Department of Biochemistry

Faculty of Medicine

Universidad Autónoma de Madrid

Deciphering regulatory elements as determinants of cardiovascular diseases

PhD program in Molecular Biosciences

Jesús Victorino Santos

BSc Biotechnology, MSc Molecular Biomedicine

Doctoral thesis directed by Dr. Miguel Manzanares Fourcade

CNIC, Madrid 2021

Madrid, 6 January 2021

I hereby certify that JESUS VICTORINO SANTOS has carried out the experimental work leading to his PhD thesis entitled "*Deciphering regulatory elements as determinants of cardiovascular diseases*" under my supervision at the Centro Nacional de Investigaciones Cardiovasculares-CNIC in Madrid.

I also declare that the work presented is novel and of great importance in the field, and of sufficient quality to merit to be presented in order to obtain a PhD degree by the Universidad Autónoma de Madrid.

Miguel Manzanares

Centro Nacional de Investigaciones Cardiovasculares (CNIC)
Melchor Fernández Almagro 3
28029 Madrid, Spain
tel: (34) 91 453 12 00
mmanzanares@cnic.es
http://www.cnic.es/en/desarrollo/genomica

Centro de Biología Molecular Severo Ochoa (CBMSO), CSIC-UAM
Nicolás Cabrera 1
28049 Madrid, Spain
mmanzanares@cbm.csic.es
http://www.cbm.uam.es/functional-genomics

Ozren Bogdanovic, PhD
Laboratory Head Developmental Epigenomics
Garvan Institute of Medical Research
384 Victoria Street, 2010 Darlinghurst
Sydney, Australia

Associate Professor
University of New South Wales
Faculty of Science
School of BABS

P: +61 (0) 2 92958340
F: +61 (0) 2 9295 8101
E: o.bogdanovic@garvan.org.au
W: www.bogdanoviclab.org

Sydney, 29th December 2020.

RE: Report for Doctoral thesis "Deciphering regulatory elements as determinants of cardiovascular diseases" by Jesús Victorino Santos

To Whom it May Concern,

It is my great pleasure to provide a summary report for Jesús Victorino Santos's PhD thesis, which deals with the identification of novel regulatory elements associated with cardiovascular diseases (CVDs). CVDs constitute one of the most significant challenges in modern-day healthcare due to interactions between a genetic predisposition and environmental / lifestyle factors. Genome-wide association studies (GWAS) have become an increasingly popular approach to resolve the polygenic architecture of CVDs. This approach entails the identification of genetic variants that are present at a higher frequency in individuals with disease than in the healthy population. However, the clinical insights derived from GWAS results have remain limited to date. This is primarily because > 90% of GWAS-identified variants reside in non-coding regions of the genome and thus do not directly affect the coding sequence of a gene. These variants are frequently found in cis-regulatory elements (CREs) where they can disrupt transcription factor binding sites thereby altering levels of their target genes.

The research presented in Jesús's PhD thesis constitutes a major advance in our understanding of how various disease variants functionally impact upon cardiovascular disease (CVD) susceptibility. The candidate has utilised diverse model systems to provide mechanistic insights into the chromatin context and molecular function of a number of such variants by applying cutting edge *in vivo* transgenesis, chromatin conformation and genome editing assays. The thesis starts with a comprehensive introduction elaborating on the non-regulatory genome, GWAS approaches, and most common CVDs, which is followed by clearly described and well-framed objectives. The first results chapter deals with the optimisation of *in vivo* reporter assays for enhancer detection. This is followed by the application of such assays both *in vitro* and *in vivo* to functionally validate a number of risk variants associated with atrial fibrillation (AF). In the third chapter the candidate explores the convergence of genetic and electrophysiological (structural) changes during AF, utilising an ovine model and mouse

transgenic assays to unravel that electrical insults might down-regulate TBX5 and GJA1 genes resulting in further cardiac defects. Finally, the fourth chapter deals with the PSCK9 gene and the functional interrogation of its regulatory landscape in liver and brain tissues through *in vivo* transgenic assays. The thesis ends with a well-rounded discussion where the candidate has done a great job of integrating all the presented advances. I particularly liked the attempt to refine the regulatory networks of TBX5 and GJA1 genes, based on the novel data discussed above.

Altogether, throughout this thesis, the candidate has managed to demonstrate significant understanding of this complex topic, which is exemplified through well-executed experiments as well as appreciation of the existing literature and consideration for the existing body of knowledge. The thesis is concise, well-written, easy to read, and I have personally learned a great deal about this exciting topic just by reading it. The work presented here is of high quality, clearly demonstrating Jesús's capability to undertake independent research and to publish high quality work. I therefore strongly believe that the thesis merits award of the PhD degree.

Sincerely,

Ozren Bogdanovic, PhD
Associate Professor
Garvan Institute of Medical Research
Faculty of Science
University of New South Wales

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

Navneet Matharu, Ph.D.
Assistant (Adjunct) Professor
Associate Professional Researcher
University of California, San Francisco
Department of Bioengineering and Therapeutic Sciences
Innovative Genomics Institute

1550 4th Street,
Rock Hall 582
San Francisco, CA 94158-2911
Tel: (415) 476-5166
Fax: (415) 514-4361
Email: Navneet.Matharu@ucsf.edu

San Francisco, January 1st 2021

RE: Summary report for Doctoral thesis titled *'Deciphering regulatory elements as determinants of cardiovascular diseases'* prepared by Mr Jesus Victorino Santos

To Whom It May Concern:

With immense appreciation I am writing this summary report of Doctoral thesis titled *'Deciphering regulatory elements as determinants of cardiovascular diseases'* prepared by Mr Jesus Victorino Santos under the direction of Dr. Miguel Manzanares Fourcade, CNIC, Madrid, Spain.

Jesus PhD work has demonstrated very high standard of scientific caliber about the subject of gene regulatory elements. He has interrogated functional aspect of non-coding genome that has been associated with the risk of two most important cardiovascular diseases (CVD), atrial fibrillation (AF) and atherosclerosis. In my opinion this work presents a framework to decipher function of GWAS risk variants for hundreds of other diseases. The impact of his work is far-reaching for other researchers who are interested in investigating the function of non-coding regulatory elements of the genome associated with disease mechanism.

Jesus has structured his thesis in a very candid flow with lucid writing, making it a thoroughly enjoyable experience to read. He has written very comprehensive and well referenced introduction with historic perspectives including state of the art tools and technologies that has advanced our understanding towards assigning function to the previously called 'junk' DNA. With this motivation Jesus presented a refreshing new approach to assign *'func of junk'* (function of junk). I admire the way Jesus has described each and every aspect of non-coding element very concisely in his introduction section.

In the first chapter Jesus developed and improved the *in vivo* enhancer reporter assay leveraging the piggyBac transposition system and named in PB-ERA system. Applying highest standards of scientific rigor (tested more 200 mouse embryos with 30 independent reporter assay constructs) he has shown that PB-ERA system yields very high transgenesis rate when compared to conventional methods used in mouse model. Jesus demonstrated the validity of this system by testing known enhancers in PB-ERA system and yield high reproducibility.
In the second chapter Jesus demonstrated the strength of this PB-ERA system in interrogating AF risk loci. Jesus, intelligently cross-validated his results with many other approaches like, deletion analysis using CRISPR/Cas9 system, transient enhancer reporter assays, 3C chromosome conformation capture analysis along with chromatin histone marks annotations. He prioritized 10 AF risk loci among 130 based topological associated domain interactions, histone marks and cardiac transcription factor binding motifs like GATA4, TBX5 and NKX2-5. Jesus diligently chose the relevant model system like HL-1 cells that are known to express all these cardiac factors for reporter assay. He assigned the functional

activity of GWAS risk loci CAV1, C9ORF3 and SYNE2 and even confirmed the biochemical histone marks and TBX-5 binding potential for the candidate enhancer regions. Of particular interest is the SYNE2-AF locus that might contain an alternative promoter for the short isoform which is specific to heart and skeletal muscles. 3C studies show that SYNE2-AF interacts with the long isoform promoter too and regulate SYNE2 is atria and ESR2 gene in lung. In this chapter Jesus also looked at a very interesting locus KCNIP1 which has CNV (copy number variant) in its first intron that is positively correlated with the KCNIP1 expression levels. Although this CNV does not show any indicative enhancer marks but did show robust enhancer activity in the PB-ERA assay system highlighting the gap in our understanding about biochemical signatures as predictive marks for enhancer function. Jesus applied CRISPR deletion analysis to assign the target gene promoter for CAV1-AF locus. Upon deletion of this candidate enhancer the expression of Cav1 and Cav2 gene got downregulated, which corroborated well with the other experiment where the CAV1-AF risk allele lowers its enhancer activity. Jesus found the CAV1-AF locus has two major modules of enhancer activity spread over 10kb region. In previous studies it is suggested that during evolutionary expansion of mammalian genome that non-coding regulatory code also expanded over large DNA segments.

Quite interestingly, Jesus deciphered the silencer activity of ZFHX3-AF locus using the PB-ERA system. The candidate region he picked has predominantly, H3K27me3 mark in all other tissue except in aorta where it is expressed. This ZFHX3-AF region did not show any enhancer activity in the PB-ERA system but interestingly Jesus observed the reduction in the ectopic expression of LacZ using this construct, which he observed for other positive candidates. This led him to validate its activity using CRISPR/Cas9 deletion analysis, upon which he found upregulation of ZFHX3 gene. He performed the 'enhancer blocker assay' generally used to assay insulator elements. Jesus designed this assay with utmost smartness and caution by changing its cloning position with respect to the enhancer element. Based on these experiments Jesus interpreted it as silencer and not insulator as this element could exert its repressive effects at any position even downstream or upstream to the enhancer-promoter-LacZ cassette. Importantly Jesus elucidated not only the role of silencer elements in AF disease risk but also highlighted the difference in the functional activity of ZFHX3-AF locus between mouse and human explaining some of the discrepancies in previous studies.

In his next chapter Jesus took advantage of chronic AF induced ovine transcriptomic data and overlayed it with the 130 AF risk loci and found 209 genes to be differentially expressed between left and right atrial appendages. Jesus prioritized top 4 genes namely, GJA1, TBX5, JMJD1C and FKBP7 for his further studies. Jesus linked nicely the TBX5 enhancer candidate region responsible for its upregulation and risk allele compromise this enhancer activity leading to downregulation of TBX5 which is cardiac regulator of many other risk loci. The search for GJA1 enhancer candidate is an interesting and quite a challenging task that Jesus undertook considering the fact that most of its risk variants lie in its gene desert region. Jesus systematically addressed this by parsing out the 20kb candidate region into three independent PB-ERA constructs. He found the GJA1-AF locus enhancer activity is spread over 20kb region which he termed as 'enhancer block' comprising of multiple binding sites for TBX5, GAT4 and NKX2-5, especially in its 600bp core region. He also could accomplish the 20kb deletion of this region to validate its effect on GJA1 transcription and not on any other distal neighboring genes. Jesus hypothesize that GJA1 transcription is TBX5 dependent and hence AF triggers that downregulate TBX5 also downregulate GJA1 expression. In this chapter Jesus shed light on the TBX5 mediated regulatory axis of AF risk.

The fourth chapter of this thesis is extremely interesting and holds huge potential not just to decipher the role of PCSK9 in brain apart from its role in lipid metabolism and atherosclerosis, but also as a potential therapeutic target for neurological conditions. In search of the cerebellum specific regulatory mechanism Jesus deconstructed out the landscape of PCSK9 gene and identified discrete enhancer regions regulating its expression in liver and cerebellum separately.

Jesus has written a well-balanced and extremely relevant discussion section highlighting both strengths and concrete future directions warranting in-depth investigations.

With such an extensive and in-depth work, in my opinion Jesus's effort is an exceptional scientific endeavor. The quality of presentation of the data is of a very high quality and the grasp on the subject is extremely exhaustive. Jesus has been able to put forth the framework for interrogating the disease risk loci and demonstrated it by generating the 'CVD specific gene regulatory network'. Jesus has been quite successful in combining many existing data sets and been able to provide meaningful explanations to many previously reported discrepancies through his stellar work. Through his thesis work Jesus has proved himself to be a seasoned scientist of top caliber and with full enthusiasm I endorse his work to be highly worthy for the award of PhD.

Sincerely,

Navneet Matharu

A mis padres.

Bien está que todos los hombres coman,
pero que todos los hombres sepan.

Federico García Lorca.

Acknoledgements

El primero de la estirpe está amarrado en un árbol y al último se lo están comiendo las hormigas.

Gabriel García Márquez, *Cien años de soledad*.

¿Quién me iba a decir que una tarea difícil del doctorado sería la de enfrentarse al folio en blanco de los agradecimientos? Sin hacerlo tremendamente aburrido, uno empieza por el principio y va recordando momentos y personas sin las cuales no habría llegado hasta el final. Pero, ¿dónde empieza una tesis? Difícil pregunta. Han pasado casi 6 años desde que fui a visitar el labo por primera vez. Un día en el que casi parecía famoso ante tanta energía y expectación por el nuevo miembro. Hoy siento que aquel día fue una buena degustación de los siguientes que vendrían.

Empezando por el principio, me gustaría agradecer a Miguel Manzanares que me eligiera para formar parte del grupo durante el máster y confiara en mí para hacer el doctorado. Sobre todo, gracias por ser un ejemplo a seguir tanto dentro como fuera del labo, por hablar tan bien en público y por dejarme ser y hacer a mi manera. Aunque el teletrabajo ha hecho que la puerta del despacho lleve casi un año cerrada, siempre estuvo abierta y no solo para “resultados espectaculares” sino también para concretar los “ya vamos viendo”. Gracias por dejarme crecer entre proyectos, becas, congresos, cursos, estancias e incluso residencias. He madurado mucho profesional y personalmente estos años, y ha sido un placer haber aprendido de ti.

La primera imagen que recuerdo del labo es de Teresa y Sergio emocionados con la llegada del nuevo. Creo que temí no estar a la altura del recibimiento. Pero todos me acogisteis muy rápido y el sentimiento de pertenencia fue inmediato. Estas líneas van para el laboratorio en general porque en mi primera (de muchas) crisis existencial(es), sabía que doctorado o no, lo que yo quería era seguir entre tanta buena gente.

Teresa, has sido y sigues siendo una de mis científicas y personas referentes. Siempre me he sentido apadrinado por ti y he agradecido tus millones de consejos en momentos importantes como laCaixa, la Arquímedes o mis proyectos frustrados. Estoy flipando al darme cuenta de que realmente coincidimos ¡menos de un año en el labo! Te admiro y espero seguir aprendiendo de ti.

Melisa, la otra veterana, siempre recordaré lo que significa “cadillo” gracias a ti. Sigo conservando la taza de Snoopy que me regalaste mi primera semana y, por supuesto, cómo olvidar que mirar en la lupa con los dos ojos te lo debo a ti y a tu boli bic. Al escucharte en el *alumni meeting* confirmé que nadie da charlas como tú.

Julio, sin duda una de las patas de esta tesis. No solo le has dado ingenio y color a mis días en Madrid, sino que has sido amigo y consejero. Gracias, sobre todo, por

hacerme practicar entrevistas y charlas, por leerme la mente, por los cafés de emergencia, por la fiesta y por estar siempre disponible para discutir sobre ciencia. Por ser claro y directo. Por llevarme a conocer el Teatro Real, por el *punting* en Cambridge y, por qué no, por nuestras discusiones.

Sergio, si he llegado hasta el final también es gracias a ti. La otra pata, muy flamenca pero más serena de esta tesis. Has sido un ejemplo de que el trabajo bien hecho, los viajes, el teatro y los amigos caben todos en una tesis. Gracias por estar ahí siempre que lo he necesitado, por no dejar pasar ni una celebración, por nuestras Ferias del CNIC y por venir a la de verdad. Por darme el mejor viaje de mi vida y por hacer que quisiera pasar más tiempo en el laboratorio.

En este proyecto han participado muchas personas, todas distintas, que han aportado trabajo, resultados, profesionalidad y ganas. ¡Ay, Isa, y qué sería de esta tesis sin ti! Sin tus ánimos y tus microinyecciones, sin que nos liáramos la manta a la cabeza con mil experimentos y especulaciones. Cuántas veces no habremos discutido los mismos resultados convenciéndonos el uno al otro y reconvenciéndonos al revés. Me has acompañado todo el camino y no sabes cuánto me alegro de que así haya sido. Claire, gracias por dejarme enseñarte lo poquito que sabía en aquel momento. Por tu interés y tus ganas, tu paciencia con mi “organización”, tu independencia y por enseñarme más tú a mí que yo a ti, gracias. Nunca te perdonaré que dijeras “yes, boss” cada vez que Miguel pasaba cerca ni que sólo cantaras en cultivos cuando yo no estaba y me enterara el último de tus dotes artísticas. Creo que nunca lo he pasado tan bien haciendo experimentos como contigo. Javi, mi otro compañero de batalla, te quiero dar las gracias por tu predisposición y tus ganas de aprender, por ayudarme a ir a cuatro manos, por las risas, los clonajes, las clases de Kung Fu y enseñarme los principios básicos del mundo interior. Aunque no hace ni un año que te fuiste parece que ha pasado una eternidad.

He aprendido que un laboratorio es como un río, que puedes reconocer después de muchos años, pero en el que nunca te puedes bañar dos veces porque va cambiando, unos vienen y otros van. Entre todos habéis hecho que disfrute el remojón incluso cuando venían nuevas corrientes. Mariajo, empezamos (y nos vamos) casi a la vez. Gracias por escucharme, por ser tan decidida y enseñarme con tu ejemplo a perseguir las cosas hasta el final. ¡Y por los calçots! Me alegro de que hayamos podido compartir el camino en su parte agria y en la dulce. Marta, llenas el laboratorio de

alegría con tu sonrisa. Gracias por hacerme pensar tanto y traerme ideas nuevas, por lo sostenible y lo vegetariano, por ponerle ilusión a lo que haces. Aprendo mucho de ti. María, intentaste alejarte de las garras del “jefesito” pero al final sucumbiste. Muchas gracias por ser mi aliada en la zona postdoc, por saber tanto, por echar una mano siempre y por aguantar mis lamentos. Antonio, gracias por el interés que pones a todo, por ser una enciclopedia andante y hacer tan buenas tortillas. Alba, haces que el labo sea más completo y crítico. Gracias por aguantar siempre hasta el final de la fiesta, por enseñarme el LL y por no dejarme conducir tu coche. Aurora, mi principio fue tu recta final y ahora que estoy acabando has vuelto para empezar de nuevo. Gracias por tus visitas a la 3Sur, por el Sector 3 y por el Polo Norte. Gracias, Raquel, por tu perseverancia; Claudio, por tu meticulosidad; Marina, por tu ayuda. Gracias a Gonzalo por ser médico y científico, y por contar conmigo en un proyecto tan interesante. Quiero agradecer al resto de estudiantes que han pasado por la familia Manzanares estos años, de todos he aprendido y disfrutado mucho. Reencontrarme con Marcos y Antonio López siempre me recuerda a la etapa en la que fuimos casi quince personas en el labo y lo divertida que fue.

Me gustaría dar las gracias a todas las personas con las que he coincidido por el CNIC, haciendo que la vida durante el doctorado sea más fácil y también divertida. A los compañeros de los antiguos departamentos y actuales áreas, a los mágicos reactivos comunes que se fueron extinguiendo y a los prehistóricos *retreats* de departamento que son el mejor *networking* que existe y que tuve la oportunidad de vivir, aunque sólo fuese el último de ellos, antes de que diversión y ciencia dejaran de ser compatibles en el CNIC. A las unidades técnicas del CNIC, especialmente a Giovanna y Elisa por su ayuda continuada y a Elena Prieto, Mariano y la unidad de celómica en los apretones finales. A las últimas sesiones de transgénesis de la unidad que me han permitido acabar. A los co-habitantes del cuarto de cultivos por compartir horas de sufrimiento y bandas sonoras. A Sergio Callejas por arreglar todo el rato la máquina de qPCR y Alberto por discutir sobre genómica. A Miguel Torres por ser el único que me preguntaba en los seminarios. Gracias a todos los del ala que dan vidilla a los días. A Jose Ángel que siempre tiene una reverencia. A Diana, que siempre sabe dónde hay que tirar los residuos. A María Galardi por las Maxis y los consejos para hacer inmunos. A Carla, Elías, María Rosaria, Cris(es), Silvia, Mariya, María, Alejandra, Geo, Sandra y a todos los demás. Gracias Sonia (y el resto de *lab*

managers y técnicos de logística) por facilitarme la vida con mi mayor enemigo: la burocracia; y por apuntarte a un bombardeo. Gracias al personal de cocina y cafetería (¿existió alguna vez la vida pre-pandemia?): Loli, Ángel, Elena, Jose, Charo, por saberos nuestros nombres, ser tan amables y hacer unas lentejas tan buenas. ¡Cuánto os echamos de menos! Al personal de limpieza y a Soriana porque no se puede ser más simpática. Muchas gracias a Ángel y los compañeros de almacén, porque él no lo sabe pero me ha traído tantos tubos de oligos que darían la vuelta al mundo si los pusiéramos en fila.

Gracias a las cañas de los miércoles, a las de los viernes, a los eventos sociales extraoficiales y a los dulces cumpleaños. Muchas gracias a los que me habéis acompañado en alguna parte del camino: Héctor, Rocío, Javito, Eli, Jose, Ana, Laura, Verdiana, Cris, Rebeca, Carles, Alberto, Sara, Fran, Maruchi, Itzíar y muchos otros. A Macarena, con quien he ido a la par desde la carrera. A las visitas de Briane a la 3Sur o las ocurrencias de Wen. Gracias a los que estuvisteis cuando comenzó la aventura en Madrid: a mis promociones de Cicerone y Máster; al club de Dory; gracias a Andrea por ser como una hermana (y por enseñarme catalán en sus largas conversaciones telefónicas con su madre), y al resto del piso (Jose, Álvaro y Anna) por nuestros momentos televisivos; a Inés por ayudarme en los inicios; y a Raúl por conectar desde el principio.

I also want to thank Nadav Ahituv for opening his lab to me, for letting me learn from amazing projects and get to know great people; and, of course, many thanks to Taka for your time, expertise and help with MPRAs. Special thanks to the people how make me miss those days in SF: Jingjing, for being so unique and fun; Navneet, you don’t know how much I learnt from you; Picard, for all the memories I keep from our trips and our foosball between experiments; to Jacob, Xujia, Serena, Ajuni, Nadjia, Aaron, Hongtao and the rest of the lab.

San Francisco ha sido un lugar muy importante en esta tesis gracias a la inmensa ayuda de Fran. Gracias por acogerme y compartir tu estancia con la mía, por presentarme a los “old dogs” y al gran Alexis. Gracias por los grandes momentos, por la ciencia y la convivencia. Muchas gracias, Laura, por hacer que tanta coincidencia se hiciera realidad. Por los cafés y las largas conversaciones. Alexis, merci beaucoup d'avoir fait mon séjour plus beau et pour nos aventures.

La recta final ha sido la más gratificante. Debí haberlo supuesto ya que “el último año todo es mágico”, como dice Sergio. Esta recta final ha coincidido con otro momento muy especial en mi vida, en el que he conocido a personas increíbles que le han dado tres o cuatro patadas a los muebles que tenía en la cabeza. A ellos, a los becarios de la generación del ‘19, no les puedo estar más agradecido. Coincidir en la Residencia de Estudiantes, ha sido el comienzo de amistades, proyectos y nuevas pasiones que me han hecho muy feliz. Muchas gracias a Fran (por hacerme disfrutar más oyéndote que hablando), Jorge (tengo debilidad por ti), Juan y Laura (por ser mis comunicadores de referencia), Elena (por ser tan dulce y descubrirme a Marjane Satrapi), Ángela, Álvaro, Miguel, Raquel, Sara, Andrea, Violeta y, por supuesto, también a los adjuntos Antonio y Guzmán. No sólo aparecisteis en el momento idóneo, sino que me habéis hecho darme cuenta de que sois imprescindibles; sin dudarlo “os sigo”. Por supuesto, gran parte de la culpa de que guarde tantos buenos recuerdos de esa etapa la tienen todos y cada uno de los trabajadores que hacían nuestra vida más fácil (incluso Jose). Muchas gracias a todos, especialmente a mis queridas Pepa y Gema y al gran Marcelo.

Violeta, muchas gracias por volarme la cabeza con tu forma de ver el mundo, por aportarme frescura y desbloquear partes que antes eran inaccesibles para mí. Gracias por tu comprensión, por regarme mientras era una planta que escribía la tesis y porque la incertidumbre del futuro para ti es una aventura. Por acercarme el cine y la literatura, por tu interés por saber qué son los *enhancers* y enseñarme la importancia de practicar. Gracias por ser un apoyo muy importante y conocerme tan bien.

Andrea, espero que tu estupidez de irte tan lejos justo antes de una pandemia mundial se pueda solucionar pronto. La admiración que te tengo se mezcla con el cariño y lo nubla todo. Gracias por hacerme disfrutar tanto cuando estamos juntos, por hacer que valore muchísimo la co-existencia, por poder hablar de todo contigo y por enseñarme a mantener el contacto.

Sevilla es, sin duda, la otra localización geográfica de esta tesis. Mi ciudad natal, donde crecí y estudié y, sobre todo, donde más me equivoqué. Me gustaría agradecer a mis amigos del barrio, por lo que representan para mí y recordarme mis raíces, especialmente David, Josué y Abraham (todo muy bíblico, sí). Y por supuesto, Natalia, mi compañera de estudios durante tanto tiempo que ya no podría hacer las cuentas.

Siempre has sido un ejemplo de perseverancia. Enrique, Adri y Ángel, gracias por disfrutar cada segundo de nuestros reencuentros como si hubieran parado el reloj entre una visita y la siguiente. A mis queridxs biotecnólogxs por lo mucho que me enseñaron en la Olavide. A Javi, María(s) y Pablo, por enseñarme que, como el vino, algunas amistades pueden envejecer mejor. Muchas gracias a Rafa Daga y su laboratorio, especialmente a Manolo Bernal, porque en el CABD fue donde empecé a degustar la ciencia en su mejor formato. A mis compis de Atlánticus agradezco que me acompañaran en aquella aventura por el desierto y a Mary O'Connell que me dejara hacer mis pinitos en genética de plantas, además de sus innumerables cartas de recomendación.

Y como no podía ser de otra forma, uno no puede renunciar a sus orígenes. Tan fácil como que yo no estaría aquí de no ser por mis padres. Y no me refiero al experimento más obvio, ese que dura nueve meses y que no hace falta ser biólogo para saber de qué trata. En un entorno no tan favorable, con medios escasos y siempre con trabajo y honradez, mis padres me enseñaron el valor del esfuerzo y la importancia de los estudios, siendo claves para que mis hermanas y yo accediéramos a la Universidad. Es gracias a vosotros que existe esta tesis doctoral.

Muchas gracias a mis hermanas por haberme aguantado toooodo este tiempo. Jessi, siempre fuiste el ejemplo a seguir, el camino ya trazado. Para mí fue más fácil porque podía seguir tu rastro. Gracias por cruzar el charco para venir a verme. Ana, gracias por tus detalles y tu inocencia; espero que pronto podamos volver a acercarnos. A mis tíos, primos y abuelos, gracias por estar siempre ahí, por no caber en los actos de graduación y por saber formar un sarao con cuatro palmas.

Sobre todo, muchas gracias a los que os atrevéis a ser diferentes y nos enseñáis con vuestro ejemplo.

A todos, de todo corazón, ¡GRACIAS!

Summary

The non-coding genome harbors cis-regulatory elements (CRE) that control gene expression in time and space. A tight control of transcription is of great importance, especially during development, and CRE disruption may lead to malformations and other congenital diseases. Genome-wide association studies (GWAS) have identified common polymorphisms associated to multifactorial disorders in humans such as cardiovascular diseases. The vast majority of these associations lay in non-coding regions. Whether these thousands of risk loci affect CREs and have a functional role in the context of disease is unknown. Cardiovascular diseases (CVDs) are common human diseases with the highest prevalence and death rate worldwide. To date, GWAS have linked hundreds of loci to a higher risk of developing two major CVD: Atrial Fibrillation (AF) and Atherosclerosis. CVDs are not an exception and for most risk loci we lack mechanistic insights into the nature of GWAS associations.

Although enhancer-reporter assays (ERAs) are a powerful tool to characterize risk-associated enhancers, these experiments are time-consuming and the throughput is very limited. This is in stark contrast with the outgrowing number of new polymorphisms associated to human diseases. In this thesis, we optimized current mouse ERA technology to achieve ~59% efficiency of transgenesis, thus enabling the scaleup of CRE discovery. We systematically interrogated a dozen risk loci strongly associated to AF in the search for disease-risk enhancers. Interestingly, we showed that the PB-ERA system that we developed is able to identify negative regulators such as silencers or insulators. Together with 3D chromatin analysis and CRISPR-mediated perturbations, we identified the targets of AF-CREs and involved new genes in arrhythmia susceptibility. Furthermore, we integrated transcriptomic data from an ovine model of AF chronification. We found that GWAS and chronification data converge on the *TBX5-GJA1* axis and identified AF-enhancers regulating the cardiac expression of both genes. These enhancers are controlled by TBX5 itself in what might be a key feedback-loop for atrial remodeling.

Last but not least, we applied our approach to a second CVD to validate it as an effective framework to understand the genetic contribution to human diseases. We interrogated the locus of the pro-atherosclerotic gene *PCSK9* and describe a dual regulation for this gene in liver and cerebellum.

El genoma no codificante contiene elementos reguladores que controlan la expression génica en el tiempo y el espacio. El control de la transcripción debe ser muy preciso, especialmente durante el desarrollo embrionario, donde la alteración de estos elementos reguladores puede dar lugar a malformaciones y otras enfermedades congénitas. Los estudios de asociación del genoma completo (de sus siglas en inglés, GWAS) han identificado polimorfismos comunes asociados a enfermedades multifactoriales como pueden ser las afecciones cardiovasculares. La gran mayoría de estas asociaciones residen en regiones no codificantes. Sin embargo, se desconoce si estas miles de regiones de riesgo están afectando a elementos reguladores y, por tanto, pueden tener un papel relevante en enfermedad. Las enfermedades cardiovasculares son comunes en humanos y tienen la mayor tasa de mortalidad. Los estudios de GWAS han asociado cientos de regiones genéticas a un mayor riesgo de padecer fibrilación auricular y arteriosclerosis, dos de las enfermedades cardiovasculares más relevantes, desconociendo los mecanismos que subyacen a estas asociaciones.

Los ensayos de *enhancer-reporter* (de sus siglas en inglés, ERA) son una herramienta muy útil para caracterizar *enhancers* en regiones genéticas asociadas a enfermedad. Sin embargo, estos experimentos son muy lentos, lo que limita su rendimiento. Esto contrasta con el creciente número de nuevos polimorfismos que cada año se asocian a enfermedades. En esta tesis doctoral, hemos desarrollado un ensayo optimizado de ERA en ratones que genera una eficiencia de transgénesis del ~59% y que nos ha permitido escalar la identificación de nuevos elementos reguladores. En total, hemos interrogado una docena de regiones genéticas asociadas a fibrilación auricular en busca de *enhancers*. Cabe destacar la capacidad del sistema que hemos desarrollado, y que hemos denominado PB-ERA, para identificar elementos reguladores negativos, como silenciadores o aisladores. Además, por medio de analizar la estructura de la cromatina y de editar el genoma hemos identificado los genes diana de estos elementos reguladores, involucrando nuevos genes en la predisposición a padecer arritmias cardíacas. Integrando datos transcriptómicos de un modelo ovino para la cronificación de la fibrilación auricular junto con genes identificados por GWAS, hemos descubierto que comparten el eje *TBX5-GJA1*, genes para los que hemos identificado *enhancers* y que podrían estar regulados por el propio *TBX5*.

Por último, hemos aplicado nuestro abordaje a una segunda enfermedad cardiovascular, para validarlo como un marco de referencia en el estudio del componente genético de enfermedades comunes. Para ello, hemos estudiado el locus del gen *PCSK9*, implicado en arteriosclerosis, descubriendo elementos reguladores que controlan su expresión específicamente en hígado y cerebelo.

Index

List of figures

List of acronyms

3C	chromosome conformation capture
3D	three-dimensional
AF	atrial fibrillation
bp	base pair
CAD	coronary artery disease
ChIP	chromatin immunoprecipitation
CNV	copy-number variation
CRE	cis-regulatory element
CRISPR	clustered regularly interspaced short palindromic repeats
CVD	cardiovascular disease
DE-*Oct4*	distal enhancer of *Oct4*
eQTL	expression quantitative trait loci
E	embryonic day
ERA	enhancer-reporter assay
GWAS	genome-wide association study
H3K27ac	acetylation histone H3 lysine 27
H3K27me3	trimethylation of histone H3 lysine 27
H3K4me1	monomethylation of histone H3 lysine 4
H3K4me3	trimethylation of histone H3 lysine 4
Hi-C	chromosome conformation capture followed by sequencing
hiPSC	human induced pluripotent stem cells
kb	kilobase
LA	left atria
LAA	left atrial appendage
LD	linkage disequilibrium
LV	left ventricle
Mb	Megabase
mESC	mouse embryonic stem cells
MPRA	massively-parallel reporter assay
PB	piggyBac
PB-ERA	piggyBac-based enhancer-reporter assay

Pol II	RNA polymerase II
qPCR	quantitative PCR
RA	right atria
RAA	right atrial appendage
RV	right ventricle
SNP	single-nucleotide polymorphism
TAD	topologically associated domain
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcription start site
UTR	untranslated region

Introduction

Can you not understand, Winston, that the individual is only a cell?
The weariness of the cell is the vigour of the organism.
Do you die when you cut your fingernails?

George Orwell, *1984*.

The Human Genome Project published the first draft of the entire human genome almost twenty years ago (Lander *et al.*, 2001). Far away from having millions of genes as preliminary estimates from the sixties indicated (Vogel, 1964), humans turned out to be 'between a chicken and a grape' in terms of gene count with a total of approximately 22,000 protein-coding units (Pertea and Salzberg, 2010). This number of genes, which occupies less than 2% of the total DNA sequence, was lower than expected and left scientists looking at that huge ~3,000-Mb portion of non-coding genetic information and wondering 'how much junk, how much func' (**Figure 1**) (Castillo-Davis, 2005).

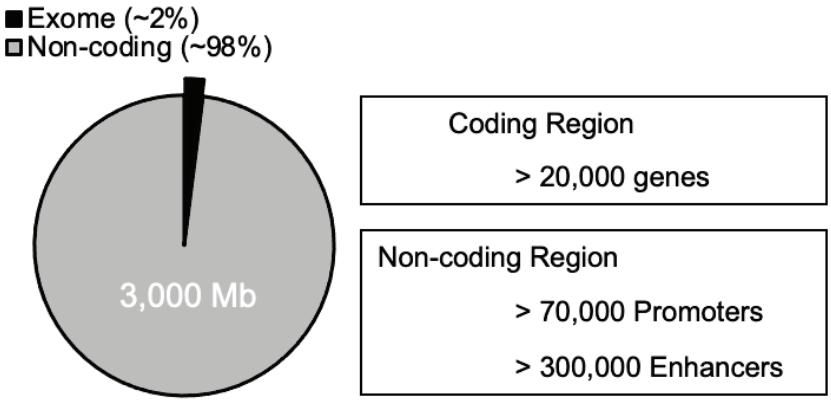

Figure 1 – Knowing the unknown: a functionally active non-coding genome. Most of our genetic information does not encode for proteins. Far from being non-functional, the ENCODE project has annotated thousands of promoters and enhancers which far exceeds the number of genes (data from Encode Project Consortium, 2012).

1. The regulatory genome.

No longer seen as 'junk DNA', the non-coding genome has been shown to harbor multiple types of modules with regulatory function. These cis-regulatory elements (CREs) control gene expression during development and homeostasis, and their disruption can lead to disease (Smith and Shilatifard, 2014; Lupiáñez, Spielmann and Mundlos, 2016; Schoenfelder and Fraser, 2019). While promoters are at the 'beginning' of the gene, near the transcription start site (TSS) from where they initiate transcription, other CREs are located distally and interact with promoters mediating the coordinated regulation of gene expression (Schoenfelder and Fraser, 2019). A key feature of CREs is their tissue specificity, which helps the organism to shape their body plan during development by switching on and off key genetic programs (Rickels and Shilatifard, 2018). Additionally, CREs are conserved across different species

(Villar *et al.*, 2015). Therefore, sequence conservation has been used to find CREs and indicates the importance of some of these regulatory elements from an evolutionary perspective. However, conversely to genes, we lack a regulatory genetic code which hinders the identification and characterization of CREs.

1.1. Enhancers, silencers & insulators.

Genetic programs change dynamically during differentiation, which requires a coordinated regulation of gene expression. There are several types of CREs that contribute differently to the transcriptional regulation of genes. Enhancers are, by far, the most studied type of CRE. The earliest studies identifying transcriptional enhancers in the 80s, defined them as DNA sequences that increase gene expression and can act in either orientation at many positions even downstream from the TSS (Banerji, Rusconi and Schaffner, 1981; Moreau *et al.*, 1981); a definition that is still widely used. Despite the ubiquitous activity of the SV40 viral enhancer described in the first studies, mammalian enhancers are normally tissue-specific and, therefore, they do not boost transcription in every cell type (Buecker and Wysocka, 2012). Instead, the classical developmental enhancers that have been identified are usually responsible for regional gene expression (**Figure 2**), where genes with complex expression patterns are the result of the activity of several enhancers (Pennacchio *et al.*, 2006).

Conversely to the boost in transcription caused by enhancers, there are also CREs that prevent gene expression or decrease it. Silencers are negative transcriptional regulators which also interact with promoters, in this case, repressing gene expression. Similar to enhancers, silencers can act in a tissue-specific manner and are independent from orientation (Pang and Snyder, 2020). However, most studies and functional approaches focus on enhancers and, thus, much less is known about silencers. The former is very surprising, especially if we take into account that we have known that silencers exist for as long as we have known about enhancing sequences (Brand *et al.*, 1985). In this occasion, the first silencer was characterized in the yeast genome (Brand *et al.*, 1985), after what mammalian examples of transcriptional silencers were also discovered affecting the *Ins1* and *Gh1* rat genes (Laimins, Holmgren-Konig and Khoury, 1986; Larsen, Harney and Moore, 1986). Despite silencers are also key for human cell differentiation and lineage specification (Sawada

et al., 1994; Donda *et al.*, 1996), they are largely understudied, possibly because of current methodology favoring the detection of enhancer-mediated upregulation (Doni Jayavelu *et al.*, 2020).

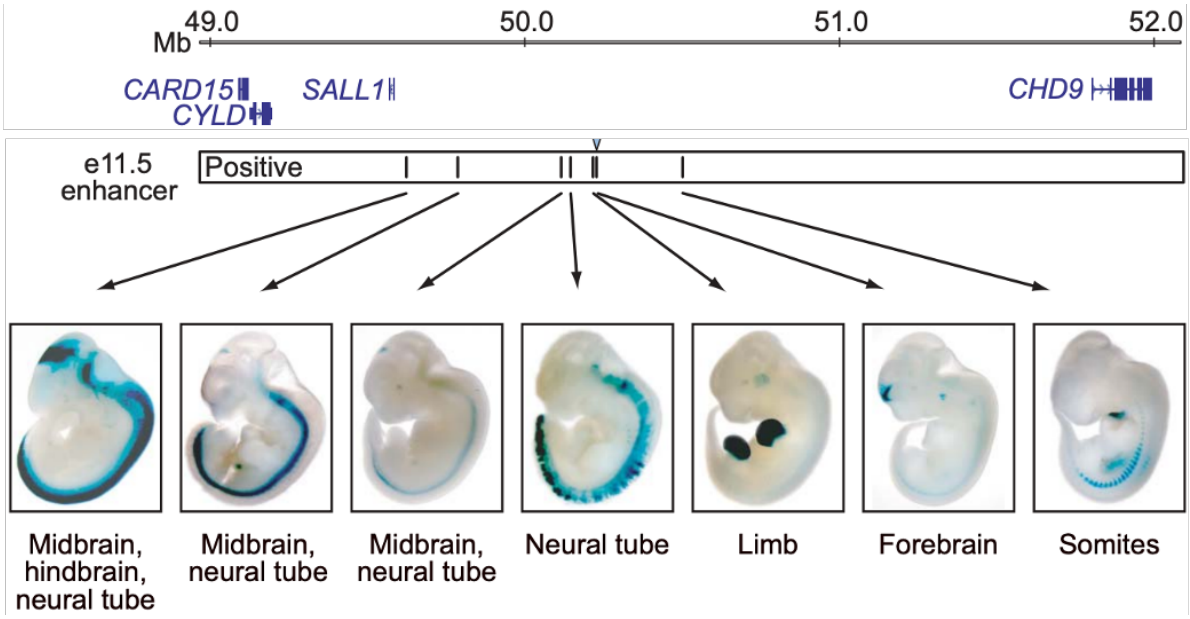

Figure 2 – Multiple regulatory elements account for complex patterns of gene expression. Overview of different tissue-specific enhancers concentrated in a large non-coding region of the human chromosome 16 near the *SALL1* gene, each of which recapitulates part of the expression domain of the endogenous gene in mouse (modified from Pennacchio *et al.,* 2006)

A third class of CREs are insulators, genomic regions that separate chromatin domains functionally. Also known as boundaries, these genomic regions firstly described in the fruit fly genome are able to establish domains of independent gene activity and insulate transgenes against chromosomal position effects (Udvardy, Maine and Schedl, 1985; Kellum and Schedl, 1991). The role of insulators in gene regulation is very important, since they constitute a mechanism to ensure specific gene expression patterns during development and lineage-specification. By preventing differentially regulated regions from interacting, they are also avoiding external regulatory elements such as enhancers or silencers to influence the expression of the genes within their boundaries. For instance, insulators play an important role modulating regulatory changes in response to environmental cues, such as reduced oxygen levels (Tiana *et al.*, 2012). On the contrary, by establishing the limits of genetic regions, insulators are also promoting interactions between genes and CREs within the boundaries. This can be especially useful to facilitate coordinated gene expression of several genes that would be respondent to the same regulatory elements (Capelson and Corces, 2004; Gaszner and Felsenfeld, 2006). Apart from preventing the genetic

communication between different regulatory domains of the chromatin, insulators are also involved in creating a barrier against the spread of heterochromatin (Giraldo *et al.*, 2003; Gaszner and Felsenfeld, 2006).

1.2. Transcription factor-mediated activity and specificity.

The activity of enhancers, silencers and insulators in a specific tissue relies on the presence of defined transcription factors (TFs) that bind their sequence and interact with RNA polymerase II (Pol II), other TFs and cofactors (**Figure 3**)(Stees *et al.*, 2012; Meng and Bartholomew, 2018; Andersson and Sandelin, 2020). Therefore, the information encoded at the non-coding regulatory DNA is exerted by TFs. For a given enhancer, e.g. a cardiac enhancer active in the heart and inactive in the limb, specificity is sustained by TFs present in the heart that are absent from the limb. Consequently, one could think that what truly makes the enhancer active is solely the presence of the single TF that binds the enhancer. If this were the case, ectopic expression of the functional TF in the non-cardiac cell would turn the enhancer active. Although some heart enhancers regulated by TBX5 can become active in human embryonic kidney (HEK) cells ectopically expressing TBX5 (Nadadur *et al.*, 2016), this is not always the case. It implies a reductionist mechanism that is not always true since not only the interaction of many TFs but also the history of the cell and the subsequent epigenetic footprint will impact the final outcome (Charest *et al.*, 2020). It is the interaction of all TFs and the integration of internal and external cues that will ultimately determine gene expression at the particular locus, and the coordinated regulation of the genome what results in the transcriptome of a specific cell.

Cardiac TFs such as TBX5, GATA4 and NKX2-5 are essential during heart development. These TFs regulate genetic programs responsible for the formation of key cardiac structures such as the chambers or the cardiac conduction system. Interestingly, these cardiac TFs coregulate many genes and often bind together to the promoter and regulatory elements of their common targets (Bruneau, 2013). This means that many of the cardiac enhancers that are involved in heart development depend on the presence of TBX5, GATA4 and/or NKX2-5 to be active and suggests a high degree of cooperativity between cardiac TFs.

However, different TFs might influence the same CREs in an opposite manner. Since it is not very clear how TFs compete for enhancers or how they interact, the effects of altering TF availability are hard to predict. In this context, the activity of different combinations of synthetic enhancers has been assessed systematically in order to explore the logic behind TF interactions (Smith *et al.*, 2013). However, while a few principles of TF cooperativity have been observed, heterotypic TF interactions remain poorly understood (Smith *et al.*, 2013; Luna-Zurita *et al.*, 2016).

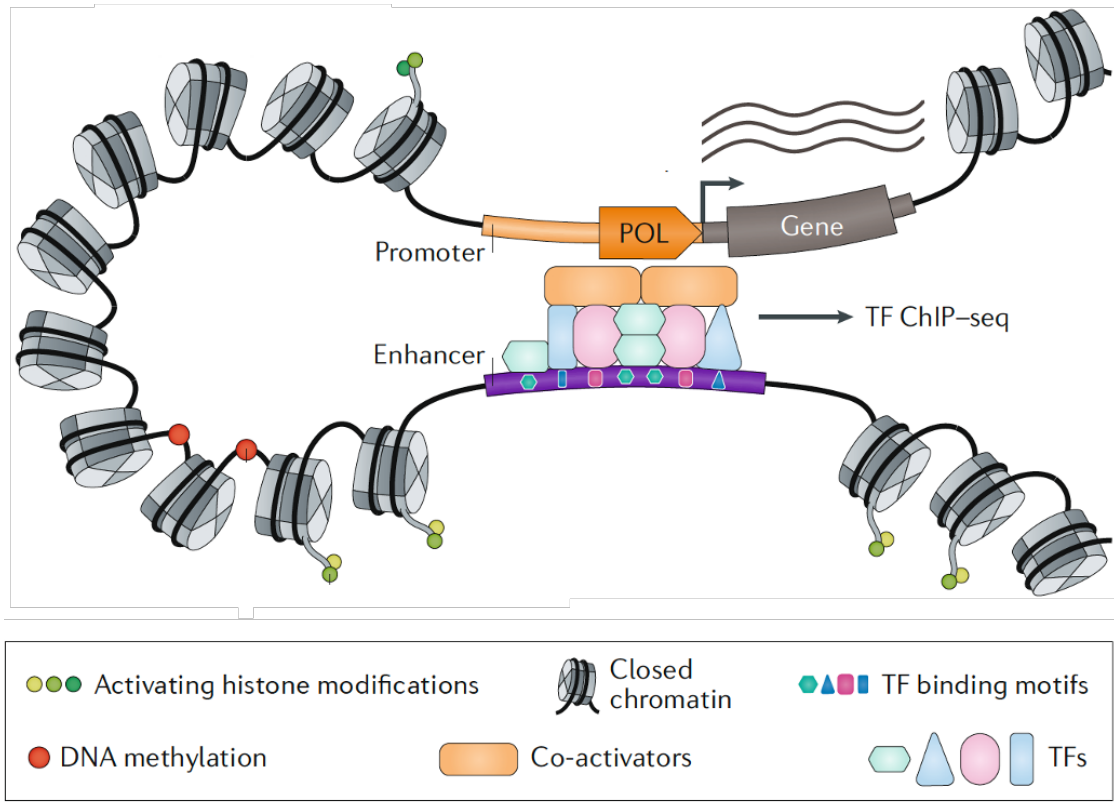

Figure 3 – Enhancer-promoter regulation is mediated by transcription factors. The regulatory potential contained in enhancers is exerted by TFs that bind specific motifs within the enhancer element and interact with its target gene promoter to control transcription (modified from Gasperini *et al.,* 2020).

2. Functional survey of the genome.

Enhancers, silencers and insulators have been identified and characterized mainly through reporter assays and the use of transgenesis. In fact, the first definitions of these regulatory elements are based on their ability to increase, decrease or protect the expression of a transgene (Banerji, Rusconi and Schaffner, 1981; Brand *et al.*, 1985; Kellum and Schedl, 1991).

Enhancer-reporter assays (ERAs) consists generally of a vector where the expression of a reporter gene such as LacZ, Luciferase or GFP, is controlled by a minimal promoter with low levels of basal expression (Manzanares *et al.*, 2000). In order to interrogate the genome for CREs, the candidate region is cloned either upstream or downstream, and reporter expression is assessed (**Figure 4**). Therefore, if the candidate region is a functional enhancer, there will be a boost in transcription. ERAs can be performed *in vitro* (e.g., plasmid transfection to tissue culture cells) or *in vivo* (e.g., pronuclear microinjection of mouse zygotes). While *in vitro* ERAs allow for the rapid assessment of candidates in a particular cell line, they only capture one cellular context. For instance, enhancers can be specifically active not only in a particular tissue but also at a precise developmental stage. Hence, using *in vitro* ERAs we are potentially missing true enhancers if the required conditions are not present (Kvon, 2015). Instead, the generation of transgenic animals carrying the ERA construct provides a powerful tool to identify enhancers in all tissues and through different developmental stages (Manzanares *et al.*, 2000). The classical way of assessing enhancer activity in mouse embryos has been zygotic microinjection of linear ERA constructs (Banerji, Olson and Schaffner, 1983; Gillies *et al.*, 1983; Mercola *et al.*, 1983; Pennacchio *et al.*, 2006; Visel *et al.*, 2007). This strategy offers an immediate answer when is used in transient and embryos are dissected prior to birth. On the other hand, establishing a mouse line enables the thorough characterization of enhancer elements in multiple tissues and individuals. Either way, this procedure is very expensive and time-consuming, thus limiting the throughput of *in vivo* enhancer characterization. Indeed, whereas hundreds of thousands of enhancers are predicted to exist in the human genome, only a small fraction of such regulatory players has been validated (Visel *et al.*, 2007; Encode Project Consortium, 2012; Gasperini, Tome and Shendure, 2020).

Despite ERAs being mainly used to characterize enhancers, identification of other types of CREs can be achieved by modifying certain aspects from the construct. For instance, in enhancer-blocking assays (EBAs) aiming to identify insulators, the testing fragment is placed between an active enhancer and the promoter (Chung, Whiteley and Felsenfeld, 1993; Lunyak *et al.*, 2007). The assay relies on the ability of insulators to block genetic communication, leading to reduced reporter expression. A similar approach that also depended on reducing existing transcription was used to describe

the first silencer. In this case, the authors tested a genomic region that drove gene expression, while a larger version of the same region actively reduced transcription (Brand *et al.*, 1985). However, both strategies largely depend on the presence of additional elements in the system, such as enhancers, making it more complex and constraining the capacity of detection.

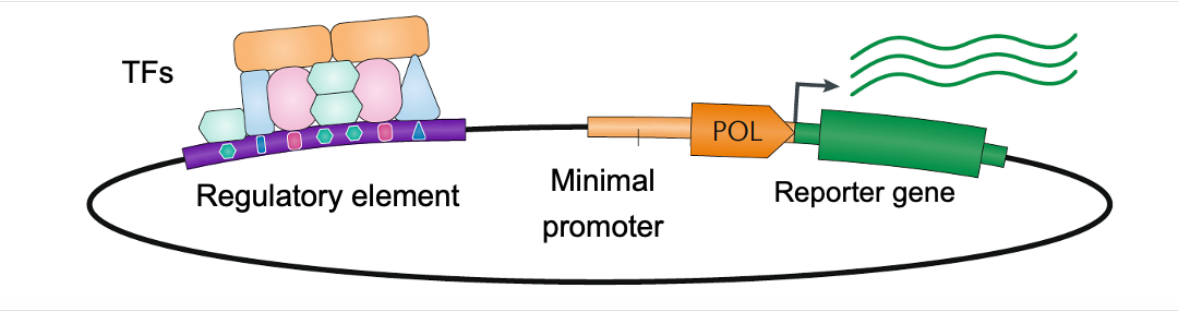

Figure 4 – Enhancer-Reporter Assays (ERAs). Schematic representation of an ERA vector containing a reporter gene, a minimal promoter and a candidate regulatory element. When a candidate genomic region contains enhancer activity in the cell type or tissue in which it is being tested, there is a boost in transcription mediated by TFs (modified from Gasperini *et al.,* 2020).

2.1. Annotated versus validated CREs.

The study of the functional genome contributes to a better understanding of physiology and disease. Since non-coding mutations affecting CREs can cause pathologies, a comprehensive catalog of all human CREs would be a major breakthrough towards precision medicine. However, a simple glance at the literature is enough to see that different reports call enhancers to different things. This is due to the existence of multiple methods that identify features of enhancers. Whereas ERAs identify genomic regions capable of induce transcription, more recent characterization of the genome has identified other features of regulatory elements. For instance, specific histone modifications mark active enhancers like monomethylation of histone H3 lysine 4 (H3K4me1) and acetylation of histone H3 lysine 27 (H3K27ac)(Rada-Iglesias *et al.*, 2011). The acetyltransferase and transcriptional coactivator p300 also associates with enhancer activity (Visel *et al.*, 2009). Additionally, enhancers are bound by tissue-specific TFs, repositioning nucleosomes and leaving DNA more accessible. Therefore, chromatin immunoprecipitation followed by sequencing (ChIP-seq) of histone marks and TFs, as well as, assessment of chromatin accessibility (e.g. ATAC-seq) provide with genome-wide maps of biochemically annotated enhancers (Encode Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). Transcription also

seems a defining feature since enhancers can initiate transcription and the product, enhancer RNA (eRNA), quantitatively correlates with enhancer activity (Kim *et al.*, 2010; Mikhaylichenko *et al.*, 2018).

Similar to enhancers, silencers also show accessible chromatin and are bound by repressor TFs, such as Polycomb repressive complex 2 (PRC2), that are different from those regulating global heterochromatin-mediated repression. In this case, the histone modifications associated with facultative silencers are more controverted but trimethylation of histone H3 lysine 27 (H3K27me3), trimethylation of histone H3 lysine 9 (H3K9me3) and monomethylation of histone H4 lysine 20 (H4K20me1) seem to associate with this feature (Ngan *et al.*, 2020; Pang and Snyder, 2020). Insulators and boundaries of locus control regions can also be annotated genome wide. CTCF binds DNA multiple times at insulators where it prevents interaction between neighbor domains (Spielmann, Lupiáñez and Mundlos, 2018).

Nevertheless, annotated enhancers, silencers and insulators do not always correlate with functional activity. For instance, despite being a first layer of evidence, only a fraction (~26%) of genomic regions annotated as enhancers turned out to have regulatory activity (Kwasnieski *et al.*, 2014). Even in ERAs, where enhancer activity is functionally assessed, there are important limitations. On the one hand, the genomic region is tested out of context and the transgene is randomly inserted into a new genomic region where it might interact with other potential CREs, thus compromising the final outcome. On the other hand, by placing the candidate enhancer nearby a promoter, we might detect regulatory activity that would not have an impact *in vivo* due to enhancer-promoter inaccessibility.

2.2. Finding the target gene.

In a huge effort to understand the regulatory genome, big international consortia like the ENCODE project or the Roadmap Epigenomics project have systematically annotated the chromatin in terms of accessibility, histone modifications and TF binding in many cell types and tissues (Encode Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). This resource has assigned biochemical signatures to ~80% of the genome and is extremely useful to predict regulatory elements. Furthermore, massively-parallel reporter assays (MPRAs) have

interrogated thousands of biochemically annotated candidate genomic regions for enhancer activity in cultivated cells (Inoue and Ahituv, 2015; Gordon *et al.*, 2020). However, either predicted or validated enhancers do not reveal information about the functional role of CREs or its potential target gene(s).

In many instances, enhancer function and target genes are assigned on a proximity basis. However, although most regulatory elements lie within and act upon nearby genes, there are examples of long-range cis-acting regulatory elements that control gene expression over hundreds of kilobases (kb) and even at the Megabase (Mb) scale. A prime example of that is the ZRS enhancer that is located ~1 Megabase (Mb) from its target but not the closest gene, *SHH*. Sequence variations in this highly conserved limb enhancer can cause polydactyly in cats, mice and humans (Lettice *et al.*, 2003; Furniss *et al.*, 2008; Kvon *et al.*, 2020). Therefore, enhancer identification through reporter assays alone does not provide sufficient information about the gene(s) they regulate. In order to overcome this limitation, three-dimensional (3D) interactions between distal enhancers and target promoters can be detected by chromosome conformation capture (3C), which is based on ligation of spatially proximal cross-linked genomic regions (Dekker *et al.*, 2002). 3C-derived techniques have rapidly evolved to 4C and Hi-C, among others, with increased throughput that allows for the interrogation of chromatin contacts in a genome-wide manner. These techniques have delimited the territory within highly interacting topologically associated domains (TADs) in which functional cis-interactions between enhancers and promoters can occur (**Figure 5**)(Dixon *et al.*, 2012; Rao *et al.*, 2014). In a recent work, Montefiori and colleagues mapped all interactions in the genome involving at least one promoter in human induced pluripotent stem cells (hiPSC)-derived cardiomyocytes (CM), generating a very useful resource for cardiovascular genetics (Montefiori *et al.*, 2018).

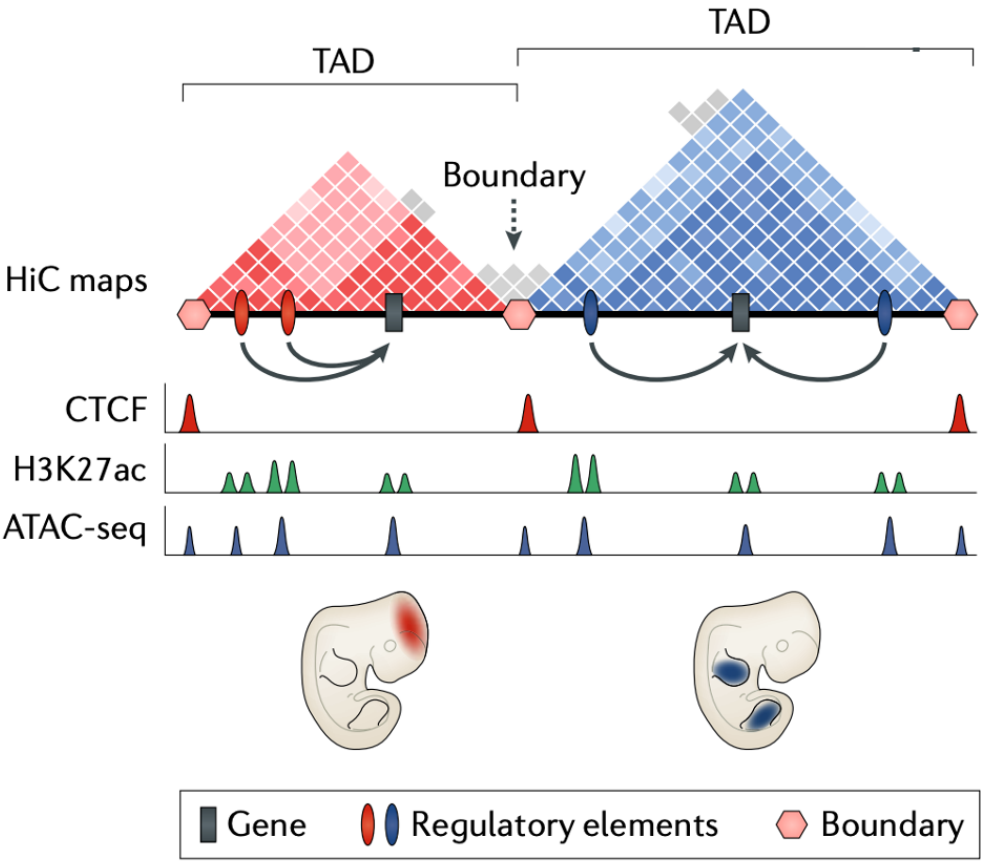

Figure 5 – Topologically associated domains (TADs) delimit enhancer-promoter communication. The chromatin is three-dimensionally organized in TADs, highly-interacting functional domains within which gene expression is tightly regulated. TADs are separated by boundary regions of low interaction which are usually enriched for CTCF binding sites and avoid crosstalk between regulatory elements and genes from neighboring domains (from Spielmann *et al.*, 2018).

Gene regulation by distal enhancers is mediated by architectural proteins such as CTCF and cohesin, that facilitate long-range physical chromatin interactions between enhancer and promoters (Spielmann, Lupiáñez and Mundlos, 2018). Although enhancer-promoter interactions, which usually occur within TADs, can be detected by Hi-C, it seems that there is not a causal relationship between enhancer-promoter interactions and direct gene regulation. Instead, enhancer-promoter interactions have been detected in tissues not showing transcription (Williamson *et al.*, 2019). More surprisingly is the fact that transcription can sometimes be regulated without detecting an enhancer-promoter interaction (Alexander *et al.*, 2019). Altogether, chromatin analysis shows an additional layer of genetic information valuable for the identification of target genes which needs to be further supported by functional evidence.

Another powerful tool to infer target genes regulated by enhancers is to study the relationship between allele genotype and gene expression. The so-called expression quantitative trait loci (eQTLs) represent SNPs associated to a change in transcription between the reference and the alternative allele. In this regard, the GTEx project has provided a substantial resource for gene expression data across multiple human tissues, including right atria and left ventricle, identifying >4 million significant eQTLs (Lonsdale *et al.*, 2013; GTEx Consortium, 2020). Despite the great utility of eQTLs to prioritize tissue-specific regulatory elements, a limitation of the data provided by GTEx resides in underrepresentation of certain cell types within complex tissues, e.g., endothelial cells in the predominantly myocyte-populated heart. Although, eQTLs are very limited to demonstrate direct regulation, when used in combination with functional assays they can be more insightful.

Current genome-editing technology allows us to perturb CREs in order to evaluate changes in gene expression and identify target genes (Gomez-Velazquez *et al.*, 2017; Sainz de Aja *et al.*, 2019). The CRISPR/Cas9 system has made possible to target virtually any loci (Jinek *et al.*, 2012, 2013; Ran *et al.*, 2013) by directing the Cas9 endonuclease to the locus of interest with a guide RNA. Differential gene expression after deletion of the candidate CRE provides experimental evidence for direct regulation (**Figure 6**). While the biochemical and three-dimensional annotation of the genome prioritizes candidate CREs and predicts target genes, enhancer perturbation functionally validates the former. However, large deletions mediated by CRISPR can alter the chromatin landscape. In other instances, CREs can be located near gene bodies like introns or even on top of coding exons (Ahituv, 2016). Deletion of a candidate CRE involving exons or TSS interferes with gene expression but it is not suitable to demonstrate cis-regulation. More recently, a catalytically inactive or dead Cas9 (dCas9) fused to either transcriptional repressors or activators showed epigenetic editing of enhancers and promoters (Gilbert *et al.*, 2013; Perez-Pinera *et al.*, 2013; Hilton *et al.*, 2015). Transcriptional interference or activation mediated by CRISPR can overcome the limitations of gene editing at gene bodies and is suitable for epigenetic therapy (Matharu *et al.*, 2019; Matharu and Ahituv, 2020).

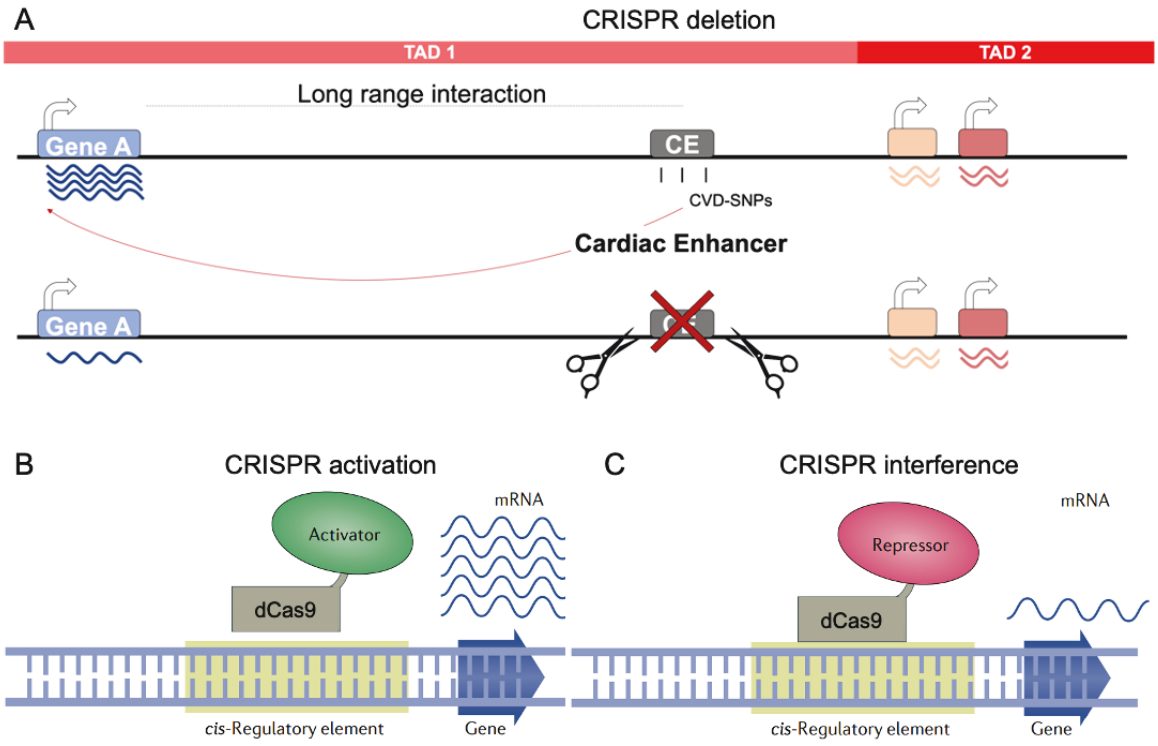

Figure 6 – Identification of target genes regulated by enhancers using (epi)genetic editing. Enhancer deletion using regular CRISPR/Cas9 technology (**A**), activation using CRISPRa (**B**), or repression using CRISPRi (**C**) are the ultimate validation of gene regulation mediated by the candidate enhancer. As depicted in (**A**), distal enhancers can interact over the long range with their target genes (Gene A) while not regulating the closest genes. Enhancer deletion would confirm specific enhancer-gene regulation. Epigenetic modulation (**B** and **C**; from Matharu and Ahituv 2020) is a powerful tool to test candidates that are located in more compromised genomic locations such as coding or promoter regions where deletions might lead to disturb the target gene without implying cis-regulation.

3. GWAS: identifying the genetic contribution to common diseases.

Individuals in a population do not have identical DNA sequences. Instead, there is a great number of places in the genome where there can be differences between individuals, generating different alleles (Craig Venter *et al.*, 2001; Lander *et al.*, 2001). Sometimes, different alleles result in phenotypic differences and cause genetic diseases. This would be the case of sequence alterations not only in gene bodies but also in their regulatory elements. Unravelling how these variants affect cellular or organismal phenotypes is a major goal to understand the genetic contribution to human diseases.

Aggressive mutations can cause heritable forms of common diseases that run in the family and are highly penetrant (Abifadel *et al.*, 2003; Chen *et al.*, 2003; Yang *et al.*, 2004). However, such mendelian versions are very rare and only account for a small

proportion of all patients with the general condition (Bapat *et al.*, 2018). On the other hand, polymorphisms are common variants present in at least 1% of the population with no apparent or mild phenotype. These common variants are nonetheless thought to be involved in the susceptibility to polygenic diseases.

Genome-wide association studies (GWAS) have allowed the exploration of the genotype-to-phenotype impact of human genome variation. GWAS leverage single-nucleotide polymorphisms (SNPs) in the genome and test whether they are linked to traits or diseases. To date, thousands of SNPs mostly in the non-coding genome have been linked to common diseases (Manolio *et al.*, 2009; Manolio, 2010; Rickels and Shilatifard, 2018). These variants might be located within CREs and affect their regulatory potential, thus affecting disease risk. Despite the large number of associations, the mechanism behind the vast majority of GWAS-SNPs remains unknown except in the case of a handful of loci (Smemo *et al.*, 2014; Gupta *et al.*, 2017). Therefore, the overall genetic contribution to common diseases are poorly understood.

3.1. What can we learn from GWAS? Regulatory variants as a basis for common diseases.

GWAS are called to identify part of the so-called ‘missing heritability’ of common diseases such as cardiovascular or neurological diseases (Manolio *et al.*, 2009). GWAS-SNPs have been pervasively found in non-coding regions and risk loci are thought to harbor regulatory potential. Variants at core of CREs can modulate their activity and affect disease-associated gene expression. However, a limitation of GWAS is that they do not have enough resolution to identify causal SNPs (Tam *et al.*, 2019). SNPs contained in a GWAS array are called tag SNPs and are representative of a group of SNPs called a haplotype. These SNPs are within the same linkage disequilibrium (LD) block which is normally inherited together. Therefore, the lead SNPs associated to a trait or disease are not necessarily the causative SNPs and rather represent the risk loci containing all SNPs in high LD. Since LD blocks can span up to a few hundred kilobases (kb), identifying the functional mechanism behind these associations is not that simple (Wall and Pritchard, 2003).

Epigenomic and chromatin interaction data generated over the last years are valuable to prioritize risk loci and candidate CREs. Additionally, the reprogramming of adult cells into human induced pluripotent stem cells (hiPSC) enables the functional assessment of disease phenotypes in cells derived directly from patients (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007). Finally, the genetic and epigenetic edition of candidate loci in disease-relevant tissues or differentiated hiPSC allows disrupting the activity of CREs and identifying novel genes associated to the disease (**Figure 7**).

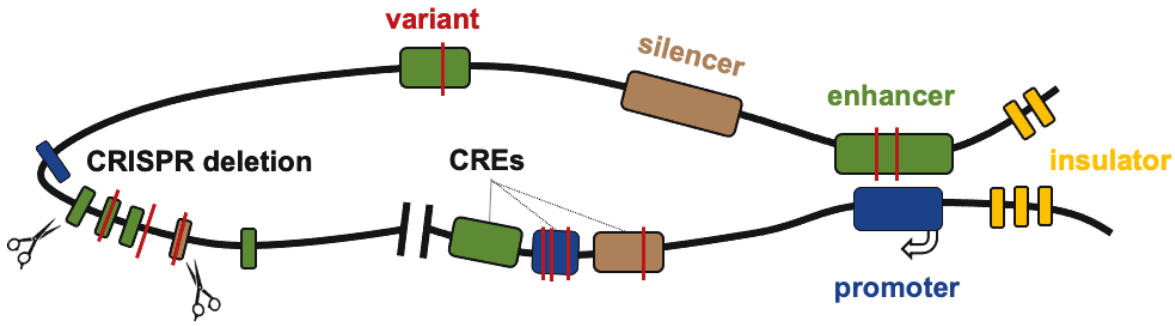

Figure 7 – Regulatory variants and disease susceptibility. GWAS-variants in the non-coding genome might disrupt cis-regulatory elements and affect the expression of disease-relevant genes.

4. (Epi)Genetics of Cardiovascular diseases.

Cardiovascular diseases (CVD) are the first cause of death in the world, a significant health burden that continues increasing and represents 31% of all deaths (Roth *et al.*, 2017; Wilkins *et al.*, 2017). GWAS have associated a number of SNPs with an increased risk of developing major cardiovascular diseases. Therefore, understanding the molecular mechanisms behind these associations opens the door for early genetic diagnosis and new therapeutic targets (Tam *et al.*, 2019). However, little is known about the regulatory and physiological mechanisms behind most of these SNPs that usually lie in uncharacterized genomic regions and far away from relevant genes. Despite the large number of associations identified, the number of SNPs that reach the widely used $5x10^{-8}$ statistical threshold keeps increasing as new studies include higher numbers of individuals.

Evidence shows that sequence changes in CREs can reduce or disrupt TF binding. Whether this is the only mechanism through which GWAS SNPs affect disease risk is

yet poorly understood. An outstanding work from the Kathiresan lab unveiled the mechanism behind a common variant associated to five vascular disease, including coronary artery disease (CAD). The SNP located in the third intron of the *PHACTR1* gene uncovered a regulatory element controlling *EDN1* gene expression. *EDN1* is located 600 kb distal from the causal SNP and encodes the vasoconstrictor protein ET-1 which is involved in atherosclerotic plaque development and promotes CAD. The risk allele showed disruption of enhancer activity and the genetic conversion of the reference allele into the risk one demonstrated that a single nucleotide change is enough to confer the vascular phenotype (Gupta *et al.*, 2017). This example illustrates the power of GWAS and how it can contribute to better understand cardiovascular diseases and develop therapies. With this in mind, my PhD pursues a clear objective: to dissect key risk loci associated to atrial fibrillation (AF) and atherosclerosis, two major cardiovascular diseases.

4.1. Atrial fibrillation

AF is the most common cardiac arrhythmia in humans, causing considerable morbidity, and contributing to overall mortality (Staerk *et al.*, 2017). AF affects over 30 million people worldwide (Chugh *et al.*, 2014) and it is estimated that the number of patients with AF will double by 2050 (Krijthe *et al.*, 2013). The prevalence of AF is about 3% of the population, being a major cause of sudden death, heart failure, cardiovascular morbidity and stroke (Kirchhof *et al.*, 2016). AF starts with abnormal electrical activity in the atria, not governed by the sinoatrial (SA) node. Uncoordinated impulses trigger re-entrant waves in a refractory region of the atria, known as the substrate. Then, atrial contraction becomes chaotic and causes fibrillation in the atria, which prevents the correct excitation of the ventricles and impairs the normal function of the heart (**Figure 8A and B**). The two events described, initial ectopic firing and re-entry, are the main mechanisms maintaining AF (Ellinor *et al.*, 2005; Arnar *et al.*, 2006; Lip *et al.*, 2016). The substrate experiments structural remodelling with time, which stabilizes the re-entry currents and diminishes the importance of ectopic firing (Christophersen *et al.*, 2009). Therefore, continuous and irreversible atrial remodelling leads to long-term perpetuation of AF.

AF is now considered a polygenic condition (Lubitz *et al.*, 2017; Bapat *et al.*, 2018) and GWAS performed in over two million people in the last fifteen years have identified

130 risk loci (Gudbjartsson *et al.*, 2007; Benjamin *et al.*, 2009; Ellinor *et al.*, 2010, 2012; Christophersen, Rienstra, *et al.*, 2017; Low *et al.*, 2017; Nielsen *et al.*, 2018; Roselli *et al.*, 2018). Small insertions and deletions (indels) and copy-number variation (CNV), less studied forms of genetic variation, have also identified genomic regions associated to AF (Gudbjartsson *et al.*, 2015; Tsai *et al.*, 2016). The first three loci to be associated to AF through GWAS were the 4q25 (Gudbjartsson *et al.*, 2007), 16q22 (Benjamin *et al.*, 2009) and 1q21 (Ellinor *et al.*, 2010) loci. Despite subsequent analysis with increased number of patients identified dozens of new loci, they remain as the most significant associations (Nielsen *et al.*, 2018; Roselli *et al.*, 2018). Nevertheless, the mechanism underlying such increased risk of AF remains elusive even for these first loci identified more than a decade ago (**Figure 8C**).

Independent signals at the 4q25 locus have been associated to AF, spanning up to 170 kb from its closest gene, *PITX2* (Gudbjartsson *et al.*, 2007, 2009; Benjamin *et al.*, 2009; Ellinor *et al.*, 2010, 2012; Lubitz *et al.*, 2014). Previous work from the Manzanares lab and others have dissected this first and most significant AF risk locus, identifying various different regulatory elements (Aguirre *et al.*, 2015; Ye *et al.*, 2016; Zhang *et al.*, 2019). Using reporter assays, a distal potentiator element containing the AF variant rs2200733 was identified (Aguirre *et al.*, 2015). This element harbours regulatory potential and assists other cardiac enhancers in the control of *Pitx2c* gene expression, the cardiac-specific isoform of a transcription factor essential during heart development (Ocaña *et al.*, 2017). 3D chromatin analysis not only showed interaction between this potentiator and the cardiac-specific *Pitx2c*, but also regulated the neighbouring *Enpep* gene, a member of the renin-angiotensin system that is involved in hypertension (Mizutani *et al.*, 2008). Interestingly, intergenic deletions regulating the expression of the *PITX2* gene resulted in animals susceptible to develop arrhythmia (Zhang *et al.*, 2019). In a work from Ye and colleagues (Ye *et al.*, 2016), the authors also implicated an intronic variant of *PITX2* in arrhythmia development. The authors introduced a point mutation in human embryonic stem cells (hESC) that changed the reference rs2595104-G allele for the risk rs2595104-T. In differentiated cardiomyocytes, the risk allele diminished binding of TFAP2-alpha and reduced the expression of the cardiac *PITX2C* isoform. These studies show how functional studies are important in order to identify new genes and mechanisms involved in the pathophysiology of the disease.

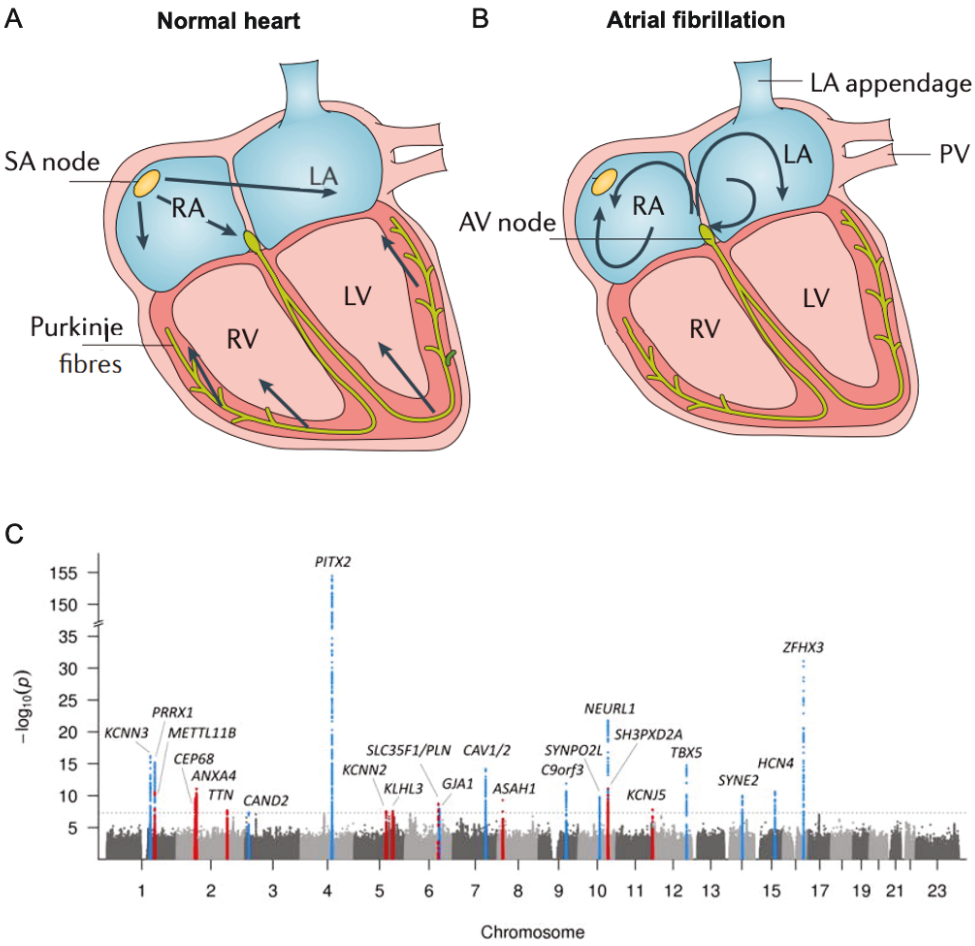

Figure 8 – Arrhythmia development and susceptibility. A) Normal heartbeat (sinus rhythm; SR) starts with the trigger from the sinoatrial (SA) node which then propagates throughout the atria and generates atria contraction. When the cardiac electrical impulse reaches the atrioventricular (AV) node, it expands to the ventricles and the Purkinje fibers generate a coordinated contraction of the ventricles which is slightly delayed from atria contraction. **B)** In the fibrillating atria, ectopic firing and reentrant waves produce inefficient atria contruction and uncoordinated electrical conduction towards the ventricle which impairs heart function (modified from Lip *et al.,* 2016). **C)** Manhantan plot showing the 30 loci that were associated to AF by 2017 that include risk loci studied in this thesis, namely *KCNN3*, *PRRX1*, *KCNIP1* (CNV), *WNT8A* (shown as *KLHL3*), *GJA1*, *CAV1*, *TBX5*, *C9orf3*, *SYNPO2L*, *SYNE2*, *HCN4* and *ZFHX3* (from Christophersen *et al.,* 2017).

The 4q25 locus has been the most extensively studied in relation to AF. However, the rest of the AF risk loci have hardly been explored and the molecular mechanisms behind associations remain unknown. The study of gene expression from human samples has detected eQTLs with mild effects in several AF risk loci including *CAV1*, *MYOZ1* and *PRRX1* (Martin *et al.*, 2015; Hsu *et al.*, 2018). These effects were based on correlations between variant genotypes and transcription levels and are very limited to find causality. Tucker and colleagues combined reporter assays, 3D chromatin

analysis and eQTLs to identify a regulatory element in the 1q24 AF locus controlling *PRRX1* gene expression (Tucker *et al.*, 2017). In an effort to understand GWAS associations, large genomic fragments associated to AF have been deleted from the mouse genome. Although some of these animal models presented altered gene expression of candidate target genes, they were viable and healthy (van Ouwerkerk *et al.*, 2019), with the exception of the deletion of a large genomic region in the *HCN4* locus which predisposed to arrhythmia (van Ouwerkerk *et al.*, 2020). More recently, in a large animal model of AF, our lab has characterized the transcriptomic and proteomic signatures of AF as the arrhythmia progresses in the sheep atria (Alvarez-Franco *et al.*, 2020). Understanding the molecular mechanisms underlying genetic predisposition to arrhythmia development and its confluence with markers of disease progression is therefore a major goal towards precision medicine.

4.2. Atherosclerosis

Atherosclerosis is the main cause of death worldwide. It is a progressive inflammatory disease of the large arteries that causes atheroma plaques, generally by accumulation of lipids and other cell types. After the ingestion of high levels of low-density lipoprotein (LDL)-cholesterol by macrophages, they become 'foam cells' and accumulate in the subendothelial layer of the arteries. As the lesion advances, the plaque can grow sufficiently to block the blood flow. However, the main complications occur after rupture of a plaque which creates a thrombus, i.e. a blood clot, that results in myocardial infarction or stroke (Lusis, 2000; Hansson and Hermansson, 2011).

Atherosclerosis is a multi-component disease affected by different cell types that act as disease players (**Figure 9 left panel**)(Glass and Witztum, 2001; Falk, 2006; Kojima, Weissman and Leeper, 2017). Apart from endothelial and vascular smooth muscle cells (VSMCs) in the arteries, macrophages from the immune system are key players. Since LDL-cholesterol is, by far, the most important and extensively studied risk factor for atherosclerosis (Ference *et al.*, 2017), diet and lifestyle are also important factors. Likewise, hepatocytes too are involved as the liver is a central organ in cholesterol metabolism (Bechmann *et al.*, 2012).

In atherosclerosis, GWAS have identified about 200 risk loci for coronary artery disease (CAD)(Ozaki *et al.*, 2002; Samani *et al.*, 2007; Willer *et al.*, 2008; Erdmann *et*

al., 2009; Schunkert *et al.*, 2011; Nikpay *et al.*, 2015; Nelson *et al.*, 2017; Van Der Harst and Verweij, 2018; Koyama *et al.*, 2020). Similar to AF and most traits and diseases, the majority of GWAS-SNPs for atherosclerosis-related phenotypes fall within non-coding regions. Risk loci might contain CREs regulating disease-relevant genes in any of the above-mentioned tissues and cell types (i.e. endothelial cells, VSMCs, macrophages and hepatocytes). Interestingly, *PCSK9*, a gene involved in familial hypercholesterolemia is found in CAD GWAS (**Figure 9 right panel**). PCSK9 is produced in the liver and secreted to the bloodstream, where it controls the metabolism of LDL-cholesterol through the turnover of LDL receptor (LDLR)(Seidah and Prat, 2007). High levels of circulating PCSK9 lead to increased LDL-cholesterol and atherosclerotic lesions with the subsequent effect in infarct and stroke risk. Conversely, low levels of PCSK9 reduce LDL-cholesterol and atherosclerosis (Cohen *et al.*, 2006). Indeed, gain-of-function mutations are found in families with hypercholesterolemia (Abifadel *et al.*, 2003), while loss-of-function mutations are protective for atherosclerosis (Rashid *et al.*, 2005; Cohen *et al.*, 2006). Due to its direct effect on cholesterol levels and artery burden, in less than fifteen years PCSK9 has come all the way from being hardly described to clinical trials, where scientists and clinicians study the way of diminishing its pro-atherosclerotic function targeting PCSK9 at the protein and mRNA levels (Shapiro, Tavori and Fazio, 2018).

The molecular mechanisms underlying atherosclerosis-SNPs in non-coding regions of the *PCSK9* locus are yet to be determined. In this regard, dissecting the *PCSK9* locus with functional genomic approaches and identifying causative SNPs will classify tissues by the risk of *PCSK9* gene expression deregulation and determine their contribution to the disease. Understanding the expression profile of *PCSK9* as well as the cell type-specific regulatory elements accounting for that will provide with powerful information to dissect the time and tissues in which the expression of this LDLR turnover regulator is key in atherosclerosis.

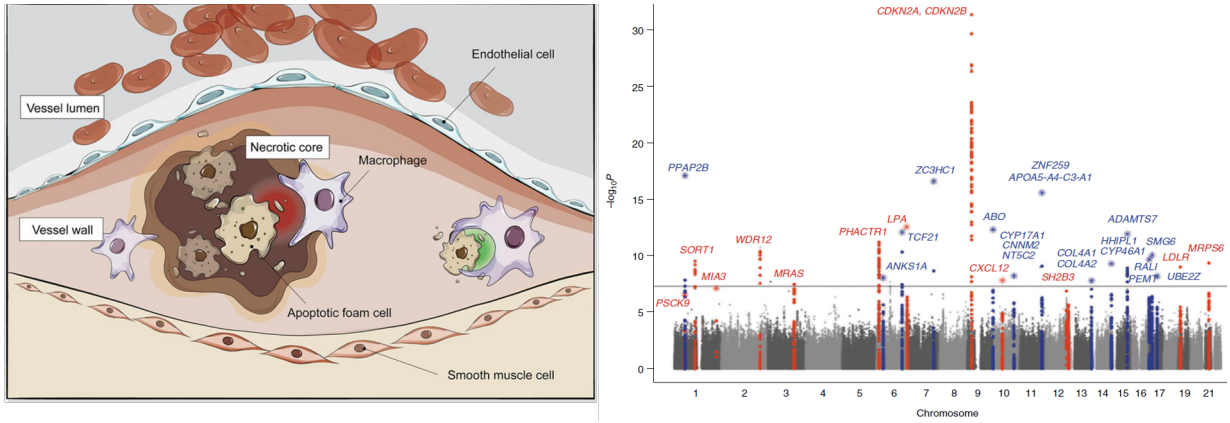

Figure 9 - Development and susceptibility of atherosclerosis. Atherosclerosis generates damage in the wall of the arteries hampering blood circulation and increasing the risk of thrombus and subsequent fatal events such as infarct or stroke (left panel; from Kojima *et al.,* 2017). Besides metabolism of LDL cholesterol in the liver, other cell types such endothelial cells, vascular smooth muscle cells and macrophages play a role locally in the atherosclerotic lesion. GWAS have associated hundreds of loci, including *PCSK9* to an increased risk of developing coronary artery disease and myocardial infarct (right panel; from Schunkert *et al.,* 2011). Molecular mechanisms behind these associations might be involved in gene regulation in the liver as well as the other local atherosclerotic cell types.

Objectives

The functional dissection of the regulatory genome associated to common diseases can shed light on the susceptibility conferred by human variation. The aim of this doctoral thesis is to explore the genetic contribution of cardiovascular diseases in search for risk-associated regulatory elements. To that end, we defined the following objectives:

- To improve current methodology for the assessment of regulatory activity in order to scale up *in vivo* interrogation of risk loci.
- To dissect the most significant AF-risk loci systematically, and functionally characterize new CREs involved in the pathophysiology of the disease.
- To explore the convergence between genetic predisposition and arrhythmia perpetuation.
- To decode the regulatory networks controlling *PCSK9* gene expression and atherosclerosis risk.

Material and methods

To love the journey is to accept no such end.
I have found, through painful experience,
that the most important step a person can take is always the next one.

Brandon Sanderson, *Oathbringer*.

1. Cloning.

The pPB- lacZ vector was obtained by inserting a cassette (3.7 kb) containing a -globin minimal promoter, a *lacZ* reporter gene and a SV40 polyadenylation signal from the p1230 plasmid (Aguirre *et al.*, 2015) into a PB-CAG-DDdCAs9-VP192-T2A-GFP vector (Weltner *et al.*, 2018 ;a gift from Diego Balboa and Timo Otonkoski, University of Helsinki), after removal of the CAG-DDdCAs9-VP192-T2A-GFP cassette (digested with SpeI and BamHI). The *lacZ* cassette was amplified using primers 'PB-Cassette Fw' and 'PB-Cassette Rv' (**Table 1**).

Commercial human DNA (Promega, Cat. No. G1521) was used for PCR amplification of all tested genome fragments from AF associated loci (for primers used see **Table 1**) using Expand High Fidelity PCR system (Roche ref. 11732650001). Primers were designed using NEBuilder assembly tool to have a minimum 20-bp homology (black lower case nucleotides) arm overhangs for Gibson cloning (Gibson *et al.*, 2009)(NEB, Cat. No. E2611) into the pPB- lacZ vector digested with SpeI and SacII for 3' cloning, or HindIII for 5' cloning. All constructs were verified by Sanger sequencing of plasmids using primers flanking the candidates cloned in the vector (see sequencing primers for 'Candidates' at **Table 6**). Large or hard-to-amplify fragments were sub-divided into several PCR fragments with shared homology between them for sequential ligation (colored lower-case nucleotides).

Chimeric constructs ASE-ZFHX3 and ZFHX3-ASE were obtained by cloning the ASE fragment upstream or downstream the ZFHX3 fragment, respectively, in the pPB- lacZ-ZFHX3 vector digested with SpeI (ASE-ZFHX3) or with SacII (ZFHX3-ASE). Specific primers were used in each chimeric design to amplify the ASE fragment with homology to each of the cloning positions.

Primer name	sequence	fragment	amp
PB-Cassette Fw	ataaagtaacaaaacttttaACTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAG	β-lacZ-SV40 cassette	p1230 vector
PB-Cassette Rv	aacatatccagtcactatggCCGCGGTGGCGGCCGCTC	β-lacZ-SV40 cassette	p1230 vector
PB-ASE Fw	cgaggtcgacggtatcgataagcttGAATTCACTAGTGATTCGC	ASE	hs
PB-ASE Rv	caggaattcgatatcaagcttGGGAATTCGATTCCAACAC	ASE	hs

PB-ZRS Fw	ctggatccccgggggatccactagtTCAAATG CTCACTTTACATG	ZRS	hs
PB-ZRS Rv	tatccagtcactatggccgcggGCTGAAGTGA TACTGAAG	ZRS	hs
PB-ZFHX3 Fw	ctggatccccgggggatccactagtATTTCTT GTAGAGACAGGG	*ZFHX3*-AF	hs
PB-ZFHX3 Rv	tatccagtcactatggccgcggTTTAAAAAAT TAAAATCAGGCCTC	*ZFHX3*-AF	hs
PB-KCNN3 Fw1	ctggatccccgggggatccactagtTACCTAC ACCAGAAGGGG	*KCNN3*-AF	hs
PB-KCNN3 Rv1	cccttcggctCGCACATCTCATCCTTAC	*KCNN3*-AF	hs
PB-KCNN3 Fw2	gagatgtgcgAGCCGAAGGGGCTGTGCA	*KCNN3*-AF	hs
PB-KCNN3 Rv2	tatccagtcactatggccgcggTACTCTCCAT TAAAGGTAGCAAAATTG	*KCNN3*-AF	hs
PB-PRRX1 Fw	ctggatccccgggggatccactagtTGTGAAA TCTGACTCCCC	*PRRX1*-AF	hs
PB-PRRX1 Rv	tatccagtcactatggccgcggGCAACTTTGG AACTGGGTAAC	*PRRX1*-AF	hs
PB-WNT8A Fw	ctggatccccgggggatccactagtGGGTCAC AGGGTCTTTCG	*WNT8A*-AF	hs
PB-WNT8A Rv	tatccagtcactatggccgcggCCTCCTTCCT TCATCCAG	*WNT8A*-AF	hs
PB-CAV1_1 Fw	ctggatccccgggggatccactagtGTGCATA ATTACTTGCAAC	*CAV1*-AF1	hs
PB-CAV1_1 Rv	tatccagtcactatggccgcggCCACACCATT CTCTTTAAC	*CAV1*-AF1	hs
PB-CAV1_2 Fw	ctggatccccgggggatccactagtGATTACA ACCTCCCTGAGG	*CAV1*-AF2	hs
PB-CAV1_2 Rv	tatccagtcactatggccgcggGGACTGACTG CACTTGCC	*CAV1*-AF2	hs
PB-C9orf3 Fw	ctggatccccgggggatccactagtGTGAAGG AGCCCTGTCTAC	*C9orf3*-AF	hs
PB-C9orf3 Rv	tatccagtcactatggccgcggTTTGGAATAT GAGACCTAGTTTAGAC	*C9orf3*-AF	hs
PB-SYNPO2L Fw	ctggatccccgggggatccactagtTACAGAA ACCAATAAATTGCAACAC	*SYNPO2L*-AF	hs
PB-SYNPO2L Rv	tatccagtcactatggccgcggTGCTCTACCA AGTCAGCAC	*SYNPO2L*-AF	hs
PB-SYNE2 Fw1	ctggatccccgggggatccactagtAAAATCA CTGATTTGGACTAG	*SYNE2*-AF	hs
PB-SYNE2 Rv1	gtctgcctatCTAGCAGTTTCCAAAAATAAC	*SYNE2*-AF	hs
PB-SYNE2 Fw2	aaactgctagATAGGCAGACTTCTCATG	*SYNE2*-AF	hs
PB-SYNE2 Rv2	tatccagtcactatggccgcggTTACACCACA CCAACATAC	*SYNE2*-AF	hs
PB-HCN4 Fw	ctggatccccgggggatccactagtGATGATGA TGCCCCAGAG	*HCN4*-AF	hs
PB-HCN4 Rv	tatccagtcactatggccgcggAAGCTCCCAAA CTGAGCTC	*HCN4*-AF	hs
PB-KCNIP1 Fw	ctggatccccgggggatccactagtATATGCCA GCGCTCTATC	*KCNIP1*-AF	hs
PB-KCNIP1 Rv	tatccagtcactatggccgcggTAGAATCATAC CCACCTTG	*KCNIP1*-AF	hs
PB-GJA1_1 Fw	ctggatccccgggggatccactagtGATATAC AAAAATGTAAAGCAATG	*GJA1*-H3K27ac	hs

PB-GJA1_1 Rv	tatccagtcactatggccgcggAAAATATTTGAGTGGCAAATATAAG	*GJA1*-H3K27ac	hs
PB-GJA1_2 Fw	ctggatccccgggggatccactagtTTAAAGAAATACTGTCTTTTGTG	*GJA1*-HiC	hs
PB-GJA1_2 Rv	tatccagtcactatggccgcggATTCTTTTTGTATGATTTTAAGATCTTAATTAAAAC	*GJA1*-HiC	hs
PB-GJA1_3 Fw	ctggatccccgggggatccactagtTTACAACATGTTATGAATTTTTAAATG	*GJA1*-SNP	hs
PB-GJA1_3 Rv	tatccagtcactatggccgcggAGAATATTTGTTCAAAGAATAGC	*GJA1*-SNP	hs
PB-Gja1mm Fw1	ctggatccccgggggatccactagtTTCTGGTCTAAATTGTTGTTC	*Gja1*-h3k27ac	mm
PB-Gja1mm Rv1	cattttcatttGAGTAGGGTGAGAGAGATATTTC	*Gja1*-h3k27ac	mm
PB-Gja1mm Fw2	caccctactcAAATGAAAATGTGCTGGC	*Gja1*-h3k27ac	mm
PB-Gja1mm Rv2	atgtgacccTGTGAAAGGGTCATTTTAC	*Gja1*-h3k27ac	mm
PB-Gja1mm Fw3	ccctttcacaGGGGTCACATACTAGATATAC	*Gja1*-h3k27ac	mm
PB-Gja1mm Rv3	TATCCAGTCACTATGGCCGCggAACATCCAGTAGTTGTACAGTCCGTGAC	*Gja1*-h3k27ac	mm
PB-minGJA1 Fw	ctggatccccgggggatccactagtCATTTCTCCCACAGGATTTTTC	min*GJA1*-H3K27ac	hs
PB-minGJA1 Rv	tatccagtcactatggccgcggTTAGGTGTTCATGCTTATCT	min*GJA1*-H3K27ac	hs
PB-CE8 Fw	ctggatccccgggggatccactagtGTGTCCCATCTGCTGCCAAG	PSCK9_CE8	hs
PB-CE8 Rv	tatccagtcactatggccgcggGAGATGTTTCTTGGGCTGGTC	PSCK9_CE8	hs
PB-CE9 Fw	ctggatccccgggggatccactagtCAGTTTGGAGGGCTCAGAAG	PSCK9_CE9	hs
PB-CE9 Rv	tatccagtcactatggccgcggCACCTCAGAAAAACCCCAAA	PSCK9_CE9	hs
PB-CE11 Fw1	ctggatccccgggggatccactagtTTGGCCTGGCTGAGAGTTTC	PSCK9_CE11	hs
PB-CE11 Rv1	caggacatgcTGTACAGAGGCCTTGCTC	PSCK9_CE11	hs
PB-CE11 Fw2	cctctgtacaGCATGTCCTGGGGCTGGC	PSCK9_CE11	hs
PB-CE11 Rv2	tatccagtcactatggccgcggGGGATCCTCACAATAACCTTATTATCCCTTTCC	PSCK9_CE11	hs
PB-CE12 Fw1	ctggatccccgggggatccactagtCACTGGGAGGTGGAGGACC	PSCK9_CE12	hs
PB-CE12 Rv1	caccttgtcaGCGAGACCTCTCCTGACC	PSCK9_CE12	hs
PB-CE12 Fw2	gaggtctcgcTGACAAGGTGGACGAAACAGGC	PSCK9_CE12	hs
PB-CE12 Rv2	ggagcttcctGGCACCTCCACCTGGGGA	PSCK9_CE12	hs
PB-CE12 Fw3	tggaggtgccAGGAAGCTCCCTCCCTCAC	PSCK9_CE12	hs
PB-CE12 Rv3	cctctgagccTGTTGCTGTTCTTTTCTCTGGAG	PSCK9_CE12	hs
PB-CE12 Fw4	aacagcaacaGGCTCAGAGGACCCACAG	PSCK9_CE12	hs
PB-CE12 Rv4	tatccagtcactatggccgcggGGGGCAAATTTTTAATCTTGCAGTAATATTAAAC	PSCK9_CE12	hs
ASE-ZFHX3 Fw	ctggatccccgggggatccaGAATTCACTAGTGATTCGC	ASE-ZFHX3AF*	hs

ASE-ZFHX3 Rv	ccctgtctctacaagaaatactagtGGGAATTCGATTCCAACAC	ASE-ZFHX3AF*	hs
ZFHX3-ASE Fw	gggaggcctgattttaatttttaaaccgcggGAATTCACTAGTGATTCGC	ZFHX3AF-ASE**	hs
ZFHX3-ASE Rv	aacatatccagtcactatggAATTCGATTCCAACACTC	ZFHX3AF-ASE**	hs

***Primers used to amplify the ASE fragment to be cloned upstream *ZFHX3*-AF.**
****Primers used to amplify the ASE fragment to be cloned downstream *ZFHX3*-AF.**

Table 1 – List of primers used to amplify and clone candidate regulatory elements. Lower case indicates homology to fragment (color coded) or vector (black). Upper case indicates sequence annealing candidate enhancer for amplification. Underlined grey nucleotides represent inserted nucleotides to generate restriction sites. Fw, forward primer; Rv, reverse primer; amp, amplifying from; hs, Homo sapiens; mm, Mus musculus.

2. Cell culture transfection and CRISPR experiments.

Mouse HL-1 atrial cardiomyocytes were cultured in Claycomb medium (Sigma) supplemented with 10% (v/v) inactive (56°C, 30 minutes) fetal bovine serum (FBS) (Sigma), 4 mmol/L L-glutamine (Sigma), 100 µmol/L norepinephrine (Sigma) and 100 U/mL penicillin-streptomycin (Sigma). All seeding supports were previously coated for 24 hours with a solution of gelatin (0.02% w/v, Sigma) and fibronectin (25 µg/mL, Sigma). Human HEK293T embryonic kidney cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma) supplemented with 10% FBS, 4 mmol/L L-glutamine and 100 U/mL penicillin-streptomycin. Human K562 erythroleukemia cells were cultured in Roswell Park Memorial Institute (RPMI, Life Technologies) 1640 medium supplemented with 10% FBS, 4 mmol/L L-glutamine and 100 U/mL penicillin-streptomycin.

For PB-ERA experiments, cells were counted one day before transfections and plated at a density of 3 × 10^5 cells per p12 well (HL-1 cells) or 1.5 × 10^5 cells per p24 (HEK293T) with complete growth medium. Cells were co-transfected with 1 µg of pPB-lacZ vector containing the appropriate genomic fragment and 1 µg of the transposase plasmid PBase (a gift from Diego Balboa and Timo Otonkoski, University of Helsinki). pPB-CAG-GFP was transfected in parallel as an internal control of transfection efficiency. Co-transfections were performed with 6 µl of Lipofectamine 2000 (Invitrogen) diluted in Opti-MEM (Sigma) reduced-serum medium, according to the manufacturer's protocol. Cells were transferred to complete medium after five hours. The empty vector pPB- lacZ or the pPB- lacZ-DEOct4 containing the pluripotent-specific distal enhancer of *Oct4* was used as a negative control. Ninety-six hours after

transfection, both DNA and RNA were isolated using AllPrep DNA/RNA kit (Qiagen, Cat. No. 80204) and kept at -80°C for qPCR analysis.

For enhancer deletion in mouse and human cells we transfected the CRISPR/Cas9 gene-editing tool, as described (Ran *et al.*, 2013). Briefly, 3 x 10^6 cells were seeded in 10-cm plates the day before transfection. Cells were transfected for five hours with 60 µl of Lipofectamine 2000 and 10 µg of each of the plasmids pSpCas9(BB)-2A-GFP (PX458, Addgene #48138) and pSpCas9(BB)-2A-Puro (PX459, Addgene #48139). Two guides were designed per enhancer (**Table 2**) using CRISPOR (http://crispor.tefor.net/) (Concordet and Haeussler, 2018) or Benchling (https://www.benchling.com/) and cloned into either the plasmid containing Cas9-GFP or Cas9-Puro. Forty-eight hours after transfection, GFP+ cells were sorted using Aria Cell Sorter (BD Biosciences) and seeded with puromycin for other four days. Isolated RNA was stored at -80°C for qPCR analysis.

Target	guide up	guide down	sp
***Cav1*-af1**	CCAGAATTCCGTTCCCAAGT	CAACTACCGAGGTTCCCGAC	mm
***Cav1*-af2**	CAACTACCGAGGTTCCCGAC	CATTGCAACTATACCTTGGT	mm
min*Cav1*-af1	GAGTAGCCTCAAAACGGCAA	CCGATCACCCTAAGAAAGAG	mm
min*Cav1*-af2	GTTAGCCCTTCAATCAGACT	TCAAGCCCTTCAAGGCATAT	mm
***ZFHX3*-AF**	AGCAACATCACCCCTCTTCGTGG	TGAAGGGTTCACCCTACCAAGGG	hs
***Zfhx3*-af**	GGTATGTACCCCACTCGATT	CTCTCTAGGGAAGAATCGCC	mm
***Tbx5*-AF**	CCTAAGCTATCTGAGCCAAA	TAAAATGGGACTAACTCACT	mm
min*Tbx5*-AF	GATGCAAGATCTCATTCGGT	CCTAGACTAATTCCCCAGAA	mm
Enh-block	GTCCCCAGGAGCTCAAAGGG	TAGGGTCTCATACACCGCCC	mm
***Gja1*-h3k27ac**	GTCCCCAGGAGCTCAAAGGG	AATACTATCTTTGATCACAG	mm
***Gja1*-hic_snp**	AATACTATCTTTGATCACAG	TAGGGTCTCATACACCGCCC	mm
min*Gja1*-h3k27ac	GCATTATGATTACTACTCTG	GTCTGTGTTTGGCTTAATAC	mm
minCE11	AGGTCATTGACCCAGGGTCA	AATAACTAGCAGCTGTAGGC	hs

Table 2 – List of guides used for CRISPR/Cas-mediated deletion of candidate regulatory elements.

3. Quantitative PCR.

Isolated RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). cDNA was used for quantitative PCR (qPCR) with Power SYBR Green (Applied Biosystems) in a 7900HT Fast Real-Time PCR System (Applied Biosystems). Expression of each gene was normalized to the

expression of the housekeeping genes *Actin* (HL-1 cells; mouse), *ACTIN* (HEK293T; human) or *GAPDH* (K562 and HepG2; human). Primers used are listed in **Table 3**.

Relative regulatory enhancer activity after PB-ERA assays in cells was calculated as the ratio of reporter *lacZ* expression (RNA) to transfection efficiency (DNA), expressed as mean ± standard deviation and statistically analyzed by unpaired Student's t-test (Graphpad Prism5). A minimum of three replicates were used to calculate enhancer activity.

The effect of enhancer deletion was calculated by comparing gene expression of experimental GFP+, Puro resistant cells versus wild type cells transfected with no guide RNAs. A minimum of three replicates were used to assess the effect of enhancer deletion.

Gene	fw	rv	sp
lacZ	GGCGACTTCCAGTTCAACAT	CATCGCCATCTGCTGCAC	-
ACTIN	TTTGAATGATGAGCCTTCGTCCCC	GGTCTCAAGTCAGTGTACAGGTAAGC	hs
GAPDH	CGCTCTCTGCTCCTCCTGTT	CCATGGTGTCTGAGCGATGT	hs
***PCSK9*_all**	ATGGTCACCGACTTCGAGAAT	GTGCCATGACTGTCACACTTG	hs
***PCSK9*_most**	CTGGTGAAGATGAGTGGCGA	GGTAATCCGCTCCAGGTTCC	hs
***PCSK9*_cb**	ACCCTAACCTTTGTCCTGCA	TCACACGAGTCACAACCTCA	hs
***ZFHX3* pair 1**	CAAGTTCACGACGGACAACCT	GCTTGCACTGGTATGAGTCCC	hs
***ZFHX3* pair 2**	GGGCAGATCTTCACCATCC	TCCTTAGCAAGCTCCTCTGG	hs
GJA1	TCCCCTCTCGCCTATGTCTC	GTTTTGCTCACTTGCTTGCTTG	hs
Actin	CAGAAGGAGATTACTGCTCTGGCT	TACTCCTGCTTGCTGATCCACATC	mm
Cav1	GCGACCCCAAGCATCTCAA	ATGCCGTCGAAACTGTGTGT	mm
***Cav2*_large**	TTGGCCTTCATTGCGGGTATC	GGCAAGACCATTAGGCAGGT	mm
***Cav2*_all**	CCACAGTGGCGTTGACTAC	AGATGAGAGTTGAGCTGGTGA	mm
Tes	AGCCCCCTGTCTAAAATGCAA	GGGTGGTGTACTTAGTGTCCTC	mm
Met	CCCCAACTTCACGGCAGAAA	GTAGTTTGTGGCTCCGAGATAAA	mm
Capza2	GGAAGCAACTGATCCAAGGC	CCCCATTCGGATAATGCTCTTTT	mm
Zfhx3	CCAATAGCCTGGAGAAGCTG	AGTTGCACAGGACACAGTGG	mm
Gja1	ACAGCGGTTGAGTCAGCTTG	GAGAGATGGGGAAGGACTTGT	mm
Hsf2	TGGACGCTTGTGGAGGAAAC	GCTCATCCAAGACCAGAAAACT	mm
Serinc1	CTTTTCTTGCTCGTCGGAGTAT	CCTTTCTCATTCTCACAGAACC	mm
Tbx5	GGCATGGAAGGAATCAAGGTG	TTTGGGATTAAGGCCAGTCAC	mm
Tbx3	TGAGGTGCTCTGGACTGGAT	ACCATCCACCGAGAGTTGTG	mm

Table 3 – List of qPCR primers. Sp, species; hs, *Homo sapiens*; mm, *Mus musculus*.

4. *In vitro* transcription of the PB transposase.

PB transposase was *in vitro* transcribed from a linear template containing a T7 promoter (T7p) and the cDNA of a hyperactive PB transposase (Yusa *et al.*, 2011).

First, linear template was obtained by PCR amplification from the PBase vector, using the primers 'PB-transcription Fw' and 'PB-transcription Rv' listed below:

Primer name	sequence	fragment	amp
PB-transcription Fw	**ttaatacgactcactatag**ATGGGCAGCA GCCTGGACGA	PB transposase cDNA	PBase
PB-transcription Rv	TCATCAGAAACAGCTCTGGC	PB transposase cDNA	PBase

Table 4 – Primers used for in vitro transcription of the PB transposase. Lower case and bold indicate the T7p added in the forward primer to allow in vitro transcription. Amp, amplifying from.

Product from PCR amplification (V=50 µl; 1 ul of vector [20 ng/ul]; 1.5 ul each primer [10 uM]; program: 94 ºC, 2 min; 10x (94 ºC, 15 sec; 65 ºC, 30 sec; 72 ºC, 2 min); 20x (94 ºC, 15 sec; 65 ºC, 30 sec; 72 ºC, 2 min + 5 sec each cycle); 72 ºC, 7 min; 4 ºC, ∞) was run in a 1% agarose gel and the desired band (1.8 kb) was purified using QIAquick gel extraction kit (Qiagen, Cat. No. 28704) and used as a template for the transcription reaction. For *in vitro* transcription, we used 'mMESSAGE mMACHINE T7 ULTRA Transcription kit' (Invitrogen, Cat. No. AM1345) at 37 ºC for 2 hours, according to manufacturer's instructions and using 500 ng of template DNA. This kit includes a step to cleave template DNA and polyadenylate the resulting mRNA. The final capped and polyadenylated mRNA was purified using RNA cleanup step from the 'RNeasy Mini Kit' (Qiagen Cat. No. 74106) and eluted in 40 µl of nuclease-free water. RNA concentration was measured using Nanodrop (approximate yield: 1.0-1.2 µg/µl) and the product was aliquoted in PCR tubes (2-3 µl each) and stored at -80 ºC. RNA is very sensitive to degradation by temperature and aliquots should not be thawed more than two times.

5. Transient transgenic assay using the PB-ERA system.

For the generation of transient transgenics, F1 (C57Bl/6 x CBA/J) females were superovulated to obtain fertilized oocytes and injected zygotes were transferred to

CD1 foster mothers following standard procedures (Behringer *et al.*, 2014). Episomal non-digested pPB- LacZ-derived constructs were microinjected into fertilized E0.5 oocytes at a concentration of 2 ng/µl. *In vitro* transcribed transposase mRNA was microinjected at a concentration of 75 ng/ul together with each pPB- lacZ construct. To ensure the correct translation of microinjected mRNA, the microinjection needle was aimed at the pronucleus and then removed slowly to ensure that mRNA was partially released in the cytoplasm. A summary of transgenic assays is shown in **Table 5**.

construct	stage	# e	# tg	% tg	# lacZ+	# tissue-specific	tissue
CAG-GFP	E10	15	NA	NA	4 (GFP+)	NA	-
ASE	E9.5	16	NA	50-100	8	8	heart
ZRS	E11.5	11	9	81.82	8	8	limb
***ZFHX3*-AF**	E11.5	26	14	53.85	1	0	-
***KCNN3*-AF**	E11.5	34	19	55.88	15	0	-
***PRRX1*-AF**	E11.5	9	5	55.56	0	0	-
***WNT8A*-AF**	E11.5	19	11	57.89	7	0	-
***CAV1*-AF1**	E11.5	20	15	75.00	11	2	heart
***CAV1*-AF2**	E11.5	20	14	70.00	11	4	heart
***C9orf3*-AF**	E11.5	13	6	46.15	3	1	heart
***SYNPO2L*-AF**	E11.5	21	9	42.86	7	0	-
***SYNE2*-AF**	E11.5	16	7	43.75	6	2 + 2	heart & lung
***SYNE2*-AF**	E14.5	1	1	100.00	1	1	lung
***HCN4*-AF**	E11.5	8	4	50.00	3	0	-
***KCNIP1*-AF**	E11.5	20	13	65.00	7	4	heart
***ZFHX3*AF-ASE**	E9.5	27	13	48.15	4	2	heart
ASE-*ZFHX3*AF	E9.5	21	13	61.90	3	3	heart
***GJA1*-H3K27ac**	E11.5	37	14	37.84	5	4	heart
***GJA1*-H3K27ac**	E14.5	6	3	50.00	3	3	heart
***GJA1*-HiC**	E11.5	17	12	70.59	10	1	heart
***GJA1*-SNP**	E11.5	20	10	50.00	9	1	heart
***Gja1*-h3k27ac** (mm)	E11.5	24	13	54.17	13	12	heart
min*GJA1*-H3K27ac	E11.5	22	13	59.09	2	0	-

Table 5 – Summary of PB-ERA transgenesis for AF candidate enhancers included to calculate the efficiency of the system. #, number; e, embryos; tg, transgenics.

6. *lacZ* staining and genotype.

At the desired stage, pregnant female mice were euthanized and embryos dissected and stained for -galactosidase activity (Behringer *et al.*, 2014). All embryos were genotyped for *lacZ* by PCR, using primers for co-amplification of Myogenin as an internal control of a single-copy gene in genomic DNA (**Table 6**). Transgenic efficiency was calculated as the percentage of embryos expressing *lacZ* of total transgenics (**Table 5**).

Gene/construct	fw	rv	comments
lacZ	GCGACTTCCAGTTCAACATC	GATGAGTTTGGACAAACCAC	
Myogenin	CCAAGTTGGTGTCAAAAGCC	CTCTCTGCTTTAAGGAGTC	
Candidates	CCGCTGTTTGGTCTGCTTTC	AAAAAGCTGAACGAGAAACG	Sanger seq

Table 6 – List of primers for genotyping and Sanger sequencing. Each candidate regulatory element was sequenced using forward and reverse primers, separately.

7. Quantitative genotyping.

In order to quantify the number of transgene insertions, we performed qPCR of genomic DNA from transgenic embryos using qPCR primers for *lacZ* (**Table 3**) and relative to a heterozygous mouse line with only one copy of the transgene.

8. Generation of a *GJA1*-H3K27ac mouse line.

After identification of a cardiac enhancer in the *GJA1*-H3K27ac fragment, a transgenic mouse line was generated for the pPB- lacZ-GJA1-H3K27ac construct using the PB-ERA system. Three males were obtained out of which only one transmitted the transgene to the offspring. All embryos reproduced the cardiac-specific expression pattern at E11.5, predominantly in the left ventricle. Offspring from this male was used to characterize the enhancer activity in other stages. The enhancer was active since E11.5 until adulthood. Genotyping was performed by PCR using LZ3 and ZT4 primers (**Table 6**). Transgenic embryos analyzed from this line and enhancer activity is summarized in **Table 7**.

Stage	# embryos	# tg	# lacZ+	heart enhancer
E9.5	14	9	8	3
E11.5	13	10	10	10
E18.5	9	6	5	5

Table 7 – Summary of *GJA1*-H3K27ac enhancer activity in the F1 of an adult transgenic male carrying the enhancer-reporter construct.

9. Animal handling.

Mice were housed and maintained in the animal facility at the Centro Nacional de Investigaciones Cardiovasculares (Madrid, Spain) in accordance with national and European Legislation. Procedures were approved by the CNIC Animal Welfare Ethics Committee and by the Area of Animal Protection of the Regional Government of Madrid (ref. PROEX 196/14).

10. Statistics.

Statistical analyses were performed with GraphPad Prism 6 or Microsoft Excel. Data are presented as means ± standard deviation (sd) unless stated otherwise. Asterisks indicate p-values < 0.05. Tests used to calculate p-value are detailed in the figure legends. In general, Student's t-test was used to compare two groups.

11. Data analysis.

Prioritization of candidate AF-CREs:

AF-associated genomic regions were classified according to the presence of regulatory features. For eQTLs, we included AF-SNPs within the candidate fragments and used GTEx publicly available data (GTEx Consortium, 2020) to annotate them if the expression of any of the genes within the risk locus associated to the genotype of the variants in heart tissue. Histone marks of active enhancers (H3K27ac and H3K4me1) were explored within human candidates and their orthologs in the mouse genome using available data from ENCODE and Roadmap Epigenomics for human left ventricle, right atria and fetal heart, as well as mouse embryonic (E14.5) and adult (8 weeks) heart tissue (Encode Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). ChIP-seq data for TBX5, GATA4 and NKX2-5 were used in

differentiated cardiomyocytes from both hiPSC (Ang *et al.*, 2016) and mESC (Luna-Zurita *et al.*, 2016) to annotate cardiac TF binding within candidate AF-CREs. Available H3K27me3 ChIP-seq data from human tissues (aorta, left ventricle, fetal heart, fetal kidney, fetal lung, fetal brain, H1 derived neuronal progenitor, H1 ESC, H9 ESC) were used to explore the repressive marks at the *ZFHX3* promoter and AF-associated genomic region. (Encode Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015).

Orthologous regions in the mouse genome:

Mouse orthologs of human AF-candidate regions assayed by PB-ERA transfection and transgenesis were obtained using UCSC liftOver tool (Kuhn, Haussler and James Kent, 2013).

Genomic interaction data:

Available promoter-capture Hi-C data from hiPSCs and differentiated CM were used to explore putative target promoters interacting with candidate regulatory regions and the specificity of their interaction (Montefiori *et al.*, 2018). Tracks were loaded to WashU epigenome browser to represent the data as arcs. For detailed assessment of the overlap between interactions and AF variants, data was represented as the mapping reads of the crosslinked interaction.

Spatial transcriptomic data (3D-Cardiomics) from the mouse heart:

Available RNA-seq data from mouse heart tissue was used to observe overall cardiac gene expression as well as atrial levels (Mohenska *et al.*, 2019).

GWAS prioritized genes:

The genomic coordinates for one GWAS-SNP at each of the 130 risk loci for AF were collected. All genes within a window of 200 kb from the variants were selected in order to look for putative target genes. In cases where the risk SNPs within a gene desert (11 loci) did not include any gene that matched the former criteria, the nearest gene at each side was included in the list of candidate genes if it was protein-coding (for this analysis, pseudogenes and non-coding RNAs were not taken into consideration).

As a result, we obtained a set of 354 genes putatively involved in AF genetic predisposition, including at least one gene per risk locus.

Intersection between GWAS and induced AF data:

Available transcriptomic data generated from a sheep model of induced AF (Alvarez-Franco *et al.*, 2020) was re-analyzed with more strict criteria, only selecting those genes differentially expressed in cardiomyocytes from both atria (comparing chronic AF sheep versus sham-operated controls). As a result, we identified a list of 209 dysregulated genes as a consequence of induced AF. Intersection of genes near GWAS hits and differentially expressed genes in sheep with chronic AF only shared four genes that were subjected to further study and functional experimentation.

Prioritization of candidate CREs at the *PCSK9* locus:

Expression of the *PCSK9* gene across adult human tissues (GTEx) and mouse embryos (Seidah *et al.*, 2003; Diez-Roux *et al.*, 2011) highlighted liver and cerebellum as relevant tissues in atherosclerosis either for its putative role in the disease or for potential adverse effects from its treatment. In order to identify putative regulatory elements of *PCSK9,* histone marks of active enhancers (H3K27ac and H3K4me1) were explored within human candidates and their orthologs in the mouse genome using available data from ENCODE and Roadmap Epigenomics for human adult liver, brain inferior temporal lobe, aorta, kidney, bone marrow mesenchymal stem cells and mouse adult liver, cerebellum, bone marrow-derived macrophages, kidney and lung (Encode Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). In order to confirm that the human HepG2 cell line was a good model to study hepatic *PCSK9* gene regulation, H3K27ac and H3K4me1 marks were included in the epigenetic analysis.

Candidate enhancers were prioritized using information from variants within the 220 kb locus that were associated to atherosclerotic feature (coronary artery disease, myocardial infarction or LDL-cholesterol levels) or correlated with *PCSK9* expression in cerebellum or liver tissue (eQTLs from GTEx; GTEx Consortium, 2020). SNPs in linkage-disequilibrium (LD) with the GWAS SNPs were imputed using SNAP (https://www.broadinstitute.org/snap/snap) and SNiPA (https://snipa.helmholtz-

muenchen.de/snipa3/) with a minimum r^2 of 0.8 (Arnold *et al.*, 2015; Pers, Timshel and Hirschhorn, 2015).

Results

While life in cell culture is far from complete,
death at any given moment is never one hundred per cent effective either.

[...]

When we throw out waste tissue culture,
we may be sure there's always something very small in there calling for help.
It's [...] the whisper of the last, lonely, useless, but nonetheless hopeful, hope.
No longer really science but still poetry.

Miroslav Holub, *Tissue culture, or about the last cell*.

1. Optimizing enhancer-reporter assays to scale *in vivo* enhancer detection.

The number of variants associated to disease keeps rising as larger number of individuals, both controls and cases, are included in the studies. CVDs are not an exception, which in the last two years have significantly increased the number of risk loci. In AF, the last two studies published in 2018, included more than one million individuals and found about one hundred new associations that made the total list of risk loci to reach 130 (Nielsen *et al.*, 2018; Roselli *et al.*, 2018). Similarly, for CAD impacting atherosclerosis, about 40 new loci have been identified in a very recent study, reaching 200 risk loci (Koyama *et al.*, 2020). Therefore, it is extremely important to focus on functional dissection of disease-associated loci which is nowadays the bottleneck after GWAS. ERAs have been the benchmark for interrogating the human genome and studying the effect of DNA variations (Pennacchio *et al.*, 2006; Kvon, 2015). Standard assessment of enhancer activity is achieved after pronuclear microinjection of mouse zygotes with linear constructs. ERAs rely on the random genomic integration of the constructs, which is a very inefficient, time-consuming and expensive process that requires a high number of animals.

In the first chapter of this thesis, we have focused on the improvement of mouse transgenesis to scale genomic interrogation.

1.1. The PB-ERA system is suitable to assess reporter expression in transient.

We aimed to increase the throughput of mouse *in vivo* ERAs in order to fast-forward GWAS-loci interrogation. For that, we developed a PB-ERA system, an enhancer assay assisted by transposition. This PB-ERA system relies on the *piggyBac* transposase (PB) and its ability to recognize PB-specific inverted terminal repeats (PBRs), integrating the inner DNA content into the genome. In particular, we generated a pPB- lacZ vector consisting of the transgene (-globin minimal promoter, lacZ reporter gene and cloning site for candidate genomic regions) which is surrounded by PBRs for the transposase recognition. Then, the pPB- lacZ vector was microinjected in an episomal way, either as an empty vector or containing enhancer sequences, together with the mRNA of a hyperactive version of PB (**Figure 10**).

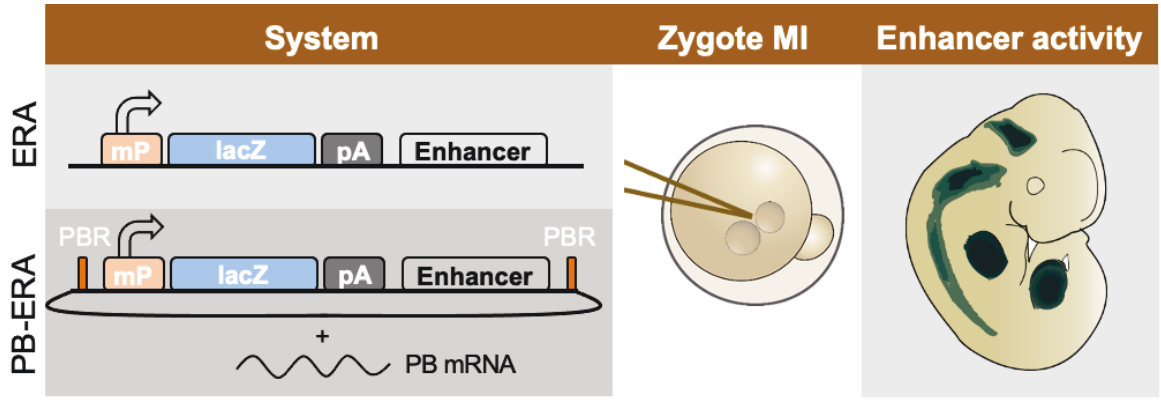

Figure 10 – Classical versus PB transgenesis. Schematics showing the ERA (linear) and PB-ERA (episomal) vectors that are microinjected into mouse zygotes and transferred to a foster mother until they reach the developmental stage of interest, usually E11.5, when we stain for β-galactosidase activity. MI, microinjection; mP, minimal promoter; pA, poly-adenylation signal; PBR, PB repeats.

In order to accelerate the transposition reaction, we microinjected the mRNA of the transposase instead of a DNA vector containing the PB sequence. The earlier the genomic integration of the construct, the lower the mosaicism of transgenic embryos. We confirmed the ability of the system to translate the PB mRNA and integrate the transgene early after zygote microinjection (Suzuki *et al.*, 2015), using a constitutive GFP cassette as a positive control (pPB-CAG-GFP) and measuring fluorescence. Collected embryos at embryonic day (E)10 were completely fluorescent (**Figure 11**), supporting transposase-mediated transgenesis as an effective system to evaluate reporter expression in transient.

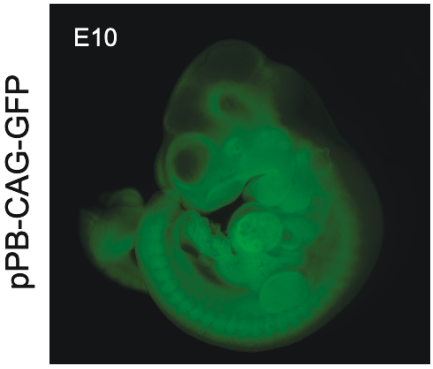

Figure 11 – Efficient transgene insertion using the PB-ERA system. Embryos microinjected with the pPB-CAG-GFP vector show high GFP signal with low degree of mosaicism.

1.2. A high-throughput method for transgenesis.

Zygote microinjection of the PB-ERA system results in reporter expression in embryos at a high rate. In order to measure the efficiency of the system in generating embryos, we genotyped transgenic embryos generated with the pPB- lacZ empty vector or with candidate enhancers. Surprisingly, we achieved an average transgenic rate of ~59% with the PB-ERA system, for transgenes as large as 14.4 kb (based on >200 transgenic mice from >20 independent transgenic constructs), compared to a ~9% of transgenic rate from classical transgenesis (**Figure 12A**). This >6-fold increase in the efficiency of transgenesis using a two-component method (one DNA vector plus one mRNA transcript) would allow to significantly reduce the time needed for multiple loci interrogation, together with number of animals and resources. However, an increased transgenesis rate does not imply an effective method for the assessment of enhacer activity. Indeed, classical transgenesis already suffers from position effects, which means that transgene expression might be affected by the place in genome where it is inserted. Therefore, if the improved transgenesis rate of the PB-ERA system is due to high number of transgene copies in the genome, this might lead to a situation in which the chances of getting at least one copy of the transgene inserted in an actively transcribed locus are also higher. While this is not a disadvantage if we wanted to obtain a gain-of-function model in which our gene of interest is expressed in the generated transgenic animal, it is indeed a problem when we interrogate the genome for functional activity. If this were the case, a candidate genomic region with no enhancer activity might also result in reporter expression (i.e. lacZ staining) leading to a high false discovery rate. In order to address this concern, we quantified the number of times that embryos had integrated the transgene. We carried out quantitative genotyping (see methods) to estimate the number of copies per embryo, and found an average of 3.4 insertions per transgenic embryos (**Figure 12B**), a rather low number. Therefore, we can rule out this as a problem of the system

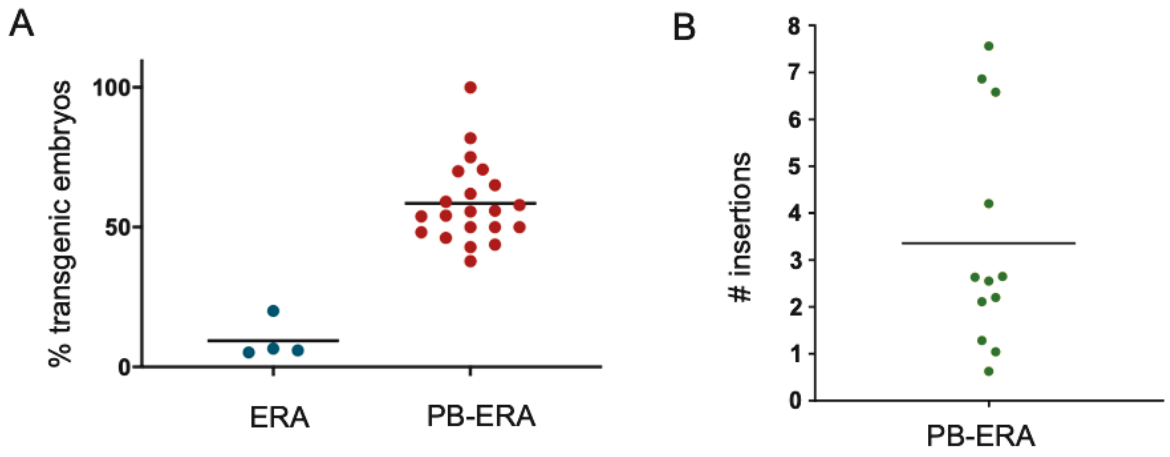

Figure 12 – Increased efficiency of transgenesis. A) Comparison between the percentage of transgenic embryos using classical transgenesis or the PB-ERA system. Each data point represents a candidate fragment tested for which several sessions of microinjection where carried out. **B)** Number of integrations in single embryos calculated by qPCR of their genomic DNA and relative to a mouse line with a single copy of the transgene. The PB-ERA system increased transgenic rate >6 fold while integrating in the genome a median of 3.36 copies of the transgene.

1.3. PB-ERA recapitulates the expression pattern of bona fide enhancers.

An increase in the efficiency of transgenesis of >6 fold, apparently not due to a huge number of integrations, seemed reasonable to assess functional activity through ERA. In order to assess the suitability of the PB-ERA system to capture enhancer activity, we cloned known enhancers in the pPB- lacZ vector and tested them by transgenesis. First, we tested a 2.2kb genomic region containing the human ZRS enhancer. This strong enhancer of *SHH* is active in the posterior limb bud and has been associated to polydactyly (Lettice *et al.*, 2003; Furniss *et al.*, 2008). As previously described, the PB-ERA system recapitulated the enhancer activity in E11.5 embryos, which we could reproduce in a large number of embryos (**Figure 13A**). Next, we cloned a heart enhancer since we wanted to explore the genetic contribution to CVD where we might potentially discover cardiac CREs. We tested a 600 bp intronic fragment containing the human left asymmetric enhancer (ASE) of the cardiac transcription factor *PITX2.* We obtained eight lacZ-stained E9.5 embryos out of sixteen recovered embryos. All positive embryos showed a left-side specific pattern, including the embryonic heart (**Figure 13B**), as it has been previously described (Shiratori *et al.*, 2001). Due to a problem extracting DNA from embryos, we could not genotype them so we could not detect transgenic embryos with no staining. Therefore, we estimated that between 9

and 10 embryos should be transgenic, taking into account our regular transgenic rate of 59% (ASE relative efficiency will be of relevance for **section 2.6** of this thesis).

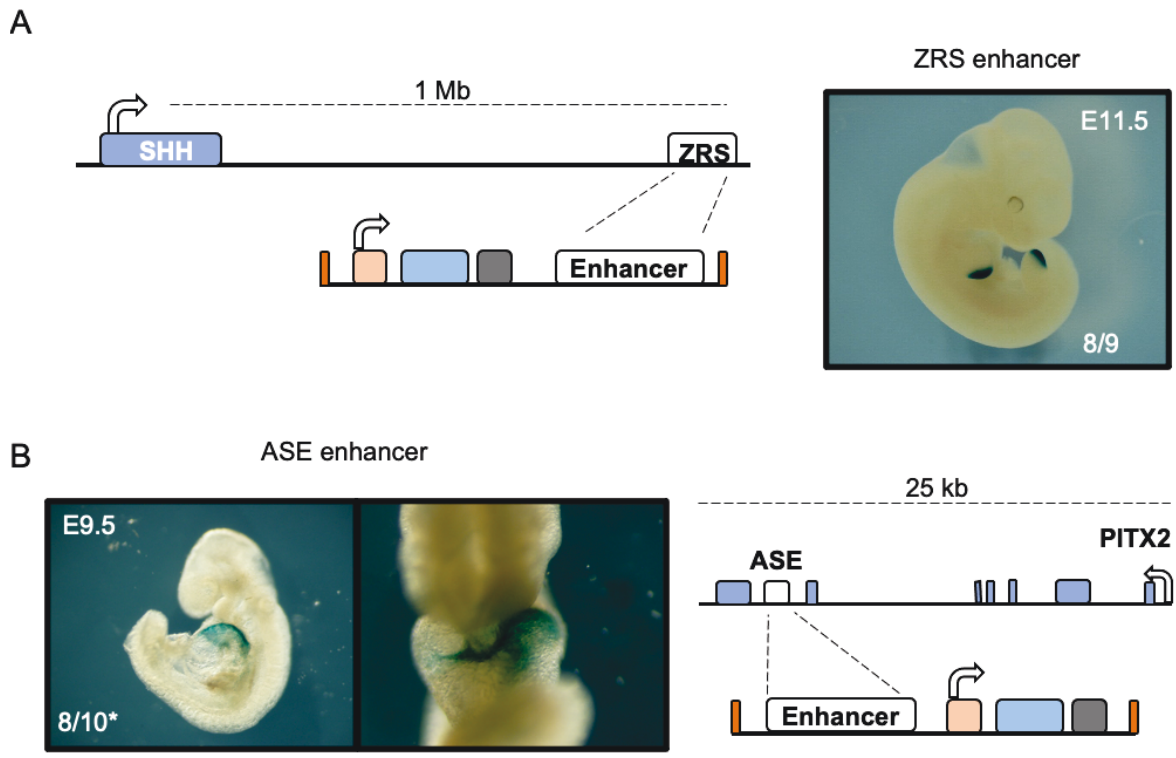

Figure 13 – The PB-ERA system is suitable to identify enhancers. The PB-ERA system recapitulates the tissue-specific patterns of bona fide human enhancers such as the limb ZRS enhancer of the *SHH* gene (**A**) and the heart asymetric enhancer (ASE) of *PITX2* (**B**). Asterisk (*) indicates estimated number of transgenic embryos as calculated taking into account the efficiency of the PB-ERA system.

In this chapter, we have focused on improving current *in vivo* tools for the assessment of regulatory activity. We have increased efficiency of transgenesis and capture enhancer activity in well-known regulatory elements involved in disease or cardiac development. Altogether, this data supports the use of the PB-ERA system for the *in vivo* interrogation of disease-risk loci, including those associated to CVDs.

2. Understanding the genetic component of AF: from GWAS signals to CREs.

Due to conserved evolutionary development, the mouse and the human are close relatives that share many developmental pathways and steps. Additionally, coding exons are much more conserved than non-coding sequences, which explains the high degree of similarity between human and mouse TFs. As a result, orthologous TFs from different organisms are expected to bind the same motifs in CREs. Indeed, in a

preprint from Ryu and collaborators, the authors characterized regulatory sequences from human and chimpanzee in iPSC-derived neural progenitors from both species, showing species-specific differences in enhancers regardless of the species of origin of the neural cells (Ryu *et al.*, 2018). Therefore, in this second chapter we used the mouse embryo as a test tube to characterize human genomic regions associated to CVDs.

2.1. Systematic *in vitro* PB-ERA finds regulatory elements in AF risk loci.

The increased output achieved by the PB-ERA system would allow us to undertake a more systematic evaluation of the regulatory activity of several genomic regions. Since SNPs and other variants such as CNVs might overlap CREs and modulate their regulatory activity, we followed this hypothesis and dissected ten loci associated with AF to functionally characterize the nature of these associations. In particular, we selected 5 kb surrounding the risk-associated SNPs for the nine strongest AF-loci detected by GWAS and the intronic CNV (4.3 kb) detected at the 5q35 locus containing the *KCNIP1* gene (**Table 8**). In order to prioritize candidate CREs, we assessed for the following criteria as predictors of regulatory activity in a tissue or cell type relevant for the disease (i.e. adult heart, fetal heart or differentiated cardiomyocytes from mouse and/or human iPSC): i) expression quantitative trait loci (eQTL) for a gene localized within the same topologically associated domain (TAD)(Dixon *et al.*, 2012); ii) H3K4me1, H3K27Ac histone marks of active enhancers (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011); iii) binding by cardiac TFs such as GATA4, TBX5 and NKX2-5 (Ang *et al.*, 2016; Luna-Zurita *et al.*, 2016). This first categorization showed that 90% of the selected loci were positive for at least one of the three predictors of enhancers, only with the exception of the *KCNIP1* locus, which suggests that many of the candidate regions might potentially be enhancers (**Figure 14A**).

Locus	Locus genes	Candidate CRE (name & coordinate-hg19)	SNPs within candidate	Ref (first identified)
16q22.3	*ZFHX3*	***ZFHX3*-AF** chr16:73049120-73054120	rs2106261 rs2359171	Benjamin *et al*. 2009
1q21.3	*KCNN3* *PMVK*	***KCNN3*-AF** chr1:154811768-154816768	rs6666258 rs13376333	Ellinor *et al*. 2010
1q24.2	*PRRX1* *GORAB*	***PRRX1*-AF** chr1:170566817-170571817	rs3903239	Ellinor *et al*. 2012
5q31.2	*WNT8A* *FAM13B*	***WNT8A*-AF** chr5:137417489-137422489	rs2040862	Ellinor *et al*. 2012

7q31.2	*CAV1* *CAV2*	***CAV1*-AF1** chr7:116183741-116188741	rs3807989	Ellinor *et al*. 2012
7q31.2	*CAV1* *CAV2*	***CAV1*-AF2** chr7:116188801-116193801	rs11773845	Ellinor *et al*. 2012
9q22	*C9orf3|AOPEP*	***C9orf3*-AF** chr9:97710959-97715959	rs10821415	Ellinor *et al*. 2012
10q22.2	*SYNPO2L* *MYOZ1*	***SYNPO2L*-AF** chr10:75419016-75423327	rs10824026 rs6480708	Ellinor *et al*. 2012
14q23.2	*SYNE2* *ESR2*	***SYNE2*-AF** chr14:64678348-64683348	rs1152591 rs2738413	Ellinor *et al*. 2012
15q24.1	*HCN4*	***HCN4*-AF** chr15:73649674-73654674	rs7164883	Ellinor *et al*. 2012
5q35.1	*KCNIP1*	***KCNIP1*-AF** chr5:170128728-170132986	CNV	Tsai *et al*. 2016

Table 8 – Summary of the AF risk loci tested using the PB-ERA system, including nine risk SNPs identified by GWAS and one CNV.

According to epigenetic predictors, three loci stood out from the rest as likely to contain disease-related enhancers: 7q31 including *CAV1*, 9q22 including *C9orf3,* and 14q23 including *SYNE2*. We cloned the human sequence of all ten selected candidates into the pPB- lacZ vector in order to functionally assess their enhancer activity. These constructs were transfected into HL-1 cells which are a cultured model for mouse atrial cardiomyocytes that recapitulates electrophysiological features of cardiomyocytes and expresses cardiac genetic programs (Claycomb *et al.*, 1998). HL-1 cells express ion channels, members of the gap junctions and TFs, including *TBX5*, *GATA4*, *NKX2-5*, and the atrial-specific *PITX2*. HL-1 cells were co-transfected with the pPBase vector that expresses PB to enable transposition. As a positive control, we used the same ASE enhancer of *PITX2* that we previously tested in embryos. As a negative control, we used the pPB- lacZ reporter construct containing the *Oct4* distal enhancer (DE) specifically active in mouse pluripotent stem cells (Yeom *et al.*, 1996). We found that two of the candidates activated reporter expression in cardiomyocytes. These candidates were *SYNE2*-AF that contains the SNPs rs2738413 and rs1152591, and *C9orf3*-AF that contains the SNP rs10821415 (**Figure 14B**; **Table 8**). Interestingly, these two genomic regions had qualified as very likely to contain enhancers in our analysis above (**Figure 14A**). This data indicates that AF risk loci can act as enhancers in cardiomyocytes.

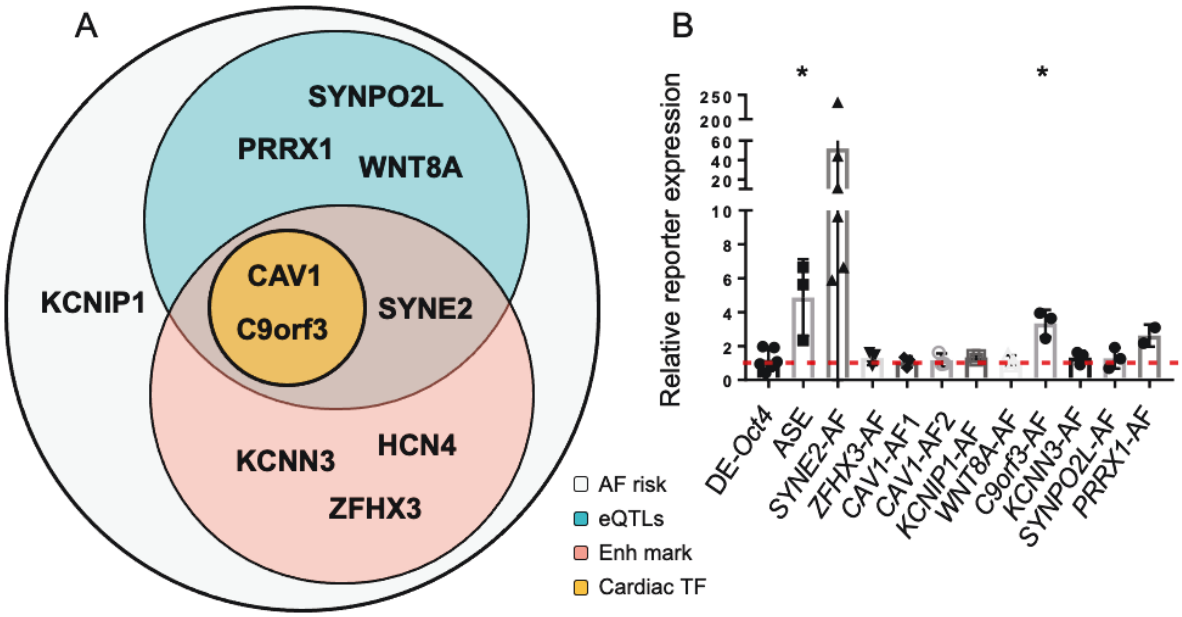

Figure 14 – *SYNE2* and *C9orf3* are candidate heart enhancers. A) Schematic representation of the regulatory features of the ten loci included in this study. The four categories included in the Venn diagram are the presence of AF SNPs or CNVs, eQTLs for the heart expression of any of the genes in the locus, histone marks of active enhancers (H3K4me1 and/or H3K27Ac), ChIPseq peak for the cardiac TFs TBX5, GATA4 or NKX2-5 in human or mouse differentiated cardiomyocytes. **B)** Enhancer activity of the candidate loci represented as RNA/DNA ratio after transfection of PB-ERA constructs in the mouse HL-1 cardiac cell line. The distal enhancer (DE) of Oct4 is used as negative control and the ASE enhancer as positive control. Asterisks indicate p-value < 0.05.

2.2. The PB-ERA system identifies *in vivo* enhancers that are not detected in tissue culture assays.

Our results suggest that AF risk loci are interesting genomic regions that contain features of enhancers and activity in cultured cardiomyocytes. However, it was surprising that other prioritized regions did not show enhancer activity in HL-1 cells. This might be due to *in vitro* tests which have important limitations as they are simpler models. Therefore, we used the PB-ERA system to test the enhancer activity of all ten selected AF-associated loci *in vivo,* by means of transient mouse embryo transgenesis. For that we dissected and examined transgenic embryos at stage E11.5, when the four-chamber heart is already formed and functional, therefore having a higher probability to capture regulatory activity related to AF. We tested a 4.3 kb fragment located in the first intron of the *KCNIP1* gene (**Figure 15A**), which has been described as a copy number variation (CNV) associated to AF (Tsai *et al.*, 2016). This fragment acted as a cardiac enhancer, driving reporter expression in scattered cells

throughout the heart (**Figure 15B and B'**). This was rather unexpected, as the *KCNIP1*-AF region did not show any of the predictors for enhancer activity we previously used (**Figure 14A**). Finding a heart enhancer in this region highlights our limited understanding of the epigenetic code.

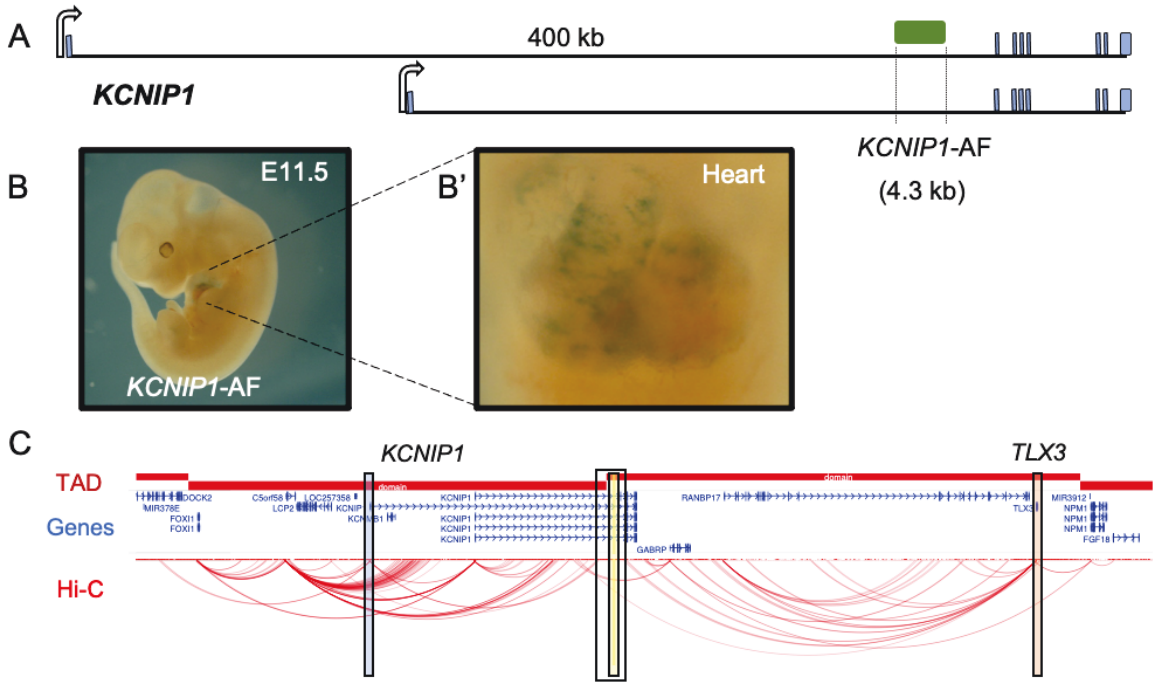

Figure 15 – *KCNIP1*-AF in the 5q35 risk locus is a heart enhancer. A) *KCNIP1*-AF is located in the first intron of the *KCNIP1* gene and 350 kb downstream its promoter region. **B)** Cardiac-specific activity of the *KCNIP1*-AF enhancer in E11.5 transgenic embryos with expression in the four chambers of the heart (**B'**). **C)** Genomic view of the 1.5 Mb region (hg19; chr5:169,450,000-170,900,000) including *KCNIP1* and *TLX3*. Indicated in the left and from top to bottom are topologically associated domains (TAD); annotated genes; promoter-capture Hi-C data of differentiated cardiomyocytes from hiPSC (data from Montefiori *et al*., 2018), represented as red arcs. The *KCNIP1*-AF enhancer (yellow rectangle) is located at the boundary region between two consecutive TADs and nearby intronic regions (wider transparent rectangle) interact with the promoters of *KCNIP1* (blue rectangle) and *TLX3* (orange rectangle).

The intron where this enhancer is located interacts with the promoter of the *KCNIP1* gene as measured by promoter-capture Hi-C in hiPSC-CM (**Figure 15C**; Montefiori *et al.*, 2018). Indeed, *KCNIP1* has been shown to be expressed in all four chambers of the heart (Tsai *et al.*, 2016), similar to the enhancer activity that we detected in our transgenic embryos (**Figure 15B'**). Chromatin analysis of the risk locus showed that the intron within which this enhancer is located also interacts with the promoter of the *TLX3* gene that is located ~600 kb downstream of the enhancer (**Figure 15C**). Further studies will be needed to address a possible role of *TLX3* in the pathophysiology of AF.

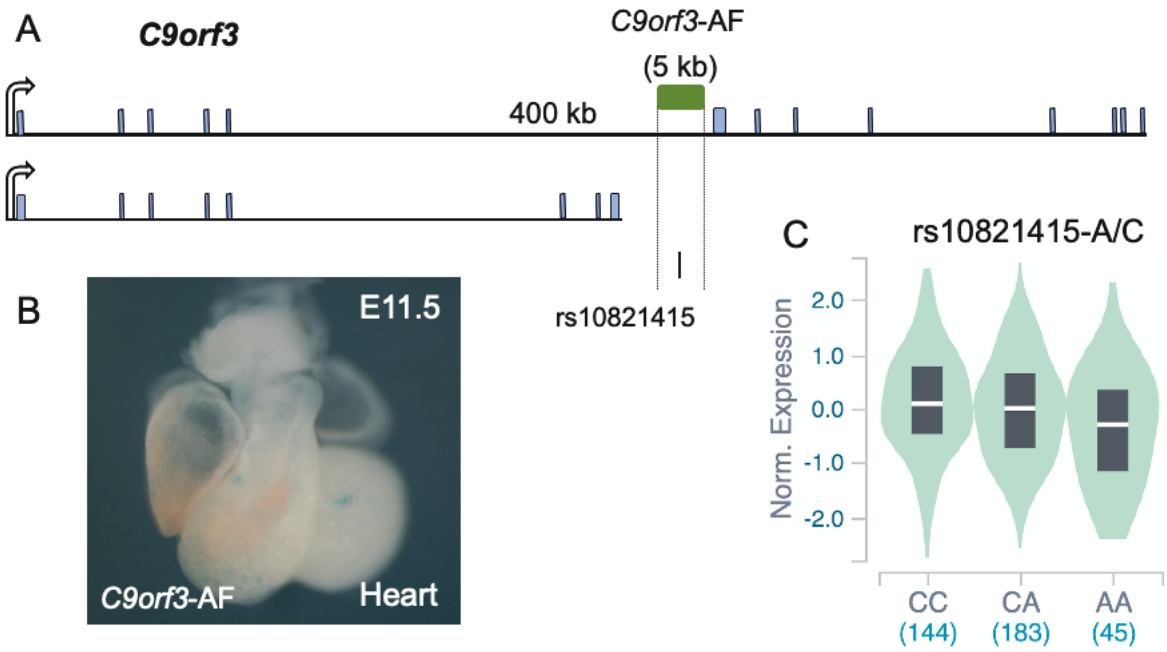

Figure 16 – Reporter assays in vivo validate the enhancer detected in the 9q22 locus. A) The *C9orf3*-AF enhancer is located in the large intron of the *C9orf3* gene and contain the variant rs10821415. **B)** Dissected heart of E11.5 transgenic embryos showing subtle reporter expression. **C)** Transcriptomic data from GTEx indicates that the risk A allele correlates with lower *C9orf3* gene expression in atrial tissue (GTEx Consortium, 2020).

Both *C9orf3*-AF and *SYNE2*-AF candidate genomic regions showed cardiac-enhancer activity in mouse embryos as well as in cell culture (**Figure 14B**). The intronic *C9orf3*-AF region (**Figure 16A**) showed only mild expression of the reporter in the heart (**Figure 16B**), but the risk associated allele at rs10821415 correlated with lower expression of *C9orf3* in atrial tissue indicating that it might be the target gene of this cardiac enhancer (**Figure 16C**; GTEx data GTEx Consortium, 2020). On the other hand, the *SYNE2*-AF region (**Figure 17A**) was active in the outflow tract (OFT) and in the lungs of both E11.5 and E14.5 embryos (**Figure 17D-F**). *SYNE2* is a large gene that produces two different transcripts (**Figure 17A**). Interestingly, the *SYNE2*-AF region overlaps the start of the shorter *SYNE2* isoform and the two SNPs associated to AF in this region lie upstream in close proximity to its TSS (**Figure 17A**). Exploring the expression data of the two *SYNE2* in GTEx, we observed that the short isoform is predominantly expressed in skeletal muscle and in the heart (**Figure 17B**). In the human atria, the expression of *SYNE2* correlated with the genotype of both rs2738413 (**Figure 17C**; GTEx data) and rs1152591 (not shown) variants.

It is precisely in the atria where the cardiac expression of the mouse *Syne2* gene is higher (**Figure 18A**; RNA-seq data from 3D-Cardiomics; Mohenska *et al*., 2019).

Although *SYNE2*-AF might contain an alternative promoter for the small muscle isoform of *SYNE2*, that putative promoter activity would be independent from the enhancer activity described here, due to the cloning strategy that we followed in which the insertion site in the pPB- lacZ vector is located downstream of the lacZ. Therefore, the *SYNE2*-AF enhancer harbors cis-regulatory potential capable of triggering reporter expression from downstream. Indeed, if we check the Hi-C data in differentiated cardiomyocytes from hiPSC we can see long-range interactions between the *SYNE2*-AF enhancer and the promoter of the large isoform of *SYNE2* as well as the promoter of *ESR2* (**Figure 18B**; Montefiori *et al.*, 2018). Interestingly, the genotype of both SNPs also correlated with *ESR2* and *MTHFD1* expression in the lungs (not shown), where we have identified that *SYNE2*-AF enhancer is also active (**Figure 17E and F**). However, the Hi-C data do not capture interactions between the *SYNE2*-AF enhancer and the promoter of *MTHFD1*. Altogether, these data suggest that *SYNE2*-AF is a human enhancer regulating *ESR2* in the lung and *SYNE2* in the atria, from which the link with AF might potentially come.

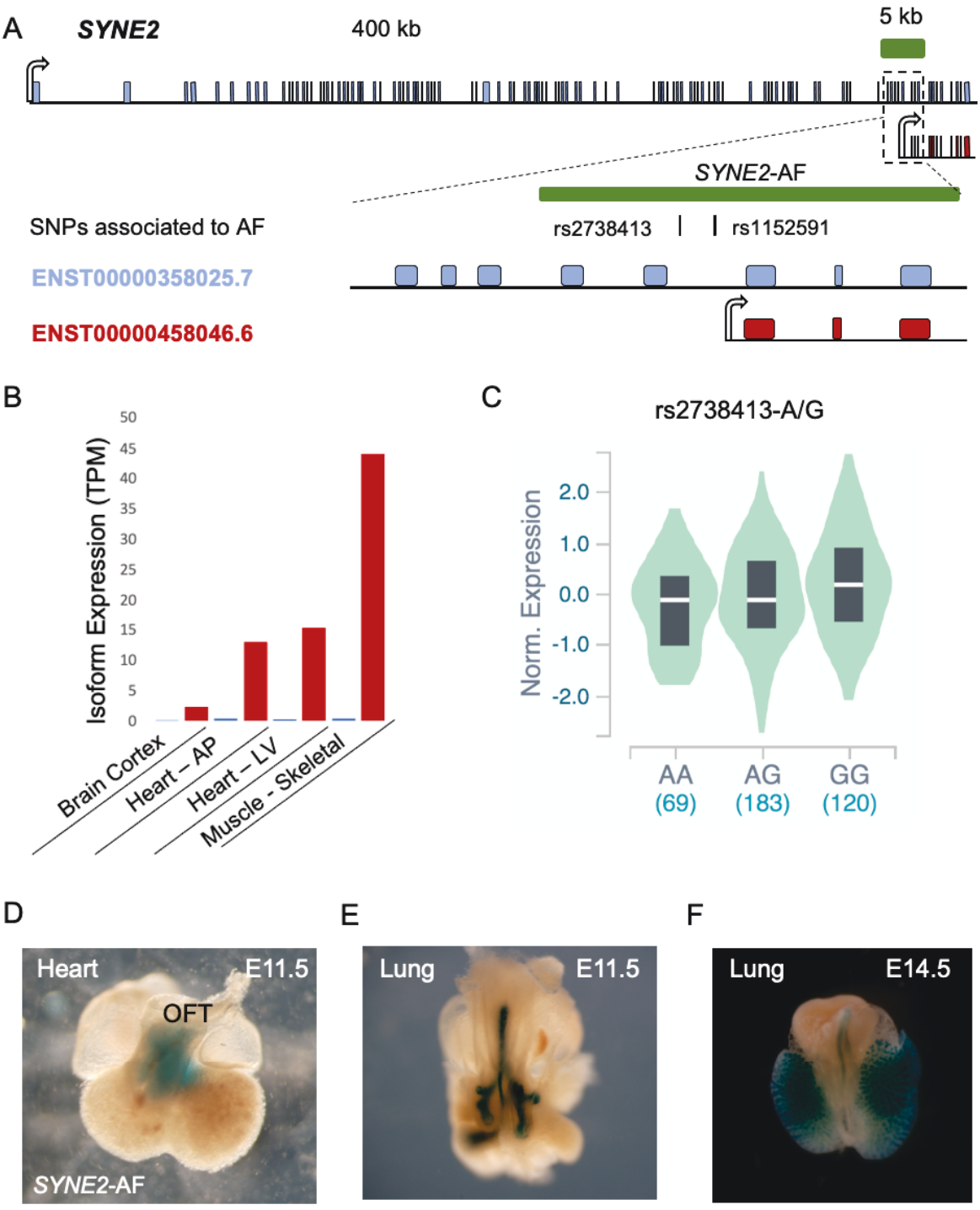

Figure 17 – The 14q23 contains a regulatory element active in heart and lung tissues. A) The *SYNE2* gene encodes a giant (blue) and a small (red) isoforms. AF variants rs2738413 and rs1152591 overlap the promoter region of the small isoform and were tested for enhancer activity (*SYNE2*-AF fragment). **B)** The small isoform (red) is mainly expressed in muscle. **C)** Risk A allele of variant rs2738413 correlated with lower *SYNE2* gene expression in atrial tissue. Variant rs1152591 showed identical results and is not shown for simplification. The PB-ERA system identified enhancer activity in the heart (**D**) and the lungs (**E** and **F**) for the *SYNE2*-AF fragment. OFT, outflow tract.

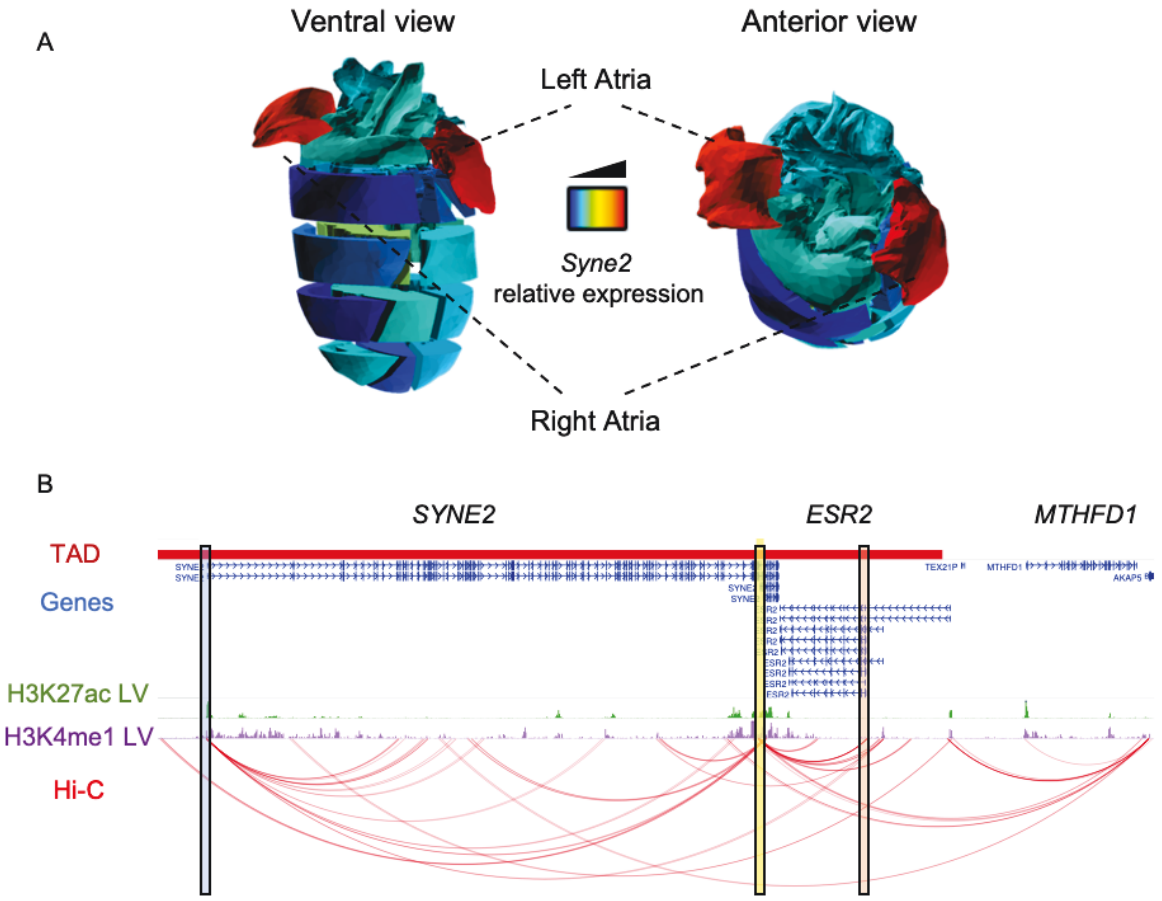

Figure 18 – Potential target genes of SYNE2-AF. A) Spatial representation of RNA-seq data from mouse heart tissue (3D-Cardiomics) indicates that *SYNE2* gene expression is maximum in the atria (data from Mohenska *et al.*, 2019). **B)** Genomic view of the 650 kb region (hg19; chr14:64,290,000-64,940,000) including *SYNE2*, *ESR2* and *MTHFD1*. Indicated in the left and from top to bottom are topologically associated domains (TAD); annotated genes; ChIP-seq data from the Roadmap epigenomics for the histone marks H3K27ac (green) and H3K4me1 (purple) of left ventricle (LV); promoter-capture Hi-C data of differentiated cardiomyocytes from hiPSC (data from Montefiori *et al.*, 2018), represented as red arcs. The *SYNE2*-AF enhancer (yellow rectangle) interacts over the long range with *SYNE2* (blue rectangle) and *ESR2* (orange rectangle) promoters.

2.3. The 7q31 locus contains CREs conserved in mammals and controlled by cardiac TFs.

We selected two candidate CREs (*CAV1*-AF1 and *CAV1*-AF2) at the 7q31 locus spanning 5 kb each and located at the large second intron of *CAV1* (**Figure 19A**). The candidates contain the variants rs3807989 and rs1173845 which are among the strongest GWAS-SNPs associated to AF (Ellinor *et al.*, 2012) as well as other variants associated to electrophysiological traits (Christophersen, Magnani, *et al.*, 2017)(**Figure 19B**). The 3D chromatin landscape of differentiated cardiomyocytes

from hiPSC indicates that this region is within a highly interactive locus, involving *CAV1*, *CAV2*, *TES*, *MET* and *CAPZA2* genes (**Figure 19A**; from Montefiori *et al*., 2018). Available epigenomic data shows the presence of enhancer marks (H3K4me1 and H3K27Ac) overlapping the candidate regions in human samples from adult left ventricle (**Figure 19B**). When tested using the PB-ERA system, *CAV1*-AF1 and *CAV1*-AF2 drove reporter expression in the heart of E11.5 transgenic embryos (**Figure 19C-D'**). Indeed, *CAV1*-AF1 is bound by cardiac TFs (GATA4 and TBX5) in human differentiated cardiomyocytes (**Figure 20 upper panel**; Ang *et al.*, 2016) as well as the orthologous mouse region, which is bound by GATA4, TBX5 and NKX2-5 in mouse differentiated cardiomyocytes (**Figure 20 bottom panel**; Luna-Zurita *et al.*, 2016). Together, this data shows that *CAV1*-AF1 and *CAV1*-AF2 are regulatory elements active in the heart, whose function is very likely conserved in mammals.

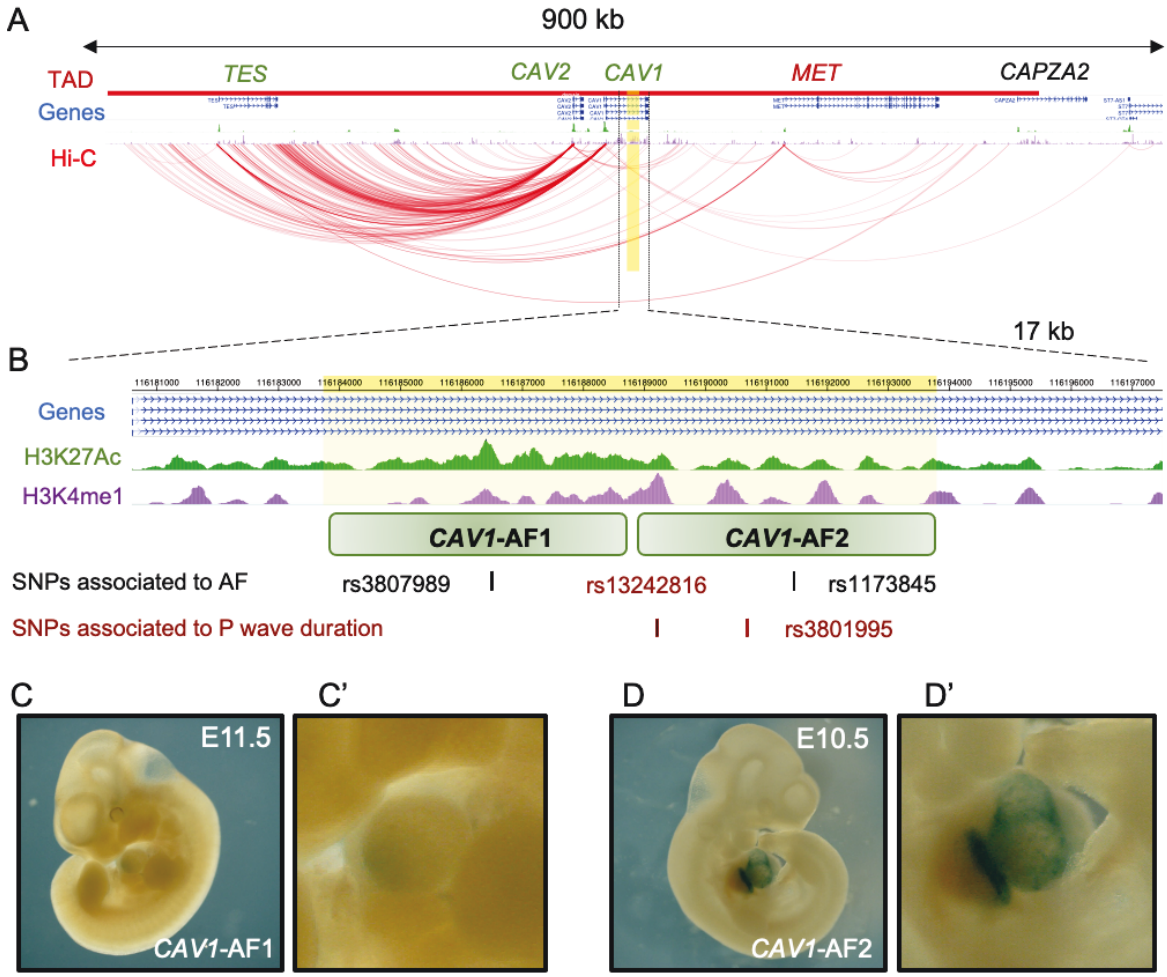

Figure 19 – The 7q31 AF risk locus contains cardiac regulatory elements. A) Genomic view of the 900 kb region (hg19; chr7:115,760,000-116,660,000) including *CAV1*, *CAV2*, *TES*, *MET* and *CAPZA2*. Indicated in the left and from top to bottom are topologically associated domains (TAD); annotated genes; promoter-capture Hi-C data of differentiated cardiomyocytes from hiPSC (data from Montefiori *et al*., 2018), represented as red arcs. Gene names in different colors indicate whether the location of their promoters are upstream (green) or downstream (red) the candidate enhancer regions (yellow rectangle), or that the genes are partially at a different TAD (black). **B)** A zoomed 17 kb view (hg19; chr7:116,180,600-116,197,400) of the genomic region in **A**, including ChIP-seq data from the Roadmap epigenomics for the histone marks H3K27ac (green) and H3K4me1 (purple) of left ventricle (LV). Marks of active enhancers are enriched in the two candidate regions *CAV1*-AF1 and *CAV1*-AF2 containing AF and electrophysiological (P wave) variants. The PB-ERA system identified enhancer activity in the heart of transgenic embryos for the candidate regions containing the rs3807989 (**C** and **C'**) and the rs1173845 (**D** and **D'**) variants.

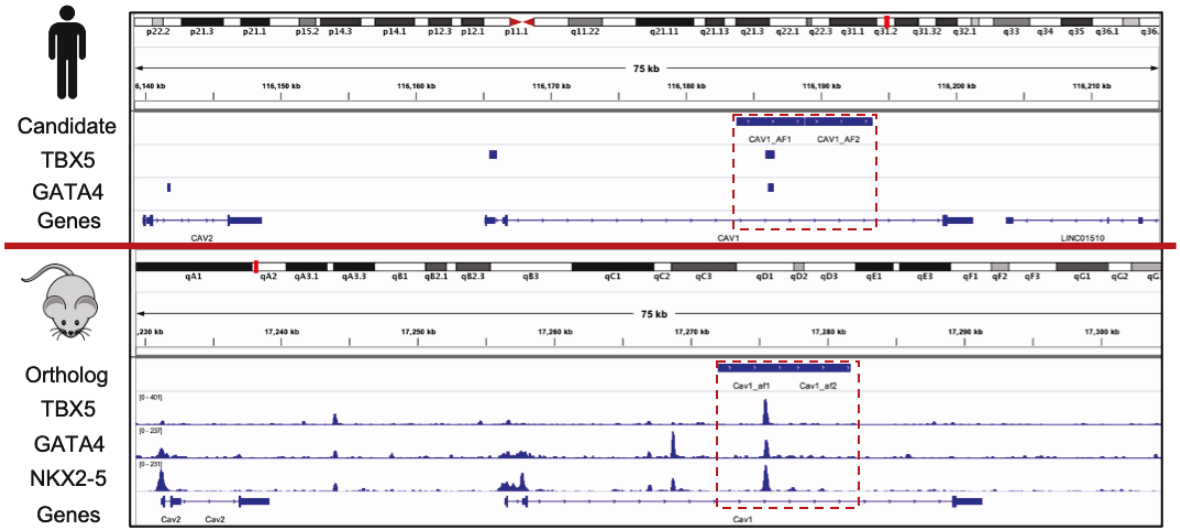

Figure 20 – Conserved cardiac TF binding at the 7q31 regulatory elements. A genomic window of 76 kb is shown for the human risk loci (upper panel; hg19; chr7:116,139,000-116,215,000) and the mouse orthologous region (bottom panel; mm9; chr6:17,229,500-17,305,000). ChIP-seq data for cardiac TFs in differentiated cardiomyocytes from hiPSC (TBX5 and GATA4 in upper panel; data from Ang *et al*., 2016) and mESC (TBX5, GATA4 and NKX2-5 in bottom panel; data from Luna-Zurita *et al*., 2016) are shown. The red dashed rectangle indicates the enhancer region which is bound by the cardiac TFs in both organisms.

2.4. AF-associated regulatory elements at the 7q31 locus differentially regulate upstream and downstream genes.

CAV1 encodes Caveolin-1, the main component of caveolae which are small membrane invaginations present in many cells, including cardiomyocytes (Parton and Del Pozo, 2013). Caveolin-1 might play a relevant role in the electrophysiological and mechanical properties of cardiomyocytes interacting with ion channels (Lin *et al.*, 2008) and gap junction proteins (Langlois *et al.*, 2008) and since caveolae regulate plasma membrane curvature which prevents membrane rupture (Cheng *et al.*, 2015). Accordingly, knockout (KO) mice for *Cav1* have cardiac conduction affected and develop ventricular arrhythmias (Yang *et al.*, 2014). The identification of cardiac enhancers at the AF-risk locus might help elucidate the mechanism behind the association. However, there is no direct evidence that these enhancers control the expression of *CAV1* or any other gene within the TAD.

The 900 kb of the 7q31 locus are highly interconnected and the large intron of *CAV1* that contains the regulatory elements is involved in long-range chromatin interactions. Nevertheless, enhancer-gene interaction is not sufficient to determine direct gene

regulation. In order to identify the target genes regulated by *CAV1*-AF1 and *CAV1*-AF2, we used the CRISPR/Cas9 system to disrupt the CREs. Since we observed human and mouse conservation both at the sequence and at the functional levels, we used HL-1 atrial-like cardiomyocytes as a cellular model. Therefore, we designed guide RNAs to delete the mouse orthologous CREs separately in HL-1 cells (**Figure 21A**). The deletion of each regulatory element led to downregulation of *Cav1*, *Cav2* and *Tes,* the three genes whose promoters where upstream, confirming that *Cav1*-af1 and *Cav1*-af2 are acting as enhancers of such genes (**Figure 21B**). More surprisingly, *Cav1*-af1 and *Cav1*-af2 seemed to be negatively regulating *Met*, as shown by its upregulation upon deletion of the CREs (**Figure 21C**). On the other hand, CRE deletion did not affect the expression of *Capza2* which is located mostly outside the TAD (**Figure 21C**). Since AF variants at the 7q31 locus are within cardiac CREs controlling gene expression, we wanted to investigate whether the variants rs3807989 and rs1173845 located within *CAV1*-AF1 and *CAV1*-AF2, respectively, were at an essential core domain of the enhancers. To do that we designed guide RNAs targeting minimal regions of ~600 bp (min*Cav1*-af1 and min*Cav1*-af2) in the mouse genome that were orthologous to the human regions surrounding each AF-SNP. We demonstrated that both minimal enhancers are essential for proper regulatory function since their deletion in HL-1 cells led to similar levels of downregulation of *Cav1*, *Cav2* and *Tes,* and upregulation of *Met* that the full *Cav1*-af1 and *Cav1*-af2 elements (**Figure 21B and C**).

Altogether, our data suggest that *CAV1*-AF1 and *CAV1*-AF2 are highly conserved heart-specific CREs that act differentially on upstream and downstream genes. These CREs contain AF variants at essential core domains necessary for their regulatory activity.

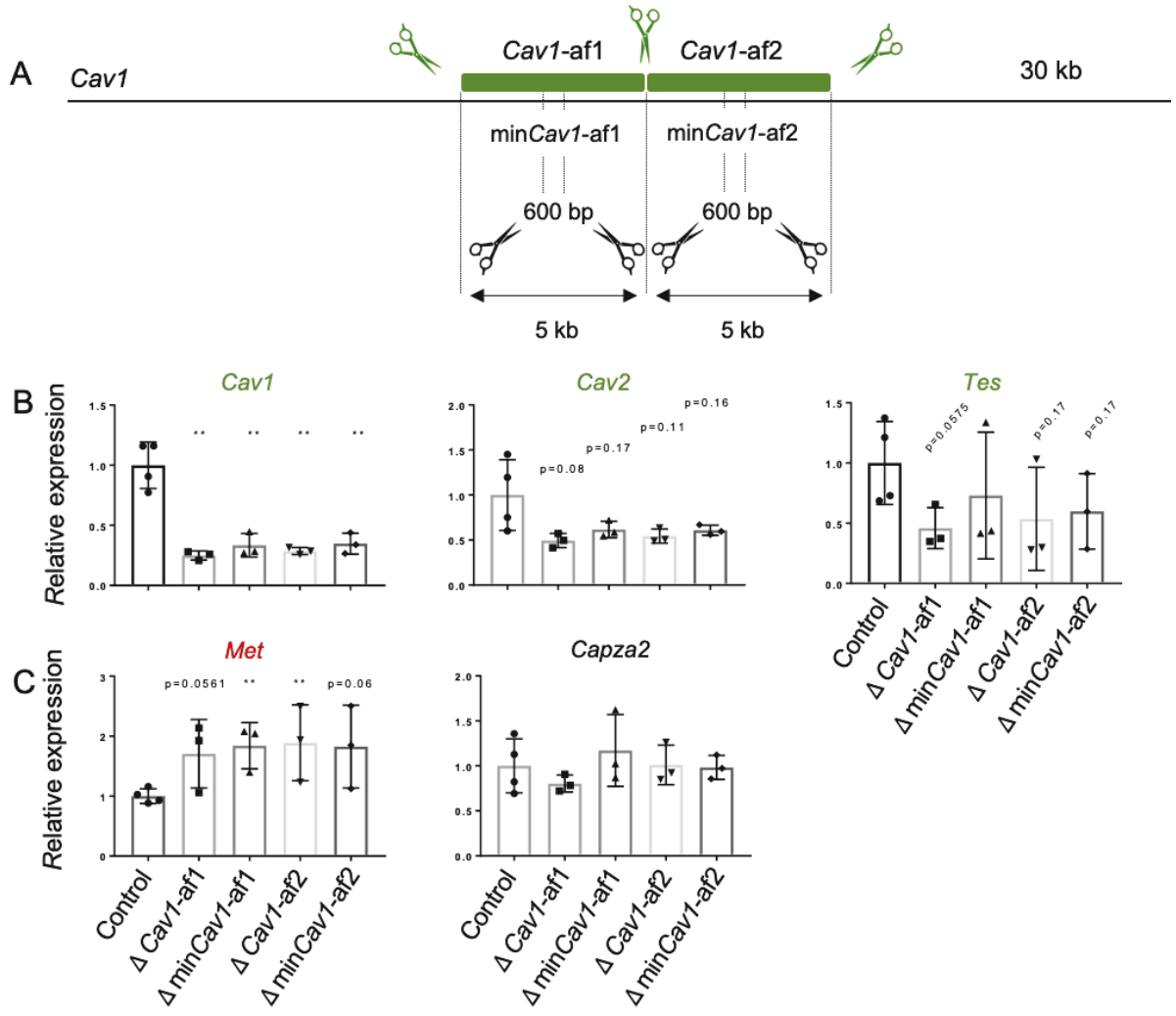

Figure 21 – Enhancer deletion involves new genes in AF. A) Schematic of the mouse orthologous region containing candidate CREs in the large intron of *Cav1*. The designed deletions for each of the orthologous enhancers are indicated. Enhancer deletions in HL-1 cardiac cells led to downregulation of upstream genes *Cav1*, *Cav2* and *Tes* (**B**) and to upregulation of the downstream *Met* gene, while not affecting *Capza2* (**C**).

2.5. Identification of a negative regulator at the 16q22 AF locus controlling *ZFHX3* gene expression.

The 16q22 locus has been associated to AF together with the 4q25 associations near *PITX2*, since the first genetic reports more than a decade ago (Gudbjartsson *et al.*, 2007; Benjamin *et al.*, 2009). Whereas variants at the 4q25 locus are intergenic, GWAS-SNPs at the 16q22 locus are located within the first intron of *ZFHX3*. Previously, dissection of the 4q25 locus has identified long-range CREs regulating *PITX2* and *ENPEP* gene expression (Aguirre *et al.*, 2015; Ye *et al.*, 2016; Zhang *et*

al., 2019). However, we ignore functional evidence that *ZFHX3* or other genes are regulated by the 16q22 AF-risk locus (van Ouwerkerk *et al.*, 2019). To test this, we investigated the regulatory role of *ZFHX3*-AF, a 5 kb region in the first intron of *ZFHX3* that contains two variants associated with AF, rs2106261 and rs2359171. Of note, the epigenetic landscape of the locus shows that the repressive chromatin mark H3K27me3 overlaps the *ZFHX3* promoter in many tissues, including the embryonic and adult heart (**Figure 22A**). The only exception to that is the aorta, which is precisely the tissue with the higher expression of *ZFHX3* (**Figure 22B**). Marks of repression extend for 50 kb spanning from the promoter to our region of interest (highlighted in yellow) in some tissues of neural fate as well as in pluripotent cells. Again, repressive marks reversely correlate with gene expression since brain samples are among the lowest *ZFHX3* expressing tissues (**Figure 22B**).

We used the PB-ERA system carrying the candidate CRE and recovered transgenic embryos at E11.5. It was not surprising to see that the intronic *ZFHX3*-AF (**Figure 23A**) did not show enhancer activity in embryos like in our *in vitro* assay in HL-1 cardiomyocytes (**Figure 14B**). Instead, the presence of H3K27me3 across multiple tissues made us wonder whether there could be negative regulators within this candidate region acting ubiquitously. We hypothesized that ubiquitous negative regulators should be able to reduce or vanish gene expression globally.

It should be noted that the high efficiency of transgenesis using the PB-ERA system led to unspecific lacZ expression and -galactosidase activity in up to 50% of embryos, as we observed when no enhancer was placed in the construct and we generated transgenics with the empty vector. Noise was easily discriminated from signal due to absence of reproducibility or lower intensity than bona-fide enhancer-driven -galactosidase activity. Taking into account this property of the PB-ERA system, it was surprising that the *ZFHX3*-AF genomic region dramatically reduced unspecific activity as compared to the empty pPB- lacZ vector (**Figure 23B**).

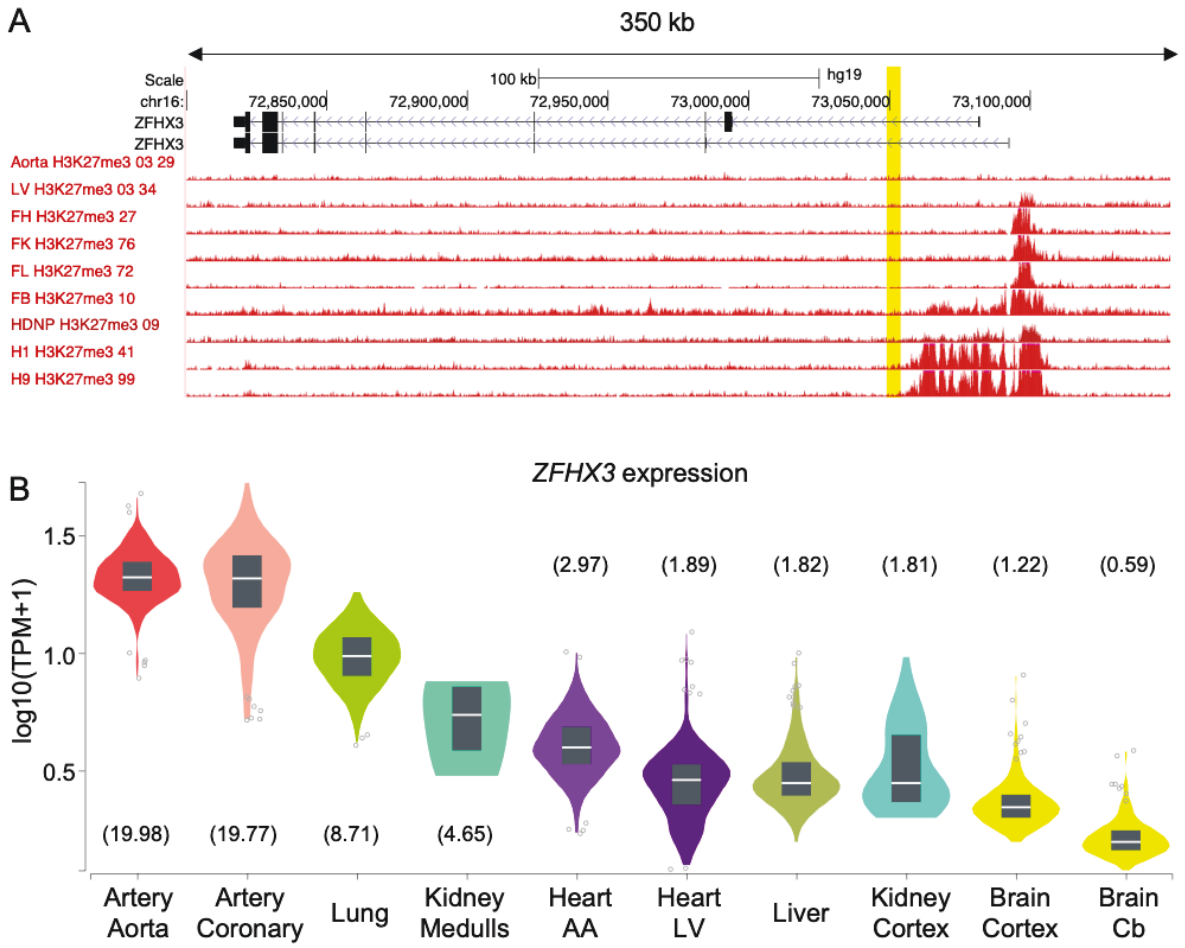

Figure 22 – Repression marks at the 16q22 locus. A) Genomic view of a 350 kb region (hg19; chr16:72,800,000-73,150,000) containing the *ZFHX3* gene and the *ZFHX3*-AF region and including the repressive H3K27me3 mark overlapping the *ZFHX3* promoter in human tissues, including the embryonic and adult heart. H3K27me3 signal extended for 50 kb spanning from the promoter to the AF association (yellow) in some tissues. Tissues from top to bottom: aorta, left ventricle, fetal heart, fetal kidney, fetal lung, fetal brain, H1 derived neuronal progenitor, H1 ESC, H9 ESC (Roadmap epigenomics data). **B)** Human *ZFHX3* gene expression across multiple tissues (data from GTEx). H3K27me3 (**A**) and gene expression (**B**) negatively correlates at the promoter of *ZFHX3*.

Next, we wanted to explore the functional relevance of this potential negative regulator by identifying its target gene. The 16q22 locus contains the *ZFHX3* gene surrounded by two gene deserts of approximately 1 Mb at each side, making *ZFHX3* itself almost the only candidate gene. Since the reduction of unspecific activity in transgenic embryos carrying the *ZFHX3*-AF genomic fragment was not tissue-specific, we hypothesized that perturbation of this putative regulatory element might affect gene expression in several cell types. To address this, we deleted *ZFHX3*-AF in two human cell lines of different origin and morphology such as the adherent embryonic kidney HEK293T cells and the lymphoblastic myeloid K562 cells. Consistently, the deletion

of the *ZFHX3*-AF region led to upregulation of *ZFHX3* in both cell lines (**Figure 23C and D**). Our data suggest that the *ZFHX3*-AF genomic region is a CRE negatively regulating *ZFHX3* gene expression presumably in different cell types. Finding a negative regulator in the 16q22 locus indicates that the nature behind this association to AF might be related to gene repression rather than enhancement. Furthermore, it brings out the PB-ERA system as a useful tool to identify negative regulators.

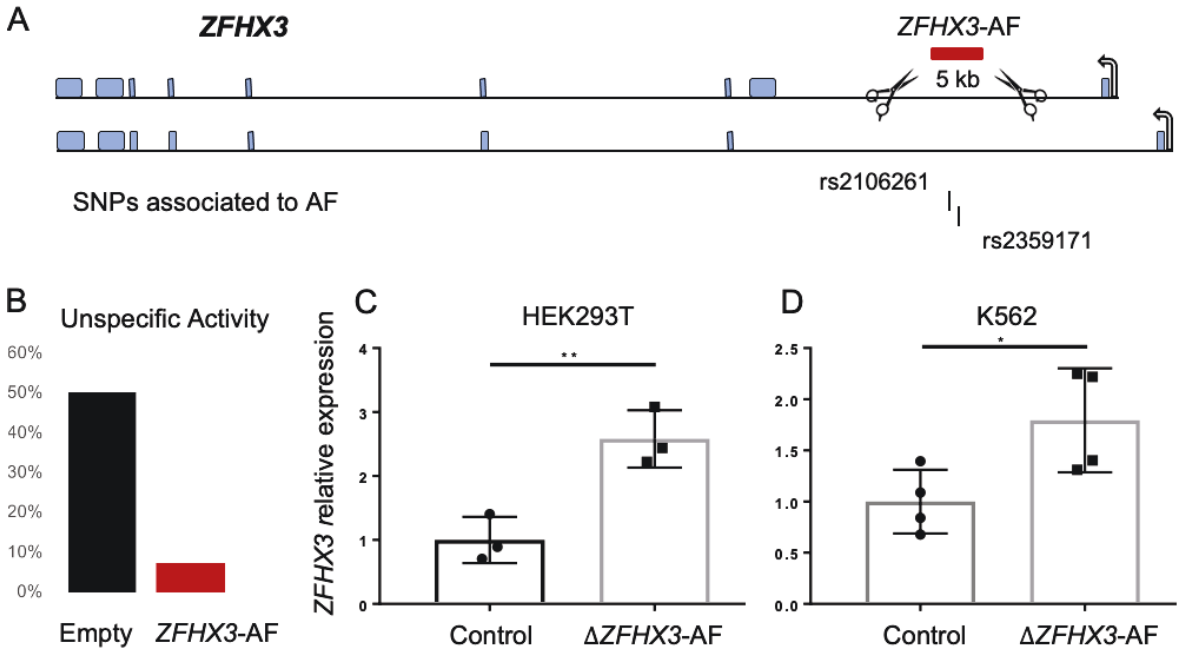

Figure 23 – The candidate ZFHX3-AF is a ubiquitous negative regulator. A) Representation of the AF-associated *ZFHX3* gene containing the risk variants rs2106261 and rs 2359171. The candidate tested region is shown in red. **B)** While unspecific activity is seen in the PB-ERA system (empty vector shown in black), this is dramatically reduced in transgenic embryos carrying the *ZFHX3*-AF candidate (red; appearing in only 1 out of 14 embryos). Deletion of the *ZFHX3*-AF element resulted in overexpression of *ZFHX3* in the human cell lines HEK293T (**C**) and K562 (**D**).

2.6. The *ZFHX3*-AF silencer is human-specific and outcompetes heart enhancers *in vivo* independently from its relative position.

We have shown that *ZFHX3*-AF at the 16q22 AF risk locus regulates gene expression at its native position and out of context, being a negative regulator of *ZFHX3* in different cell types. In order to see whether the mouse orthologous region of this regulatory element was active in mouse cardiac cells, we deleted a 6.3 kb intronic region (*Zfhx3*-af) from the genome of HL-1 cells (**Figure 24A**). However, no

differences were found between control and cells lacking *Zfhx3*-af (**Figure 24B**), suggesting no conservation of this regulatory element in the mouse.

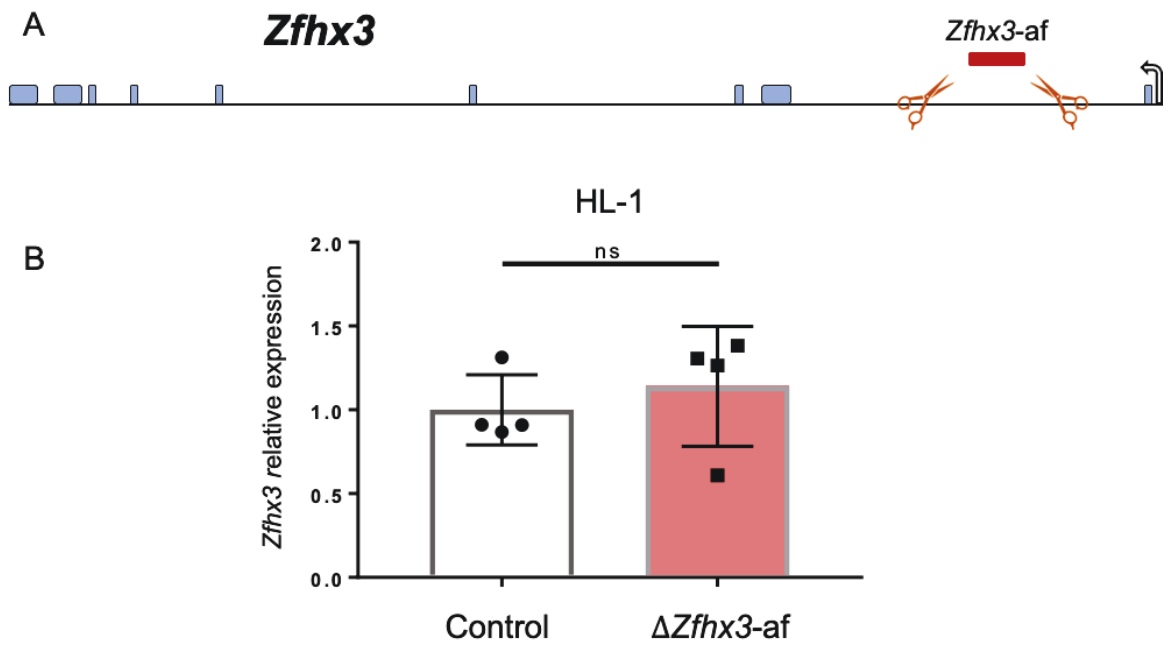

Figure 24 - The 16q22 silencer is not active in mouse cells. A) Schematic representation of the mouse *Zfhx3* gene locus containing the mouse ortholog *Zfhx3*-af (red bar). **B)** Deletion of the intronic element in mouse HL-1 cardiac cells did not result in *Zfhx3* gene expression changes, conversely to deletion of the human orthologous sequence.

To gain insight into the genetic principles governing the activity of *ZFHX3*-AF we used the PB-ERA to perform *in vivo* assays of enhancer blockade. Based on the ability of insulators to block enhancer activity when located between the enhancer and the promoter (Lunyak *et al.*, 2007), we tested whether *ZFHX3*-AF might be able to block enhancer activity. Since we wanted to confirm that the negative regulator was able to exert its function in a tissue of relevance for AF, we chose the cardiac ASE enhancer for the assay (see **Figure 13B**). Accordingly, enhancer activity was assessed in E9.5 embryos carrying one of the following three conditions: i) ASE enhancer alone (pPB-ASE- lacZ); ii) ZFHX3-AF between the globin promoter and ASE (pPB- lacZ-ZFHX3AF-ASE); or iii) ZFHX3-AF outside (pPB- lacZ-ASE-ZFHX3AF)(**Figure 25A**). This experiment helped us to better characterize the type of activity driven by the *ZFHX3*-AF negative regulator and showed very robustly that it is a silencer able to downregulate enhancer-driven gene expression. Rather than being an insulator-like element, *ZFHX3*-AF is able to downregulate gene expression independently from its relative position to the promoter and enhancer (**Figure 25B**;

Table 9). In addition, and importantly in the context of AF, the silencer identified here functioned in the developing heart.

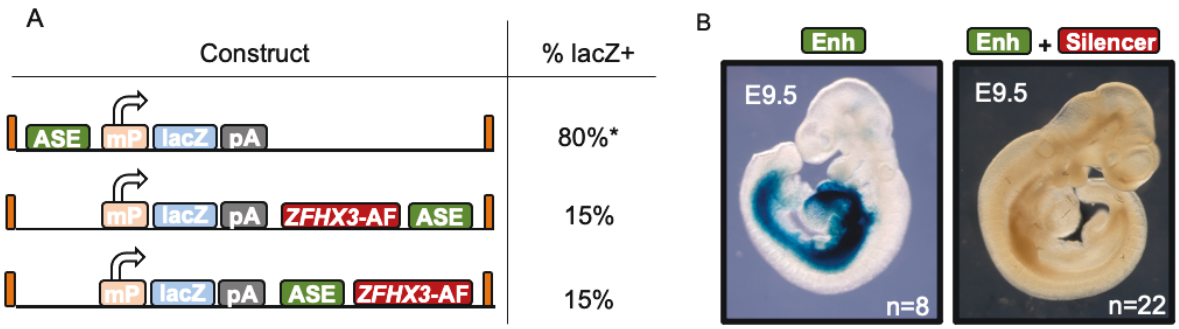

Figure 25 – The *ZFHX3*-AF is a silencer that outcompetes heart enhancers. A) Schematic representation of the chimeric and control constructs tested. Asterisk indicates the estimated proportion of lacZ positive embryos (between 50-100%) for the ASE element which we could not genotype. **B)** Representative embryos showing the ASE enhancer pattern (left) or complete silencing (right).

Construct	# embryos	# tg	% tg	# lacZ+	% lacZ+
ASE-mP-lacZ	16	NA	50-100	8	50-100
mP-lacZ-ZFHX3AF-ASE	27	13	48.15	2	15.38
mP-lacZ-ASE-ZFHX3AF	21	13	61.90	2	15.38

Table 9 – Summary of transgenesis for the enhancer-blocking assay.

In this second chapter, we have surveyed ten loci associated to the most common arrhythmia in human. Using the PB-ERA we have uncovered new enhancers and silencers with regulatory activity in heart tissue. Indeed, chimeric constructs showed that an AF-silencer is able to outcompete heart enhancers. Furthermore, enhancer perturbations identified target genes for several AF-enhancers involving new genes in AF and characterizing complex enhancer-promoter communication. Altogether, our study establishes a new framework for the efficient dissection of the genetic contribution to common human diseases and characterizes the function of new genetic elements that might be of relevance for the better understanding of gene regulation and AF.

3. Convergence between AF genetic predisposition and induced chronic arrhythmia.

Despite AF is considered polygenic (Lubitz *et al.*, 2017; Bapat *et al.*, 2018) the individual contribution of genes is moderate. Therefore, identifying AF-enhancers does not imply that their contribution to arrhythmia susceptibility is high. Whereas knock-out mice for some of the target genes that we identified in the previous chapter to be regulated by AF-CREs develop cardiac pathologies (Yang *et al.*, 2014), we have also shown that deletion of major enhancers cause partial loss of gene expression and thus the effect of single-nucleotide variants might be even milder (see **Figure 21B** in Chapter 2 of Results).

In this chapter, we have explored the convergence between genetic predisposition and the mechanisms governing atrial remodeling in order to identify regulatory elements with a higher phenotypic impact in disease progression.

3.1. Deciphering the susceptibility to atrial remodeling.

AF is a progressive disease that causes electrophysiological and structural changes to the atria. This progressive remodeling of the atrial substrate leads to long-term perpetuation of AF (Christophersen *et al.*, 2009). Several studies in animals showed that the artificial trigger of AF through a pacemaker led to increasingly longer episodes of sustained arrhythmia (Wijffels *et al.*, 1995; Filgueiras-Rama *et al.*, 2012). This observation of self-perpetuation is known as the concept ‘AF begets AF’ by which the molecular consequences of the disease are also causative in the sense that they reinforce arrhythmia, creating an aberrant feedback loop. From the genetic point of view, this model of atrial remodeling generating more remodeling might mean that, during fibrillation, changes in gene expression could potentially perpetuate themselves epigenetically.

Since genetic risk and electrical stimuli can drive AF, exploring the common elements shared between both mechanisms could be very insightful. GWAS and subsequent functional studies bring a list of candidate genes with a potential causative role in AF. Similarly, measuring transcriptomic and proteomic levels in patients as compared with healthy individuals can provide us with a valuable picture of atrial substrate remodeling

as a consequence of AF. Are there common deregulated genes? Are there core elements that change as a cause for and consequence of AF?

To address that, we took advantage of available transcriptomic data generated from a sheep model of induced AF and intersected it with GWAS genes. In this study, recently published by our lab, disease progression correlates with changes in the cardiomyocyte expression of genes encoding structural proteins of the myofibril, ion channels, cell-to-cell communication proteins, chromatin remodelers and developmental TFs (Alvarez-Franco *et al*., 2020). We searched the RNA-seq data coming from isolated cardiomyocytes for specific transcriptional changes between sheep with a normal sinus rhythm and those in chronic AF. In order to identify robust markers of atrial remodeling, we filtered the data coming from the right (R) and the left (L) atrial appendage (AA) keeping only the common differentially expressed genes, which restricted the list of candidates. Induced AF altered the expression of 209 genes shared between RAA and LAA cardiomyocytes (**Figure 26A**). Next, we collected the genomic coordinates for a representative GWAS-SNP at each of the 130 risk loci for AF and looked for putative target genes. All genes within a window of 200 kb from the variants were selected (**Figure 26C**). As expected, risk SNPs within a gene desert (11 loci) did not include any gene that matched the former criteria. In those cases, we looked for the nearest gene at each side and included them in the list of candidate genes if it was protein-coding (for this analysis, pseudogenes and non-coding RNAs were not taken into consideration). As a result, we obtained a set of 354 genes putatively involved in AF genetic predisposition, including at least one gene per risk locus. Interestingly, prioritized genes from GWAS and differentially expressed genes in sheep with chronic AF only shared four hits: *GJA1*, *TBX5*, *JMJD1C* and *FKBP7* (**Figure 26B; Table 10**).

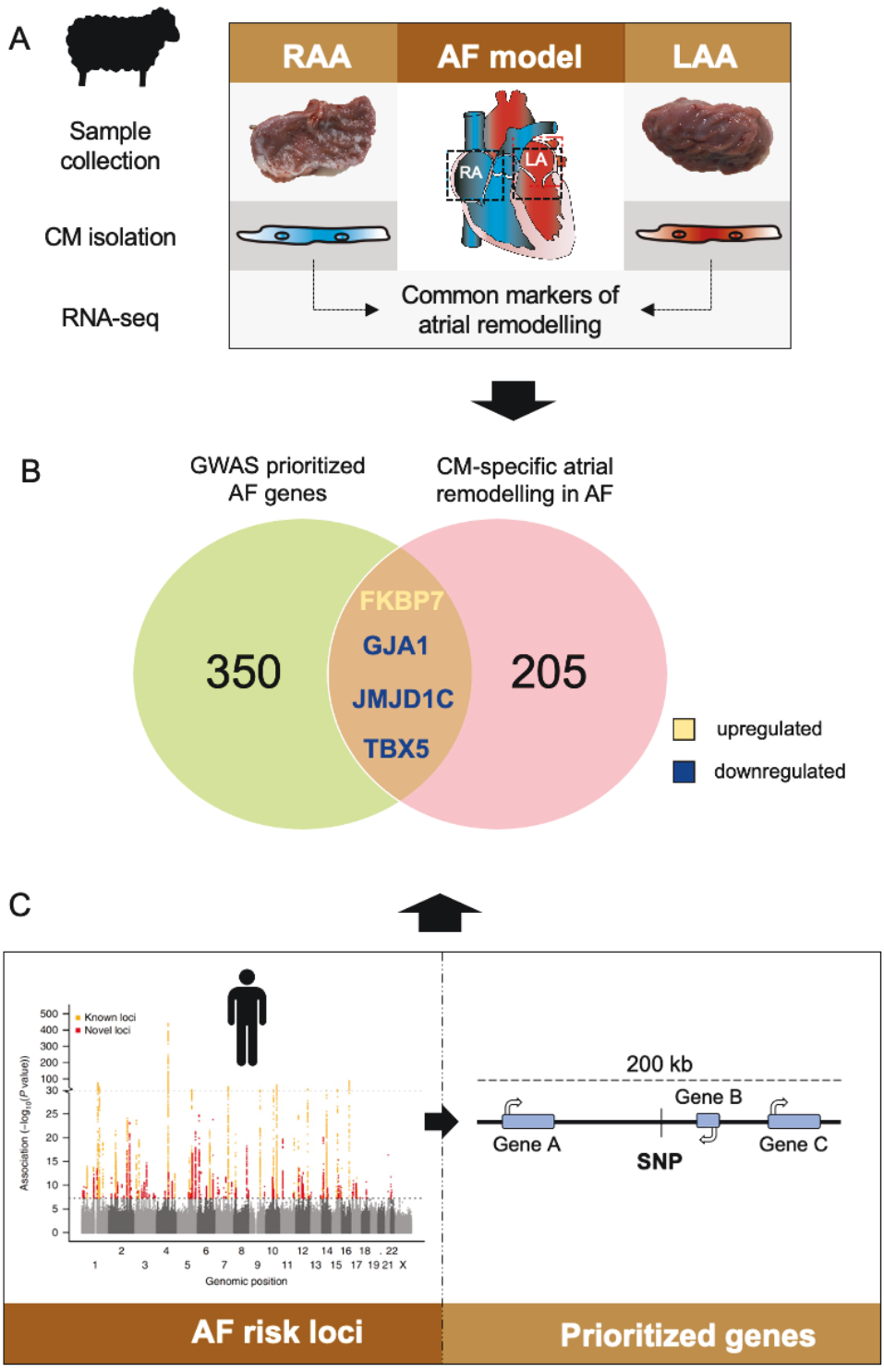

Figure 26 – Intersection between GWAS and induced AF. A) Representation of the experimental design for the transcriptomic data obtained from a sheep model of induced AF (Alvarez-Franco *et al.*, 2020). **B)** Venn diagram showing the four GWAS genes differentially expressed in both atria from the induced AF model. **C)** GWAS prioritized genes were obtained selecting all genes within a window of 200 kb from the GWAS variants.

	CM - RAA		CM - LAA		Mean expression
Gene	**logFC**	**adj. p-value**	**logFC**	**adj. p-value**	**logCPM**
GJA1	-1.0978	0.0025	-0.9129	0.0035	8.0105
TBX5	-0.7786	0.0277	-1.0084	0.0035	7.3429
JMJD1C	-0.9948	0.0370	-1.1617	0.0112	7.1610
FKBP7	0.8712	0.0462	1.0262	0.0097	4.6678

Table 10 – RNA-seq data (Alvarez-Franco *et al.*, 2020) **for the four genes identified at the intersection between susceptibility and AF-induced genes.**

3.2. SNPs for AF and electrophysiological traits accumulate at a *TBX5* conserved intronic enhancer in cardiomyocytes.

We hypothesized that genes found at the intersection between genetic susceptibility and AF induction would play an important role in disease. While there is not much information about *FKBP7*, it was very interesting to see downregulation of the histone demethylase *JMJD1C* and essential genes for cardiomyocyte function, such as *GJA1* and *TBX5*, in the diseased atria (**Table 10**). Indeed, deletions of either the *GJA1* gene involved in cell-to-cell communication, or the cardiac TF-encoding *TBX5* gene lead to severe heart phenotypes (Reaume *et al.*, 1995; Gutstein *et al.*, 2001; Nadadur *et al.*, 2016; Dai *et al.*, 2019).

TBX5 is a T-box-containing transcription factor TF that is essential for proper heart development (Bruneau *et al.*, 2001). Haploinsufficiency of *TBX5* causes Holt-Oram syndrome (Basson *et al.*, 1997), a rare autosomal dominant human disease characterized by upper limb malformations and congenital heart disease. Additionally, its deletion in adult mice triggers irregular heartbeat (Nadadur *et al.*, 2016). In this regard, *TBX5* expression is diminished in both atria after induced AF (**Table 10**), what might contribute to perpetuating AF. On the other hand, GWAS have associated the 12q24 locus to AF. Variants in this locus fall near the *TBX5* gene which is why it was included in our prioritized list of genes related to AF predisposition. The three variants associated to AF in this locus (rs883079-T, rs3825214-G, rs10507248-T) are located at the 3' part of the *TBX5* gene very close to each other. Interestingly, other variants

associated to electrocardiogram (ECG) traits like PR interval and QRS duration also localize in this genomic region (**Table 11**). The epigenetic landscape of the risk locus showed that, apart from being a hotspot for polymorphisms associated to cardiac conduction defects, the *TBX5*-AF region located at the last intron of *TBX5* is very likely containing a heart enhancer (H3K27ac and H3K4me1 peaks in left ventricle) that interacts with the promoter of *TBX5* in differentiated cardiomyocytes (**Figure 27A**). Indeed, we defined an 800-bp minimal region (min*TBX5*-AF) overlapping the Hi-C interaction with the promoter of *TBX5* and containing the variant rs3825214 (**Figure 27B**).

SNP ID	Risk Allele	Trait	Relative location	Coordinates hg19	Refs
rs883079	T	AF & ECG	TBX5 3'-UTR	chr12:114793240	Christophersen 2017
rs3825214	G	AF & ECG	TBX5 last intron	chr12:114795443	Zhang *et al*., 2016
rs10507248	T	AF	TBX5 last intron	chr12:114797093	Sinner *et al*., 2014
rs7312625	A	ECG	TBX5 last intron	chr12:114799974	Smith *et al*., 2011
rs7135659	G	ECG	TBX5 last intron	chr12:114801772	Hong *et al*., 2014
rs1895585	A	ECG	TBX5 last intron	chr12:114802138	Butler *et al*., 2012

Table 11 – List of SNPs associated to arrhythmia in the *TBX5* risk locus.

Enhancer marks seemed in fact to be conserved in the mouse for the risk locus, indicating that orthologous *Tbx5* gene expression might be controlled by the same heart enhancers than in human (**Figure 28A**). Interestingly, TBX5 itself together with GATA4 and NKX2-5 bind *Tbx5*-af, the mouse orthologous enhancer region, in mESC-CM (**Figure 28A**). In order to gain insight into the mechanisms underlying the non-coding genetic associations in the 12q24 locus to AF, we deleted either the 9.7 kb Tbx5-af (**Figure 28A**; highlighted in yellow) or the 800 bp minTbx5-af **(Figure 28A**; highlighted in blue) in the atrial-like HL-1 mouse cell line. Deletion of both intronic region led not only to downregulation of the *Tbx5* gene (**Figure 28B**), but also to upregulation of the >200 kb distal *Tbx3* gene (**Figure 28C**) although we do not see a clear interaction between the promoter of the human *TBX3* gene and the intronic regulatory elements in the Hi-C data from human cardiomyocytes (**Figure 27A**). Our data indicates that the last intron of the *TBX5* gene contains cis-regulatory activity over

two relevant TFs involved in cardiac development, such as TBX5 and TBX3, and regulates them in an opposite manner. The binding of this functional enhancer by TBX5 itself suggests a putative self-regulation that might contribute to arrhythmia susceptibility and perpetuation via a positive feedback loop.

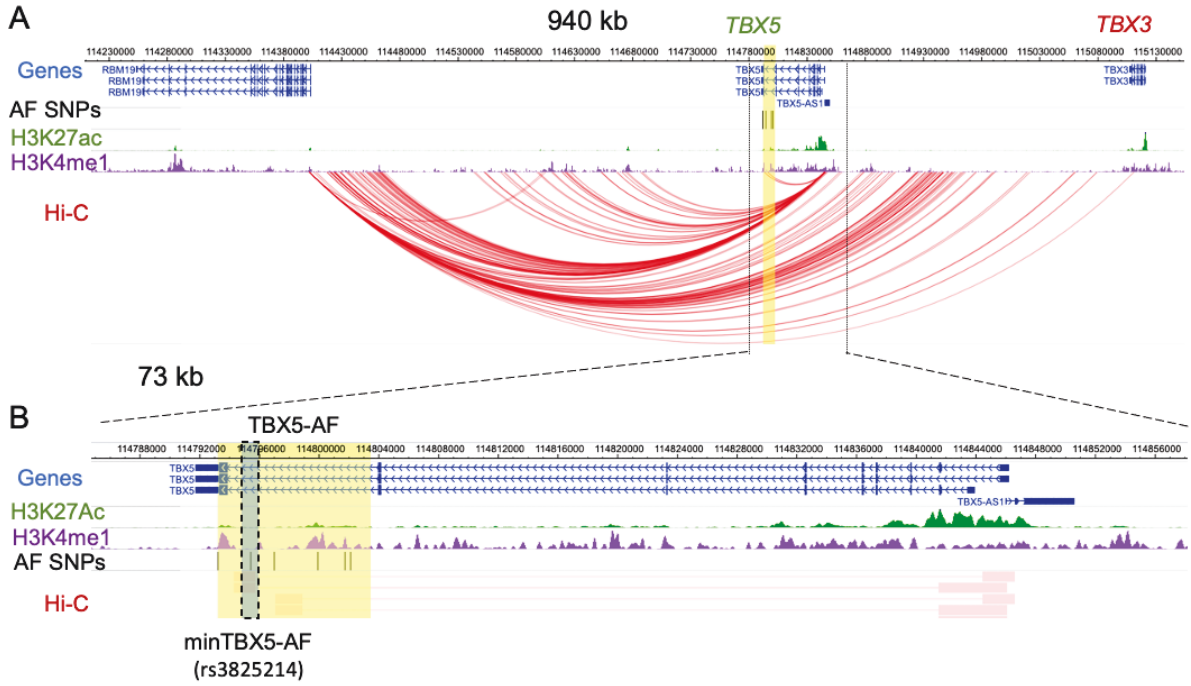

Figure 27 – Epigenetic and chromatin features of the 12q24 AF risk locus. A) Genomic view of 940 kb in the 12q24 locus (hg19; chr12:114,210,000-115,150,000) showing the candidate intronic enhancer (yellow) interacting with the promoter of *TBX5* in differentiated cardiomyocytes. Indicated in the left and from top to bottom are WashU tracks for annotated genes, AF SNPs (black vertical bars), H3K27ac (green) and H3K4me1 (purple) ChIP-seq signal from human left ventricle, promoter-capture Hi-C data from Montefiori *et al*., 2018 indicating three-dimensional chromatin contacts in differentiated cardiomyocytes from hiPSC and represented by red arcs. **B)** Magnification of the 73 kb region (hg19; chr12:114,785,000-115,858,000) spanning from the candidate enhancer until the promoter of *TBX5* shows marks of active enhancer (H3K27ac in green and H3K4me1 in purple) in the region. The chromatin data is shown here as the mapping reads of the crosslinked interaction (red horizontal bars connected by a line). The *TBX5*-AF region (highlighted in yellow) contains enhancer marks and two different regions interacting with the promoter of *TBX5*. The variant rs3825214 contained in the min*TBX5*-AF region (highlighted in blue) overlaps two interaction with the promoter of *TBX5*. WahsU; Washington University epigenome browser.

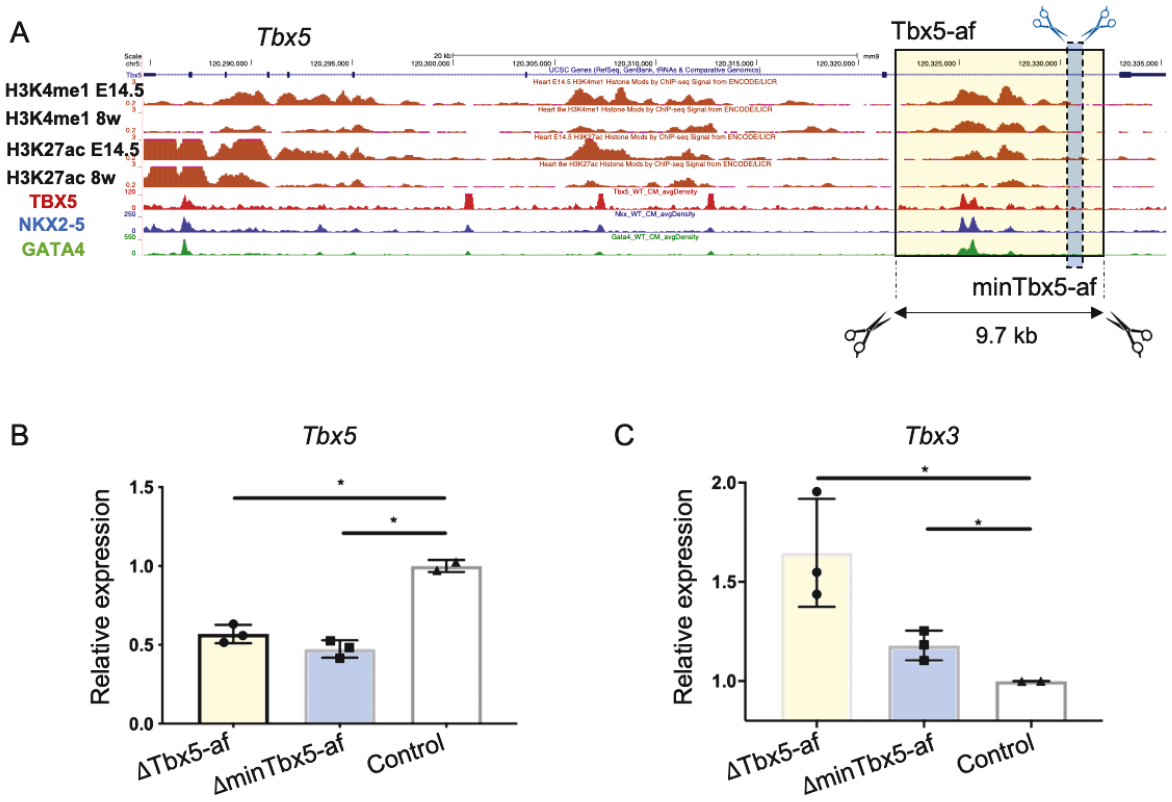

Figure 28 – A regulatory element at the mouse orthologous region to the 12q24 risk locus diferentially regulates *Tbx5* and *Tbx3*. A) Genomic view of the 50 kb mouse orthologous *Tbx5* region (mm9; chr5:120,285,000-120,335,000) including the following epigenetic information from top to bottom: genes (*Tbx5* exons and introns), ChIP-seq signal of H3K4me1 in mouse embryonic (E14.5) and adult (8 weeks) heart, H3K27ac in embryonic (E14.5) and adult (8 weeks) heart (ENCODE data), TBX5, NKX2-5 and GATA4 in differentiated cardiomyocytes from mESC (Luna-Zurita *et al*. 2016). The intronic *Tbx5*-af (highlighted in yellow) contains enhancer marks and cardiac TF binding. The ~800 bp min*Tbx5*-af (highlighted in blue) is the orthologous region to the human min*TBX5*-AF containing the variant rs3825214. Deletion of either the *Tbx5*-af (9.7 kb) region or the min*Tbx5*-af in HL-1 cells using CRISPR, resulted in the downregulation of *Tbx5* (**B**) and the upregulation of *Tbx3* (**C**).

3.3. An AF susceptibility locus is a distal enhancer that controls the cardiac expression of the *GJA1* gene in mammals.

GJA1 was one of the only four shared signatures between AF predisposition and perpetuation datasets (**Figure 26B**). Indeed, this gene was the most significantly downregulated one in cardiomyocytes from sheep in chronic AF (**Table 10**). *GJA1* encodes Connexin43 (Cx43), the major component of gap junctions that connect the cytoplasm of two adjacent cells. This structure is present in the cardiac intercalated disc and provides a low-resistance pathway for cell-to-cell passage of electrical charge (Leo-Macias, Agullo-Pascual and Delmar, 2016).

Therefore, proper *GJA1* expression is essential for correct excitability and cardiac conduction velocity (Beauchamp *et al.*, 2012; Desplantez *et al.*, 2012). In line with this, the volume fraction of Cx43 is diminished in patients with AF (Luo, Li and Yang, 2007). Besides its relevance in the heart, *GJA1* is expressed in many tissues and mutations of Cx43 result in oculodentodigital dysplasia (ODDD), a pleiotropic, autosomal dominant disorder that in humans affects primarily the eye, dentition, digits of the hands and feet, and also the heart (Paznekas *et al.*, 2009). Indeed, mutant mice for *GJA1* with ODDD have increased susceptibility to arrhythmias (Kalcheva *et al.*, 2007; Tuomi, Tyml and Jones, 2011).

As for the *TBX5* locus, GWAS have associated the 6q22 locus, in which *GJA1* is located, to AF. However, instead of intronic associations, variants at this locus are intergenic and fall within a 1 Mb gene desert (**Table 12**). Indeed, no gene was prioritized at the 6q22 locus in the first round of candidate gene selection (within a 200 kb window surrounding GWAS associations). As we did for all risk loci with no candidate gene selected, we included the first gene upstream (*GJA1*) and downstream (*HSF2*) the tag SNP (rs12664873). *HSF2* is the closest gene to the risk locus, located ~250 kb from the association. However, interaction data from differentiated cardiomyocytes suggested that *HSF2* and the tag SNP were in different TADs and not interacting with each other (**Figure 29A**). Instead, a long-range interaction is detected between the promoter of the *GJA1* gene and a genomic region (*GJA1*-HiC) ~5 kb from rs12664873. This Hi-C interaction is specific for differentiated cardiomyocytes, being absent in pluripotent stem cells (**Figure 29A**). We explored epigenetic marks in the gene desert with special focus on the proximities

SNP ID	Risk Allele	Trait	Relative location	Coordinates hg19	Refs
rs9401451	G	AF	Intergenic	chr6:122099152	Nielsen *et al*., 2018
rs868155	C	AF	Intergenic	chr6:122389906	Roselli *et al*., 2018
rs13191450	A	AF	Intergenic	chr6:122392136	Roselli *et al*., 2018
rs13195459	G	AF	Intergenic	chr6:122403559	Nielsen *et al*., 2018
rs13219206	C	AF	Intergenic	chr6:122414157	Low *et al*., 2017
rs12664873	**T**	**AF**	**Intergenic**	**chr6:122463191**	**Christophersen, 2017**

Table 12 - List of SNPs associated to AF arrhythmia in the 6q22 risk locus.

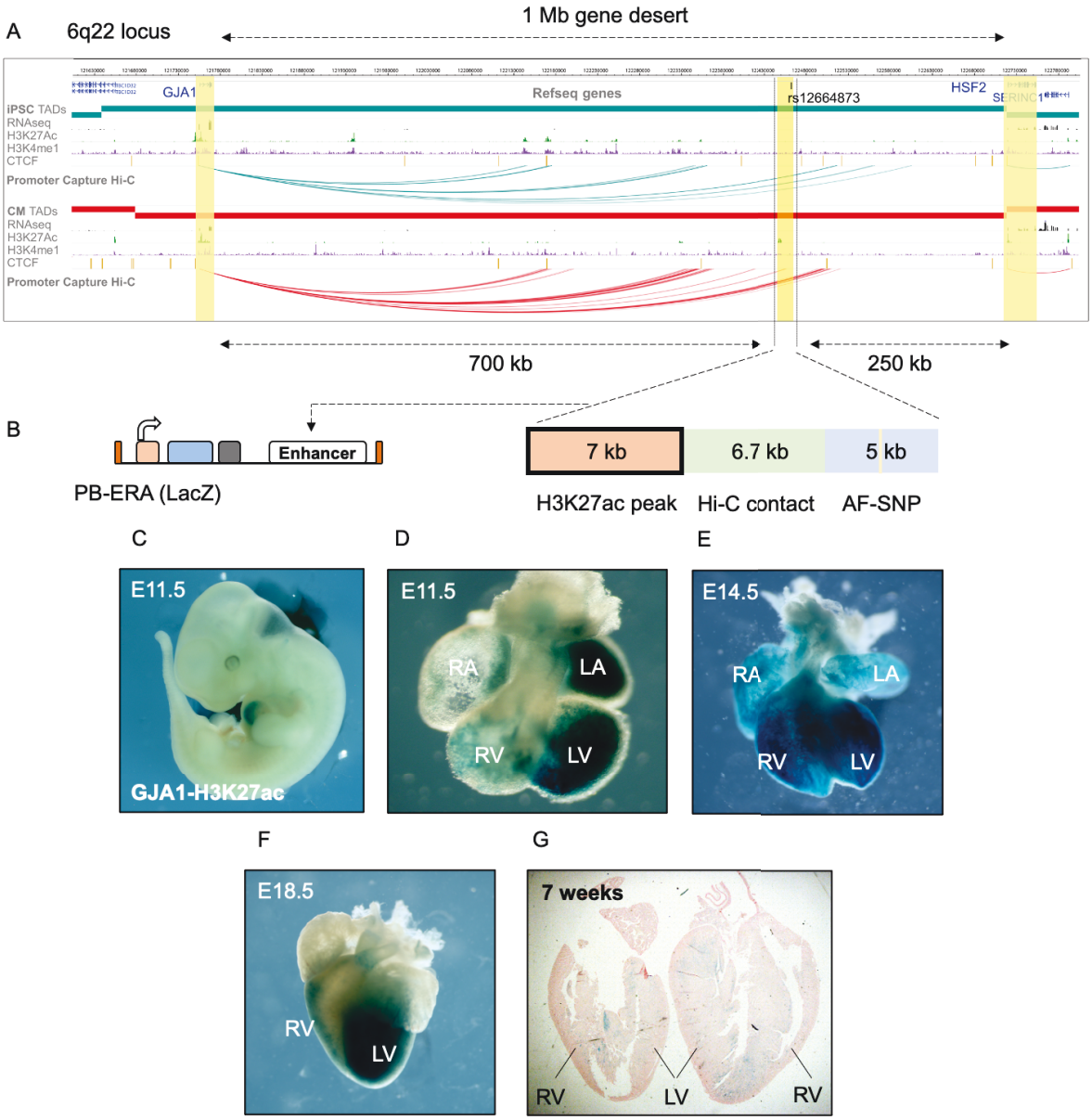

Figure 29 – Identifying a cardiac-specific element in a gene desert of the 6q22 AF risk locus. A) Genomic view of the 6q22 locus (hg19; chr6:121,605,000-122,805,000) including (top to bottom) Refseq genes, TADs, left ventricle H3K27ac and H3K4me1, CTCF binding and interaction data (promoter-capture Hi-C from Montefiori *et al.*, 2018) in hiPSC (top) and differentiated cardiomyocytes (bottom). The AF variant rs12664873 is located within a candidate region that interacts with the *GJA1* gene, while not interacting with its closest gene, *HSF2*. **B)** Schematic representation of the approach in which the 18.7 kb region of interest is divided in three sub-domains with different annotations (peak of H3K27ac, Hi-C contact or AF SNP) that were tested for enhancer activity using the PB-ERA system. Cardiac-specific activity was detected in the fragment containing a H3K27ac peak (and called *GJA1*-H3K27ac) in the embryonic (C-F) and adult heart (G). TAD, topologically associated domains; RA, right atria; LA, left atria; RV, right ventricle; LV, left ventricle.

of the rs12664873 variant and found a genomic region (*GJA1*-H3K27ac) enriched for H3K27ac near the SNP and adjacent to the promoter-interacting region. All in all, this analysis highlighted an 18.7 kb intergenic fragment as the best candidate regulatory element in this 1 Mb gene desert (**Figure 29B**).

In order to identify potential CREs regulating *GJA1* gene expression, we used the PB-ERA system (see Chapter 1 of Results above) in order to assess the regulatory potential of this large prioritized region. We subdivided the 18.7 kb region into three fragments (**Figure 29B**) according to the previous annotations and generated transgenic embryos that we examined at E11.5 or E14.5 (**Table 13**). While *GJA1*-SNP and *GJA1*-HiC fragments did not show regulatory activity, we identified a specific heart enhancer in *GJA1*-H3K27ac, the genomic fragment enriched for H3K27ac in human left ventricle (**Figure 29C-E).** Next, we tested the enhancer activity of this regulatory element in the offspring of an adult transgenic male carrying the enhancer-reporter construct. We found that the *GJA1*-H3K27ac enhancer regulates gene expression throughout lifespan, been also active in the adult heart (**Figure 29G**). This would mean that the incorrect or diminished functioning of the *GJA1*-H3K27ac enhancer might result in a sustained alteration of *GJA1* expression levels. Interestingly, the *GJA1*-H3K27ac enhancer drove reporter expression predominantly in the left ventricle, although variable expression can be detected in the right ventricle and in the atria (**Figure 29D-G**). These results suggest a role for the *GJA1*-H3K27ac enhancer in the atria that might indeed be involved in AF. Additionally, due to its prominent left ventricle activity, its role in other cardiomyopathies should not be unnoticed.

Candidate enhancer	Species	coordinates hg19/mm9	size	# tg	# lacZ+ (heart)
***GJA1*-H3K27Ac**	hs	chr6:122447000-122454000	7 kb	17	7
***GJA1*-HiC**	hs	chr6:122454000-122460686	6.7 kb	12	1
***GJA1*-SNP (rs12664873)**	hs	chr6:122460686-122465695	5 kb	10	1
***Gja1*-h3k27ac**	mm	chr10:56830666-56837165	6.5 kb	13	12
min*GJA1*-H3K27ac	hs	chr6: 122451039-122451610	572 bp	13	0

Table 13 – List of tested human and mouse candidate enhancers in the 6q22 AF risk locus.

The 18.7 kb intergenic region at the 6q22 locus is composed of three modules that contain a heart enhancer, the AF variant rs12664873 and a region interacting with the promoter of *GJA1* in differentiated cardiomyocytes. This 3D chromatin interaction suggests that *GJA1* might be the candidate target gene to be regulated by the *GJA1*-H3K27ac enhancer. Sequence similarity indicated conservation of the *GJA1*-H3K27ac regulatory region in the mouse genome (**Figure 30A and B**) and assaying the activity of the mouse *Gja1*-h3k27ac enhancer with the PB-ERA system showed the same pattern of reporter expression in the heart with preferential activity in the left-ventricle (**Figure 30C**). In order to gain evidence of direct gene regulation, we decided to delete the *Gja1*-h3k27ac cardiac enhancer from the mouse HL-1 cardiac cells. Deletion of the large 18.7 kb region as well as the *Gja1*-h3k27ac enhancer led to downregulation of the distal *Gja1* expression levels while not affecting *Hsf2*, the nearest neighboring gene to the assayed regions (**Figure 30D).** Interestingly, deletion of the region encompassing both the *Gja1*-hic and *Gja*-snp mouse orthologous modules also led to similar downregulation of *Gja1* while not affecting *Hsf2* (**Figure 30D**). Our data suggest that although the enhancer activity of this risk locus resides primarily in *GJA1*-H3K27ac, the whole 18.7 kb region might carry regulatory potential, harboring several elements needed for the proper expression of *GJA1*.

The enhancer block at the 6q22 AF locus controls the expression of *GJA1* specifically in the heart and, while enhancer activity is encoded at the *GJA1*-H3K27ac element, there seem to be additional elements essential for correct enhancer-mediated gene regulation. In an attempt to shed light on what might be controlling the cardiac specificity of the *GJA1*-H3K27ac enhancer we searched for cardiac TF binding in this region by examining public datasets. Available TBX5 and GATA4 ChIP-seq data on differentiated cardiomyocytes from hiPSC (Ang *et al.*, 2016) showed binding of both cardiac TFs at the same ~600 bp region of the *GJA1*-H3K27ac enhancer that we called min*GJA1*-H3K27ac (**Figure 31A**). Cardiac TF binding site (TFBS) search in this region predicted ten independent binding sites for TBX5, four for GATA4 and eight for NKX2-5. We used the PB-ERA to assess enhancer activity, but we did not detect cardiac lacZ staining in transgenic animals carrying the minimal enhancer (**Table 13**). In order to test whether this minimal enhancer might impact gene expression, we deleted the mouse orthologous min*Gja1*-h3k27ac, which is also bound by cardiac TFs (**Figure 31B**; data from (Luna-Zurita *et al.*, 2016)), in HL-1 cells causing

a significant downregulation of *Gja1* expression to a 25% (**Figure 31C**). Altogether our results suggest that the 6q22 risk locus contains a conserved heart-specific distal enhancer of *GJA1* which is controlled by TBX5, among other cardiac TFs, throughout lifespan. This element is located within a larger region with regulatory activity, whose integrity is important to maintain proper *GJA1* regulation. Therefore, mutations at this enhancer block might be of relevance for AF and other cardiac arrhythmias.

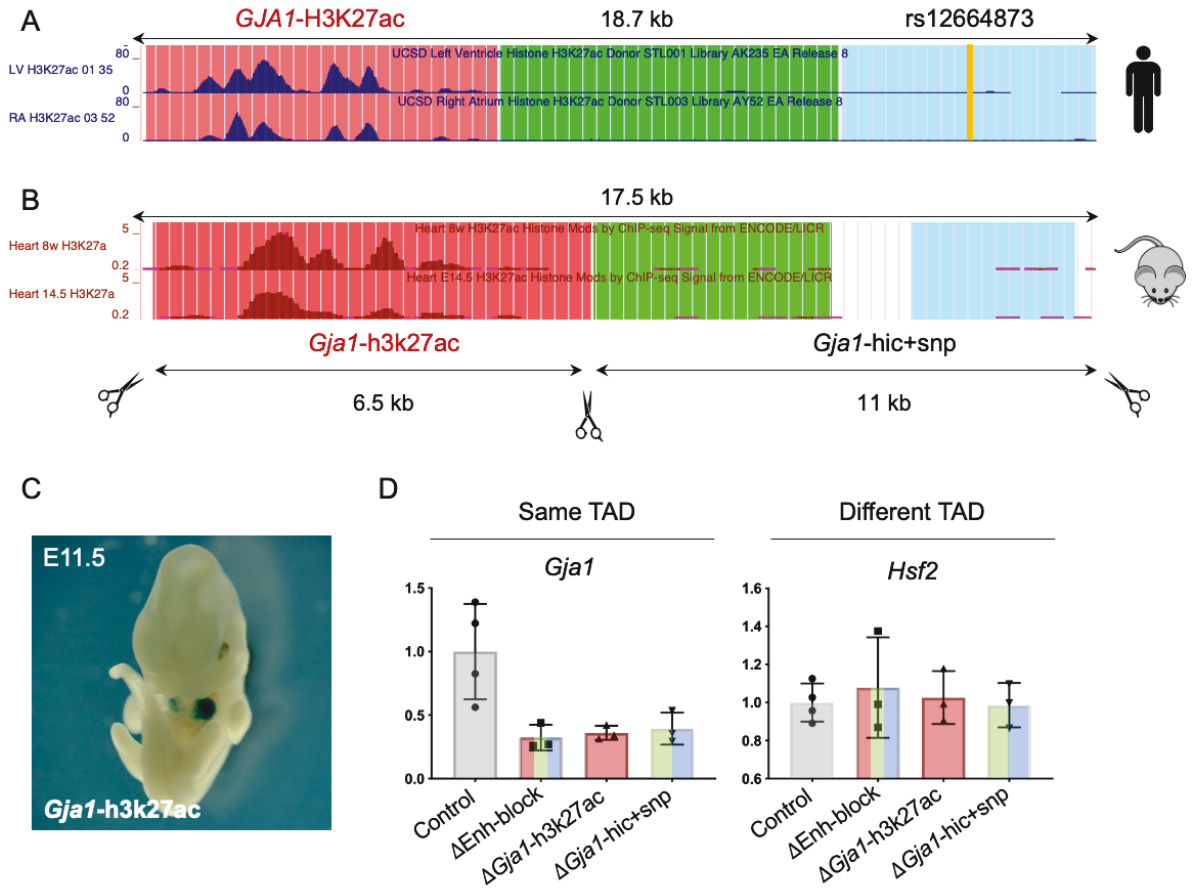

Figure 30 – A regulatory block at the 6q22 locus regulates GJA1 specifically. A) Genomic view of the human enhancer block at the 6q22 locus (hg19; chr6:122,447,000-122,465,695) showing ChIP-seq signal for H3K27ac in human LV and RA. **B)** Genomic view of the mouse orthologous regions for the enhancer block, as calculated with LiftOver (mm9; chr10: 56,830,666-56,848,166). **C**) The mouse *Gja1*-h3k27ac orthologous region is a heart enhancer. **D**) Deletion of the different modules of the regulatory block in HL-1 cells resulted in the specific downregulation of the *Gja1* gene while not affecting its closest neighboring gene Hsf2. LV, left ventricle; RA, right atria.

In this third chapter, we have explored the regulatory potential of two AF risk loci and identified cardiac enhancers controlling the expression of *TBX5* and *GJA1*. It is important to highlight that the 12q24 and 6q22 risk loci were prioritized because they

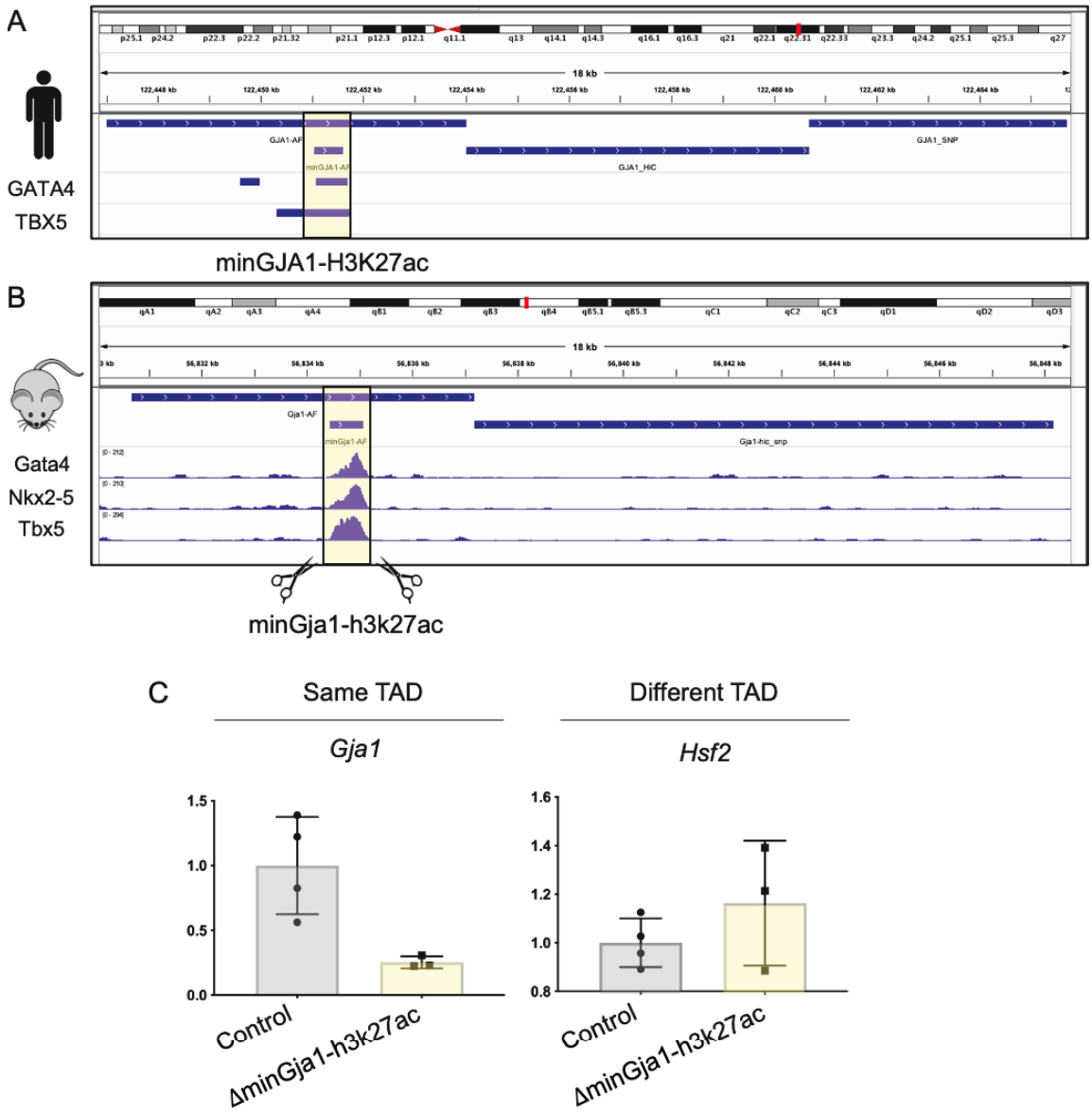

Figure 31 – Identification of a 600-bp minimal enhancer region. Genomic view of the human regulatory block at the 6q22 risk locus (**A**) and its mouse ortholog (**B**) showing ChIP-seq data for cardiac transcription factors of differentiated cardiomyocytes from hiPSC or mESC (data from Ang *et al.*, 2016 and data from Luna-Zurita *et al.*, 2016). A minimal region (min*GJA1*-H3K27ac) with conserved binding is highlighted in yellow. Deletion of the orthologous minimal enhancer (min*Gja1*-h3k27ac) in HL-1 cells was sufficient to downregulate the expression of its target *Gja1* gene, while not affecting the nearest neighboring gene, *Hsf2* (**C**).

contained genes which were downregulated in a chronic model of AF. The downregulation of *TBX5* and *GJA1* in the fibrillating atria, together with evidence supporting causal roles for both genes in arrhythmia beyond GWAS associations (Kalcheva *et al.*, 2007; Tuomi, Tyml and Jones, 2011; Nadadur *et al.*, 2016), suggest an important role in AF for the two regulatory elements that we have identified to regulate *TBX5* and *GJA1*.

4. Dissecting the regulatory landscape of the pro-atherosclerotic *PSCK9* gene: from relevant cis-regulatory elements to disease.

PCSK9 has been linked to atherosclerosis both by coding and non-coding mutations in familial hypercholesterolemia (Abifadel *et al.*, 2003) and GWAS (Nelson *et al.*, 2017; Van Der Harst and Verweij, 2018), respectively. PCSK9 protein is produced in the liver and secreted to the bloodstream where it controls the metabolism of LDL-cholesterol (Seidah and Prat, 2007). High levels of circulating PCSK9 lead to increased LDL-cholesterol, atherosclerotic lesions and high risk of infarct and stroke. Conversely, low levels of PCSK9 reduce LDL-cholesterol and atherosclerosis (Cohen *et al.*, 2006). In less than fifteen years, PCSK9 has come all the way from its initial description in relation to lipid metabolism to clinical trials, where scientists and clinicians study the way of diminishing its pro-atherosclerotic effect by targeting PCSK9 at the protein and mRNA level (Shapiro, Tavori and Fazio, 2018).

Therefore, understanding the expression profile of PCSK9 as well as the cell type-specific regulatory elements accounting for it will provide us with useful information to dissect the time and tissues in which the expression of this LDLR turnover regulator is key in atherosclerosis. In this fourth chapter, we performed a candidate enhancer search, characterized prioritized candidates and identified key regulatory elements controlling *PCSK9* gene expression not only in the liver but also in the cerebellum, where the role of *PCSK9* has been poorly studied.

4.1. Epigenomic mapping of the *PCSK9* locus identifies candidate tissue-specific enhancers.

We first explored the organs in which *PCSK9* is actively transcribed. Gene expression profiles from GTEx showed high levels of *PCSK9* transcripts not only in the liver (25 transcripts per million - TPM), which is the main organ known for proprotein convertase expression and activity, but also in cerebellum (22-25 TPM), medium expression in lung (~8 TMP) and low levels (>1.5 TPM) in colon, esophagus, pancreas and small intestine (**Figure 32A**). Due to the role of PCSK9 in atherosclerosis via LDLR turnover regulation (Horton, Cohen and Hobbs, 2007; Lagace, 2014), and the increasing number of claimed interactions between cardiovascular and neurodegenerative disease (Casserly and Topol, 2004; Dardiotis *et al.*, 2012), it is very interesting to find

PCSK9 gene expression in the brain. A more detailed examination of transcription of the *PCSK9* gene shows a major isoform present in most tissues (ENST00000302118.5), with the exception of the cerebellum, where a specific isoform (ENST00000490692.1) is detected (**Figure 32B**). In fact, almost 50% of *PCSK9* transcripts in the cerebellum come from this isoform. This is in stark contrast to the liver, in which there is only expression of the major isoform (**Figure 32C and D**). Whether this cerebellum isoform is functional and may have a role in atherosclerosis or even neurological disorders is yet to be determined.

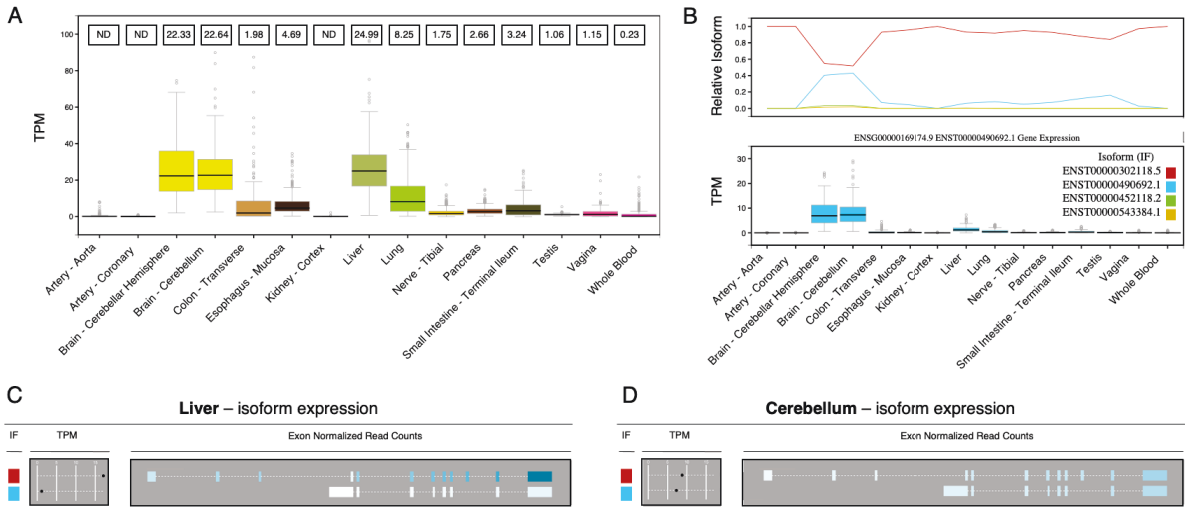

Figure 32 – *PCSK9* gene expression throughout the body in humans. A) *PCSK9* expression values from GTEx data shows expression in several tissues including the liver and the cerebellum. Values are represented in TPM using box plots and median values are shown on the upper part above each tissue. ND, not detected. **B)** *PCSK9* isoform-specific visualizations showing relative isoform expression (upper panel) and the cerebellum-specific isoform levels by tissue (bottom panel). Colour legend for different transcripts are shown on the right. **C**, **D**) Diagrams showing isoform expression levels in liver (**C**) and cerebellum (**D**). Colour boxes in the IF (isoform) column corresponds to the transcripts in **B**. All data was obtained from GTEx (GTEx Consortium, 2020).

In order to select candidate regulatory elements for *PCSK9*, we used ENCODE ChIP-seq data for H3K27ac and H3K4me1 histone marks in *PCSK9* expressing tissues (liver, cerebellum, kidney and lungs (**Figure 33).** Since *PCSK9* has been shown to be expressed in macrophages, playing a role in pro-inflammatory response at the atherosclerotic lesion, we also used epigenomic data from bone marrow-derived macrophages (BMDM)(Ricci *et al.*, 2018). Genomic conservation of regulatory mechanisms and elements across mammals (Villar *et al.*, 2015) can overcome technological limitations by inferring results from experiments in other species. Therefore, we hypothesized that *PCSK9* expression should be conserved between

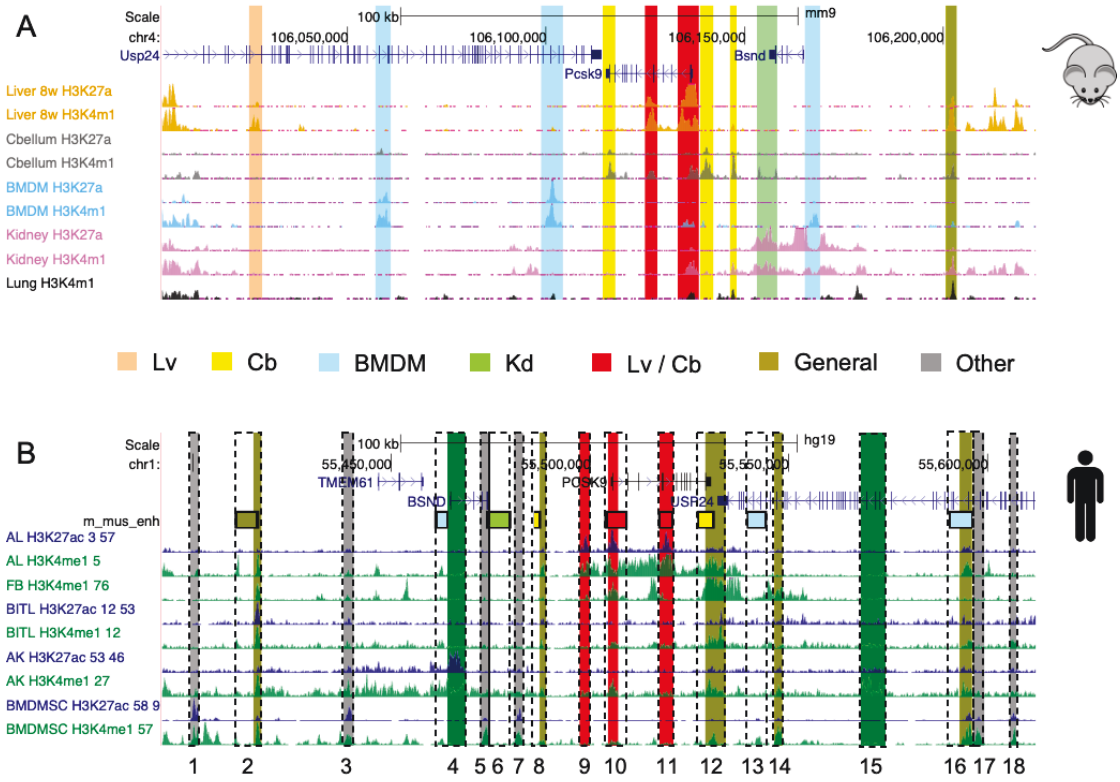

Figure 33 – Genomic features of the PCSK9 locus. A) Mouse *Pcsk9* locus (mm9; chr1:55,392,216-55,612,215) view of candidate regulatory elements used to select conserved regulatory elements. Indicated on the left and from top to bottom H3K27ac and H3K4me1 ChIP-seq signal from ENCODE and Roadmap Epigenomics are shown for liver, cerebellum, bone marrow derived macrophages (BMDM), kidney and lung. **B)** Human *PCSK9* locus (hg19; chr1:55,392,216-55,612,215) view of candidate regulatory elements selected for functional assays (dashed line rectangles; number of candidate is indicated below each rectangle). On the left, and from top to bottom, are: genes present in the region examined; human orthologous regions to mouse candidate enhancers to evaluate conservation at the genetic and epigenetic level (m_mus_enh; calculated with LiftOver tool); H3K27ac and H3K4me1 ChIP-seq signal from ENCODE and Roadmap Epigenomics in adult liver (AL), fetal brain (FB), brain inferior temporal lobe (BITL), adult kidney and bone marrow mesenchymal stem cells (BMDMSC). Figure legend showing a color for each different category of candidate regulatory elements based on tissue specificity. For detailed information on definitive candidate regions see **Table 14**.

human and mouse, and that its CREs should behave similarly. We took advantage of the detailed catalog of mouse epigenomic information (**Figure 33A**) and used it to compare enhancer features in the orthologous human regions (**Figure 33B**, coloured box track). This information was especially valuable when there was no epigenetic information for the human tissue (e.g. cerebellum and BMDM). For instance, although we are dissecting the human *PCSK9* locus, there are no ChIP-seq data on histone marks in cerebellum as opposed to the ENCODE mouse data for H3K27ac and H3K4me1. Data from other brain parts such as the brain inferior temporal lobe (BITL) or the fetal brain may not be good indicators of the epigenetic state in the cerebellum.

On one hand, mouse epigenomic data from cortex, olfactory bulb and whole brain tissues differ significantly from cerebellum data in the *Pcsk9* locus (not shown). On the other hand and conversely to what happens with the liver data, enhancer features from human BITL and fetal brain tissues hardly overlap with the mouse orthologous candidate regions harboring enhancer marks.

Epigenomic information shows that there are both ubiquitous and tissue-specific marks of active enhancers in both species (colored highlighted regions, **Figure 33A and B**). Eighteen candidate enhancers (CE) were selected combining mouse and human data (dashed line rectangles in **Figure 33B**) and annotated for potential tissue/s of activity (**Table 14**).

ID	Category	Size (bp)	Start Coordinate	End Coordinate
CE_01	Other	2,199	chr1:55,399,487	chr1:55,401,685
CE_02	General	6,842	chr1:55,410,532	chr1: 55,417,373
CE_03	Other	2,545	chr1:55,437,946	chr1:55,440,490
CE_04	BMDM/Kidney	8,072	chr1:55,460,752	chr1:55,468,823
CE_05	Other	2,545	chr1:55,472,363	chr1:55,474,907
CE_06	Kidney	6,355	chr1:55,473,781	chr1:55,480,135
CE_07	Other	2,246	chr1:55,481,042	chr1:55,483,287
CE_08	**General/Cerebellum**	**3,184**	**chr1:55,485,790**	**chr1:55,488,973**
CE_09	**Liver/Cerebellum**	**2,569**	**chr1:55,497,438**	**chr1:55,500,006**
CE_10	Liver/Cerebellum	5,821	chr1:55,503,648	chr1:55,509,468
CE_11	**Liver/Cerebellum**	**4,016**	**chr1:55,517,085**	**chr1:55,521,100**
CE_12	**General/Cerebellum**	**7,248**	**chr1:55,527,059**	**chr1:55,534,306**
CE_13	BMDM	5,041	chr1:55,539,515	chr1:55,544,555
CE_14	General	2,096	chr1:55,546,435	chr1:55,548,530
CE_15	Kidney	6,237	chr1:55,568,240	chr1:55,574,476
CE_16	General/BMDM	5,825	chr1:55,590,352	chr1:55,596,176
CE_17	Other	2,703	chr1:55,596,513	chr1:55,599,215
CE_18	Other	1,647	chr1:55,605,693	chr1:55,607,339

Table 14 – List of candidate regulatory elements of *PCSK9*. Tested fragments are shown in bold. hg19 assembly was used to calculate coordinates.

4.2. Assessing candidate CREs *in vivo* reveals a dual regulation of *PCSK9* gene expression.

Since we ignore human *PCSK9* gene expression patterns during development, we searched for its mouse orthologue. Mouse *Pcsk9* has been reported to be expressed in fetal liver and cerebellum as early as E14.5 (**Figure 34** left panel), later being also expressed in small intestine and kidney at E17 (**Figure 34** right panel), which are the same tissues in which *PCSK9* is expressed in humans. Since the liver and the cerebellum were the most prominent tissues with *PCSK9* expression in the data coming from human adult tissues and mouse embryos, we decided to focus our analysis in candidate enhancers with such annotations (CE8-12; **Table 14**). We explored GWAS-SNPs near *PCSK9* or eQTLs associated to its expression. We gathered SNPs associated to an atherosclerotic-trait such as coronary artery disease, myocardial infarction or LDL-cholesterol levels (disease SNPs) and imputed all SNPs in LD (r^2=0.8) with the disease SNPs (LD SNPs) to find whether they overlapped with our short-listed candidate enhancers. We found SNPs that fell within CE9, CE10 and CE12. Since eQTLs inform us of polymorphisms linked to gene expression in a tissue-dependent manner, we collected 316 different variants from the GTEx database affecting *PCSK9* gene expression and mapped them to the *PCSK9* locus. Cerebellum-specific eQTLs were enriched (32 out of 316) and most of them fell within CE8, CE9 and CE10. Remarkably, the seven liver eQTLs did not overlap with cerebellum eQTLs. Instead, they fell in a particular intronic region partly overlapping CE11 (**Figure 34B**). This analysis indicated that sequence variation at the candidates CE8, CE9, CE10, CE11 and CE12 affected *PCSK9* expression and might be associated to atherosclerosis.

We interrogated candidate enhancers using the PB-ERA system and collecting mouse embryos at E14.5 (**Table 15**), which is the earliest developmental timepoint at which we have evidence for mouse *Pcsk9* gene expression in the liver and cerebellum. CE10 was excluded from the enhancer assay due to proximity to *PCSK9* promoter and TSS. Assaying the enhancer activity of CE8, CE9, CE11, and CE12, allowed us to identify *PCSK9* major CREs regulating cerebellum (CE9; **Figure 34C**) and liver (CE11; **Figure 34D**) expression.

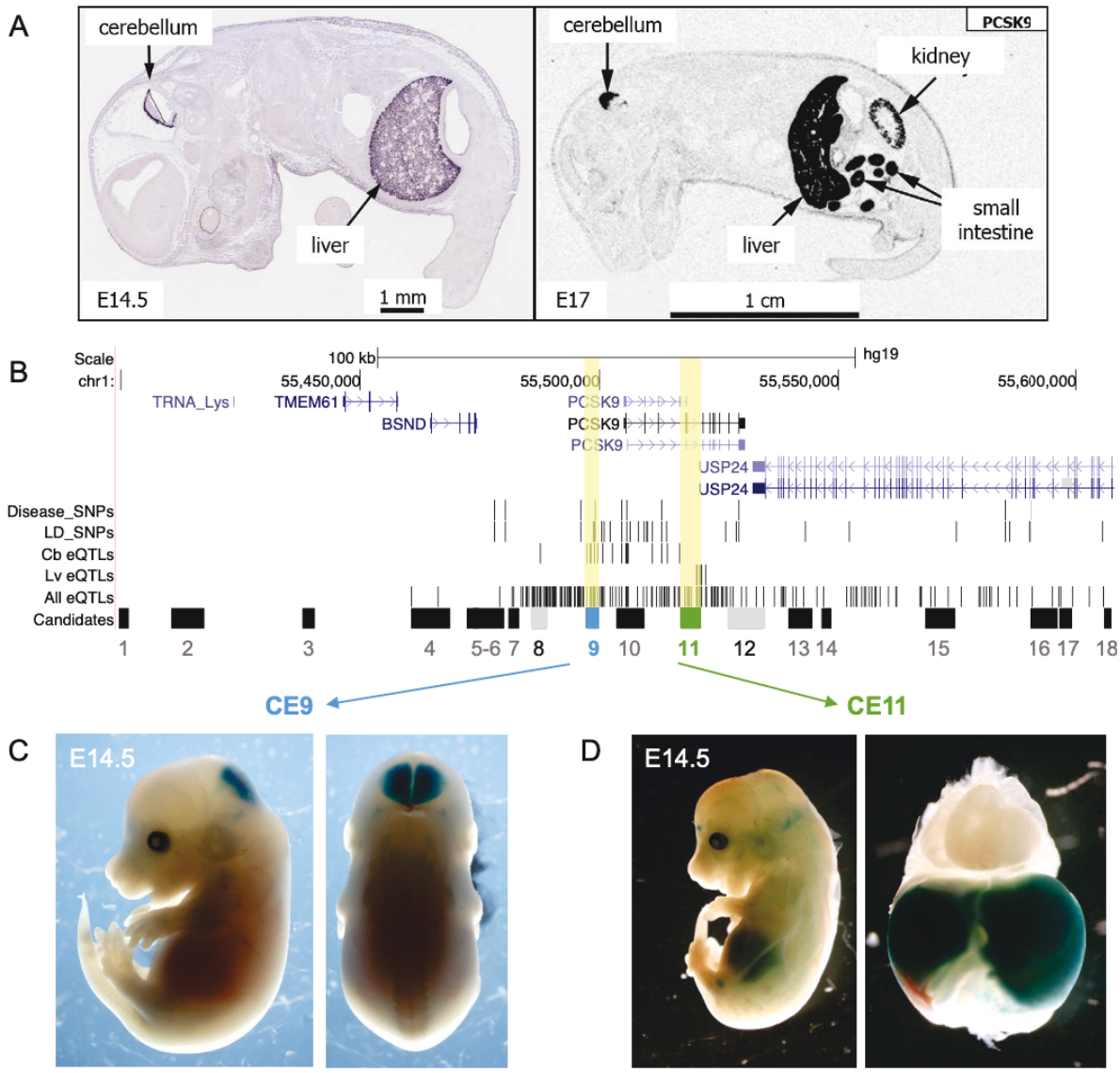

Figure 34 – Identification of regulatory elements driving *PCSK9* expression. A) Published in situ hybridization of mouse Pcsk9 showing liver and cerebellum expression (left from Diez-Roux G. *et al*., 2020 and right from Seidah *et al*., 2003). **B)** Genomic view of the *PCSK9* locus (hg19; chr1:55,399,000-55,608,000) showing annotated genes and from top to bottom (as indicated in the left) are disease SNPs associated to atherosclerotic traits, SNPs in linkage disequilibrium (LD SNPs) with disease SNPs, *PCSK9* eQTLs in cerebellum (Cb), liver (Lv) or any tissue (all). Non-prioritized candidates are represented as black boxes, tested elements with no detected activity by grey boxes, cerebellum enhancer by a blue box and liver enhancer by a green box. Enhancer assay of the positive elements captured cerebellar (**C**) or liver (**D**) activity.

candidate enhancer	stage	# tg	# lacZ+	liver	cerebellum
CE8	E14.5	2	0	0	0
CE9	E14.5	1	1	0	1
CE11	E14.5	14	3	2	0
CE12	E14.5	6	0	0	0

Table 15 – List of tested human candidate enhancers of PCSK9.

4.3. CE11 regulates the liver isoform of *PCSK9*.

Mutations in non-coding regulatory regions controlling liver expression of *PCSK9* might affect enhancer activity, alter *PCSK9* expression levels and, ultimately, influence atheroma plaque formation. In order to prove the direct regulation of *PCSK9* by the CE11 regulatory element, we used CRISPR/Cas9 genome editing in the human hepatocarcinoma HepG2 cell line which recapitulates the epigenetic landscape of *PCSK9* gene locus in the adult liver (**Figure 35A**). Since the original tested sequence (4 kb) spanned two exons, we deleted a smaller core region of CE11 (minCE11; 1.1 kb) included in the intron. The minCE11 fragment includes peaks of H3K27ac and H3K4me1, indicative of enhancer activity, in both adult human liver and HepG2 cells (**Figure 35A**). Deletion of the minCE11 region in HepG2 cells affected the expression levels of the major PCSK9 isoform, ENST00000302118.5 (**Figure 35B**), while not affecting the cerebellum-specific transcript, ENST00000490692.1 (**Figure 35C**). Unexpectedly, disruption of the regulatory element increased *PCSK9* gene expression instead of downregulating its transcription, which suggests that the logic underlying the mechanism of this regulatory element is more complex than initially thought.

Together, these findings show a differential regulation of *PCSK9* gene expression in two main domains of expression: the liver and the cerebellum. Here, we identified two regulatory elements contributing to this dual regulation and showed that the CE11 specifically affect the levels of the major isoform of *PCSK9* in hepatic cells, while not affecting the cerebellum isoform. Further dissection of these elements will lead to a better understanding of the regulatory networks controlling *PCSK9* gene expression and its cell type-specific contribution to LDL-cholesterol levels.

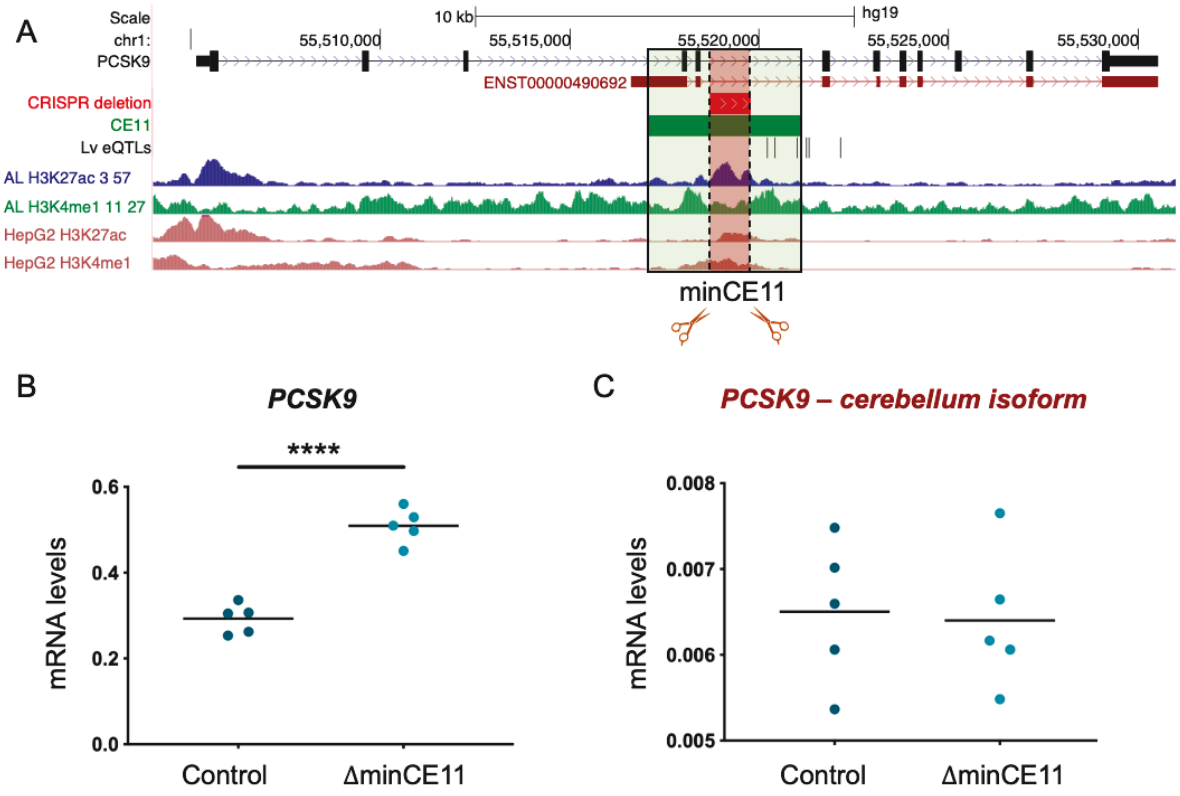

Figure 35 – CE11 regulates *PCSK9* major isoform in hepatic cells. A) UCSC genomic view of the *PCSK9* locus (hg19; chr1:55,504,000-55,531,000) indicating the major (black) and cerebellum-specific (red) isoforms. ChIP-seq data for the histone marks H3K27Ac and H3K4me1 are shown for human liver tissue (Roadmap epigenomics) and HepG2 human cells (ENCODE). *PCSK9* Liver eQTLs are contained in the CE11 enhancer (green) which spans two exons. Histone marks of active enhancers peaked at a fully-intronic minimal fragment (minCE11; indicated in red). Deletion of the minCE11 from the genome of HepG2 cells affected transcript levels of the major *PCSK9* isoform (**B**), while not affecting the cerebellum isoform (**C**).

Discussion

Dime que no tienes dudas
Sobre ninguna cosa
Confirmaré que eres una persona sospechosa

Los Punsetes, *Una persona sospechosa*.

Precise control of gene expression is essential for the correct development and functioning of tissues and organs (Smith and Shilatifard, 2014; Lupiáñez, Spielmann and Mundlos, 2016; Rickels and Shilatifard, 2018). Therefore, identifying and characterizing regulatory elements and how genetic variation affects their activity is crucial towards achieving precision medicine. In this context, CVDs are an important health burden to which the genomics field will contribute to a better understanding. AF is a major arrhythmia that affects over 30 million people worldwide (Chugh *et al.*, 2014) and is estimated to double by 2050 (Krijthe *et al.*, 2013). On the other hand, severe cases of atherosclerosis lead to most cases of ischemic heart disease and stroke which collectively killed ~10 million people in 2010 (Lozano *et al.*, 2012). In this work, we aimed to elucidate the genetic mechanism behind GWAS associations to CVDs for which we had to overcome existing limitations of current *in vivo* enhancer assays in mammals.

1. Towards higher-throughput discovery of regulatory elements.

After the discovery of the first viral enhancer (Banerji, Rusconi and Schaffner, 1981; Moreau *et al.*, 1981), classical transgenesis allowed for the *in vivo* characterization of mammalian enhancers (Banerji, Olson and Schaffner, 1983; Gillies *et al.*, 1983; Mercola *et al.*, 1983). While mouse transgenesis has effectively identified hundreds of regulatory elements involved in mammalian development and disease (Manzanares *et al.*, 2000; Lettice *et al.*, 2003; Nobrega *et al.*, 2003; Pennacchio *et al.*, 2006; Visel *et al.*, 2007), transient methods of *in vivo* ERAs are rather inefficient (Kvon, 2015) and has very limitedly evolved in forty years (Brinster *et al.*, 1981, 1985). Whereas transposases, such as the Tol2 system (Kawakami *et al.*, 2004), have increased ERA performance in zebrafish (Kawakami *et al.*, 2004; Bessa *et al.*, 2008, 2009), mice were not suitable for rapid and higher scale experiments (Bessa *et al.*, 2009). The Sleeping Beauty (Ivics *et al.*, 1997) or *piggyBac* (Cadiñanos and Bradley, 2007) systems have been used to generate mouse lines with integrated sensors of enhancer activity (Ruf *et al.*, 2011; Symmons *et al.*, 2014, 2016; Uslu *et al.*, 2014; Shima *et al.*, 2016). However, segregation of the multiple insertions and enhancer analysis ultimately requires high periods and costs of animal breeding and maintenance.

Our findings that the PB-ERA system is a convenient method to systematically assess enhancer activity in disease-associated genomic regions, opens the door to in-depth

interrogation of the ever-increasing number of risk loci. This not only applies to disease-associated regulatory regions, but also it is of relevance for biochemically annotated enhancers. Since only a fraction of predicted enhancers have regulatory activity after functional validation (Kvon, 2015) and high-throughput techniques of enhancer detection are very constrained to cell culture (Inoue and Ahituv, 2015) we need to make rapid progress in the *in vivo* characterization of the regulatory genome. Recently, another study aiming to scale up mouse ERAs developed 3-component method assisted by CRISPR in order to test the impact of human variants in the ZRS enhancer (Kvon *et al.*, 2020). This once again stresses the need for mouse *in vivo* ERAs of higher throughput. Here, we showed that the 2-component PB-ERA system yields an average 59% of transgenic embryos, being the most efficient system reported to the best of our knowledge. The number of integrations per transgenic embryo were not too high, thus enabling the capture of enhancer patterns with a minimized position effect. Furthermore, the PB-ERA system that we implemented was able to detect genomic regions with both enhancer and repressor activity, overcoming an existing limitation in the field, which is mostly focused on positive regulators of gene expression. We have also used the PB-ERA system to perform assays of enhancer blockade, being able to discriminate between silencers and insulators and showing the versatility of the system.

2. The regulatory potential behind GWAS susceptibility.

In the context of AF, the study of the nature of non-coding genetic associations has paid special attention to transcriptional enhancers. Yet, insights into how GWAS associations contribute to AF remains a challenge even for the most significant SNPs at the 4q25 locus. As previously mentioned, AF-SNPs in this locus lie within a gene desert where distal enhancer elements interact with the promoters of *PITX2* and *ENPEP* (Aguirre *et al.*, 2015; Ye *et al.*, 2016; Zhang *et al.*, 2019). On the one hand, *PITX2* encodes a developmental TF expressed in the atria (Kirchhof *et al.*, 2011) found to be decreased in AF patients (Chinchilla *et al.*, 2011) and also in a sheep model of induced AF (Alvarez-Franco *et al*., 2020). On the other hand, *ENPEP* encodes aminopeptidase A, a member of the renin-angiotensin system involved in hypertension (Mizutani *et al.*, 2008). Nevertheless, the variants in the locus do not correlate with atrial expression of *PITX2* in patients (Gore-Panter *et al.*, 2014), illustrating how

complex it is to ascertain the contribution of human variation to complex diseases. In this regard, individual lead-SNPs might not completely disrupt or create regulatory elements, what supports the notion that many of them might have a mild effect on gene expression.

In the last decade, GWAS performed in over two million people, including 200,000 AF cases, have identified 130 risk loci, including additional genomic regions associated through less characterized forms of genetic variation such as indels and CNVs. Of these, our study addresses twelve of the most significant associations and assesses their regulatory potential and target genes. While we observe an overlap between enhancers identified in cultured atrial myocytes and mouse transgenic embryos, it is however not surprising that our *in vivo* approach outperforms cell culture experiments. Our study shows that prioritization of candidate loci increases the success rate of enhancer identification as we found enhancers in all three loci presenting three layers of enhancer marks, i.e. the 7q31locus including *CAV1*, the 9q22 locus including *C9orf3* and the 14q23 locus including *SYNE2*. It is important to highlight that we are nonetheless still far from completely understanding the regulatory genome and cataloguing all CREs. The cardiac enhancer identified at the *KCNIP1* locus is an example of the former, since no other predictor supported this candidate beyond the previous association to AF of a 4.4 kb CNV (Tsai *et al.*, 2016). In this case, chromatin analysis of the region involved not only *KCNIP1* but also the TF-encoding *TLX3* gene in AF. Considering that the presence of the CNV in AF patients positively correlated with *KCNIP1* mRNA levels (Tsai *et al.*, 2016), it would be very interesting to also explore *TLX3* expression in patients and its putative role in AF.

Our study brings along again proteases of the renin-angiotensin system in association with AF (Healey and Connolly, 2003; Kumagai *et al.*, 2003; Zhang *et al.*, 2010; Martin *et al.*, 2015) since, similarly to *ENPEP* in the 4q25 locus that encodes aminopeptidase A (Aguirre *et al.*, 2015), the cardiac expression of the C9orf3|AOPEP gene, encoding aminopeptidase O, seems to be influenced by a heart enhancer in the 9q22 locus. On the other hand, the association at the 14q23 locus seems to be involved in response to mechanical stress and signal transduction between the nucleus and the cytoplasm of cardiomyocytes. Here, we found a cardiac enhancer regulating *SYNE2*, which encodes Nesprin-2, a giant nuclear envelope protein that links the lamins to the

cytoskeleton and is involved in myocyte nuclear positioning (Davidson *et al.*, 2020). Mice lacking Nesprin-1 and 2 developed cardiomyopathy, as well as mice lacking the C-terminal KASH domain of Nesprin-1 (Puckelwartz *et al.*, 2010; Banerjee *et al.*, 2014). Precisely, the short isoform of *SYNE2* that produces the alternative promoter that is close proximity with the *SYNE2*-AF enhancer contains the KASH domain. In a recent publication, similar results indicate that AF variants at the 14q23 locus affect the short isoform of *SYNE2* and implicate Nesprin2α1 in nuclear stiffness (Liu *et al.*, 2019).

The functional analysis performed at the 7q31 locus indicates that the large second intron of the *CAV1* gene might be a hub of enhancers or a cardiac 'super-enhancer' (Pott and Lieb, 2015). Although we do not endorse the term super-enhancer as conferring special properties to a new class of CREs, here we found a large region with regulatory activity. The 10 kb spanning the two transcriptional regulators that we described (*CAV1*-AF1 and *CAV1*-AF2) might harbor smaller modules that cooperatively regulate gene expression. In fact, deletion of several regions of this hub of CREs containing AF variants resulted in significant misregulation of target genes. Recent genetic screenings *in vitro* showed putative enhancer regions at this and other loci (van Ouwerkerk *et al.*, 2020), however this methodology still have high false discovery rates (FDR) since only 3 out of the best 10 putative enhancers that contained variants replicated the results in luciferase assays performed in the same cell line (van Ouwerkerk *et al.*, 2020). Furthermore, the 7q31 locus is an example of how important is to assess enhancer activity in its native chromatin region. Enhancer perturbation using CRISPR technology not only showed that the regions containing the variants rs3807989 and rs1173845 are true core modules of enhancers but also identified their target genes. Noteworthy, two of the target genes encode CAV1 and CAV2 proteins that are members of caveolae and are involved in mechanosensing. More surprising was that the AF enhancers in 7q31 also regulated *MET* and *TES* genes. On the one hand, *MET* encodes hepatocyte growth factor (HGF) receptor that plays a physiological cardio-protective role in adult cardiomyocytes preventing cardiomyocyte hypertrophy, heart fibrosis, and heart dysfunction (Arechederra *et al.*, 2013). On the other hand, *TES* encodes Testin, a member of the focal adhesions that connects the cell to the extracellular matrix and is involved in mechanical and regulatory signal transduction (Coutts *et al.*, 2003; Wang *et al.*, 2005). *TES* has been

associated to other cardiovascular diseases such as atherosclerosis and aneurism, playing important roles in endothelial and vascular smooth muscle cells where Testin can be found in the nucleus putatively co-regulating gene expression (Archacki *et al.*, 2012; Li *et al.*, 2020). Interestingly, *TES* is found upregulated in cardiomyocytes from the chronic sheep model of induced AF (Alvarez-Franco *et al.*, 2020). Therefore, further evaluation of *MET* and *TES* roles in AF might potentially be of clinical relevance.

Silencers are also essential in the coordinated regulation of gene expression and while recent reports have developed methods for their high-throughput identification in cell culture (Ngan *et al.*, 2020; Pang and Snyder, 2020), currently there were no mouse *in vivo* tools for their efficient characterization. The *ZFHX3*-AF silencer identified at the 16q22 locus directly regulates *ZFHX3* gene expression in a negative fashion and is able to outcompete heart enhancers *in vivo*. The genetic mechanism behind the AF associations at the 16q22 has remained elusive after a decade of research and even some reports have suggested that the first intron of *Zfhx3* have no regulatory potential in mice (van Ouwerkerk *et al.*, 2019). We showed that the silencer activity of this risk-associated CRE is not conserved between human and mouse, since deletion of the *Zfhx3*-af mouse ortholog region do not affect gene expression in cardiac cells. This might explain previous negative results using mouse models and emphasizes the need for new models in biomedical research. *ZFHX3*, also known as *ATBF1*, encodes a developmental transcription factor that has been involved in myogenic and neuronal differentiation (Berry *et al.*, 2001; Jung *et al.*, 2005). *ZFHX3* knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes (Kao *et al.*, 2016), where it might also have a role in tachypacing-induced inflammation through the regulation of *STAT3* (Jiang *et al.*, 2014). Altogether, our work on the 16q22 locus stresses out the importance of other types of CRE, such as silencers, when understanding the genetic contribution to disease risk.

3. *TBX5* might govern arrhythmia predisposition and perpetuation

In order to explore the convergence between intrinsic (genetic) and extrinsic (environmental) cues in AF perpetuation, we analyzed in detail transcriptomic data

from an induced model of AF. We found a core set of three genes that were downregulated in both atria upon acquisition of long-standing arrhythmia. For *TBX5* and *GJA1*, the two most significant downregulated genes, we identified AF-risk enhancers (*TBX5*-AF and *GJA1*-H3K27ac) controlling their cardiac gene expression that contained GWAS variants associated to the disease. These two AF-enhancers as well as *CAV1*-AF and *C9orf3*-AF are bound by TBX5 itself, an essential TF for cardiac development whose depletion from the adult heart can cause cardiac conduction defects (Nadadur *et al.*, 2016). Our data suggest that TBX5 might govern a gene regulatory network that contributes to AF susceptibility through atrial remodeling and starts with the downregulation of *TBX5* after electrical insults of the atria. We propose a model in which *TBX5* self-regulation is able to generate changes in the expression of other AF genes such as *GJA1*, *CAV1*, *CAV2*, *TES*, *MET*, *C9orf3*|*AOPEP*, as well as its paralog *TBX3* (**Figure 36**), a cardiac TF implicated in the development of the pacemaker cardiomyocytes of the sinoatrial node (Hoogaars *et al.*, 2007).

Conversely to the intronic *TBX5*-AF enhancer, the associations in the gene desert of the 6q22 locus represent a more challenging example of functional characterization. We found that a large genomic block of 18.7 kb controls the expression of the distal *GJA1* gene. Enhancer activity resides in the *GJA1*-H3K27ac sub-fragment, which confers heart specificity and contains a ~600 bp minimal enhancer (min*GJA1*-H3K27ac) essential for its function. This minimal enhancer is bound by cardiac TFs, containing multiple TFBS for GATA4, NKX2-5 and TBX5 as well as three polymorphisms (rs78437352-G>A, rs80105958 C>A, rs76014281 A>G). However, these variants do not overlap any of the consensus motifs for the previous TFs and we do not have evidence that they are linked to the AF tag-SNPs. Strikingly, deletion of the rest of the genomic block containing a Hi-C contact domain and the tag-SNP for AF (rs12664873) also led to downregulation of *GJA1*. Since this AF risk locus is located ~700 kb from its target gene, our data suggest that the *GJA1*-H3K27ac enhancer confers tissue-specificity, while the rest of the block might be required for enhancer-promoter 3D interaction.

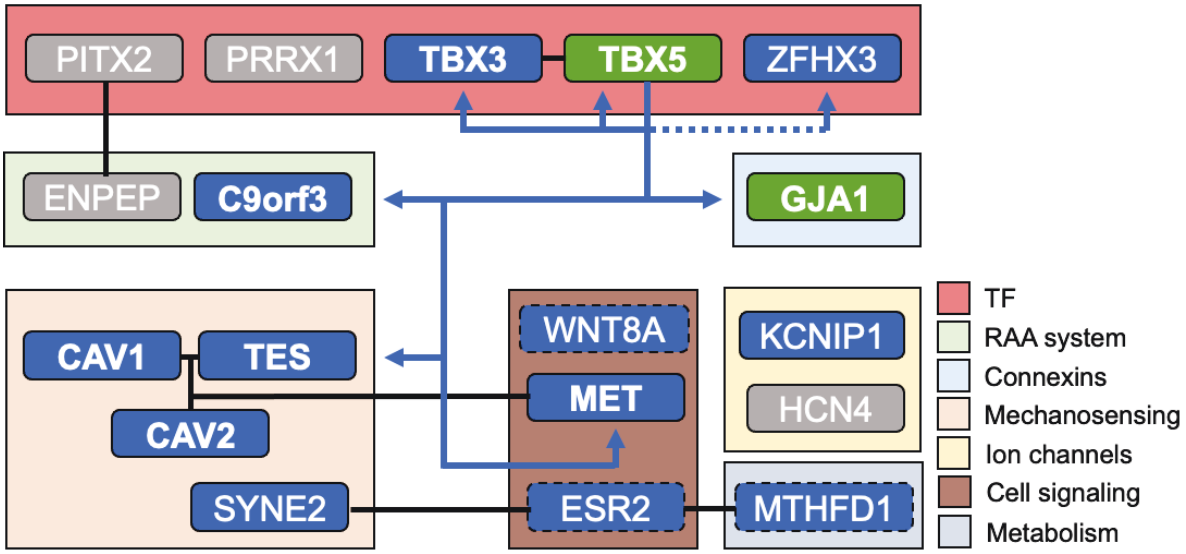

Figure 36 – Proposed gene regulatory network in AF. Beyond ion channels, several genes have been involved in AF with very different molecular functions from transcriptional regulators, cell signaling molecules and receptors or the RAA system. Genes belonging to the same risk locus are connected with black lines. In this thesis (blue and green rectangles) as well as in previous studies (grey rectangles), AF-regulatory elements have been found to control transcriptional levels of multiple genes, involving new players in arrhythmia pathophysiology. Dash lines in *WNT8A*, *ESR2* and *MTHFD1* represent a lower degree of evidence in disease-relevant tissues. *TBX5* and *GJA1* are found at the convergence between genetic susceptibility and arrhythmia perpetuation (green rectangles). *TBX5*, a cardiac TF, seems to govern part of this network by binding to AF-regulatory elements (blue continuous arrows) and promoters (blue dashed arrows) of many AF susceptibility genes, including *TBX5* itself and *GJA1*. TF, transcription factors; RAA, renin-angiotensin-aldosterone system.

A recent study tested the effect of large genomic deletions of AF risk loci in mouse models (van Ouwerkerk *et al.*, 2019). Indeed, the deletion of a 40-kb region overlapping our 600-bp minimal enhancer led to downregulation of *Gja1* in the adult atria. While we cannot assure that deleting the min*GJA1*-H3K27ac would phenocopy such a large deletion, it suggests that disrupting this minimal enhancer might affect *GJA1* expression throughout lifespan. Although the min*GJA1*-H3K27ac is necessary, the PB-ERA system showed that it is however not sufficient to drive reporter expression. Additional ChIP-seq datasets indicate that there might be another essential module contained in the larger *GJA1*-H3K27ac enhancer. This second module is also bound by cardiac TFs and, together with our min*GJA1*-H3K27ac, might be sufficient to fire reporter expression. A similar type of regulation has been described for *GJA5*, a paralog of *GJA1* regulated by two ~600 bp domains that are only able to drive reporter expression when a chimeric construct including both domains is microinjected in mouse embryos (Yang *et al.*, 2020).

Single-cell (sc)RNA-seq data from the developing mouse heart shows that *Gja1* is expressed in atrial myocytes from E9.5 although at very low levels, which then increase as the heart keeps developing (DeLaughter *et al.*, 2016). Although the *GJA1*-H3K27ac enhancer is fully active from E11.5 onwards, we captured activity starting at the earlier E9.5 stage (not shown), suggesting that *GJA1*-H3K27ac is the main enhancer of *GJA1* in the heart, including the atria. Since mice with oculodentodigital dysplasia (ODDD) have reduced *GJA1* expression in the atria and are more susceptible to develop sustained atrial arrhythmia after electrical stimulation (Tuomi, Tyml and Jones, 2011), a sustained lower expression of *GJA1* due to defects in the *GJA1*-H3K27ac enhancer might also contribute to arrhythmia development. Additionally, we found that the *GJA1*-H3K27ac enhancer is not only active in atria and atrial myocytes, but also it is a very strong ventricular enhancer. Precisely, the scRNA-seq data shows that *TBX5* is expressed predominantly in the left ventricle at E11.5 where the enhancer is more active, which further supports that TBX5 is conferring the cardiac specificity. Since *GJA1* mutations can lead to ventricular arrhythmia, studying the putative role of *GJA1*-H3K27ac in other forms of arrhythmic disorders might be of additional relevance.

The third common gene between chronic AF and GWAS predisposition was *JMJD1C*, encoding a histone demethylase which might also be involved in perpetuating AF through epigenetic mechanisms. Indeed, a global dysregulation of chromatin was described in the cardiomyocytes from sheep with induced AF. Chromatin remodelers were downregulated in the sheep AF model that also showed lower amounts of histones 3 and 4. This chromatin de-compaction led to increased expression of transposable elements in cardiomyocytes from both atria (Alvarez-Franco *et al.,* 2020), which has been recently proposed to happen in some pathological states and during aging (Wood and Helfand, 2013). Therefore, the implication of *JMJD1C* in AF deserves to be further investigated.

4. Dual regulation of *PCSK9* points towards a possible implications in neurological diseases.

PCSK9 is a circulating protein that mediates LDLR turnover via targeting its lysosomal degradation instead of recycling. The canonical pathway involves secreted PCSK9 binding to LDLR on the surface, internalization, and degradation. However,

intracellular endogenous PCSK9 can also target LDLR to lysosomes during the secretory pathway. Therefore, increased PCSK9 leads to low LDLR at the surface of hepatic cells and reduced LDL-cholesterol removal from the bloodstream, which ultimately induces atherosclerosis (Shapiro and Fazio, 2017). In addition to LDLR, PCSK9 also regulates the levels of other receptors like VLDLR, ApoeER2, LRP-1, CD36 and BACE1 at the membrane (Horton, Cohen and Hobbs, 2007; Jonas, Costantini and Puglielli, 2008; Poirier *et al.*, 2008; Canuel *et al.*, 2013; Demers *et al.*, 2015; Tang *et al.*, 2020).

At the transcriptional level, *LDLR* and *PCSK9* genes are coordinately regulated in the liver by sterol regulatory element-binding protein-2 (SREBP-2), a transcription factor that activates many genes involved in cholesterol metabolism in response to feeding (Horton *et al.*, 2003; Maxwell *et al.*, 2003). Thus, higher levels of *LDLR* correlate with higher levels of *PCSK9*, a mechanism that seems to limit cholesterol uptake. However, *PCSK9* is not only expressed in the liver and secreted to the bloodstream, but also expressed in the cerebellum from where it is secreted to the cerebrospinal fluid (CSF) (Rousselet *et al.*, 2011; Chen, Troutt and Konrad, 2014). The importance of this fluid in maintaining, feeding and cleaning the brain together with the cholesterol-rich membranes of neurons, made us hypothesized that *PCSK9* could be involved in stroke as well as other neurological disorders. For instance, Alzheimer's disease (AD) has been associated with dysregulation of brain cholesterol. However, the role of *PCSK9* in AD is controversial (Zimetti *et al.*, 2016; Courtemanche *et al.*, 2018). We performed an epigenetic analysis where we selected candidate regulatory elements potentially regulating *PCSK9* gene expression in a tissue-specific way. Using our PB-ERA technology, we have uncovered a dual regulation of *PCSK9* controlled by a liver- (CE11) and a cerebellum-specific (CE9) enhancer. Therefore, genetic variation associated to atherosclerosis could be increasing or decreasing the activity of such enhancers resulting in aberrant levels of *PCSK9* .

Treatment with two approved antibody-based PCSK9 inhibitors (evolocumab and alirocumab) reduces cholesterol levels and improves atherosclerosis (Raal *et al.*, 2012; Stein *et al.*, 2012). However, treatment with these antibodies is rather expensive and should be administered every two weeks (Shapiro, Tavori and Fazio, 2018). Conversely to previous therapy based on generic statins, the annual antibody-based

therapy costs thousands of euros. Since we have identified the possible key regulatory elements of *PCSK9* gene expression, now the question is: can we reduce atherosclerosis by (epi)genetically targeting *PCSK9* regulatory elements? Epigenetic therapy has been successfully applied to treat haploinsufficient obesity in mice (Matharu *et al.*, 2019). In this elegant work, the authors overexpressed the remaining functional copy of the haploinsufficient gene with CRISPR-mediated activation (CRISPRa). Epigenetic modulation of tissue-specific regulatory elements is useful to avoid undesired effects in other tissues. Here, we propose the opposite strategy: liver-specific downregulation of *PCSK9* by targeting CRISPR-mediated inhibition (CRISPRi) to the liver-specific CE11 regulatory element. LDL-cholesterol reduction through epigenetic silencing of *PCSK9* liver expression would open a new window of therapeutic potential, especially for those patients with resistance to PCSK9 inhibitor treatment (Shapiro *et al.*, 2018). Additionally, understanding how genomic variants in tissue-specific CREs affect *PCSK9* gene expression will enable assessing the atherosclerotic risk of people with non-coding mutations on such enhancers, which might improve diagnosis.

The discovery of a dual mechanism of expression of *PCSK9* (liver and cerebellum) raises another important question: what is the role of this protein in the brain? This is of great relevance for several reasons: i) PCSK9 in the brain can have a similar role as in the arteries which can have important implications in stroke and neurological diseases; ii) PCSK9 can have a different role in the brain which will uncover new functions of this protein; iii) since inhibition of PCSK9 is the aim of antibody-based treatments, understanding the role of PCSK9 in the brain can help predict potential adverse reactions caused by these treatments. Neither cholesterol nor PCSK9 cross the blood-brain barrier (BBB) under normal conditions (Chen, Troutt and Konrad, 2014; O'Connell and Lohoff, 2020). Similarly, PCSK9 inhibitors unlikely cross the BBB either (Shapiro, Tavori and Fazio, 2018; O'Connell and Lohoff, 2020). However, several studies have raised concern on neurocognitive adverse events caused by PCSK9 inhibitors (Robinson *et al.*, 2015; Lipinski *et al.*, 2016; Khan *et al.*, 2017). The EBBINGHAUS study, designed to assess neurocognitive adverse events, did not detect differences between patients treated with PCSK9 inhibitors or placebo after two years. However, the long-term effect of extreme cholesterol lowering by PCSK9

inhibitors remains unknown as well as the presence of depressive symptoms which were not evaluated at EBBINGHAUS (Giugliano *et al.*, 2017; Mannarino *et al.*, 2018).

The importance of assessing the long-term effect of the treatment with inhibitors resides in non-canonical functions of PCSK9 that might lead to other complications. Recent studies have implicated PCSK9 in inflammation, apoptosis, and immunity among other pathways (Apaijai *et al.*, 2019; Liu X, Bao X, Hu M, Chang H, Jiao M, Cheng J, Xie L, Huang Q, Li F, 2020; O'Connell and Lohoff, 2020; Tang *et al.*, 2020). For instance, high-fat diet caused severe hepatic steatosis, ER stress, inflammation and insulin resistance in a *Pcsk9* knockout mouse model (Lebeau *et al.*, 2019). Additional roles of PCSK9 are associated to liver regeneration, neurogenesis and neuronal differentiation (Seidah *et al.*, 2003; Rousselet *et al.*, 2011). Indeed, low maternal PCSK9 serum levels during pregnancy are associated with fetal neural tube defects (An *et al.*, 2015). Altered lipid metabolism has been extensively implicated in Alzheimer's disease (AD) and genes involved in cholesterol transport and metabolism are among the strongest associated loci to AD (Lambert *et al.*, 2013; Beecham *et al.*, 2014). However, whether PCSK9 levels are altered in AD as well as its putative role in this disease are controverted (Jonas, Costantini and Puglielli, 2008; Liu *et al.*, 2010; Zimetti *et al.*, 2016; Courtemanche *et al.*, 2018).

Due to the specific separation between cholesterol homeostasis in the brain from that in the rest of the organism, the specific role of PCSK9 in the nervous system needs to be elucidated. The presence of an evolutionary-conserved enhancer specific of cerebellum suggests a role in the brain. In line with this finding, we also observed a smaller isoform specifically expressed in this tissue. However, it still needs to be ascertained whether this isoform is protein coding. Altogether, our study of *PCSK9* gene regulation will contribute to a better understanding of lipid metabolism through the regulatory elements identified here, which might potentially be useful to reduce cholesterol levels through tissue-specific epigenetic therapy.

In this thesis we have improved the current technology for the *in vivo* interrogation of the genome, which is a major bottleneck in the field of genetics. With the aim of understanding the genetic contribution to cardiovascular diseases, we have applied our methodology to GWAS associated loci in an attempt to shed light on the genetic susceptibility to AF and atherosclerosis. We have followed a systematic approach to

study a dozen AF risk loci, identifying regulatory elements and their target genes for many of them. In particular, it is important to highlight the characterization of regulatory elements that regulate gene expression in a negative fashion and the involvement of new genes into the AF gene regulatory network where the *TBX5*-*GJA1* axis might play an important role. On the other hand, we have focused on the pro-atherosclerotic *PCSK9* gene in order to understand its regulation in the liver and the cerebellum. All in all, we present a framework to decipher the function of disease-associated loci, having generated a catalog of regulatory elements involved in disease-risk that we envision might be of help to understand the pathophysiology of cardiovascular diseases.

Conclusions

We accept her, we accept her
Gooble, Gobble,
One of us

Tod Browning, *Freaks*.

1. The PB-ERA system is an efficient tool to interrogate the genome which increases the throughput of mouse transgenesis and is suitable for the characterization of enhancers, silencers and insulators.

2. AF-risk variants are often part of cardiac-specific regulatory elements controlling the expression of cardiovascular-related genes.

3. Cardiac-specific regulatory elements at the 7q31 locus differentially control gene expression of target genes located upstream or downstream the regulatory elements and suggest a role for the genes *CAV1*, *CAV2*, *TES* and *MET* in AF susceptibility.

4. The 16q22 AF locus contains a human-specific silencer that controls *ZFHX3* gene expression and acts in a different cell types, including the heart.

5. *GJA1* and *TBX5* are putative core genes for AF perpetuation as found at the intersection between genetic susceptibility and atrial transcriptomic changes in a chronic model of the disease.

6. *GJA1* is regulated by a long-range conserved cardiac enhancer in the 6q22 AF-risk locus whose activity is mediated by cardiac TFs, including TBX5.

7. *TBX5* gene expression is controlled by an intronic enhancer associated to AF and other ECG traits. TBX5 regulates its own enhancer as well as many other AF enhancers, putatively creating a positive feedback loop of disease relevance.

8. The liver/cerebellum dual expression of the pro-atherosclerotic gene *PCSK9* is controlled by different enhancers such as the CE11 and the CE9 identified here that might be involved in atherosclerosis and suggest a role of *PCSK9* in the brain.

1. El sistema PB-ERA es una herramienta eficiente para interrogar el genoma, que aumenta el rendimiento de la generación de ratones transgénicos y es adecuada para la caracterización de potenciadores, silenciadores y aisladores genéticos.

2. Muchas de las variantes asociadas a fibrilación auricular forman parte de elementos reguladores cardíacos implicados en la regulación de genes involucrados en procesos cardiovasculares.

3. Los elementos reguladores cardíacos identificados en el locus 7q31 regulan de forma diferente la expresión de genes localizados en su extremo 5' de los localizados en su extremo 3' y sugieren que los genes *CAV1*, *CAV2*, *TES* and *MET* podrían estar implicados en fibrilación auricular.

4. El locus 16q22 contiene un silenciador específico de humanos que controla la expresión del gen *ZFHX3* en múltiples tejidos, incluyendo el corazón.

5. *GJA1* y *TBX5* son genes candidatos a estar implicados en perpetuar la fibrilación auricular como sugiere el hecho de encontrarlos en la intersección entre genes con predisposición genética (GWAS) y que cambian transcripcionalmente en un modelo crónico de la enfermedad.

6. La expresión del gen *GJA1* está regulada por elementos reguladores cardíacos del locus 6q22 que actúan desde larga distancia y están conservados evolutivamente. A su vez, la actividad de estos reguladores está mediada por factores de transcripción cardíacos, entre los que encontramos a TBX5.

7. La expression del gen *TBX5* está controlada por un element regulador localizado en uno de sus introns que contiene polimorfismos asociados a fibrilación auricular y otros fenotipos de electrocardiograma. TBX5 regula a su propio enhancer, así como otros enhancers asociados a fibrilación auricular, en lo que podría constituir un mecanismo de retroalimentación positiva involucrado en la cronificación de la enfermedad.

8. La expresión del gen pro-arteriosclerótico *PCSK9* en el hígado y cerebelo está controlada por medio de elementos reguladores como el CE11 y el CE9 identificados en esta tesis, que podrían estar implicados en arteriosclerosis y que sugieren un papel de *PCSK9* en el cerebro.

Bibliography

Abifadel, M. *et al.* (2003) “Mutations in PCSK9 cause autosomal dominant hypercholesterolemia,” *Nature Genetics*. doi: 10.1038/ng1161.

Aguirre, L. A. *et al.* (2015) “Long-range regulatory interactions at the 4q25 atrial fibrillation risk locus involve PITX2c and ENPEP,” *BMC Biology*. BioMed Central Ltd., 13(1), p. 26. doi: 10.1186/s12915-015-0138-0.

Ahituv, N. (2016) “Exonic enhancers: proceed with caution in exome and genome sequencing studies,” *Genome medicine*. doi: 10.1186/s13073-016-0277-0.

Alexander, J. M. *et al.* (2019) “Live-cell imaging reveals enhancer-dependent sox2 transcription in the absence of enhancer proximity,” *eLife*. doi: 10.7554/eLife.41769.

Alvarez-Franco, A. *et al.* (2020) “Transcriptome and proteome mapping in the sheep atria reveal molecular features of atrial fibrillation progression,” *Cardiovascular Research*. doi: 10.1093/cvr/cvaa307.

An, D. *et al.* (2015) “Identification of PCSK9 as a novel serum biomarker for the prenatal diagnosis of neural tube defects using iTRAQ quantitative proteomics,” *Scientific Reports*. doi: 10.1038/srep17559.

Andersson, R. and Sandelin, A. (2020) “Determinants of enhancer and promoter activities of regulatory elements,” *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0173-8.

Ang, Y.-S. *et al.* (2016) “Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis,” *Cell*. Cell Press, 167(7), pp. 1734-1749.e22. doi: 10.1016/J.CELL.2016.11.033.

Apaijai, N. *et al.* (2019) “Pretreatment with PCSK9 inhibitor protects the brain against cardiac ischemia/reperfusion injury through a reduction of neuronal inflammation and amyloid beta aggregation,” *Journal of the American Heart Association*. doi: 10.1161/JAHA.118.010838.

Archacki, S. R. *et al.* (2012) “Comparative gene expression analysis between coronary arteries and internal mammary arteries identifies a role for the TES gene in endothelial cell functions relevant to coronary artery disease,” *Human Molecular Genetics*. doi: 10.1093/hmg/ddr574.

Arechederra, M. *et al.* (2013) “Met signaling in cardiomyocytes is required for normal

cardiac function in adult mice," *Biochimica et Biophysica Acta - Molecular Basis of Disease*. doi: 10.1016/j.bbadis.2013.08.008.

Arnar, D. O. *et al.* (2006) "Familial aggregation of atrial fibrillation in Iceland," *European Heart Journal*. doi: 10.1093/eurheartj/ehi727.

Arnold, M. *et al.* (2015) "SNiPA: An interactive, genetic variant-centered annotation browser," *Bioinformatics*. doi: 10.1093/bioinformatics/btu779.

Banerjee, I. *et al.* (2014) "Targeted Ablation of Nesprin 1 and Nesprin 2 from Murine Myocardium Results in Cardiomyopathy, Altered Nuclear Morphology and Inhibition of the Biomechanical Gene Response," *PLoS Genetics*. doi: 10.1371/journal.pgen.1004114.

Banerji, J., Olson, L. and Schaffner, W. (1983) "A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes," *Cell*. doi: 10.1016/0092-8674(83)90015-6.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) "Expression of a -globin gene is enhanced by remote SV40 DNA sequences," *Cell*. doi: 10.1016/0092-8674(81)90413-X.

Bapat, A. *et al.* (2018) "Genomic basis of atrial fibrillation," *Heart*. BMJ Publishing Group, pp. 201–206. doi: 10.1136/heartjnl-2016-311027.

Basson, C. T. *et al.* (1997) "Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome," *Nature Genetics*. doi: 10.1038/ng0197-30.

Beauchamp, P. *et al.* (2012) "Electrical coupling and propagation in engineered ventricular myocardium with heterogeneous expression of connexin43," *Circulation Research*. doi: 10.1161/CIRCRESAHA.111.259705.

Bechmann, L. P. *et al.* (2012) "The interaction of hepatic lipid and glucose metabolism in liver diseases," *Journal of Hepatology*. doi: 10.1016/j.jhep.2011.08.025.

Beecham, G. W. *et al.* (2014) "Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer's Disease and Related Dementias," *PLoS Genetics*. doi: 10.1371/journal.pgen.1004606.

Behringer, R. *et al.* (2014) "Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition," *Cold Harbor Laboratory Press*.

Benjamin, E. J. *et al.* (2009) “Variants in ZFHX3 are associated with a trial fibrillation in individuals of European ancestry,” *Nature Genetics*, 41(8), pp. 879–881. doi: 10.1038/ng.416.

Berry, F. B. *et al.* (2001) “Positive and Negative Regulation of Myogenic Differentiation of C2C12 Cells by Isoforms of the Multiple Homeodomain Zinc Finger Transcription Factor ATBF1,” *Journal of Biological Chemistry*. doi: 10.1074/jbc.M010378200.

Bessa, J. *et al.* (2008) “meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye,” *Development*. doi: 10.1242/dev.011932.

Bessa, J. *et al.* (2009) “Zebrafish Enhancer Detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish,” *Developmental Dynamics*. doi: 10.1002/dvdy.22051.

Brand, A. H. *et al.* (1985) “Characterization of a ‘silencer’ in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer,” *Cell*. doi: 10.1016/0092-8674(85)90059-5.

Brinster, R. L. *et al.* (1981) “Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs,” *Cell*. doi: 10.1016/0092-8674(81)90376-7.

Brinster, R. L. *et al.* (1985) “Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.82.13.4438.

Bruneau, B. G. *et al.* (2001) “A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease,” *Cell*. doi: 10.1016/S0092-8674(01)00493-7.

Bruneau, B. G. (2013) “Signaling and transcriptional networks in heart development and regeneration,” *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a008292.

Buecker, C. and Wysocka, J. (2012) “Enhancers as information integration hubs in development: Lessons from genomics,” *Trends in Genetics*. doi: 10.1016/j.tig.2012.02.008.

Cadiñanos, J. and Bradley, A. (2007) “Generation of an inducible and optimized

piggyBac transposon systemy," *Nucleic Acids Research*. doi: 10.1093/nar/gkm446.

Canuel, M. *et al.* (2013) "Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Can Mediate Degradation of the Low Density Lipoprotein Receptor-Related Protein 1 (LRP-1)," *PLoS ONE*. doi: 10.1371/journal.pone.0064145.

Capelson, M. and Corces, V. G. (2004) "Boundary elements and nuclear organization," *Biology of the Cell*. doi: 10.1016/j.biolcel.2004.06.004.

Casserly, I. and Topol, E. (2004) "Convergence of atherosclerosis and Alzheimer's disease: Inflammation, cholesterol, and misfolded proteins," *Lancet*. doi: 10.1016/S0140-6736(04)15900-X.

Castillo-Davis, C. I. (2005) "The evolution of noncoding DNA: How much junk, how much func?," *Trends in Genetics*. doi: 10.1016/j.tig.2005.08.001.

Charest, J. *et al.* (2020) "Combinatorial Action of Temporally Segregated Transcription Factors," *Developmental Cell*. doi: 10.1016/j.devcel.2020.09.002.

Chen, Y. H. *et al.* (2003) "KCNQ1 gain-of-function mutation in familial atrial fibrillation," *Science*. doi: 10.1126/science.1077771.

Chen, Y. Q., Troutt, J. S. and Konrad, R. J. (2014) "PCSK9 is present in human cerebrospinal fluid and is maintained at remarkably constant concentrations throughout the course of the day," *Lipids*. doi: 10.1007/s11745-014-3895-6.

Cheng, J. P. X. *et al.* (2015) "Caveolae protect endothelial cells from membrane rupture during increased cardiac output," *Journal of Cell Biology*. doi: 10.1083/jcb.201504042.

Chinchilla, A. *et al.* (2011) "PITX2 insufficiency leads to atrial electrical and structural remodeling linked to arrhythmogenesis," *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.110.958116.

Christophersen, I. E. *et al.* (2009) "Familial aggregation of atrial fibrillation: A study in danish twins," *Circulation: Arrhythmia and Electrophysiology*. NIH Public Access, 2(4), pp. 378–383. doi: 10.1161/CIRCEP.108.786665.

Christophersen, I. E., Magnani, J. W., *et al.* (2017) "Fifteen Genetic Loci Associated with the Electrocardiographic P Wave," *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.116.001667.

Christophersen, I. E., Rienstra, M., *et al.* (2017) “Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation,” *Nature Genetics*. Nature Publishing Group, 49(6), pp. 946–952. doi: 10.1038/ng.3843.

Chugh, S. S. *et al.* (2014) “Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study,” *Circulation*. Lippincott Williams & WilkinsHagerstown, MD, 129(8), pp. 837–847. doi: 10.1161/CIRCULATIONAHA.113.005119.

Chung, J. H., Whiteley, M. and Felsenfeld, G. (1993) “A 5′ element of the chicken -globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila,” *Cell*. doi: 10.1016/0092-8674(93)80052-G.

Claycomb, W. C. *et al.* (1998) “HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.95.6.2979.

Cohen, J. C. *et al.* (2006) “ Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease ,” *New England Journal of Medicine*. doi: 10.1056/nejmoa054013.

Concordet, J. P. and Haeussler, M. (2018) “CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens,” *Nucleic Acids Research*. doi: 10.1093/nar/gky354.

Courtemanche, H. *et al.* (2018) “PCSK9 Concentrations in Cerebrospinal Fluid Are Not Specifically Increased in Alzheimer’s Disease,” *Journal of Alzheimer’s Disease*. doi: 10.3233/JAD-170993.

Coutts, A. S. *et al.* (2003) “TES is a novel focal adhesion protein with a role in cell spreading,” *Journal of Cell Science*. doi: 10.1242/jcs.00278.

Craig Venter, J. *et al.* (2001) “The sequence of the human genome,” *Science*. doi: 10.1126/science.1058040.

Creyghton, M. P. *et al.* (2010) “Histone H3K27ac separates active from poised enhancers and predicts developmental state,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1016071107.

Dai, W. *et al.* (2019) “A calcium transport mechanism for atrial fibrillation in Tbx5-mutant mice,” *eLife*. eLife Sciences Publications Ltd, 8. doi: 10.7554/eLife.41814.

Dardiotis, E. *et al.* (2012) “Cognitive impairment in heart failure,” *Cardiology Research and Practice*. doi: 10.1155/2012/595821.

Davidson, P. M. *et al.* (2020) “Nesprin-2 accumulates at the front of the nucleus during confined cell migration,” *EMBO reports*. doi: 10.15252/embr.201949910.

Dekker, J. *et al.* (2002) “Capturing chromosome conformation,” *Science*. doi: 10.1126/science.1067799.

DeLaughter, D. M. *et al.* (2016) “Single-Cell Resolution of Temporal Gene Expression during Heart Development.,” *Developmental cell*. NIH Public Access, 39(4), pp. 480–490. doi: 10.1016/j.devcel.2016.10.001.

Demers, A. *et al.* (2015) “PCSK9 Induces CD36 Degradation and Affects Long-Chain Fatty Acid Uptake and Triglyceride Metabolism in Adipocytes and in Mouse Liver,” *Arteriosclerosis, Thrombosis, and Vascular Biology*. doi: 10.1161/ATVBAHA.115.306032.

Desplantez, T. *et al.* (2012) “Connexin43 ablation in foetal atrial myocytes decreases electrical coupling, partner connexins, and sodium current,” *Cardiovascular Research*. doi: 10.1093/cvr/cvs025.

Diez-Roux, G. *et al.* (2011) “A high-resolution anatomical atlas of the transcriptome in the mouse embryo,” *PLoS Biology*. doi: 10.1371/journal.pbio.1000582.

Dixon, J. R. *et al.* (2012) “Topological domains in mammalian genomes identified by analysis of chromatin interactions,” *Nature*. doi: 10.1038/nature11082.

Donda, A. *et al.* (1996) “Identification and characterization of a human CD4 silencer,” *European Journal of Immunology*. doi: 10.1002/eji.1830260232.

Doni Jayavelu, N. *et al.* (2020) “Candidate silencer elements for the human and mouse genomes,” *Nature Communications*. doi: 10.1038/s41467-020-14853-5.

Ellinor, P. T. *et al.* (2005) “Familial aggregation in lone atrial fibrillation,” *Human Genetics*. Springer, 118(2), pp. 179–184. doi: 10.1007/s00439-005-0034-8.

Ellinor, P. T. *et al.* (2010) “Common variants in KCNN3 are associated with lone atrial fibrillation,” *Nature Genetics*, 42(3), pp. 240–244. doi: 10.1038/ng.537.

Ellinor, P. T. *et al.* (2012) “Meta-analysis identifies six new susceptibility loci for atrial fibrillation,” *Nature Genetics*. Nature Publishing Group, 44(6), pp. 670–675. doi:

10.1038/ng.2261.

Encode Project Consortium (2012) “An integrated encyclopedia of DNA elements in the human genome,” *Nature*.

Erdmann, J. *et al.* (2009) “New susceptibility locus for coronary artery disease on chromosome 3q22.3,” *Nature Genetics*. doi: 10.1038/ng.307.

Falk, E. (2006) “Pathogenesis of Atherosclerosis,” *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2005.09.068.

Ference, B. A. *et al.* (2017) “Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement fromthe European Atherosclerosis Society Consensus Panel,” *European Heart Journal*. doi: 10.1093/eurheartj/ehx144.

Filgueiras-Rama, D. *et al.* (2012) “Long-term frequency gradients during persistent atrial fibrillation in sheep are associated with stable sources in the left atrium,” *Circulation: Arrhythmia and Electrophysiology*. NIH Public Access, 5(6), pp. 1160–1167. doi: 10.1161/CIRCEP.111.969519.

Furniss, D. *et al.* (2008) “A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb,” *Human Molecular Genetics*. doi: 10.1093/hmg/ddn141.

Gasperini, M., Tome, J. M. and Shendure, J. (2020) “Towards a comprehensive catalogue of validated and target-linked human enhancers,” *Nature Reviews Genetics*, pp. 292–310. doi: 10.1038/s41576-019-0209-0.

Gaszner, M. and Felsenfeld, G. (2006) “Insulators: Exploiting transcriptional and epigenetic mechanisms,” *Nature Reviews Genetics*. doi: 10.1038/nrg1925.

Gibson, D. G. *et al.* (2009) “Enzymatic assembly of DNA molecules up to several hundred kilobases,” *Nature Methods*. doi: 10.1038/nmeth.1318.

Gilbert, L. A. *et al.* (2013) “XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes,” *Cell*. doi: 10.1016/j.cell.2013.06.044.

Gillies, S. D. *et al.* (1983) “A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene,” *Cell*. doi: 10.1016/0092-8674(83)90014-4.

Giraldo, P. *et al.* (2003) “Functional dissection of the mouse tyrosinase locus control region indentifies a new putative boundary activity,” *Nucleic Acids Research*. doi: 10.1093/nar/gkg793.

Giugliano, R. P. *et al.* (2017) “Cognitive Function in a Randomized Trial of Evolocumab,” *New England Journal of Medicine*. doi: 10.1056/nejmoa1701131.

Glass, C. K. and Witztum, J. L. (2001) “Atherosclerosis: The road ahead,” *Cell*. doi: 10.1016/S0092-8674(01)00238-0.

Gomez-Velazquez, M. *et al.* (2017) “CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart,” *PLoS Genetics*. doi: 10.1371/journal.pgen.1006985.

Gordon, M. G. *et al.* (2020) “lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements,” *Nature Protocols*. doi: 10.1038/s41596-020-0333-5.

Gore-Panter, S. R. *et al.* (2014) “Atrial Fibrillation Associated Chromosome 4q25 Variants Are Not Associated with PITX2c Expression in Human Adult Left Atrial Appendages,” *PLoS ONE*. Edited by N. H. Bishopric. Public Library of Science, 9(1), p. e86245. doi: 10.1371/journal.pone.0086245.

GTEx Consortium (2020) “The GTEx Consortium atlas of genetic regulatory effects across human tissues,” *Science*. doi: 10.1126/science.aaz1776.

Gudbjartsson, D. F. *et al.* (2007) “Variants conferring risk of atrial fibrillation on chromosome 4q25,” *Nature*. Nature Publishing Group, 448(7151), pp. 353–357. doi: 10.1038/nature06007.

Gudbjartsson, D. F. *et al.* (2009) “A sequence variant in ZFHX3 on 16q22 associates with a trial fibrillation and ischemic stroke,” *Nature Genetics*. doi: 10.1038/ng.417.

Gudbjartsson, D. F. *et al.* (2015) “Large-scale whole-genome sequencing of the Icelandic population,” *Nature Genetics*. doi: 10.1038/ng.3247.

Gupta, R. M. *et al.* (2017) “A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression,” *Cell*. Elsevier, 170(3), pp. 522-533.e15. doi: 10.1016/J.CELL.2017.06.049.

Gutstein, D. E. *et al.* (2001) “Conduction Slowing and Sudden Arrhythmic Death in

Mice With Cardiac-Restricted Inactivation of Connexin43," *Circulation Research*. Lippincott Williams & Wilkins, 88(3), pp. 333–339. doi: 10.1161/01.RES.88.3.333.

Hansson, G. K. and Hermansson, A. (2011) "The immune system in atherosclerosis," *Nature Immunology*. doi: 10.1038/ni.2001.

Van Der Harst, P. and Verweij, N. (2018) "Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease," *Circulation Research*. doi: 10.1161/CIRCRESAHA.117.312086.

Healey, J. S. and Connolly, S. J. (2003) "Atrial fibrillation: Hypertension as a causative agent, risk factor for complications, and potential therapeutic target," *American Journal of Cardiology*. doi: 10.1016/S0002-9149(03)00227-3.

Hilton, I. B. *et al.* (2015) "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers," *Nature Biotechnology*. doi: 10.1038/nbt.3199.

Hoogaars, W. M. H. *et al.* (2007) "Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria," *Genes and Development*. doi: 10.1101/gad.416007.

Horton, J. D. *et al.* (2003) "Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes," *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1534923100.

Horton, J. D., Cohen, J. C. and Hobbs, H. H. (2007) "Molecular biology of PCSK9: its role in LDL metabolism," *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2006.12.008.

Hsu, J. *et al.* (2018) "Genetic Control of Left Atrial Gene Expression Yields Insights into the Genetic Susceptibility for Atrial Fibrillation," *Circulation. Genomic and precision medicine*. NLM (Medline), 11(3), p. e002107. doi: 10.1161/CIRCGEN.118.002107.

Inoue, F. and Ahituv, N. (2015) "Decoding enhancers using massively parallel reporter assays," *Genomics*. doi: 10.1016/j.ygeno.2015.06.005.

Ivics, Z. *et al.* (1997) "Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells," *Cell*. doi: 10.1016/S0092-

8674(00)80436-5.

Jiang, Q. *et al.* (2014) “Down-regulation of ATBF1 activates STAT3 signaling via PIAS3 in pacing-induced HL-1 atrial myocytes,” *Biochemical and Biophysical Research Communications*. doi: 10.1016/j.bbrc.2014.05.041.

Jinek, M. *et al.* (2012) “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science*. doi: 10.1126/science.1225829.

Jinek, M. *et al.* (2013) “RNA-programmed genome editing in human cells,” *eLife*. doi: 10.7554/eLife.00471.

Jonas, M. C., Costantini, C. and Puglielli, L. (2008) “PCSK9 is required for the disposal of non-acetylated intermediates of the nascent membrane protein BACE1,” *EMBO Reports*. doi: 10.1038/embor.2008.132.

Jung, C. G. *et al.* (2005) “Homeotic factor ATBF1 induces the cell cycle arrest associated with neuronal differentiation,” *Development*. doi: 10.1242/dev.02098.

Kalcheva, N. *et al.* (2007) “Gap junction remodeling and cardiac arrhythmogenesis in a murine model of oculodentodigital dysplasia,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0705472105.

Kao, Y. H. *et al.* (2016) “ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes,” *International Journal of Cardiology*. doi: 10.1016/j.ijcard.2016.02.091.

Kawakami, K. *et al.* (2004) “A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish,” *Developmental Cell*. doi: 10.1016/j.devcel.2004.06.005.

Kellum, R. and Schedl, P. (1991) “A position-effect assay for boundaries of higher order chromosomal domains,” *Cell*. doi: 10.1016/0092-8674(91)90318-S.

Khan, A. R. *et al.* (2017) “Increased Risk of Adverse Neurocognitive Outcomes with Proprotein Convertase Subtilisin-Kexin Type 9 Inhibitors,” *Circulation: Cardiovascular Quality and Outcomes*. doi: 10.1161/CIRCOUTCOMES.116.003153.

Kim, T. K. *et al.* (2010) “Widespread transcription at neuronal activity-regulated enhancers,” *Nature*. doi: 10.1038/nature09033.

Kirchhof, P. *et al.* (2011) “PITX2c is expressed in the adult left atrium, and reducing

Pitx2c expression promotes atrial fibrillation inducibility and complex changes in gene expression," *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.110.958058.

Kirchhof, P. *et al.* (2016) "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal*. doi: 10.1093/eurheartj/ehw210.

Kojima, Y., Weissman, I. L. and Leeper, N. J. (2017) "The Role of Efferocytosis in Atherosclerosis," *Circulation*. doi: 10.1161/CIRCULATIONAHA.116.025684.

Koyama, S. *et al.* (2020) "Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease," *Nature Genetics*. doi: 10.1038/s41588-020-0705-3.

Krijthe, B. P. *et al.* (2013) "Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060," *European Heart Journal*. doi: 10.1093/eurheartj/eht280.

Kuhn, R. M., Haussler, D. and James Kent, W. (2013) "The UCSC genome browser and associated tools," *Briefings in Bioinformatics*. doi: 10.1093/bib/bbs038.

Kumagai, K. *et al.* (2003) "Effects of angiotensin II type 1 receptor antagonist on electrical and structural remodeling in atrial fibrillation," *Journal of the American College of Cardiology*. doi: 10.1016/S0735-1097(03)00464-9.

Kvon, E. Z. (2015) "Using transgenic reporter assays to functionally characterize enhancers in animals," *Genomics*. doi: 10.1016/j.ygeno.2015.06.007.

Kvon, E. Z. *et al.* (2020) "Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants," *Cell*. Elsevier, 0(0), pp. 1–10. doi: 10.1016/j.cell.2020.02.031.

Kwasnieski, J. C. *et al.* (2014) "High-throughput functional testing of ENCODE segmentation predictions," *Genome Research*. doi: 10.1101/gr.173518.114.

Lagace, T. A. (2014) "PCSK9 and LDLR degradation: Regulatory mechanisms in circulation and in cells," *Current Opinion in Lipidology*. doi: 10.1097/MOL.0000000000000114.

Laimins, L., Holmgren-Konig, M. and Khoury, G. (1986) "Transcriptional 'silencer'

element in rat repetitive sequences associated with the rat insulin 1 gene locus," *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.83.10.3151.

Lambert, J. C. *et al.* (2013) "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease," *Nature Genetics*. doi: 10.1038/ng.2802.

Lander, E. S. *et al.* (2001) "Initial sequencing and analysis of the human genome," *Nature*. doi: 10.1038/35057062.

Langlois, S. *et al.* (2008) "Caveolin-1 and -2 interact with connexin43 and regulate gap junctional intercellular communication in keratinocytes," *Molecular Biology of the Cell*. doi: 10.1091/mbc.E07-06-0596.

Larsen, P. R., Harney, J. W. and Moore, D. D. (1986) "Repression mediates cell-type-specific expression of the rat growth hormone gene," *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.83.21.8283.

Lebeau, P. F. *et al.* (2019) "Pcsk9 knockout exacerbates diet-induced non-alcoholic steatohepatitis, fibrosis and liver injury in mice," *JHEP Reports*. doi: 10.1016/j.jhepr.2019.10.009.

Leo-Macias, A., Agullo-Pascual, E. and Delmar, M. (2016) "The cardiac connexome: Non-canonical functions of connexin43 and their role in cardiac arrhythmias," *Seminars in Cell and Developmental Biology*. doi: 10.1016/j.semcdb.2015.12.002.

Lettice, L. A. *et al.* (2003) "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly," *Human Molecular Genetics*. doi: 10.1093/hmg/ddg180.

Li, Yang *et al.* (2020) "Variants of Focal Adhesion Scaffold Genes Cause Thoracic Aortic Aneurysm," *Circulation Research*. doi: 10.1161/CIRCRESAHA.120.317361.

Lin, J. *et al.* (2008) "The regulation of the cardiac potassium channel (HERG) by caveolin-1," *Biochemistry and Cell Biology*. doi: 10.1139/O08-118.

Lip, G. Y. H. *et al.* (2016) "Atrial fibrillation," *Nature Reviews Disease Primers*. doi: 10.1038/nrdp.2016.16.

Lipinski, M. J. *et al.* (2016) "The impact of proprotein convertase subtilisin-kexin type 9 serine protease inhibitors on lipid levels and outcomes in patients with primary

hypercholesterolaemia: A network meta-analysis," *European Heart Journal*. doi: 10.1093/eurheartj/ehv563.

Liu, M. *et al.* (2010) "PCSK9 is not involved in the degradation of LDL receptors and BACE1 in the adult mouse brain," *Journal of Lipid Research*. doi: 10.1194/jlr.M006635.

Liu, N. *et al.* (2019) "Atrial fibrillation associated common risk variants in SYNE2 lead to lower expression of nesprin-2 and increased nuclear stiffness," *bioRxiv*. doi: 10.1101/708057.

Liu X, Bao X, Hu M, Chang H, Jiao M, Cheng J, Xie L, Huang Q, Li F, L. C. (2020) "Inhibition of PCSK9 potentiates immune checkpoint therapy for cancer," *Nature*. doi: 10.1038/s41586-020-2911-7.

Lonsdale, J. *et al.* (2013) "The Genotype-Tissue Expression (GTEx) project," *Nature Genetics*. doi: 10.1038/ng.2653.

Low, S. K. *et al.* (2017) "Identification of six new genetic loci associated with atrial fibrillation in the Japanese population," *Nature Genetics*. Nature Publishing Group, 49(6), pp. 953–958. doi: 10.1038/ng.3842.

Lozano, R. *et al.* (2012) "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*. doi: 10.1016/S0140-6736(12)61728-0.

Lubitz, S. A. *et al.* (2014) "Novel genetic markers associate with atrial fibrillation risk in Europeans and Japanese," *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2013.12.015.

Lubitz, S. A. *et al.* (2017) "Genetic risk prediction of atrial fibrillation," *Circulation*. Lippincott Williams and Wilkins, 135(14), pp. 1311–1320. doi: 10.1161/CIRCULATIONAHA.116.024143.

Luna-Zurita, L. *et al.* (2016) "Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis," *Cell*. doi: 10.1016/j.cell.2016.01.004.

Lunyak, V. V. *et al.* (2007) "Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis," *Science*. doi: 10.1126/science.1140871.

Luo, M. H., Li, Y. S. and Yang, K. P. (2007) "Fibrosis of collagen I and remodeling of

connexin 43 in atrial myocardium of patients with atrial fibrillation," *Cardiology*. doi: 10.1159/000095501.

Lupiáñez, D. G., Spielmann, M. and Mundlos, S. (2016) "Breaking TADs: How Alterations of Chromatin Domains Result in Disease," *Trends in Genetics*. doi: 10.1016/j.tig.2016.01.003.

Lusis, A. J. (2000) "Atherosclerosis," *Nature*. doi: 10.1038/35025203.

Mannarino, M. R. *et al.* (2018) "PCSK9 and neurocognitive function: Should it be still an issue after FOURIER and EBBINGHAUS results?," *Journal of Clinical Lipidology*. doi: 10.1016/j.jacl.2018.05.012.

Manolio, T. A. *et al.* (2009) "Finding the missing heritability of complex diseases," *Nature*. doi: 10.1038/nature08494.

Manolio, T. A. (2010) "Genomewide Association Studies and Assessment of the Risk of Disease," *New England Journal of Medicine*. Edited by W. G. Feero and A. E. Guttmacher. Massachussetts Medical Society, 363(2), pp. 166–176. doi: 10.1056/NEJMra0905980.

Manzanares, M. *et al.* (2000) "Conservation and elaboration of Hox gone regulation during evolution of the vertebrate head," *Nature*. doi: 10.1038/35048570.

Martin, R. I. R. *et al.* (2015) "Genetic variants associated with risk of atrial fibrillation regulate expression of PITX2, CAV1, MYOZ1, C9orf3 and FANCC," *Journal of Molecular and Cellular Cardiology*. Academic Press, 85, pp. 207–214. doi: 10.1016/j.yjmcc.2015.06.005.

Matharu, N. *et al.* (2019) "CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency," *Science*. doi: 10.1126/science.aau0629.

Matharu, N. and Ahituv, N. (2020) "Modulating gene regulation to treat genetic disorders," *Nature Reviews Drug Discovery*. doi: 10.1038/s41573-020-0083-7.

Maxwell, K. N. *et al.* (2003) "Novel putative SREBP and LXR target genes identified by microarray analysis in liver of cholesterol-fed mice," *Journal of Lipid Research*. doi: 10.1194/jlr.M300203-JLR200.

Meng, H. and Bartholomew, B. (2018) "Emerging roles of transcriptional enhancers in

chromatin looping and promoter-proximal pausing of RNA polymerase II," *Journal of Biological Chemistry*. doi: 10.1074/jbc.R117.813485.

Mercola, M. *et al.* (1983) "Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus," *Science*, 221(4611), pp. 663–665. doi: 10.1126/science.6306772.

Mikhaylichenko, O. *et al.* (2018) "The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription," *Genes and Development*. doi: 10.1101/gad.308619.117.

Mizutani, S. *et al.* (2008) "New insights into the importance of aminopeptidase A in hypertension," *Heart Failure Reviews*. doi: 10.1007/s10741-007-9065-7.

Mohenska, M. *et al.* (2019) "3D-Cardiomics: A spatial transcriptional atlas of the mammalian heart," *bioRxiv*. doi: 10.1101/792002.

Montefiori, L. *et al.* (2018) "A promoter interaction map for cardiovascular disease genetics," *eLife*, 7. doi: 10.7554/eLife.35788.

Moreau, P. *et al.* (1981) "The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants," *Nucleic Acids Research*, 9(22), pp. 6047–6068. doi: 10.1093/nar/9.22.6047.

Nadadur, R. D. *et al.* (2016) "Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm," *Science Translational Medicine*. doi: 10.1126/scitranslmed.aaf4891.

Nelson, C. P. *et al.* (2017) "Association analyses based on false discovery rate implicate new loci for coronary artery disease," *Nature Genetics*. doi: 10.1038/ng.3913.

Ngan, C. Y. *et al.* (2020) "Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development," *Nature Genetics*. doi: 10.1038/s41588-020-0581-x.

Nielsen, J. B. *et al.* (2018) "Biobank-driven genomic discovery yields new insight into atrial fibrillation biology," *Nature Genetics*. Nature Publishing Group, pp. 1234–1239. doi: 10.1038/s41588-018-0171-3.

Nikpay, M. *et al.* (2015) "A comprehensive 1000 Genomes-based genome-wide

association meta-analysis of coronary artery disease," *Nature Genetics*. doi: 10.1038/ng.3396.

Nobrega, M. A. *et al.* (2003) "Scanning Human Gene Deserts for Long-Range Enhancers," *Science*. doi: 10.1126/science.1088328.

O'Connell, E. M. and Lohoff, F. W. (2020) "Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) in the Brain and Relevance for Neuropsychiatric Disorders," *Frontiers in Neuroscience*. doi: 10.3389/fnins.2020.00609.

Ocaña, O. H. *et al.* (2017) "A right-handed signalling pathway drives heart looping in vertebrates," *Nature*. Nature Publishing Group, 549(7670), pp. 86–90. doi: 10.1038/nature23454.

van Ouwerkerk, A. F. *et al.* (2019) "Identification of atrial fibrillation associated genes and functional non-coding variants," *Nature Communications*. Nature Publishing Group, 10(1), p. 4755. doi: 10.1038/s41467-019-12721-5.

van Ouwerkerk, A. F. *et al.* (2020) "Identification of Functional Variant Enhancers Associated with Atrial Fibrillation.," *Circulation research*. American Heart Association Bethesda, MD , p. CIRCRESAHA.119.316006. doi: 10.1161/CIRCRESAHA.119.316006.

Ozaki, K. *et al.* (2002) "Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction," *Nature Genetics*. doi: 10.1038/ng1047.

Pang, B. and Snyder, M. P. (2020) "Systematic identification of silencers in human cells," *Nature Genetics*. Nature Research, 52(3), pp. 254–263. doi: 10.1038/s41588-020-0578-5.

Parton, R. G. and Del Pozo, M. A. (2013) "Caveolae as plasma membrane sensors, protectors and organizers," *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3512.

Paznekas, W. A. *et al.* (2009) "GJA1 mutations, variants, and connexin 43 dysfunction as it relates to the oculodentodigital dysplasia phenotype," *Human Mutation*. doi: 10.1002/humu.20958.

Pennacchio, L. A. *et al.* (2006) "In vivo enhancer analysis of human conserved non-coding sequences," *Nature*. Nature Publishing Group, 444(7118), pp. 499–502. doi: 10.1038/nature05295.

Perez-Pinera, P. *et al.* (2013) “RNA-guided gene activation by CRISPR-Cas9-based transcription factors,” *Nature Methods*. doi: 10.1038/nmeth.2600.

Pers, T. H., Timshel, P. and Hirschhorn, J. N. (2015) “SNPsnap: A Web-based tool for identification and annotation of matched SNPs,” *Bioinformatics*. doi: 10.1093/bioinformatics/btu655.

Pertea, M. and Salzberg, S. L. (2010) “Between a chicken and a grape: Estimating the number of human genes,” *Genome Biology*. doi: 10.1186/gb-2010-11-5-206.

Poirier, S. *et al.* (2008) “The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2,” *Journal of Biological Chemistry*. doi: 10.1074/jbc.M708098200.

Pott, S. and Lieb, J. D. (2015) “What are super-enhancers?,” *Nature Genetics*. doi: 10.1038/ng.3167.

Puckelwartz, M. J. *et al.* (2010) “Nesprin-1 mutations in human and murine cardiomyopathy,” *Journal of Molecular and Cellular Cardiology*. doi: 10.1016/j.yjmcc.2009.11.006.

Raal, F. *et al.* (2012) “Low-density lipoprotein cholesterol-lowering effects of AMG 145, a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 serine protease in patients with heterozygous familial hypercholesterolemia: The reduction of LDL-C with PCSK9 inhibiti,” *Circulation*. doi: 10.1161/CIRCULATIONAHA.112.144055.

Rada-Iglesias, A. *et al.* (2011) “A unique chromatin signature uncovers early developmental enhancers in humans,” *Nature*. doi: 10.1038/nature09692.

Ran, F. A. *et al.* (2013) “Genome engineering using the CRISPR-Cas9 system,” *Nature Protocols*. doi: 10.1038/nprot.2013.143.

Rao, S. S. P. *et al.* (2014) “A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping,” *Cell*. doi: 10.1016/j.cell.2014.11.021.

Rashid, S. *et al.* (2005) “Decreased lipid synthesis in livers of mice with disrupted Site-1 protease gene,” *PNAS*. National Academy of Sciences, 98(24), pp. 13607–13612. doi: 10.1073/pnas.201524598.

Reaume, A. G. *et al.* (1995) “Cardiac malformation in neonatal mice lacking connexin43.,” *Science (New York, N.Y.)*, 267(5205), pp. 1831–4. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/7892609 (Accessed: April 22, 2019).

Ricci, C. *et al.* (2018) “PCSK9 induces a pro-inflammatory response in macrophages,” *Scientific Reports*. doi: 10.1038/s41598-018-20425-x.

Rickels, R. and Shilatifard, A. (2018) “Enhancer Logic and Mechanics in Development and Disease,” *Trends in Cell Biology*. doi: 10.1016/j.tcb.2018.04.003.

Roadmap Epigenomics Consortium *et al.* (2015) “Integrative analysis of 111 reference human epigenomes,” *Nature*. Nature Publishing Group, 518(7539), pp. 317–329. doi: 10.1038/nature14248.

Robinson, J. G. *et al.* (2015) “Efficacy and Safety of Alirocumab in Reducing Lipids and Cardiovascular Events,” *New England Journal of Medicine*. doi: 10.1056/nejmoa1501031.

Roselli, C. *et al.* (2018) “Multi-ethnic genome-wide association study for atrial fibrillation,” *Nature Genetics*. Nature Publishing Group, 50(9), pp. 1225–1233. doi: 10.1038/s41588-018-0133-9.

Roth, G. A. *et al.* (2017) “Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015,” *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2017.04.052.

Rousselet, E. *et al.* (2011) “PCSK9 reduces the protein levels of the LDL receptor in mouse brain during development and after ischemic stroke.,” *Journal of lipid research*. American Society for Biochemistry and Molecular Biology, 52(7), pp. 1383–91. doi: 10.1194/jlr.M014118.

Ruf, S. *et al.* (2011) “Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor,” *Nature Genetics*. doi: 10.1038/ng.790.

Ryu, H. *et al.* (2018) “Massively parallel dissection of human accelerated regions in human and chimpanzee neural progenitors,” *bioRxiv*. doi: 10.1101/256313.

Sainz de Aja, J. *et al.* (2019) “ The pluripotency factor NANOG controls primitive hematopoiesis and directly regulates Tal1 ,” *The EMBO Journal*. doi: 10.15252/embj.201899122.

Samani, N. J. *et al.* (2007) “Genomewide Association Analysis of Coronary Artery Disease,” *New England Journal of Medicine*. doi: 10.1056/nejmoa072366.

Sawada, S. *et al.* (1994) “A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development,” *Cell*. doi: 10.1016/0092-8674(94)90140-6.

Schoenfelder, S. and Fraser, P. (2019) “Long-range enhancer–promoter contacts in gene expression control,” *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0128-0.

Schunkert, H. *et al.* (2011) “Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease,” *Nature Genetics*. doi: 10.1038/ng.784.

Seidah, N. G. *et al.* (2003) “The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): Liver regeneration and neuronal differentiation,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0335507100.

Seidah, N. G. and Prat, A. (2007) “The proprotein convertases are potential targets in the treatment of dyslipidemia,” *Journal of Molecular Medicine*. doi: 10.1007/s00109-007-0172-7.

Shapiro, M. D. *et al.* (2018) “Diagnosing resistance to a proprotein convertase subtilisin/kexin type 9 inhibitor,” *Annals of Internal Medicine*. doi: 10.7326/M17-2485.

Shapiro, M. D. and Fazio, S. (2017) “PCSK9 and atherosclerosis - lipids and beyond,” *Journal of Atherosclerosis and Thrombosis*. doi: 10.5551/jat.RV17003.

Shapiro, M. D., Tavori, H. and Fazio, S. (2018) “PCSK9 from basic science discoveries to clinical trials,” *Circulation Research*. doi: 10.1161/CIRCRESAHA.118.311227.

Shima, Y. *et al.* (2016) “A mammalian enhancer trap resource for discovering and manipulating neuronal cell types,” *eLife*. doi: 10.7554/eLife.13503.

Shiratori, H. *et al.* (2001) “Two-step regulation of left-right asymmetric expression of Pitx2: Initiation by nodal signaling and maintenance by Nkx2,” *Molecular Cell*. doi: 10.1016/S1097-2765(01)00162-9.

Smemo, S. *et al.* (2014) “Obesity-associated variants within FTO form long-range functional connections with IRX3,” *Nature*. doi: 10.1038/nature13138.

Smith, E. and Shilatifard, A. (2014) “Enhancer biology and enhanceropathies,” *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2784.

Smith, R. P. *et al.* (2013) “Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model,” *Nature Genetics*. doi: 10.1038/ng.2713.

Spielmann, M., Lupiáñez, D. G. and Mundlos, S. (2018) “Structural variation in the 3D genome,” *Nature Reviews Genetics*. doi: 10.1038/s41576-018-0007-0.

Staerk, L. *et al.* (2017) “Atrial Fibrillation: Epidemiology, Pathophysiology, Clinical Outcomes,” *Circulation Research*. doi: 10.1161/CIRCRESAHA.117.309732.

Stees, J. S. *et al.* (2012) “Recruitment of transcription complexes to enhancers and the role of enhancer transcription,” *Biology*. doi: 10.3390/biology1030778.

Stein, E. A. *et al.* (2012) “Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: A phase 2 randomised controlle,” *The Lancet*. doi: 10.1016/S0140-6736(12)60771-5.

Suzuki, S. *et al.* (2015) “A hyperactive piggyBac transposon system is an easy-to-implement method for introducing foreign genes into mouse preimplantation embryos,” *Journal of Reproduction and Development*. doi: 10.1262/jrd.2014-157.

Symmons, O. *et al.* (2014) “Functional and topological characteristics of mammalian regulatory domains,” *Genome Research*. doi: 10.1101/gr.163519.113.

Symmons, O. *et al.* (2016) “The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances,” *Developmental Cell*. doi: 10.1016/j.devcel.2016.10.015.

Takahashi, K. *et al.* (2007) “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors,” *Cell*. doi: 10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors,” *Cell*. doi: 10.1016/j.cell.2006.07.024.

Tam, V. *et al.* (2019) “Benefits and limitations of genome-wide association studies,” *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0127-1.

Tang, Y. *et al.* (2020) “Research progress on alternative non-classical mechanisms of

PCSK9 in atherosclerosis in patients with and without diabetes," *Cardiovascular Diabetology*. doi: 10.1186/s12933-020-01009-4.

Tiana, M. *et al.* (2012) "A role for insulator elements in the regulation of gene expression response to hypoxia," *Nucleic Acids Research*. doi: 10.1093/nar/gkr842.

Tsai, C. T. *et al.* (2016) "Genome-wide screening identifies a KCNIP1 copy number variant as a genetic predictor for atrial fibrillation," *Nature Communications*. Nature Publishing Group, 7, p. 10190. doi: 10.1038/ncomms10190.

Tucker, N. R. *et al.* (2017) "Diminished PRRX1 Expression Is Associated with Increased Risk of Atrial Fibrillation and Shortening of the Cardiac Action Potential," *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.117.001902.

Tuomi, J. M., Tyml, K. and Jones, D. L. (2011) "Atrial tachycardia/fibrillation in the connexin 43 G60S mutant (Oculodentodigital dysplasia) mouse," *American Journal of Physiology - Heart and Circulatory Physiology*. doi: 10.1152/ajpheart.01094.2010.

Udvardy, A., Maine, E. and Schedl, P. (1985) "The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains," *Journal of Molecular Biology*. doi: 10.1016/0022-2836(85)90408-5.

Uslu, V. V. *et al.* (2014) "Long-range enhancers regulating Myc expression are required for normal facial morphogenesis," *Nature Genetics*. doi: 10.1038/ng.2971.

Villar, D. *et al.* (2015) "Enhancer evolution across 20 mammalian species," *Cell*. doi: 10.1016/j.cell.2015.01.006.

Visel, A. *et al.* (2007) "VISTA Enhancer Browser - A database of tissue-specific human enhancers," *Nucleic Acids Research*. doi: 10.1093/nar/gkl822.

Visel, A. *et al.* (2009) "ChIP-seq accurately predicts tissue-specific activity of enhancers," *Nature*. doi: 10.1038/nature07730.

Vogel, F. (1964) "A PRELIMINARY ESTIMATE OF THE NUMBER OF HUMAN GENES.," *Nature*. doi: 10.1038/201847a0.

Wall, J. D. and Pritchard, J. K. (2003) "Haplotype blocks and linkage disequilibrium in the human genome," *Nature Reviews Genetics*. doi: 10.1038/nrg1123.

Wang, Y. *et al.* (2005) "Visualizing the mechanical activation of Src," *Nature*. doi:

10.1038/nature03469.

Weltner, J. *et al.* (2018) “Human pluripotent reprogramming with CRISPR activators,” *Nature Communications*. doi: 10.1038/s41467-018-05067-x.

Wijffels, M. C. E. F. *et al.* (1995) “Atrial fibrillation begets atrial fibrillation: A study in awake chronically instrumented goats,” *Circulation*. doi: 10.1161/01.CIR.92.7.1954.

Wilkins, E. *et al.* (2017) “European Cardiovascular Disease Statistics 2017,” *European Heart Network*.

Willer, C. J. *et al.* (2008) “Newly identified loci that influence lipid concentrations and risk of coronary artery disease,” *Nature Genetics*. doi: 10.1038/ng.76.

Williamson, I. *et al.* (2019) “Developmentally regulated Shh expression is robust to TAD perturbations,” *Development (Cambridge)*. doi: 10.1242/dev.179523.

Wood, J. G. and Helfand, S. L. (2013) “Chromatin structure and transposable elements in organismal aging,” *Frontiers in Genetics*. doi: 10.3389/fgene.2013.00274.

Yang, K. C. *et al.* (2014) “Caveolin-1 modulates cardiac gap junction homeostasis and arrhythmogenecity by regulating csrc tyrosine kinase,” *Circulation: Arrhythmia and Electrophysiology*. doi: 10.1161/CIRCEP.113.001394.

Yang, T. *et al.* (2020) “Conjugated activation of myocardial-specific transcription of Gja5 by a pair of Nkx2-5-Shox2 co-responsive elements,” *Developmental Biology*. doi: 10.1016/j.ydbio.2020.07.003.

Yang, Y. *et al.* (2004) “Identification of a KCNE2 gain-of-function mutation in patients with familial atrial fibrillation,” *American Journal of Human Genetics*. doi: 10.1086/425342.

Ye, J. *et al.* (2016) “A Functional Variant Associated with Atrial Fibrillation Regulates PITX2c Expression through TFAP2a,” *American Journal of Human Genetics*. Cell Press, 99(6), pp. 1281–1291. doi: 10.1016/j.ajhg.2016.10.001.

Yeom, Y. I. *et al.* (1996) “Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells,” *Development*, 122(3).

Yusa, K. *et al.* (2011) “A hyperactive piggyBac transposase for mammalian applications,” *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1008322108.

Zhang, M. *et al.* (2019) “Long-range Pitx2c enhancer–promoter interactions prevent predisposition to atrial fibrillation,” *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(45), pp. 22692–22698. doi: 10.1073/pnas.1907418116.

Zhang, Y. *et al.* (2010) “The role of renin-angiotensin system blockade therapy in the prevention of atrial fibrillation: A meta-analysis of randomized controlled trials,” *Clinical Pharmacology and Therapeutics*. doi: 10.1038/clpt.2010.123.

Zimetti, F. *et al.* (2016) “Increased PCSK9 cerebrospinal fluid concentrations in Alzheimer’s disease,” *Journal of Alzheimer’s Disease*. doi: 10.3233/JAD-160411.

Publications

The work described in this thesis is included in the following manuscripts that will be submitted this year:

Victorino J, Rollan I, Rouco R, Adan J, Manzanares M (2021). *Systematic in vivo interrogation identifies novel enhancers and silencers associated to Atrial Fibrillation*. Nucl Acid Res (to submit).

Victorino J, Rollan I, Manzanares M (2021). *Genetic susceptibility and induced atrial fibrillation converge on the TBX5-GJA1 axis*. PLOS Genet (to submit).

During the development of the thesis, the following review has been submitted for publication:

Victorino J†, Alvarez A†, Manzanares M (2021). *Epigenomic drivers and the regulatory component of atrial fibrillation*. J Mol Cell Cardiol (submitted, invited review; †co-first authors).

The collaboration in other research projects during the development of the thesis has resulted in the following publication:

Lopez-Jimenez E†, Sainz de Aja J†, Badia-Careaga C#, Barral A#, Rollan I#, Rouco R#, Santos E#, Tiana M#, **Victorino J**#, Sanchez-Iranzo H, Acemel RD, Torroja C, Adan J, Andres-Leon E, Gomez-Skarmeta JL, Giovinazzo G, Sanchez-Cabo F, Manzanares M (2019). *Pluripotency factors regulate the onset of Hox cluster activation in the early embryo*. bioRxiv doi: https://doi.org/10.1101/564658 (†co-first authors; #equal contribution).