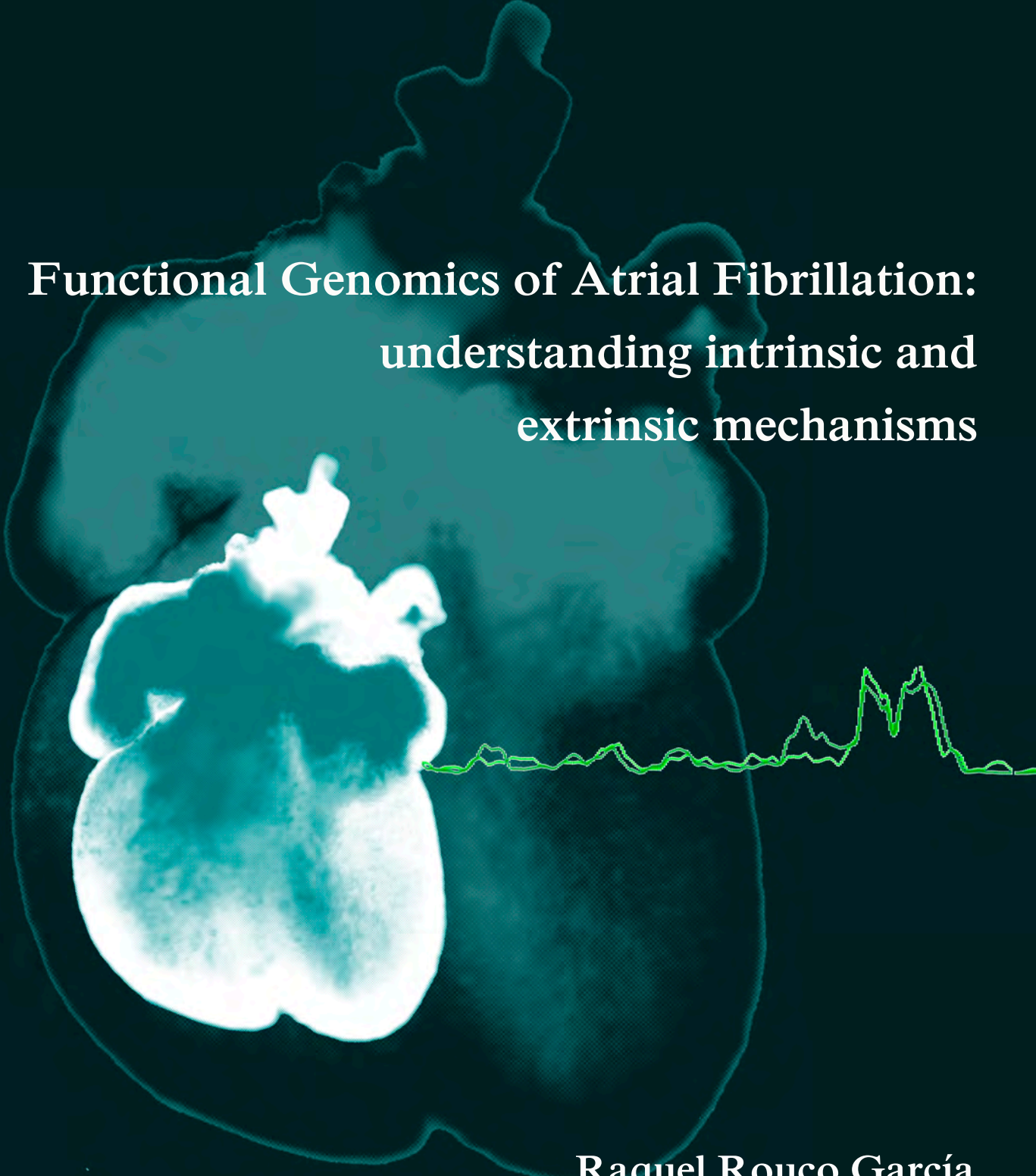


Universidad Autónoma de Madrid

Postgraduate Programme in Molecular Biosciences



**Functional Genomics of Atrial Fibrillation:  
understanding intrinsic and  
extrinsic mechanisms**

**Raquel Rouco García**

**Madrid, 2019**





**Department of Biochemistry**

**Faculty of Medicine**

**Universidad Autónoma de Madrid**

**Functional Genomics of Atrial Fibrillation:  
understanding intrinsic and  
extrinsic mechanisms**

**Raquel Rouco García, BSc Biology**

**Doctoral thesis directed by Miguel Manzanares Fourcade**

**CNIC, Madrid 2019**





I hereby certify that RAQUEL ROUCO GARCÍA has carried out the experimental work leading to his PhD thesis entitled "*Functional Genomics of Atrial Fibrillation: understanding intrinsic and extrinsic mechanisms*" 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.

Madrid, 22<sup>nd</sup> February 2019



Miguel Manzanares  
Centro Nacional de Investigaciones Cardiovasculares-CNIC  
Melchor Fernández Almagro, 3  
28029 Madrid, Spain

tel: (34) 91 453 12 00  
fax: (34) 91 453 13 04  
mmanzanares@cnic.es  
<http://www.cnic.es/en/desarrollo/genomica>



Todd J. Herron, PhD  
Associate Research Scientist  
Internal Medicine-Cardiovascular Medicine and Molecular & Integrative Physiology  
Lefkofsky Scholar  
Director, Cardiovascular Regeneration Core Laboratory



February 23, 2019

RE: PhD Thesis Report for Raquel Rouco Garcia

Dear Sir/Madam:

I write to provide a report on Raquel Rouco Garcia's PhD Thesis entitled, "*Functional Genomics of Atrial Fibrillation: understanding intrinsic and extrinsic mechanisms*". In addition, I am excited to provide my enthusiastic endorsement of her candidacy for the PhD degree. Under the direction of Miguel Manzaneres Fourcade, she has performed a detailed analysis of molecular and genetic changes that occur in the process of atrial fibrillation, a common cardiac arrhythmia that is the number one cause of embolic stroke. This thesis work encompasses a multi-faceted study relying on the use of a diverse array of model systems and approaches. Remarkably, her thesis work has also relied on contributions from an international team of investigators that has coalesced to provide new mechanistic insight into disease mechanisms.

The extensive genetic analysis provided by Genome Wide Association Studies (GWAS) has shed light on genetic defects that may cause disease. However, the majority of single nucleotide polymorphisms (SNPs) identified for atrial fibrillation reside in non-coding regions of genome. This has made it difficult to understand how these mutations contribute to disease mechanisms since they are not present in the proteome. To address this problem, Ms. Rouco Garcia hypothesized in her thesis work that these common non-coding SNPs reside in functional regions of the genome that regulate the expression of other genes that contribute to the disease phenotype. In silico modeling was used to screen for common variant associations and Circular Chromosome Conformation Capture followed by deep sequencing (4C-seq) identified that locus 7q31 contains a potential regulatory element inside intron2 of *CAV1* that through chromatin folding could establish long-range interactions with *CAV1*, *TES*, *MET*, and *CAPZA2*. In human induced pluripotent stem cell derived cardiomyocytes, deletion of this regulatory element using CRISPR/Cas9 resulted in reduced expression of the *CAV1* gene. This may lead to electrophysiological remodeling associated with atrial fibrillation because the *CAV1* gene encodes for caveolin, a protein essential for cell membrane structure and pooling of ion channels. Importantly, the thesis work also shows the important expression of caveolin 1 in the mouse heart, which persists through adulthood. This approach using hiPSC derived cardiomyocytes was largely performed in my laboratory at the University of Michigan Cardiovascular Regeneration Core Laboratory. The application of genome editing to hiPSCs is a cutting edge approach that provides novel mechanistic insight into disease mechanisms. In her pursuit of the aims of this thesis, Ms. Rouco Garcia developed the gene editing approach, executed complex gene transfer experiments and laborious single cell cloning and genetic screening. The successful execution of these complex approaches underscores her suitability for receiving the PhD degree.

826, Room 223M  
2800 Plymouth Rd.  
Ann Arbor, MI 48106  
P: (734) 998-0460  
F: (734) 998-7711  
[todsherr@umich.edu](mailto:todsherr@umich.edu)



Todd J. Herron, PhD  
Associate Research Scientist  
Internal Medicine-Cardiovascular Medicine and Molecular & Integrative Physiology  
Lefkolsky Scholar  
Director, Cardiovascular Regeneration Core Laboratory



FRANKEL CARDIOVASCULAR CENTER  
MICHIGAN MEDICINE

Further, this thesis work performed careful genetic analysis of the changes that occur in atrial fibrillation disease progression. These studies relied on a sheep model of persistent atrial fibrillation. The combination of approaches using human, mouse and sheep models provides a comprehensive study on AF genetics and disease mechanisms. In addition to providing new technical advances, this thesis work provides novel hypotheses to the underlying molecular changes that may prove to be therapeutic targets to treat atrial fibrillation. I hope that you will accept my recommendation for awarding Ms. Rouco Garcia a well-deserved PhD degree.

Best regards,

Todd Herron, PhD



## The University of Chicago

---

Pritzker School of Medicine  
Biological Sciences Division

Marcelo A. Nobrega M.D., Ph.D.  
Professor  
Department of Human Genetics  
920 East 58<sup>th</sup> Street CLSC 515-E  
Chicago, IL 60637  
Phone: (773) 834-7919  
Fax (773) 834-8470  
nobrega@uchicago.edu  
<http://www.genes.uchicago.edu/faculty.html>

February 25, 2019

**RE: Report for the thesis entitled "Functional Genomics of Atrial Fibrillation: understanding intrinsic and extrinsic mechanisms" by Raquel Rouco Garcia.**

To Whom It May Concern,

It is my pleasure to write a summary report for Raquel Rouco Garcia's thesis, and strongly endorsing her request of a PhD. I met Raquel two years ago, when I visited CNIC and gave a seminar. On that occasion, she presented her ongoing work to me, exchanging ideas and troubleshooting. It was a very productive interaction, I was impressed by the sense of ownership she had on her project, being able to defend her experimental design, analysis of data and interpretation of her results. She left the impression of a mature PhD student, pursuing a very interesting and innovative project, in an excellent environment in Dr Manzanera's lab. Below is a summary report of her thesis work.

Spotting the genetic variation that is associated with heart diseases, as Atrial Fibrillation, has shown that most of disease-linked variants are located in the non-coding regions of the genome. This work explores the possibility that disease associated variants are located in regulatory areas whose alteration could lead to a change in gene function. Moreover, this thesis uses a large animal model of atrial fibrillation to examine the molecular mechanisms that are modified during disease progression and therefore alters the normal physiology programs of the organism.

The introduction provides a thorough review of the large number of topics combined in this thesis. Overall, it introduces the main questions and the ambitious objectives addressed in this work. The results first describe the experiments performed by the candidate towards unveiling interactions in *loci* associated to the disease and suggesting novel candidate genes whose expression alteration could contribute to atrial fibrillation. In this sense, *CAV1* is identified as an interesting candidate gene whose expression is tracked during heart development using a LacZ reporter mouse line. This reveals a clear pattern of expression in the atria, the main tissue affected in the disease. More interestingly, CRISPR/Cas9 deletion in human derived cardiomyocytes from induced pluripotent stem cells, of a candidate regulatory element identified by the candidate and located in *CAV1* causes reduction of its expression. In the second part of the results section, the transcriptomic and proteomic changes that occurs during atrial fibrillation progression in a sheep induced model of the disease have been characterized by RNA-seq and

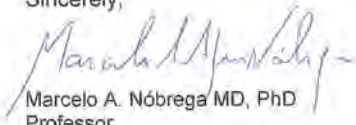


LC-MS/MS. Analysis of the data unveil changes in gene expression involved in the structural remodeling that takes place in the left atrium as a consequence of the disease progression. It also has led to the identification of profound alteration of mitochondrial protein levels that led to an upregulation of mitochondrial complexes and supercomplexes that are the key components of OXPHOS system. This work has broadly studied several relevant components of the molecular mechanisms underlying atrial fibrillation and through the discussion, the value of the findings have been clearly put in context.

Overall, this thesis has addressed a complex and common disease with sophisticated tools, proving that Raquel has been able to manage different and complex technical approaches to fulfill the objectives of her thesis. Her work will be of great interest to the cardiovascular genetics community, and I have no doubt will be highly cited. For all of this, Raquel Rouco Garcia deserves to be awarded a PhD.

Please do not hesitate to contact me if you need further information.

Sincerely,



Marcelo A. Nóbrega MD, PhD  
Professor  
Department of Human Genetics  
University of Chicago  
Chicago, IL, USA  
[nobrega@uchicago.edu](mailto:nobrega@uchicago.edu)

A mi familia.  
Especialmente a Tere, José,  
David y Amadeo.





## ACKNOWLEDGEMENTS

---

*El ser humano no es un participio, es un gerundio.*

Ortega y Gasset

*Science and everyday life should not be separated.*

Rosalind Franklin



Hace nueve años por estas fechas comenzó mi aventura en el mundo de la investigación. Acababa de terminar la asignatura de Genética y tenía ganas de aprender más sobre esa asombrosa molécula que es el ADN. Para mi fortuna mi profesor César Benito (Mario), buscaba algún alumno que quisiera unirse a su laboratorio para investigar. Tras dos años y medio comencé a conocer el significado de ser científica, siempre de la mano de mi primer mentor, **Mario**, que a día de hoy sigue siendo un gran consejero. Gracias por aquella oportunidad y por todo lo que me has enseñado.

A mi segundo gran mentor, **Miguel**, lo conocí en el verano de tercero de carrera. Mi curiosidad por profundizar más en el universo de la genética me llevo claramente a querer formar parte de su laboratorio. Miguel, gracias por apostar por mi, por animarme y ayudarme a desarrollarme como científica. Por todo el tiempo que me has dedicado tanto para los buenos como para los malos resultados, por dejar que me equivoque y que encuentre soluciones, por confiar en mi para múltiples colaboraciones, porque durante toda la tesis he aprendido mucho por el camino.

Gracias por no ser un jefe sin más, por ser también una gran persona. Estoy segura de que esa es la combinación perfecta para que hayas sido capaz de reunir en un mismo laboratorio a un grupo de personas increíbles que conforman el legado de los “Manzanares”. Todos y cada uno de los miembros presentes y pasados me habéis marcado de alguna manera, y espero que esta tesis sea solo un punto y seguido de las historias que hemos compartido y de las que podemos compartir.

**Luis**, tu fuiste la primera persona que me dedicó su tiempo y atención en el laboratorio, gracias por tus consejos y por enseñarme qué es la querencia. **Elena**, contigo aprendí a ver las cosas desde diferentes puntos de vista y el lado crítico de la vida. **Eva Fernández-Cáceres** gracias por darme a conocer la bioinformática y por descubrirme Sevilla y la calle feria (¿o era la calle fiesta?). **Aurora**, aunque tu etapa con nosotros fue breve, nos contagiaste tu alegría y optimismo. **Teresa**, siempre serás un referente en el laboratorio y tu vocación didáctica y divulgadora me enseñó muchas cosas y aún hoy sigo aprendiendo gracias a ti. **Melisa**, contigo comenzó toda la experiencia de los 4Cs, gracias por introducirme en ese interesante mundo de la estructura de la cromatina. **Julio**, admiro tu inagotable e incansable sed de ciencia. **Sergio**, atento y resolutivo, eres un gran científico y una mejor persona, siempre has sido un ejemplo que seguir, gracias por toda tu ayuda. **Claudio**, gracias por compartir todo tu conocimiento conmigo, por ayudarme y escucharme, y aunque digas lo contrario eres buen profesor. **Isa**, eres una incansable trabajadora, gracias por tu esfuerzo e implicación en el laboratorio para que

todo funcione, seguro que el futuro sabrá recompensar tu esfuerzo. **Majo**, observadora y atenta, gracias por preocuparte por mi, por ayudarme siempre que lo he necesitado y sobretodo gracias por tus consejos, siempre muy útiles, tu tenacidad seguro que te lleva lejos. **Alba**, gracias por compartir proyecto conmigo porque, aunque seguro en muchos momentos no ha sido fácil, creo que has conseguido hacer maravillas, tú vales mucho. **María**, te has metido en el bolsillo al laboratorio entero desde que llegaste, y aunque ha sido breve el periodo de tiempo que hemos trabajado juntas me alegro enormemente de ello, he aprendido mucho y mejor aún me he divertido, haces que sea muy fácil trabajar a tu lado, gracias también por saber escuchar y aconsejar. **Jesús**, eres un luchador siempre dispuesto a sacar tus ideas adelante, mucho ánimo con la recta final, mi consejo es que busques el lado positivo de todo lo que has hecho, que es mucho. **Antonio**, o más bien debería decir Pionono estoy segura de que acabarás tu tesis firmando como tal, gracias por hacerme reír, por compartir ese rincón de té y muchas conversaciones a través de nuestro hueco de la estantería, has hecho que mi día a día sea mucho más divertido, estoy segura de que serás un gran científico, pero eres aún mejor cocinero. **Gonzalo**, contigo he compartido muchas aventuras, primero fue en aquel workshop en el CRG y luego seguimos liándonos la manta a la cabeza con tus 4Cs, soy optimista y seguro que todo tu esfuerzo llega a convertirse en una fantástica tesis. **Sara**, supiste implicarte para que intentáramos entender un poco más de bioinformática contigo. **Antonio López**, el ingeniero del grupo, tú también supiste adaptarte a nosotros para que te entendiéramos, eres el alma de la fiesta. **Marcos**, el becario de la becaria, como tu mismo te autodenominaste, fue una alegría poder trabajar contigo, me hiciste sentir mayor al ver en ti y en tus inquietudes como me sentía yo al empezar la tesis, eres único, estoy segura de que llegarás lejos y no me lo quiero perder. **Marta**, gracias por tu eterna sonrisa y por tu espíritu aventurero, es contagioso y estoy segura de que te llevará muy lejos. **Marina**, a pesar del poquito tiempo que llevas en el laboratorio, ya has sabido adaptarte a nosotros, sea donde sea seguro que consigues hacer una tesis fantástica. **Javi**, para ti solo tengo palabras de agradecimiento por dejarte enseñar y por todo lo que me has ayudado, y porque a pesar de los múltiples problemas en cultivos no te has desanimado y has seguido dando lo mejor de ti.

El mundo de los cultivos celulares me ha llevado también a conocer a gente maravillosa con la que ha sido fácil trabajar. El equipo de células pluripotentes (**Giovanna, Elisa, Nines y Belén**) gracias por vuestro soporte y ayuda en el día a día de cultivos, es un placer trabajar con vosotras. **Jarek**, gracias por querer compartir conmigo tus

conocimientos. **Celia**, intensa y bellísima persona, gracias por preocuparte siempre de mi, por ser una incansable trabajadora y poner siempre un toque de humor en cultivos. **Marilina**, gracias por tu ayuda desde el minuto cero, por dejarte enseñar, por compartir tus historias conmigo, por todo tu esfuerzo, te deseo todo lo mejor.

También quiero agradecer a todo el grupo del LAB-JAC que me ha acogido como una más, incluido **Jorge** que además de buen científico, tiene buen ojo para fichar a gente excepcional. **Elías**, te admiro por tu sosiego en el día a día de la ciencia, gracias por tener siempre las palabras adecuadas para cada momento, desde una broma a buen consejo. **Diana**, la enciclopedia del grupo, entregada a la ciencia y siempre dispuesta a resolver mis dudas y a enseñarme. **Carmen**, siempre perfeccionista, mucho ánimo con todo lo que está por venir, tú puedes con ello. **Carol**, gracias por tu aleatoriedad. **Natalia**, entrañable y tenaz supiste ganarte a todo el mundo en poquito tiempo, espero verte pronto. **Inés**, eres la promesa del grupo, hablar contigo siempre es muy fácil. **María Rosaria**, MRP, o debería decir doctora Pricolo, gracias a ti descubrimos Nápoles y me contagiaste tu energía. **Carla**, eres una gran amiga, son muchos los buenos momentos que hemos compartido juntas y espero que haya muchos más, estoy segura de que encontrarás tu lugar en la ciencia, sabes que siempre podrás contar conmigo sin importar la parte del mundo en la que estemos, mucho ánimo con la recta final. **Ángel**, gracias por tu buen humor, por tener siempre el GIF adecuado y dejarme que echara gasolina al fuego de tus ideas, fue muy divertido trabajar contigo en el PhDay.

Gracias, a todos los miembros del “pichiday”. Desde el primer año que me uní he conocido a muchas personas fantásticas, pero me quedo especialmente con **Ana Paredes**, siempre con tu sonrisa dispuesta a arrasar con las redes sociales o con lo que se ponga por delante y con **María Gómez**, gracias por tu ilusión y tu incansable trabajo por sacar lo mejor del PhDay.

Gracias a todas las personas que forman o han formado parte de la 3 Sur y otras plantas del CNIC con los que he tenido la fortuna de pasar el día a día o trabajar codo con codo. **Cris Sánchez**, sin miedo a lo que venga por delante, me has enseñado que siempre es posible seguir adelante. **Oliver**, echo de menos esas conversaciones filosóficas sobre la vida. **Sandra**, gracias por tu risa contagiosa. **Sara González**, siempre alegrándonos con tus cumplidos, has sabido adaptarte a las circunstancias y seguro que llegarás lejos. **María Sánchez**, independientemente de la hora del día siempre tienes una sonrisa que ofrecer. **Sara Cogliati**, gracias por enseñarme el maravilloso mundo de la mitocondria. **Miguel Sánchez**, siempre dispuesto a echarme una mano. **Antonio Quilez** y **José**,

vosotros junto con Alba habéis hecho que esas largas noches aislando cardiomiocitos fueran más divertidas y llevaderas. Gracias a los CICERONE de mi promoción, **Noelia, Laura, Ángel, Dani, Víctor, Fran, Marcos Sande, Marcos Siguero, Carlos, Amanda, Arturo, Marichu** por las cervezas y fiestas que hemos compartido, siempre recordaré como empezamos. Gracias a todos los científicos que me he cruzado durante estos años, especialmente a todos aquellos con los que he colaborado porque he aprendido mucho con vosotros (**David Filgueiras, Miguel Ángel del Pozo, Rafael Domínguez, Juan Tena** y los que ahora mismo no recuerdo).

Gracias a las múltiples unidades técnicas del CNIC que han hecho mi trabajo más fácil (genómica, bioinformática, celómica, vectores virales...). Especialmente a **Alberto Benguria** y a **Carlos Torroja** por ayudarme a sacar adelante los 4Cs. Gracias **Luis Miguel, Chema** y **Juande**, vosotros me acogisteis desde el primer día que pisé el CNIC, y desde entonces os habéis preocupado por mí. Chema y Juande, siempre habéis sido buenos consejeros (especialmente Chema si de organizar un viaje se trataba) y me habéis escuchado cuando lo he necesitado.

Gracias a todas esas personas que trabajáis en la logística del CNIC (bioseguridad, recepción, cafetería, almacén, informática, compras, formación, instrumentación, seguridad y un largo etcétera). Especialmente gracias a **Sonia** y a **Cristina Giménez** por hacer todo el papeleo mucho más llevadero, y a **Juan Carlos** por estar dispuesto siempre a echarme una mano siempre con los problemas informáticos.

También te quiero dar las gracias Miguel, porque gracias a ti conocí a mi tercer mentor, José Jalife. **Pepe** gracias por estar siempre dispuesto a enseñarme, gracias por tu confianza y darme la oportunidad de trabajar en tu laboratorio. Gracias a ti y a **Paloma**, por hacer mi aterrizaje en Michigan mucho más llevadero y por preocuparos por mí.

At this point, I will change to English to give thanks to all these wonderful people that I met in Ann Arbor. **Todd**, you taught me the value of working with hiPSCs and you have always advised me well. **Andre**, I have been very lucky working with you, thank you for teaching me everything about cell culture and about life, I really admire you. **Kate**, thanks for your patience and everyday smile. **Jeff**, thanks for being always ready to help me. **Justin**, thanks to you I improved my English listening. **Guadalupe**, thanks for all the time that you have spent on me, thanks for taking care of me and share your knowledge with me. **Kuljeet**, thanks for being a helping hand. **Daniela**, thanks for your help inside and outside the lab, I always remember all the experiences that we shared. Thanks to you I met Valdivia's lab, **Hector, Carmen** and **Francisco**, you have also treated me as

one more of you, thanks to you parties I always felt at home. Thanks to the sheep people and other members of the lab, of the CAR or the UMICH (**Rafael, Andy, Steve, Roberto, Eric, Sergey, Lakshmi, Evan, Ede, Don, Raso, Allison, Mohammed, Omer, Keita, Bildge**), that were there to teach me and help me or to enjoy time outside the lab.

I also want to thank the “Spanish team in Ann Arbor”, gracias por todos los buenos momentos que pasamos juntos. **Ricardo**, eres un gran investigador y una mejor persona, siempre humilde y dispuesto a ayudar. **Pablo**, aunque breve fue una alegría conocerte, estoy segura de que el futuro te llevará muy lejos. **Ismael**, rápidamente te convertiste en un amigo y compañero de viajes, aunque fuera por A2. **Óscar y Sephora** no hay palabras suficientes para agradecerlos lo mucho que me habéis ayudado, gracias por acogerme las dos ocasiones que estuve por allí, gracias por transmitirme vuestra alegría y felicidad familiar para que me sintiera como en casa os deseo que seáis muy felices.

Y por supuesto, no puedo terminar sin dar las gracias a mi familia, la de toda la vida y la que he elegido con el tiempo.

Gracias a mis amigos, a esos que conozco desde que teníamos poco más de tres años. **Adri, Mari**, gracias porque vosotras más que nadie habéis estado pendientes de mi, dispuestas a darme unas palabras de ánimo cuando más lo he necesitado. **Lucía**, porque a pesar de que nos vemos muy de vez en cuando nuestra amistad no caduca. Gracias a **Jorge, Marga, Irene, Mar, Diego, Champi, Tapi, Ester, Cimarra, Puebla, Monti** por todos los buenos momentos que hemos pasado juntos. Gracias a **María Fernandez Lera**, porque aunque los años pasan y hemos estado muy cerca y muy lejos, cada vez que hablamos y nos vemos es como si el tiempo no hubiera pasado.

Gracias a mis amigos de la Universidad por las aventuras que compartimos durante la carrera y las que hemos compartido más tarde. A **Marina** porque fuiste tu la primera en hablarme de Miguel y su laboratorio. **Kike**, gracias por demostrar que con esfuerzo se puede. **Leti**, gracias por tu alegría y por mantenernos unidos. **Marta**, siempre dispuesta a quedar y a hablar de la vida. **Ana**, mi primera compañera de piso, gracias por las peripecias que pasamos juntas fuera y dentro del labo. **Bea Sánchez**, nunca olvidaré nuestras conversaciones científicas, pero especialmente nunca olvidaré aquellas sobre la vida.

Gracias a toda mi familia porque vosotros habéis sido la pieza clave de que haya llegado hasta aquí, siempre habéis creído en mi y en que este momento y esta tesis llegarían.

Especialmente, quiero dar las gracias a mis padres **José y Tere**, y a mi hermano **David**, porque gracias a vosotros, a vuestro cariño y educación he podido llegar hasta aquí. Gracias por comprender mis momentos de malhumor y por vuestra paciencia, por seguirme hasta el otro lado del charco y estar siempre preocupados porque sea feliz. Os quiero.

Gracias a los amigos de la familia de toda la vida. A **Paco** por entenderme y crear la fantástica portada de esta tesis y **Mari Carmen** por ser una luchadora y un ejemplo a seguir.

Gracias a mis tíos (**José y Beni**) por preocuparos por mi. A mi abuela **Josefa**, que siempre preguntará por mi, más que yo por ella. Gracias a mi prima **Carmen**, por estar siempre dispuesta a ayudarme de todas las maneras posibles y junto con **Ernesto** dispuestos a asesorar en la organización de un buen viaje. Gracias a **Maria José y Ben** por vuestro cariño, por todos esos veranos de inglés intensivo con vosotros. Gracias a mi tía **Leni**, por estar siempre preocupada por mi.

Gracias a todos los miembros de Famileando: mi abuelo, **Antonio**, por querer mantener unidos a todos los miembros; **Satur**, la líder del grupo y para mi una abuela más; **Celes** (siempre optimista) y **Tere** (mi bióloga favorita y un ejemplo a seguir); **Nacho** (más que un primo eres ya un hermano para mi) y **Javi** (gracias por hacernos pensar); los franceses **María** (la primera doctora de la familia), **Fran** (estoy segura de que eres un gran profesor e investigador, gracias por tus consejos), **Eva y Diego** (unos primos fantásticos y divertidos); los valencianos, **Pío y Mónica** (gracias por ser como sois); los ginebrinos, **Alex** (gracias por tu esfuerzo por mantener a este grupo siempre unido y por hacernos viajar) y **David** (eres uno más de la familia y lo sabes), a los dos gracias por vuestra ayuda estos años y por lo que está por venir. Gracias, porque todos sois maravillosos y aunque nos puedan separar muchos km siempre estamos juntos.

Por último, y no por ello menos importante, gracias **Amadeo**, por ser como eres, por complementarme, por intentar entenderme y ayudarme, por tu paciencia, especialmente en estos últimos meses. Gracias por seguirme hasta el confín de mundo, sabes que siempre estoy deseando que llegue nuestro próximo viaje. Mwen éméw.

A todos vosotros mi más sincero:

**Gracias!**

**Thank you!**

**Grazie!**

**Merci!**



## SUMMARY

---

*DNA provides the music. Our cells and the environment provide the orchestra.*

J. Craig Venter



## SUMMARY

Atrial Fibrillation (AF) is the most common cardiac arrhythmia, affecting over 33 million people worldwide. However, its underlying molecular mechanisms and genetic networks are still poorly understood. Recently, Genome Wide Association Studies (GWAS) have identified several polymorphic variants linked to an increased risk of AF. Almost all such variants are located in non-coding regions, which led us to hypothesize that they could reside in functional elements that interact and regulate genes involved in the origin and development of AF. To test that hypothesis, we have performed *in silico* screening of AF associated *loci* toward selecting candidate regulatory elements to study the physical interactions that they establish with neighbouring regions by Circular Chromosome Conformation Capture followed by deep sequencing (4C-seq). From all selected regions, 7q31 *locus* contains a potential regulatory element inside intron 2 of *CAV1* that through chromatin folding could establish long-range interactions with *CAV1*, *TES*, *MET*, and *CAPZA2*. Deletion of this candidate regulatory element by CRISPR/Cas9 mediated genome-editing in human induced pluripotent stem cell derived cardiomyocytes (hiPSC-CMs) leads to a reduction of *CAV1* expression. This finding could be very relevant for AF since *CAV1* has a key role on the formation of caveolae, membrane invaginations with multiple roles as holding ion-channels involved in action potential generation. Moreover, we show that in the mouse *Cav1* is expressed in the atrial myocardium from early stages of cardiac development up to adulthood.

To further study the molecular mechanisms behind AF, we have also used a well-characterized sheep model of persistent AF. This model mimics AF progression from paroxysmal episodes to persistent and long-standing persistent AF as it occurs in many patients. Transcriptome analysis of the posterior left atrial wall evidences a strong structural remodelling and suggests that during AF development there could be a remodelling of the innervation of the autonomous nervous system. Additionally, gene expression profile of isolated cardiomyocytes from the left and right atria have revealed that, during AF progression, there is a significant alteration of gene transcription regulation that can lead to cardiomyocyte dedifferentiation. Protein expression profile of those cell populations unveils an overexpression of the subunits of the respiratory chain complexes that can produce an increase in oxidative stress that could promote AF development. Furthermore, we find that changes identified in long-standing persistent sheep (one year in AF) are already present only one week of self-sustained AF. All these results contribute to a better understanding of AF progression.



## RESUMEN

---

*La originalidad consiste en el retorno al origen, así pues,  
original es aquello que vuelve  
a la simplicidad de las primeras soluciones.*

Antoni Gaudí



## RESUMEN

La Fibrilación Auricular (FA) es una de las arritmias más comunes, afectando a unos 33 millones de personas en todo el mundo. A pesar de su relevancia, los mecanismos moleculares subyacentes no son muy conocidos. Recientemente, estudios de asociación del genoma completo (GWAS, por sus siglas en inglés) han identificado múltiples variantes localizadas en regiones no codificantes del genoma. De modo que, es posible que dichas variantes estén incluidas en regiones que interaccionan y regulan genes implicados en el origen y desarrollo de la FA. Para testar esta hipótesis, hemos llevado a cabo un estudio *in silico* de varios *loci* asociados a FA y tras seleccionar regiones candidatas a ser elementos reguladores hemos testado las interacciones físicas que establecen con regiones circundantes del genoma, mediante un ensayo de la conformación circular de la cromatina seguido de secuenciación masiva. De todas las regiones estudiadas, el *locus* 7q31 contiene un potencial elemento regulador en el intrón 2 del gen *CAV1* que puede establecer interacciones con el propio gen *CAV1*, así como con los genes *TES*, *MET* y *CAPZA*. La eliminación de esta región mediante editado genético en cardiomiocitos derivados de células humanas pluripotentes produce una reducción en la expresión de *CAV1*. Así mismo, hemos observado en este trabajo como *Cav1* se expresa en el miocardio auricular desde comienzos del desarrollo y permanece en el corazón adulto, pudiendo tener un importante papel en la FA.

Para profundizar en los mecanismos moleculares que subyacen a la FA, también hemos trabajado con un modelo de FA persistente en oveja. Este modelo mimetiza la progresión de la FA desde episodios paroxísticos a episodios de FA persistente, como ocurre en muchos pacientes. El estudio transcripcional de la pared posterior de la aurícula izquierda nos muestra un remodelado de su estructura, y sugiere un posible remodelado de la inervación del sistema nervioso autónomo. Del mismo modo, el estudio de los cambios en la expresión genética de cardiomiocitos aislados de las aurículas izquierda y derecha de este modelo apunta a que durante la progresión de la FA hay una importante alteración de la regulación de la transcripción que lleva a su desdiferenciación. Por otro lado, el estudio de los cambios proteómicos nos revela que, en estas poblaciones celulares, se produce una sobreexpresión de las subunidades que componen los complejos de la cadena respiratoria mitocondrial, lo que puede producir un aumento del estrés oxidativo y promover el desarrollo de la FA. También hemos observado que cambios en la expresión detectados en ovejas con un año de desarrollo de la enfermedad, ya estaban presentes en ovejas con una semana en FA. Así, todos estos resultados contribuyen a un mejor entendimiento de esta enfermedad.





# INDEX

ACKNOWLEDGEMENTS .....	11
SUMMARY.....	19
RESUMEN.....	23
INDEX.....	27
LIST OF FIGURES .....	31
LIST OF TABLES .....	33
LIST OF ACRONYMS .....	35
INTRODUCTION.....	37
What is Atrial Fibrillation? .....	39
Molecular mechanisms underlying AF initiation, maintenance and progression.....	40
The genetic contribution to AF .....	43
Dissecting AF by GWAS.....	44
What is the role of non-coding variants in common human diseases? .....	48
Functional analysis of genomic non-coding elements .....	50
<i>In vitro</i> AF models for studying AF.....	54
<i>In vivo</i> models for studying AF.....	56
OBJECTIVES .....	59
MATERIALS AND METHODS .....	63
1.Circularized Chromosome Conformation Capture: 4C-seq .....	65
1.1 Viewpoint design .....	65
1.2 Sample collection and 4C template preparation.....	66
1.3 Library generation.....	67
1.4 Library sequencing .....	67

1.5 4C-seq data analysis .....	68
<b>2. Mouse analysis.....</b>	<b>69</b>
2.1 Mouse model .....	69
2.2 Embryos and adult heart collection .....	70
2.3 $\beta$ -Galactosidase staining.....	70
2.4 Histological sections .....	71
<b>3. hiPSC culture, differentiation and functional characterization.....</b>	<b>71</b>
3.1 hiPSC culture and differentiation .....	71
3.3 CRISPR-Cas9 deletion .....	71
<b>4. RNA-extraction.....</b>	<b>72</b>
<b>5. RT-qPCR .....</b>	<b>73</b>
<b>6. Western Blot.....</b>	<b>73</b>
<b>7. Statistics.....</b>	<b>74</b>
<b>8. AF induced Sheep model.....</b>	<b>74</b>
8.1 Experimental animals.....	74
8.2 Heart removal and tissue isolation.....	75
8.3 Cardiomyocytes isolation .....	75
8.4 RNA-seq.....	76
8.5 LC-MS/MS proteomics.....	76
<b>9 Immunofluorescence staining.....</b>	<b>76</b>
<b>10. Imaging.....</b>	<b>77</b>
<b>11 Mitochondrial functional assays .....</b>	<b>77</b>
11.1 Blue native gel electrophoresis.....	77
11.2 Enzymatic in gel activity determination.....	78
<b>RESULTS .....</b>	<b>79</b>
<b>1. Regulatory region screening on AF associated <i>loci</i> .....</b>	<b>81</b>
1.1 Chromatin structure profile of the 4q25 <i>locus</i> .....	81
1.2. Analysis of the genomic landscapes of AF associated <i>loci</i> through 4C-seq assay...	83
<b>2. <i>Cav1</i> is expressed during heart development and in the adult.....</b>	<b>88</b>

<b>3. Functional characterization of <i>CAV1</i> hiPSC-CMs.....</b>	<b>91</b>
3.1 <i>CAV1</i> is mainly expressed in atrial-like hiPSC-CMs.....	91
3.2 Deletion of 7q31 AF-associated variants in hiPSCs.....	92
3.3 Effect of the deletion of the candidate <i>CAV1</i> regulatory region in hiPSC-CMs and non-cardiomyocyte cells.....	94
3.4 hiPSC-CMs deleted for the candidate AF-related regulatory region have a reduction in <i>CAV1</i> protein expression .....	96
<b>4. Understanding AF disease progression in a sheep model of persistent AF .....</b>	<b>97</b>
4.1 Posterior Left Atria genomic expression profile during AF progression.....	97
4.2 Transcriptomic and proteomic profiling of isolated atrial cardiomyocytes during AF progression.....	102
<b>DISCUSSION .....</b>	<b>109</b>
<b>1. Understanding genotype modulation in common human traits and diseases...</b>	<b>111</b>
<b>2. The 3D genome structure of AF associated <i>loci</i> .....</b>	<b>111</b>
<b>3. <i>CAV1</i> is expressed in the mouse heart from embryos to adults .....</b>	<b>115</b>
<b>4. Potential regulation of <i>CAV1</i> expression in human cardiac cells by a novel regulatory element .....</b>	<b>116</b>
<b>5. Underlying molecular mechanisms changes during AF progression.....</b>	<b>117</b>
5.1 Studying PLA structural and neural remodelling during AF progression .....	118
5.2 Molecular mechanisms altered in LA and RA cardiomyocytes during AF progression. ....	118
<b>6. GWAS associated <i>loci</i> modulation in AF progression.....</b>	<b>120</b>
<b>CONCLUSIONS .....</b>	<b>123</b>
<b>CONCLUSIONES.....</b>	<b>127</b>
<b>BIBLIOGRAPHY .....</b>	<b>131</b>
<b>APPENDIXES.....</b>	<b>149</b>
<b>PUBLICATIONS.....</b>	<b>169</b>



## LIST OF FIGURES

Figure 1: Understanding normal and abnormal electrical activity of the heart .....	39
Figure 2: Schematic representation of normal and abnormal atrial AP .....	41
Figure 3: Manhattan plot from AF GWAS meta-analysis .....	46
Figure 4: Chromatin 3D organization and regulation of the genome .....	49
Figure 5: Interaction between a regulatory elements and gene promoters .....	50
Figure 6: Chromosome Conformation Capture Technologies.....	54
Figure 7: Viewpoint design requirements .....	65
Figure 8: Cav1 LacZ cassette .....	69
Figure 9: Chromatin interaction profiles in 4q25 mouse syntenic region.....	82
Figure 10: Chromatin interaction profiles in 1q24 syntenic region of mouse.....	84
Figure 11: Chromatin interaction profiles in 16q22 syntenic region of mouse.....	85
Figure 12: Chromatin interaction profiles in 1q21 syntenic region of mouse.....	85
Figure 13: Chromatin interaction profile in 14q23 syntenic region of mouse .....	86
Figure 14: Chromatin interaction profiles in 7q31 syntenic region of mouse.....	87
Figure 15: Hi-C promoter interaction map of 7q31 <i>locus</i> in human cells.....	87
Figure 16: Expression of the LacZ reporter under the control of the <i>Cav1</i> promoter in whole embryos .....	89
Figure 17: <i>Cav1</i> LacZ reporter expression during mouse heart development.....	89
Figure 18: <i>Cav1LacZ/+</i> E18.5 mouse embryo heart sections.....	90
Figure 19: <i>Cav1</i> LacZ reporter expression in postnatal and adult hearts .....	91
Figure 20: Correlations <i>CAV1</i> expression with atrial and ventricular markers in hiPSC-CMs.....	92
Figure 21: Experimental design for CRISPR/Cas9 deletion at the <i>locus</i> 7q31 .....	93
Figure 22: Genotyping of deleted hiPSC clones.....	93
Figure 23: Selected clones' sequences and successful differentiation to cardiomyocytes.....	94
Figure 24: 7q31 <i>locus</i> gene expression in hiPSCs, hiPSC-CMs and Non- cardiomyocytes populations.....	95
Figure 25: Western Blot shows a decrease in <i>CAV1</i> in hiPSC-CMs deleted clones	96
Figure 26: An induced sheep model for the progression of AF.....	97
Figure 27: PLA RNA-seq differentially expressed genes analysis.....	99
Figure 28: Gene Ontology analysis of PLA RNA-seq Differentially Expressed Genes .....	100

Figure 29: Adrenergic and cholinergic study on PLA tissue during AF development .....	101
Figure 30: Principal Component Analysis isolated cardiomyocytes RNA-seq .....	102
Figure 31: Differentially Expressed Genes and Proteins on isolated cardiomyocytes .....	103
Figure 32: Correlations of genes and proteins differentially expressed during AF progression .....	104
Figure 33: Gene Ontology of genes and proteins differentially expressed in LA and RA isolated cardiomyocytes .....	105
Figure 34: Changes to mitochondrial supercomplexes during AF progression in isolated atrial cardiomyocytes .....	107
Figure 35: Genomic interaction landscape, epigenetic marks and transcriptional regulation of the 7q31 syntenic region in the mouse .....	114

## LIST OF TABLES

Table 1: GWAS meta-analysis associated variants .....	47
Table 2: 4C-seq viewpoint primers .....	66
Table 3: Genotyping primers for <i>Cav1LacZ</i> mouse line alleles .....	69
Table 4: PCR conditions for the <i>Cav1LacZ</i> mouse line genotyping.....	69
Table 5: Summary of total embryos collected .....	70
Table 6: gRNAs using for CRISPR/Cas9 mediated deletion .....	71
Table 7: Primer sequence for genotyping the deletion.....	72
Table 8: Primers sequences used for RT-qPCR assays.....	73
Table 9: Antibodies used in WB assay.....	74
Table 10: Antibodies used on Immunofluorescence assay .....	77
Table 11: Antibodies used for Blue Native assays .....	78
Table 12: GWAS AF associated genes differentially expressed in AF induced model .....	121





## LIST OF ACRONYMS

<b>3C</b>	Chromosome conformation capture
<b>3D</b>	Three-dimensional
<b>4C</b>	Circular chromosome conformation capture
<b>4C-seq</b>	Circular chromosome conformation capture with high-throughput sequencing
<b>5C</b>	Chromosome conformation capture carbon copy
<b>AF</b>	Atrial fibrillation
<b>ANS</b>	Autonomic nervous system
<b>AP</b>	Action potential
<b>APD</b>	Action potential duration
<b>AV</b>	Atrioventricular
<b>bp</b>	Base pairs
<b>C LA vs C RA</b>	Chronic left versus chronic right
<b>C vs S</b>	Chronic versus sham
<b>C vs T</b>	Chronic versus transition
<b>ChAT</b>	Choline acetyltransferase
<b>ChIA-PET</b>	Chromatin interaction analysis with paired-end tag
<b>ChIP</b>	Chromatin immunoprecipitation
<b>ChIP-loop</b>	Chromatin immunoprecipitation-combined loop
<b>CRISPR</b>	Clustered regularly interspaced short palindromic repeats
<b>CVD</b>	Cardiovascular diseases
<b>DAD</b>	Delay afterdepolarization
<b>DEG</b>	Differentially Expressed gene
<b>DEP</b>	Differentially Expressed protein
<b>dpc</b>	Days post coitum
<b>DSB</b>	Double strand break
<b>E</b>	Embryonic day
<b>EAD</b>	Early afterdepolarization
<b>eQTL</b>	Expression quantitative trait <i>loci</i>
<b>EU</b>	European Union
<b>FA</b>	Fibrilación auricular
<b>FH</b>	Foetal heart
<b>FISH</b>	Fluorescent in situ hybridization
<b>GFP</b>	Green fluorescent protein
<b>GO</b>	Gene ontology
<b>GWAS</b>	Genome wide association studies
<b>Hi-C</b>	Chromosome conformation coupled with next-generation sequencing
<b>hiPSC</b>	Human induced pluripotent stem cell
<b>hPSC-CMs</b>	Human pluripotent stem cells derived cardiomyocytes
<b>iPSCs</b>	Induced pluripotent stem cells
<b>Kb</b>	Kilobase

<b>LA</b>	Left atrium
<b>LADs</b>	Lamina-associated domains
<b>LD</b>	Linkage disequilibrium
<b>LV</b>	Left ventricle
<b>Mb</b>	Megabase
<b>MPRAs</b>	Massively parallel reporter assays
<b>NADs</b>	Nucleolar-associated domains
<b>NGS</b>	Next generation sequencing
<b>OXPHOS</b>	Oxidative phosphorylation system
<b>PCA</b>	Principal component analysis
<b>PCHi-C</b>	Promoter capture Hi-C
<b>PLA</b>	Posterior left atria
<b>PVDF</b>	Polyvinylidene difluoride
<b>RA</b>	Right atrium
<b>RE</b>	Restriction enzyme
<b>RFP</b>	Red fluorescent protein
<b>ROS</b>	Reactive oxygen species
<b>RV</b>	Right ventricle
<b>S LA vs S RA</b>	Sham left versus sham right
<b>SA</b>	Sinoatrial
<b>SCs</b>	Supercomplexes
<b>SNP</b>	Single nucleotide polymorphism
<b>SR</b>	Sarcoplasmic reticulum
<b>T LA vs T RA</b>	Transition left versus transition right
<b>T vs S</b>	Transition versus sham
<b>TAD</b>	Topologically associated domain
<b>TALEN</b>	Transcription activator-like effector nucleases
<b>TF</b>	Transcription factors
<b>TFBS</b>	Transcription factor-binding site
<b>TH</b>	Tyrosine hydroxylase
<b>TSS</b>	Transcription start site
<b>US</b>	United States
<b>UTR</b>	Untranslated region
<b>w</b>	Weeks
<b>WB</b>	Western blot
<b>WT</b>	Wildtype
<b>ZFN</b>	Zinc finger nuclease

## INTRODUCTION

---

*Thinking always ahead, thinking always of trying to do more,  
brings a state of mind in which nothing is impossible.*

Henry Ford

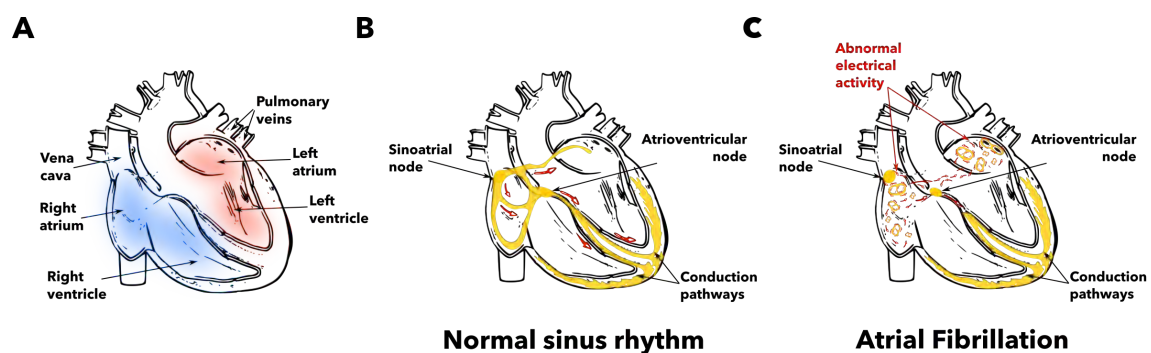


**Cardiovascular diseases (CVD)** are the main cause of morbidity and death around the world, even more than cancer, representing 31% of all global deaths. Currently, their incidence and prevalence are increasing. For example, in Europe there are more than 11 million new cases every year. It is estimated that CVD total cost to the European Union (EU) economy is €210 billion a year and \$329.7 billion in the United States (US)<sup>1,2</sup>.

Among CVD, cardiac arrhythmias, that are electrical disorders of the heart, strongly contribute to this financial burden. Between all of them, **Atrial Fibrillation (AF)** is the most common sustained arrhythmia<sup>3</sup>. AF has a prevalence of 3% of the population and is a major cause of sudden death, heart failure, stroke and cardiovascular morbidity<sup>4</sup>. This percentage can increase up to 5% or even more with age<sup>3,5</sup> and its prevalence and incidence has been progressively increasing worldwide, affecting nowadays more than 30 million people<sup>6,7</sup>.

### What is Atrial Fibrillation?

A normal heart beat starts at the endogenous peacemaker, the sinoatrial (SA) node. It is located at the top right region of the right atrium, at the junction with the vena cava. Sinus node cells initiate an electrical wave that is spread through the right (RA) and left (LA) atria making them to contract and pump blood into the ventricles. This electrical signal also spreads to the atrioventricular (AV) node, where it is delayed before being propagated across the ventricles to make them contract and pump blood to the body (Figure 1 A-B)<sup>8</sup>.



**Figure 1: Understanding normal and abnormal electrical activity of the heart**

(A) Main heart structures. (B) Normal heart electrical activity starting at the SA node.

(C) Uncoordinated electrical activity at right and left atria causes AF.

However, sometimes the heart does not follow this organized pattern of electrical activity. AF is defined as a supraventricular tachyarrhythmia<sup>9</sup>. That means that uncoordinated atrial electrical activity starts causing fast and irregular atria contraction (fibrillation).

This occurs when abnormal electrical signals appear at different parts of the atria, aside from SA node, thus being unable to follow the pacemaker instructions (**Figure 1C**). This fast beating causes a deterioration of atrial function due to a heart rate that can reach 400-600 beats per minute, as compared to a normal resting heart rate (sinus rhythm) that is 60-80 beats per minute<sup>10</sup>.

Atrial structural and electrical changes in AF patients do not appear suddenly, neither in the same way in all the patients. AF is a progressive disorder that usually starts as short and infrequent episodes. Nevertheless, in many patients it evolves to longer and more frequent events. Depending on its pattern, AF is classified as **paroxysmal AF** (short episodes that end spontaneously or with intervention, lasting less than 7 days), **early persistent AF** (continuous episodes that last more than 7 days and less than 3 months), **persistent AF** (lasting longer than 7 days and that does not stop spontaneously), **long-standing persistent AF** (continuous AF episode extending longer than 12-months), and **permanent AF** (present all the time and sinus rhythm is not restored)<sup>4,11</sup>.

Understanding this progression and variability between patients is of clear clinical interest but has been challenging due to our lack of knowledge of the underlying mechanisms and the pathophysiology behind AF.

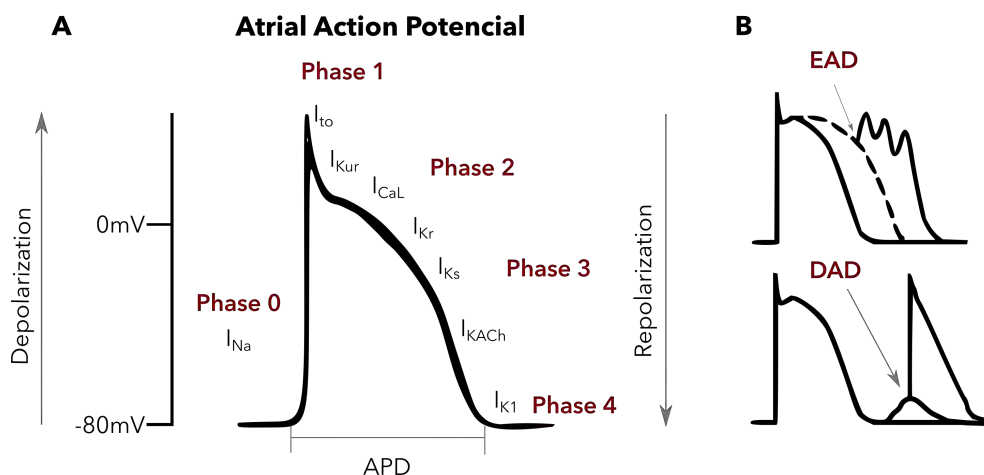
### **Molecular mechanisms underlying AF initiation, maintenance and progression**

To fathom the complexity of the mechanisms behind the initiation, maintenance and progression of AF understanding the normal atrial electrical activity of the cells is needed. Activity of cardiomyocytes is measured by their action potential (AP), that is the electrical potential difference between the inside and the outside of the cell, as a result of ion exchange<sup>12</sup>. The cardiac AP is divided on five phases<sup>8</sup> (**Figure 2A**):

- **Phase 0.** AP, in normal sinus rhythm atrial cells, starts with an activation of Na<sup>+</sup> channels (encoded mainly by *SCN5A* gene), that originates a current ( $I_{Na}$ ) that cause Na<sup>+</sup> ions to move into the cell, making it more positive intracellularly and causing its **depolarization**. Subsequently, an activation of L-type Ca<sup>2+</sup> current ( $I_{CaL}$ ) produces Ca<sup>2+</sup> entrance into the cells. (L-type Ca<sup>2+</sup> channels are encoded by *CACNA1C*, *CACNB2*).
- **Phase 1.** Na<sup>+</sup> channels are rapidly inactivated to favour a **rapid repolarization** together with the activation of two K<sup>+</sup> channels that generate the transient-outward K<sup>+</sup> current ( $I_{to}$ ) (encoded by *KCND2*, *KCND3* and *KCNA4* genes) and ultra-

rapid delayed-rectifier  $K^+$  current ( $I_{kur}$ ) (encoded by *KCNA5* gene) to promote  $K^+$  moving out of the cell, making the inside/outside difference more negative.

- **Phase 2.** It is also called “plateau” because of a reduction on the repolarization velocity, causing AP prolongation what determines AP duration (APD). This is produced by a fine equilibrium between the ions that are entering inside the cell by  $I_{Na}$  (it is inactivated slowly),  $I_{CaL}$  (it was activated at phase 0) currents and the  $K^+$  ions that are leaving the cell by  $I_{kur}$ , rapid delayed-rectifier  $K^+$  current ( $I_{kr}$ ) (channels of this current are encoded by *KCNQ1*, *KCNE1*, *KCNE2* genes) and slow delayed-rectifier  $K^+$  current ( $I_{ks}$ ) (encoded by *KCNH2* gene). At this point,  $Ca^{2+}$  entrance into the cells, generates  $Ca^{2+}$  release from the sarcoplasmic reticulum (SR) through the ryanodine receptor channel type 2 (RYP2). The levels of intracellular  $Ca^{2+}$  activates cytoskeleton filaments to promote cell contraction.
- **Phase 3.** A rapid repolarization occurs. L-type  $Ca^{2+}$  and  $Na^+$  channels are closed, meanwhile some  $K^+$  currents are still active, like  $I_{kr}$  and  $I_{ks}$ , and also new ones are open such as acetylcholine-dependent  $K^+$  current ( $I_{KACH}$ ).
- **Phase 4.** It is the period between two AP, also called refractory period. It starts when the cell recovers its (resting) membrane potential, and the inward-rectifier  $K^+$  current ( $I_{K1}$ ) is responsible of its maintenance (this channel is encoded by *KCNJ2*). However,  $Na^+/Ca^{2+}$  exchanger (NCX) and  $Na^+/K^+$  ATPase pump are also needed for  $Ca^{2+}$  extrusion and ion balance. In addition,  $Ca^{2+}$  ATPase type2 (SERCA2) takes  $Ca^{2+}$  back into the SR. All these processes allow atria relaxation.



**Figure 2: Schematic representation of normal and abnormal atrial AP**

**A)** Representation of a normal atrial AP with the main ionic currents that are involved during all the phases of the depolarization and repolarization of the cell. **B)** Types of electrical ectopic activity. Early afterdepolarization (EAD), the result of an AP prolongation and delay afterdepolarization (DAD) generated by an abnormal depolarization.

During AF this normal electrical activity is altered. Hence the main mechanisms involved in the AF **initiation** are related with atrial electrical dysfunction. Both re-entry (circulation of a cardiac impulse leading a continuous excitation and dissemination of the impulse) and focal or ectopic activity in the atria have been suggested<sup>13</sup>.

**Re-entry** usually happens when conduction is slow. That can happen when  $I_{Na}$  is reduced, when there is abnormal cardiac cell coupling due to problems with connexin protein function, or when there is fibrotic structural remodelling of atrial tissue. Also, re-entry can occur when the refractory period, which is determined by APD, is short, which can be caused by a reduction on the inward  $Ca^{2+}$  current or an increase in the outward  $K^+$  current<sup>12</sup>.

**Ectopic (trigger) activity** is likely the result of **early afterdepolarizations (EADs)** and **delay afterdepolarizations (DADs)**. EADs are usually the result of a prolonged AP, that allows  $Ca^{2+}$  current to be active again generating a new depolarization before the cell returns to the refractory period (**Figure 2B top**). This is caused by abnormalities in ion channels, mainly gain of function of genes coding for ion-channels that mediates inward  $Na^+$  and  $Ca^{2+}$  currents, or loss of function of  $K^+$  channels coding genes responsible of outward  $K^+$  current. DADs are promoted by abnormal  $Ca^{2+}$  handling that generates irregular intracellular release during the refractory period causing an abnormal depolarization and a rapid abnormal beat<sup>12</sup> (**Figure 2B bottom**).

Once AF is initiated, the mechanisms involved in its **maintenance** are less clear. Initially, Moe and Abildskov (1959)<sup>14</sup> proposed the multiple wavelet hypothesis, that considers that AF is maintained by multiple coexisting electrical waves that are propagated randomly through the atria. Nowadays, there are two main theories that attempt to explain AF maintenance. On one hand, Jalife and colleagues have proposed the idea that AF wave propagation is not random and it is sustained by a small number of localized re-entrance sources called rotors, that generates spiral waves<sup>15</sup>. On the other hand, de Groot and colleagues have postulated that AF depends on an endo-epicardial electrical dissociation that promotes complex bidirectional conduction between two layers of the atrial walls<sup>16</sup>. While rotors have been identified in patients, the debate is still open whether both models may even coexist<sup>17</sup>. Therefore, more experimental data is needed to fully comprehend AF maintenance.

If understanding how AF maintains itself has been difficult, deciphering the underlying players behind AF **progression** has been even more challenging. It has been described that AF progression is a combination of slow and fast signalling pathways that are



activated with different kinetics<sup>18</sup>. Initial players seem to be mitochondrial reactive oxygen species (ROS), which produce a rapid electrical remodelling by inducing a shortening of APD and the refractory period. Also, ROS can promote a Ca<sup>2+</sup> overload (that promotes triggering activity) that together with slower players such as atrial dilatation, activation of profibrotic pathways and inflammatory processes, change gene expression to produce ion channel remodelling, myocyte hypertrophy and fibrosis changes that are all necessary for AF progression.

A deeper knowledge of these mechanisms and of the many targets that remain to be discovered will result in an improvement of AF diagnosis and treatment.

### **The genetic contribution to AF**

AF onset has been associated to various diseases such as valvular disease, coronary artery disease, sleep apnea, hypertension, hyperthyroidism, diabetes, and to many other risk factors such as aging, obesity, sex or smoking<sup>19</sup>. However, in many AF patients no specific risk factor has been found. This suggests that these risk factors are not the only cause of AF initiation and progression opening the door to take a look at genetics.

The first familial cases of AF were described in 1936-43<sup>20,21</sup>. Since then, epidemiological studies such as the Framingham Heart Study and Iceland Study have corroborated a similar pattern of AF heritability (1.85 fold increase risk in offspring<sup>22</sup> and 1.77 fold increase on first degree relatives<sup>23</sup>), suggesting a genetic contribution to AF. Also, in 2009 a study in Danish twins confirmed that AF heritability contributes to AF variability<sup>24</sup>.

Once AF heritability was accepted a new field for research was opened to be explored. First studies to unravel the patterns of AF heredity took advantage of these family studies and **linkage genetics** (that are based on genetic markers located near to each other on a chromosome have a higher probability of being inherited together due to meiotic recombination). Through this approach many *loci* have been identified<sup>25-27</sup>. However, it took some time to describe the first AF causal variant in a family with hereditary persistent AF. In 2003, *KCNQ1*, which encodes for a potassium channel subunit of the repolarising I<sub>Ks</sub> current, was the first gene on to be linked to AF<sup>28</sup>.

Since then, many ion channel genes have been sequenced on AF patients as part of **candidate genes studies** looking for new mutations to explain AF heritability<sup>29</sup>. Ion channel genes were good candidates since they were related with the disruption of the electrical signal and many other arrhythmias. Anyhow, many other genes encoding proteins for gap junctions, transcription factors, muscle contraction, cell and nucleus

structure have been shown to be associated to AF<sup>30</sup>. Through these classical studies more than 30 **rare variants** (Variants with a frequency less than 1% in the population) located in the coding regions of genes have been identified as associated to AF<sup>31</sup>.

However, these studies based on monogenic AF families, isolated AF cases and candidate gene studies, are limited. Identifying the actual causal variant and know how often are these mutations on AF patients is difficult<sup>29</sup>. Therefore, these classical approaches have not allowed to fully appreciate the genetic contribution to AF. Indeed, many of genes that have been tested, have both gain and loss of function mutations related with AF, showing that the genetics behind AF is poorly understood.

### **Dissecting AF by GWAS**

The improvements and reduction in costs of genome sequencing has enabled large-scale studies such as Genome Wide Association Studies (GWAS) to be achieved and light up the way to understand the gene regulatory network underlying AF and other human diseases, by discovering the role of **common variants** (Variants with a frequency higher than 5% in the population).

GWAS are based on genotyping a panel of Single Nucleotide Polymorphisms (SNPs) on hundreds or thousands healthy controls and cases, looking for variants significantly associated to a given trait or disease (threshold higher than  $P=5 \times 10^{-8}$ )<sup>32</sup>.

In 2007, the first GWAS for AF was published, describing a highly associated variant (rs2200733) located in a gene desert (large intergenic region) at **4q25 locus** upstream of the *PITX2* gene<sup>33</sup>. It encodes for the paired-like homeodomain transcription factor 2. *Pitx2* is implicated in the establishment of the left-right axis during heart development<sup>34</sup>. It also has a relevant role in the formation of the SA node and the pulmonary veins<sup>35,36</sup>, which is very interesting due to the role of the SA node on normal electrical activity of the heart and the prevalence of abnormal electrical activity arising from the pulmonary veins in AF patients. Indeed, pulmonary veins isolation through catheter ablation is the regular approach to treat recurrent AF. Also, it is known that *Pitx2* depletion is lethal by causing mouse embryo death due to cardiac malformations<sup>37</sup>. Furthermore, mice with an haploinsufficiency of *Pitx2c*, the main isoform in the heart, have shorter AP<sup>38</sup>, suggesting that *PITX2* could have a role on AF susceptibility.

Subsequent GWAS with higher number of cases and controls found also associations on this *locus* (rs6843082, rs17042171)<sup>39,40</sup> and on chromosomes 16q22 and 1q21, in the intronic regions of *ZFHX3* (rs2106261 and rs7193343) and *KCNN3* (rs13376333) respectively<sup>39-41</sup>.

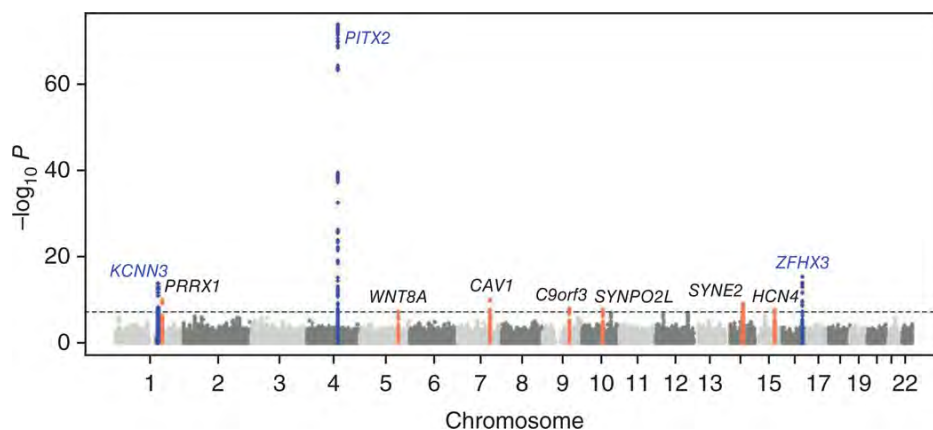
Variants on **16q22 locus** marked *ZFHX3* gene as a candidate for AF causality. *ZFHX3* (formerly known as *ATBF1*) encodes Zinc Finger Homeobox 3, that is a transcription factor involved in the regulation of myogenic and neural differentiation<sup>42</sup>. Over-expression of *ZFHX3* produces complexes with the protein inhibitor of activated STAT3 (PIAS3) and represses signalling by the signal transducer and activator of transcription 3 (STAT3)<sup>43</sup>. In tachypaced HL-1 cells, *ZFHX3* and PIAS3 expression are decreased, meanwhile STAT3 is increased. The same effect has been observed when *ZFHX3* is downregulated<sup>44</sup>, suggesting that an inhibition of *ZFHX3* and an enhanced expression of STAT3 could mediate an inflammatory process that contribute to AF. Also, it has been shown that *ZFHX3* knockdown HL-1 cells have an altered calcium homeostasis that causes an arrhythmogenesis<sup>45</sup>.

Meanwhile variants on **1q21 locus** point *KCNN3*, that encodes for a calcium activated potassium channel, SK3, responsible for membrane repolarization as a putative causal gene. Its role in neurons well known, where it is involved in the late repolarization to decrease the neurons excitability<sup>46</sup>. In the heart it is less widely studied, but recent publication has shown that in a mouse model the SK3 channel contributes to repolarization of AP on atrial myocytes and SK3 overexpression causes a shortening on APD and therefore makes the mouse predisposed to atrial arrhythmias<sup>47</sup>.

To increase the number of associated variants, a combination of data from different GWAS through meta-analysis was needed. Meta-analysis uses the combined genetic information from hundreds of thousand or even millions of individuals. This approach evaluates the consistency or inconsistency of multiple datasets increasing the power of detection<sup>48</sup>. The first AF meta-analysis was conducted by the AFGen Consortium and identified ten AF associated *loci* in individuals of European and Japanese ancestry (**Figure 3**)<sup>49</sup>. Among them are included the three previously identified *loci*, being the variant (rs6817105) identify on 4q25 *locus* the most highly associated in the study ( $P=1.8 \times 10^{-74}$ ). **Table 1** summarizes the *loci* associated by this meta-analysis, its lead SNP and the putative causal gene. 5q31 *locus* association was not replicated on independent studies performed by the authors. Despite it can be interesting candidate *locus* since a recent study has showed that *Wnt8a* overexpression damages calcium handling and regulates microRNA signature expression in a *Pitx2* mouse model of loss of function, triggering in that way AF<sup>50</sup>.

Among, the new identified *loci* it is interesting the **1q24 locus** which its associated variant is located in a gene desert similar to the one of 4q25 *locus*. The causal gene

associate to this *locus* *PRRX1*, as *PITX2* and *ZFH3*, also encodes for a transcription factor, the paired related homeobox1. It is highly expressed in the developing heart, mainly in mesenchymal tissues<sup>51</sup>, being involved in epithelial to mesenchymal transition<sup>52</sup>. It is also expressed during development of great vessels<sup>53</sup> and pulmonary veins formation<sup>54</sup>, which suggests a possible role of *PRRX1* on AF onset, as *PITX2* was, due to the abnormal electrical activity arising from the pulmonary veins in AF patients. Very recently, it has been shown that in human embryonic stem cell-derived cardiomyocytes and zebrafish *PRRX1* suppression leads to APD shortening, which is a very common phenotype of AF<sup>55</sup>.



**Figure 3: Manhattan plot from AF GWAS meta-analysis**

This first AF meta-analysis Manhattan plot shows 10 significant AF associated *loci* (P-value higher than  $5 \times 10^{-8}$ ) marked by several SNPs (Figure from Ellinor *et al.* 2012)<sup>49</sup>.

**10q22 locus** is also interesting since it was previously associated to AF by Brugada and colleagues in a familial AF population<sup>25</sup>. The associated variant is located 5kb upstream *SYNPO2L*, Synaptopodin 2 like, gene and 20kb, upstream of *MYOZ1*, Myozenin 1, gene. Both genes express proteins located in skeletal and cardiac muscle, placed at Z-disc interacting with many other proteins making both of them interesting candidates as AF associated genes. A recent eQTL analysis from left and right atria tissue found that *MYOZ1* was most probably the gene related to AF based on this association<sup>56</sup>. Additionally, **15q24 locus** could have a relevant role on AF since the nearest associated gene *HCN4* encodes for the hyperpolarization-activated cyclic nucleotide gated potassium channel, that is in charge of the funny current ( $I_f$ ) and therefore controls the cardiac pacemaker at the SA node. Moreover, it is expressed in the AV node. *HCN4* mutations have been found in families with AF.<sup>57</sup> However, it is still needed to determine if the associated variant produces an abnormal *HCN4* expression or not. In addition, *C9orf3*, located at **9q22 locus**, codes for the protein aminopeptidase O (ONPEP). ONPEP is expressed in the heart and mediates the generation of Angiotensin IV by the

cleavage of Angiotensin III<sup>58</sup>. Hence, it could have a role on AF development through the renin-angiotensin axis, that has been previously related to AF<sup>59</sup>.

<b>Locus</b>	<b>SNP</b>	<b>Candidate causal gene</b>
<b>1q21</b>	rs6666258	<i>KCNN3</i>
<b>1q24</b>	rs3903239	<i>PRRX1</i>
<b>4q25</b>	rs6817105	<i>PITX2</i>
<b>5q31</b>	rs2040862	<i>WNT8A</i>
<b>7q31</b>	rs3807989	<i>CAV1</i>
<b>9q22</b>	rs10821415	<i>C9orf3</i>
<b>10q22</b>	rs10824026	<i>SYNPO2L</i>
<b>14q23</b>	rs1152591	<i>SYNE2</i>
<b>15q24</b>	rs7164883	<i>HCN4</i>
<b>16q22</b>	rs2106261	<i>ZFHX3</i>

**Table 1: GWAS meta-analysis associated variants**

AF associated *loci* by GWAS meta-analysis<sup>49</sup>, putative causal SNPs and candidate genes.

**7q31 locus** associated gene *CAV1* can also have a key role on AF. *CAV1* gene encodes Caveolin-1, a key protein in the formation and maintenance of caveola, an invagination of the plasma membrane which is involved in membrane trafficking, mechanosensing, extracellular matrix remodelling, lipid and cholesterol homeostasis, tissue regeneration and cell signalling and migration<sup>60</sup>. In addition, caveolae hold many ion channels involved in the different phases of the AP<sup>61</sup>. *CAV1* is also be found in lipid droplets, peripheral adhesions, endomembrane system and even mitochondria<sup>62</sup>. Several studies performed in knockout mice showed that these animals have dilated cardiomyopathy and pulmonary hypertension or cardiac hypertrophy<sup>63</sup>. In a similar way, there are evidences that suggest that **14q23 locus** associated gene *SYNE2* could have a potential role on causing arrhythmia, as mutations in this gene occur in some cases of Emery-Dreifuss muscular dystrophy (EDMD)<sup>64</sup>. *SYNE2* encodes Nesprin2, an actin-binding and nuclear-envelop associated protein relevant for the preservation of the nuclear envelop structure<sup>65</sup>. Moreover, Nesprin2 is needed for centrosome-nucleus coupling during neurogenesis and neuronal migration<sup>66</sup>. Therefore, *SYNE2* could have a key role on AF development by uncoupling of the nucleus and cytoskeleton.

Many of this associations has been confirmed by subsequent GWAS and meta-analysis performed in recent years on different populations to AF<sup>67-74</sup>. Nevertheless, in 2018 Roselli et al. conducted a large meta-analysis for AF using more than half a million individuals and identifying 67 new AF associated *loci*<sup>73</sup>. At the same time Nielsen et al. performed a similar scale meta-analysis on more than 1,000,000 individuals, leading to the identification of 80 new *loci*<sup>75</sup>. Therefore, and as of today, more than one-hundred different genomic *loci* have been associated to AF.

Nevertheless, the function of all AF *loci* and their associated variant is still uncertain, particularly since most of them are located in non-coding regions. Moreover, SNPs associated to different traits or complex diseases by GWAS are not necessarily the causal variant of the *locus* where is located, and neither the nearest gene has to be necessarily the target gene of the association<sup>76</sup>. It has been shown on AF 4q25 *locus* and other complex disease as diabetes that mere proximity to a variant does not mean causality or the correct identification of the causal gene<sup>77,78</sup>. Therefore, further studies are needed to decipher their link to AF.

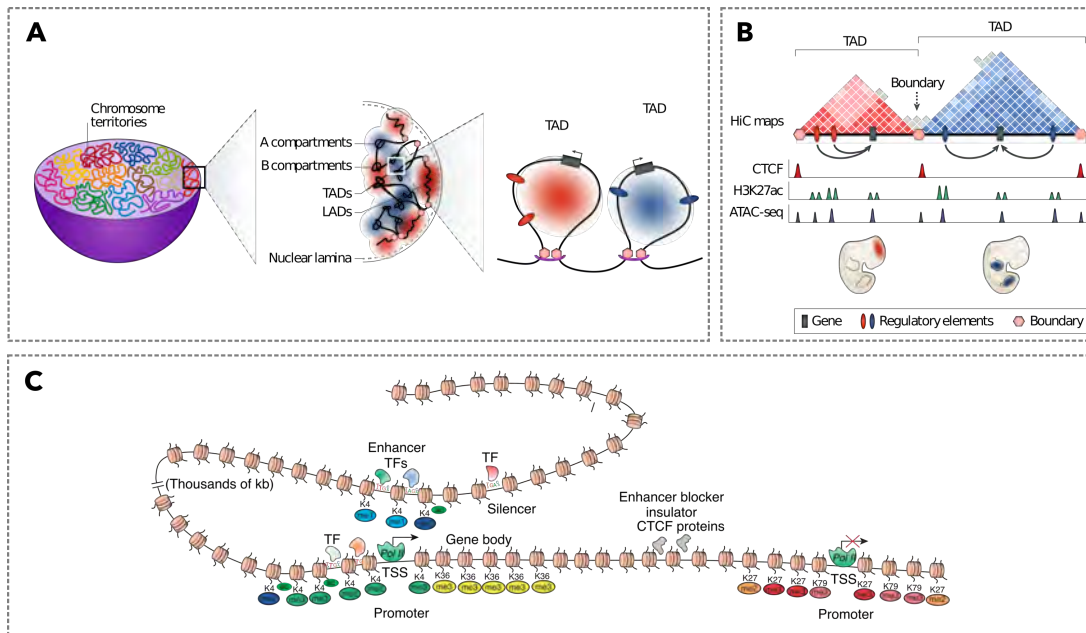
### **What is the role of non-coding variants in common human diseases?**

Non-coding regions represent the 98% of our genome<sup>79</sup> and they contain different genomic elements involved in gene regulation. Gene expression requires a fine regulation which is achieved by a collaborative arrangement of multiple *cis* (within chromosomes) regulatory elements<sup>80</sup>. The basic element that regulates transcription is the **gene promoter**, which is composed by the core promoter and nearby regulatory elements, that are named as **proximal promoter**. The **core promoter** is the DNA region where the transcription preinitiation complex and the transcriptional machinery are assembled, defining the transcriptional start site (TSS). Binding sites for different transcriptional regulators can be located in the immediate vicinity of the promoter<sup>81</sup> (**Figure 4C**).

However, precise regulation of gene expression in time and space is accomplished by a set of regulatory elements, such as insulators, silencers or enhancers (**Figure 4C**). These regulatory elements are located in non-coding regions, that can be at an intronic region, or at any distance upstream or downstream from the target gene's TSS playing their role over long-distances<sup>80</sup>. **Insulators** are DNA elements that isolate a DNA region from the positive or negative regulation of silencers or enhancers that are acting on other genes. **Silencers** are DNA segments where transcription factors (TF) acting as repressors bind to silence or repress a gene promoter. **Enhancers** are DNA regions able to recruit TF that work together to enhance gene expression<sup>81,82</sup>.

In this broad regulatory landscape, it has been shown that enhancer elements can be located more than 1 megabase (Mb) away from their target gene, not affecting nearby genes that are closer along the linear DNA. This is possible by long-range physical chromatin interactions, where DNA looping, mediated by architectural proteins such as CTCF or cohesin, facilitate enhancer-promoter interaction<sup>83</sup>. This reveals a new layer of

regulation, where the three-dimensional structure of the genome will not merely be a packaging instrument, but important for the proper control of gene expression.



**Figure 4: Chromatin 3D organization and regulation of the genome**

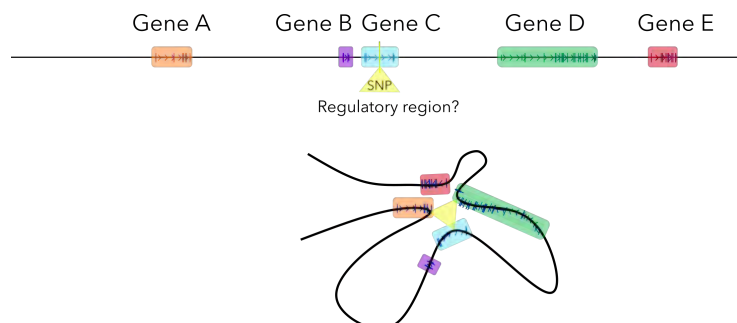
**A** and **B**) Representation of the different layers of DNA organization. **A**) In the nucleus, chromosomes are located in the A, B compartments and within them chromatin is organized LADs and TADs. **B**) TADs contain inside sub-TADs with the last layer of DNA organization the nucleosomes. Sub-TADs organization can vary from one tissue to another (Figure modified from Spielmann et al. 2018)<sup>83</sup>. **C**) Scheme of gene expression regulatory elements. The promoter on the bottom left is activated by a distal enhancer that contains transcription factor binding sites where TFs are bound to regulate the gene expression, also active histone marks helps to the positive regulation. On the other hand, the promoter in the right bottom is not expressed due to the insulator located between the two promoters that avoid the enhancer to activate this second promoter. Near to the enhancer there is a silencer that could compete with it to negatively regulate gene expression (Figure modified from Sakabe and Nobrega 2010)<sup>80</sup>.

At the nuclear level chromosomes are distributed in different territories (**Figure 4A**). Then individually, chromosomes are segregated into different compartments called A (open and transcriptionally active euchromatin) or B (close and silent heterochromatin) compartments. B compartments can be found at the periphery of the nucleus as lamina-associated domains (LADs) or as nucleolar-associated domains (NADs)<sup>83</sup>. Within A and B compartments chromatin is organized in a packaged form called topologically associated domains (TADs), long genomic regions of high interaction between themselves with a reduced interaction at the boundary regions and with neighbouring TADs and other regions on the genome<sup>83,84</sup>. Almost 60-70% of TADs are pretty much constant between cell types and between species, suggesting that these domains are inherent to the mammalian genome<sup>85</sup>. Inside TADs, chromatin is folded into sub-TADs,

which are more variable among tissues (**Figure 4B**). At last, genome is packaged on the nucleosomes establishing the first level of DNA folding<sup>84</sup>.

Nucleosomes are formed by an octamer of histone proteins (H2A, H2B, H3 and H4) around which DNA is wrapped. These histones can be acetylated, methylated or phosphorylated, among other post-transcriptional modifications, introducing an extra layer for regulation of gene expression<sup>80</sup>. The best characterized histone marks for active promoters and enhancers are methylation of lysine 4 of histone 3 (H3K4me3 for promoters and H3K4me1 for enhancers), acetylation of histone 3 lysine 27 (H3K27ac) related to active enhancers and promoters, and methylation of histone 3 lysine 27 (H3K27me3) associated to repressed regions<sup>86,87</sup> (**Figure 4C**).

All these different layers of chromatin organization are cooperating together with regulatory elements to modulate gene expression over long distances. Therefore, non-coding regions have a relevant role in how gene transcription is arranged, in each cell over time and space, and any alteration on them could involve an important change on gene expression. That is why a non-coding variant located inside a regulatory element could affect its activity and consequently could have a role on disease development by changing gene expression regulation (**Figure 5**).



**Figure 5: Interaction between a regulatory elements and gene promoters**

Disease associated variants located in non-coding regions can be inside a regulatory region. Therefore, they may modulate gene expression of neighbouring genes or over long distances genes that are physically relocate closer through DNA folding and looping.

Hence, many SNPs associated to AF could be indicating that, at the position where they are located, there is a regulatory element that controls the expression of one or more genes involved in AF onset, maintenance or progression. To address this important issue, functional studies of those variants and their genomic context are needed to establish the functional effect of any disease associated SNP.

**Functional analysis of genomic non-coding elements**

To address if an associated variant, located in a non-coding region, could have a function, a common approach is first performing an *in silico* analysis<sup>76</sup>. Usually, this



analysis consists on examining available public databases to assess if that region has active or repressive histone marks, known transcription factors binding sites (TFBSs) or if there is any already published data about that region that could give a hint on understanding its potential function of that region.

To empirically assess the function of a potential regulatory region classically it has been addressed by reporter assays. This approach addresses whether a specific non-coding region acts or not as a regulatory element based on placing it together with a minimal promoter and a reporter gene such as green fluorescent protein (GFP), red fluorescent protein (RFP),  $\beta$ -galactosidase or luciferase. Then this construct is assayed for reporter gene activity *in vitro* or *in vivo* to check if it alters its expression<sup>81</sup>. Nowadays, it is possible to test multiple regions at the same time by means of massively reporter assays (MPRAs)<sup>88</sup>.

Once a region with regulatory activity is identified, further assays are required to gain a better knowledge of its function. Using genomic tools, deletions and other mutagenesis strategies can be done to define its function more accurately. These techniques allow to edit a whole regulatory region or simply mutate an individual SNP<sup>76</sup>. First genome editing tools were zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), these chimeric nucleases composed by a programmable sequence-specific DNA-binding module linked to a non-specific DNA cleavage domain, are able to induce DNA double-strand breaks (DSBs) that activate DNA repair mechanisms<sup>89</sup>. Recently the nucleic acid-based clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-based has been developed. In bacteria, the CRISPR/Cas system provides an acquired immunity against foreign DNA via RNA-guided DNA cleavage. However, since 2012 when Doudna and Charpentier first demonstrated that CRISPR/Cas9 could target DNA *in vitro*<sup>90</sup> and one year later its editing capacity was tested *in vivo*<sup>91</sup>, CRISPR/Cas technology has evolved to a system where CRISPR can be targeted easily to almost any genome by a short RNA guide that helps to direct the Cas9 nuclease to the target *locus* and perform DSBs. This genome engineering technology has transformed the genome editing field, empowering researchers to perform almost any type of modification, from performing a small mutation, elucidating the structural organization of the genome or establishing linkage between a genetic variation and a particular phenotype<sup>92</sup>.

While these experiments allow us to study whether a non-coding region acts or not as a regulatory element do not enable to explain how an enhancer plays its functional role

over its target gene. Chromatin folding and looping is the key for enhancer-promoter interactions therefore experiments for studying that three-dimensional (3D) genome structure were needed.

Historically, fluorescent in situ hybridization (FISH) was the main method to study nuclear organization. This cytogenetic approach is based on a genomic DNA complementary nucleotide sequence (probe) that can hybridize and labelling a target DNA region. Then the fluorescently labelled *loci* can be visualized by conventional fluorescence microscopy or superresolution light microscopy<sup>84</sup>. However, this method is limited to a few selected *loci* that are studied in single cells. To have an insight from cell population has been very relevant the development of chromosome conformation capture (3C)<sup>93</sup> techniques and its derivatives (**Figure 6**). These assays rely on quantification of interaction frequencies between DNA segments that are in close spatial proximity in the nucleus independently of their linear genomic distance<sup>83</sup>. Indeed, 3C approaches are based on the idea of that genomic regions that are physically closer in the nucleus are more likely to be captured by cross-linking than those that are far away<sup>84</sup>.

All 3C approaches start by capturing the native physical interactions established by the genome in the cell nucleus by formaldehyde fixation. This is followed by digestion and ligation steps to generate chimeric DNA templates formed by DNA fragments that can be far away on the linear genome but not in its 3D conformation<sup>84,94</sup> (**Figure 6** upper horizontal panel). In 3C, the number of ligation events between non-neighbouring sites is estimated by semiquantitative<sup>93</sup> or quantitative<sup>95</sup> PCR amplification of pre-selected ligation junctions that are part of the chimeric DNA template. In that way, 3C approach is considered a “one versus one” strategy because it allows to look into the interactions between two pre-selected *loci*, to check if there are interactions between them or not<sup>94</sup> (**Figure 6** first vertical panel).

3C has been subsequently modified to develop other conformation capture based methods in which the main differences are the number of genomic regions that are interrogated in parallel and the size of the genomic region that it is explored.

Circular chromosome conformation capture (4C) uses the ligated 3C template for a second round of digestion and ligation to obtain small DNA circles that are amplified by inverse PCR using primers for a selected genomic region (viewpoint). These amplified fragments were initially analysed by microarrays<sup>96</sup>. Nowadays, this is performed by next generation sequencing (NGS) and therefore is called 4C-seq<sup>97</sup> (circular chromosome

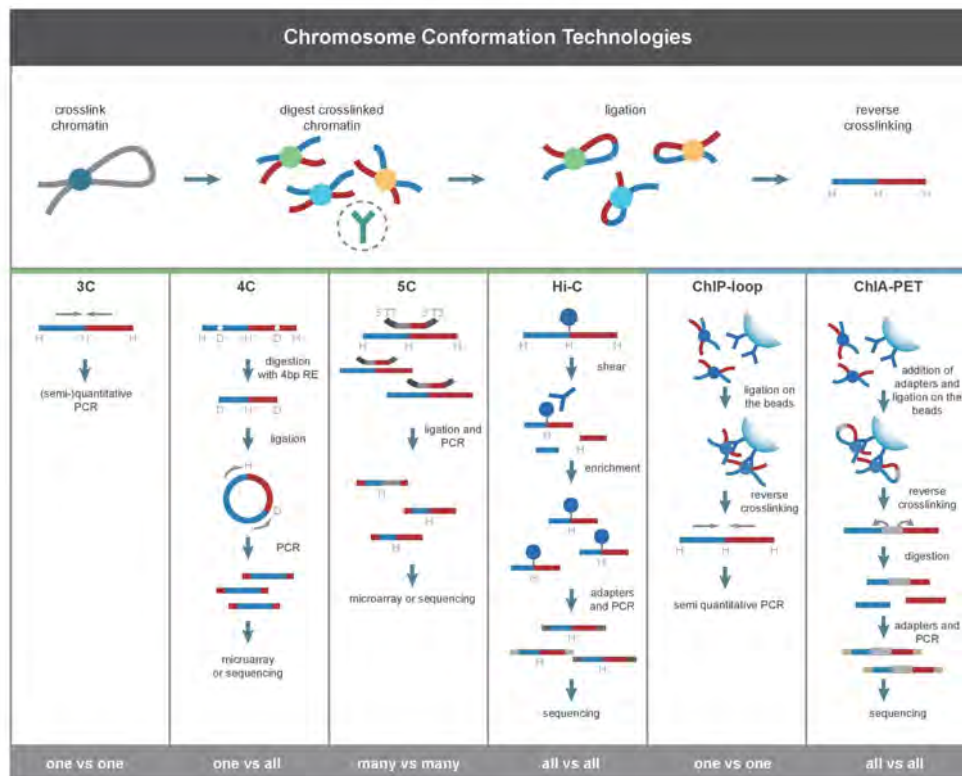
conformation capture with high-throughput sequencing). 4C is known as “one versus all” strategy because a defined viewpoint region is screened for sequences in the whole genome that can be contacting with it<sup>94</sup> (**Figure 6** second vertical panel). 4C allows to answer the question of which regions of the genome are interacting with the region of interest or viewpoint.

In chromosome conformation capture carbon copy (5C) technology the 3C template is hybridized with a mix of oligonucleotides that overlap with different restriction enzyme sites in the genomic region of interest. In that way, pairs of oligonucleotides that match with interacting fragments are juxtaposed to the 3C template being ligated together. Subsequently, they are amplified by multiplex PCR reaction and the readout can be addressed by microarray or by high-throughput sequencing<sup>98</sup>. Hence, 5C is described as “many versus many” technology because it is focused on the capture of all the interactions that occur in a region of the genome of a few megabases in an unbiased manner<sup>99</sup> (**Figure 6** third vertical panel).

Finally, NGS have led to 3C technology to evolve to the chromosome conformation coupled with next-generation sequencing (Hi-C) technology. Hi-C is based on high-throughput sequencing to quantify 3C ligated products to study whole genome 3D architecture. In Hi-C assays, after the restriction enzyme digestion the restrictions ends are filled with biotinylated nucleotides, followed by a blunt end ligation. In that way, purified DNA is pulled-down with biotin to ensure that only ligation junctions are selected<sup>100</sup>. Accordingly, Hi-C is considered “all versus all” method because its resolution extends to the entire genome enabling the detection of all possible cross-linked contacts in the genome<sup>99</sup> (**Figure 6** forth vertical panel).

However, 3C technologies only provide information about 3D chromatin structure and genome contacts, but not about the function itself of these interactions. A new set of techniques have been developed by combining 3C technologies and chromatin immunoprecipitation (ChIP). The first one to be developed was chromatin immunoprecipitation-combined loop (ChIP-loop)<sup>101</sup>. ChIP-loop allows to determine which genomic sites interact and the candidate proteins mediating the interaction<sup>102</sup> (**Figure 6** fifth vertical panel). A few years later chromatin interaction analysis with paired-end tag (ChIA-PET) was developed. ChIA-PET offers the possibility of analysing exclusively the chromatin interactions that are formed between sites bound by a given DNA or chromatin-interacting protein<sup>103</sup> (**Figure 6** last vertical panel). More recently, a new combination of ChIP and Hi-C (HiChIP) have allowed to increase yield of captured

contacts associated with a protein of interest and therefore the specificity of the interactions sequenced after a Hi-C experiment<sup>104</sup>.



**Figure 6: Chromosome Conformation Capture Technologies.**

The first horizontal panel shows the common step in all “C” methods and the vertical panels the specific steps of each method. In the lower panels, each technique is classified depending on the number of regions that are interrogated (3C, 4C, ChIP-loop one, 5C many, Hi-C and ChIA-PET the whole genome) and the number of genomic regions that are explored (3C and ChIP-loop one, 5C many and 4C, Hi-C and ChIA-PET the whole genome). (Figure from Li et al. 2014)<sup>105</sup>

Thus, 3C-based technologies have been crucial to understand how the linear organization of different functional elements in the genome interact resulting in functional 3D regulatory DNA networks<sup>94</sup>.

### ***In vitro* AF models for studying AF**

To elucidate the underlying mechanism of AF and linking AF associated variants to its phenotype, functional assays are needed. In the cardiovascular field accessing to patient material for gathering more knowledge it is limited, since it can only be obtained by invasive procedures or post-mortem analysis<sup>106</sup>.

The generation of induced pluripotent stem cells (iPSCs) by Takahashi and Yamanaka in 2006 by TF mediated reprogramming<sup>107</sup> opened a new field of research. One year later, it was possible to make human induced pluripotent stem cells (hiPSCs)<sup>108,109</sup> and therefore it become feasible to generate patient-specific hiPSCs. Furthermore, several

years later it was possible to generate hiPSC derived cardiomyocytes (hiPSC-CMs)<sup>110,111</sup>, filling the lack of a tissue culture model to study cardiac diseases and providing a great source to study cardiac mechanisms, drug discovery, cardio toxicity testing and to perform gene targeting assays to test the role of common and rare variants in cardiac cells.

To generate hiPSC-CMs, the first step consists in reprogramming somatic cells from a healthy donor or a patient carrying a genetic mutation to generate hiPSCs. Initially, reprogramming was performed by using retrovirus, that are stably integrated in the genome. However in the last years non-integrative delivery methods have been developed such as adenoviruses, Sendai viruses, episomal vectors, mRNAs or small molecules<sup>112</sup>. Another improvement is that now, hiPSCs can be derived from almost any cell, being blood cells widely used due their easy obtaining, instead of fibroblasts from skin biopsies that were originally used as somatic cells source<sup>106</sup>.

Once, hiPSCs are generated they are pushed to differentiate towards cardiomyocytes by mimicking endogenous developmental signalling. Therefore, several signalling pathways involved in mesoderm induction and cardiac specification, such as activin-Nodal, BMP, Wnt, TGF- $\beta$  and FGF pathways have to be controlled<sup>111</sup>. Different protocols have been published with different permutations of growth factors, cytokines, small molecules and media to reach an efficient hiPSC-CMs production<sup>110,111,113</sup>. However, the result of these differentiation protocols is a heterogeneous population of nodal-, atrial- and ventricular-like cells. Recently published protocols efficiently differentiate hiPSCs towards the ventricular<sup>114</sup> and auricular<sup>115</sup> lineages by altering retinoic acid signals. Nevertheless, there is still a major concern from these protocols in that hiPSC-CMs resemble embryonic or foetal cardiomyocytes not fully differentiated to the adult phenotype. Hence, there is a big interest on the field to improve hiPSC-CMs maturation. In 2016, Herron et al. successfully achieved a rapid maturation of hiPSC-CMs that were able to propagate an electric impulse as fast as  $\approx 50\text{cm/s}$ , similar to the velocity of an adult cardiac ventricle, by plating them on a softer substrate than the regular culture cell plates<sup>116</sup>. Nowadays, other laboratories have improved this maturation process by electromechanical conditioning of early-stage hiPSC-derived cardiomyocytes with increasing intensity over time<sup>117</sup> or using three-dimensional culture systems.

Even though a greater understanding of molecular mechanism underlying the cardiomyocyte differentiation and maturation is needed, hiPSC-CMs are a very useful tool together with genome editing approaches to improve the knowledge between a

genotype (e.g. a disease associated variant) and its phenotype (e.g. a disease). Thereby, patient-specific hiPSC-CMs can be generated for testing the ability of drugs to treat the phenotype and study the molecular mechanism in a personalized context. Even more, despite different genetic backgrounds between cell lines generated from different patients, nowadays genome editing tools such as ZFNs, TALENs and CRISPR/Cas9 allow to create isogenic (essentially identical genotype) control cell lines by replacing the mutated allele by the healthy one<sup>106</sup>. Although, this approach may seem only applicable to monogenic diseases, hiPSC-CMs are also an excellent model to explore the role of SNPs and other variants located in non-coding region.

Although, development of hiPSC-CMs have a great potential, it still remains to be proven if they are able to recapitulate complex diseases.

### ***In vivo* models for studying AF**

Since the access to AF clinical samples is difficult, the development of robust animal models would provide full access to cells and tissue from different heart regions and cell types. *In vivo* preclinical models also provide a source for following up the disease and studying the underlying mechanisms in a controlled experimental setting and without confounding factors.

In contrast to other diseases, an appropriate AF mouse model is difficult to generate since the mice heart rate is up to ten times faster than the human heart and their APD is shorter and lacks the plateau phase<sup>118</sup>. Nevertheless, there is a high conservation of the cardiac developmental pathways, morphological structures and signalling pathways between mouse and human. This makes mice still a useful model to understand genetic components of AF by allowing dissection of the molecular mechanism of action of genes and proteins linked to AF. In this regard, in 2016 a mouse model was described where adult-specific *Tbx5* deletion (a transcription factor associated to AF by GWAS) causes a rapid onset of arrhythmic phenotypes by disrupting expression of AF-susceptibility genes such as *Scn5a*, *Gja1*, *Ryr2*, *Dsp* and *Atp2a2*. Moreover, the authors showed that reduced *Pitx2* expression (another gene associated to AF by GWAS) can rescue this phenotype and that TBX5 drives PITX2 expression<sup>119</sup>. Likewise, many other mice models have been generated to unravel molecular pathways related to AF such as ion channel dynamics, calcium homeostasis, anchoring and junction complexes, transcriptional, post-transcriptional and epigenetic regulation<sup>118</sup>.

Since the heart of large animal more closely resemble the human condition, they have been widely used for understanding the mechanisms and management of AF. Several

species have been used to this aim, such as dog, goat, sheep, pig and rabbit, and there have been many different ways of generating AF by acute atrial insults and in the presence of autonomic nervous system modulation<sup>120</sup>.

In this work, we have used a sheep tachypacing model of long-term atrial fibrillation, generated by Martins et al.<sup>121</sup>. In this model, a pacemaker is implanted over the RA. Upon activation, AF is induced by continuous pacing until self-sustained AF is detected by the loop recorder implanted over the LA and pacemaker activation it is not needed any more. This sheep model has unveiled that AF progression from paroxysmal to persistent states involves structural and electrical remodelling of the atria. Structural changes include atria dilatation, hypertrophy and fibrosis, while electrical remodelling is characterized by ion channel protein expression and ion current changes and by a progressive increase of dominant frequency (DF) (the electrical signal with the highest frequency, number of electrical events per second, compared with its surroundings) in both atria<sup>121,122</sup>.

Nevertheless, in spite of the interesting and physiological results obtained from this large animal models, there is still a lack of understanding of the genetic and molecular mechanisms underlying AF establishment and progression.

In this works, we have used a combination of different approaches, from patients GWAS, to *in vitro* and *in vivo* studies that will enable us to better understand the underlying genetic mechanisms behind AF and will offer an opportunity to validate new therapies and diagnosis.





## OBJECTIVES

---

*All our dreams can come true,  
if we have the courage of pursue them.*

Walt Disney



## OBJECTIVES

Understanding how the molecular and genetic networks behind atrial fibrillation (AF) changes during disease progression and comprehend how they are regulated may contribute to improve diagnosis and treatment of AF patients. To address this general objective, we have established the following specific objectives:

- To study the 3D genomic landscape of AF associated *loci*.
- To functionally characterize candidate regulatory elements that hold AF associated variants.
- To identify transcriptomic changes during AF progression in the posterior left atrial wall of a sheep model of persistent AF.
- To identify transcriptomic and proteomic alterations in the cardiomyocyte cell population of left and right atria of a sheep model of persistent AF.
- To experimentally extend the knowledge of the most relevant transcriptomic and proteomic changes observed in the sheep model of persistent AF.



## MATERIALS AND METHODS

---

*Nothing in life is to be feared, it is only to be understood.  
Now is the time to understand more, so that we may fear less.*

Marie Curie



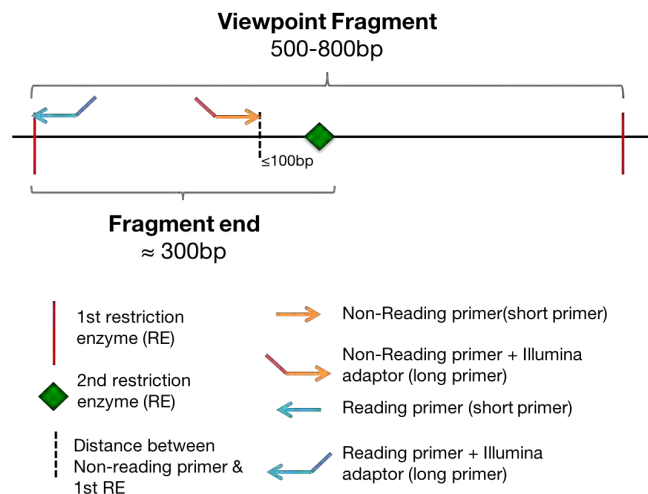
## 1. Circularized Chromosome Conformation Capture: 4C-seq

### 1.1 Viewpoint design

4C technology allows us to identify the interactions that take place between the viewpoint and the rest of the genome. The size of this fragment of interaction is defined by the first restriction enzyme (RE) that is used to digest chromatin. It should have a size of around 500-800bp, to ensure efficiency in the cross-linked step to fix the interactions with other regions. Additionally, the distance between the first and the second enzyme digestion sites should be approximately 300-350bp to improve the ligation step after digestion with the second RE (**Figure 7**).

The 4C products are amplified by inverse PCR, which requires primers designed outward from on view point. These primers are named as reading primer and non-reading primer. Although both are necessary in the PCR amplification, only the reading primer is going to be sequenced, so all the reads have it as part of their sequence. That allows, in the case of performing a multiple study, to identify the reads from different samples.

The reading primer is design from the first RE digestion site that is farthest from the second RE site. The non-reading primer is design around the 100bp to the second RE cut (**Figure 7**).



#### Figure 7: Viewpoint design requirements

View point size is bounded by the digestion of the chromatin by restriction enzymes (RE). The reading primer is designed from the 1<sup>st</sup> RE site, while the non-reading primer is designed in the 100bp near to the 2<sup>nd</sup> RE site. These primers need to have in their sequence the Illumina adapters to be able to sequence the 4C products.

To correctly address the genomic regions of interest an *in silico* analysis of all the *loci* was performed. First the evolutionary conservation and synteny was analysed. Once we identified this synteny, we searched on ENCODE data in the UCSC browser<sup>123</sup> histone marks related with active enhancers (H3K27ac and H3K4me1) to select the region with an optimal viewpoint and look for the desired first and second RE. Once, the viewpoint fragment was defined, primers for each viewpoint, (**Table 2**) were designed with the online program primer3 v.0.4.0<sup>124,125</sup>. To find the optimal pair of primers the default settings were optimized. Differences in temperature between primers it should be as lower as possible, and the %GC in the primer sequence should be between 40 and 60. Primers were also checked for matches elsewhere in the genome using Blast tool from the Ensembl Genome Browser for the Mouse Genome Reference (GRCm38.p6)<sup>126</sup>.

Primer	Sequence 5'-3'
mVP1_Kcnn3_DPNII_R	CTCTGCTGCTGGGAGGATC
mVP1_Kcnn3_NLAIII_NR	AGGAGGAGACTCAGGAGAGACC
mVP1_Prrx1_DPNII_R	TCTAGAACCCCTAGGAAATGATC
mVP1_Prrx1_CSP6I_NR	GAACTGTGTAAACTGTTGCTACCC
mVP2_Prrx1_DPNII_R	AGCAAAGCCAAAGTGATC
mVP2_Prrx1_NLAIII_NR	TGCTTCTTTAACCCTTACCC
mVP1_Pitx2_NlAIII_R	GACCACCTTGCTGGCCATG
mVP1_Pitx2_DpnII_NR	ATAATTTAGCTGCAGGGCTTTGG
mVP2_Pitx2_NlAIII_R	CTTTTATACTTCTTCACCTCATG
mVP2_Pitx2_DpnII_NR	TTTCTTAGACTCCCTTCTGC
mVP3_Pitx2_DpnII_R	ACAAATAGGAAGGAGTTCTGATC
mVP3_Pitx2_Csp6I_NR	TAGAAGTTTACCGGCCTGG
mVP1_Cav1_DpnII_R	CCTCGTACCTGCGTGATC
mVP1_Cav1_NlAIII_NR	CTGGTCCAGTTGTCTTGG
mVP1_Zfhx3_DpnII_R	ACCTGGAAAGACTTCTTGATC
mVP1_Zfhx3_NLAIII_NR	AGTGTAGACACTGGTTCAAGC
mVP1_Syne2_DpnII_R	ACATAGAAGACCCAGGGAGATC
mVP1_Syne2_Csp6I_NR	TGGCTCAAATCTCAGCACTCC

**Table 2: 4C-seq viewpoint primers**

Mouse (m) reading (R) and non-reading (NR) pair of primers for each viewpoint (VP) tested in this study. Primers are named with the closest gene to the viewpoint and the first (reading primer) or second (non-reading) restriction enzyme whose RE site is underlined.

## 1.2 Sample collection and 4C template preparation

4C was performed as previously described<sup>127,128</sup> with minor modifications. Pools of 20 atria or 4 ventricles dissected from 8 weeks BL6C57 male mice were disaggregated and crosslinked with 2% paraformaldehyde. Once chromatin was fixed, cells were lysed in



50mM Tris-HCl pH7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% TX-100 and 1x complete protease inhibitor (Roche). Afterwards, chromatin was digested with the first RE (DpnII or NlaIII, depending on viewpoint design; **Table 2**). Subsequently, the first ligation was performed with T4 DNA ligase (Promega), followed by the reversal cross-links to remove the previously fixed interactions. After the second digestion (DpnII, NlaIII or Csp6I depending on viewpoint design; **Table 2**) and second ligation, 4C templates were purified using the PCR purification kit QIAquick (Qiagen) and quantified by spectrophotometry using a NanoDrop equipment (Thermo Scientific).

### 1.3 Library generation

Before performing the final inverse PCR of 4C templates, a control PCR using short primers (without Illumina sequencing adaptors) was performed to check primers specificity and range of amplification. The PCR was performed using the Expand Long Template Polymerase enzyme (Roche) and serial dilutions (100ng; 50ng; 25ng;12,5ng) of the 4C templates and a standard 4C-PCR program of 30 cycles (Splinter et al., 2012). Once primers were checked, the same PCR was performed using the long primers (including Illumina adaptors). These were P5 (the reading primer, ≈75nt) and P7 (the non-reading primer, ≈40nt). After checking functionality and the efficiency of long primers, the final PCR of 4C templates was performed. For each sample (atria and ventricle pools) and pair of primers, a single mix reaction was prepared with 500ng of the 4C template. This mix was distributed in 16 tubes, and after PCR was completed reactions were pooled and purified with the High Pure PCR Product Purification Kit (Roche), followed by a second purification with Agencourt AMPure Kit to fully remove primers and small fragments. Finally, the purity and concentration of each sample was measured using the Qubit® dsDNA HS Assay kit for 100 Assys, 0.2-100ng (Invitrogen) and Qubit fluorometer® 2.0.

### 1.4 Library sequencing

Libraries were diluted to 10nM using Tris-Cl 10mM, pH8.5 with 0.1% Tween 20. Libraries were applied to an Illumina flow cell for cluster generation (HiSeq SR Cluster Kit v4 cBot) and sequenced-by-synthesis single reads of 50 bp using the HiSeq SBS Kit v4 (Illumina) on the Illumina HiSeq 2500 following standard protocols. Sequence reads were processed using the CASAVA package (Illumina) to produce fastq files.

### 1.5 4C-seq data analysis

#### 1.5.1 Reads Normalization

4C-seq reads were mapped using Bowtie-1.1.2<sup>129</sup> against the Mouse Genome Reference NCBI37/mm9 allowing a maximum of three mismatches. Reads mapping to fragments flanked by two restriction sites of the same enzyme, fragments smaller than 40 bp, or within a window of 10 kb surrounding the viewpoint were filtered out. Mapped reads were converted to reads per first enzyme fragment ends and smoothed using a 30-fragment mean running window algorithm. Smoothed scores of each experiment were normalized by the total number of reads before visualization of 1-2Mb around each viewpoint.

#### 1.5.2 Statistical Contact Estimation

Fastq files were demultiplexed using Cutadapt<sup>130</sup> using the viewpoint sequences as indexes. Potential Illumina adaptor contaminants were removed from the 3' end of the reads and the viewpoint from the 5' end but without removing the restriction enzyme site. To avoid chimeric reads derived from ligation of short fragments, a final step was included to remove a second cutter sequence close to the 3' end, keeping sequences with a minimum output length of 18 nucleotides.

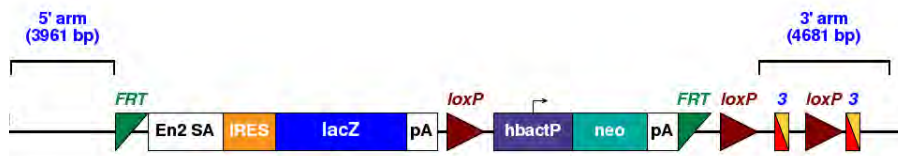
For contact estimation, the reference genome was virtually cut according to the first and second restriction enzymes. Then processed reads were assigned to their corresponding first cutter digested genome fragment, filtering out reads on fragments located in 5kb around the viewpoint. Quantification was performed considering each fragment end as one captured site if one or more sequences mapped to it. The number of captured sites was summarized per 30 fragments window (max of 60 captured sites per window). The frequency of captured sites per window was used to fit a distance decreasing monotone function and Z-scores were calculated from its residuals using a modified version of FourCSeq<sup>131</sup>. Significant contacts were considered in cases where the Z-score was  $>2$  in both replicates and deviated significantly (adjusted p value  $<0.05$ ) from its normal cumulative distribution in at least one of the replicates.

After data analysis processed reads and interactions were visualized at the WashU EpiGenome Browser Mouse genome (mm9)<sup>132</sup>.

## 2. Mouse analysis

### 2.1 Mouse model

In this work, we have used the mouse line *Cav1<sup>tm1a(KOMP)Mbp</sup>* (named from now on as *Cav1LacZ*) to characterize *Cav1* promoter activity through a *LacZ* reporter. This mouse line was originally obtained from UC DAVIS-KOMP Repository from the Knock-out first project<sup>133</sup>, and kindly provided by Dr. Miguel Ángel del Pozo (CNIC). The allele generated combines the advantage of both a reporter allele and a conditional knockout allele, following exposure to site-specific recombinases Flp or Cre (**Figure 8**).



**Figure 8: Cav1LacZ cassette**

Diagram of the *Cav1<sup>tm1a(KOMP)Mbp</sup>* mouse line cassette between intron 2 and exon3. It has *LacZ* as reporter allele, however it also can be used to produce conditional knockout allele.

The mouse line was maintained on heterozygosity (*Cav1LacZ/+*). Adult mice were genotyped by PCR of tail-tip DNA using primers listed in **Table 3** and using PCR conditions listed in **Table 4**. For genotyping, two separated PCR, with a common reverse primer, were performed one for the wildtype (WT) allele that produces a 279 base pairs (bp) product and another one for the *Cav1LacZ* allele that has a size of 308bp

Primer	Sequence 5'-3'
Intron II F cKO 2	GTCATTTCTGTGCCTCAACTCG
Common-en2 R	CCAACTGACCTTGGGCAAGAACAT
Intron II R cKO	CCAGGTGTTTCTTTGCTGTGTC

**Table 3: Genotyping primers for Cav1LacZ mouse line alleles**

Primers for the WT and cassette allele of *Cav1LacZ* mouse line.

Step	Temperature	Time	Cycles
Initial denaturation	94 °C	1 minute	
Denaturation	94 °C	0.5 minutes	
Annealing	60 °C	0.5 minutes	x35
Extension	72 °C	0.5 minutes	
Final extension	72 °C	10 minutes	
Hold	14 °C	indefinitely	

**Table 4: PCR conditions for the Cav1LacZ mouse line genotyping**

Mice were housed and maintained in the animal facility at the Centro Nacional de Investigaciones Cardiovasculares, CNIC, (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.

## 2.2 Embryos and adult heart collection

*Cav1LacZ*<sup>+</sup> heterozygous male mice were crossed with WT female mice (CD1 or BL6C57) to collect embryos from all stages between embryonic day (E) 9.5 and E18.5.

**Table 5** summarize all collected embryos. Early staged embryos (E9.5, E10.5 and E11.5) were analysed in whole mount. Hearts from E12.5 to E18.5 embryos and onwards were manually dissected, except for a few embryos at E12.5, E13.5 and E14.5 that were collected for whole embryo staining. Day 3 postnatal embryos (P3) and adult mice (14-15 weeks old) were euthanized for heart extraction.

dpc	N° total embryos collected	N° of <i>Cav1LacZ</i> + embryos	% of <i>Cav1LacZ</i> +
<b>E9.5</b>	17	0	0%
<b>E10.5</b>	17	9	53%
<b>E11.5</b>	14	9	64%
<b>E12.5</b>	11	7	64%
<b>E13.5</b>	19	11	58%
<b>E14.5</b>	19	9	47%
<b>E15.5</b>	10	4	40%
<b>E16.5</b>	25	10	40%
<b>E17.5</b>	6	4	67%
<b>E18.5</b>	18	7	39%

**Table 5: Summary of total embryos collected**

## 2.3 $\beta$ -Galactosidase staining

Whole embryos and whole hearts were placed in a fixing solution (PBS-IGEPAL 0.2%, 1% formaldehyde, 0.2% glutaraldehyde, 2mM MgCl<sub>2</sub>, 5 mM EGTA) at room temperature for 20, 30 and 45 minutes for E9.5, E10.5 and E11.5 respectively; 30 minutes for hearts from E12.5 to E18.5 embryos and 1hour and 30 minutes for adult hearts and whole embryos from E12.5. After two washes in PBS-IGEPAL 0.2% for 10min (embryos) and 20 minutes (adults) at room temperature embryos or dissected hearts were stained overnight with LacZ staining solution (5 mM K<sub>3</sub>Fe, 5 mM K<sub>4</sub>Fe, 2 mM MgCl<sub>2</sub>, 1.04 mg/mL X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside) at room temperature. Samples were washed twice with PBS-IGEPAL 0.2% for 10min (embryos) and 20 minutes (adults) at room temperature and overnight at 4°C. Finally, samples were fixed with 4% PFA.

## 2.4 Histological sections

For histology sections, E18.5 hearts were paraffin embedded by the CNIC Histopathology Unit. 10µm sections were obtained with a Leica RM2245 semi-automatic microtome. After drying, sections were stained with Nuclear Fast Red.

## 3. hiPSC culture, differentiation and functional characterization

### 3.1 hiPSC culture and differentiation

In this study we have worked with the established BJ-hiPSC line, which was originally reprogrammed by using mRNA<sup>134</sup>. hiPSCs were maintained in feeder free conditions on matrigel (Corning Matrigel Growth Factor Reduced) with iPS Brew Media (Miltenyi Biotec, Germany) with manual marking and picking to remove spontaneous differentiation. Cardiac directed differentiation to obtain hiPSC-derived cardiomyocytes (hiPSC-CMs) was performed as previously described using the small molecule protocol<sup>110</sup>. Briefly, hiPSCs were individualized with versene, plated as monolayers in 6 well plates and maintained in stem cell media until reaching ~90% confluence. At day 0, cardiac directed differentiation was initiated by adding to RPMI media supplemented with B27 (-insulin) the glycogen synthase kinase-3 inhibitor CHIR99021 (12µM). On day 1 CHIR99021 was removed and cells were cultured in RPMI with B27 (-insulin) until day 3 when the small molecule Wnt inhibitor IWP4 (5µM) was added to the cells and not removed until day 5. From day 5 to 7 the cells were maintained in RPMI with B27 (-insulin). On day 7 RPMI media was supplemented with B27+ insulin and cells were maintained in this media with daily changes until day 27 approximately (15 day after first beating was detected). At day 27, hiPSC-CMs were trypsinized and cardiomyocytes were purified using magnetic activated cell sorting to target and deplete non-myocytes (Miltenyi Biotec, PSC-derived cardiomyocyte isolation kit, human) and autoMACS Pro Separator to purify cardiomyocyte and non-cardiomyocyte populations.

### 3.3 CRISPR-Cas9 deletion

We aimed to delete a region at the intron2 of *CAV1* gene containing the SNP rs3807989 that has been associated to AF by using two guides RNA (**Table 6**), designed using the CRISPR Design Tool from <https://benchling.com/crispr>.

sgRNA	Target site 5'-3'	PAM
gCAV1_SNP1	GGGACGGAACTTTCAACACT	TGG
gCAV1_SNP2	GCGGTTTTAGATTCCGATGT	GGG

**Table 6: gRNAs using for CRISPR/Cas9 mediated deletion**

Each guide was cloned in a pGEM<sup>®</sup>-T Easy Vector System (Promega #A137A) under U6 promoter together with the tracrRNA.

8x10<sup>5</sup> hiPSCs were electroporated with 1250ng of each gRNA containing plasmid, together with 2500ng of the plasmid pX458 (pSpCas9(BB)-2A-GFP, Addgene #48138). This plasmid is ready to express the spCas9 protein and a GFP reporter. Electroporation was performed using program A-23 at Nucleofector<sup>®</sup> 1 Device (Lonza) following the standard protocol for the Human Stem Cell Nucleofector Kit 2 (Lonza, VPH-5022).

GFP-positive cells were enriched by sorting in a MoFlo<sup>®</sup> Astrios<sup>™</sup> (Beckman Coulter) cell sorter. On average, over 16% of the transfected cells were positive for GFP. Sorted cells were replated and cultured for 2 days in post-sorting media (50% StemMACs iPSBrew XF media, 50% conditioned StemMACs iPSBrew XF media, 10μM Rock-Inhibitor Y-27632, 20ng/ml βFGF) changed daily, and then cultured with regular StemMACs iPSBrew XF media (MACs Miltenyi Biotech, # 130-104-368).

22 individual clones were grown and genotyped using primers of **Table 7** of which 10 were mutant, 6 were heterozygous and 6 were wildtype.

Oligos for cloning	Sequence 5'-3'
Cav1_SNP_F	CATGCAGACCAATGGGTAGA
Cav1_SNP_R	TTCTTGTTGCCAGTCTGAA

**Table 7: Primer sequence for genotyping the deletion**

#### 4. RNA-extraction

RNA from LA and RA isolated cardiomyocytes and from cells in tissue culture was extracted and purified using the RNeasy Mini Kit (Qiagen) following manufacturer's instructions and digested with DNaseI (Qiagen) to remove genomic DNA.

RNA from 30-35mg of posterior left atria (PLA) wall frozen tissue was extracted using Quiagen RNeasy Mini Kit. Disruption and homogenization of the tissue was performed using T10 ULTRA-TURRAX (IKA Works Inc., Alemania) and 600ul of RLT buffer supplemented with β-mercaptoethanol. DNase digestion was performed always during RNA extraction using Qiagen RNase-Free DNase Set.

RNA tissue concentration was checked with NanoDrop ND-1000 and RNA cardiomyocyte concentration was determined with NanoDrop 2000. RNA quality control for RNA-seq was checked with Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

## 5. RT-qPCR

For quantitative-RT PCR analysis from 1µg to 400ng of total RNA of each sample was reversed-transcribed to cDNA using Quantitech Reverse kit (Applied Biosystems) and diluted 1:20 as template for amplification reactions, carried out with the Power SYBR green PCR Master Mix (Applied Biosystems) following manufacturer's instructions. PCR amplifications were carried on an AB 7900-Fast-384 machine. Expression values were normalized to the expression of the housekeeping gene GAPDH. Primer used are detailed in **Table 8**.

Gene	Primer F	Primer R
<b><i>Cav1</i></b>	CGCGACCCTAAACACCTCAA	GAAGCTGGCCTTCCAAATGC
<b><i>Cav2</i></b>	G TTCCTGACGGTGTTCCTGG	AGAAGCATCGTCCTACGCTC
<b><i>Tes</i></b>	GCTTAGGTACGAGCAAGGA	TGCTCTTCTTGCCACACTT
<b><i>Met</i></b>	TGAGAGCTGCACCTTGACTT	AAGTGCCACCAGCCATAGGA
<b><i>Capza2</i></b>	TACCCGAATGGAGTCTGCAC	TGACCTCCAACGACCATTCC
<b><i>Tnnt2</i></b>	AAGAGGAAGCAAAGGAGGCT	TCCATCGGGGATCTTG GGAG
<b><i>MyI7</i></b>	CACCGTCTTCCTCACGCTCT	GGGTGTCAGGGCGAACATCT
<b><i>MyI2</i></b>	G TTCGGGAAATGCTGACCAC	AAGTTGCCAGTCACGTCAGG
<b><i>GAPDH</i></b>	CGCTCTCTGCTCCTCCTGTT	CCATGGTGTCTGAGCGATGT

**Table 8: Primers sequences used for RT-qPCR assays**

## 6. Western Blot

hiPSCs, hiPSC-CMs and non-myocytes were collected, and the pellet was frozen in liquid nitrogen. For protein extraction, pellets were resuspended in RIPA buffer (50mM Tris-HCl pH 8, 150 mM NaCl, 0,02% NaN<sub>3</sub>, 0,1% SDS, 1% NP-40, 1% Sodium Deoxycholate) containing protease inhibitors (Complete ULTRA tablet; Roche; 06538304001). Lysates were sonicated to shear genomic DNA and cellular debris was removed by centrifugation at 4°C for 10 minutes at 13000 rpm. Protein concentration was quantified using BioRad DC protein assay (BioRad, 5000112) and 30 µg of each sample was resolved in 10% SDS-polyacrilamide gels.

Sheep PLA tissue samples were washed with protease inhibitors (Roche, protease inhibitor tablet) containing PBS and flash frozen in liquid nitrogen. Frozen tissue (50–100mg) was homogenized in 1ml of lysis buffer containing (in mM): Tris•HCl (25), NaCl (150), EDTA (1), NaF (4), Sodium ortho-vanadate (2), Triton X-100 1% and protease inhibitor. The homogenate was centrifuged at 10000 rpm for 5 minutes; the supernatant was used for western blotting. The tissue lysates (20 µg) was resolved in 10% SDS-polyacrilamide gels.

Cell proteins extracts were transferred to Polyvinylidene difluoride (PVDF) membrane (Immobilon®-P, Millipore, IPVH100010) and PLA tissue extracted proteins were transferred to Nitrocellulose Membrane. Membranes were blocked with 5% non-fat dry milk in TBS-T (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween-20) or PBST (PBS, 0.1% Tween-20) respectively, for 1 hour and next incubated overnight at 4°C with the indicated antibodies diluted in the blocking solution (**Table 9**). Next day, the membrane was washed 5 times with TBS-T or PBST for 5 minutes and incubated with the secondary antibody (**Table 9**) diluted in the blocking solution, for 1 hour at room temperature. Then the membrane was washed again 5 times with TBS-T or PBST for 5 minutes.

Immunolabeling was detected by enhanced chemiluminescence (Amersham ECL Western Blotting Detection Kits, GE Healthcare, RPN2108) and exposed to medical X-Ray film (Agfa, ENKMV) in the case of cell protein extracts. In the case of tissue protein immunolabeling was detected by Chemiluminescent Western Blot Detection Kit: Pierce ECL, SuperSignal West Pico, SuperSignal West Fempto (Thermo Scientific) and exposed to UV light on a Gel Doc (BioRad).

Antibody	Reference	Dilution
Rabbit anti-CAV1	(D46G3) XP 3267 Cell Signalling	1:1000
Mouse anti-GAPDH	MAB374 Merk	1:10000
Goat anti-ChAT	AB144P Chemicon	1:500
Rabbit anti-TH	AB152 Chemicon	1:1000
Goat anti-Mouse Immunoglobuling HRP	P0447 Dako	1:2000
Goat anti-Rabbit Immunoglobuling HRP	P00448 Dako	1:2000

**Table 9: Antibodies used in WB assay**

## 7. Statistics

Statistical analysis was performed with GraphPad Prism 7. Data are presented as means or means  $\pm$  s.d. as indicated. Differences were considered statistically significant at p-value  $< 0.05$ . Tests used to calculate p-value are detailed in the figure legends. Unpaired two-tailed T-test was used to compare two groups and Pearson correlation for pairwise correlations.

## 8. AF induced Sheep model

### 8.1 Experimental animals

Atrial fibrillation was induced using a tachypacing device placed at right atria. Pacemaker implantation and pacing protocol were performed as described<sup>121</sup>.



Samples for this work were taken from sheep used in this previous study. Three groups of three animals each were used for each analysis on this study: transition ( $11.33 \pm 4.04$  days of self-sustained AF without reversal to sinus rhythm), chronic or long-standing persistent AF ( $365.33 \pm 10.50$  days of self-sustained AF without reversal to sinus rhythm) and control group (two sham operated animals and one non-operated, always in sinus rhythm).

## 8.2 Heart removal and tissue isolation

Hearts were removed by thoracotomy and placed in cold cardioplegic solution. Left Atria (LA) and Right Atria (RA) were removed and divided in three parts: posterior, medial, and anterior. P was used for molecular biology; M was used for histology and A was used for cardiomyocyte isolation. In addition, Posterior Left Atria (PLA) was taken for histology and molecular biology analysis.

All procedures were approved by the University of Michigan Committee on Use and Care of Animals and complied with National Institutes of Health guidelines.

## 8.3 Cardiomyocytes isolation

Cardiomyocyte isolation was performed as previously described<sup>121</sup> from anterior regions of LA and RA. Tissue samples were placed into a stock solution containing (in mM): NaCl (120), KCl (5.4), MgSO<sub>4</sub> (5), Pyruvate (5), Glucose (20), Taurine (20), HEPES (20) and nitrilotriacetic acid (5). Then tissue was cut with scissors into 1mm<sup>3</sup> pieces. Pieces were shaken at 37°C for 12 min bubbling with 100% O<sub>2</sub>. Every 3 minutes, tissue was transferred to fresh stock solution through gauze. Then pieces were transferred to a calcium free protease digestion solution for 45 minutes containing (in mM): NaCl (120), KCl (5.4), MgSO<sub>4</sub> (5), Pyruvate (5), Glucose (20), Taurine (20), HEPES (20) and protease type XXIV (Sigma) for 45 min. After, protease digestion pieces were transferred collagenase digestion solution containing (in mM): NaCl (120), KCl (5.4), MgSO<sub>4</sub> (5), Pyruvate (5), Glucose (20), Taurine (20), HEPES (20), CaCl<sub>2</sub> (0.05) and collagenase type I (Worthington) for 3 digestion time points. The first time point was taken after 15min of digestion, second one after 30 min and last one at 45 min. Myocyte suspension was decanted and centrifuged 2 min at 500 rpm. Supernatant was discarded, and pellets were resuspended on KB solution containing (in mM): L-Glutamic Acid (50), KOH (70), KCl (30), L-Aspartic Acid-K (10), KH<sub>2</sub>PO<sub>4</sub> (10), MgSO<sub>4</sub>-7H<sub>2</sub>O (2), Glucose (20), Taurine (20), Creatine (5), EGTA (0.5) and HEPES (10). and centrifuged one more time. Supernatant was aspirated, and pellets were frozen in liquid nitrogen.

### 8.4 RNA-seq

RNA-seq from LA and RA isolated cardiomyocytes were performed by the CNIC Genomics Unit using the Illumina HiSeq 2500 sequencer. PLA RNA-seq was done, also by CNIC Genomics Unit, using the Illumina GAIIx sequencer.

RNA-seq analysis was performed by CNIC Bioinformatics Unit. Sequencing reads were pre-processed by means of a pipeline that used Cutadapt 1.7.1 (1) to trim sequencing reads, eliminating Illumina adaptor remains, and to discard reads shorter than 30 bp, and FastQC (2), to assess read quality. The resulting reads were mapped against the sheep transcriptome (Oar v3.1, release 75, feb2014 archive) and quantified using RSEM v1.17. (for PLA reads) and RSEM v1.2.20 (for LA and RA isolated cardiomyocytes). In PLA RNA-seq only genes with at least 1 count per million in at least 3 samples were considered for statistical analysis. Data were then normalized, and differential expression tested using the bioconductor package EdgeR. In LA and RA isolated cardiomyocytes, data were then processed with a differential expression analysis pipeline that used Bioconductor package limma for normalization and differential expression testing.

Functional enrichment analysis was performed with Gene Ontology (GO) for Biological Processes using Enrichr<sup>135,136</sup>. GO were ranked by a combined score, that is calculated multiplying the log of the p-values from the Fisher exact test by the Z-score of the deviation from the expected rank.

### 8.5 LC-MS/MS proteomics

LC-MS/MS proteomic profiling was performed on LA and RA isolated cardiomyocytes of the same sheep as the ones used for CMs RNA-seq by the CNIC Proteomics Unit. Quantification and normalization of the data was carried out using the open source QuiXoT software<sup>58</sup> developed by the Unit. The TMT 10-plex system was used, which make it possible to compare up to 10 different isobaric compounds in a single run. We used a pool of the sham-left and sham-right samples as reference in both runs, one for the left atrial appendage and other for the right, to be able to compare the 18 samples of CMs proteome. For differential abundance analysis and visualization, we used in-house scripts.

## 9 Immunofluorescence staining

Tissue samples were sectioned longitudinally to the atrial wall plane at 4  $\mu$ m, fixed in 10% buffered formalin, and embedded in paraffin. For immunofluorescence staining,

sections were deparaffinized and rehydrate through a graded ethanol series and boiled in 0.01M citrate buffer, pH=6.0. Washed slices were blocked with donkey serum in PBS+0.1% Triton-X-100. Drained slides were incubated in primary antibody (**Table 10**) overnight. On the following day, slides were incubated with secondary antibody (**Table 10**). Finally, DAPI was applied at 1 $\mu$ g/ml. Slides were mounted using FluoromountG (SouthernBiotech).

Antibody	Reference	Dilution
<b>Goat anti-ChAT</b>	AB144P Chemicon	1:100
<b>Mouse anti-Actinin</b>	A7811 Sigma-Aldrich	1:300
<b>Donkey anti-Goat IgG Cy3 conjugate</b>	AP180C Chemicon	1:250
<b>Alexa Fluor 488 Donkey anti-mouse IgG</b>	715-545-150 Jackson ImmunoResearch	1:250

**Table 10: Antibodies used on Immunofluorescence assay**

## 10. Imaging

$\beta$ -galactosidase staining was observed using a Leica MZ FLIII Scope and photographed with a Nikon digital camera DXM 1200F and NIS-Elements D 3.2 software.

Histological sections imaging was performed using a Nikon eclipse 90i microscope with a Nikon digital camera DXM 1200F and NIS-Elements D 3.2 software.

Immunofluorescence images were obtained on a Zeiss Axioplan 2 imager equipped with structured illumination (Apotome).

## 11 Mitochondrial functional assays

### 11.1 Blue native gel electrophoresis

Mitochondria extraction from LA and RA isolated cardiomyocytes from our AF induced model were performed in triplicates. Briefly, cardiomyocytes frozen pellets were resuspended in PBS and solubilized with 1 volume of digitonin (8mg/ml) (Sigma D5628), incubated on ice for 10 minutes, diluted with PBS and centrifugated 5 minutes at 10000g at 4°C. Supernatant containing cell membranes and debris was discarded, pellet was washed again with PBS. After a new centrifugation at 10000g at 4°C supernatant was discarded and pellet enriched in mitochondria was resuspended in 1.5M aminocaproic acid, 50mM Bis-Tris HCl pH7. At this point enriched mitochondria fraction was quantify by Bradford<sup>137</sup> (Bio-Rad protein assay). 30-35  $\mu$ g of the sample was collected for future SDS assays and solubilized in 4x Laemmli Sample Buffer (BioRad) and frozen. The rest of the sample was solubilized with 1% digitonin adjusting the volume to the total amount of protein of each sample. After 5 minutes of ice

incubation samples were centrifuged 30 minutes at 18000g at 4°C. Supernatant enriched in mitochondrial proteins were collected and after adding Blue G Sample buffer were run in a 3%-13% gradient Blue Native gel as described before<sup>138</sup>. After electrophoresis, native mitochondrial proteins were electroblotted onto PVDF membrane (Merck-milipore). Membranes were blocked with 1% non-fat dry milk in PBS-T (PBS, 0.1% Tween-20) for 30 minutes and next incubated overnight at 4°C with the indicated antibodies diluted in the blocking solution (**Table 11**). The secondary antibody was Alexa Fluor™ 680 goat anti-mouse (Invitrogen A21057). The signals were acquired with the ODYSSEY Infrared Imaging System (LI-COR)

Antibody	Reference	Dilution
<b>Mouse anti-CORE1</b>	Abcam ab110252	1:1000
<b>Mouse anti-FP70</b>	Invitrogen 459200	1:1000
<b>Mouse anti-COX1</b>	Invitrogen 35-8100	1:1000
<b>Mouse anti-NDUFA9</b>	Abcam ab14713	1:1000
<b>Alexa Fluor™ 680 Goat anti-Mouse</b>	Invitrogen A21057	1:10000

**Table 11: Antibodies used for Blue Native assays**

## 11.2 Enzymatic in gel activity determination

Determination of enzymatic in gel activity of NADH-dehydrogenase (Complex I activity) was carried out as previously reported<sup>139</sup> in one sample of each condition. Native mitochondrial proteins were run in a 3%-13% gradient Blue Native gel. The gel was incubated in 0.1M Tris-HCl, 0.14mM NADH and 1.0mg/ml NBT (nitro blue tetrazolium), pH 7.4. Gel was scanned at 1h, 2h, 4h 30 minutes and after overnight incubation.

## RESULTS

---

*To build may have to be the slow and laborious task of years.  
To destroy can be the thoughtless act of a single day.*

Wiston Churchill



## 1. Regulatory region screening on AF associated *loci*

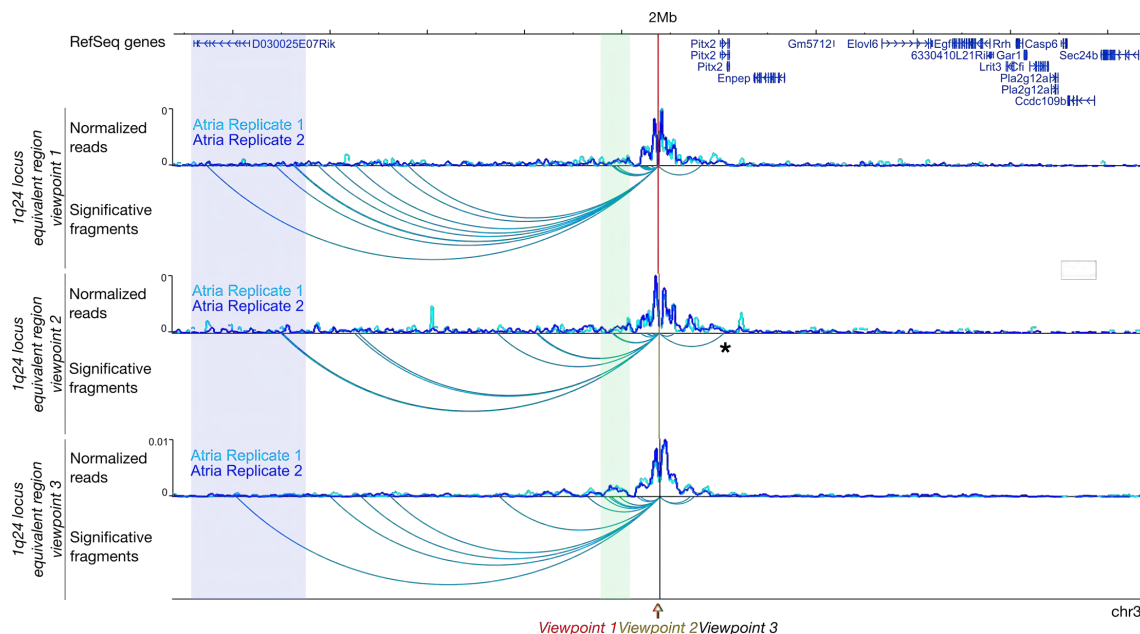
### 1.1 Chromatin structure profile of the 4q25 *locus*

It has been shown that more than 90% of disease associated variants are located in non-coding regions<sup>76</sup>, and AF associated SNPs are not an exception. Previously in our lab, we have explored 4q25 *locus* which harbours highly AF associated variants inside a 1.5Mb gene desert. Within that *locus* there is a regulatory region, conserved between human and mouse, that includes the first SNP associated to AF (rs2200733) and that establishes a long-range interaction with *Pitx2c* and *Enpep* genes promoters<sup>78</sup>. Despite these two genes having different roles, *Pitx2c* encodes a transcription factor relevant in the establishment of left-right asymmetry during development<sup>34</sup> and *Enpep* is involved in the renin-angiotensin system<sup>140</sup>, we observed how they are co-expressed in the pulmonary veins during heart development. Therefore, an alteration of a shared regulatory element could affect their expression on very relevant tissues for AF development.

These enhancer-promoter interactions were identified by 3C, a technique that is driven by the specific hypothesis of assess whether two predefined regions are interacting among them or not. Since, other SNPs have been associated to this *locus* (rs10033464, rs6843082, rs17042171, rs6817105)<sup>33,40,41,49</sup> we extended this study to explore if there could be other active regulatory elements within 4q25 *locus* and which genes could be their targets. To further examine this *locus* and study several candidate regulatory regions at the same time we decided to perform a 4C-seq assay. However, access to human heart samples is limited and this assay requires a considerable amount of starting material. Therefore, we carried out this analysis on atrial and ventricular cardiac tissue from adult mice. In order to select new potential regulatory elements in this *locus* and establish our viewpoints, we performed an *in silico* analysis. To correctly address the genomic region to be examined in mouse, we first analysed the evolutionary conservation and synteny of genes and non-coding elements between human and mouse (Appendix 1). Thereby, the gene desert that in humans harbours AF associated SNPs has a size of 1.5Mb, and in the mouse genome it has a similar size of 1.3Mb. Once we identified this synteny, we searched on ENCODE data in the UCSC browser<sup>123</sup> histone marks related with active enhancers (H3K27ac and H3K4me1) at 4q25 *locus*. After this *in silico* analysis, we selected three viewpoints in the mouse genome equivalent to the human regions that contain rs6817105, rs2200733, rs2129977 AF associated variants (Appendix 1). These viewpoints were separated by a distance of

≈3kb. In this way we also wanted to evaluate the resolution of 4C-seq to distinguish different chromatin profiles between viewpoints placed in a close range.

In all the cases in atrial tissue, we identified multiple interactions with other regions of the gene desert (**Figure 9**). We did not succeed in obtaining useful data from ventricles, since one of the replicates failed. All three viewpoints interact with a long non-coding RNA (D030025E07Rik) located ≈840kb upstream of the viewpoints (**Figure 9**, blue highlighted region). This long non-coding RNA has been identified as *Playrr* (*Pitx2* locus-asymmetric regulated RNA)<sup>141</sup>. It is expressed at the right side of the dorsal mesentery while its expression at the left side is repressed by *Pitx2*. Welsh and co-workers evidenced that *Pitx2* and *Playrr* mutual antagonism is coordinated by asymmetric chromatin interactions<sup>141</sup>. However, in the human genome any long-coding RNA has been described at this *locus* so far. In the human genome, 70kb upstream rs2200733 has been identified a microRNA, miR297, that recently has been involved in cardiomyocyte hypertrophy<sup>142</sup>. At this distance we detected several interactions upstream of our viewpoints (**Figure 9**, green highlighted region), however, any microRNA has been annotated in the mouse genome in this region.



**Figure 9: Chromatin interaction profiles in 4q25 mouse syntenic region**

4C-seq from three viewpoints, located at the mouse syntenic regions to the human 4q25 *locus*, shows long-range interactions in mouse atria. The three viewpoints interact with the long non-coding RNA D030025E07Rik (*Playrr*; blue highlight) and have other upstream interactions (green highlight). Viewpoint 2 also interacts with different *Pitx2c* promoter (asterisk). Genomic window shown corresponds to mm9; chr3:127778946-129778946.

Less interactions were found downstream of the viewpoints and are mostly limited to the *Pitx2* gene. Viewpoint 1 and 3 have interactions 40-50kb away from *Pitx2a* and *b*



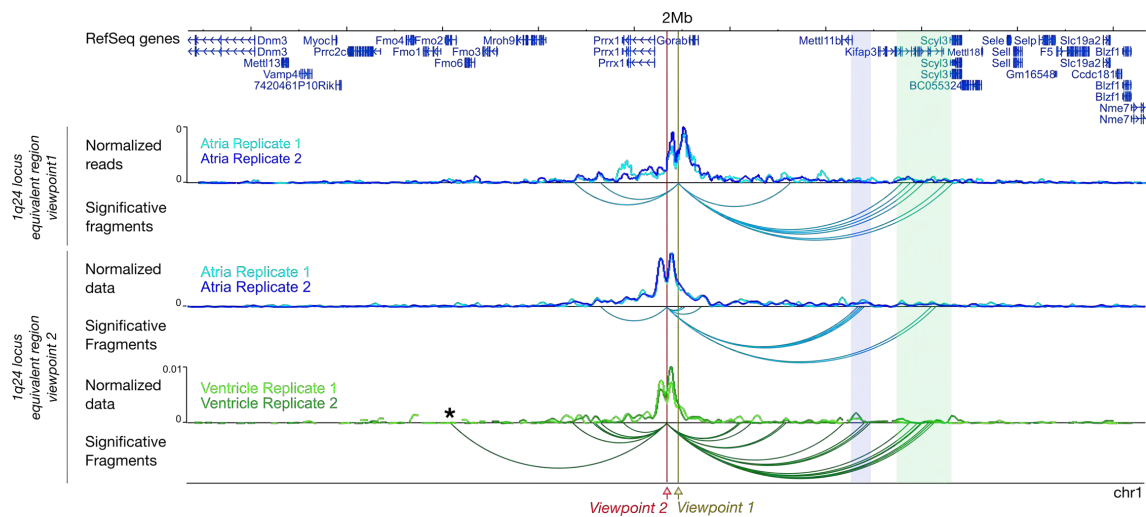
isoforms (two out of the three isoforms of *Pitx2* gene that are generated by alternative splicing) and 50-60kb from *Pitx2c* isoform promoter (third isoform of *Pitx2* gene that uses an alternative promoter and that is mainly expressed in the heart)<sup>143</sup>. Viewpoint 2 showed an interaction inside *Pitx2c* promoter region (**Figure 9**, marked with an asterisk). Therefore, we are able to capture subtle differences on chromatin interaction profiles of nearby viewpoints that could indicate different promoter-regulatory element combinations. The interaction map that we have generated based on 4C-seq confirms previous findings in relation to the regulatory landscape of this region<sup>141</sup> despite we could not confirm previously described interaction with *Enpep*<sup>78</sup>.

## 1.2. Analysis of the genomic landscapes of AF associated *loci* through 4C-seq assay

We next aimed to address in a more systematic fashion the genomics landscapes of other AF-associated variants (**Table 1**), to shed light on putative regulatory elements and their target genes with a predicted role in AF. After *in silico* analysis, we discarded 10q22 and 9q22 *loci* for further study because of lack of synteny between human and mouse (Appendix 1). We also discarded 15q24 and 5q31 *loci* because disease-associated variants are located less than 10kb away from a promoter (Appendix 1). At that distance, by 4C-seq, is not possible to distinguish between interactions due to the promoter or to nearby regulatory elements.

We tested the 1q24 *locus*, where AF associated variant rs3903239 is located in a gene desert of 100kb between *GORAB* gene, that codes for a member of the Golgin family<sup>144</sup>, and *PRRX1*, that encodes for a transcription factor expressed during heart development and involved in epithelial-to-mesenchymal transition<sup>51,52</sup>. In the mouse genome this intergenic region spans 70kb but the precise and exact position where the AF SNP lies is not conserved between mouse and human. Therefore, we defined two different viewpoints in two nearby regions flanking upstream and downstream of the variant, that do show conservation and harbour histone marks for active enhancers in both species (Appendix 1). The first region we selected is located 12kb upstream of the lead SNP (the most significant SNP) in the human genome. Our 4C-seq profile from mouse atria (no useful data was obtained from ventricles) showed interactions 55kb downstream from the viewpoint with *Kifap3* and *Scyl3* promoter (**Figure 10**, green highlighted region). *Kifap3* codes for a small G-protein involved in the interaction of chromosomes with Kinesin ATPase motor proteins<sup>145</sup>, and *Scyl3* has a role in regulating cell adhesion and migration<sup>145</sup>. Upstream interactions from this viewpoint spanned the *Prrx1* gene. The

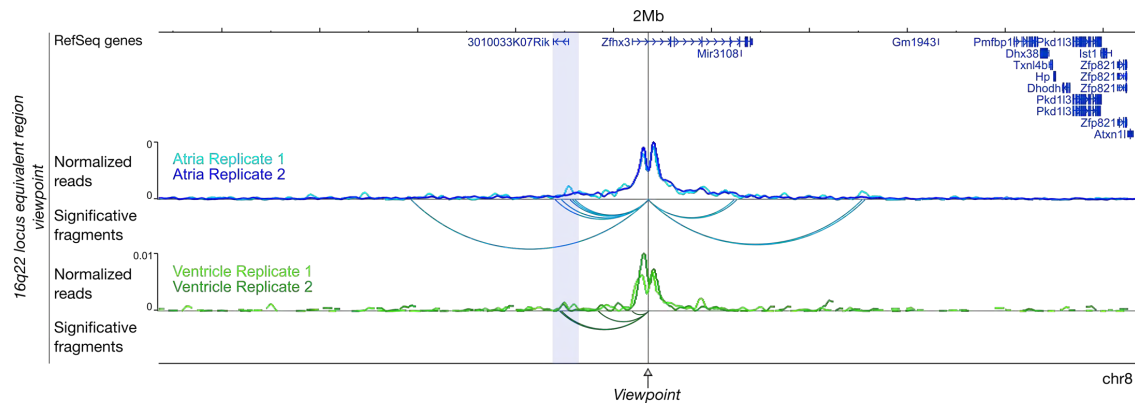
second viewpoint mapped 18kb downstream of rs3903239 in the human genome, and includes the variant rs577676, which after starting this study has also been associated to AF<sup>146</sup>. This viewpoint also interacts, in both ventricles and atria, with *Kifap3* and interactions are also established with the promoter of *Mettl11b*, whose product is involved in substrate methylation<sup>147</sup> (**Figure 10**, blue highlighted region). Similar to the previous viewpoint, there are upstream *Prrx1* interactions. We also detected interactions with the promoter of *Fmo2* region, coding for a flavin-containing monooxygenase<sup>148</sup>, but only in ventricles (**Figure 10**, interaction marked with an asterisk). Hence, we have identified in 1q24 *locus* new 3D chromatin contacts that could unveil the interaction between potential regulatory elements and new target genes that may play a role on AF development.



**Figure 10: Chromatin interaction profiles in 1q24 syntenic region of mouse** 4C-seq from two viewpoints, located at mouse syntenic regions to the human 1q24 *locus* show upstream and downstream long-range interactions. Viewpoint 1 interact with *Scyl3* and *Kifap3* (green highlight) in mouse atria. Viewpoint 2 interacts with *Kifap3* and *Mett11b* (blue highlight), in atria and ventricle. Viewpoint 2 also shows interactions with *Fmo2* gene only in ventricular tissue (asterisk). Genomic window shown corresponds to mm9; chr1:164269259-166269259.

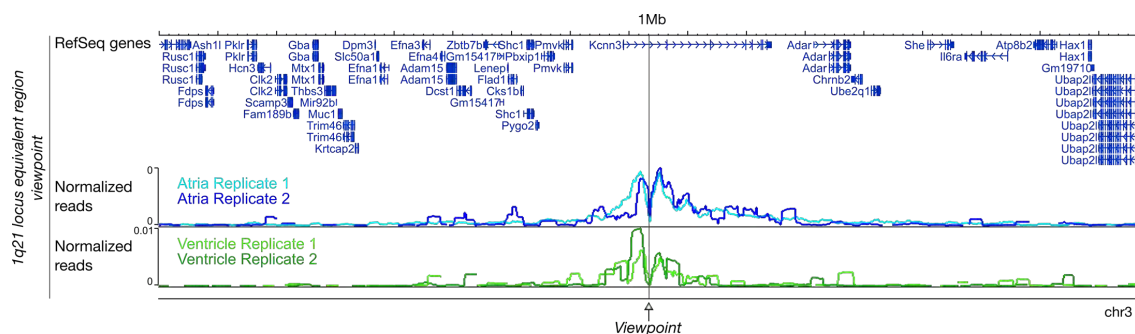
We next analysed 16q22 *locus* that holds an associated variant (rs210626) located in the first intron of the *ZFH3*, coding for a transcription factor involved in the regulation of myogenic and neural differentiation<sup>42</sup>. This gene is similarly sized in humans and mouse (265kb and 246kb respectively), the first intron included. We designed a viewpoint in the mouse genome close to a conserved region that contain the associated SNP in humans (Appendix 1). 4C-seq from this viewpoint identified significant interactions upstream of *Zfhx3*, in both atria and ventricles, towards the long non-coding RNA 3010033K07Rik also named as lnc-BATE1 (**Figure 11**, blue highlighted region). Lnc-BATE1 has been established as a regulator of brown adipose tissue development

and maintenance in mice<sup>149</sup>. In humans, has not been described any long non-coding RNA at an equivalent distance. In addition, there are significant interactions towards intron 5 of *Zfx3* and downstream regions, but only in the atria.



**Figure 11: Chromatin interaction profiles in 16q22 syntenic region of mouse** 4C-seq from a viewpoint, located at mouse regions syntenic to the human 16q22 *locus*, show long-range interaction with the long non-coding RNA 3010033K07Rik (*IncBATE1*; blue highlight) on mouse atrial and ventricular tissue and downstream interaction in atria samples with *Zfx3*. Genomic window shown corresponds to mm9; chr8:110271749-112271749.

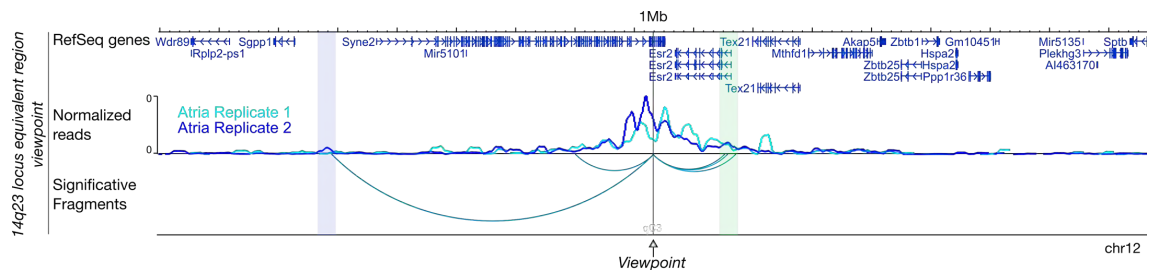
We also aim to test the chromatin landscape at 1q21 *locus* that holds several AF associated variants (rs13376333 and rs6666258) located within *KCNN3* first intron, that codes for a calcium activated potassium channel and is a prime candidate to be linked to AF<sup>47</sup>. This gene is very well conserved between human and mouse (162kb and 152kb respectively) (Appendix 1). Unfortunately, we were not able to obtain reproducible 4C data from either atria and ventricles for this *locus* (Figure 12).



**Figure 12: Chromatin interaction profiles in 1q21 syntenic region of mouse** 4C-seq normalized reads profiles of two atria and two ventricles replicates from the viewpoint located at mouse syntenic region to the human 1q21 *locus* show a lack of reproducibility between samples. Genomic window shown corresponds to mm9; chr3:88850434-89851434.

We next analysed 14q23 *locus* which harbours an AF associated variant (rs1152591) in the intronic region of *SYNE2*, that encodes a protein required for centrosome-nucleus coupling<sup>65</sup>. This gene has several isoforms in mouse and humans, and since this

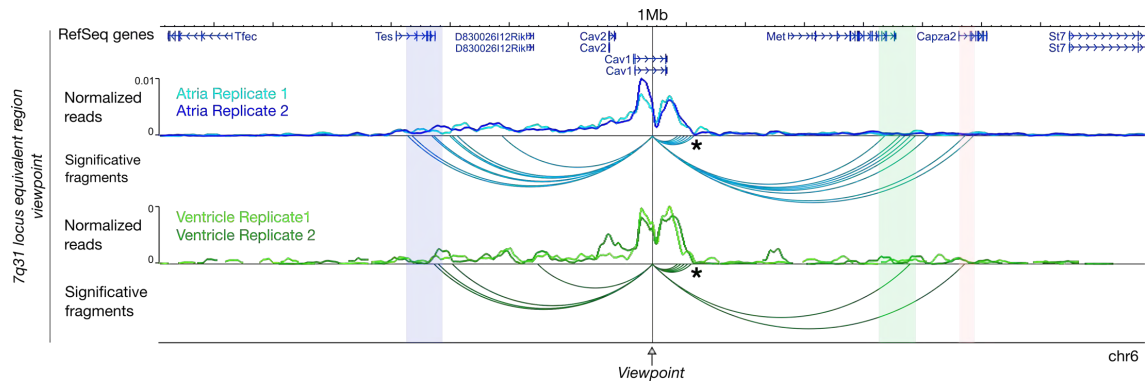
polymorphism is located in a small intronic region of 1.2kb, we defined our viewpoint flanking the orthologous intron that in mouse has size of 1.1kb (Appendix 1). This 4C-seq display fewer but interesting significant interactions in atria samples than previous viewpoints (we did not obtain valid data from ventricular tissue). There is a long-range interaction between the promoter of *Sgpp1*, which encodes for a sphingolipid metabolite that regulates diverse biological processes<sup>150</sup>, and also the promoter region of the biggest *Syne2* isoform (**Figure 13**, blue highlighted region). Downstream of the viewpoint interactions were found on non-coding regions and on the promoter region of *Esr2*, which encodes for an estrogen receptor<sup>151</sup> (**Figure 13**, green highlighted region).



**Figure 13: Chromatin interaction profile in 14q23 syntenic region of mouse**

4C-seq assay performed in mouse atria from a viewpoint located at mouse syntenic region to the human 14q23 locus. Upstream interactions are found between *Syne2* and *Sgpp1* promoters (blue highlight) and downstream with *Esr2* promoter (green highlight). Genomic window shown corresponds to mm9; chr12:76698975-77698975.

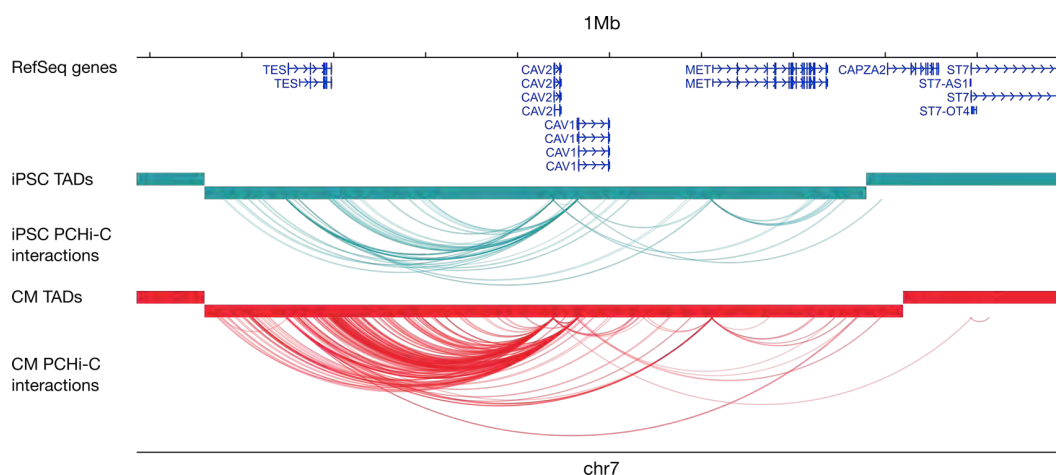
Finally, we explored 7q31 locus. AF associated variant rs3807989 is located at the second intron of *CAV1* that encodes Caveolin-1, one of the main components of caveolae<sup>152</sup>. This gene is well conserved (having a size of 36kb on human and 34kb in mouse), so we designed a viewpoint in the mouse genome region syntenic to the human 7q31 locus that includes AF-associated SNP (Appendix 1). We detected upstream of our viewpoint significant interactions in both atria and ventricle samples with *Tes* (**Figure 14**, blue highlighted region) which encodes for Testin, a scaffold protein that has a role in cell adhesion, cell spreading and in the reorganization of the actin cytoskeleton<sup>153</sup>. Downstream of our viewpoint we find interactions at the 3' untranslated region (UTR) of *Cav1* itself (**Figure 14**, interactions marked with asterisks) and with *Met* and *Capza2* genes (**Figure 14**, green and red highlighted regions). *Met* codes for the hepatocyte growth factor receptor<sup>154</sup> and *Capza2* encodes an F-actin capping protein that regulates the growth of the actin filaments. Since from all 4C-seq chromatin interaction maps obtained, the mouse syntenic region of the human 7q31 locus was showing possible interactions between regulatory elements and target genes that may play a role on AF development, we further analysed the genomic landscape surrounding rs3807989.



**Figure 14: Chromatin interaction profiles in 7q31 syntenic region of mouse**

4C-seq from a viewpoint located at mouse syntenic region to the human 7q31 *locus*, show long-range upstream interactions with *Tes* (blue highlight) and downstream interactions with *Met* (green highlight) and *Capza2* (red highlight) genes, together with *Cav1* 3'UTR (interactions marked with asterisks). Genomic window shown corresponds to mm9; chr6:16775633-17776530.

To gain a better insight of the chromatin architecture of this region, we took advantage of the recent genome-wide promoter interaction map published by Montefiori et al.<sup>155</sup> Here, the authors performed promoter capture Hi-C on hiPSCs and on hiPSC-CMs, as a source of human samples to study chromatin interactions in a cardiovascular context. As it can be observed in **Figure 15**, there is a clear increase in the number of interactions established within this region if we compare hiPSCs and hiPSC-CMs, although all interactions seen by 4C-seq are also detected in both cell types. Therefore, this chromatin profile from human cells confirms our 4C-seq performed in the mouse syntenic region to the human 7q31 *locus*. On this work we will try to dissect and understand this regulation.



**Figure 15: Hi-C promoter interaction map of 7q31 *locus* in human cells**

Promoter capture Hi-C (PCHI-C) interactions in hiPSCs and hiPSC-CMs at 7q31 *locus* within TAD organizational units. This data from Motefiori et al.<sup>155</sup> shows strong interactions between *CAV1* and *TES*, as well as with other interactions between *CAV1* and *MET* and *CAV1* and *CAPZA2* promoters. Genomic window shown corresponds to hg19; chr7:115686241-116686241.

## 2. *Cav1* is expressed during heart development and in the adult

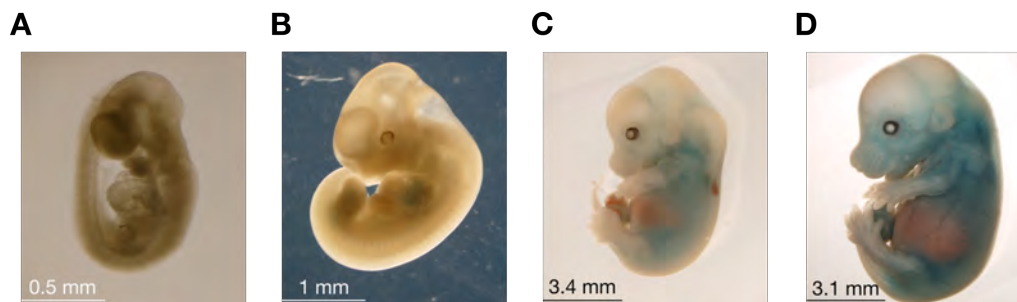
CAV1 is a member of the Caveolin gene family, which also includes by CAV2 and CAV3. All of them are involved in the formation of the caveolae, invaginated membrane nanodomains (60-80nm) with multiple functions<sup>62</sup>, among them holding a large number of ion channels<sup>156</sup>. *Cav1* and *Cav3* knockout mice loss caveolae, while *Cav2* knockout do not. However, there is a strong reduction of *Cav2* expression in *Cav1* knockout mouse<sup>157</sup>. Moreover, *Cav1* knockout mice show dilated cardiomyopathy and pulmonary hypertension, and in some circumstances cardiac hypertrophy<sup>63</sup>, suggesting that CAV1 could have a specific role in the heart caveolae. Therefore, CAV1 is an excellent candidate to perform functional analysis to better understand its role and association to AF.

CAV1 and CAV2 have been reported to be present in non-muscle cells (including adipocytes, endothelial and epithelial cells and fibroblast) and in smooth muscle-cells, while CAV3 is found in skeletal muscle and smooth-muscle cells<sup>152</sup>. Hence, CAV3 has been the one considered as specific for caveolae formation in the heart. However, there are some evidences for CAV-1 expression in the heart, as it is detected by whole-mount in situ hybridization in E9.5 and E10.5 mouse embryos<sup>158,159</sup>. Expression is also seen in adult, mouse and rat endothelial cells of blood vessels in the ventricles, and in endothelial cells and cardiomyocytes in the atria, where it colocalizes with caveolin-3<sup>160</sup>. Aside from this data, there is not a deep characterization of where and when CAV1 is expressed from the onset of heart development up to adult stages.

We asses in more detail the expression of *Cav1* in the heart by using the mouse line *Cav1<sup>tm1a(KOMP)Mbp</sup>* which includes a *LacZ* reporter that allows us to follow *Cav1* expression by  $\beta$ -galactosidase activity. We analysed heterozygous mouse embryos (*Cav1LacZ/+*) were analysed from E9.5 to E18.5. We did not observe reporter expression at E9.5 (**Figure 16A**), but from E10.5 to E18.5 we observed a gradual increase in expression, starting in the heart in left and right atria at E10.5 (**Figure 17A**, black arrow). At E11.5 we started detecting  $\beta$ -galactosidase staining in other regions of the embryos such as the hindlimb (**Figure 16B** and **Figure 17B**, marked with an asterisk), which is even more noticeable at later stages like E13.5 and E14.5 (**Figure 16C-D**). Focusing in the heart, at E12.5 reporter expression is mainly detected in the atria but starts to be spread to the ventricles (**Figure 17C**). Expression in the ventricles is even more clear at E13.5 and E14.5, when it can be observed that CAV1 is present in the upper part of the ventricles, in the cells that are forming the vasculature of the heart (**Figure 17D-E**).

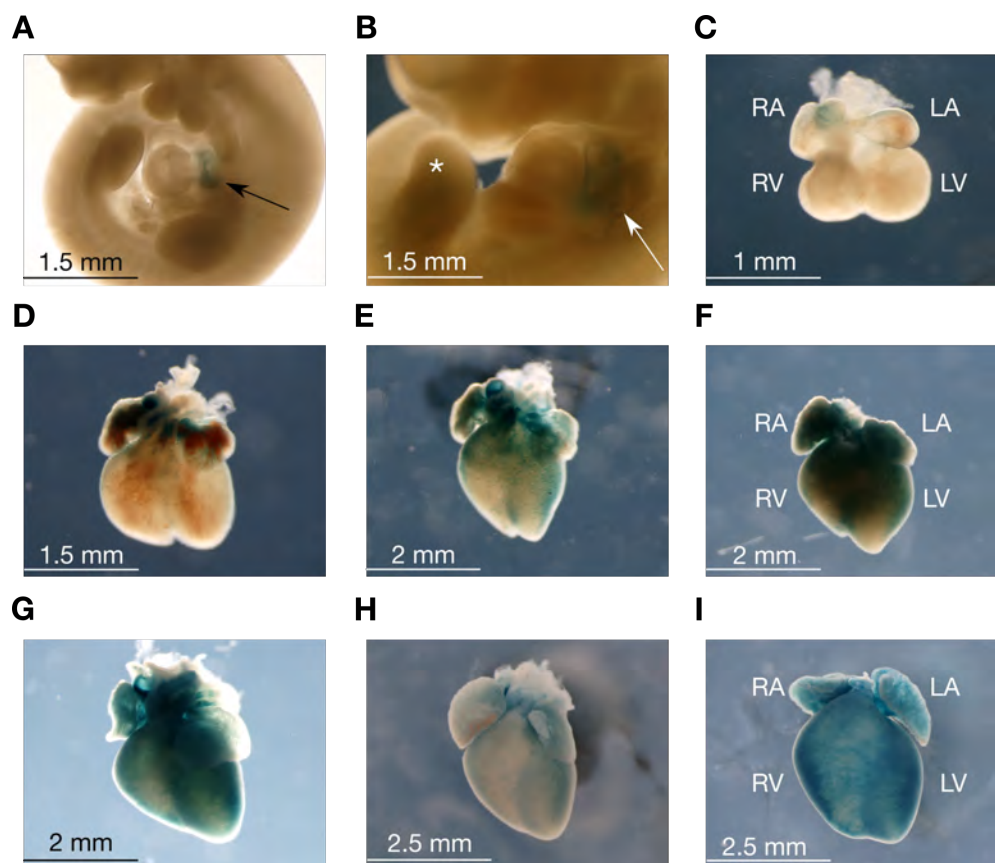


At E15.5 and E16.5 this pattern extends through all the ventricles (**Figure 17F-G**). In E17.5 and E18.5 embryos *LacZ* staining is present over the whole heart. However, it seems that some cells in the ventricle are not positive for the staining (**Figure 17H-I**).



**Figure 16: Expression of the *LacZ* reporter under the control of the *Cav1* promoter in whole embryos**

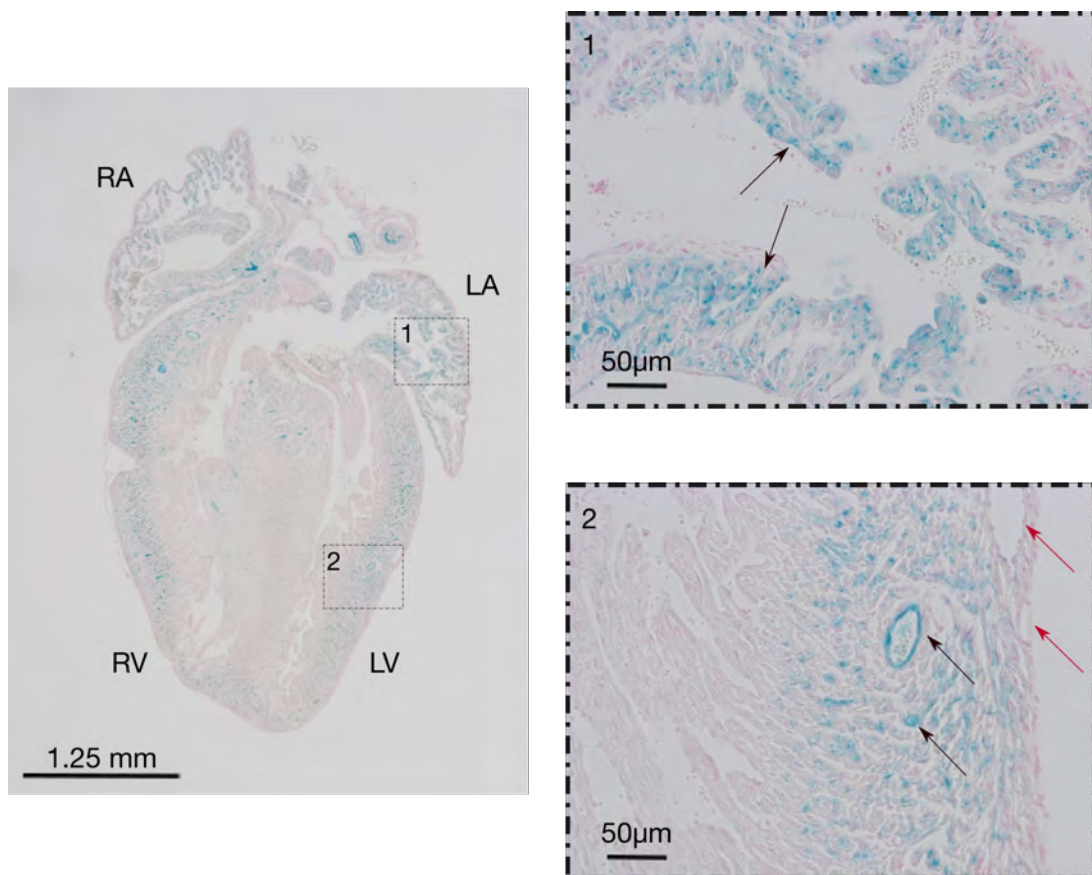
**A)** E9.5, **B)** E11.5, **C)** E13.5, **D)** E14.5 *Cav1LacZ*<sup>+</sup> embryos. No expression is observed at E9.5 (**A**), and gradually expression starts in the heart at E11.5 (**B**) to then be spread to other tissues at later stages such as E13.5 (**C**) and E14.5 (**D**).



**Figure 17: *Cav1* *LacZ* reporter expression during mouse heart development**

**A)** E10.5, **B)** E11.5, **C)** E12.5, **D)** E13.5, **E)** E14.5, **F)** E15.5, **G)** E16.5, **H)** E17.5, **I)** E18.5 mouse embryonic hearts. *Cav1LacZ*<sup>+</sup> reporter signal can be observed in the heart atria at E10.5 (**A**; black arrow) where it can be also observed at E11.5 (**B**; white arrow) and E12.5 (**C**). At E13.5 (**D**) its expression is extended through the ventricles being more evident at E14.5 (**E**), E15.5 (**F**) and E16.5 (**G**). At later stages, *CAV1* reporter activity can be found in the whole heart (**I**). However, there are some cells in the ventricles that seems do not to have *CAV1* expression.

To have a better resolution to study positive cells for *LacZ* staining we performed histological sections from E18.5 hearts (**Figure 18**). CAV1 reporter expression can be observed in cardiomyocytes of the atrial myocardium (**Figure 18.1** black arrows) while in ventricles expression is present in the vascular endothelium, but not in all the vessels (**Figure 18.2** red arrows). Vessels that are localized in the most superficial layer of the ventricular myocardium do not show reporter expression, that is only present in those localized to the interior myocardium (**Figure 18.1** black arrows). As the coronary plexus is differentially localized in the myocardium<sup>161</sup>, our observations suggest that *Cav1* could be differentially expressed between veins and arteries.



**Figure 18: *Cav1LacZ/+* E18.5 mouse embryo heart sections**

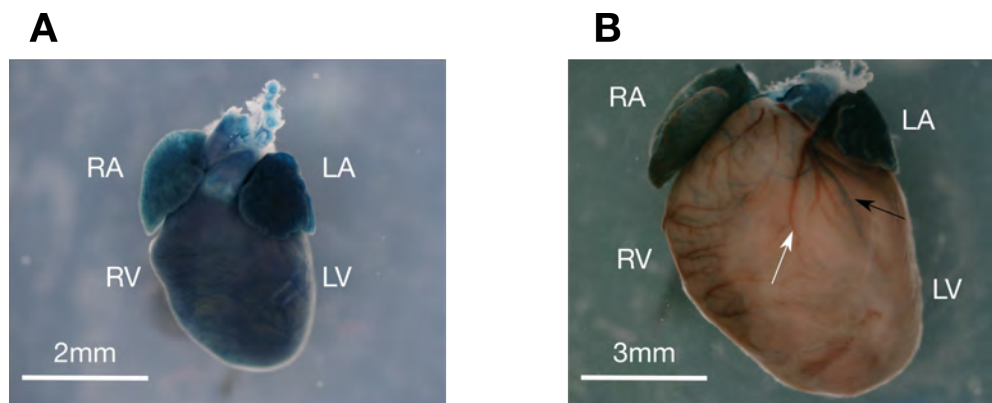
Whole heart section shows that atria have higher levels of *LacZ* reporter staining than ventricles. Higher magnification of the atria (**1**) allows to appreciate how cardiomyocytes from atrial myocardium are expressing the reporter, while in ventricles (**2**) not all cells are positive for *LacZ* staining and this is restricted to endothelial cells lining the coronary vessels. Black arrows mark positive endothelial cells surrounding blood vessels, while red arrows show superficial vessels without positive staining.

Finally, we wanted to analyse CAV1 expression in new born and adult mice. In post-natal mice (P3) the expression is apparently stronger than in E18.5 (**Figure 17I**), and higher in LA than in RA (**Figure 19A**). Adult mice maintain high reporter expression in both atria, being slightly higher in LA than in RA (**Figure 19B**). In addition, expression is



clearly detected in the coronary vasculature although not in all the vessels. In **Figure 19B**, the black arrow points out an example of a vessel stained blue for *LacZ* reporter and the white arrow marks an empty vessel negative for expression that shows a reddish hue due to blood.

These results suggest that, at least in mice, the developing heart is the first site of expression of *Cav1* during development. Expression is initially located in the atria, most probably in cardiomyocytes, and later appears in endothelial cells surrounding vessels of the coronary vasculature. This pattern of expression is strongly suggestive for a putative role of *CAV1* in atrial cardiomyocyte physiology.



**Figure 19: *Cav1* *LacZ* reporter expression in postnatal and adult hearts**

**A)** Day 3 postnatal mouse shows positive reporter expression over the whole heart. **B)** Adult mouse heart, with clear expression in the atria and coronary vessels of the ventricles (black arrow). Other vessels (white arrow) show no expression.

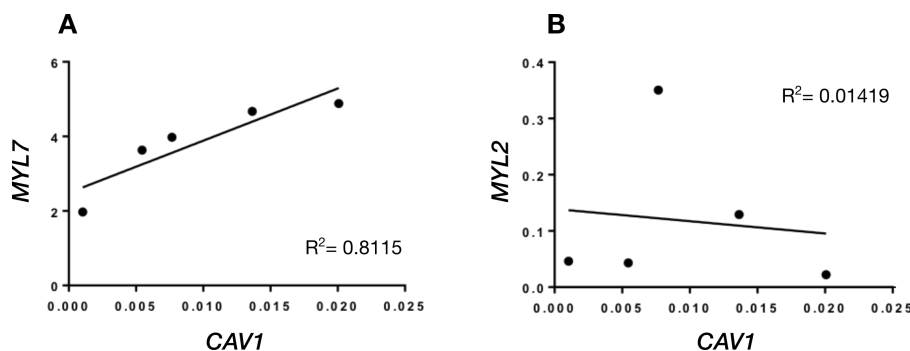
### 3. Functional characterization of *CAV1* hiPSC-CMs

The above results suggest that *CAV1* could have a role during AF. Indeed, in 2014, it was published that Caveolin1 mRNA and protein levels were reduced in atrial tissue from AF patients<sup>162</sup>. These evidences made us wonder which could be the role of *CAV1* in AF, and which are the regulatory mechanisms underlying the association by GWAS of the *CAV1* locus to the disease. In order to explore these issues, and in the absence of patient samples, we moved to a tissue culture system for human cardiac cells, the hiPSC-CMs.

#### 3.1 *CAV1* is mainly expressed in atrial-like hiPSC-CMs

We differentiated the pluripotent control BJ-hiPSCs cell line towards cardiomyocytes following small molecule protocol<sup>110</sup>. Once generated, we purified hiPSC-CMs by magnetic activated cell sorting to isolate cardiomyocytes from the non-cardiomyocyte population (endothelial, smooth muscle, fibroblasts and not fully differentiated cells<sup>163</sup>) that is obtained during this differentiation protocol and processed the cells for RNA

extraction. The differentiation protocol of hiPSCs towards cardiomyocytes also results in a heterogeneous population that includes nodal, atrial and ventricular-like cells populations<sup>164</sup>. To more precisely characterized if *CAV1* was preferentially expressed in some of these cardiomyocyte subtypes, we carried out in parallel RT-qPCR for *CAV1* and atrial (*MYH7* gene coding for MLC2A which is expressed in atrial cells and immature ventricular cardiomyocytes) or ventricular (*MYH2* gene coding for MLC2V mature ventricular cardiomyocytes) markers<sup>110</sup>. *CAV1* expression highly correlates with *MYL7* ( $R^2=0.8115$ ; P-value=0.0369) but not with *MYL2* ( $R^2=0.01419$ ; P-value=0.01419). Therefore, in humans *CAV1* expression is likely to be occurring in atrial-like cardiomyocytes (**Figure 20**) as we have observed in mice.



**Figure 20: Correlations *CAV1* expression with atrial and ventricular markers in hiPSC-CMs**

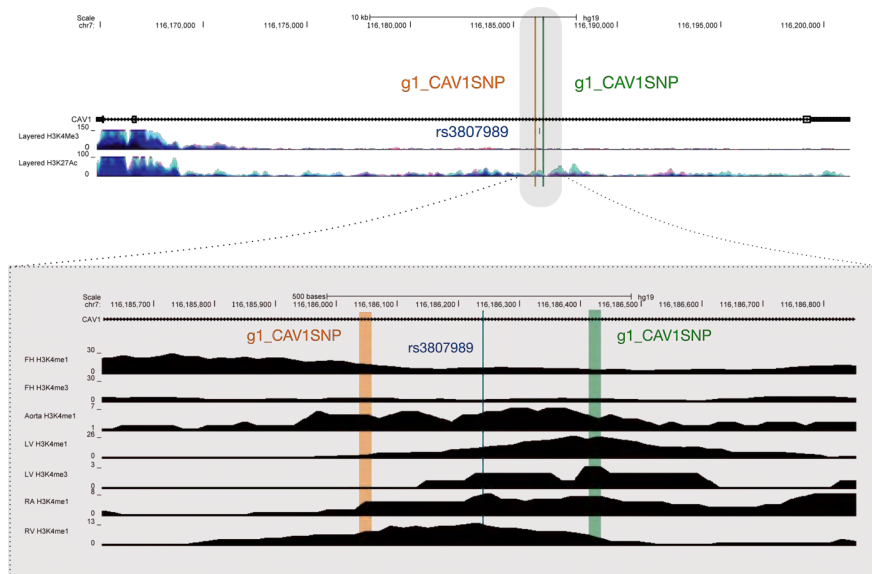
(A-B) Pairwise correlations of BJ control hiPSC-CMs expression levels for *CAV1*, *MYL7* and *MYL2*. Pearson correlation ( $R^2$ ) is indicated for each correlation. A) *CAV1* relative expression levels highly correlate with *MYL7* expression levels ( $R^2=0.8115$ ) B) *CAV1* relative expression levels do not correlate with *MYL2* expression levels ( $R^2=0.01419$ ).

### 3.2 Deletion of 7q31 AF-associated variants in hiPSCs.

After confirming that *CAV1* is expressed predominantly in atrial-like hiPSC-CMs, we decided to use the CRISPR/Cas9 genome editing tool to delete the region containing AF-associated SNP (rs3807989) in hiPSCs to later differentiate them towards cardiomyocytes. In this way, we could determine if the expression of the genes that we have observed that are interacting with our candidate region is modified in its absence. We used two guide RNAs (**Figure 21**) to delete our candidate region of 357bp that including the AF associated SNP. This region contains histone marks for active enhancers (H3K4me1) in heart related tissues (RA, LV and aorta), suggestive of being a regulatory element for the *locus* (**Figure 21**).

After hiPSCs electroporation with the guide RNAs and Cas9 protein, sorting and growing, we obtained 22 independent hiPSC clones, 6 of which were wildtype (WT), another 6 were heterozygous for the deletion and 10 were homozygous for the deletion

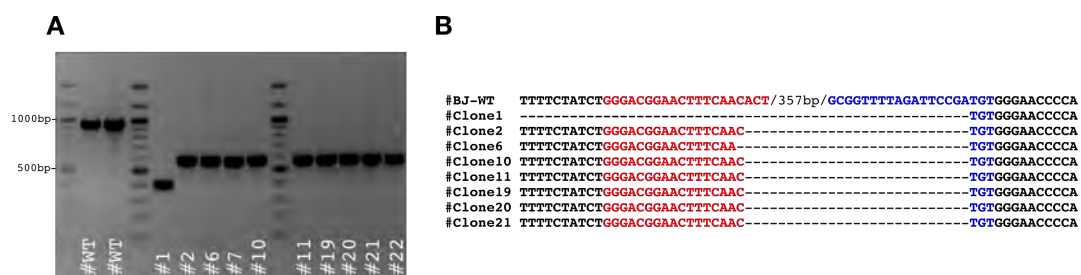
as determined by PCR genotyping. As can be observed in the gel shown in **Figure 22A**, deleted clones show a smaller band of an expected size of 568bp, than WT controls, where PCR amplification results in a 965bp expected band.



**Figure 21: Experimental design for CRISPR/Cas9 deletion at the locus 7q31**

The upper panel indicates the location of the lead SNP associated to AF, rs3807989, in the second intron of the *CAV1* gene and the two flanking gRNAs used to delete the region. The lower panel is a zoom of the region to show the histone marks associated to active enhancer (H3K4me1) that has been mapped to this region in cardiovascular tissues.

Through Sanger sequencing we confirmed that 6 homozygous clones have the same deletion of 377bp, clone6 has a 378bp deletion and clone1 has a larger deletion of 556bp (**Figure 22B**). In addition, we detected that clone 7 and clone 22 had different deletions in each allele meanwhile the rest of the clones showed the same deletion in both alleles.

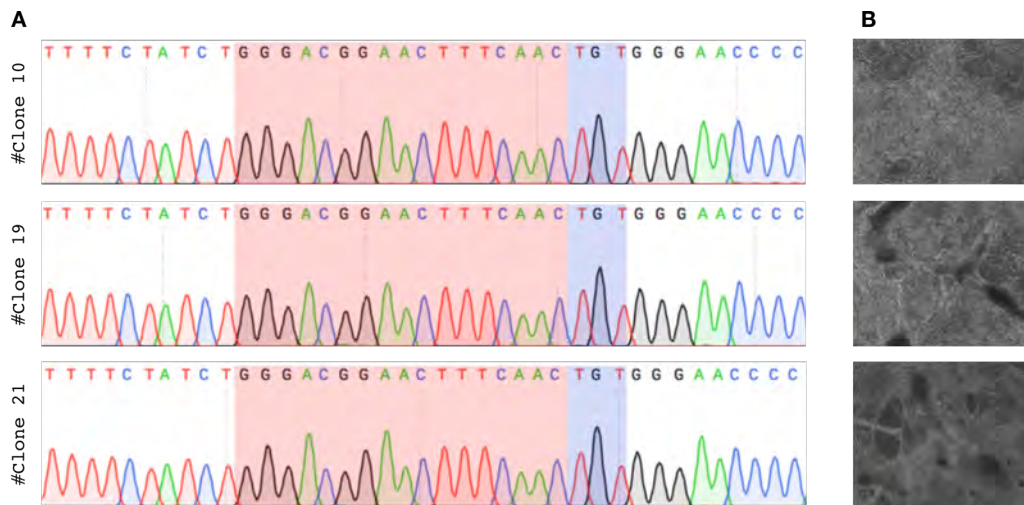


**Figure 22: Genotyping of deleted hiPSC clones**

**A)** Genotyping gel where the first two samples are controls (#WT) that have the expected band size of 965bp. Clones #2 to #22 are positive clones for the deletion with an expected band size of 568bp and clone #1 is also positive for a larger deletion. **B)** Genomic sequence of the positive clones that after Sanger sequencing have the same deletion in both alleles.

We next selected three clones that had the same deletion (clones 10, Clone 19 and Clone 2; **Figure 23A**) for further analysis. These three clones, together with the BJ-hiPSC control line, could be efficiently differentiated to hiPSC-CMs (**Figure 23B**).

Therefore, the deletion does not impair the differentiation capacity of pluripotent cells towards the cardiomyocyte lineage.

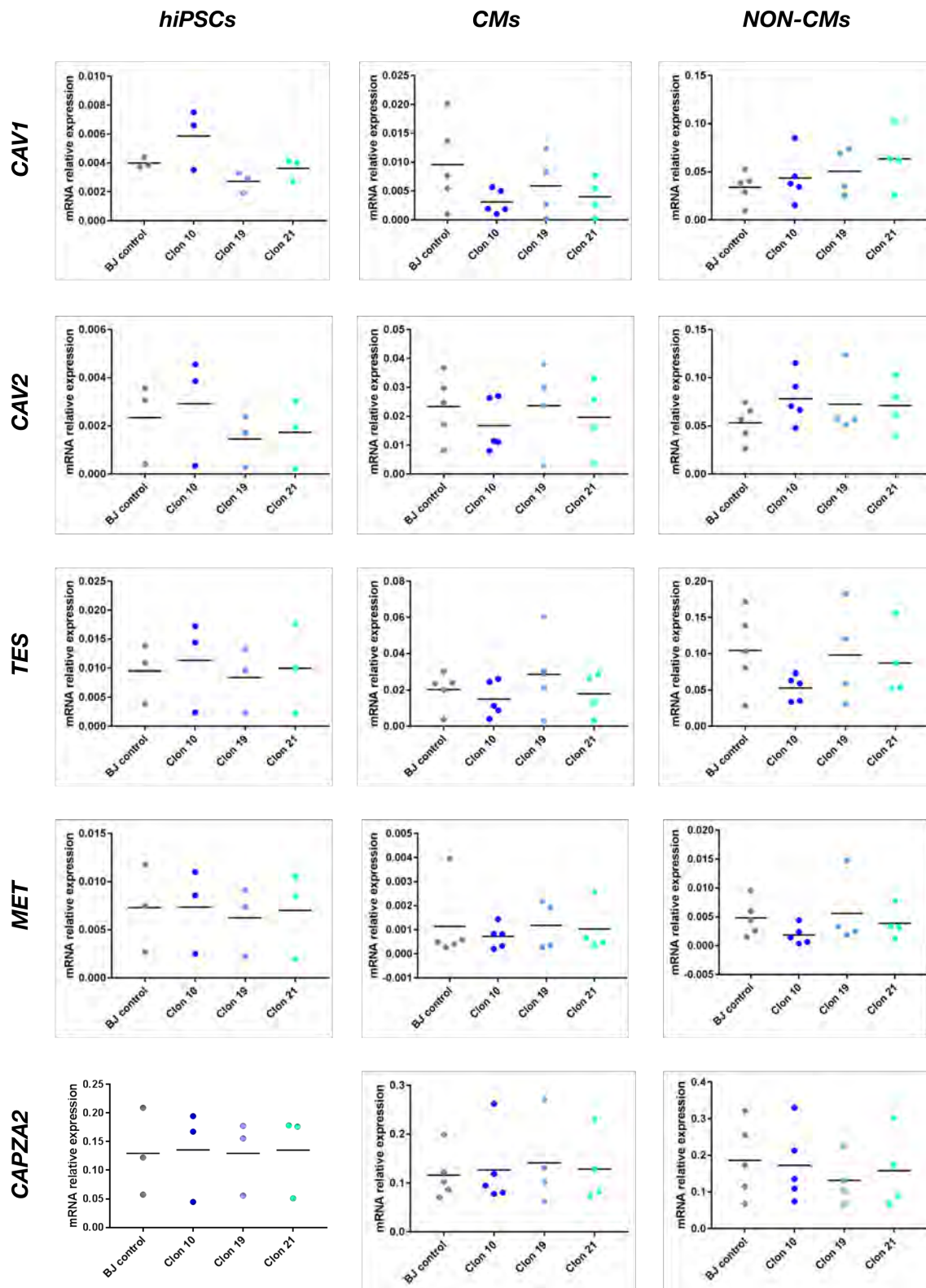


**Figure 23: Selected clones' sequences and successful differentiation to cardiomyocytes**

**A)** Sanger sequence of clones 10,19 and 21 showing that the same deletion is present in all. **B)** Representative images from the supplementary videos 1-3 of successful differentiation of deleted hiPSCs clones to cardiomyocytes

### 3.3 Effect of the deletion of the candidate *CAV1* regulatory region in hiPSC-CMs and non-cardiomyocyte cells.

To investigate if the lack of our candidate regulatory region affects the expression of the nearby genes, we first studied the gene expression of *CAV1*, *CAV2*, *TES*, *MET* and *CAPZA2* genes in hiPSCs before their differentiation. These genes were selected based on the chromatin interaction profiles we observed in mouse atria by 4C-seq, and in human cells by analysis of previous promoter capture Hi-C data (see above). We did not observe any differences in expression of the genes examined by RT-qPCR between BJ-hiPSC controls and isogenic mutant clones by RT-qPCR (**Figure 24**). We next differentiated the four cell lines towards hiPSC-CMs and purified cardiomyocytes 15 days after beating of the cell cultures was observed (approximately 27 day since starting the differentiation protocol). At this stage, we also collected the non-cardiomyocytes. We performed RT-qPCR of our panel of genes in both cell population (**Figure 24**). Although we did not find statistically significant differences, we do observe a tendency for a reduction in *CAV1* expression in deleted hiPSC-CMs as compared to controls. On the other hand, there is a trend for increased expression in non-cardiomyocytes for both *CAV1* and *CAV2*. Despite the lack of significant differences, this data is suggestive of a role of our candidate regulatory region in the regulation of *CAV1* and merits further studies.

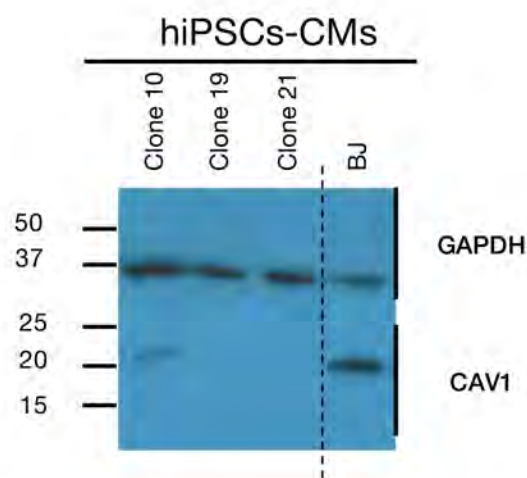


**Figure 24: 7q31 locus gene expression in hiPSCs, hiPSC-CMs and Non-cardiomyocytes populations**

mRNA relative expression levels of genes *CAV1*, *CAV2*, *TES*, *MET* and *CAPZA2* in hiPSCs (n=3), hiPSC-CMs (CMs; n=4-5) and non-cardiomyocytes (NON-CMs; n=4-5) from BJ control cell line and on deleted clones 10,19 and 21. Horizontal black bars indicate mean values. No statistical differences with the BJ control cells were observed after performing unpaired two-tailed Student t-test.

### 3.4 hiPSC-CMs deleted for the candidate AF-related regulatory region have a reduction in CAV1 protein expression

In order to have a more complete picture of the effects of the candidate regulatory region deletion we examine expression levels of CAV1 protein. We have carried preliminary experiments by Western Blot (WB) of purified protein from controls and deleted hiPSC-CMs. As is shown in **Figure 25**, levels of CAV1 protein are highly reduced in all three clones that lack the regulatory element as compared to the control. We will further confirm this finding and also examine the levels of protein expression of other candidates from the region (CAV2, TES, MET and CAPZA2) in the near future.



**Figure 25: Western Blot shows a decrease in CAV1 in hiPSC-CMs deleted clones**

Western blot of protein extracted from hiPSC-CMs of control and deleted clones (10,19 & 21). Deleted clones show a reduction in CAV1 protein meanwhile GAPDH control proteins is stable between them (n=1).

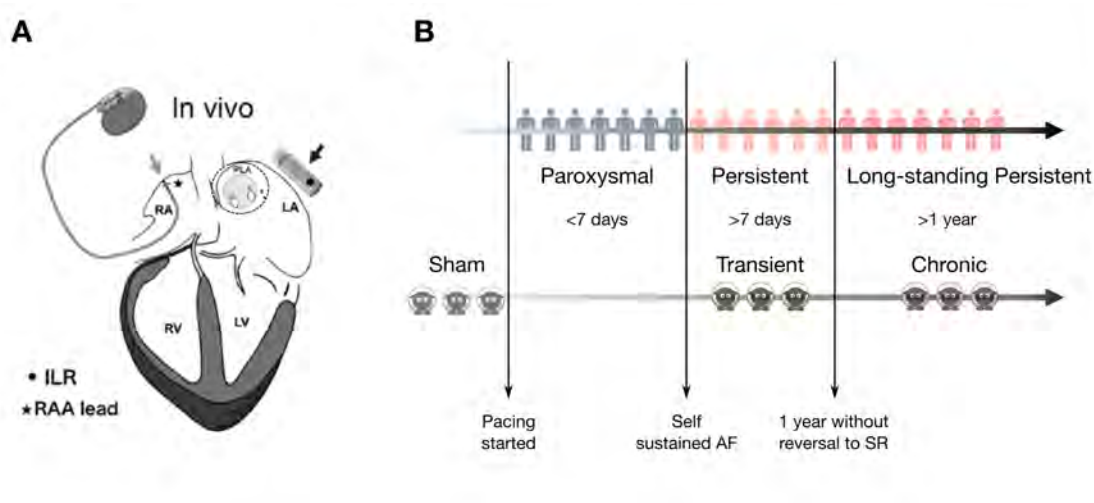
Although, this result needs to be validated, together with the expression trend that we have observed by RT-qPCR, our data suggests that the deleted region may play a role as regulatory element for CAV1 in cardiomyocytes and possibly in other non-myocyte cells.



## 4. Understanding AF disease progression in a sheep model of persistent AF

The different approaches described above have allowed us to start understanding the intrinsic genetic basis of AF. However, since AF is a complex disease where other gene networks and pathways could be involved and deregulated as a consequence of external factors, we drew on a sheep model of induced AF to have a better insight of the underlying molecular mechanism responsible for progression of the disease.

For this study, we took advantage of sheep samples from a long-standing persistent AF model, previously established by Martins et al.<sup>121</sup>, in which self-sustained AF is induced by atrial tachypacing (**Figure 26A**). We have used nine sheep from this study, three controls, three transition sheep (1-2 weeks of self-sustained persistent AF) and three chronic sheep (long standing persistent AF model after a year on self-sustained persistent AF) (**Figure 26B**).



**Figure 26: An induced sheep model for the progression of AF**

**A**) Scheme of the model showing the pacemaker and the lead for tachypacing at the RA and the implantable loop recorder (ILR) in the LA for registering heart activity. **B**) Diagram that summarizes the three groups of sheep used in the study sham (control), transient (persistent) and chronic (Long-standing Persistent).

### 4.1 Posterior Left Atria genomic expression profile during AF progression

We took samples of tissue from the posterior left atria (PLA) wall (**Figure 26**) from the three groups of sheep (control, transition and chronic) to perform a transcriptomic profile by RNA-seq. Principal component analysis (PCA)<sup>165</sup>, that groups samples based on unbiased components that explain the variance of the samples, showed that one transition sample (T3, **Figure 27A**) did not cluster together with the other two transition sheep. Examination of the track records of the original study<sup>121</sup> showed that this sheep

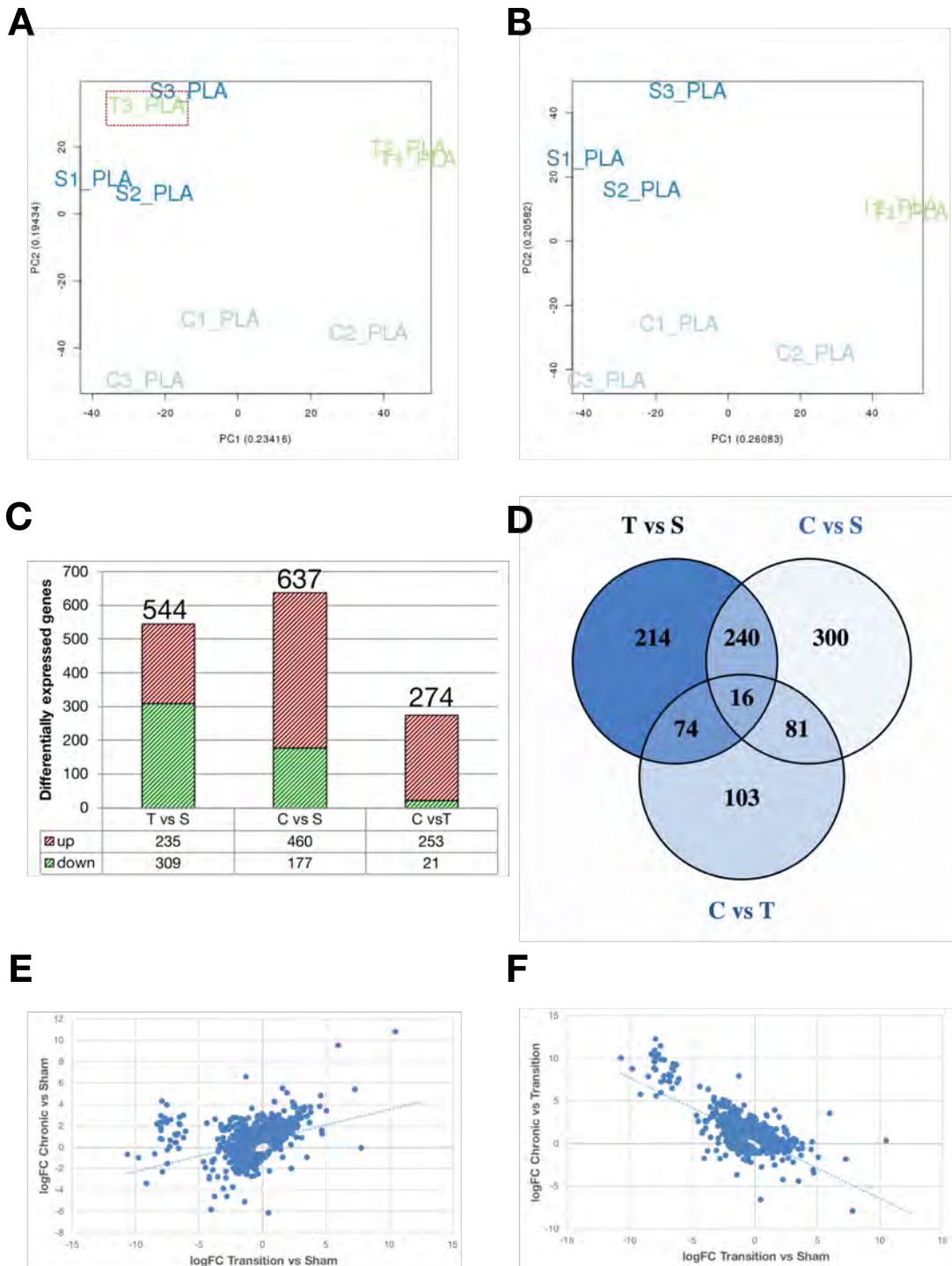
had a problem with the battery of its pacemaker and it may have not worked properly. Therefore, we removed this sample from further analysis. PCA without T3 sheep showed clustering of the sample in the three expected groups: sham, transition and chronic (**Figure 27B**).

In this PLA RNA-seq we detected 12440 transcribed genes combining all samples, of which 1028 were differentially expressed (adjusted p-value <0.05) in at least one of the three comparisons that we performed: transition versus sham (T vs S), chronic versus sham (C vs S), and chronic versus transition (C vs T). This analysis showed us that during the first stage of disease progression (T vs S comparison) 544 genes were differentially expressed, with similar number of upregulated (235) and downregulated (309) genes (**Figure 27C**). One year after AF becomes self-sustained (C vs S contrast) there are 637 differentially expressed genes (460 upregulated and 177 downregulated) (**Figure 27C**).

From those 637 differentially expressed genes 40% (256) were already differentially expressed at the first stage of the disease (T vs S comparison) (**Figure 27D**) meaning that many of the changes that are observed after one year of AF progression were already present very early, in the first week of persistent AF. This is even more obvious when we examine the degree of change (measured as log-fold change) of the differentially expressed genes between transition and sham groups (T vs S) and the chronic versus sham (C vs S) ones (**Figure 27E**).

There is a positive linear relationship between the degree of change initial phases of the disease (T vs S) and those occurring later (C vs S). However, when we compare the degree of change occurring during the shift from transition to chronic stage (C vs T) with the one that happened before (T vs S) we observed that many of those changes levelled out. This is evident for example if we look at **Figure 27F** where there is a group of genes that are downregulated during sham to transition (T vs S) step and upregulated on transition to chronic (C vs T), and this might explain why there are more upregulated (253) than downregulated (21) genes in this comparison. Therefore, in PLA tissue quickly after AF initiation there are many changes that are mostly maintained during AF progression.

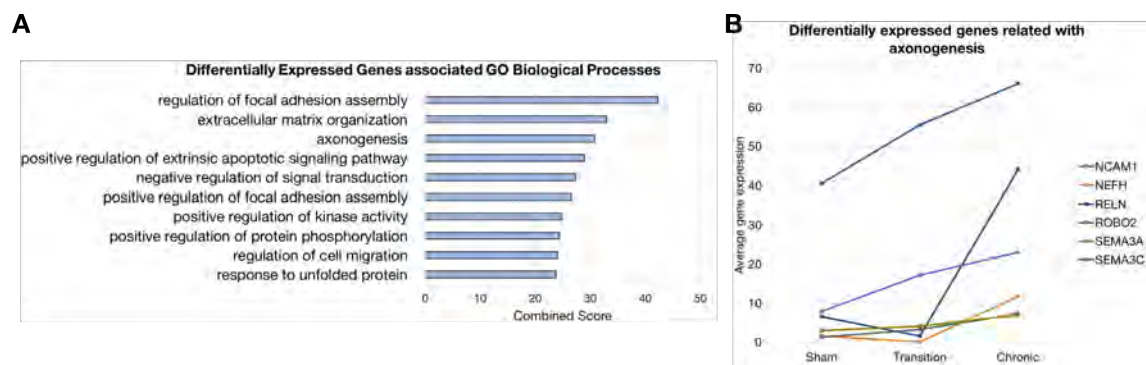




**Figure 27: PLA RNA-seq differentially expressed genes analysis**

**A**) Principal component analysis (PCA) of three sham (S1, S2, S3 dark blue labelled), transition (T1, T2, T3 green labelled) and chronic (C1, C2, C3 light blue labelled) sheep showed that one sheep, T3, cluster together with sham sheep instead of being with the other sheep. **B**) PCA without T3 sheep, all the sheep cluster in the three expected groups: sham (S1, S2, S3), transition (T1, T2) and chronic (C1, C2, C3). **C**) Differentially expressed genes on each one of the contrasts that we have perform to compare the three groups of sheep. Transition vs Sham (T vs S), Chronic vs Sham (C vs S) and Chronic vs Transition (C vs T). **D**) Venn diagram of the differentially expressed genes on each contrast. **E-F**) logFC correlations between differentially express genes.

To gain more insight on the genes and pathways that were altered during AF progression we performed a functional enrichment analysis using Gene Ontology (GO) for Biological Processes using Enrichr<sup>135,136</sup> on the list of differentially expressed genes in at least one comparison (**Figure 28A**). The first two most significant enriched GO terms, regulation of focal adhesion and extracellular matrix organization, are closely related with extracellular matrix remodelling which is known to be induced by AF<sup>18</sup>.

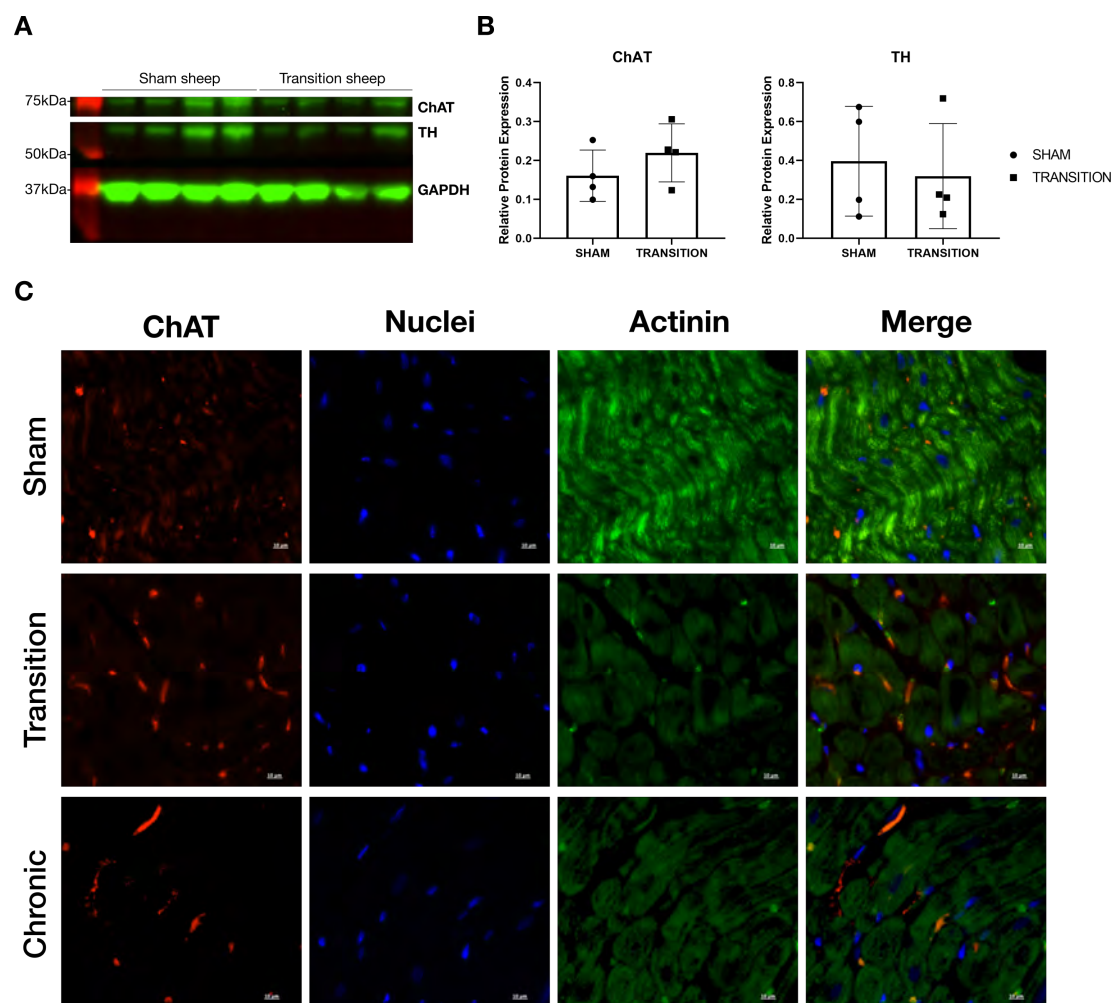


**Figure 28: Gene Ontology analysis of PLA RNA-seq Differentially Expressed Genes**  
**A)** First ten Gene Ontology terms enriched for the list of differentially expressed genes in the PLA RNA-seq ordered by Combined Score significance. **B)** Average gene expression of a representative list of genes belonging to axonogenesis GO term.

Rather unexpectedly, the third most significant GO term was related to axonogenesis that refers to the morphogenesis or creation of shape of the developing axon (**Figure 28A-B**). Within this term we found differentially expressed genes involved in nervous system development (*NCAM1*), neuronal migration (*RELN*), neurofilaments (*NEFH*) or axon guidance (*ROBO2*). Moreover, it includes *SEMA3A* and *SEMA3C*. *Sema3a* is highly expressed during heart development in the subendocardial cells of the trabeculated myocardium, being later restricted to the Purkinje fibers (specialized conduction cells) along the ventricular wall<sup>166</sup>. *Sema3a* overexpression leads to upregulation of beta adrenergic receptors and APD prolongation<sup>167</sup>.

The autonomic nervous system (ANS) plays an important role in the heart regulating heart rate, rhythm and contractility. The heart receives innervation from both sympathetic and parasympathetic systems and proper balance from both inputs is crucial for a proper cardiac function<sup>167</sup>. Hence, we decided to explore if there was any difference at the sympathetic and parasympathetic innervation by studying the adrenergic and cholinergic neurons that are present in the PLA during AF progression. In order to do so, we first performed a WB for choline acetyltransferase (ChAT), an enzyme needed for the synthesis of acetylcholine (the main neurotransmitter of parasympathetic nervous system), and for tyrosine hydroxylase (TH), which is required

for the synthesis of norepinephrine (the main neurotransmitter of sympathetic nervous system)<sup>168</sup>.



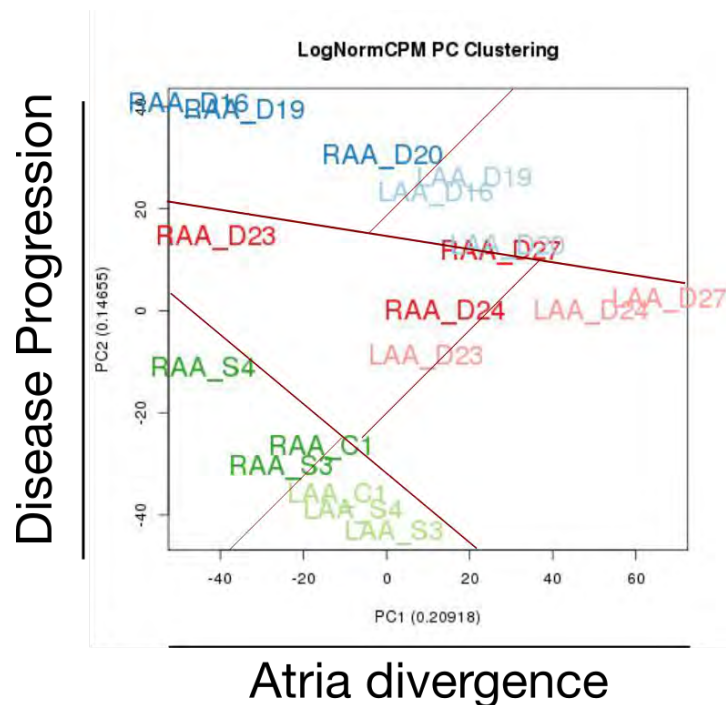
**Figure 29: Adrenergic and cholinergic study on PLA tissue during AF development**

**A**) Western Blot for choline acetyltransferase (ChAT) and for tyrosine hydroxylase (TH) markers of adrenergic and cholinergic neurons respectively, in PLA samples from sham and transition sheep. **B**) Quantification of protein expression relative to GAPDH for ChAT and TH ( $n=4$  sheep in each group). Data are means  $\pm$  s.e.m by unpaired two tail Student t-test. **C**) Immunofluorescent staining of sections from PLA tissue of sham, transition and chronic sheep ( $n=1$  for each group) for ChAT (first column). Nucleus were stained with DAPI (second column). Actinin staining (third column) was used as to label cells cytoskeleton.

However, as can be observed in **Figure 29A-B**, we did not find any significant difference, but only a trend of higher levels of ChAT in samples from transition sheep. Therefore, in order to evaluate if there may be a greater number of cholinergic neurons during AF progression we carried out an immunofluorescent staining for ChAT in sections of PLA tissue from sham, transition and chronic sheep (**Figure 29C**). ChAT staining reveals a probable increase of cholinergic neurons in the transition sections compared with the sham and chronic. However, further experiments need to be done to increase the sample size in and test if it possible to confirm this result.

## 4.2 Transcriptomic and proteomic profiling of isolated atrial cardiomyocytes during AF progression.

To further explore the underlying molecular mechanisms behind AF progression we isolated cardiomyocytes from right and left atria of sham, transition and chronic sheep. We performed a transcriptomic and proteomic profiling of both populations by RNA-seq and liquid chromatography tandem-mass spectrometry (LC-MS/MS) respectively. We performed PCA as above, finding that the first two components of cardiomyocytes RNA-seq data explained more than 30% of sample variability and perfectly separated the samples in left and right (PC1) and sham, transition and chronic sheep (PC2) (**Figure 30**). We also attempted to perform PCA with LC-MS/MS data. However, we could not obtain robust resolution as 4506 peptides were identified compared to the 13262 transcripts detected by RNA-seq.



**Figure 30: Principal Component Analysis isolated cardiomyocytes RNA-seq**

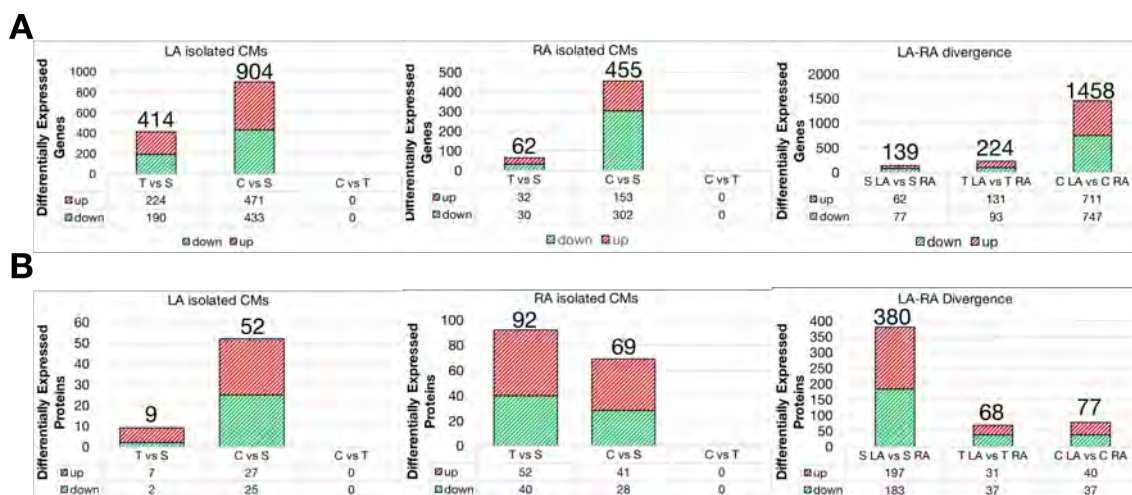
Principal Component Analysis of RNA-seq from isolated cardiomyocytes of left (light colours labeled) and right (dark colours labelled) atria from sham (S4, S3, C1), transition (D23, D24, D27) and chronic (D20, D19, D16) sheep. Principal component 1 (PC1) explains 20% of variability and groups the samples in left and right atria (Atria Divergence). PC2 explains 15% of variability and cluster samples by the state of disease progression (Sham, Transition or Chronic).

As we previously did with samples from PLA tissue, we compared sham, transition and chronic groups for both transcriptomic and proteomic datasets (T vs S, C vs S and C vs T). We also performed additional comparisons to explore differences between atria: sham left versus sham right (S LA vs S RA), transition left versus transition right (T LA



vs T RA) and chronic left versus chronic right (C LA vs C RA). We identified 2479 genes and 581 proteins differentially expressed (adjusted p-value <0.05) in at least one of the comparisons.

During AF progression the number of differentially expressed genes increase in both atria from 414 (T vs S) genes to 904 (C vs S) in the LA and from 62 (T vs S) to 455 (C vs S) in the RA (**Figure 31A**). These numbers also show how there are more transcriptional changes in left than in the right atria cardiomyocytes. Regarding protein changes, we found an increase of differentially expressed proteins in the LA during AF progression. This may suggest that LA suffers more and faster changes than RA. However, the total number of differentially expressed proteins, is bigger in the right (161 proteins) than in the left atria (61 proteins) (**Figure 31B**).



**Figure 31: Differentially Expressed Genes and Proteins on isolated cardiomyocytes**

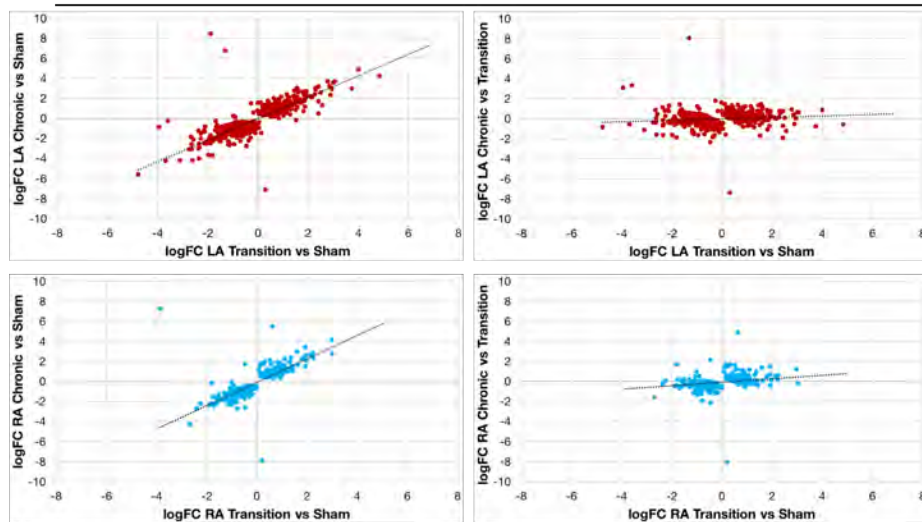
Differentially expressed genes (**A**) and proteins (**B**) in the comparisons performed between AF progression groups at LA (first panel) and RA (second panel) and between the comparison performed to explore atria divergence (third panel). Transition vs Sham (T vs S), Chronic vs Sham (C vs S), Chronic vs Transition (C vs T), Sham Left vs Sham Right atria (S LA vs S RA), Transition Left vs Transition Right atria (T LA vs T RA), Chronic Left vs Chronic Right atria (C LA vs C RA).

Interestingly we did not find any differentially expressed gene or protein when comparing data from chronic versus transition (**Figure 31A-B**), again suggesting that the mayor molecular changes that accompany AF progression occur during the early initial phases of the disease. We also confirmed that when we plot the degree of change, measure as logFC, of each differentially express gene or protein, as we did previously with the PLA RNA-seq data (**Figure 32A-B**). There is a positive linear relationship between the degree of change at initial phases of the disease (T vs S) and those occurring later (C vs S). However, when we compare the degree of change occurring

during the shift from transition to chronic stage (C vs T) with the one that happened before (T vs S) we observed that many of those changes are maintained without big changes in both atria.

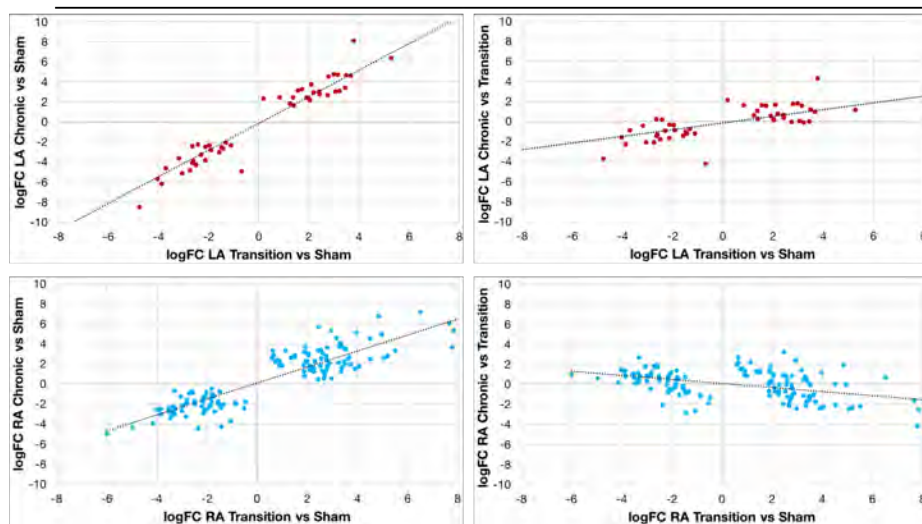
**A**

### Gene changes during disease progression



**B**

### Protein changes during disease progression

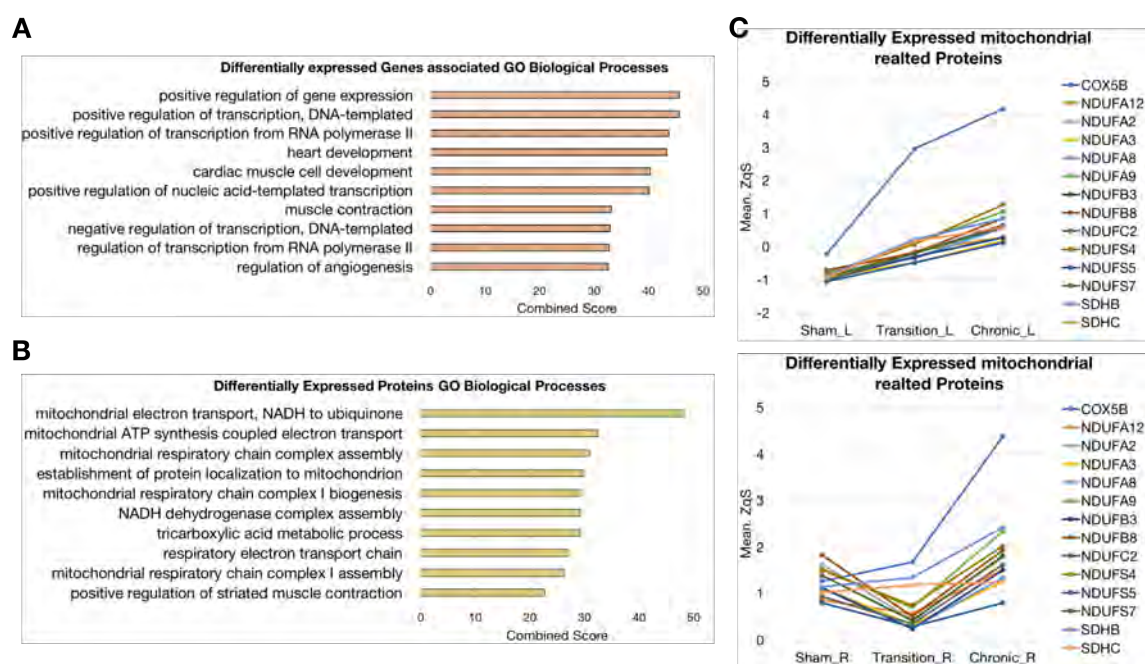


**Figure 32: Correlations of genes and proteins differentially expressed during AF progression**

**A)** Gene correlations **B)** Protein correlation. LA (red dots) and RA (light blue dots) correlations between Transition vs Sham contrast and Chronic vs Sham one, shows how changes that start happening at the early stage of the disease (Transition) are maintained at later stages (Chronic). Meanwhile, the correlation between Transition vs Sham and Chronic vs Transition does not showed any difference suggesting that those changes that were initiated during transition are maintained during chronic AF without changes.

Analysis of atria divergence contrasts reveals an increase on differentially expressed genes during AF progression giving to huge differences on the transcriptomic profile of both atria at chronic stage. Meanwhile proteomic profile suggests that the basal differences between both atria are diluted during AF progression (**Figure 31A-B**).

To further explore the data, we next carried out a GO analysis for Biological Processes by using Enrichr, using as input the list of genes and proteins differentially expressed in at least one of the above comparisons, as previously with the PLA RNA-seq data. The main biological processes that showed up from the list of differentially expressed genes are related with regulation of gene transcription (**Figure 33A**) which includes many transcription factors such as *ATF1*, *ATF2*, *PITX2*, *MEF2A* or *TBX5*, among others. In addition, there are GO terms related with heart and cardiac muscle development containing genes such as *RARB*, *NKX2-5*, *MYOM1*, *TTN* or *MYL2*. A more comprehensive and detailed analysis of this data is currently underway in the lab.



**Figure 33: Gene Ontology of genes and proteins differentially expressed in LA and RA isolated cardiomyocytes**

**A)** First ten Gene Ontology (GO) terms enriched for the list of differentially expressed genes in the RNA-seq data from LA and RA cardiomyocytes, ordered by combined score significance. **B)** First ten GO terms terms enriched for the list of differentially expressed proteins in the LC-MS/MS data from LA and RA cardiomyocytes, ordered by combined score significance. **C)** Average expression levels of selected mitochondrial proteins that are upregulated during AF progression, particularly in LA cardiomyocytes.

On the other hand, from the list of differentially expressed proteins nine out of the first ten most significant GO terms for biological processes are linked to mitochondria (**Figure 33B**). Most of the subunits of the mitochondrial respiratory chain complex I

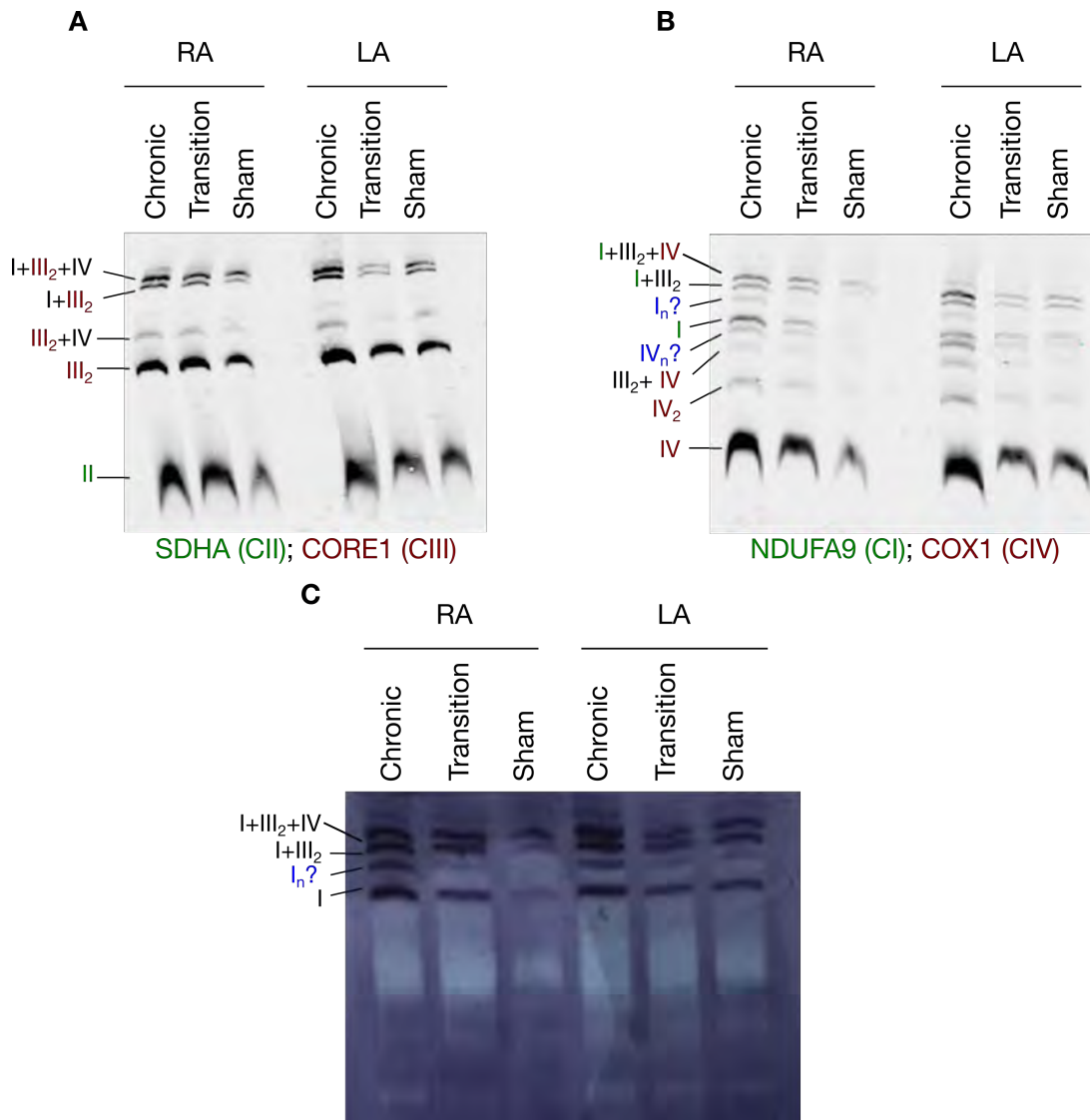
(NADH-ubiquinone oxidoreductase, CI) are differentially expressed (E.g., NDUFA9, NDUFA8, NDUF8, NDUF5, NDUF4), although proteins from other complexes are also altered, such as COX5B, from complex IV (cytochrome c oxidase, CIV) and SDHC and SDHB, from complex II (succinate-quinone oxidoreductase, CII). Interestingly, all these proteins are overexpressed in cardiomyocytes from both atria during AF progression although this trend is much more evident in LA cardiomyocytes (**Figure 33C**). No differentially expressed proteins were found for components of complex III (ubiquinol-cytochrome c oxidoreductase, cytochrome bc<sub>1</sub> complex, CIII) or complex V (ATP synthase, CV).

Mitochondrial function requires correct synthesis, transport and assembly of proteins and cofactors of the electron transport chain into functional complexes and supercomplexes (SCs)<sup>169</sup>. To explore whether the upregulation of mitochondrial proteins, mainly from CI, alters this process we carried out a blue-native gel electrophoresis, that allow separation of SCs and complexes in native condition allowing so the analysis of the electron transport chain organization. We performed this analysis with native mitochondrial proteins from LA and RA cardiomyocytes of sham, transition and chronic sheep. As can be observed in **Figure 34A**, immunodetection of CII and CIII, using antibodies for SDHA and CORE1 respectively, shows how these are distributed among well characterized macromolecular structures (II and III<sub>2</sub> complexes; and III<sub>2</sub>+IV, I+III<sub>2</sub> and I+III<sub>2</sub>+IV supercomplexes). Moreover, it can be seen that meanwhile individual complexes (II and III<sub>2</sub>) do not seem to change, there is a clear trend within the supercomplexes to increase along disease progression.

In a similar way, immunodetection of CI by NDUFA9 antibody and CIV by COX1 (**Figure 34B**) shows the correct formation of complex CI and complex IV, IV<sub>2</sub> and the supercomplexes previously shown in **Figure 34A** (III<sub>2</sub>+IV, I+III<sub>2</sub> and I+III<sub>2</sub>+IV supercomplexes). In this case, we observe a clear trend to increase in the formation of both individual complexes and supercomplexes along disease progression. Most interestingly, we detect a supernumerary band positive for COX1 (IV<sub>n</sub>), typically found on heart mitochondria extracts<sup>170</sup>, and another one positive for NDUFA9 (I<sub>n</sub>) that has not been previously identify in the literature. To study if this new I<sub>n</sub> was functional or not, we performed a blue-native gel electrophoresis followed by in-gel activity assay (**Figure 34C**). This assay revealed that all the complexes and supercomplexes that we identify as containing CI (as detected by anti-NDUFA9 antibody) have NADH dehydrogenase activity including the supernumerary band.



Therefore, during AF progression, the mitochondrial supercomplexes that form the oxidative phosphorylation system (OXPHOS) are increased along with an increase of complex I and IV. This finding has a great relevance on the understanding of underlying mechanism behind AF development. They are pointing out a possible cause for an increase in oxidative stress during AF progression, that could be the result of cardiomyocyte work overload that induce an overexpression of the subunits of the respiratory chain complexes to produce more energy.



**Figure 34: Changes to mitochondrial supercomplexes during AF progression in isolated atrial cardiomyocytes**

Immunodetection of (A) CII by SDHA and CIII by CORE1, and (B) CI by NDUFA9 and CIV by COX1, after blue-native gel electrophoresis with native protein from LA and RA isolated cardiomyocytes of sham, transition and chronic sheep. The presence of the novel supernumerary bands positive for CI is indicated in (B) in purple lettering and followed by a question mark. (C) Blue-native gel electrophoresis followed by in-gel activity assay reveals that all the complexes and supercomplexes identified have NADH dehydrogenase activity due to the presence of CI including the supernumerary band for CI. (All blots are representative of the result obtained in three biological replicates).

Therefore, by using a tachypacing AF induced sheep model we have unveiled new genetic networks and molecular pathways that are affected during AF progression. PLA transcriptomic profiling has suggested that during AF development there is a remodelling of the PLA innervation of the autonomous nervous system. Moreover, proteomic and transcriptomic changes on left and right isolated cardiomyocytes have showed us that during AF progression there is an overexpression of mitochondrial complexes and supercomplexes, as well as, a significant remodelling of gene transcription regulation.

## DISCUSSION

---

*Intelligence is the ability to adapt to change.*

Stephen Hawking



## 1. Understanding genotype modulation in common human traits and diseases

In 1911 Wilhelm Johannsen coined the terms gene, phenotype and genotype. He described the genotype as the total sum of genes, the “elements” in the gametes that were being studied by modern Mendelian researchers at that time, and phenotype as the appearance of an organism<sup>171</sup>. Since that date those concepts have remained with us. Nowadays, we have a broader view of these terms and it is accepted that the phenotype is the set of characters of an organism that can be observed, that are the result of the action of the genotype modulated by the environment. Commonly, these ideas have been summarized in the following simple equation:

$$\text{Phenotype} = \text{Genotype} + \text{Environment}$$

The study of monogenic diseases has established direct associations between a phenotype (the disease) and a genotype (the causal gene), with little or no contribution of the environment in some cases. However, common and multifactorial human diseases, such as AF, can be influenced by multiple extrinsic factors as well as numerous genetic variants each one with small phenotypic effect. This has made their genetic study extremely difficult<sup>106</sup>. Since 2001, when the first draft of the human genome was released<sup>172,173</sup>, novel genetics techniques and approaches have rapidly evolved allowing new analysis such as GWAS that have helped to identify novel common variants associated with common diseases<sup>174</sup>. However, understanding how those variants contribute to a complex phenotype is not obvious, especially when the vast majority of them are located in non-coding regions of uncertain function<sup>175</sup>. Therefore, it is still challenging to unveil their role and to identify the causal genes behind the disease.

## 2. The 3D genome structure of AF associated *loci*

Since non-coding regions can harbour regulatory elements that are able to modulate gene expression, we have hypothesized that AF associated variants could be located in regulatory elements and therefore be modifying nearby or long-distance gene transcription thanks to chromatin folding and looping. To have a better insight on putative regulatory regions and their target genes with a predicted role in AF, we have studied the genomic landscape of several AF-associated variants.

4C-seq assays, on mouse atria and ventricles of regions syntenic to those that in the human genome contain AF associated SNP, captured significant interactions from five variant-containing regions in at least one of the tissues analyzed. Three of them were

located near transcription factor coding genes (*Pitx2*, *Prrx1* and *Zfhx3*) which make them very interesting candidates as AF causal genes since they will be regulating downstream cascades of genes potentially involved in AF. Indeed, it has been shown how specific loss of *Pitx2* in the atria (using an *NppaCre<sup>+/+</sup>Pitx2<sup>-/-</sup>* mouse model) modifies a large number of other candidate AF GWAS-associated genes such as *Zfhx3*, *Syne2*, *Kcnn3*, *Ilf6r*, *Wnt8* or *Cav1*<sup>50</sup>. Our 4C-seq data from the syntenic 4q25 locus in mouse has confirmed previous interaction data with *Pitx2c* promoter. However we could not confirm the previously reported interaction with the *Enpep* promoter<sup>78</sup>. This can be due to 3C assay (as used in that case) can be a more sensitive technique than 4C-seq since it is directly asking if two regions in the genome are interacting between each other. Meanwhile, 4C-seq can have lower resolution in finding significant interactions if there is a change on chromatin conformation between left and right atria (that were pooled together). Therefore, *Enpep* interactions in the left atria could be diluted and levelled out by the absence of interaction in the right. However, we captured interactions of our viewpoint with the lncRNA *Playrr*, that have been described to be expressed only in the right side but not in the left because *Pitx2* represses its expression mediated by asymmetric left/right chromatin topology<sup>141</sup>. Therefore, 4C-seq from left and right separated pools would help to bring more light on differential interactions and chromatin structure at this locus in which an alteration mediated by a variant can have a key role in left-right differences in atrial physiology and hence AF development.

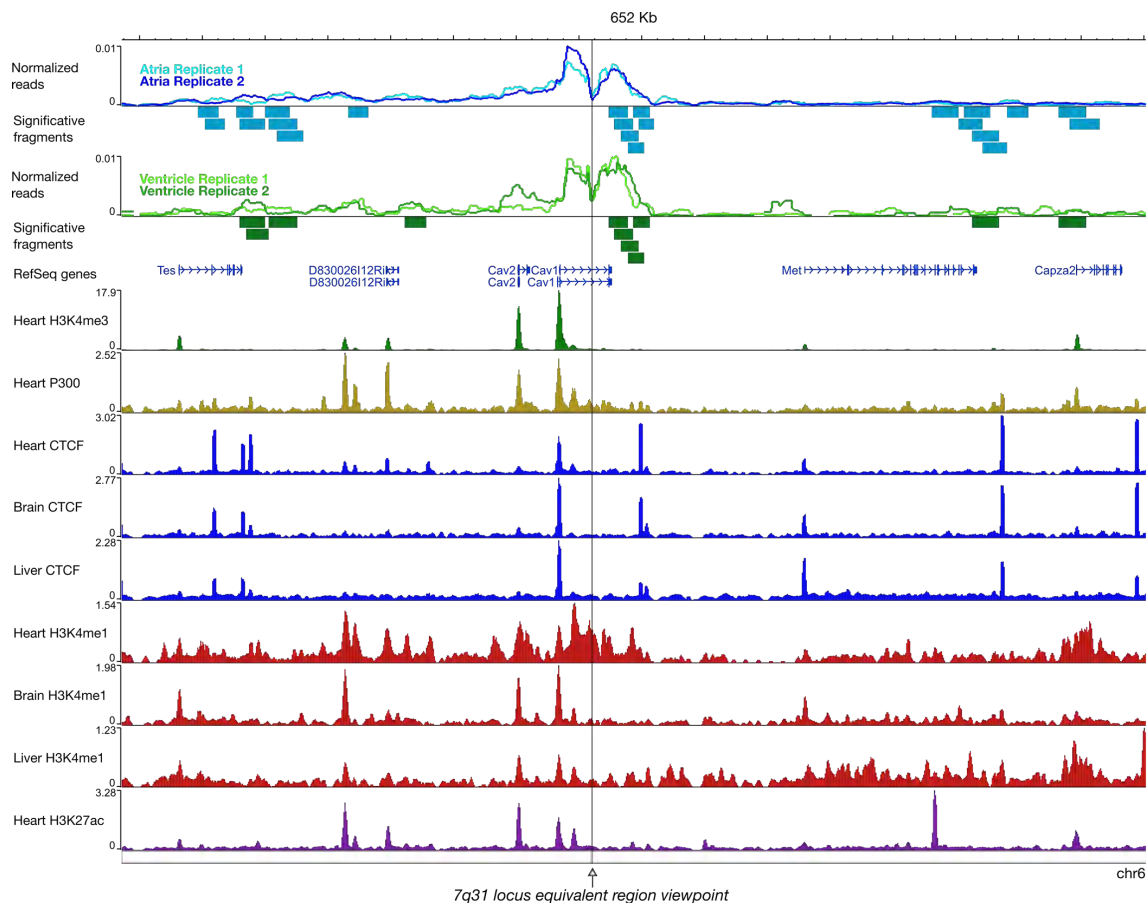
Regarding *PRRX1* locus, it has been recently published that AF associated variant rs577676 is included within an enhancer region whose activity is altered by this SNP in mouse HL-1 cell line<sup>55</sup>. Moreover, this enhancer is able to interact in human cardiac fibroblast with *PRRX1*, whose transcription is decreased in LA of individuals homozygous for the risk allele and therefore this variant could be altering *PRRX1* expression through this enhancer<sup>55</sup>. In our hands, 4C-seq from the region in mouse syntenic to that including rs577676 in humans did not show direct interaction with *PRRX1*, but did interact with other neighboring genes (*Mettl11b*, *Kifap3* and *Fmo2*). However, since this locus is not very well conserved between mouse and human there could be species-specific differences in regulatory elements. Hence, functional studies in human cells would be necessary to test the role of our candidate regulatory elements from the *PRRX1* locus on other genes in the region and explore their potential role in AF.

Although different functional approaches in model systems have established *PITX2* and *PRRX1* as highly likely AF causal genes, there is still a high fraction of AF inheritance to

be explained. In this regard, some of the *loci* that have not been as well explored could contain several genes with a role on AF. This could be the case of *SYNE2* and *ESR2* at 14q23, or *TES*, *CAV1*, *MET* and *CAPZA2* at 7q31. Throughout this work we have focused on the significant interactions captured from the candidate region at *Cav1*. Interestingly, detailed analysis of 4C-seq data shows how, both in atria and ventricle samples, all captured interactions are near to CTCF binding site identified in heart and also in other tissues samples by the Mouse ENCODE consortium<sup>176</sup> (**Figure 35**). This suggests that in this *locus*, there is a stable 3D conformation of the chromatin that could be mediated by CTCF and that locates in close interaction our putative regulatory region with genes of the *locus*. One of the interactions captured by 4C-seq, in atria and ventricles occurs between the viewpoint and *Cav1*'s 3' UTR and nearby region (around 25kb). 3' UTRs can contain important regulatory elements to modulate gene function, determine protein levels and even they can act like long noncoding RNA<sup>177</sup>. Moreover, it has been described that 3'UTRs can contain enhancer elements that interact with gene promoters by chromatin looping to regulate gene expression<sup>178</sup>. *Cav1*'s 3'UTR region is enriched for active enhancer histone marks (H3K4me1) in the heart (**Figure 35**). Therefore, this 3'UTR could be regulating *Cav1* gene expression in the heart through the interaction with regulatory elements and mediated by chromatin looping. Further experiments are needed to support this regulatory condition. Moreover, we also captured interactions from our viewpoint with the 3'UTR of the *Tes* and *Met* genes, which could also be involved in their regulation. Hi-C data from hiPSC-CMs also confirms an interaction between the promoter of *CAV1* and *TES* and *MET* genes<sup>155</sup>, suggesting that there is complex chromatin conformation of this *locus* both in human and mouse. Hence, *CAV1*, *TES* and *MET* could share similar regulatory mechanism in the heart. However, little is known about their co-expression and regulation, what could help to identify novel candidate genes with a role on AF.

*TES* and *CAV1* are co-expressed in melanoma cancer progression<sup>179</sup> and *MET* and *CAV1* in basal-like breast cancers<sup>180</sup>. In the heart they have been studied independently. Recently, it was shown that *TES* is expressed in the heart where it seems to have a protective role against cardiac hypertrophy and fibrosis by impairing Angiotensin II effect through a blockade of calcineurin-NFAT axis<sup>181</sup>. *MET*, together with its ligand HGF, have multiple roles in the cardiovascular system depending on the cell type: it is important for autophagy regulation in cardiomyocytes, induces endothelial cells proliferation and migration, antagonizes with TGF $\beta$ 1 and angiotensin II in fibroblast preventing fibrosis, and influences the inflammatory response of macrophages

providing protection against atherosclerosis<sup>182</sup>. Moreover, *Met* deletion in cardiomyocytes impairs their protection against oxidative stress and normal cardiac function<sup>183</sup>. *CAV1* expression in the heart has been detected<sup>159,160</sup> and *Cav1* knockout mouse have important cardiac defects<sup>157</sup>. However, its precise role in cardiac cells is still unclear. It has been suggested that *CAV1* has a protective role against cardiac fibrosis through negative regulation of Smad, PI3K/AKT, MAPK and JNK signalling pathways<sup>184</sup>. The protective role of *TES*, *MET* and *CAV1* against fibrosis, one of the main changes during AF progression that provides an excellent substrate (source that contribute to conduction alterations that facilitate AF maintenance) for AF sustenance<sup>185</sup>, makes them as candidates to have a potential common function during AF and therefore they could be under similar regulatory mechanisms. Further experiments are needed to have a better understanding of the regulation of this interesting genomic region.



**Figure 35: Genomic interaction landscape, epigenetic marks and transcriptional regulation of the 7q31 syntenic region in the mouse**

Above the Ref-seq track, 4C-seq normalized reads profile and significant interacting fragments in atria and ventricles are shown. Below, transcriptional (p300 and CTCF) and epigenetic marks (H3K4me3, active promoters; H3K4me1, active enhancers; H3K27ac, active enhancers and promoters) from adult heart, brain and liver from the Mouse ENCODE Consortium<sup>176</sup>. Genomic window shown corresponds to mm9; chr6:16978957-17631569.



### 3. CAV1 is expressed in the mouse heart from embryos to adults

GWAS together with our 4C-seq data has provided us a list of genes that could be associated to AF but have been poorly studied in this context, *CAV1* is one of them. It has been described that AF patients have a decreased level of *CAV1* mRNA and protein in the atria<sup>162</sup> but this study, on a rather small cohort, was limited and there are no more studies about its potential role on AF. Therefore, our studies are the first to address the relevance of *CAV1* in cardiomyocytes in relation to AF. *Cav1*<sup>-/-</sup> mice present right ventricular dilatation, decreased systolic function, and an increase in heart-to-body weight ratio<sup>157</sup>, and they finally develop dilated cardiomyopathy and pulmonary hypertension or cardiac hypertrophy<sup>63</sup>. Moreover, *CAV1* is a key component of caveolae membrane invaginations where multiple ion channels relevant for AP have been localized<sup>156</sup>. Indeed, it has been shown how *CAV1* can negatively regulate Kir2.1 channel which is involved in  $I_{K1}$  current needed for the resting membrane potential<sup>186</sup>. Also, it has been shown how *CAV1* can modify  $Ca^{2+}$  channels in arterial smooth muscle cells<sup>187,188</sup>. Despite these evidences point to a relevant role of *CAV1* in the heart, little is known about when and where it is expressed in the heart. We therefore performed a characterization of *Cav1* expression during heart development and adult mice using a knock-in allele in mice where expression of the *LacZ* reporter acts as a read out of the endogenous gene. We observed that *LacZ* staining is first present in the atria at E10.5, when the four chambers of the heart are well-defined<sup>189</sup>. However, it has been previously shown by whole mount in-situ hybridization that E9.5 embryos express *Cav1* in the heart<sup>158</sup>. This discrepancy can be as a consequence of the use of our model in heterozygosis (to be able to maintain both reporter and *Cav1* expression). Therefore, if at E9.5 there are low levels of *CAV1* protein we can be missing them since we are tracking just half of the activity of *Cav1* promoter and we may not have enough signal intensity. Hence, immunohistochemical staining at E9.5 would help to solve this disparity but in any case, it is clear that *CAV1* is expressed in the heart at early stages in development.

Moreover, we have observed that *Cav1* reporter expression in the atria is sustained during heart development and this expression spreads to the vascular plexus of the ventricles at later stages (E12.5-E13.5) to finally be expressed in all the atrial myocardium and in some coronary vessels in the ventricles. Future immunohistochemical staining will help to confirm *CAV1* expression in atrial cardiomyocytes and to distinguish which vessels express *CAV1*, since our histological

sections suggest that *CAV1* could be expressed in the coronary arteries but not in veins. In this way we could better understand the role of *CAV1* in the heart development and consequently understand the potential relation to AF. It is also interesting to notice that there are other AF associated genes that have a role during development. Thereby, *Pitx2* is needed for left/right determination during embryonic development being finally restricted to the left side in different organs like the heart<sup>190</sup>. Moreover, it has also been recently shown that the epithelial-mesenchymal transition process that is also required for left/right asymmetry, and needed for proper heart looping, is mediated by *Prrx1a* in the zebrafish, *PRRX1* and *SNAIL1* in the chick and *SNAIL1* but not *PRRX1* in the mouse<sup>191</sup>. Hence, it would be interesting to unveil if their association to AF could be due to their role during development in humans or because of their expression in cardiac cells of an adult organism. It is already known that *Pitx2* it is also relevant for cardiac function in adult mice and its atrial specific deletion causes APD abnormalities and resting membrane potential alterations<sup>192</sup>. Hence, these evidences increase its potential role in AF.

#### **4. Potential regulation of *CAV1* expression in human cardiac cells by a novel regulatory element**

AF associated non-coding variant rs3807989, located at intron 2 of *CAV1* gene in 7q31 *locus*, could be inside a regulatory element and therefore modifying nearby or distant gene transcription thanks to chromatin folding and looping. 4C-seq in mouse and Hi-C human cells data have shown us that at this *locus* there is a complex chromatin conformation that could be favoring enhancer-promoter interactions. Hence, we tested its regulatory role by deleting a 357bp region that contains the lead SNP, by CRISPR/Cas9 mediated genome editing<sup>193</sup> in hiPSCs. Expression levels of the potential target genes (*CAV1*, *CAV2*, *TES*, *MET* and *CAPZA2*) did not change in hiPSCs deleted clones. However, in hiPSC-CMs and non-cardiomyocyte population, *CAV1* expression could be modified by the deletion of this region. Despite changes do not reach statistical significance there is a trend for *CAV1* to be downregulated in hiPSC-CMs. The large variability between replicates that we observed in these experiments could be the consequence of heterogeneity of the protocol itself. The result of the standard differentiation of hiPSCs towards hiPSC-CMs is usually an heterogeneous population of nodal, atrial and ventricular-like cardiomyocytes<sup>116</sup>, whose proportions are not possible to easily control and can differ even between cells differentiating the same day with the same reagents in parallel. Therefore, since *CAV1* is specific to atrial cardiomyocytes and its expression correlates with atrial markers in hiPSC-CMs, the

effect of this deletion on *CAV1* could be masked in a heterogenous culture. More specific and directed differentiation experiments towards atrial hiPSC-CMs would help solve this problem<sup>115</sup>. On the other hand, WB assay from hiPSC-CMs deleted clones supports the trend of reduced *CAV1* expression in the absence of the regulatory region. More experiments are currently being done to increase the sample size to further confirm this result.

Regarding 7q31 there is also an interesting observation to be considered. The two possible alleles of the *CAV1* rs3807989 SNP have been independently associated to different cardiac electrical traits, namely AF and the duration of the PR interval (measure of the period that extends from the atrial and ventricles depolarization). While the G-allele at this position is associated to increased risk for AF<sup>49,71</sup>, the A-allele is associated to a prolonged PR interval<sup>194–198</sup>. Further confirming this observation, rs11773845 that has recently associated to AF<sup>71,73,75</sup> and is located 5kb upstream rs3807989, has also this double allele association to AF and the PR interval<sup>199,200</sup>. These two variants are in linkage disequilibrium (they are inherited together), therefore both variants have the same probability to have a role in AF. This double and reverse association of the two possible alleles of a single variant to AF and PR interval has been also described for *SCN5A* and *SCN10A* *loci*<sup>194</sup>. PR interval has been considered as a predictor of increase AF risk<sup>201</sup> and even more as an intermediate phenotype for AF<sup>194</sup>. Therefore, this opposite allele effect is unexpected; however, the way through which conduction abnormalities could affect AF is complex and not well known. Hence, future studies are needed to unveil the role of these variants. Understanding the underlying mechanisms that link a given variant with its disease will help to a better diagnosis of patients<sup>202</sup>.

## **5. Underlying molecular mechanisms changes during AF progression**

Since candidate regions studies give us only information about how a few *loci* could contribute to AF we have also worked with a tachypacing AF induced sheep model, previously established by Martins et al.<sup>121</sup> to have a broader view of the mechanisms underlying AF. Whole tissue transcriptomics together with transcriptomic and proteomic profiling of isolated atrial cardiomyocytes of three different conditions (sham, transition and chronic) have revealed that most of the changes identified after one year of disease progression (chronic condition) have already taken place after changing in the first week of the disease (transition condition). Therefore, there is a very fast molecular remodelling upon AF establishment. Identification of markers of these molecular mechanisms could help early diagnosis and classification of AF patients. Moreover, AF treatment of

patients could be adjusted based on the pace of the molecular remodelling that is underway.

### **5.1 Studying PLA structural and neural remodelling during AF progression**

The posterior wall of the left atria (PLA) is the closest tissue to the pulmonary veins, where the vast majority of AF initiation triggers are usually found<sup>17</sup>. That is why surgical isolation of pulmonary veins has become the standard method to treat AF progression<sup>11</sup>. Moreover, PLA is the cardiac tissue that suffers the strongest structural remodelling during AF progression, as can be observed by an increase in fibrosis in the long-standing persistent AF sheep model<sup>121</sup>. In this animal model, functional analysis of gene expression changes showed an upregulation of GO genes related to focal adhesion and extracellular matrix. These results are also in concordance with previously described structural remodelling and extracellular matrix gene expression changes in the RA of AF patients<sup>203</sup> and in protein extracted from LA of AF patients<sup>204</sup>, and surely reflect the increased fibrosis observed *in vivo*. More surprisingly, we found a group of genes involved in axonogenesis that suggests remodelling of the autonomous nervous system of the PLA during AF progression. It has been previously shown by Arora et al. that in normal dog hearts, the PLA is enriched in sympathetic and parasympathetic nerve bundles in comparison with LA and pulmonary veins, providing an excellent substrate for AF maintenance<sup>205</sup>. Indeed, it has been shown in vagal (vagus nerve is the main source of parasympathetic fibers to the heart) stimulated AF dog models that parasympathetic blockade inhibits AF inducibility and atrial remodelling in PLA<sup>206</sup>. We have tried to address by WB and by immunofluorescence if there was an increase of parasympathetic and sympathetic innervation in our model during AF progression. However, our results were not conclusive, and they only suggest that there could be an increase of cholinergic neurons (parasympathetic component). Further experiments to include also chronic sheep, an increase of the sample size and the use of other sympathetic and parasympathetic markers may help to elucidate if there is a real neural remodelling in our AF model.

### **5.2 Molecular mechanisms altered in LA and RA cardiomyocytes during AF progression.**

Over 70% of the total cardiac mass is formed by cardiomyocytes due to their large volume. However, they only represent 30% of the total cells of the heart. Therefore, we considered important to study isolated atrial cardiomyocytes in order to understand specific changes in the cell type that shows altered electrical behaviour in AF. Our

transcriptomic profiling of isolated cardiomyocytes from LA and RA showed that transcriptomic changes in LA are more abundant than in the RA, suggesting a faster gene expression regulation at the LA than in RA which leads to an increase on differences between both atria at chronic stage. Meanwhile, proteomic profiling of this cells displays more changes in RA than LA at transition stage but similar number of differentially expressed proteins at chronic stage. This proteomics changes level out the proteomics differences that were present between atria in sham sheep. This apparent lack of correlation between mRNA and proteins could be favour by the different amount of genes and proteins captured by each assay, however, it has been previously reported that usually only 40% of protein variation can be explained by measurement of mRNA<sup>207</sup>. Anyhow, our data is pointing out that cardiomyocytes in both atria are changing with the disease progression and there not seem to be atria specific changes.

Gene Ontology analysis for biological processes enriched in differentially expressed genes during AF progression indicate changes in genes associated to heart and cardiac muscle development. Similar changes have been reported in a goat model of sustained AF. Electron microscopy on these goats have revealed heterochromatin alterations and a nuclear structure resembling the interphase nuclei of embryonic myocytes<sup>208</sup>. Further evidences for dedifferentiation of atrial myocytes have also been reported in human RA samples by identifying  $\beta$ -MHC and  $\alpha$ -smooth muscle actin re-expression<sup>209</sup>. Moreover, in RA tissue samples from permanent AF patients, a reprogramming of atrial transcriptome to a ventricular-like signature has been described<sup>203</sup>. This result is supported by previous experiments in a porcine model of AF where it was found MLCV (ventricular myosin light chain) mRNA and protein was detected in atrial tissue<sup>210</sup>. In our data, both mRNA and protein levels of ventricular myosin light chain are differentially upregulated, confirming observations in other models. A detail analysis of these changes is being done in our lab to further explore this transcriptional alteration and dedifferentiation of atrial myocytes.

On the other hand, enrichment analysis of functional annotations for differentially expressed proteins revealed a clear alteration in OXPHOS related proteins, more specifically an upregulation of proteins that are part of CI, CII and CIV complexes. Similar protein alterations have been described in the proteomic profile LA of AF patients<sup>204</sup>. Changes in mitochondria size and shape has been reported by electron microscopy in sustained AF goats<sup>208</sup> and mitochondria aggregates have been also found in atrial myocytes of patients with atrial arrhythmias<sup>211</sup>. However, these changes did not explain whether upregulation of mitochondrial proteins have a role or not, since to be

functional they have to be assembled into complexes and supercomplexes. Blue-native gel electrophoresis and in-gel activity assays have allowed us to conclude that this increase expression in mitochondrial proteins results in an increase of functional complexes and supercomplexes. Moreover, we have identified novel complexes formation that are positive for CI. In the mitochondria CI is the main source of reactive oxygen species (ROS) generation and therefore significantly contributes to cellular oxidative stress<sup>212</sup>. Therefore, an increase on CI complex and containing supercomplexes as we observed could be producing an increase on oxidative stress during AF progression. This is a very relevant finding, since it has been reported that mitochondrial oxidative stress can promote AF<sup>213,214</sup>. Eventually, the overall increase of OXPHOS subunits leading to more complexes and SCs can be induced by the overload of work that cardiomyocytes have to sustain during AF even though it is still to clarify whether these events are in the framework of a physiological homeostasis or they can damage the cardiomyocyte and worsen the disease progression. Hence, a better understanding of OXPHOS complexes and SCs upregulation during AF progression will help to understand the disease and eventually develop possible AF treatment, since it has already been described that inhibition of mitochondrial ROS production prevents AF<sup>213</sup>.

## **6. GWAS associated *loci* modulation in AF progression**

Throughout this work, we have used different approaches to elucidate the different contributions genetic (intrinsic) and environmental (extrinsic) factors to AF. Therefore, we have studied how several non-coding variants could be involved in AF and identified putative target genes, and at the same time we have worked with an AF induced model to unveil the genetic and proteomic networks that are modified because of the disease. These two approaches have the same final aim, to understand AF origin and development, so it would be interesting to explore if there are common alterations between them. Sure enough, transcriptomic and proteomic profiling from atrial samples have showed us that many of the genes that are part of the genomic landscape of AF-associated variants are also differentially expressed in the sheep AF model.

For example, *PITX2* is downregulated during AF progression in the LA cardiomyocytes, while *ENPEP* is downregulated in the PLA. Here, *PITX2* shows a trend to decrease during AF progression (adj. p-value=0.06) (**Table 12**). In the case of genes at 1q24 *locus*, despite *PRRX1* not being differentially expressed, *FMO2* and *METTL11B* genes that in our 4C-seq studies were interacting with the mouse syntenic region to the human one that contains the risk variant, are downregulated in the PLA during AF progression.

Finally, at 7q31 *locus* all candidate genes except *CAPZA2* are differentially expressed in some condition. *MET* and *TES* are both upregulated during AF progression in the PLA, however in LA cardiomyocytes only *TES* is significantly upregulated. This suggest that different regulation depending on the cell types of the same genomic *locus* could be having independent contributions to AF. On the other hand, *CAV1* is only differentially expressed between left and right protein samples in controls, but this difference is lost during the disease progression.

Gene	Locus	PLA RNA-seq	CM RNA-seq	CM LC-MS/MS
<i>ESR2</i>	14q23	NOT FOUND	NOT FOUND	NOT FOUND
<i>SYNE2</i>	14q23	NoDEG	NoDEG	NoDEP
<i>ENPEP</i>	4q25	DEG	NoDEG	NoDEP
<i>PITX2</i>	4q25	P=0.06	DEG	NOT FOUND
<i>FMO2</i>	1q24	DEG	NoDEG	NoDEP
<i>GORAB</i>	1q24	NoDEG	NoDEG	NoDEP
<i>KIFAP3</i>	1q24	NoDEG	NoDEG	NOT FOUND
<i>METTL11B</i>	1q24	DEG	DEG	NOT FOUND
<i>PRRX1</i>	1q24	NoDEG	NoDEG	NOT FOUND
<i>SCYL3</i>	1q24	NoDEG	NoDEG	NoDEP
<i>CAPZA2</i>	7q31	NOT FOUND	NoDEG	NoDEP
<i>CAV1</i>	7q31	NoDEG	NoDEG	DEP
<i>MET</i>	7q31	DEG	NoDEG	NOT FOUND
<i>TES</i>	7q31	DEG	DEG	NOT FOUND
<i>KCNN3</i>	1q21	NoDEG	NoDEG	NOT FOUND
<i>ZFH3</i>	16q22	NoDEG	NoDEG	NOT FOUND

**Table 12: GWAS AF associated genes differentially expressed in AF induced model**  
 GWAS AF associated *locus* and its potential causal genes that have been shown to interact by 4C-seq between them or with a potential regulatory element status in the different datasets of AF induced sheep model. Differentially expressed gene (DEG), differentially expressed protein (DEP) and NOT FOUND means that has not been detected in this dataset.

Our analysis shows that the study of a complex disease such as AF requires the integration of data obtained from multiple sources and experimental models.

GWAS has clearly helped in identification of multiple candidate AF associated genes, especially in recent years when meta-analysis that have led to the identification of more than one hundred different genomic *loci* associated to AF<sup>73,75</sup>. However, variant identification is as important as understanding the functional link for their association, especially when they will have a small phenotypic effect. Therefore, understanding the mechanistic link between an associated variant and AF can contribute to a better patient diagnosis and personalized treatment. Moreover, for a better treatment it is also

essential to comprehend the underlying mechanisms behind AF progression, otherwise current inefficient drug treatment will remain.

Through this work we have aimed to contribute to this knowledge by studying not only the genotype behind the disease but also exploring different layers of regulation and the crosstalk with extrinsic factors that also modulates AF origin and progression.

We have attempted to unveil novel mechanism underlying AF progression through different *in vitro* and *in vivo* system models to have a better notion on how a given phenotype such as AF is the result of a complex crosstalk of the genotype with environmental factors that we do not fully understand yet.



## CONCLUSIONS

---

*Success is the ability to go from one failure  
to another with no loss of enthusiasm.*

Winston Churchill



## CONCLUSIONS

- We have described the chromatin interaction map of mouse genomic region syntenic to human *loci* associated to AF by GWAS.
- *CAV1* starts to be expressed in the atria since early stages of heart development, expanding its expression at later stages of development to the coronary vasculature.
- Deletion of a genomic region containing an AF-associated SNP in intron 2 of *CAV1* causes a reduction of *CAV1* expression hiPSC-CMs.
- Transcriptomic and proteomic profiling from tissue and isolated cardiomyocytes of an AF induced model in sheep shows that most molecular changes occur at early stages of the disease and are then maintained during chronification of the disease.
- Transcriptomic profiling of PLA from an AF induced model in sheep show an structural remodelling and suggests a remodelling of the neuronal innervation of the autonomous nervous system during AF development at the PLA.
- Proteomic profiling from LA and RA isolated cardiomyocytes of an AF induced model in sheep shows that mitochondrial OXPHOS proteins are overexpressed during AF progression leading to more complexes and SCs assembly. Moreover, novel complexes containing Complex I appears in diseased cardiomyocytes.



## CONCLUSIONES

---

*Observar sin pensar es tan peligroso como pensar sin observar.*

Santiago Ramón y Cajal



## CONCLUSIONES

- Hemos realizado una descripción del perfil de interacciones de la cromatina que tiene lugar en ratón en aquellas regiones sinténicas a las que en humano contienen varios *loci* asociados a FA.
- CAV1 comienza a expresarse en las aurículas al comienzo del desarrollo y su expresión se extiende en estadios posteriores a la vasculatura coronaria.
- La eliminación de la región que contiene en el intrón 2 de *CAV1* la variante asociada a FA produce una reducción de la expresión de CAV1 en cardiomiocitos derivados de células humanas pluripotentes.
- El estudio del perfil transcripcional y proteómico de tejido y cardiomiocitos aislados de un modelo de FA inducida en oveja muestra que la mayoría de los cambios moleculares ocurren en las primeras etapas de la enfermedad y se mantienen durante la progresión de la misma.
- El estudio del perfil transcripcional de la pared posterior de la aurícula izquierda en un modelo de FA inducida en oveja sugiere que junto con el remodelado estructural se produce un remodelado de la inervación del sistema nervioso autónomo durante la progresión de la enfermedad.
- El estudio del perfil proteómico de cardiomiocitos aislados de las aurículas izquierda y derecha del modelo de FA inducida en oveja revela un incremento en las proteínas del sistema de fosforilación oxidativa durante la progresión de la FA dando lugar a un mayor ensamblaje de complejos y supercomplejos, así como, a la aparición de un nuevo complejo.





## BIBLIOGRAPHY

---

*Ignorance more frequently begets confidence than does knowledge:  
it is those who know little, and not those who know much,  
who so positively assert that this or that problem  
will never be solved by science.*

Charles R. Darwin



1. Wilkins, E. *et al.* European Cardiovascular Disease Statistics 2017. *Eur. Hear. Network, Brussels* 8 (2017).
2. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2018 Update A Report From the American Heart Association. *Circulation* **137**, 67–492 (2018).
3. Kannel, W. ., Wolf, P. ., Benjamin, E. . & Levy, D. Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates. *Am. J. Cardiol.* **82**, 2N–9N (1998).
4. Kirchhof, P. *et al.* 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016).
5. Schnabel, R. B. *et al.* 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: A cohort study. *Lancet* **386**, 154–162 (2015).
6. Chugh, S. S. *et al.* Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study. *Circulation* **129**, 837–847 (2014).
7. Miyasaka, Y. *et al.* Secular Trends in Incidence of Atrial Fibrillation in Olmsted County, Minnesota, 1980 to 2000, and Implications on the Projections for Future Prevalence. (2006).
8. Pérez-Villacastín, J. & Sociedad Española de Cardiología Sección de Electrofisiología y Arritmias. *Cardio arritmias*. Grupo CTO. (2017).
9. Fuster, V. *et al.* ACC / AHA / ESC Practice Guidelines ACC / AHA / ESC 2006 Guidelines for the Management of Patients With Atrial Fibrillation – Executive Summary A Report of the American College of Cardiology / American Heart Association Task Force on Practice Guidelines a. (2006).
10. Waktare, J. E. P. Atrial Fibrillation. *Circulation* **106**, 14–16 (2002).
11. Calkins, H. *et al.* 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Hear. Rhythm* **14**, e275–e444 (2017).
12. Andalib, A., Brugada, R. & Nattel, S. Atrial fibrillation: evidence for genetically determined disease. *Curr. Opin. Cardiol.* **23**, 176–183 (2008).
13. Heijman, J., Voigt, N., Nattel, S. & Dobrev, D. Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression. *Circ. Res.* **114**, 1483–1499 (2014).
14. Moe, G. K. & Abildskov, J. A. Atrial fibrillation as a self-sustaining arrhythmia independent of focal discharge. *Am. Heart J.* **58**, 59–70 (1959).
15. Jalife, J., Berenfeld, O. & Mansour, M. Mother rotors and fibrillatory conduction:

- a mechanism of atrial fibrillation. *Cardiovasc. Res.* **54**, 204–216 (2002).
16. de Groot, N. M. S. *et al.* Electropathological Substrate of Longstanding Persistent Atrial Fibrillation in Patients With Structural Heart Disease. *Circulation* **122**, 1674–1682 (2010).
  17. Calvo, D., Filgueiras-Rama, D. & Jalife, J. Mechanisms and Drug Development in Atrial Fibrillation. *Pharmacol. Rev.* **70**, 505–525 (2018).
  18. Jalife, J. & Kaur, K. Atrial remodeling, fibrosis, and atrial fibrillation. *Trends in Cardiovascular Medicine* **25**, 475–484 (2015).
  19. Anumonwo, J. M. B. & Kalifa, J. Risk Factors and Genetics of Atrial Fibrillation. *Heart Fail. Clin.* **12**, 157–166 (2016).
  20. Wolff, L. Familial Auricular Fibrillation. *N. Engl. J. Med.* **229**, 396–398 (1943).
  21. ORGAIN, E. S., WOLFF, L. & WHITE, P. D. Uncomplicated auricular fibrillation and auricular flutter. *Arch. Intern. Med.* **57**, 493 (1936).
  22. Fox, C. S. *et al.* Parental Atrial Fibrillation as a Risk Factor for Atrial Fibrillation in Offspring. *JAMA* **291**, 2851 (2004).
  23. Arnar, D. O. *et al.* Familial aggregation of atrial fibrillation in Iceland. *Eur. Heart J.* **27**, 708–712 (2006).
  24. Christophersen, I. E. *et al.* Familial aggregation of atrial fibrillation: A study in danish twins. *Circ. Arrhythmia Electrophysiol.* **2**, 378–383 (2009).
  25. Brugada, R. *et al.* Identification of a Genetic Locus for Familial Atrial Fibrillation. *N. Engl. J. Med.* **336**, 905–911 (1997).
  26. Ellinor, P. T., Shin, J. T., Moore, R. K., Yoerger, D. M. & MacRae, C. A. Locus for Atrial Fibrillation Maps to Chromosome 6q14–16. *Circulation* **107**, 2880–2883 (2003).
  27. Oberti, C. *et al.* Genome-Wide Linkage Scan Identifies a Novel Genetic Locus on Chromosome 5p13 for Neonatal Atrial Fibrillation Associated With Sudden Death and Variable Cardiomyopathy. *Circulation* **110**, 3753–3759 (2004).
  28. Chen, Y. H. *et al.* KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science (80-. ).* **299**, 251–254 (2003).
  29. Christophersen, I. E. & Ellinor, P. T. Genetics of atrial fibrillation: From families to genomes. *Journal of Human Genetics* **61**, 61–70 (2016).
  30. Tucker, N. R. & Ellinor, P. T. Emerging Directions in the Genetics of Atrial Fibrillation. 1469–1482 (2014).
  31. Pérez-Serra, A., Campuzano, O. & Brugada, R. Update about atrial fibrillation genetics. *Curr. Opin. Cardiol.* **32**, 246–252 (2017).

32. Manolio, T. A. Genomewide Association Studies and Assessment of the Risk of Disease. *N. Engl. J. Med.* **363**, 166–176 (2010).
33. Gudbjartsson, D. F. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
34. Franco, D. & Campione, M. The Role of Pitx2 during Cardiac Development: Linking Left–Right Signaling and Congenital Heart Diseases. *Trends Cardiovasc. Med.* **13**, 157–163 (2003).
35. Mommersteeg, M. T. M. *et al.* Molecular Pathway for the Localized Formation of the Sinoatrial Node. *Circ. Res.* **100**, 354–362 (2007).
36. Mommersteeg, M. T. M. *et al.* *Pitx2c* and *Nkx2-5* Are Required for the Formation and Identity of the Pulmonary Myocardium. *Circ. Res.* **101**, 902–909 (2007).
37. Franco, D., Christoffels, V. M. & Campione, M. Homeobox transcription factor Pitx2: The rise of an asymmetry gene in cardiogenesis and arrhythmogenesis. *Trends Cardiovasc. Med.* **24**, 23–31 (2014).
38. Kirchhof, P. *et al.* PITX2c Is Expressed in the Adult Left Atrium, and Reducing Pitx2c Expression Promotes Atrial Fibrillation Inducibility and Complex Changes in Gene Expression. *Circ. Cardiovasc. Genet.* **4**, 123–133 (2011).
39. Gudbjartsson, D. F. *et al.* A sequence variant in ZFH3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
40. Ellinor, P. T. *et al.* Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat. Genet.* **42**, 240–244 (2010).
41. Benjamin, E. J. *et al.* Variants in ZFH3 are associated with atrial fibrillation in individuals of European ancestry. *Nat. Genet.* **41**, 879–881 (2009).
42. Berry, F. B. *et al.* Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. *J. Biol. Chem.* **276**, 25057–65 (2001).
43. Nojiri, S. *et al.* ATBF1 enhances the suppression of STAT3 signaling by interaction with PIAS3. *Biochem. Biophys. Res. Commun.* **314**, 97–103 (2004).
44. Jiang, Q. *et al.* Down-regulation of ATBF1 activates STAT3 signaling via PIAS3 in pacing-induced HL-1 atrial myocytes. *Biochem. Biophys. Res. Commun.* **449**, 278–283 (2014).
45. Kao, Y.-H. *et al.* ZFH3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes. *Int. J. Cardiol.* **210**, 85–92 (2016).
46. Köhler, M. *et al.* Small-conductance, calcium-activated potassium channels from

- mammalian brain. *Science* **273**, 1709–14 (1996).
47. Zhang, X.-D. *et al.* Critical roles of a small conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel (SK3) in the repolarization process of atrial myocytes. *Cardiovasc. Res.* **101**, 317–25 (2014).
  48. Zeggini, E. & Ioannidis, J. P. A. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191–201 (2009).
  49. Ellinor, P. T. *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, (2012).
  50. Lozano-Velasco, E. *et al.* Pitx2 impairs calcium handling in a dose-dependent manner by modulating Wnt signalling. *Cardiovasc. Res.* **109**, 55–66 (2016).
  51. Libório, T. N. *et al.* In situ hybridization detection of homeobox genes reveals distinct expression patterns in oral squamous cell carcinomas. *Histopathology* **58**, 225–233 (2011).
  52. Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–24 (2012).
  53. Bergwerff, M. *et al.* Patterns of paired-related homeobox genes PRX1 and PRX2 suggest involvement in matrix modulation in the developing chick vascular system. *Dev. Dyn.* **213**, 59–70 (1998).
  54. Ihida-Stansbury, K. *et al.* Paired-Related Homeobox Gene *Prx1* Is Required for Pulmonary Vascular Development. *Circ. Res.* **94**, 1507–1514 (2004).
  55. Tucker, N. R. *et al.* Diminished *PRRX1* Expression Is Associated With Increased Risk of Atrial Fibrillation and Shortening of the Cardiac Action Potential. *Circ. Cardiovasc. Genet.* **10**, (2017).
  56. Lin, H. *et al.* Gene expression and genetic variation in human atria. *Hear. Rhythm* **11**, 266–271 (2014).
  57. Duhme, N. *et al.* Altered HCN4 channel C-linker interaction is associated with familial tachycardia–bradycardia syndrome and atrial fibrillation. *Eur. Heart J.* **34**, 2768–2775 (2013).
  58. Díaz-Perales, A. *et al.* Identification of human aminopeptidase O, a novel metalloprotease with structural similarity to aminopeptidase B and leukotriene A4 hydrolase. *J. Biol. Chem.* **280**, 14310–7 (2005).
  59. Shahid, F., Lip, G. Y. H. & Shantsila, E. Renin–angiotensin blockade in atrial fibrillation: where are we now? *J. Hum. Hypertens.* **31**, 425–426 (2017).
  60. Sohn, J., Brick, R. M. & Tuan, R. S. From embryonic development to human diseases: The functional role of caveolae/caveolin. *Birth Defects Res. Part C -*

- Embryo Today Rev.* **108**, 45–64 (2016).
61. MAGUY, A., HEBERT, T. & NATTEL, S. Involvement of lipid rafts and caveolae in cardiac ion channel function. *Cardiovasc. Res.* **69**, 798–807 (2006).
  62. Echarri, A. & Del Pozo, M. A. Caveolae - mechanosensitive membrane invaginations linked to actin filaments. *J. Cell Sci.* **128**, 2747–58 (2015).
  63. Zhao, Y.-Y. *et al.* Defects in caveolin-1 cause dilated cardiomyopathy and pulmonary hypertension in knockout mice. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11375–80 (2002).
  64. Zhang, Q. *et al.* Nesprin-1 and -2 are involved in the pathogenesis of Emery–Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Hum. Mol. Genet.* **16**, 2816–2833 (2007).
  65. Lüke, Y. *et al.* Nesprin-2 Giant (NUANCE) maintains nuclear envelope architecture and composition in skin. *J. Cell Sci.* **121**, 1887–98 (2008).
  66. Zhang, X. *et al.* SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice. *Neuron* **64**, 173–87 (2009).
  67. Sinner, M. F. *et al.* Analyses to Identify 5 Novel Genes for Atrial Fibrillation. (2015).
  68. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
  69. Tsai, C.-T. *et al.* Genome-wide screening identifies a KCNIP1 copy number variant as a genetic predictor for atrial fibrillation. *Nat. Commun.* **7**, 10190 (2016).
  70. Lin, H. *et al.* Methylome-wide Association Study of Atrial Fibrillation in Framingham Heart Study. *Sci. Rep.* **7**, 1–9 (2017).
  71. Christophersen, I. E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952 (2017).
  72. Low, S.-K. *et al.* Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat. Genet.* **49**, 953–958 (2017).
  73. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nature Genetics* **50**, 1–9 (2018).
  74. Nielsen, J. B. *et al.* Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development. *Am. J. Hum. Genet.* **102**, 103–115 (2018).
  75. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature Genetics* **50**, 1234–1239 (2018).

## BIBLIOGRAPHY

---

76. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *American Journal of Human Genetics* **102**, 717–730 (2018).
77. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).
78. Aguirre, L. A. *et al.* Long-range regulatory interactions at the 4q25 atrial fibrillation risk locus involve PITX2c and ENPEP. *BMC Biol.* **13**, 26 (2015).
79. J. Craig Venter, Mark D. Adams, Eugene W. Myers, Peter W. Li, R. J. M. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).
80. Sakabe, N. J. & Nobrega, M. A. Genome-wide maps of transcription regulatory elements. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 422–437 (2010).
81. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
82. Spitz, F. & Furlong, E. E. M. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics* **13**, 613–626 (2012).
83. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
84. Fraser, J., Williamson, I., Bickmore, W. A. & Dostie, J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **79**, 347–372 (2015).
85. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
86. García-González, E., Escamilla-Del-Arenal, M., Arzate-Mejía, R. & Recillas-Targa, F. Chromatin remodeling effects on enhancer activity. *Cell. Mol. Life Sci.* **73**, 2897–2910 (2016).
87. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
88. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159–164 (2015).
89. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
90. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–21 (2012).
91. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23 (2013).
92. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-



- Cas9 for genome engineering. *Cell* **157**, 1262–78 (2014).
93. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–11 (2002).
  94. Wit, E. De *et al.* A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
  95. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–54 (2006).
  96. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
  97. Splinter, E. *et al.* The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* **25**, 1371–83 (2011).
  98. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–309 (2006).
  99. Bickmore, W. A. The Spatial Organization of the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
  100. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–93 (2009).
  101. Horike, S., Cai, S., Miyano, M., Cheng, J.-F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* **37**, 31–40 (2005).
  102. Gavrillov, A. *et al.* Chromatin Immunoprecipitation Assays. **567**, 171–188 (2009).
  103. Fullwood, M. J. *et al.* An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
  104. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
  105. Li, G. *et al.* Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics* **15**, S11 (2014).
  106. Goedel, A., My, I., Sinnecker, D. & Moretti, A. Perspectives and Challenges of Pluripotent Stem Cells in Cardiac Arrhythmia Research. *Current Cardiology Reports* **19**, 23 (2017).
  107. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–76 (2006).

## BIBLIOGRAPHY

---

108. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–72 (2007).
109. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–20 (2007).
110. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci.* **109**, E1848–E1857 (2012).
111. Burridge, P. W. *et al.* Chemically defined generation of human cardiomyocytes. *Nat. Methods* **11**, 855–860 (2014).
112. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183–193 (2016).
113. Kattman, S. J. *et al.* Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8**, 228–40 (2011).
114. Weng, Z. *et al.* A Simple, Cost-Effective but Highly Efficient System for Deriving Ventricular Cardiomyocytes from Human Pluripotent Stem Cells. *Stem Cells Dev.* **23**, 1704–1716 (2014).
115. Cyganek, L. *et al.* Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight* **3**, (2018).
116. Herron, T. J. *et al.* Extracellular matrix-mediated maturation of human pluripotent stem cell-derived cardiac monolayer structure and electrophysiological function. *Circ. Arrhythmia Electrophysiol.* **9**, (2016).
117. Ronaldson-Bouchard, K. *et al.* Advanced maturation of human cardiac tissue grown from pluripotent stem cells. *Nature* **556**, 239–243 (2018).
118. Riley, G., Syeda, F., Kirchhof, P. & Fabritz, L. An Introduction to Murine Models of Atrial Fibrillation. *Front. Physiol.* **3**, 296 (2012).
119. Nadadur, R. D. *et al.* Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci. Transl. Med.* **8**, 354ra115 (2016).
120. Nishida, K., Michael, G., Dobrev, D. & Nattel, S. Animal models for atrial fibrillation: Clinical insights and scientific opportunities. *Europace* **12**, 160–172 (2010).
121. Martins, R. P. *et al.* Dominant Frequency Increase Rate Predicts Transition from Paroxysmal to Long-Term Persistent Atrial Fibrillation. (2014).
122. Filgueiras-Rama, D. *et al.* Long-Term Frequency Gradients During Persistent Atrial Fibrillation in Sheep Are Associated With Stable Sources in the Left Atrium.

- Circ. Arrhythmia Electrophysiol.* **5**, 1160–1167 (2012).
123. Rosenbloom, K. R. *et al.* ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **41**, D56–D63 (2012).
  124. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
  125. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).
  126. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
  127. Noordermeer, D. *et al.* The dynamic architecture of Hox gene clusters. *Science* (80-. ). **334**, 222–225 (2011).
  128. Van De Werken, H. J. G. *et al.* 4C technology: Protocols and data analysis. *Methods in Enzymology* **513**, (Elsevier Inc., 2012).
  129. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
  130. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
  131. Klein, F. A. *et al.* FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085–3091 (2015).
  132. Zhou, X. *et al.* Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. *Nat. Biotechnol.* **33**, 345–346 (2015).
  133. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–42 (2011).
  134. Bizy, A. *et al.* Myosin light chain 2-based selection of human iPSC-derived early ventricular cardiac myocytes. *Stem Cell Res.* **11**, 1335–1347 (2013).
  135. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
  136. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
  137. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
  138. Schägger, H. [12] Native electrophoresis for isolation of mitochondrial oxidative phosphorylation protein complexes. *Methods Enzymol.* **260**, 190–202 (1995).
  139. Jung, C., Higgins, C. M. J. & Xu, Z. Measuring the Quantity and Activity of

- Mitochondrial Electron Transport Chain Complexes in Tissues of Central Nervous System Using Blue Native Polyacrylamide Gel Electrophoresis. *Anal. Biochem.* **286**, 214–223 (2000).
140. Mizutani, S. *et al.* New insights into the importance of aminopeptidase A in hypertension. *Heart Fail. Rev.* **13**, 273–284 (2008).
141. Welsh, I. C. *et al.* Chromatin Architecture of the Pitx2 Locus Requires CTCF- and Pitx2-Dependent Asymmetry that Mirrors Embryonic Gut Laterality. *Cell Rep.* **13**, 337–49 (2015).
142. Bao, Q. *et al.* MicroRNA-297 promotes cardiomyocyte hypertrophy via targeting sigma-1 receptor. *Life Sci.* **175**, 1–10 (2017).
143. Cox, C. J. *et al.* Differential regulation of gene expression by PITX2 isoforms. *J. Biol. Chem.* **277**, 25001–10 (2002).
144. Yang, Z.-P. *et al.* SCYL1BP1 has tumor-suppressive functions in human lung squamous carcinoma cells by regulating degradation of MDM2. *Asian Pac. J. Cancer Prev.* **15**, 7467–71 (2014).
145. Shimizu, K., Shirataki, H., Honda, T., Minami, S. & Takai, Y. Complex formation of SMAP/KAP3, a KIF3A/B ATPase motor-associated protein, with a human chromosome-associated polypeptide. *J. Biol. Chem.* **273**, 6591–4 (1998).
146. Elena Dolmatova, N. R. T. H. L. R. R. C. J. Y. M. F. S. M. I. S. A. L. J. L. G. V. E. J. B. K. L. L. L. M. D. J. M. P. T. E. Abstract 18865: Identification of a Functional Snp Regulating Prrx1 at the 1q24 Locus for Atrial Fibrillation. *Circulation* **130**, A18865 (2014).
147. Petkowski, J. J. *et al.* NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J.* **456**, 453–62 (2013).
148. Siddens, L. K., Krueger, S. K., Henderson, M. C. & Williams, D. E. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem. Pharmacol.* **89**, 141–7 (2014).
149. Alvarez-Dominguez, J. R. *et al.* De Novo Reconstruction of Adipose Tissue Transcriptomes Reveals Long Non-coding RNA Regulators of Brown Adipocyte Development. *Cell Metab.* **21**, 764–776 (2015).
150. Mandala, S. M. *et al.* Sphingoid base 1-phosphate phosphatase: A key regulator of sphingolipid metabolism and stress response. *Proc. Natl. Acad. Sci.* **95**, 150–155 (1998).
151. Mosselman, S., Polman, J. & Dijkema, R. ER $\beta$ : Identification and characterization of a novel human estrogen receptor. *FEBS Lett.* **392**, 49–53 (1996).

152. Parton, R. G. & Simons, K. The multiple faces of caveolae. *Nat. Rev. Mol. Cell Biol.* **8**, 185–194 (2007).
153. Sala, S. *et al.* Expanding the Interactome of TES by Exploiting TES Modules with Different Subcellular Localizations. *J. Proteome Res.* **16**, 2054–2071 (2017).
154. Atanelishvili, I. *et al.* M10, a caspase cleavage product of the hepatocyte growth factor receptor, interacts with Smad2 and demonstrates antifibrotic properties in vitro and in vivo. *Transl. Res.* **170**, 99–111 (2016).
155. Montefiori, L. E. *et al.* A promoter interaction map for cardiovascular disease genetics. *Elife* **7**, (2018).
156. Balijepalli, R. C. & Kamp, T. J. Caveolae, ion channels and cardiac arrhythmias. *Prog. Biophys. Mol. Biol.* **98**, 149–160 (2008).
157. Le Lay, S. & Kurzchalia, T. V. Getting rid of caveolins: Phenotypes of caveolin-deficient animals. *Biochim. Biophys. Acta - Mol. Cell Res.* **1746**, 322–333 (2005).
158. Bullejos, M., Bowles, J. & Koopman, P. Extensive vascularization of developing mouse ovaries revealed by caveolin-1 expression. *Dev. Dyn.* **225**, 95–99 (2002).
159. Li, G. *et al.* Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Dev. Cell* **39**, 491–507 (2016).
160. Volonte, D., McTiernan, C. F., Drab, M., Kasper, M. & Galbiati, F. Caveolin-1 and caveolin-3 form heterooligomeric complexes in atrial cardiac myocytes that are required for doxorubicin-induced apoptosis. *Am. J. Physiol. Circ. Physiol.* **294**, H392–H401 (2008).
161. Sharma, B., Chang, A. & Red-Horse, K. Coronary Artery Development: Progenitor Cells and Differentiation Pathways. *Annu. Rev. Physiol.* **79**, 1–19 (2017).
162. Yi, S. L., Liu, X. J., Zhong, J. Q. & Zhang, Y. Role of caveolin-1 in atrial fibrillation as an anti-fibrotic signaling molecule in human atrial fibroblasts. *PLoS One* **9**, (2014).
163. Yoshida, Y. & Yamanaka, S. iPS cells: A source of cardiac regeneration. *J. Mol. Cell. Cardiol.* **50**, 327–332 (2011).
164. Burridge, P. W. *et al.* Chemically defined generation of human cardiomyocytes. *Nat. Methods* **11**, 855–860 (2014).
165. Jolliffe, I. T. *Principal component analysis.* (Springer, 2002).
166. Epstein, J. A., Aghajanian, H. & Singh, M. K. Semaphorin Signaling in Cardiovascular Development. *Cell Metab.* **21**, 163–173 (2015).
167. Franciosi, S. *et al.* The role of the autonomic nervous system in arrhythmias and

- sudden cardiac death. *Auton. Neurosci.* **205**, 1–11 (2017).
168. Rysevaite, K. *et al.* Immunohistochemical characterization of the intrinsic cardiac neural plexus in whole-mount mouse heart preparations. *Hear. Rhythm* **8**, 731–738 (2011).
169. Cogliati, S., Lorenzi, I., Rigoni, G., Caicci, F. & Soriano, M. E. Regulation of Mitochondrial Electron Transport Chain Assembly. *J. Mol. Biol.* **430**, 4849–4873 (2018).
170. Cogliati, S. *et al.* Mechanism of super-assembly of respiratory complexes III and IV. *Nature* **539**, 579–582 (2016).
171. Johannsen, W. The Genotype Conception of Heredity. *Am. Nat.* **45**, 129–159 (1911).
172. Consortium, I. H. G. S. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
173. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–51 (2001).
174. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
175. Bapat, A., Anderson, C. D., Ellinor, P. T. & Lubitz, S. A. Genomic basis of atrial fibrillation. *Heart* **104**, 201–206 (2018).
176. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
177. Mayr, C. Regulation by 3'-Untranslated Regions. *Annu. Rev. Genet. Vol 51* **51**, 171–194 (2017).
178. Jash, A., Yun, K., Sahoo, A., So, J.-S. & Im, S.-H. Looping Mediated Interaction between the Promoter and 3' UTR Regulates Type II Collagen Expression in Chondrocytes. *PLoS One* **7**, e40828 (2012).
179. Vizkeleti, L. *et al.* Prognostic relevance of the expressions of CAV1 and TES genes on 7q31 in melanoma. *Front. Biosci. (Elite Ed)*. **4**, 1802–12 (2012).
180. Xu, K. *et al.* Lunatic Fringe Deficiency Cooperates with the Met/Caveolin Gene Amplicon to Induce Basal-like Breast Cancer. *Cancer Cell* **21**, 626–641 (2012).
181. Gao, L. *et al.* Testin protects against cardiac hypertrophy by targeting a calcineurin-dependent signalling pathway. *J. Cell. Mol. Med.* **23**, 328–339 (2019).
182. Gallo, S., Sala, V., Gatti, S. & Crepaldi, T. Cellular and molecular mechanisms of HGF/Met in the cardiovascular system. *Clin. Sci.* **129**, 1173–1193 (2015).
183. Arechederra, M. *et al.* Met signaling in cardiomyocytes is required for normal

- cardiac function in adult mice. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1832**, 2204–2215 (2013).
184. Shihata, W. A., Putra, M. R. A. & Chin-Dusting, J. P. F. Is There a Potential Therapeutic Role for Caveolin-1 in Fibrosis? *Front. Pharmacol.* **8**, 567 (2017).
185. Lau, D. H. *et al.* Pathophysiology of Paroxysmal and Persistent Atrial Fibrillation: Rotors, Foci and Fibrosis. *Hear. Lung Circ.* **26**, 887–893 (2017).
186. Han, H. *et al.* Silencing of Kir2 channels by caveolin-1: Cross-talk with cholesterol. *J. Physiol.* **592**, 4025–4038 (2014).
187. Cheng, X. & Jaggar, J. H. Genetic ablation of caveolin-1 modifies Ca<sup>2+</sup> spark coupling in murine arterial smooth muscle cells. *Am. J. Physiol. Circ. Physiol.* **290**, H2309–H2319 (2006).
188. Mu, Y. P. *et al.* Alterations in caveolin-1 expression and receptor-operated Ca<sup>2+</sup> entry in the aortas of rats with pulmonary hypertension. *Cell. Physiol. Biochem.* **39**, 438–452 (2016).
189. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. *Nat. Rev. Genet.* **6**, 826–835 (2005).
190. Campione, M. *et al.* The homeobox gene Pitx2: mediator of asymmetric left-right signaling in vertebrate heart and gut looping. *Development* **126**, (1999).
191. Ocaña, O. H. *et al.* A right-handed signalling pathway drives heart looping in vertebrates. *Nature* **549**, 86–90 (2017).
192. Chinchilla, A. *et al.* PITX2 insufficiency leads to atrial electrical and structural remodeling linked to arrhythmogenesis. *Circ. Cardiovasc. Genet.* **4**, 269–279 (2011).
193. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
194. Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153–159 (2010).
195. Holm, H. *et al.* Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117–122 (2010).
196. Verweij, N. *et al.* Genetic Determinants of P Wave Duration and PR Segment. *Circ. Cardiovasc. Genet.* **7**, 475–481 (2014).
197. Sano, M. *et al.* Genome-wide association study of electrocardiographic parameters identifies a new association for PR interval and confirms previously reported associations. *Hum. Mol. Genet.* **23**, 6668–6676 (2014).
198. Seyerle, A. A. *et al.* Genome-wide association study of PR interval in

- Hispanics/Latinos identifies novel locus at ID2. *Heart* **104**, 904–911 (2018).
199. Butler, A. M. *et al.* Novel Loci Associated With PR Interval in a Genome-Wide Association Study of 10 African American Cohorts. *Circ. Cardiovasc. Genet.* **5**, 639–646 (2012).
200. Hong, K.-W. *et al.* Identification of three novel genetic variations associated with electrocardiographic traits (QRS duration and PR interval) in East Asians. *Hum. Mol. Genet.* **23**, 6659–6667 (2014).
201. Soliman, E. Z., Prineas, R. J., Case, L. D., Zhang, Z. & Goff, D. C. Ethnic Distribution of ECG Predictors of Atrial Fibrillation and Its Impact on Understanding the Ethnic Distribution of Ischemic Stroke in the Atherosclerosis Risk in Communities (ARIC) Study. *Stroke* **40**, 1204–1211 (2009).
202. Tada, H., Kawashiri, M., Yamagishi, M. & Hayashi, K. Atrial fibrillation: an inherited cardiovascular disease—a commentary on genetics of atrial fibrillation: from families to genomes. *J. Hum. Genet.* **61**, 3–4 (2016).
203. Barth, A. S. *et al.* Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: Expression of a ventricular-like genomic signature. *Circ. Res.* **96**, 1022–1029 (2005).
204. Doll, S. *et al.* Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469 (2017).
205. Arora, R. *et al.* Neural substrate for atrial fibrillation: implications for targeted parasympathetic blockade in the posterior left atrium. *Am. J. Physiol. Circ. Physiol.* **294**, H134–H144 (2008).
206. Qin, M., Li, L., Liu, X., Liu, T. & Shi, S. B. Neural substrate of posterior left atrium: A novel modulation for inducibility and remodeling of atrial fibrillation in canine. *PLoS One* **12**, e0176626 (2017).
207. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
208. Ausma, J. *et al.* Structural Changes of Atrial Myocardium due to Sustained Atrial Fibrillation in the Goat. *Circulation* **96**, 3157–3163 (1997).
209. Rucker-Martin, C., Pecker, F., Godreau, D. & Hatem, S. N. Dedifferentiation of atrial myocytes during atrial fibrillation: role of fibroblast proliferation in vitro. *Cardiovasc. Res.* **55**, 38–52 (2002).
210. Lai, L. P. *et al.* Functional Genomic Study on Atrial Fibrillation Using cDNA Microarray and Two-Dimensional Protein Electrophoresis Techniques and Identification of the Myosin Regulatory Light Chain Isoform Reprogramming in



- Atrial Fibrillation. *J. Cardiovasc. Electrophysiol.* **15**, 214–223 (2004).
211. Mary-Rabine, L. *et al.* The relationship of human atrial cellular electrophysiology to clinical function and ultrastructure. *Circ. Res.* **52**, 188–199 (1983).
212. Kussmaul, L. & Hirst, J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7607–12 (2006).
213. Xie, W. *et al.* Mitochondrial oxidative stress promotes atrial fibrillation. *Sci. Rep.* **5**, 11427 (2015).
214. Samman Tahhan, A. *et al.* Association between oxidative stress and atrial fibrillation. *Hear. Rhythm* **14**, 1849–1855 (2017).



## APPENDIXES

---

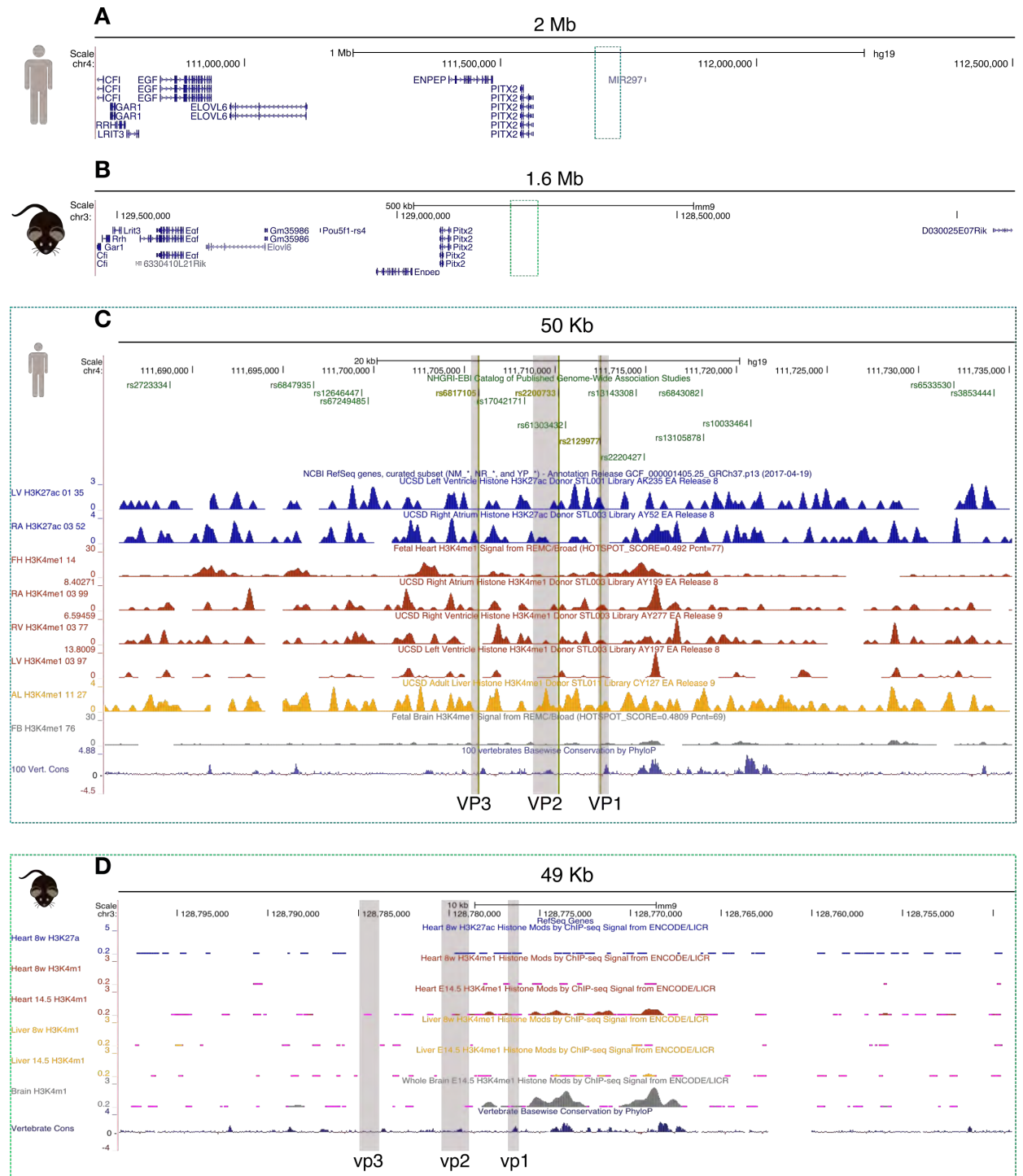
*Evolution is no linear family tree,  
but change in the single multidimensional being that  
has grown to cover the entire surface of Earth.*

Lynn Margulis



## APPENDIX 1

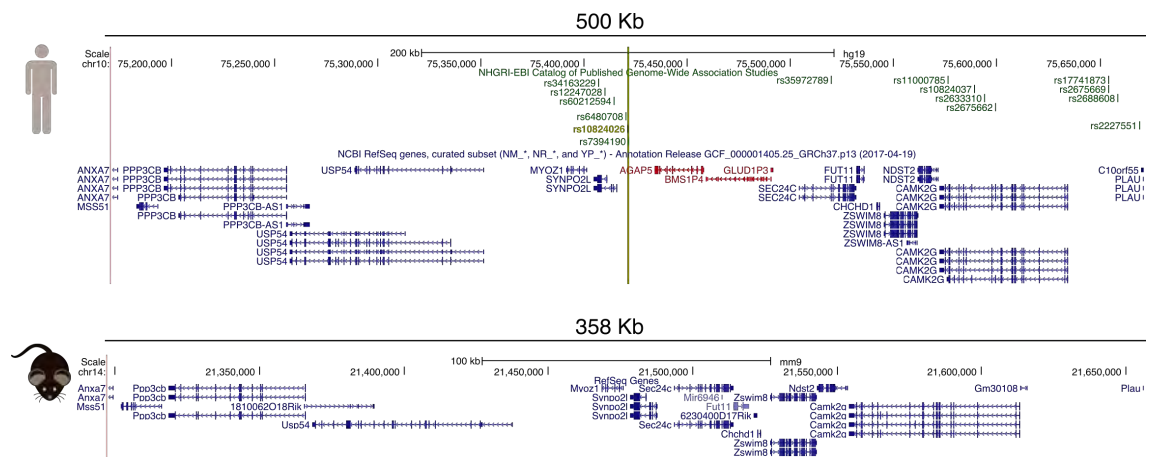
## 4q25 locus genome analysis



**A)** Genomic window of 2Mb of the 4q25 locus (hg19; chr4:110,710,169-112,710,169). **B)** Syntenic 1.6 Mb genomic window of the mouse genome (mm9; chr3:127,899,121-129,540,220). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in A (hg19; chr4:111,685,169-111,735,169) showing AF associated SNPs in light brown lettering and other SNPs in green. Below, epigenetic marks (H3K27ac, active enhancers

and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA), fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The three regions equivalent to the viewpoints selected for 4C-seq in the mouse are highlighted in light grey. **D)** Genomic window of 49 Kb corresponding to the light green rectangle in B (mm9; chr3:128,748,962-128,798,204). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart; 8 w mouse liver, E14.5 mouse liver; and E14.5 mouse whole brain are shown. Lower track shows vertebrate conservation. The three viewpoints selected for 4C-seq are highlighted in light grey. These viewpoints correspond to the syntenic regions that in the human genome contain AF associated SNPs.

## 10q22 locus genome analysis

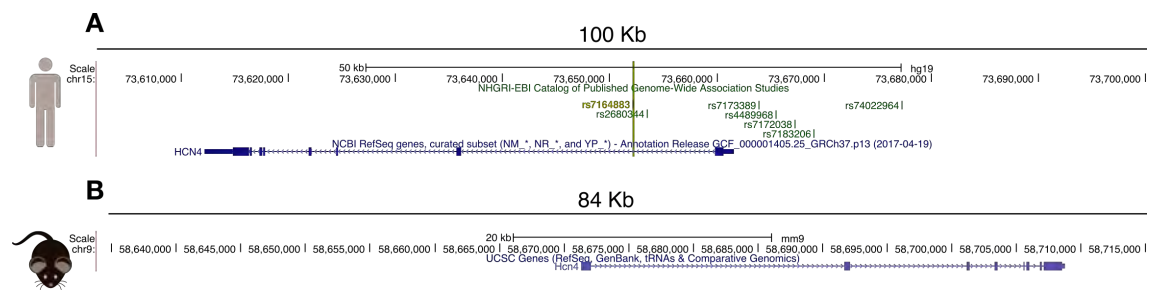


**A)** Genomic window of 500Kb of the 10q22 locus (hg19; chr10:75,171,208-75,671,208). First track shows AF associated SNPs in light brown lettering and other SNPs in green. Genes in red are not conserved with mouse. **B)** Syntenic 358 Kb genomic window of the mouse genome (mm9; chr14:21,297,554-21,656,177). Genes that are present in the human genome but not in the mouse are located between *SYNPO2L* and *SEC24C*, thereby due to lack of synteny over the region containing AF-associated SNPs, this locus was discarded for 4C-analysis.





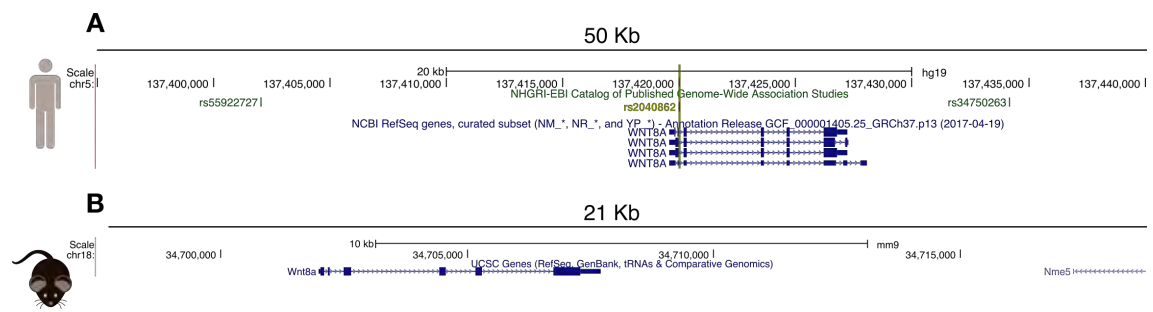
## 15q24 locus genome analysis



**A)** Genomic window of 100 Kb of the 15q24 locus (hg19; chr15:73,602,174-73,702,174). First track shows AF associated SNPs in light brown lettering and other SNPs in green.

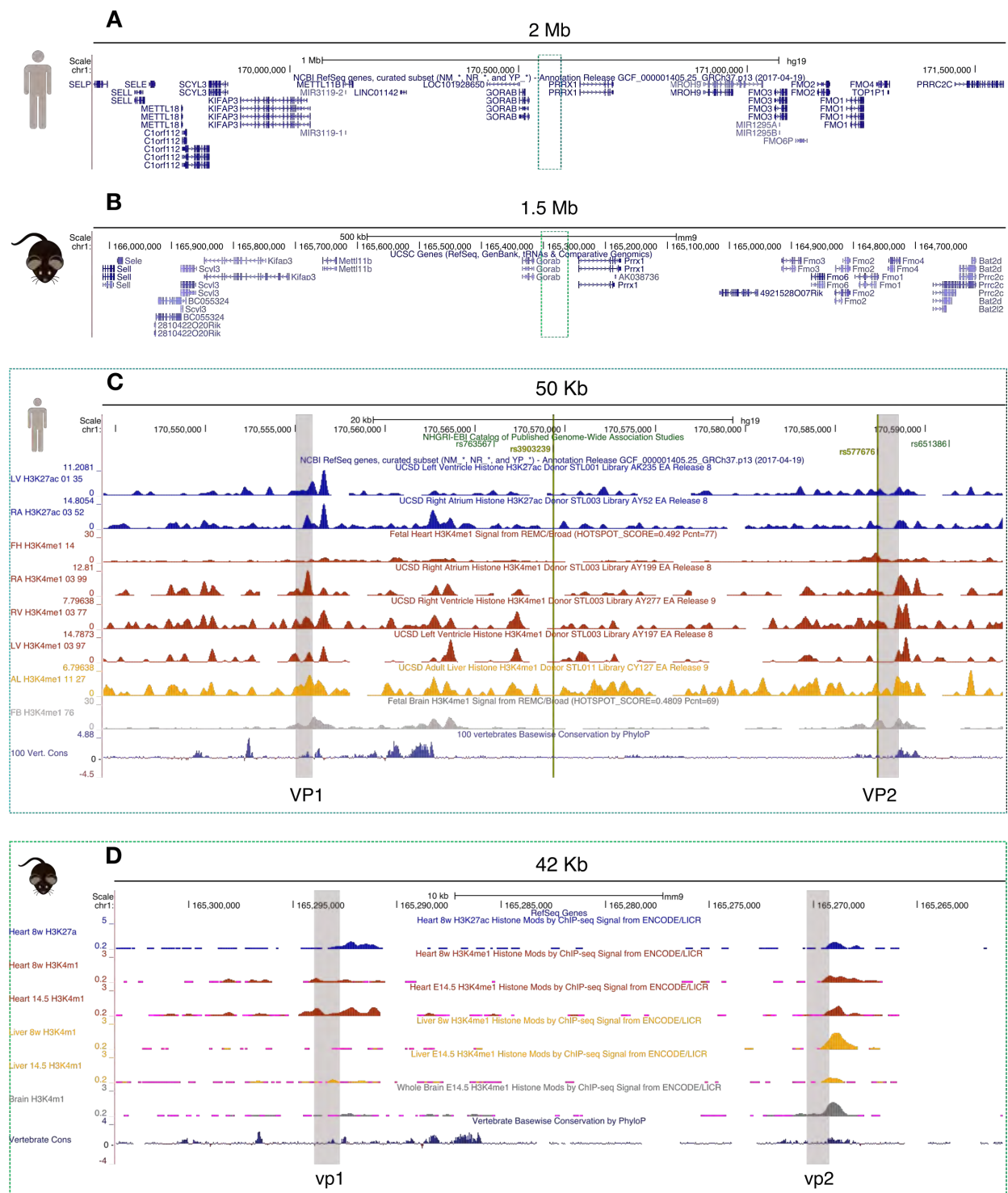
**B)** Syntenic 84 Kb genomic window in the mouse genome (mm9; chr9:58,633,900-58,718,236). AF associated SNP is located at 9 Kb from *HCN4* promoter this being not possible to distinguish by 4C-seq interactions due to the promoter or to nearby regulatory elements. Thereby, this region was discarded for 4C-analysis.

## 5q31 locus genome analysis



- A)** Genomic window of 50 Kb of the 5q31 locus (hg19; chr5:137,394,989-137,444,989). First track shows AF associated SNPs in light brown lettering and other SNPs in green.
- B)** Syntenic 21 Kb genomic window in the mouse genome (mm9; chr18:34,697,477-34,718,787). AF associated SNP is located at 200 bp from *WNT8A* promoter this being not possible to distinguish by 4C-seq interactions due to the promoter or to nearby regulatory elements. Thereby, this region was discarded for 4C-analysis.

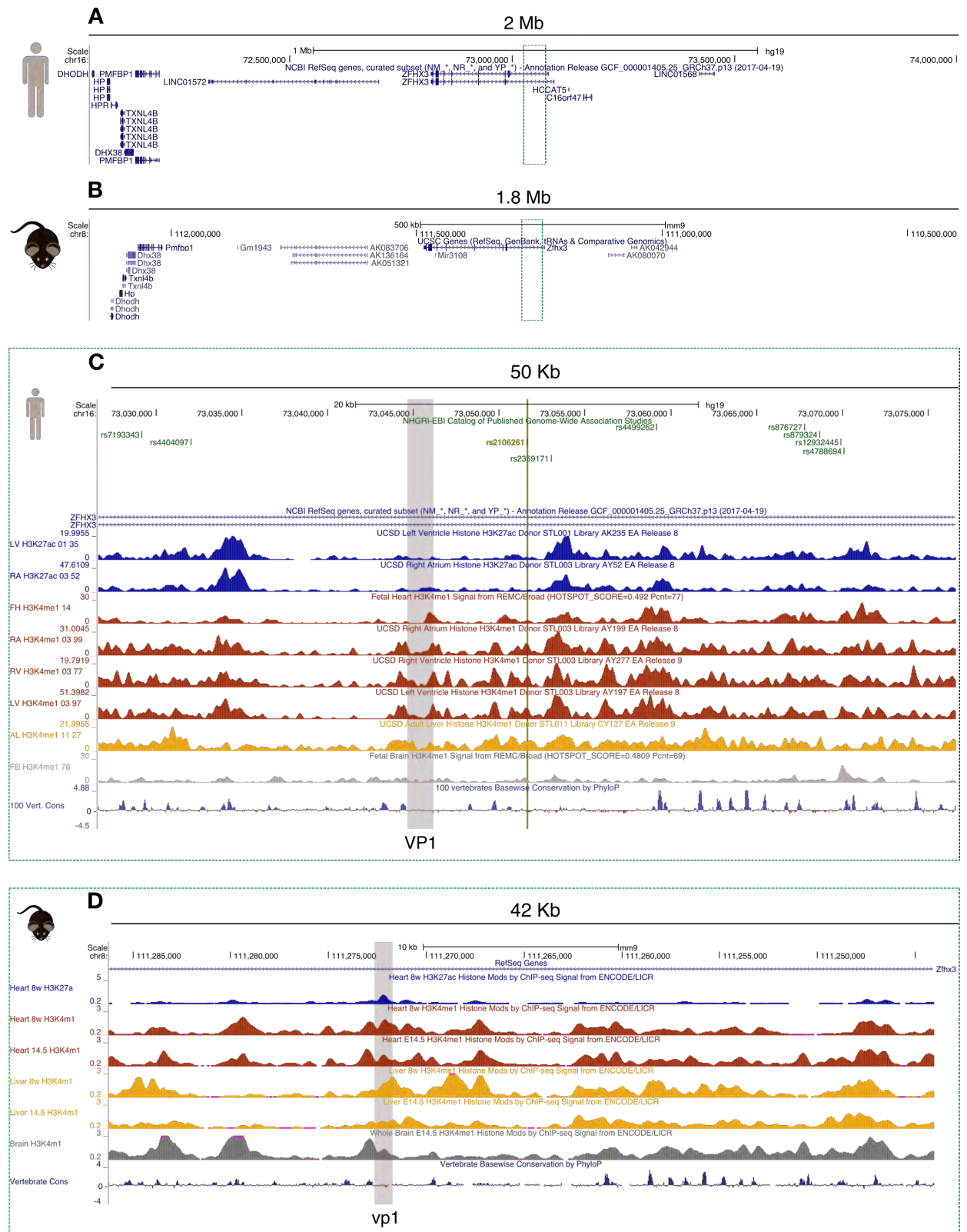
## 1q24 locus genome analysis



**A)** Genomic window of 2Mb of the 1q24 locus (hg19; chr1:169,569,317-171,569,317).  
**B)** Syntenic 1.5 Mb genomic window of the mouse genome (mm9; chr1:164,600,792-166,069,863). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in A (hg19; chr1:170,544,317-170,594,317) showing AF associated SNPs in light brown lettering and other SNPs in green. Below epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA),

fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The two equivalent regions to the viewpoint selected for 4C-seq in mouse are highlighted in light grey. **D)** Genomic window of 50 Kb corresponding to the light green rectangle in B (mm9; chr1:165,260,906-165,303,500). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart, 8 w mouse liver, E14.5 mouse liver, and E14.5 mouse whole brain are shown. Bottom track shows vertebrate conservation. The two viewpoints selected for 4C-seq experiments are highlighted in light grey. These viewpoints correspond to the syntenic regions that in the human genome contain AF associated SNPs.

16q22 *locus* genome analysis



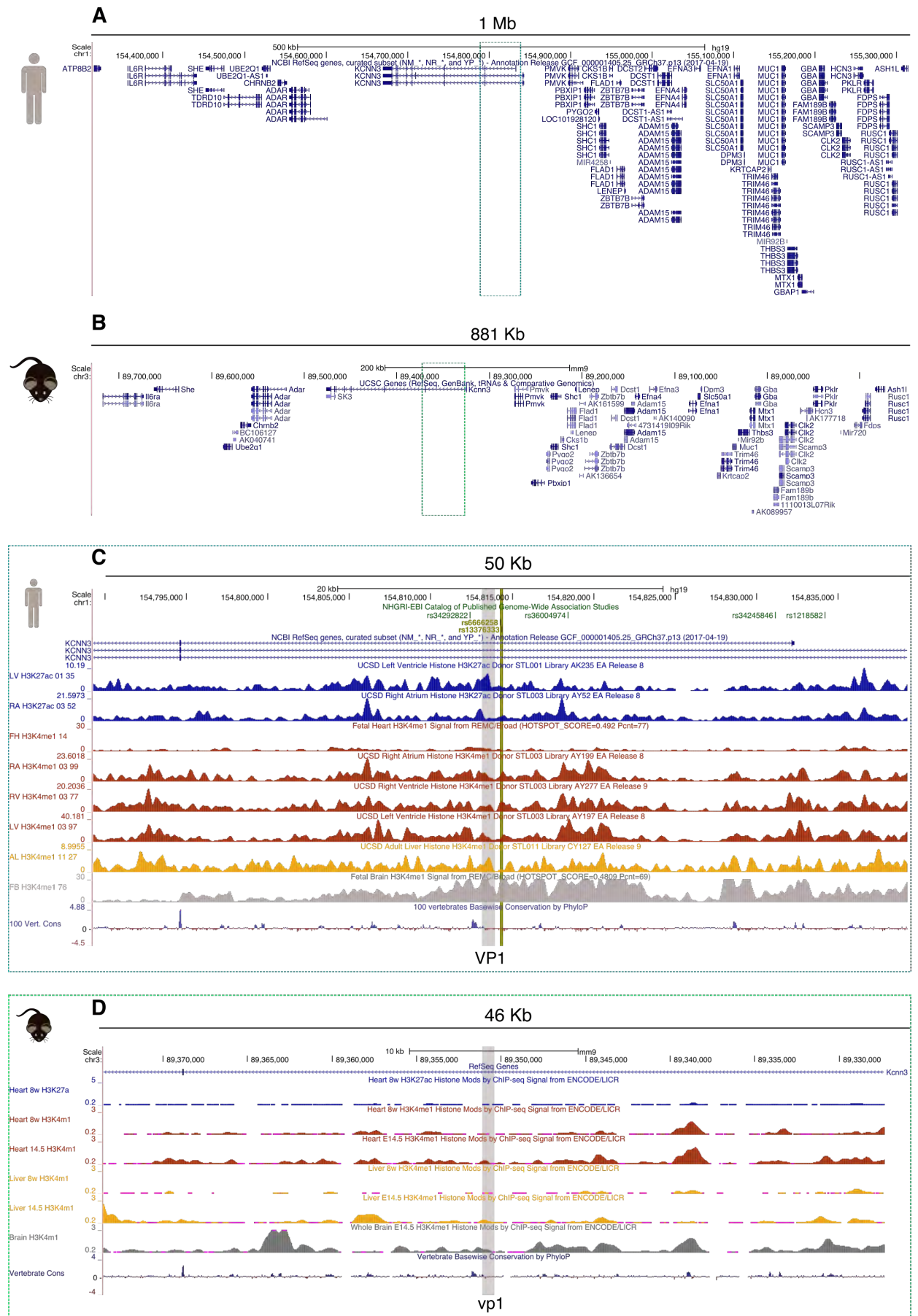
**A)** Genomic window of 2Mb of the 16q22 *locus* (hg19; chr16:72,051,620-74,051,620).

**B)** Syntenic 1.8 Mb genomic window of the mouse genome (mm9; chr8:110,361,166-

112,124,918). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in

A (hg19; chr16:73,026,620-73,076,620) showing AF associated SNP in light brown lettering and other SNPs in green. Below epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA), fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The equivalent region to the viewpoint selected for 4C-seq in mouse is highlighted in light grey. **D**) Genomic window of 42 Kb corresponding to the light green rectangle in B (mm9; chr8:111,244,027-111,286,260). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart, 8 w mouse liver, E14.5 mouse liver, and E14.5 mouse whole brain are shown. Bottom track shows vertebrate conservation. The viewpoint selected for 4C-seq experiments is highlighted in light grey. This viewpoint corresponds to the syntenic region that in the human genome contain AF associated SNP.

# 1q21 locus genome analysis



**A)** Genomic window of 1Mb of the 1q21 *locus* (hg19; chr1:154,314,268-155,314,268). **B)** Syntenic 880 Kb genomic window of the mouse genome (mm9; chr3:88,872,702-89,753,591). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in A (hg19; chr1:154,789,268-154,839,268) showing AF associated SNP in light brown lettering and other SNPs in green. Below epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA), fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The equivalent region to the viewpoint selected for 4C-seq in mouse is highlighted in light grey. **D)** Genomic window of 46 Kb corresponding to the light green rectangle in B (mm9; chr3:89,327,302-89,373,558). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart, 8 w mouse liver, E14.5 mouse liver, and E14.5 mouse whole brain are shown. Bottom track shows vertebrate conservation. The viewpoint selected for 4C-seq experiments is highlighted in light grey. This viewpoint corresponds to the syntenic region that in the human genome contain AF associated SNP.





77,642,796). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in A (hg19; chr14:64,655,848-64,705,848) showing AF associated SNP in light brown lettering and other SNPs in green. Below epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA), fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The equivalent region to the viewpoint selected for 4C-seq in mouse is highlighted in light grey. **D)** Genomic window of 53 Kb corresponding to the light green rectangle in B (mm9; chr12:77,175,847-77,229,724). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart, 8 w mouse liver, E14.5 mouse liver, and E14.5 mouse whole brain are shown. Bottom track shows vertebrate conservation. The viewpoint selected for 4C-seq experiments is highlighted in light grey. This viewpoint corresponds to the syntenic region that in the human genome contain AF associated SNP.



17,722,331). **C)** Genomic window of 50 Kb corresponding to the light blue rectangle in A (hg19; chr7:116,161,241-116,211,241) showing AF associated SNP in light brown lettering and other SNPs in green. Below epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from left ventricle (LV), right atrium (RA), fetal heart (FH), right ventricle (RV), adult liver (AL), and Fetal Brain (FB) from the NIH Roadmap Epigenomics Project. Lower track shows vertebrate conservation. The equivalent region to the viewpoint selected for 4C-seq in mouse is highlighted in light grey. **D)** Genomic window of 45 Kb corresponding to the light green rectangle in B (mm9; chr6:17,253,680-17,298,207). Epigenetic marks (H3K27ac, active enhancers and promoters; H3K4me1, active enhancers) from 8 weeks (w) mouse heart, E14.5 mouse embryonic heart, 8 w mouse liver, E14.5 mouse liver, and E14.5 mouse whole brain are shown. Bottom track shows vertebrate conservation. The viewpoint selected for 4C-seq experiments is highlighted in light grey. This viewpoint corresponds to the syntenic region that in the human genome contain AF associated SNP.

This thesis includes a CD with the following appendixes and extra information:

**APPENDIX 2:** Table with differentially expressed genes and proteins obtained after RNA-seq and LC/MS-MS data analysis in posterior left atria and isolated cardiomyocytes of an AF induced sheep model.

**APPENDIX 3:** Functional enrichment analysis gene ontology terms from RNA-seq and LC/MS-MS differentially expressed genes and proteins respectively in posterior left atria and isolated cardiomyocytes of an AF induced sheep model.

**MOVIE 0:** Representative video of BJ control hiPSCs-CMs successful differentiation beating.

**MOVIE 1:** Representative video of clone 10 hiPSCs-CMs successful differentiation beating.

**MOVIE 2:** Representative video of clone 19 hiPSCs-CMs successful differentiation beating.

**MOVIE 3:** Representative video of clone 21 hiPSCs-CMs successful differentiation beating.

**WashU EpiGenome Browser Sessions:** Files for 4C-seq data visualization in the WashU EpiGenome Browser mm9.



## PUBLICATIONS

---

*If you know that you are on the right track,  
if you have this inner knowledge,  
then nobody can turn you off...  
no matter what they say.*

Barbara McClintock





## PUBLICATIONS

The work described in this thesis regarding AF induced sheep model is included in the following manuscript that will be submitted this year:

Alvarez A†, **Rouco R**†, Guerrero-Serna G, Tiana M, Cogliati S, Kaur K, Saeed M, Enriquez JA, Jalife J\*, Manzanares M\* (2019). Transcriptome and Proteome Mapping in the Sheep Reveal Molecular Features of Atrial Fibrillation Progression. *Circulation Res*, to submit (†co-first authors; \*co-corresponding authors).

The collaboration in other research projects at Miguel Manzanares lab during the development of the thesis has resulted in the following publication already submitted:

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).





