experiments included 4–6 weeks-old wild-type and knockout mouse strains (C57BL/6). Mice selected in each experiment were littermates. Primary pro-B and pre-B lymphocytes were isolated by using cell sorting. Tet2 was identified as a direct target of HDAC7 with chromatin immunoprecipitation and expression analysis experiments. 5-hydroxymethylation levels in pro-B cells were quantified by ELISA and hydroxymethyl-DIP experiments. 5-hmeDIP-seq, ATAC-seq, H3K27ac and H3K27me3 ChIP-seq were performed to determine global and specific changes. All results were validated by qPCR assays and were successfully reproduced. No sample size calculations were performed, since these were selected on the basis of previous studies performed in our lab. The numbers of experimental replicates are included in the figure legends.

Mice

Hdac7$^{fl/-}$ on C57BL/6 mice have been previously described (28) and were kindly provided by Dr Eric Olson (UT Southwestern Medical Center, Dallas, TX, USA). Mb1-Cre$^{ki/+}$ on C57BL/6 (B6.C(Cg)-Cd79a$^{tm1(cre)Reth}$/EhobJ) mice were kindly provided by Dr Michael Reth (Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany). Experiments were performed with 4–6 week-old mice. Littermate controls were used for *Hdac7*$^{fl/-}$ mb1-cre mice. Animal housing and handling and all procedures involving mice, were approved by the Bellvitge Biomedical Research Institute (IDIBELL) ethics committee and the Animal Experimentation Ethics Committee (CEEA) of the Comparative Medicine and Bioimage Centre of Catalonia (CMCiB), at Germans Trias i Pujol Research Institute (IGTP), in accordance with Spanish national guidelines and regulations.

Cells

HAFTL pre-B cell line transduced with a MSCV-GFP-C/EBPα retroviral vector (to generate C10 cells) and with a MSCV-hCD4-C/EBPαER retroviral vector (to generate C11 cells), were described previously described (29). C10-MSCV and C10-HDAC7 cells were generated as previously described (30).

shRNA-Tet2 infection of primary B cells and transduced pre-B cell line (C11)

Retroviral vector for PGK-shRNATet2-GFP retroviral vector has been described in (31). $CD19^+$ B cells from $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ mice or C11 cells were infected with the shRNA Tet2 targeting vector(shTet2) or with an empty retroviral vector (shCtrl). Cells were infected twice, in a time gap of 24 h, and then, 48 h after second infection, GFP^+ cells were sorted using a FACSAria™ Fusion cell sorter (BD Biosciences). After isolation, $CD19^+$ B cells were cultured on RPMI media supplemented with 2% FBS, 0,03% Primatone RL (Sigma), 1 mM penicillin/streptomycin, 50 μM β-mercaptoethanol and 1% IL-7 (Peprotech), whereas C11 cells were cultured in RPMI media supplemented with 10% FBS and 1mM penicillin/streptomycin. *Tet2* knockdown was confirmed by qRT-PCR using SYBR green quantification.

β-Estradiol treatment of cell lines

C10-MSCV, C10-HDAC7, C11-shCtrl and C11-shTet2 cells were cultured and treated as previously described (30).

Flow cytometry analysis and cell-sorting

Cells were extracted from bone marrow (femur and tibia of both hind legs) of $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ mice. Red blood cells were lysed with ACK lysis buffer. Cell counts were determined using a manual cell counter and Türk's staining to facilitate the counting of white cells nuclei. Isolated cells were incubated with anti-CD16/CD32 (2.4G2, Fc Block) (BD Bioscience) for 10 min on ice to reduce non-specific staining. The following antibodies were used for analysis (from BD Biosciences): anti-B220 (RA3-6B2), anti-CD43 (S7), anti-IgM (R6-60.2), anti-CD19 (1D3) and anti-Cd11b (M1/70). For Cd11b, Streptavidin-V50 (560797, BD Biosciences) was used as a secondary antibody. Cells were stained with primary antibodies for 30 min on ice in the dark. Cells were analyzed on a BDFACS Canto II (BD Biosciences) or sorted on BD FACSAria™ Fusion cell sorter (BD Biosciences). Data were analyzed using FlowJo software (Tree Star, Inc.). See Supplementary Figure S1A for sorting gating strategy (pro-B cells: IgM^-, $CD19^+$, $B220^+$, $CD43^+$; pre-B cells: IgM^-, $CD19^+$, $B220^+$, $CD43^-$).

Magnetic cell separation (MACS)

Cells were extracted from bone marrow (femur and tibia of both hind legs) of $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ mice. Red blood cells were lysed with ACK lysis buffer. Cell counts were determined using a manual cell counter and Türk's staining to facilitate the counting of white cells nuclei. Isolated cells were incubated with anti-CD16/CD32 (2.4G2, Fc Block) (BD Bioscience) for 10 min on ice to reduce non-specific staining. The following antibodies were used for separation (from Miltenyi Biotec): anti-CD19-Microbeads (mouse), anti-Cd11b biotin (M1/70, BD Biosciences), and Streptavidin-Microbeads. Samples were incubated for 20 min at 4°C in the dark. CD11b needed double incubation, first with anti-Cd11b and second with Streptavidin-beads. Samples were put into Ls columns (Miltenyi Biotec) to perform magnetic cell separation. After three washes, cell from positive fractions (CD19+ and Cd11b+) were kept for further experimentation.

RNA-sequencing and analysis

Total RNA was extracted from HDAC7-deficient and control pro-B and pre-B cells in the Genomics facility of Institute for Research in Biomedicine (IRB) in Barcelona. Samples were quantified and subjected to quality control using a Bioanalyzer apparatus (IRB, Barcelona). Samples were processed at BGI Genomics Service, (China). Briefly, low input library was performed in all samples. Later, they were sequenced in paired-end mode with a read length of 100 bp. Thirty-five million paired-end reads were generated for each sample. Quality control of the samples was performed with the FastQC tool (available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired-end reads from RNA-Seq were aligned to the murine reference genome (GRCm38) using Hisat2 (version 2.0.5). Quality of the alignments was assessed using FastQC (v0.11.2). A count table file indicating the number of reads per gene in each sample was generated using HTSeq (version 0.6.0) (32). Genes with no or very low expression were filtered out and differentially expressed genes were identified using DESeq (33), requiring a minimum adjusted *P*-value of 0.05 and a $|\log_2 FC|$ value >1. Functional analysis was performed using gene set enrichment analysis (GSEA) (34) using a pre-ranked list of human orthologs genes and the gene set database c5.all.v7.2.symbols.gmt (Gene Ontology). GSEA analyses were performed from the $Hdac7^{+/-}$ pro-B versus $Hdac7^{+/-}$ pre-B cells, $Hdac7^{+/-}$ pro-B versus $Hdac7^{fl/-}$ pro-B cells and $Hdac7^{+/-}$ pro-B versus $Hdac7^{fl/-}$ pre-B cells comparisons. Genes were ranked using this formula: $-\log_{10}(\text{FDR}) \times \text{sign}[\log(\text{FC})]$. As genesets collection, hallmarks (H) from the Molecular Signatures Database (MSigDB) were selected, adding the specified custom genesets. Data from RNA-seq is available under accession code GSE171855.

RT-qPCR analysis

RNA from sorted pro-B and pre-B cells was extracted with an RNeasy Mini kit (Qiagen) and subsequently converted into cDNA using the High Capacity cDNA Reverse Transcription Kit (AB Applied Biosystems) according to the manufacturer's instructions. Real-time-quantitative PCR (RT-qPCR) was performed in triplicate using SYBR Green I Master (Roche). PCR reactions were run and analyzed using the LightCycler 480 Detection System (Roche). RT-qPCR primer pairs are shown in supplemental information Supplementary Table S1.

Micrococcal nuclease assay

Two million of B cells from $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ mice were resuspended in 500 μl of lysis buffer (10 mM Tris, 10 mM NaCl, 3 mM $MgCl_2$, 1% Triton, pH 7.5) and incubated on ice for 10 min. Nuclei were collected by centrifugation at 300 g for 5 min at 4°C. The nuclear pellet was then

resuspended in 400 μl of nuclear lysis buffer (20 mM Tris, 20 mM KCl, 70 mM NaCl, 3 mM $CaCl_2$ and protease inhibitors, pH 7.5). Aliquots of 100 μl were incubated with 9 units of micrococcal nuclease (MNAse, ThermoFisher) and digested at room temperature for 0, 1, 2 and 5 min, respectively. Then 3μl of 0.5M EDTA was added to stop digestion, and DNA was purified by using the QIAquick PCR purification kit (Qiagen). About 500 ng were used for gel electrophoresis and 12 ng of DNA were used for qPCR analysis using SybrGreen (Roche). Primers obtained from (35) were designed to target the β-globin gene to obtain PCR amplicons longer than the length of a single nucleosome. Mononucleosomes generated during MNAse digestion cannot be amplified by qPCR. Then, reduced amplification involves more open/accessible chromatin state, at least at this locus.

Western Blot

White cells from bone marrow of control and *Hdac7*$^{fl/-}$ mice were extracted as in sorting procedure. Next, cells were stained with anti-CD19-microBeads (Miltenyi Biotec) according to manufacturer's instructions. CD19^{+} B cells were isolated by magnetic separation with LS column adapters (Miltenyi Biotec). Purified cells were lysed with RIPA buffer. Lysates were resolved on 8–15% SDS-PAGE (Mini-Protean electrophoresis chamber, Bio-Rad) and transferred on nitrocellulose membranes (Amersham Biosciecnes). Membranes were blocked in 5% milk in TBS with 0.1% Tween (TBS-T) and incubated overnight at 4°C, with primary antibodies (anti-Tet2 ab94580, abcam 1:1000; anti-HDAC7 sc-11421, Santa Cruz Biotechnology 1:1000; anti-H3K9me3 ab8898 (Abcam) 1:1000; anti-PUMA 12450 (Cell Signaling) 1:500, anti-IRF4 sc-48338 (Santa Cruz Biotechnology) 1:500; anti-c-MYB sc-74512 (Santa Cruz Biotechnology) 1:500; anti-H3 ab1791 (Abcam) 1:1000; anti-Lamin B1 ab16048 (Abcam) 1:1000, and anti-Actin AC-15 (Sigma-Aldrich) 1:40000). Secondary antibody incubations (HRP-anti mouse, P0260, or anti rabbit, P0448, Dako 1:3000), were carried out for 1h at room temperature. Protein signal was detected using ECL western detection kit (Amersham Biosciences).

Chromatin immunoprecipitation assays (ChIP)

For chromatin immunoprecipitation (ChIP) assays, purified pro-B cells from the bone marrow of *Hdac7*$^{+/-}$ and *Hdac7*$^{fl/-}$ mice were crosslinked for 15 min in 1% formaldehyde, followed by inactivation in 125 mM glycine for 5 min and by two washes in cold PBS. Afterward, samples were incubated in cell lysis buffer from LowCell# ChIP kit (Diagenode) for 30 min at 8°C and sonicated with M220-Focused Ultra Sonicator (Covaris) according to manufacturer's instructions. Next steps of ChIP experiments were performed using resources from the LowCell# ChIP kit (Diagenode) according to the manufacturer's instructions. The following antibodies were used for immunoprecipitation: anti-HDAC7 (Abcam, HDAC7-97, 2.5 μg), anti-H3K9/K14ac (Millipore, 06-599 2.5 μg), anti-H3K27me3 (Millipore, 07-449, 2.5 μg), anti-H3K27Ac (Abcam, ab4729, 2.5 μg) and anti-H3k9me3 (Abcam, ab8898, 2.5 μg). Real-time quantitative PCR (RT-qPCR) was performed in triplicate and the results analyzed. Data are presented as the ratio between the HDAC7-bound fraction and histone modification antibody relative to the input control. ChIP-qPCR primer pairs are shown in supplemental information Supplementary Table S1.

ChIP-seq experiments and analysis

H3K9/K14ac ChIP-seq data were extracted from ChIP-seq experiments, as described elsewhere (26), whose data are available under the accession code: SRA submission SUB1614653. For new ChIP-seq experiments, purified pro-B and pre-B cells from the bone marrow of *Hdac7*$^{+/-}$ and *Hdac7*$^{fl/-}$ mice were crosslinked for 15 min in 1% formaldehyde, followed by inactivation in 125 mM glycine for 5 min and by two washes in cold PBS. Afterward, samples were lysed and sonicated with M220-Focused Ultra Sonicator (Covaris) to obtain fragments of 250–500 bp. Samples were processed according to Blueprint Histone ChIP-Seq protocol (https://www.blueprint-epigenome.eu/). The following antibodies were used for immunoprecipitation: 2.5 μg of anti-H3K27me3 (07449) and 2.5 μg of anti-H3K27ac (ab4729). As experimental control we used input sonicated chromatin (not immunoprecipitated) in all experimental conditions. For the analysis, reads were checked for quality using FastQC (0.11.5) and then trimmed using trim galore (v.0.6.6) to remove the sequencing adapters. Reads were aligned to the mouse reference genome GRCm38 using Bowtie v2.3.2 with '–very-sensitive' parameters (36). Aligned reads were then filtered based on ENCODE standards and removed those mapping to the blacklist and duplicates using samtools (v.1.9) and sambamba (v.0.7.0). Peaks were called using MACS2 v2.2.7.1 (37) with default parameters, providing an input sample to avoid false positives. Background correction was applied by first defining a set of non-redundant enriched regions for all samples by taking the union of all peaks from both replicates of all samples. Bigwig files were generated using bamCoverage v3.2.1 from deeptools (38). Genomic peak annotation was performed with HOMER software (v4.11). Intensity plots were performed using computeMatrix in a window of ±1 kb center in the TSS from deeptools (38). Data from H3K27ac and H3K27me3 ChIP experiments are available under accession code: GSE204673.

ATAC-seq experiment and analysis

50 000 purified pro-B and pre-B cells from the bone marrow of *Hdac7*$^{+/-}$ and *Hdac7*$^{fl/-}$ mice were isolated and freshly lysed using cold lysis buffer (10 mM Tris−HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500 g for 10 min using a refrigerated centrifuge, and pellet was resuspended in the transposase reaction mix (12.5 μl 2 × TD buffer, 2 μl transposase (Illumina) and 5.5 μl nuclease-free water). The transposition reaction was carried out for 1 h at 37°C, followed by addition of clean up buffer (900 mM NaCl, 300 mM EDTA, 2 μl 5%SDS, 20 ng Proteinase K) and incubation for 30 min at 40°C. Tagmented DNA was isolated using 2× SPRI beads from Beckman–Coulter. Following purification, we amplified library fragments using

1× NEBnext PCR master mix and 1.25 μM of Nextera PCR primers as described elsewhere (39). Sequence reads quality was assessed with MultiQC v1.12 (40), and adapter content (if any) was trimmed using Trimmomatic v0.39 (41). Paired-end reads were aligned to GRCm38 using Bowtie2 v2.2.3 with a maximum insert size of 2000 (-X 2000) (36). These parameters ensured that fragments up to 2 kb were allowed to align and that only unique aligning reads were collected (-m1). For all data files, duplicates and mtDNA were removed using picard (http://picard.sourceforge.net) and sambamba (v.0.7.0), respectively. For peak-calling, MACS2 v2.2.7.1 (37) tool was used with default parameters. Background correction was applied by first defining a set of non-redundant enriched regions for all samples by taking the union of all peak summits from both replicates of all samples. We then quantified the signal at all summits in each sample by counting the number of fragments (using the R bioconductor package csaw, v. 1.0.7) (42). The resulting counts matrix file was analyzed for differential peaks with DESeq2 (43). Genomic peak annotation was performed with HOMER software (v4.11). Intensity plots were performed using computeMatrix in a window of ±1 kb center in the TSS from deepTools (38). Data from ATAC-seq are available under accession code GSE204672.

Quantification of global 5-hydroxymethylation levels

To quantify 5-hmC, a Quest 5hmC DNA ELISA kit (Zymo Research) was used according to the manufacturer's protocol. First, genomic DNA from sorted cells was extracted using Quick-DNA Miniprep Plus kit (D4068, Zymo Reseach). Next, the bottom of the provided well was coated with anti-5-hmC polyclonal antibody (pAb) for 1 h at 37°C in the dark. Wells were then blocked and 100 ng of denatured genomic DNA was added for 30 min at 37°C in the dark. After corresponding washes, anti-DNA HRP antibody was applied to wells for 30 min at 37°C in the dark. After corresponding washes, HRP developer (3,3',5,5'-tetramethylbenzidine (TMB) (Sigma-Aldrich)) was added to detect the DNA bound to the anti-5-hmC pAb for 20–30 min at room temperature in the dark. Afterward, the color reaction was stopped by the addition of sulfuric acid and the resulting color was analyzed at 450 nm by using a Glomax microplate reader (Promega). The percentage of 5-hmC DNA was estimated from linear regression.

hMeDIP-qPCR experiments

Genomic DNA was purified by using the same kit as in ELISA assay. 1 μg of genomic DNA from wild-type and HDAC7-deficient sorted pro-B cells was sonicated with M220-Focused Ultra Sonicator (Covaris) to obtain fragments of 300–400 bp. Fragmented DNA were incubated with 2 μg anti-5hmC (Active Motif, 39769) and 20 μl of Dynabeads G (Life Technologies) for 16 h at 4°C, and 10% of DNA was kept to be used as input. After incubation, Dynabeads were washed 3 times with IP buffer (10 mM Na-Phosphate pH 7, 0.14 M NaCl, 0.05% Triton X-100) and then were resuspended in Proteinase K digestion buffer (50 mM Tris pH8, 10 mM EDTA, 0.5% SDS) for 30 min at 55°C. DNA from immunocomplexes was purified with the QIAquick MinElute kit (Qiagen). Real-time quantitative PCR was performed and the results analyzed. Data are presented as the ratio of the enrichment of 5-hmC relative to the input control.

hMeDIP sequencing experiments and analysis

Purified genomic DNA (1 μg) from wild-type and HDAC7-deficient pro-B cells was sonicated to obtain fragments of 300–400 bp. Adaptor ligations were performed and libraries constructed by qGenomics Laboratories (Barcelona, Spain). 2 μg of anti-5hmC (Active Motif, 39769) were incubated with 20 μl of Dynabeads G (Life Technologies) for 2 h at 4°C. Fragmented DNA was incubated with Dynabeads and antibody for 16 h at 4°C, and 10% of DNA was kept to be used as input. DNA was purified as described in hMeDIP qPCR assay. Amplified libraries were constructed and sequenced at qGenomics Laboratories (Barcelona, Spain). Fastq data were obtained with Trim Galore-0.4.2 and Cutadapt-1.6. Reads were mapped with bwa-0.7.12. Sorting Sam to Bam was carried out with Picard-2.8.0 SortSam and duplicates were removed with Picard-2.8.0 MarkDuplicates. Bigwig files were made with deeptools and normalized based on RPKM. Peak calling was performed using MACS2 bdgpeakcall option (-c 250 -l 10 -g 10). To avoid false positives, peaks that intersect with peaks in the corresponding input samples were removed. DESEQ analysis (DESeq2 v1.20.0) was then used to define peaks and perform quantitative analyses. The Diffbind-2.6.6 R package was used for differential binding analysis. Differential enrichment was defined by a threshold value of $P = 0.005$ and a >1-fold difference in KO relative to WT samples. Motif enrichment was analyzed and peak depth quantified with HOMER software (v4.10). Data from hMeDIP-seq is available under accession code GSE135263.

Co-immunoprecipitation (Co-IP)

Co-IP assays were performed using $CD19^+$ cells from control and HDAC7 deficient mice. Cell extracts were prepared in lysis buffer [50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1% Triton-X-100, protease inhibitor cocktail (complete™, Merck)] with corresponding units of Benzonase (Sigma) and incubated at 4°C for 6 h. 50 μl of supernatant was saved as input and diluted with 2× Laemmli sample buffer (4% SDS, 20% glycerol, 120 mM Tris–HCl, pH 6.8). Supernatant was first incubated with PureProteome™ Protein A/G agarose suspension (Merck Millipore) for 1 h to remove background signal. Samples were then incubated overnight at 4°C with corresponding antibodies against TET2 (ab124297, Abcam) and rabbit (12-371, Merck Millipore) IgGs (negative control) and A agarose beads. After that, beads were washed three times with lysis buffer. For sample elution, 100 μl of 1× Laemmli sample buffer was added to the beads. Samples and inputs were denatured at 95°C in the presence of 1% β-mercaptoethanol.

Expression profiling of microRNAs

We used miRCURY LNA™ Universal RT microRNA PCR System (Exiqon) to determine miRNA expression profiles.

The miRNA annotation of mirBase version 20.0 was used. Single-stranded cDNA was synthesized by reverse transcription of 8 μl of RNA, using the universal cDNA Synthesis Kit II (Exiqon). Diluted cDNA was mixed with ExiLENT SYBR® Green master mix (Exiqon), and quantitative PCR was performed using the Roche LightCycler® 480 RealTime PCR system (Roche). Primers design for validations of miRNA's differential expression was performed with miRprimer software (https://sourceforge.net/projects/mirprimer/).

Immgen data analysis

Immgen is a public resource that is the result of a collaborative group of immunology and computational biology laboratories that share knowledge and expertise to perform a broad and deep dissection of the genome's activity and its regulation in the immune system. This public resource is broadly used by many immunology laboratories to interrogate biological questions of a gene of interest or more broad mechanistic insights within the hematopoietic system.

Gene Skyline tool: expression data from microarrays collected from several participating laboratories is normalized with Robust Multiarray Average (RMA) algorithm (44,45). RMA is a three-step algorithm, which removes background noise (normally distributed), performs quantile normalization and log2 transformation. Raw reads from microarray are normalized by median of ratios with DeSeq2 package from Bioconductor. Spike-in data is also used to evaluate the RMA method in order to address questions about the correctness of microarray data. Immgen developers ensure that the comparison of expression levels of one gene among different cell types is sensitive and quantitative.

Modules and regulators tool: this tool uses an algorithm called Ontogenet in order to outline the regulatory networks that drive the hematopoietic cell differentiation (see detailed explanation in (46). Briefly, it classifies and clusters genes in groups called 'modules' based on expression levels of genes among all cell types from all lineages and developmental stages from the hematopoietic system given by several microarray analyses. Each module is called 'coarse module', which can be further divided into 'fine modules'. The algorithm assumes that genes included in one module are regulated in the same way. For these modules, Immgen create automatically an expression matrix represented as a heatmap, showing potential regulators of the module and their expression pattern among blood cell types. Immgen also provides a table of expression level and regulatory activity (weight), classifying each regulator into positive and negative.

My GeneSet tool: this browser allows you to examine the pattern of expression of a selected set of genes over some or all of the ImmGen expression data (obtained from either RNA-seq or microarray experiments).

Statistical analysis

Data were analyzed by Student's two-tailed unpaired *t*-test and Mann–Whitney test using GraphPad Prism (v7). *P*-values lower than 0.05 were considered statistically significant. Statistical methods for analysis of genome wide datasets involving hMeDIP-seq, RNA-seq, ChIP-seq and ATAC-seq are explained in detail under the respective section.

RESULTS

HDAC7 regulates chromatin condensation and 5-hmC levels in pro-B cells

We previously identified HDAC7 as a critical transcriptional repressor of lineage-inappropriate genes that ensures correct early B cell development. However, its exact molecular mechanisms remained to be addressed (26,30). To unveil the mechanisms and functions that HDAC7 exerts during early B lymphocyte development along the different cell stages of differentiation, we performed RNA-sequencing (RNA-seq) experiments on pro-B and pre-B cells from HDAC7 conditional knockout mice (*Hdac7*$^{fl/-}$; hereafter, HDAC7-deficient) and their control littermates (*Hdac7*$^{+/-}$; hereafter, wild-type) (Figure 1A; see also gating strategy, Supplementary Figure S1A). First, we performed an unsupervised cluster analysis of the samples based on their gene expression profiles. Notably, pro-B and pre-B wild-type cell populations were grouped into two different clusters, confirming that early B cell differentiation steps were affected by gene transcription changes, a hallmark of early B lymphocyte development. In contrast, HDAC7-deficient pro-B and pre-B cells were more similar and grouped closer, indicating that the B cell developmental block observed upon HDAC7 deficiency is a consequence of altered gene transcription programs (Figure 1B). Based on RNA-seq analyses, a comparison of the differentially expressed genes between the cell populations showed that 1992 and 2140 genes were up-regulated and down-regulated, respectively, during the cellular transition from pro-B to pre-B wild-type cells (Supplementary Figure S1B). The lack of HDAC7 from the pro-B or pre-B cells resulted in a downregulation of 272 and 1576 genes, respectively, as compared to the wild-type pro-B cells (Supplementary Figure S1B). Among the downregulated genes, we observed a dramatic disruption in heavy chain immunoglobulin (Igh) production (Supplementary Figure S1C), which is a critical process in early B cell development; this demonstrated that HDAC7 deficiency would drastically affect the immune response (Supplementary Figure S1D).

HDAC7 is a transcriptional repressor. Accordingly, 393 and 1609 genes in pro-B and pre-B cells, respectively, were upregulated in the HDAC7-deficient cells as compared to pro-B wild-type cells (Figure 1C). Gene Ontology (GO) analysis of up-regulated genes at both cell stages showed that they were related, on one hand, to cell proliferation, survival and immune processes (Supplementary Figure S1E) and, on the other hand, to enhanced protein binding (Supplementary Figure S1F). To specifically analyze the gene expression changes exclusively depending on HDAC7, and omit differences due to different cell stages, we overlapped upregulated genes from pro-B and pre-B cells. GO analysis revealed that overlapped upregulated genes upon HDAC7 deficiency were associated with an altered immune response (Figure 1D, upper panel), with a clear tendency for increased protein binding and transcription factor (TF) activity, which may correlate with a more open chromatin

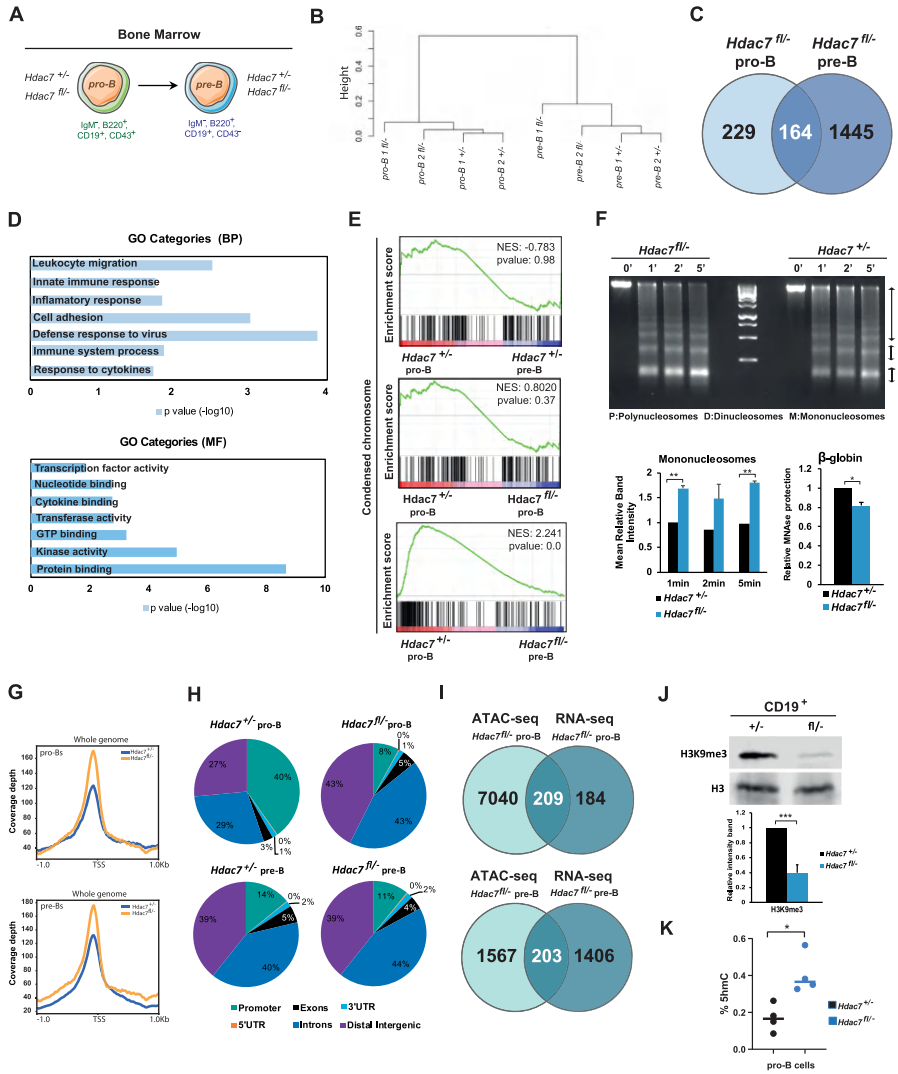


Figure 1. HDAC7 deficiency results in global chromatin de-condensation and increase 5-hmC levels. (**A**) Experimental design for RNA-seq experiments, showing the type of cells used. B cell progenitors (pro-B); B cell precursors (pre-B). (**B**) Cluster dendogram obtained from RNA-seq gene expression profiles for each replicate and cell type. (**C**) Venn Diagram comparing up-regulated genes in $Hdac7^{fl/-}$ pro-B cells respect to control pro-B cells (left) and in $Hdac7^{fl/-}$ pre-B cells respect to control pro-B cells (right). (**D**) Gene ontology (GO) enrichment analysis of up-regulated genes in HDAC7 deficient pro-B and pre-B cells respect to control pro-B cells. BP refers to Biological Processes and MF corresponds to Molecular Functions. (**E**) GSEA analysis comparing wild-type pre-B and HDAC7-deficient pro-B and pre-B cells respect to wild-type pro-B cells in expression profiles related to condensed chromosome signatures. (**F**) Chromatin accessibility assay in control and HDAC7-deficient B cells. Isolated nuclei from these cells were digested with 5 units of micrococcal nuclease (MNase) for 0, 1, 2 or 5 min. Different nucleosomal fractions (mono, di and polynucleosomes) were separated by gel electrophoresis (upper panel). Band density quantification from chromatin accessibility assay using ImageJ software. Results are expressed as the mean ± SEM of three independent experiments. $^{**}P < 0.01$ by unpaired *t*-test (left lower panel). qPCR showing protection of MNase-digested DNA after 5 min incubation in control or HDAC7-deficient B cells (right lower panel). Data represent mean ± SEM of three independent experiments per condition after normalization to *Actb* gene expression. $^{*}P < 0.05$ by unpaired T-test. (**G**) ATAC-seq coverage depth (per base pair per peak per 10 million mapped reads) of peaks located in the TSS (−1 kb to +1 kb) in wild-type and HDAC7-deficient pro-B (upper panel) and pre-B cells (lower panel). (**H**) Genomic distribution of chromatin accessible peaks in $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ pro-B and pre-B cells determined by ATAC-seq experiments. (**I**) Venn Diagram comparing differentially enriched peaks in ATAC-seq in $Hdac7^{fl/-}$ pro-B cells and up-regulated genes in RNA-seq in $Hdac7^{fl/-}$ pro-B cells (upper panel) and pre-B cells (lower panel) compared to control cells. (**J**) H3K9me3 global levels decrease upon HDAC7 deficiency, analyzed by Western blot representative assay (upper panel). Total H3 levels were used as a control. Band quantification of the Western blot experiment was performed using ImageJ program. Results are expressed as the mean ± SEM of three independent experiments. $^{***}P < 0.001$ by unpaired *t*-test (lower panel). (**K**) ELISA assay showing global levels of DNA 5-hmC in pro-B cells from wild-type and HDAC7-deficient mice. Each dot represents one animal ($n = 4$, test); the line represents the mean. Data is represented as mean ± SEM of four independent experiments. $^{*}P < 0.05$, Mann–Whitney test.

state (Figure 1D, lower panel). Gene set enrichment analysis (GSEA) showed that the gene signatures in HDAC7-deficient pro-B or pre-B cells were associated to a deficiency in chromosome condensation, suggesting a potential loss of the heterochromatin state in the absence of HDAC7 (Figure 1E). This de-condensation was independent of the cell stage, as there was no difference in these gene signatures between control pro-B and pre-B cells, when separately analyzed for each genotype (Figure 1E). These results indicate that HDAC7 deficiency may result in chromatin decompaction. Accordingly, HDAC7-deficient B cells presented increased chromatin accessibility, as genomic DNA from these cells was more sensitive to micrococcal nuclease (MNase) digestion than that from control B cells (Figure 1F, upper and lower left panels). These results were reinforced by qPCR showing protection of MNase-digested DNA at β-globin site (Figure 1F, lower right panel). To definitively prove the potential role of HDAC7 in maintaining proper chromatin status in pro-B and pre-B cells, we performed transposase-accessible chromatin assay coupled with next generation sequencing (ATAC-seq). We found that the global chromatin accessibility intensity was higher in both HDAC7-deficient pro-B and pre-B cells compared to control cells (Figure 1G). Consistent with its role as a transcriptional repressor, we observed ∼13 000 more accessible peaks in HDAC7-deficient pro-B cells. Next, we determined ATAC-seq differential regions between HDAC7-deficient and wild-type pro-B (Figure 1H, upper panel) and pre-B cells (Figure 1H, lower panel). We found that the genomic distribution of the accessible sites differed significantly in the absence of HDAC7. In particular, we observed a reduction of peaks located at promoters and, more notably in pro-B cells, a dramatic increase of accessible peaks at distal intergenic regions (Figure 1H). To interrogate a potential functional link between the altered chromatin landscape and transcriptional programming in the absence of HDAC7, we integrated the data obtained in ATAC-seq and RNA-seq experiments. We compared the open chromatin regions in HDAC7-deficient pro-B and pre-B cells versus upregulated genes in the absence of HDAC7 (Figure 1I, upper panel). Half of the differentially upregulated genes (209 genes) in pro-B cells also presented a more open chromatin state in HDAC7-deficient cells, whereas 203 genes among the pre-B cells upregulated genes set showed a more accessible chromatin (Figure 1I, lower panel). Additionally, and given that chromatin condensation is associated to heterochromatin state, we next isolated $CD19^{+}$ B cells from HDAC7-deficient or wild-type mice and assessed the levels of H3K9me3, a well-known epigenetic mark involved in the establishment of heterochromatin, genome stability, and cell identity maintenance (47). We found that HDAC7-deficient B cells showed a significant and dramatic decrease in H3K9me3 as compared to their wild-type counterparts (Figure 1J). Increases in 5-hmC have been associated with chromatin decompaction and, consequently, with alterations in hematopoietic cell differentiation and the maintenance of cell identity (48). Importantly, 5-hmC and H3K9me3 have an opposite genomic localization pattern, with TET proteins recruited mostly to euchromatin regions (49,50). We next carried out ELISA assays to determine the global 5-hmC levels; which revealed that HDAC7 loss in pro-B cells led to a significant increase in the global levels of 5-hmC (Figure 1K). Thus, altogether our data indicated that HDAC7 controls chromatin compaction and DNA 5-hmC levels at early stages of B cell development.

HDAC7 regulates TET2 expression in pro-B and pre-B cells

Our results suggested that an essential function of HDAC7 during early B cell development could be the regulation of TET proteins expression, which are responsible for incorporating 5-hmC into DNA. Among the different family members, TET2 appears to play crucial roles during hematopoiesis and is highly expressed in myeloid cells (20–22). Therefore, we hypothesized that HDAC7 may be responsible for tightly regulating and fine-tuning the physiological levels of TET2 in pro-B and pre-B cells, thereby ensuring proper cell differentiation and identity. We directly addressed our hypothesis firstly by analyzing whether TET2 is expressed at different levels in the lymphoid and myeloid compartments within the hematopoietic system. For this purpose, we examined transcriptomic data from the Immunological Genome Project Database (Immgen) (http://www.immgen.org/). Indeed, TET2 was expressed at much higher levels in myeloid cells than in lymphoid populations (Figure 2A). Next, taking advantage of the 'modules and regulators' interactive tool in the Immgen database, we searched for putative positive and negative TET2 regulators. This tool uses all expression data of hematopoietic cell populations deposited by several groups and enables the search for putative positive and negative regulators of the gene of interest. Strikingly, HDAC7 appeared to be the only negative regulator or transcriptional repressor controlling TET2 expression within the hematopoietic system (Figure 2B). Further analysis of data from Immgen database to compare the expression pattern between HDAC7 and TET enzymes coding genes in lymphoid and myeloid populations revealed an inverse correlation with gene expression between HDAC7 and TET2 (Figure 2C), indicating a fully opposite expression pattern of both genes throughout hematopoiesis. In contrast, TET1 had an expression pattern similar to that of HDAC7, while TET3 displayed uniform expression levels throughout all tested cell populations (Supplementary Figure S2A and B). These data suggested that HDAC7 was responsible for controlling TET2 levels in lymphoid cells. We next examined data from RNA-seq in Figure 1B–E, as well as from our published microarray (26). We found that *Tet2* was upregulated in HDAC7-deficient pro-B and pre-B cells with respect to control pro-B cells (Figure 2D and E) reaching similar levels to that of macrophages (Supplementary Figure S2C), whereas the expression of the essential B cell genes *E2a* and *Pax5* remained unaltered (Supplementary Figure S2D and E). RT-qPCR assays revealed that the absence of HDAC7 from pro-B and pre-B cells did not alter the expression levels of the other class IIa HDACs family members (HDAC4, 5 and 9) indicating the specific requirement of HDAC7 during early B cell development (Supplementary Figure S2F). According to mRNA levels, analysis of our ATAC-seq revealed a more open chromatin at *Tet2* gene loci in HDAC7-deficient pro-B cells and pre-B cells (Figure 2F). Western blot experiments confirmed that the absence of HDAC7 from B cells

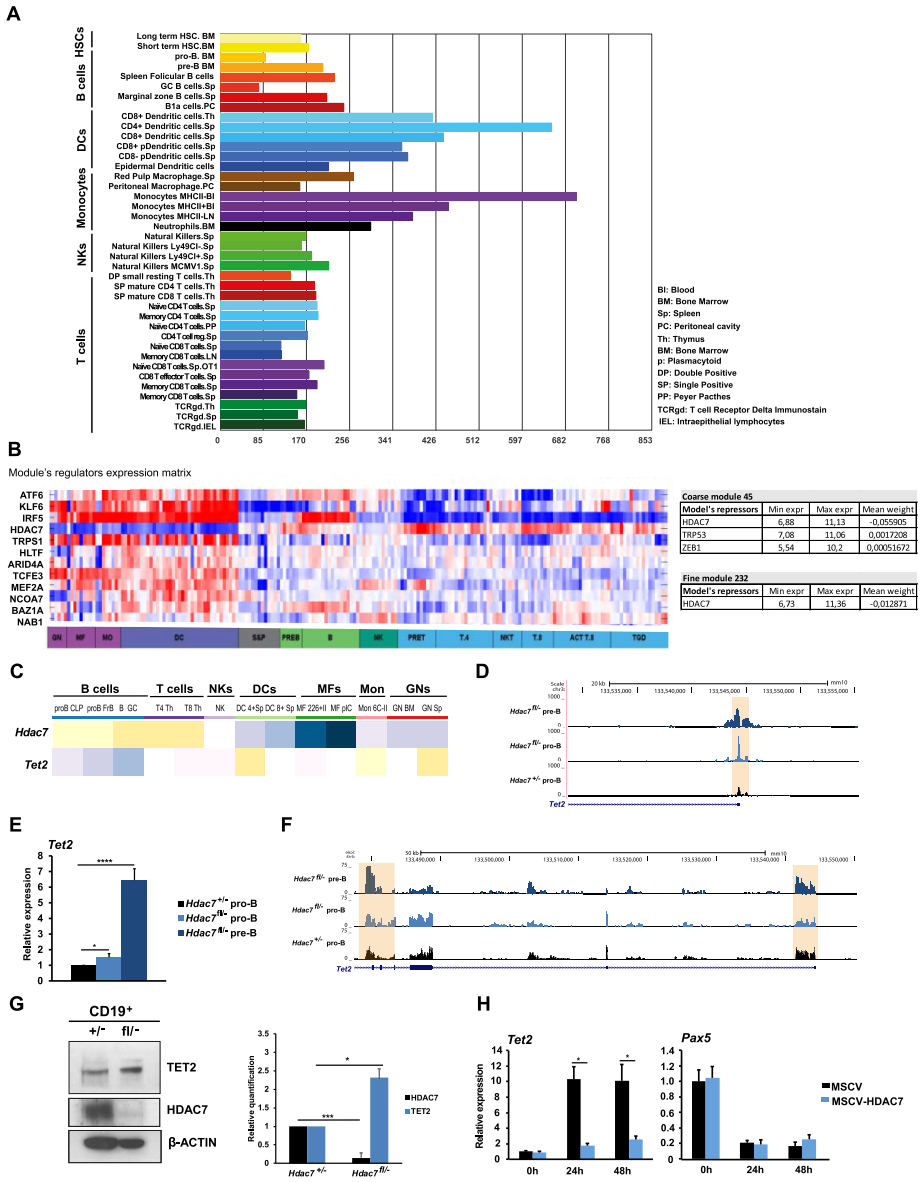


Figure 2. (A) HDAC7 controls DNA 5-hmC through *Tet2* regulation in pro-B and pre-B cells. Gene skyline showing the *Tet2* gene expression profile along different murine hematopoietic cell lineages (obtained from 'gene skyline' tool of Immgen database). (**B**) Heatmap showing the expression of potential *Tet2* gene regulators in hematopoietic cell subsets (left panel). Tables showing mean weight (enzymatic activity) of potential negative regulators of *Tet2* gene in hematopoietic cells. *Tet2* gene is included in two modules or cluster of genes: coarse module includes 315 genes and three negative regulators and fine module contains 35 genes and one negative regulator (right panel). Data were extracted from Immgen public database using *Modules and Regulators* tool. (**C**) Heatmap showing the inverse expression of *Hdac7* and *Tet2* in the indicated cell subsets. Data was extracted from Immgen public database using *My Gene Set* tool. (**D**) Genome browser snapshot of the *Tet2* gene showing signal for RNA-seq in wild-type and HDAC7-deficient pro-B and pre-B cells. Orange shadow indicates *Tet2* promoter. (**E**) Analysis by RT-qPCR of the *Tet2* mRNA levels in wild-type and HDAC7-deficient pro-B and pre-B cells. (**F**) Genome browser snapshot of the *Tet2* gene showing signal for ATAC-seq in $Hdac7^{+/-}$ pro-B cells, $Hdac7^{fl/-}$ pro-B and pre-B cells. Orange shadows indicate *Tet2* promoter and additional exonic regions. (**G**) TET2 upregulation in $Hdac7^{fl/-}$ $CD19^{+}$ B cells compared to $Hdac7^{+/-}$ B cells analyzed by Western Blot assay (left panel). Quantification of two Western Blot experiments was performed by using ImageJ program (right panel). (**H**) RT-qPCR experiments for gene expression changes for *Tet2* and *Pax5* in the absence or presence of HDAC7 during the conversion of pre-B cells into macrophage-like cells. 24 and 48 h time points indicate time since transdifferentiation induction with β-estradiol. Data from E and H is represented as mean ± SEM of four independent experiments. $*P < 0.05$, $****P < 0.0001$, unpaired *t*-test. Data from G is represented as mean ± SEM of two independent experiments. $*$ $P < 0.05$, unpaired *t*-test.

corresponded with significantly increased TET2 protein levels (Figure 2G). Therefore, different physiological levels of TET2 protein may have an effect and specific function in hematopoietic cell populations. Next, using a transdifferentiation system developed by Graf laboratory (24), we found that HDAC7 exogenous expression blocked the upregulation of *Tet2* during pre-B cells conversion into macrophages (Figure 2H). Forced expression of HDAC7 did not interfere with the downregulation of *Pax5* during cellular conversion (Figure 2H), confirming that its expression is totally independent of that of HDAC7. Together, these findings indicated that HDAC7 may govern early B cell development in the bone marrow through the control of the physiological levels of TET2.

Tet2 is a direct HDAC7 target gene in pro-B cells

We next performed chromatin immunoprecipitation (ChIP) assays to test whether *Tet2* is a direct HDAC7 target gene in B cells; we found that HDAC7 was recruited to the promoter of the *Tet2* gene, which contains MEF2 binding sites in pro-B cells (Figure 3A, upper panel). This was consistent with the previously demonstrated requirement of HDAC7 to interact with MEF2C to repress its target genes in pro-B cells (26), and with *Tet2* being a MEF2C direct target in pro-B cells (6). We also demonstrated HDAC7 recruitment at a previously described *Tet2* enhancer (24) (Figure 3A, lower panel). We previously reported by ChIP-seq experiments that HDAC7 loss in pro-B cells causes an increase of H3K9/K14ac at the promoter and enhancers of its target genes (26). ChIP-seq data examination and ChIP-qPCR revealed an increase of H3K9/K14ac at both promoter and enhancer regions of *Tet2* in the absence of HDAC7 from pro-B cells (Figure 3B and C). As expected, the absence of HDAC7 from pro-B cells had no effect on H3K9/K14ac enrichment at *Pax5* gene (Supplementary Figure S3A). Next, integration of ATAC-seq and H3K9/K14ac ChIP-seq data from HDAC7-deficient pro-B cells revealed an overlapping of 4487 enriched regions, remarkably including *Tet2* (Supplementary Figure S3B). To further gain insight into the mechanisms altered by HDAC7 deficiency, we performed ChIP-seq of H3K27ac and H3K27me3 histone marks in wild-type and HDAC7-deficient pro-B cells. The analysis revealed an increased global intensity of the activating histone mark H3K27ac in the HDAC7-deficient pro-B cells (Supplementary Figure S3C), as well as at both specific promoter and enhancer regions of *Tet2* gene (Figure 3D, E). In addition, the absence of HDAC7 from pro-B cells resulted in a decrease of global H3K27ac mark in promoter regions and an increase in distal intergenic regions (Supplementary Figure S3D). Increased number of H3K27ac peaks in distal regions upon HDAC7 deficiency agrees with previous results of more open chromatin regions under the same conditions, reinforcing the repressive role of HDAC7. The integration of ATAC-seq and H3K27ac-seq data revealed 3113 common peaks (representing ∼40% of the enriched regions in both assays) in the absence of HDAC7, in which *Tet2* was also included (Supplementary Figure S3E). In parallel, H3K27me3 ChIP-seq data unveiled similar coverage intensity between wild-type and HDAC7 deficient pro-B cells (Supplementary Figure S3F). The genomic distribution of H3K27me3 remained also unaltered in the absence of HDAC7 (Supplementary Figure S3G). In the case of H3K27me3 enrichment in HDAC7-deficient pro-B cells, we performed the integrative analysis comparing with ATAC-seq data from wild-type cells and observed an overlap of ∼25% of enriched peaks (Supplementary Figure S3H). ChIP-qPCR demonstrated a significant decrease of H3K27me3 at *Tet2* gene loci in HDAC7-deficient pro-B cells (Figure 3F, G). Finally, we aimed to determine if the global decrease in H3K9me3 in the absence of HDAC7 shown in Figure 1J could be observed at the *Tet2* gene loci. Indeed, ChIP-qPCR showed that HDAC7 deficiency in pro-B cells led to lower enrichment levels of the repressive histone mark H3K9me3, at both the promoter and enhancer of *Tet2* gene (Supplementary Figure S3I). Collectively, our findings demonstrated that HDAC7 controls the proper deposition of epigenetic marks and that *Tet2* is a direct HDAC7-target gene in pro-B cells.

HDAC7 deficiency alters 5-hmC at specific loci

We next performed a 5-hmC DNA immunoprecipitation (hMeDIP) followed by next-generation sequencing (hMeDIP-seq) in pro-B cells purified from *Hdac7*$^{fl/-}$ mice and their *Hdac7*$^{+/-}$ control littermates. While the total frequencies of 5-hmC peaks were similar, the global 5-hmC coverage was higher in HDAC7-deficient pro-B cells than in wild-type pro-B cells, in concordance with the results from Figure 1K (Figure 4A and Supplementary Figure S4A). The genomic distribution of the peaks did not differ significantly between the two genotypes, with most peaks located in intergenic and intronic regions as well as within LINE-1 elements (Figure 4B). Further, we found differential peaks in 5-hmC enrichment: HDAC7-deficient pro-B cells had an increase of ∼13 000, and a decrease in ∼15 800, 5-hmC peaks, as compared to wild-type B cells. Next, we performed an integrative analysis of the ATAC-seq and hMeDIP-seq data obtained. In particular, we overlapped the genes presenting a more open chromatin with the genes associated with higher 5-hmC enrichment in the HDAC7-deficient pro-B cells. We found that around 40% of genes with a more accessible chromatin harbored increased 5-hmC, indicating that HDAC7 may regulate a proper DNA compaction to avoid aberrant deposition of this epigenetic mark (Figure 4C). Focusing on alterations in potential lineage-inappropriate genes, we observed an increase in the enrichment of 5-hmC at the *Jun* gene in HDAC7-deficient pro-B cells as compared to wild-type cells, indicating that *Jun* may be aberrantly overexpressed in the absence of HDAC7 (Figure 4D). Indeed, RNA-seq data showed a higher expression of *Jun* in HDAC7-deficient pro-B and pre-B cells as compared to wild-type pro-B cells (Figure 4D). This correlated with a more open chromatin of the *Jun* gene in the absence of HDAC7 in both cell types (Figure 4D). Similar results were observed for the *Fosb* gene (Supplementary Figure S4B). The results from hMeDIP-seq and RNA-seq were confirmed by hMeDIP-qPCR and RT-qPCR experiments (Figure 4E, F, and Supplementary Figure S4C). Upregulation of additional lineage inappropriate genes, *Cd69* and *Itgb2,* in HDAC7-deficient pro-B and pre-B cells is shown in Supplementary Figure S4D. Importantly, we found in-

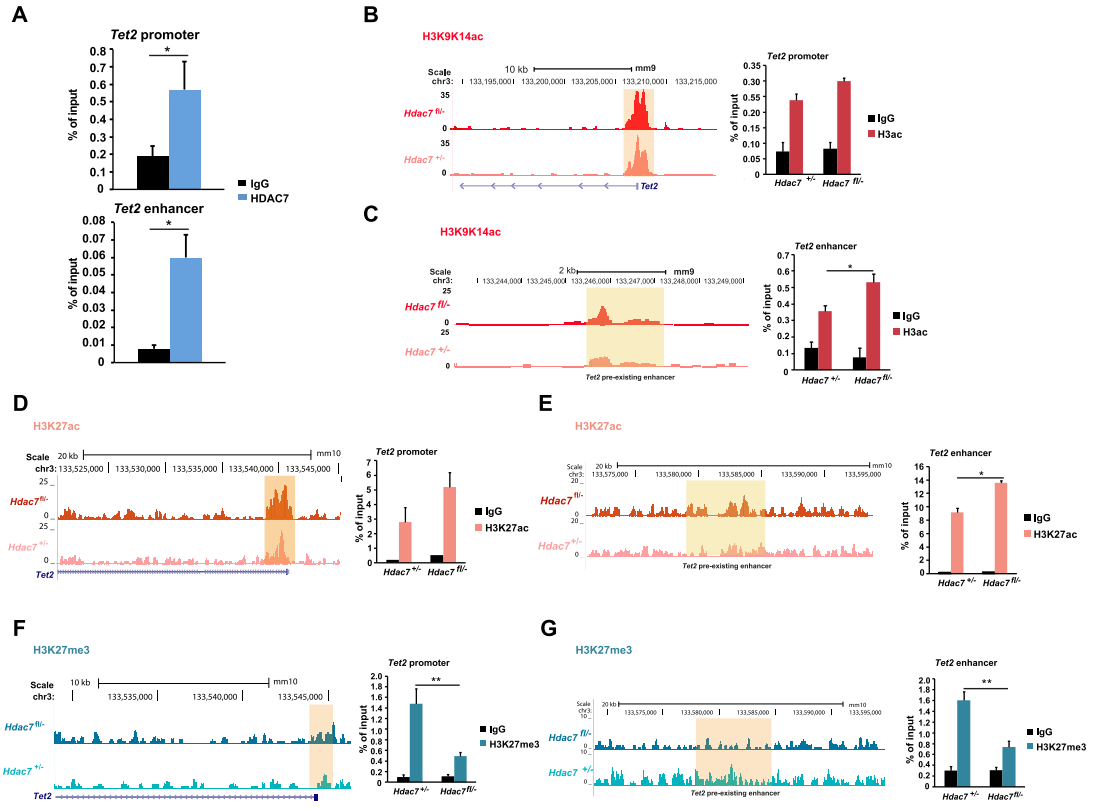


Figure 3. *Tet2* is a direct HDAC7 target gene in pro-B and pre-B cells. (**A**) Chromatin immunoprecipitation (ChIP) experiments showing the recruitment of HDAC7 to the *Tet2* promoter or the *Tet2* enhancer, quantified as % of input. (**B**) Genome browser snapshot of H3K9/K14ac ChIP-seq (left panel) showing differential enrichment levels at the promoter region of *Tet2* gene for HDAC7-deficient pro-B and control cells. ChIP-qPCR assay validating ChIP-seq data after immunoprecipitation with anti-H3K9/14ac antibody is shown in right panel, quantified as % of input. (**C**) As in (B), but for a *Tet2* pre-existing enhancer region. (**D**, **E**) As in (B, C), but for H3K27ac ChIP-seq enrichment and after immunoprecipitation with H3K27ac antibody. (**F**, **G**) As in (B, C), but for H3K27me3 ChIP-seq enrichment and after immunoprecipitation with H3K27me3 antibody. All data are represented as mean ± SEM of $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, unpaired *t*-test.

creased TET2 recruitment to the *Jun* gene loci identified in the hMeDIP-seq analysis, as well as to further myeloid gene promoters (*Fosl2*, *Ahnak* and *Itgb2*), in HDAC7-deficient pro-B cells (Figure 4G and Supplementary Figure S4E).

To definitively prove the functionality of HDAC7–TET2 axis, we performed rescue analysis (gain and loss of function) using three different experimental approaches. A graphical scheme depicting the three experimental approaches is shown in Supplementary Figure S4F. First, we transduced purified B cells from bone marrow of wild-type and HDAC7-deficient mice with a retroviral vector for specific shRNA against *Tet2* (shTet2) and compared them to counterpart cells transduced with control retroviral vector (shCtrl). We found that knocking down *Tet2* prevented the upregulation of *Jun* and *Fosl2* in HDAC7 deficient B cells (Figure 4H). Therefore, as a second approach, we took advantage of the HAFTL murine pre-B cell line engineered to transdifferentiate into functional macrophages by addition of β-estradiol (C11 cells) and previously reported in (29). Similarly to the case of C10 cells, HDAC7 and TET2 become downregulated and upregulated during cellular conversion, respectively. C11 were transduced with shCtrl or shTet2 retroviral vectors and sorted GFP-positive cells were later induced to macrophage transdifferentiation by the addition of β-estradiol, in order to achieve double HDAC7 and TET2 deficiency (Figure 4I). RT-qPCR assays showed that *Jun* and *Fosl2* were upregulated after cellular conversion, concomitant to HDAC7 downregulation. Importantly, *Tet2* knockdown resulted in a significant block of both inappropriate genes induction (Figure 4I). And third, we determined the expression of *Jun* and *Fosl2* in C10 samples from Figure 2H similarly to the case of C11 cells, *Jun* and *Fosl2* were upregulated during the conversion of pre-B cells into macrophages. Importantly, HDAC7 exogenous expression blocked their increased expression (Figure 4J). These data demonstrate that the HDAC7–TET2 axis is involved in the proper control of the expression of lineage inappropriate genes in B cells.

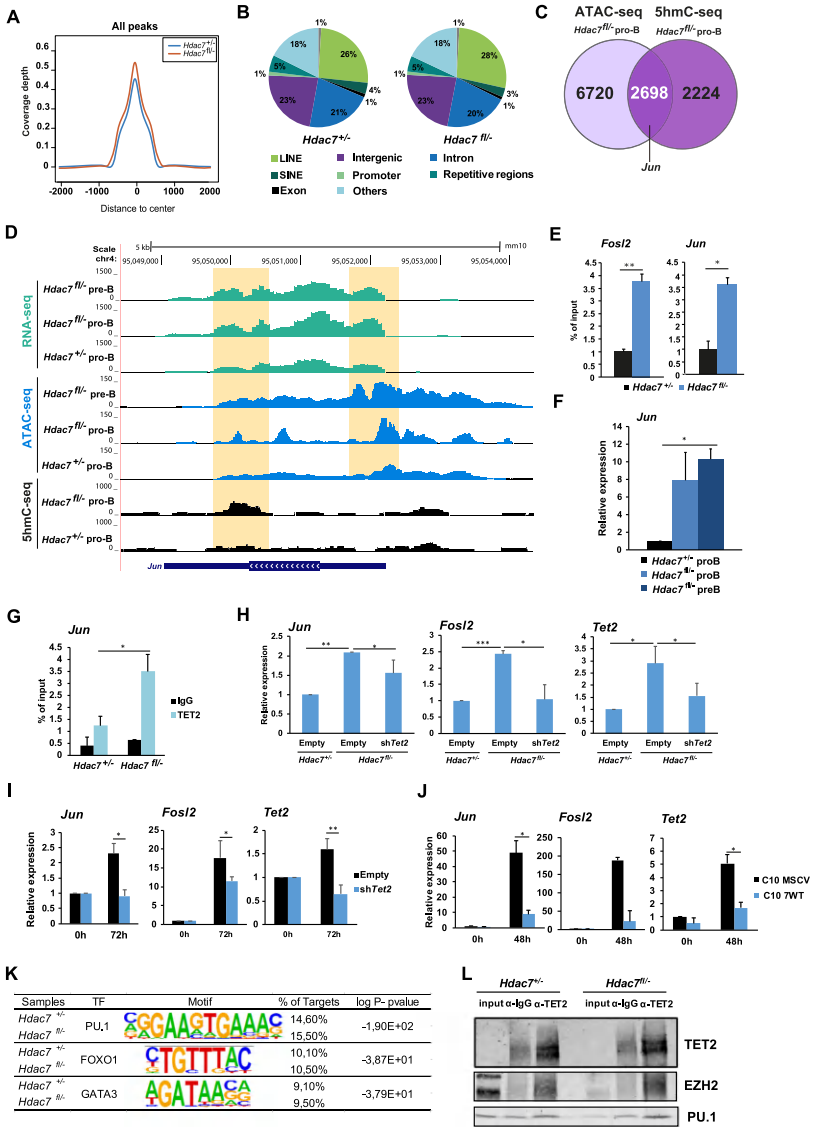


Figure 4. HDAC7 deficiency results in increased recruitment of TET2 and altered 5-hmC enrichment at B cell lineage inappropriate genes. (**A**) 5-hmC coverage depth (per base pair per peak per 10 million mapped reads) of 5-hmC peaks (−2 kb to +2 kb) in control and HDAC7-deficient pro-B cells. (**B**) Genomic distribution of 5-hmC enrichment in $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ pro-B cells. (**C**) Venn Diagram comparing enriched peaks associated genes in ATAC-seq from $Hdac7^{fl/-}$ pro-B cells and 5hmC enriched genes in hMeDIP-seq in $Hdac7^{fl/-}$ pro-B cells compared to control cells. (**D**) Genome browser snapshot of the *Jun* gene showing signal for 5-hmC from hMeDIP-seq, ATAC-seq and RNA-seq data in wild-type or HDAC7-deficient pro-B and pre-B cells. The peak locations are indicated in the orange-shaded rectangles. Orange shadows indicate *Jun* gene regions with differential peaks. (**E**) 5hmC-qPCR experiments showing the enrichment of 5-hmC in wild-type or HDAC7-deficient pro-B cells in the *Fosl2* and *Jun* genes. (**F**) Analysis by RT-qPCR of *Jun* mRNA levels in control or HDAC7-deficient pro-B cells and pre-B cells. (**G**) Analysis of TET2 recruitment at the *Jun* gene by ChIP-qPCR in control and HDAC7-deficient pro-B cells. (**H**) *Ex vivo* shCtrl and shTet2 retroviral infection in B cells from HDAC7-deficient mice. As control, wild-type B cells were transduced with shCtrl vector. Cells were infected with sh*Tet2* and expression levels of target genes were analyzed by RT-qPCR 72h after retroviral infection. (**I**) *In vitro* shTet2 infection of C11 pre-B cells. Cells were infected with shCtrl or shTet2 and then treated with β-estradiol to induce trans-differentiation to macrophages. Expression levels of target genes were analyzed after 72 h of transdifferentiation (**J**) C10 pre-B cells expressing empty plasmid (MSCV) and HDAC7-expressing plasmid (7WT) were treated with β-estradiol to induce transdifferentiation to macrophages. Expression levels of target genes were analyzed after 48 h of transdifferentiation. (**K**) The most relevant TF binding motifs in wild-type or HDAC7-deficient pro-B cells, based on their enrichment in 5-hmC using the HOMER database of known motifs. (**L**) Co-immunoprecipitation assays performed in $Hdac7^{+/-}$ and $Hdac7^{fl/-}$ $CD19^{+}$ B cells. Protein extracts were immunoprecipitated using an anti-TET2 antibody, using IgG as a negative control and total protein extract as input. Data in (E, F) are represented as mean ± SEM of $n = 4$. Data in (G–J) are represented as mean ± SEM of $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ unpaired *t* test.

Finally, we performed a motif enrichment analysis to determine whether HDAC7 deficiency produces an alteration in chromatin positioning that could lead to changes in TF occupancy. Although we found no differences associated with HDAC7, we did note enrichment of relevant TFs in the hematopoietic system, such as PU.1 (Figure 4K), which has been previously reported to interact with TET2 (51,52). The occupancy of PU.1 under both conditions is consistent with its relevance in both lymphoid and myeloid lineages. Accordingly, we corroborated that TET2 interacts with PU.1 in HDAC7-deficient B cells, as well as in wild-type B cells. As expected, EZH2 (a known TET2 partner) was also identified as an interactor (Figure 4L). Together, our results demonstrate an essential role of HDAC7 in silencing B cell-inappropriate genes by its regulation of TET2 expression and, consequently, of the DNA 5-hmC levels.

HDAC7 directs 5-hmC and expression of specific miRNA in pro-B cells

Further examination of our hMeDIP-seq data revealed that the coverage depth of 5-hmC peaks at miRNAs in HDAC7-deficient pro-B cells was higher than in control pro-B cells (Figure 5A). Additionally, integrative analysis with our ATAC-seq obtained data demonstrated that more than 50% of miRNA-related peaks that are enriched in 5-hmC mark in the absence of HDAC7, are also located in open chromatin regions (Figure 5B). In fact, we found that several miRNAs involved in leukemia and lymphoma, as well as in myeloid differentiation, such as miR-125b and miR34a, were more enriched in 5-hmC and located in more open chromatin regions in pro-B cells from *Hdac7*$^{fl/-}$ mice (53) (see an example of miR125b in Figure 5C). Using ChIP-qPCR, we found that TET2 recruitment increased at both miR125b- and miR34a-associated loci, which correlated to the enhanced 5-hmC enrichment in HDAC7-deficient pro-B cells (Figure 5D). Additionally, GSEA analysis of our RNA-seq data confirmed that gene sets related to miR-34a and miR-125b functions were more expressed upon HDAC7 deficiency (Figure 5E). To examine a potential connection between changes in 5-hmC and chromatin condensation and HDAC7-mediated control of miRNA expression, we performed a miRNA profiling using a qPCR-based panel containing over 375 miRNAs (miRCURY LNA™ microRNA Array [Exiqon]) in wild-type or HDAC7-deficient pro-B cells (Figure 5F and Supplementary Table S2). We found 25 miRNAs whose levels of expression differed significantly between wild-type and HDAC7-deficient pro-B cells, which correlated with differential 5hmC enrichment and chromatin state. miRNAs that were up-regulated and 5hmC-enriched under HDAC7-deficient conditions included miR-125b-5p, miR-126, miR-29b, miR-34a and miR-99a (Figure 5F and Supplementary Figure S5A). On the contrary, B-cell related miRNAs that were down-regulated upon HDAC7 deficiency, such as miR-150a and miR-181a, also presented decreased 5hmC enrichment and more closed chromatin state (Figure 5F and Supplementary Figure S5B). The differential expression observed in HDAC7-deficient pro-B cells of several miRNAs involved in the hematopoietic system or related disorders were validated by RT-qPCR (Figure 5G). Thus, aberrant microRNAs (such as miR-125b and miR-34a) were upregulated, while B cell–specific miRNAs (such as miR-28a, miR150, miR-142 and miR181a) were downregulated in HDAC7-deficient pro-B cells. Finally, we tested whether targets of these miRNAs are altered as a consequence of HDAC7 deficiency. On one hand, we observed that protein levels of MYB, target of miR-142 and miR-150, were increased in *Hdac7*$^{fl/-}$ cells (Figure 5H). On the other hand, we found decreased expression of PUMA and IRF4, both targets of miR-125b, upon HDAC7 deficiency (Figure 5I). Globally, our data indicate that, through the regulation of *Tet2*, HDAC7 controls the expression levels of crucial miRNAs of the immune system.

HDAC7 regulates 5-hmC levels and expression of LINE-1 elements

Our hMeDIP-seq approach also revealed that, according to the average signal from all the peaks obtained, the signal intensity of 5-hmC peaks associated with LINE-1 elements in HDAC7-deficient pro-B cells was higher than in their wild-type counterparts (Figure 6A). Gene ontology analysis of genes located in regions associated to LINE-1 elements revealed that 5-hmC enriched genes in *Hdac7*$^{fl/-}$ pro-B cells were more associated to gene activation and cell proliferation, whereas down-regulated LINE-1-related genes were involved in negative regulation of cell proliferation and cell differentiation (Figure 6B). In addition, we found that specific LINE-1-related loci were more susceptible to MNAse treatment in HDAC7-deficient B cells than wild-type cells and, therefore, more predisposed for aberrant upregulation (Figure 6C). Accordingly, we found that >50% of LINE-1 related peaks (associated 514 to genes) with 5-hmC enrichment were located in more open chromatin regions (Figure 6D). Hence, HDAC7 depletion was correlated with increased chromatin accessibility and enhanced or uncontrolled gene activation in LINE-1 associated loci (see examples in Figure 6E and Supplementary Figure S6A). Aberrant expression of LINE-1 elements is associated with genome instability. Accordingly, GSEA analysis of our RNA-seq data showed that the gene signatures in HDAC7-deficient pro-B or pre-B cells were associated to impaired DNA repair mechanisms (Figure 6F), suggesting a potential increase of genomic instability in the absence of HDAC7 that could correlate with LINE-1 aberrant expression. In fact, genes included in DNA repair geneset (form GSEA) that were down-regulated in HDAC7-deficient pro-B cells were related to DNA damage response and DNA repair abilities (Supplementary Figure S6B, C), supporting the potential affection of DNA repair capacity upon HDAC7 deficiency. Regions with differential 5-hmC peaks were validated by hMeDIP-qPCR assays, confirming that the absence of HDAC7 from pro-B cells resulted in higher levels of 5-hmC in LINE-1 elements (Figure 6G). Moreover, we observed a significant increase in the expression of LINE-1 transcripts from the most active families in HDAC7-deficient pro-B cells (Figure 6H), reinforcing the correlation between 5-hmC enrichment and gene activation. Previously published data revealed that TET1 and TET2 are recruited to the 5′ UTR of young LINE-1 elements in embryonic stem cells (54). We confirmed by ChIP-qPCR that

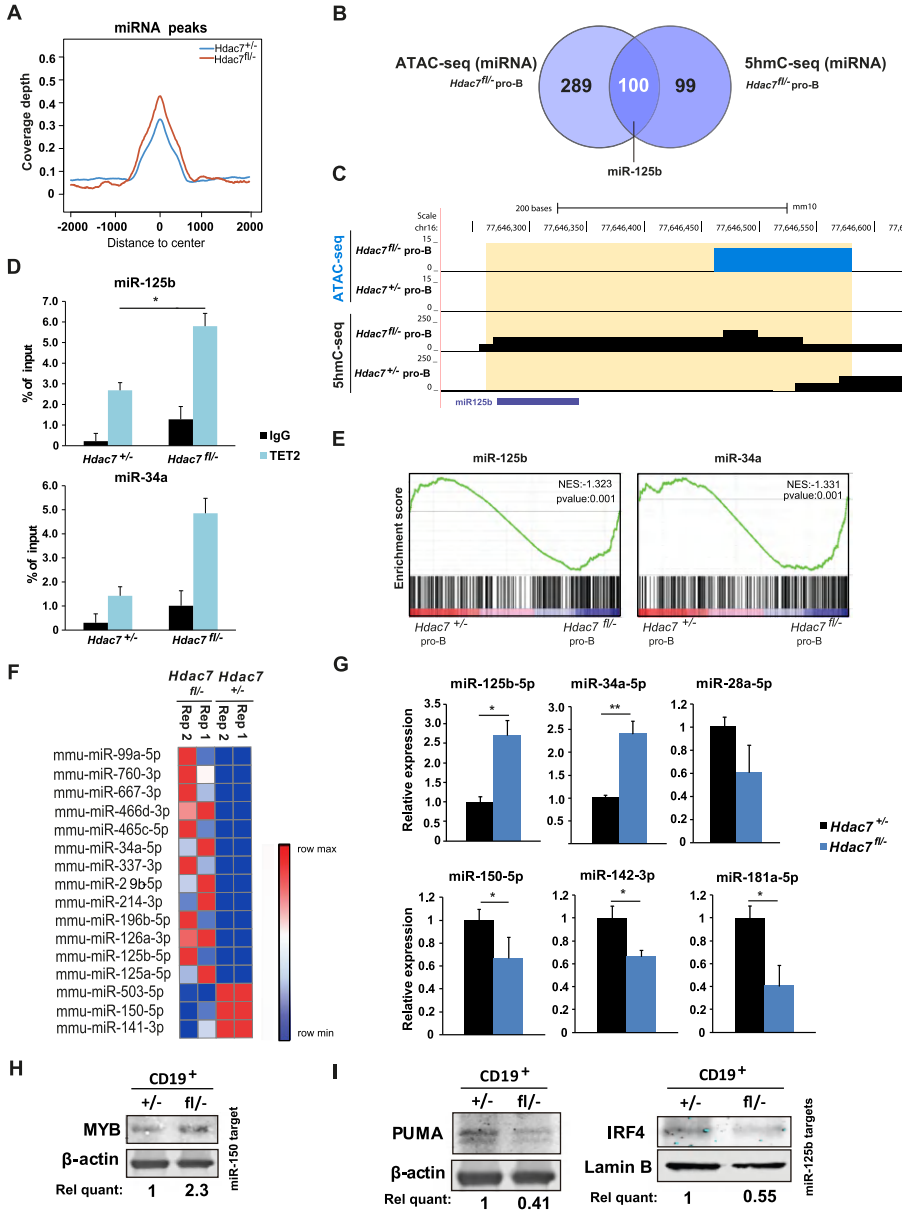


Figure 5. HDAC7 directs B cell–specific miRNA hydroxymethylation and expression patterns. (**A**) 5-hmC coverage depth (per base pair per peak per 10 million mapped reads) of 5hmC peaks located in microRNAs (−2 kb to +2 kb) from control or HDAC7-deficient pro-B cells. (**B**) Venn diagram comparing miRNA-related enriched peaks in ATAC-seq from *Hdac7*$^{fl/-}$ pro-B cells and miRNA-related peaks enriched in 5hmC-seq (or hMeDIP-seq) in *Hdac7*$^{fl/-}$ pro-B cells compared to control cells. (**C**) Example of 5-hmC and open chromatin enriched peaks at miR-125b from hMeDIP-seq and ATAC-seq experiments. (**D**) ChIP-qPCR analysis of TET2 recruitment to hydroxymethylated miRNAs in HDAC7-deficient pro-B cells. (**E**) GSEA analysis showing gene sets related to miR-34a and miR-125b functions, comparing HDAC7-deficient pro-B cells to control pro-B cells. miR-125B and miR-34a datasets were retrieved from www.gsea-msigdb.org. (**F**) Heat map of the differential expression of miRNAs for two HDAC7-deficient vs. wild-type replicates. Only miRNAs with a FC > 2 or FC < 0.5 2 from the miRCURY LNA™ Universal RT panel were selected. See also Supplementary Table S2. (**G**) RT-qPCR analysis of selected microRNAs from the miRCURY LNA™ Universal RT panel in HDAC7-deficient and control pro-B cells. The levels of U6 RNA were used for normalization. (**H**) Protein levels of miR-150 and miR-142 target MYB were assessed by western blot assays in control and HDAC7-deficient CD19^{+} cells. (**I**) Protein levels of miR-125b targets PUMA (left panel) and IRF4 (right panel) were assessed by western blot assays in control and HDAC7-deficient CD19^{+} cells. Using β-actin or Lamin-B as housekeepings, relative quantification of protein levels in G and H was performed with ImageJ software and is indicated below each panel. Data in (G) is represented as mean ± SEM of $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, unpaired t test.

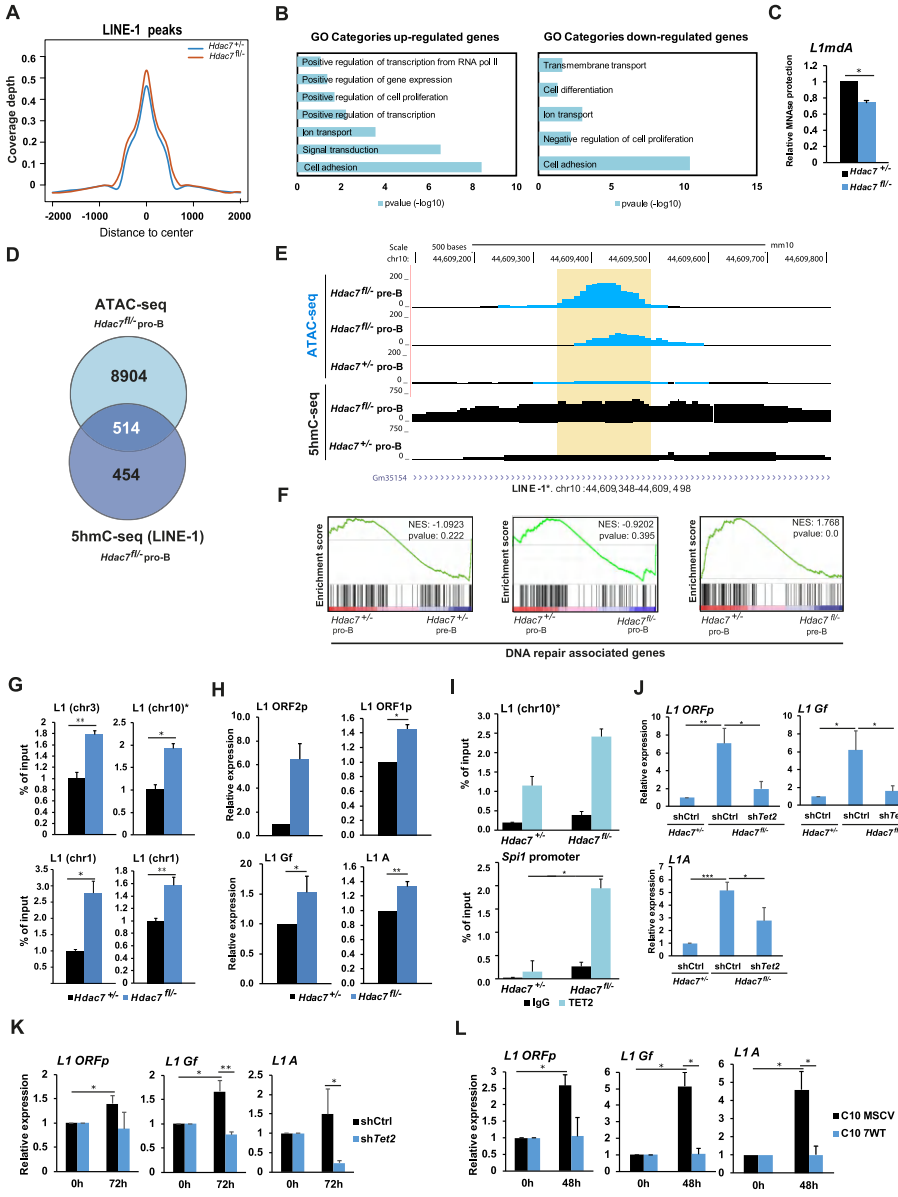


Figure 6. HDAC7 regulates 5-hydroxymethylation levels of transposable LINE-1 elements. (**A**) 5-hmC coverage depth (per base pair per peak per 10 million mapped reads) of 5-hmC peaks located in LINE-1 elements (−2 kb to +2 kb) in wild-type or HDAC7-deficient pro-B cells. (**B**) Gene Ontology (GO) enrichment analysis of genes associated to up-regulated (left panel) and down-regulated (right panel) regions in HDAC7-deficient pro-B cells respect to control pro-B cells in hMeDIP-seq shown in (A). (**C**) RT-qPCR showing protection of MNase-digested DNA at the 5min time point in wild-type or HDAC7-deficient B cells. (**D**) Venn diagram comparing enriched peaks associated genes in ATAC-seq from *Hdac7*$^{fl/-}$ pro-B cells and LINE-1 enriched peaks associated genes in 5hmC-seq (or hMeDIP-seq) in *Hdac7*$^{fl/-}$ pro-B cells compared to control cells. (**E**) Example of 5-hmC and ATAC-seq enrichment in young retrotransposon (L1) from peaks detected in hMeDIP-seq and ATAC-seq data. The peak location found common in the two omics analyses is located in the orange-shaded rectangle. (**F**) GSEA analysis comparing wild-type pre-B or HDAC7-deficient pro-B and pre-B cells to wild-type pro-B cells in expression profiles related to DNA repair signatures. DNA repair dataset was retrieved from www.gsea-msigdb.org. (**G**) hMeDIP-qPCR analysis of 5-hmC enrichment at LINE-1 retrotransposable elements in wild-type or HDAC7-deficient pro-B cells. (**H**) RT-qPCR analysis of the expression of proteins encoded by LINE-1 (ORFp1, ORFp2) and the most active L1 elements subfamilies (Gf, A) in wild-type or HDAC7-deficient pro-B cells. (**I**) ChIP-qPCR analysis of TET2 recruitment at hydroxymethylated L1 elements. TET2 enrichment at the *Spi1* promoter was used as a positive control. (**J–L**) As in Figure 4H-J, but analyzing the expression levels of L1 associated regions (L1 ORFp, L1 Gf and L1A). Data from (**G–L**) are represented as mean ± SEM of $n = 3$. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, unpaired t test.

TET2 was recruited to LINE-1 elements with enhanced 5-hmC in HDAC7-deficient pro-B cells; TET2 recruitment to the *Spi1* promoter was used as a positive control (Figure 6I). Finally, using samples from the gain and loss of function experimental approaches shown in Figure 4H–J, we further demonstrated that HDAC7-mediated LINE-1 regulation depends on *Tet2* expression (Figure 6J-L). Overall, our data indicated that HDAC7 plays a role in preserving the chromatin state and genome integrity in B cells by restricting the expression levels of TET2, which consequently leads to the maintenance of physiological levels of 5-hmC at retrotransposon elements.

DISCUSSION

Here we reveal an unprecedented HDAC7-mediated molecular mechanism that preserves the correct chromatin conformation, histone marks deposition and DNA 5-hydroxymethylation state. Notably, this state is essential for B cell identity and, consequently, for a correct gene expression pattern during early B cell development. HDAC7 deficiency resulted in a global chromatin de-compaction that significantly increased its accessibility. This correlated with a global increased of H3K27ac in the absence of HDAC7 from pro-B cells. Chromatin organization is dynamically reshaped during B cell development, obtaining unique populations at each differentiation stage (55); however, this process is highly controlled, and alterations in chromatin state (such as that observed here in HDAC7-deficient cells) can alter the transcriptional regulation and gene expression patterns, which can drive malignant transformation (56). In line with the increased chromatin accessibility, HDAC7 deficiency also caused a significant decrease in H3K9me3, a hallmark of heterochromatin state, which is involved in maintaining lineage stability and preventing cell reprogramming (47,57,58). TET enzymes are mainly recruited to open chromatin regions; therefore, euchromatin (unlike heterochromatin) is enriched in 5-hmC (49). HDAC7 deficiency results in a global decrease of heterochromatin regions and enhanced TET2 recruitment. It has been reported that TET2 loss at stem cell stages produced an aberrant number of myelomonocytic cells and impairment in the expression of macrophage markers such as Mac-1 in myeloid cells (20,22). Within the B cell lineage, TET2 conditional deficiency at pro-B cell stage does not cause any phenotype during development and differentiation. Only conditional deletion of both TET2 and TET3 lead to defective B cell development (52). Even though there is no phenotype observed by TET2 deficiency *in vivo*, TET2 has been reported to play a critical role in mediating the hydroxymethylation of cytosine residues from myeloid genes during pre-B cells conversion into macrophages (24,29). Despite loss of TET2 and TET3 leads to aberrant lymphocyte development and related disorders (59,60), and loss of TET2 enzymatic activity appears to mainly affect myeolopoiesis (23,60), our results strongly suggest that HDAC7 is a critical factor that preserves B cell identity and correct DNA hydroxymethylation state via *Tet2* gene silencing.

Previous studies have established a close relationship between transcription regulators and dynamic changes in DNA methylation during B cell development and commitment, specifically at the pro-B to pre-B cell transition (14,59). Hematopoietic cells present low hydroxymethylation levels (∼0.2%) compared to other cell types, such as Purkinje cells or embryonic stem cells (∼5%) (50,61). However, we observed a significant decrease of heterochromatin, and an increase in 5-hmC, upon HDAC7 deficiency, leading to several molecular and biological consequences. First, we found a high percentage of 5-hmC peaks located in intergenic and distal promoter regions. These results may imply that distal regulatory regions with enhancer features are dependent on TET2-mediated DNA demethylation and may correlate with the presence of additional mechanisms that control DNA methylation status at promoter-associated regions (62,63). *Jun* and *Fosl2* were found among the myeloid and T cell genes with increased 5-hmC levels in HDAC7-deficient B cells. *Fosl2* is a TET2-activated gene during the transdifferentiation of pre-B cells to macrophages (24), and *Jun* undergoes enhancer demethylation prior to B cell being reprogrammed into induced pluripotent stem cells (iPSCs) (64). The finding that lineage-inappropriate genes are already marked with 5-hmC in B cell progenitors supports the notion that they may be epigenetically poised in early stages of development.

HDAC7 deficiency also led to differential 5-hmC levels in some regions containing miRNAs. miRNAs are epigenetic players that have crucial roles in multiple developmental processes, and their de-regulation is involved in many biological disorders. Recent studies have demonstrated that miRNAs have a role in normal and malignant B cell development, by modulating the expression of key regulatory genes (65–67). For instance, miR-34a is strongly expressed in myeloid cells, and its constitutive expression in B cells blocks the pro-B to pre-B cell transition (26,68); notably, this is the biological effect that we observed upon HDAC7 deficiency. miR-150, which is down-regulated in pro-B cells from *Hdac7*-null mice, is related to B cell development and performs tumor-suppressor functions in leukemic cells. miR-142 and miR-181 are highly expressed in a cell-specific manner (69), such as in hematopoietic cells. Specifically, B cell function is impaired in miR-142 deficiency conditions (67), and miR181a regulates positively the B lymphocyte differentiation (70). Previous studies indicated that miR-126 is downregulated in lymphoid cells, (71,72), which is consistent with our results. Moreover, miR-29b is activated by C/EBPα and represses *Tet2* expression, which concords with C/EBPα and *Tet2* up-regulation when HDAC7 is deficient (73). miR-34a is strongly expressed in myeloid cells, and its constitutive expression in B cells blocks the pro-B to pre-B cell transition (68). Finally, some members of the miR-99 family, such as miR-99b, are abundant in macrophages, neutrophils, and monocytes. Here, we observed that another member of the family related to leukemic stem cells, miR-99a, was upregulated in HDAC7-deficient pro-B cells. Thus, our results demonstrate that HDAC7 can also exert its gene repressive function during early B cell development by regulating gene expression, presumably by interacting with its classical partner MEF2C, which may be recruited to miRNA regulatory regions, as it does at other specific miRNA regions in the skeletal muscle (74).

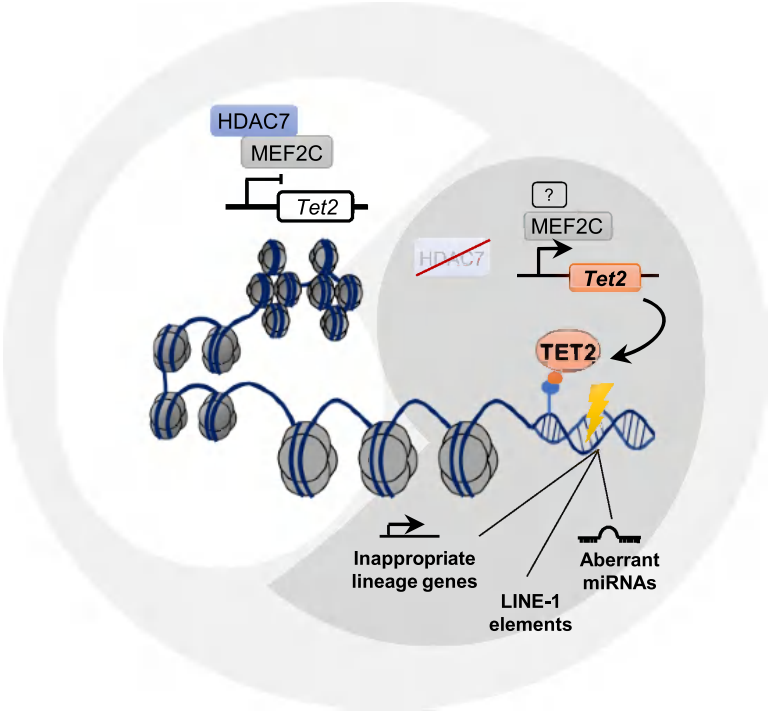


Figure 7. Visual representation summarizing the role of HDAC7–TET2 axis in the regulation of proper gene silencing in early B cell development.

We have detected increased 5-hmC enrichment and a more open chromatin state in some regions containing LINE-1 transposable elements in HDAC7-deficient pro-B cells. LINE-1 elements are the only active autonomous retrotransposons in the mammalian genome and, consequently, a potential disturber of chromatin stability (75). In fact, LINE-1 transcripts from the A and Gf subfamilies, which present increased expression in HDAC7-deficient pro-B cells, contain some members that still have the full-length transcript, which maintains its retrotransposon activity in the mouse genome (76). The fact that increased 5-hmC is associated to TET2 recruitment suggests that HDAC7 might be required to preserve chromatin integrity by mediating the silenced status of LINE-1 elements. Of note, recent studies have shown that tight regulation of TET2 activity is essential for correct maintenance of genome stability: TET2 deficiency produces defects in DNA damage response, and its overexpression produces chromosome instability and aneuploidy due to a collapse in BER activity (77,78). Thus, results from this paper suggest that TET2 aberrant expression by HDAC7-deficient B cells may impair their capacity to repair DNA damage, which agree with the observed higher cell death rates in these cells that we reported in our previous published work (26).

Significant loss of H3K9me3 enrichment upon HDAC7 deficiency could also correlate with LINE-1 deregulated expression, since this heterochromatin mark is required to repress aberrant expression of retrotransposons in mammal embryonic stem cells (79,80). However, given that DNA methylation is the main source of LINE-1 repression in more differentiated cells, we suggest that LINE-1 are silenced due to DNA demethylation caused by TET2 upregulation; this would reinforce the effects of DNA methylation loss following TET2 up-regulation upon HDAC7 deficiency. This result indicates that LINE-1 deregulation is produced as a consequence of HDAC7 deficiency.

Our results represent a significant step forward in our understanding of how B cells acquire their genetic identity, from three different perspectives. First, we identified HDAC7 as a chromatin modulator that regulates the heterochromatin state and histone marks deposition in early B cell development. Second, we demonstrated that HDAC7 is the specific transcription repressor that controls TET2 activity, which it achieves by fine-tuning its physiological expression levels in pro-B cells. This may represent the mechanistic explanation for the different TET2 expression levels observed in myeloid and lymphoid cells. Third, our results reveal an unexpected role for HDAC7 in controlling proper DNA 5-hydroxymethylation status and expression of lineage- or functionally-inappropriate genes, microRNAs, and non-coding elements (such as LINE-1 elements) in pro-B cells. We recently identified HDAC7 to be a novel biomarker and prognostic factor in infants (<1-year-olds) with pro-B acute lymphoblastic leukemia (pro-B-ALL) and MLL-AF4 rearrangement (27). This subgroup of pediatric patients presents an extremely adverse outcome, with sur-

vival rate below 35%, and the loss of HDAC7 is associated with a worse prognosis. Therefore, the elucidation of the exact molecular mechanisms that HDAC7 exert during physiological early B cell development will be crucial to understand how their deregulation can result in B cell–associated malignancies, with potential implications in the clinics.

Altogether, our findings lead us to a proposed model by which HDAC7 functions during early B cell development are not restricted to controlling expression by direct recruitment to its target genes. Rather, HDAC7 governs the expression of another crucial epigenetic regulator, TET2. The identified HDAC7–TET2 epigenetic axis is essential to preserve proper 5-hmC and histone marks levels, chromatin compaction, and expression of miRNAs and LINE-1 elements (Figure 7). We anticipate that our findings may open new avenues to understanding the consequences of HDAC7 deregulation in altering the molecular mechanism found in B cell–related malignancies, eventually leading to strategies to develop therapies for these pathologies.

DATA AVAILABILITY

Data are available in GEO repository as follows: RNA-seq (GSE171855), ATAC-seq (GSE204672), ChIP-seq (GSE204673) and hMeDIP-seq (GSE135263).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. We thank Dr Eric Olson (UT Southwestern Medical Center, Dallas, TX, USA) and Dr Michael Reth (Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany) for kindly providing the *Hdac7*$^{loxp/-}$ and mb1-Cre mice, respectively. We thank Lucía Fanlo for her assistance in technical issues and bioinformatics analysis of ChIP-seq and ATAC-seq experiments. We thank Alberto Bueno for deep analysis of our RNA-seq and hMeDIP-seq data, in order to assess the presence of differentially expressed dsRNA species. We also thank Drs Pura Muñoz Cánoves and Tokameh Mahmoudi for helpful comments on the manuscript.

FUNDING

Spanish Ministry of Economy and Competitiveness (MINECO) [SAF2017-87990-R]; Spanish Ministry of Science and Innovation (MICINN) [EUR2019-103835]; Josep Carreras Leukaemia Research Institute (IJC, Badalona, Barcelona); IDIBELL Research Institute (L'Hospitalet de Llobregat, Barcelona); A.M. is funded by the Spanish Ministry of Science, Innovation and Universities, which is part of the Agencia Estatal de Investigación (AEI) [PRE2018-083183] (cofunded by the European Social Fund]; OdB. was funded by a Juan de la Cierva Formación Fellowship from the Spanish Ministry of Science, Innovation and Universities [FJCI-2017-32430]; Postdoctoral Fellowship from the Asociación Española Contra el Cáncer (AECC) Foundation [POSTD20024DEBA]; B.M. is awardee of the Ayudas para la formación del profesorado universitario [FPU18/00755, Ministerio de Universidades]; B.M.J. is funded by La Caixa Banking Foundation Junior Leader project [LCF/BQ/PI19/11690001]; FEDER/Spanish Ministry of Science and Innovation [RTI2018-094788-A-I00]; L.T.-D. is funded by the FPI Fellowship [PRE2019-088005]; L.R. is funded by an AGAUR FI fellowship [2019FI-B00017]; J.L.S. is funded by ISCIII [CP19/00176], co-funded by ESF, 'Investing in your future' and the Spanish Ministry of Science, Innovation and Universities [PID2019-111243RA-I00]. CRG acknowledge the support of the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033). Funding for open access charge: Spanish Ministry of Science, Innovation and Universities (MICIU) [SAF2017-87990-R, EUR2019-103835].

Conflict of interest statement. None declared.

REFERENCES

1. Nimmo,R.A., May,G.E. and Enver,T. (2015) Primed and ready: understanding lineage commitment through single cell analysis. *Trends Cell Biol.*, **25**, 459–467.
2. Parra,M. (2009) Epigenetic events during b lymphocyte development. *Epigenetics*, **4**, 462–468.
3. Ramírez,J., Lukin,K. and Hagman,J. (2010) From hematopoietic progenitors to b cells: mechanisms of lineage restriction and commitment. *Curr. Opin. Immunol.*, **22**, 177–184.
4. Cobaleda,C., Schebesta,A., Delogu,A. and Busslinger,M. (2007) Pax5: the guardian of b cell identity and function. *Nat. Immunol.*, **8**, 463–470.
5. Kwon,K., Hutter,C., Sun,Q., Bilic,I., Cobaleda,C., Malin,S. and Busslinger,M. (2008) Instructive role of the transcription factor E2A in early b lymphopoiesis and germinal center b cell development. *Immunity*, **28**, 751–762.
6. Kong,N.R., Davis,M., Chai,L., Winoto,A. and Tjian,R. (2016) MEF2C and EBF1 co-regulate B cell-specific transcription. *PLoS Genet.*, **12**, e1005845.
7. Ikawa,T., Kawamoto,H., Wright,L.Y.T. and Murre,C. (2004) Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity*, **20**, 349–360.
8. Nutt,S.L. and Kee,B.L. (2007) The transcriptional regulation of b cell lineage commitment. *Immunity*, **26**, 715–725.
9. Pongubala,J.M.R., Northrup,D.L., Lancki,D.W., Medina,K.L., Treiber,T., Bertolino,E., Thomas,M., Grosschedl,R., Allman,D. and Singh,H. (2008) Transcription factor EBF restricts alternative lineage options and promotes b cell fate commitment independently of pax5. *Nat. Immunol.*, **9**, 203–215.
10. Stehling-Sun,S., Dade,J., Nutt,S.L., DeKoter,R.P. and Camargo,F.D. (2009) Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor mef2c. *Nat. Immunol.*, **10**, 289–296.
11. Boller,S. and Grosschedl,R. (2014) The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunol. Rev.*, **261**, 102–115.
12. Vilarrasa-Blasi,R., Soler-Vila,P., Verdaguer-Dot,N., Russiñol,N., Di Stefano,M., Chapaprieta,V., Clot,G., Farabella,I., Cuscó,P., Agirre,X. *et al.* (2021) Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat. Commun.*, **12**, 651.
13. Martin-Subero,J.I. and Oakes,C.C. (2018) Charting the dynamic epigenome during B-cell development. *Semin. Cancer Biol.*, **51**, 139–148.
14. Benner,C., Isoda,T. and Murre,C. (2015) New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing b cell progenitors. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 12776–12781.

15. Almamun,M., Levinson,B.T., Gater,S.T., Schnabel,R.D., Arthur,G.L., Wade Davis,J. and Taylor,K.H. (2015) Genome-wide DNA methylation analysis in precursor B-cells. *Epigenetics*, **9**, 1588–1595.
16. Schuyler,R.P., Merkel,A., Raineri,E., Altucci,L., Vellenga,E., Martens,J.H.A., Pourfarzad,F., Kuijpers,T.W., Burden,F., Farrow,S. *et al.* (2016) Distinct trends of DNA methylation patterning in the innate and adaptive immune systems. *Cell Rep.*, **17**, 2101–2111.
17. Boller,S., Li,R. and Grosschedl,R. (2018) Defining b cell chromatin: lessons from EBF1. *Trends Genet.*, **34**, 257–269.
18. Azagra,A., Marina-Zárate,E., Ramiro,A.R., Javierre,B.M. and Parra,M. (2020) From loops to looks: transcription factors and chromatin organization shaping terminal b cell differentiation. *Trends Immunol.*, **41**, 46–60.
19. Wu,X., Li,G. and Xie,R. (2018) Decoding the role of TET family dioxygenases in lineage specification. *Epigenetics Chromatin*, **11**, 58.
20. Moran-Crusio,K., Reavie,L., Shih,A., Abdel-Wahab,O., Ndiaye-Lobry,D., Lobry,C., Figueroa,M.E., Vasanthakumar,A., Patel,J., Zhao,X. *et al.* (2011) Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, **20**, 11–24.
21. Quivoron,C., Couronné,L., Della Valle,V., Lopez,C.K., Plo,I., Wagner-Ballon,O., Do Cruzeiro,M., Delhommeau,F., Arnulf,B., Stern,M.-H. *et al.* (2011) TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*, **20**, 25–38.
22. Ko,M., Bandukwala,H.S., An,J., Lamperti,E.D., Thompson,E.C., Hastie,R., Tsangaratou,A., Rajewsky,K., Koralov,S.B. and Rao,A. (2011) Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 14566–14571.
23. Ko,M., Huang,Y., Jankowska,A.M., Pape,U.J., Tahiliani,M., Bandukwala,H.S., An,J., Lamperti,E.D., Koh,K.P., Ganetzky,R. *et al.* (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, **468**, 839–843.
24. Kallin,E.M., Rodríguez-Ubreva,J., Christensen,J., Cimmino,L., Aifantis,I., Helin,K., Ballestar,E. and Graf,T. (2012) Tet2 facilitates the derepression of myeloid target genes during CEBPα-Induced transdifferentiation of Pre-B cells. *Mol. Cell*, **48**, 266–276.
25. Cull,A.H., Snetsinger,B., Buckstein,R., Wells,R.A. and Rauh,M.J. (2017) Tet2 restrains inflammatory gene expression in macrophages. *Exp. Hematol.*, **55**, 56–70.
26. Azagra,A., Román-González,L., Collazo,O., Rodríguez-Ubreva,J., de Yébenes,V.G., Barneda-Zahonero,B., Rodríguez,J., Castro de Moura,M., Grego-Bessa,J., Fernández-Duran,I. *et al.* (2016) In vivo conditional deletion of HDAC7 reveals its requirement to establish proper b lymphocyte identity and development. *J. Exp. Med.*, **213**, 2591–2601.
27. de Barrios,O., Galaras,A., Trincado,J.L., Azagra,A., Collazo,O., Meler,A., Agraz-Doblas,A., Bueno,C., Ballerini,P., Cazzaniga,G. *et al.* (2021) HDAC7 is a major contributor in the pathogenesis of infant t(4;11) proB acute lymphoblastic leukemia. *Leukemia*, **35**, 2086–2091.
28. Chang,S., Young,B.D., Li,S., Qi,X., Richardson,J.A. and Olson,E.N. (2006) Histone deacetylase 7 maintains vascular integrity by repressing matrix metalloproteinase 10. *Cell*, **126**, 321–334.
29. Bussmann,L.H., Schubert,A., Vu Manh,T.P., De Andres,L., Desbordes,S.C., Parra,M., Zimmermann,T., Rapino,F., Rodriguez-Ubreva,J., Ballestar,E. *et al.* (2009) A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell*, **5**, 554–566.
30. Barneda-Zahonero,B., Román-González,L., Collazo,O., Rafati,H., Islam,A.B.M.M.K., Bussmann,L.H., di Tullio,A., De Andres,L., Graf,T., López-Bigas,N. *et al.* (2013) HDAC7 is a repressor of myeloid genes whose downregulation is required for transdifferentiation of Pre-B cells into macrophages. *PLoS Genet.*, **9**, e1003503.
31. Figueroa,M.E., Abdel-Wahab,O., Lu,C., Ward,P.S., Patel,J., Shih,A., Li,Y., Bhagwat,N., Vasanthakumar,A., Fernandez,H.F. *et al.* (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*, **18**, 553–567.
32. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
33. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq-A python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
34. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
35. Vazquez,B.N., Thackray,J.K., Simonet,N.G., Chahar,S., Kane-Goldsmith,N., Newkirk,S.J., Lee,S., Xing,J., Verzi,M.P., An,W. *et al.* (2019) SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina. *Nucleic Acids Res.*, **47**, 7870–7885.
36. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
37. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nussbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of chip-Seq (MACS). *Genome Biol.*, **9**, R137.
38. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
39. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
40. Ewels,P., Magnusson,M., Lundin,S. and Käller,M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
41. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
42. Thierion,E., Le Men,J., Collombet,S., Hernandez,C., Coulpier,F., Torbey,P., Thomas-Chollier,M., Noordermeer,D., Charnay,P. and Gilardi-Hebenstreit,P. (2017) Krox20 hindbrain regulation incorporates multiple modes of cooperation between cis-acting elements. *PLoS Genet.*, **13**, e1006903.
43. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
44. Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U. and Speed,T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
45. Painter,M.W., Davis,S., Hardy,R.R., Mathis,D., Benoist,C. and and Immunological Genome Project Consortium.and Immunological Genome Project Consortium. (2011) Transcriptomes of the b and t lineages compared by multiplatform microarray profiling. *J. Immunol.*, **186**, 3047–3057.
46. Jojic,V., Shay,T., Sylvia,K., Zuk,O., Sun,X., Kang,J., Regev,A., Koller,D., Best,A.J., Knell,J. *et al.* (2013) Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.*, **14**, 633–643.
47. Nicetto,D. and Zaret,K.S. (2019) Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Curr. Opin. Genet. Dev.*, **55**, 1–10.
48. Izzo,F., Lee,S.C., Poran,A., Chaligne,R., Gaiti,F., Gross,B., Murali,R.R., Deochand,S.D., Ang,C., Jones,P.W. *et al.* (2020) DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.*, **52**, 378–387.
49. Kubiura,M., Okano,M., Kimura,H., Kawamura,F. and Tada,M. (2012) Chromosome-wide regulation of euchromatin-specific 5mC to 5hmC conversion in mouse ES cells and female human somatic cells. *Chromosom. Res.*, **20**, 837–848.
50. Ficz,G., Branco,M.R., Seisenberger,S., Santos,F., Krueger,F., Hore,T.A., Marques,C.J., Andrews,S. and Reik,W. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, **473**, 398–402.
51. de la Rica,L., Rodríguez-Ubreva,J., García,M., Islam,A.B., Urquiza,J.M., Hernando,H., Christensen,J., Helin,K., Gómez-Vaquero,C. and Ballestar,E. (2013) PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated

methylation in monocyte-to-osteoclast differentiation. *Genome Biol.*, **14**, R99.
52. Lio,C.-W., Zhang,J., González-Avalos,E., Hogan,P.G., Chang,X. and Rao,A. (2016) Tet2 and tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. *Elife*, **5**: e18290.
53. Chaudhuri,A.A., So,A.Y.-L., Mehta,A., Minisandram,A., Sinha,N., Jonsson,V.D., Rao,D.S., O'Connell,R.M. and Baltimore,D. (2012) Oncomir miR-125b regulates hematopoiesis by targeting the gene lin28A. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 4233–4238.
54. de la Rica,L., Deniz,Ö., Cheng,K.C.L., Todd,C.D., Cruz,C., Houseley,J. and Branco,M.R. (2016) TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.*, **17**, 234.
55. Schoonhoven,A.van, Huylebroeck,D., Hendriks,R.W. and Stadhouders,R. (2020) 3D genome organization during lymphocyte development and activation. *Brief. Funct. Genomics*, **19**, 71–82.
56. Vilarrasa-Blasi,R., Soler-Vila,P., Verdaguer-Dot,N., Russiñol,N., Di Stefano,M., Chapaprieta,V., Clot,G., Farabella,I., Cuscó,P., Kulis,M. *et al.* (2021) Dynamics of genome architecture and chromatin function during human b cell differentiation and neoplastic transformation. *Nat. Commun.*, **12**, 651.
57. Becker,J.S., Nicetto,D. and Zaret,K.S. (2016) H3K9me3-Dependent heterochromatin: barrier to cell fate changes. *Trends Genet.*, **32**, 29–41.
58. Allan,R.S., Zueva,E., Cammas,F., Schreiber,H.A., Masson,V., Belz,G.T., Roche,D., Maison,C., Quivy,J.-P., Almouzni,G. *et al.* (2012) An epigenetic silencing pathway controlling t helper 2 cell lineage commitment. *Nature*, **487**, 249–253.
59. Orlanski,S., Labi,V., Reizel,Y., Spiro,A., Lichtenstein,M., Levin-Klein,R., Koralov,S.B., Skversky,Y., Rajewsky,K., Cedar,H. *et al.* (2016) Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 5018–5023.
60. Ito,K., Lee,J., Chrysanthou,S., Zhao,Y., Josephs,K., Sato,H., Teruya-Feldstein,J., Zheng,D., Dawlaty,M.M. and Ito,K. (2019) Non-catalytic roles of tet2 are essential to regulate hematopoietic stem and progenitor cell homeostasis. *Cell Rep.*, **28**, 2480–2490.
61. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-Hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, **324**, 929–930.
62. Rasmussen,K.D. and Helin,K. (2016) Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.*, **30**, 733–750.
63. Rasmussen,K.D., Berest,I., Keβler,S., Nishimura,K., Simón-Carrasco,L., Vassiliou,G.S., Pedersen,M.T., Christensen,J., Zaugg,J.B. and Helin,K. (2019) TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Res.*, **29**, 564–575.
64. Sardina,J.L., Collombet,S., Tian,T.V., Gómez,A., Di Stefano,B., Berenguer,C., Brumbaugh,J., Stadhouders,R., Segura-Morales,C., Gut,M. *et al.* (2018) Transcription factors drive tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell*, **23**, 727–741.
65. Zhang,J., Jima,D.D., Jacobs,C., Fischer,R., Gottwein,E., Huang,G., Lugar,P.L., Lagoo,A.S., Rizzieri,D.A., Friedman,D.R. *et al.* (2009) Patterns of microRNA expression characterize stages of human B-cell differentiation. *Blood*, **113**, 4586.
66. Chen,C.-Z., Li,L., Lodish,H.F. and Bartel,D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83–86.
67. Zheng,B., Xi,Z., Liu,R., Yin,W., Sui,Z., Ren,B., Miller,H., Gong,Q. and Liu,C. (2018) The function of microRNAs in B-cell development, lymphoma, and their potential in clinical practice. *Front. Immunol.*, **9**, 936.
68. Rao,D.S., O'Connell,R.M., Chaudhuri,A.A., Garcia-Flores,Y., Geiger,T.L. and Baltimore,D. (2010) MicroRNA-34a perturbs b lymphocyte development by repressing the forkhead box transcription factor foxp1. *Immunity*, **33**, 48–59.
69. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
70. Li,J., Wan,Y., Ji,Q., Fang,Y. and Wu,Y. (2013) The role of microRNAs in B-cell development and function. *Cell. Mol. Immunol.*, **10**, 107–112.
71. Lechman,E.R., Gentner,B., Ng,S.W.K., Schoof,E.M., van Galen,P., Kennedy,J.A., Nucera,S., Ciceri,F., Kaufmann,K.B., Takayama,N. *et al.* (2016) miR-126 regulates distinct self-renewal outcomes in normal and malignant hematopoietic stem cells. *Cancer Cell*, **29**, 214–228.
72. Petriv,O.I., Kuchenbauer,F., Delaney,A.D., Lecault,V., White,A., Kent,D., Marmolejo,L., Heuser,M., Berg,T., Copley,M. *et al.* (2010) Comprehensive microRNA expression profiling of the hematopoietic hierarchy. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15443–15448.
73. Amodio,N., Rossi,M., Raimondi,L., Pitari,M.R., Botta,C., Tagliaferri,P. and Tassone,P. (2015) miR-29s: a family of epi-miRNAs with therapeutic implications in hematologic malignancies. *Oncotarget*, **6**, 12837–12861.
74. Liu,N., Williams,A.H., Kim,Y., McAnally,J., Bezprozvannaya,S., Sutherland,L.B., Richardson,J.A., Bassel-Duby,R. and Olson,E.N. (2007) An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 20844–20849.
75. Richardson,S.R., Doucet,A.J., Kopera,H.C., Moldovan,J.B., Garcia-Perez,J.L. and Moran,J.V. (2015) The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol. Spectr.*, **3**, MDNA3-0061-2014.
76. Goodier,J.L., Ostertag,E.M., Du,K. and Kazazian,H.H. (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.*, **11**, 1677–1685.
77. Kafer,G.R., Li,X., Horii,T., Suetake,I., Tajima,S., Hatada,I. and Carlton,P.M. (2016) 5-Hydroxymethylcytosine Marks sites of DNA damage and promotes genome stability. *Cell Rep.*, **14**, 1283–1292.
78. Mahfoudhi,E., Talhaoui,I., Cabagnols,X., Della Valle,V., Secardin,L., Rameau,P., Bernard,O.A., Ishchenko,A.A., Abbes,S., Vainchenker,W. *et al.* (2016) TET2-mediated 5-hydroxymethylcytosine induces genetic instability and mutagenesis. *DNA Repair (Amst).*, **43**, 78–88.
79. Bulut-Karslioglu,A., De,La, Rosa-Velázquez,I.A., Ramirez,F., Barenboim,M., Onishi-Seebacher,M., Arand,J., Galán,C., Winter,G.E., Engist,B. *et al.* (2014) Suv39h-Dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell*, **55**, 277–290.
80. Kato,M., Takemoto,K. and Shinkai,Y. (2018) A somatic role for the histone methyltransferase setdb1 in endogenous retrovirus silencing. *Nat. Commun.*, **9**, 1683.