

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

CARACTERIZACIÓN DE PERFILES TRANSCRIPTÓMICOS Y EPIGENÓMICOS EN LA DIFERENCIACIÓN DE CELULAS TRONCALES EMBRIONARIAS A NEURONAS DOPAMINÉRGICAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

CÉSAR DANIEL MELÉNDEZ RAMÍREZ

TUTOR J. IVÁN VELASCO VELÁZQUEZ Instituto de Fisiología Celular, UNAM

COMITÉ TUTOR:

ERNESTO SOTO REYES SOLÍS UAM Cuajimalpa

VÍCTOR JULIÁN VALDEZ GONZÁLEZ Instituto de Fisiología Celular, UNAM

Ciudad de México. Noviembre, 2024



Universidad Nacional Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



PROTESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL

(Graduación con trabajo escrito)

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado:

Caracterización de perfiles transcriptómicos y epigenómicos en la diferenciación de células troncales embrionarias humanas a neuronas dopaminérgicas

que presenté para obtener el grado de doctor es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

César Daniel Meléndez Ramírez No. de cuenta 311314283

Contenido

1 Agradecimientos
2 Dedicatoria
3Resumen 9
4 Abstract
5 Introducción11
5.1 Potencialidad y epigenética11
5.2 Metilación del DNA
5.3 Modificaciones post-traducionales en las histonas12
5.4 Regiones reguladoras del genoma del genoma14
5.5 Enhancers transcripcionales14
5.6 Accesibilidad de la cromatina en enhancers15
5.7 Modificaciones post-traduccionales en las histonas y otras proteínas asociadas a regiones enhancer16
5.8 Trascripción para la detección de enhancers18
5.9 Estructura tridimensional de la cromatina y enhancers transcripcionales 18
5.10 Enhancers transcripcionales y enfermedades
5.11 Enhancers transcripcionales y enfermedades del sistema nervioso central 21
5.12 Enhancers transcripcionales y desarrollo del sistema nervioso central 22
5.13 Células troncales
6 Antecedentes
7 Planteamiento del problema:
8 Hipótesis
9 Objetivo general
10 Objetivos particulares:
11 Resultados:
11.1 Ensayos de ChIP-Seq en neuronas dopaminérgicas
11.2 Definición de enhancers putativos en células troncales embrionarias y neuronas dopaminérgicas del mescencefalo
11.3 Identificación de enhancers putativos en <i>loci</i> de genes relevantes para la pluripotencia y para la inducción dopaminérgica

11.4 Analisis de inducción dop	e enhancers putativ aminérgica	vos candidatos a	a ser reguladore	s de la
11.5 El locus de	e TAGLN3 contiene	e un enhancer es	pecifico de la ir	nducción neural
11.6 Enriquecir celulas troncal	niento diferencial c es pluripotentes y ⊧	le las histonas H neuronas dopam	I3K27ac y H3K4 hinérgicas del m	me1 entre escencefalo 34
11.7 Análisis de neuronas dopa	e regiones con enri minérgicas del me	iquecimiento de scencefalo huma	marcadores de anas	enhancers en
11.8 Análisis de dopaminérgica	e la distribución ge	nómica de RFX4	a través de la c	diferenciación 39
12Discusión:				
13Perspectivas:				
14 Conclusión:				
15 Materiales y	métodos:			
15.1 Cultivo ce	lular y diferenciacio	ón dopaminergio	a	
15.2 Extracciór	n de cromatina			
15.3 Inmunopre	ecipitación de la cro	omatina		
15.4 Secuencia	ción de Nueva Ger	eración		
15.5 Análisis b	ioinformático del C	hIP-Seq		
15.6 Ensayo de	CUT&RUN			
15.7 Preparació	ón de las biblioteca	s de CUT&RUN.		51
15.8 Procesam	iento de los datos o	de CUT&RUN		
16 Referencias:				53
17Anexo:				

1.- Agradecimientos

Este trabajo se realizó en el laboratorio del Doctor Iván Velasco Velázquez en el Instituto de Fisiología Celular de la UNAM.

Agradezco a mi comité tutoral por el tiempo dedicado en la revisión y discusión de mi proyecto. Siempre tuvieron la mejor disposición:

Dr. J. Iván Velasco VelázquezDr. Ernesto Soto Reyes SolísDr. Víctor Julián Valdés Rodríguez

A mis sinodales, por la rápida revisión de mi trabajo de tesis y sus valiosos comentarios:

Dra. Rosa Estela Navarro González	Dr. Felix Recillas Targa
Dra. Selma Eréndira Avendaño Vázquez	Dr. David Valle García
Dra. María Antonieta Chávez González	

A la Dirección General de Asuntos del Personal Académico-UNAM (Papiit-UNAM IN 213719) por el financiamiento otorgada para la realización de este proyecto.

Al Consejo Nacional de Ciencia y Tecnología (Conacyt 256092 y 272815) por el financiamiento otorgada para la realización de este trabajo.

A la Dra. Itzel Escobedo Ávila del Instituto de Fisiología celular por todo el apoyo técnico proporcionado durante la realización este trabajo.

A la Dra. Raquel Cuevas Diaz Duran de la Escuela de Medicina y Ciencias de la Salud, Tecnológico de Monterrey, por todo el apoyo en el análisis de datos y a lo largo del desarrollo del proyecto.

A la Dra. Rosa Rebollar y a Aylin del Moral por todo el apoyo técnico brindado al proyecto.

A Mayela Giacoman Lozano por todo el apoyo conceptual, experimental y bioinformático a lo largo del desarrollo del proyecto.

Al Dr. Wolfgang Fischle por permitirme hacer una estancia en KAUST, Arabia Saudita.

A Jorge Landgrave y a la Fernanda Vargas, por todos los recursos que aportaron a este proyecto, así como su apoyo técnico.

A el Programa de Apoyo a los Estudios del Posgrado (PAEP) por brindarme apoyo para asistir a un congreso nacional y uno internacional durante la realización de mi doctorado.

2.- Dedicatoria

Para mi mamá y mi papá, no solo por haberme dado la guianza necesaria para volverme un adulto integro, sino por haberme brindado apoyo incondicional en cada etapa de mi formación academica. Espero algun dia ser una persona tan extraordinaria como ustedes.

Para mis abuelitos, Yolanda, Antonio, María y Luis. Es un privilegio compatir este logro con todos ustedes. Gracias por siempre haberme tratado con un cariño incondicional. Me considero extramadamente afortunado de que sus familias se unieran para que se formara la mia.

Para mi hermana Paola, porque siempre estas ahí para darme consejos sobre sobre las cosas de la vida que no entiendo bien, como que regalarle a la gente en sus cumpleaños. Nunca cambies tu forma de ser.

Para el resto de mi familia, en especial Eduardo Ramírez, Luis Alvaréz, Irma Trejo, Isela Soto, Lourdes Trejo, Gloria Trejo y Gina Garcia, Fabian Meléndez, David Meléndez y Fabiola Meléndez.

Para mi tutor, Iván Velasco, por toda lo que hiciste por mi, la confianza y el apoyo. Gracias por permitir que el AL-101 fuera mi segundo hogar por casi nueve años. Sé que en ti siempre encontrare un valioso colega y amigo.

Para Raquel Cuevas, por haber sido mi tutora no oficial. Sin tu apoyo y guianza no hubiese sido posible hacer proyectos con este alcance.

Para Mayela Giacoman, no me pudo haber tocado una mejor colaboradora para este proyecto. Se siente bonito habernos formado como científicos a la par, 300 reuniones de zoom despues.

Para Ernesto Soto y Julian Valdez, porque su presencia en el proyecto me hizo sentir siempre respaldado, tanto a nivel tecnico como conceptual. No se me olvidaran ninguna de sus enseñanzas.

Para Rodrigo Barrios y Lissania Guerra, por compartirme su visión de la biologia molecular y la epigenetica.

Para Itzel Escobedo A. y a Daniel Cortes, mas que tecnicos de laboratorio, fueron valiosos compañeros y amigos.

Para Fernanda Vargas y Jorge Langrave, por haberme apoyando tanto durante mi estanica en KAUST. Esa experiencia forma parte importante de lo que soy hoy. Espero aplicar correctamente todas sus enseñansas.

Para Karla Mendez y Alejandro Arce, por haberse interesado en mi proyecto, echarme porras y confiar en mi.

Para Manuel León, por tu amistad y porque sin ti nunca hubiera echado a andar mi proyecto de licenciatura. Me ayudaste a darle forma y a no deprimirme tras 40 Western blots fallidos.

Para Adolfo López, por haberme enseñado a cultivar celulas troncales humanas de la manera mas divertida posible.

Para los miembros anteriores del AL-101, en especial Alonso Ortega, Rolando Lara, Jose Luis Rodriguez, Angel Polanco, Beetsi Urieta, María Jose Castellanos y Karla Itzel. Todos hemos tomado caminos distintos, pero que alregia saber que ahí donde nos encontremos, habra un pedacito de todo lo que vivimos juntos.

Para Brenda Flores y Luis Delgado. Con ustedes comenzó este camino, en aquellas interminables charlas de área II en la prepa. Gracias a su compañía y esas conversaciones, tomé la decisión de seguir por este rumbo. Sin ustedes, no sé si habría llegado hasta aquí.

Para Cristian Garcia F., Alonso Villareal, Daniel Mireles, Omar Proo, Zyanya Alanis, Ricardo Gonzalez, Oscar Salinas, Chucho Salinas, Cristian G. Garcia, Diego Garcia, Oscar Solis, Fernanda Trejo, Antonio Navedo, Alonso Suarez, Aldo Martinez y Franco Giordani. La amistad que tenemos esta mas alla del tiempo y la distancia. Espero que la vida siempre los ponga en mi camino, poque ninguna pena sera mas fuerte que unas cervezas con ustedes.

Para Isabel Correa, por siempre estar segura de que iba a poder con lo que me propusiera, tu compañía ha sido una fuente invaluable de motivación a lo largo de este camino.

Para Luis Angel, Lili y Andy. Gracias por ser mi otra familia.

Para los miembros actuales del AL-101, en especial Juan Jair Santillan, Fernando Becerra, y Margareht Reyes. El AL-101 se queda en buenas manos.

Para Arturo Soto, Ignacio Alvirde, Alexis Monje y Juan P. Segura. Gracias por las oportunidades y la confianza.

Para Olivia Pineda, Ricardo Nilo, Ana Morales, Karen Castellanos, por ser grandes compañeros de trabajo.

Para mis otros colegas del Instituto de fisiologia celular, es especial Fernando, Nuri, Alex e Itzel.

Para Jose Reyes, Pamela Rendon y Miguel Franco, por integrarme en su grupo y ser miembros recurrentes de los miercoles de podcast.

Para la biologia y la ciencia méxicana. Gracias.

3.-Resumen

El cerebro humano cuenta con más de 100 tipos celulares, todos generados mediante mecanismos muy finos de control de la expresión génica durante el desarrollo. Uno de esos tipos celulares son las neuronas dopaminérgicas del mesencéfalo (mNDA); un subtipo neuronal importante por ser el principal tipo celular afectado en la Enfermedad de Parkinson (PD). Las células troncales pluripotentes humanas (hPSCs) se han diferenciado con éxito en mNDA. Se ha demostrado la funcionalidad de las mNDA resultantes de estos protocolos. Sin embargo, es necesaria una caracterización epigenómica y transcriptómica de estas neuronas para comprender los mecanismos reguladores que inducen la especificación dopaminergica. La expresión génica está controlada por la interacción de factores de transcripción, promotores y enhancers. Se ha demostrado que las modificaciones de histonas y la accesibilidad de la cromatina están relacionadas con el estado de activación de enhancers. Para comprender cómo se logran los programas específicos de expresión génica en distintas estirpes celulares, el proyecto ENCODE y el proyecto Roadmap Epigenomics han generado información sobre los perfiles epigenómicos de diversos tejidos humanos y linajes celulares. Sin embargo, hasta donde sabemos, la epigenómica de las neuronas mDA humanas en desarrollo no ha sido abordada aún. Para obtener información sobre los mecanismos reguladores involucrados en la especificación celular de las mNDA, evaluamos la transcripción (RNA-Seg) y el estado de accesibilidad de la cromatina (ATAC-Seq) en distintas etapas del proceso de diferenciación in vitro de mNDA. Encontramos cambios significativos en la expresión génica y la accesibilidad de la cromatina. Para caracterizar más a fondo las regiones de cromatina abierta, realizamos ChIP-Seq de modificaciones de histonas (H3K27ac y H3K4me1) en este mismo modelo de diferenciación in vitro, para localizar posibles regiones enhancers. La integración de esta información epigenómica nos permitió identificar enhancers putativos que orquestan la inducción dopaminérgica. Encontramos una correlación entre la posición de los enhancers putativos y la expresión génica. El análisis de los motivos de unión de factores de transcripción sugirió el papel del factor de transcripción RFX4 en los enhancers putativos específicos de mNDA. Mapeamos los sitios de unión de la cromatina de RFX4 mediante CUT&RUN a través del protocolo de diferenciación dopaminérgica in vitro. Encontramos que los sitios de unión de RFX4 están asociados con regiones accesibles en mNDA y con un aumento en la transcripción de genes cercanos, algunos de los cuales son reguladores bien conocidos de la neurogénesis y del desarrollo del sistema nervioso central.

9

4.- Abstract

The human brain has over 100 cell types, generated and maintained through finely tuned gene expression mechanisms during development, adulthood, and in response to environmental stimuli. One of these cell types is midbrain dopaminergic neurons (mDA), an important neuronal subtype, as it is the primary affected cell type in Parkinson's Disease (PD). Human pluripotent stem cells (hPSCs) have been successfully induced into mDA. The functionality of the resulting mDA neurons has been demonstrated. Nevertheless, a comprehensive epigenomic and transcriptomic characterization is needed to understand the regulatory mechanisms driving mDA neuron cell specification. Gene expression is driven by an interplay of transcription factors, promoters and enhancers. It has been demonstrated that histone modifications and chromatin accessibility are related to the activation status of enhancers. To understand how cell-specific gene expression programs are achieved, the ENCODE Project Consortium and the Roadmap Epigenomics Project have generated epigenomic signatures of diverse human tissues and cell lineages. However, to the best of our knowledge, the epigenomics of differentiated human mDA neurons has not been addressed. To gain insight into the regulatory mechanisms involved in mDA neuron cell specification, we obtained mDA expression profiles (RNA-Seq) and chromatin accessibility maps (ATAC-Seq) at different stage of the in vitro differentiation protocol. After data analysis we found significant changes in gene expression and chromatin accessibility. To further characterize open chromatin regions, we performed histone modification (H3K27ac and H3K4me1) ChIP-seg on in vitro generated mDA, to locate possible enhancer regions. The integration of this epigenomic information allowed us to obtain putative active enhancers driving dopaminergic induction. We found a correlation between putative enhancer position and gene expression. Transcription factor binding motif analysis suggested a role of the RFX4 transcription factor on mDA-specific enhancer regions. We mapped RFX4 chromatin binding sites by CUT&RUN on embryonic stem cells, and in vitro generated neural precursors and mDA neurons. We found that RFX4 binding sites were associated with mDA accessible regions and increased transcription of nearby genes, some of which are wellknown regulators of neurogenesis and central nervous system development.

5.- Introducción

5.1 Potencialidad y epigenética

El ciclo de vida de los metazoarios comienza tras la fertilización del óvulo (**Fig. 1A**), dando origen al cigoto (**Fig. 1B**) y posteriormente al embrión (**Fig. 1C**). A partir de éste, se generarán cientos de tipos celulares con virtualmente idéntico material genético pero distinto potencial de diferenciación (Riveiro & Brickman, 2020) y distinta actividad biológica (Pladevall-Morera & Zylicz, 2022) (**Fig. 1D**). Los mecanismos por los cuales un mismo genoma es capaz de producir esta variabilidad funcional son producto de procesos epigenéticos, es decir, el establecimiento de firmas moleculares heredables que modifican la actividad transcripcional en respuesta a estímulos ambientales sin alterar la secuencia de bases nitrogenadas del DNA (Misteli, 2007).



Mismo genoma

Figura 1. En los metazoarios, el desarrollo de todos los tejidos sucede a partir de una célula inicial llamada cigoto. A) Momento de la fecundación del óvulo por un espermatozoide. B) El cigoto es la célula resultante de la fecundación. C) El cigoto se desarrolla y segmenta, formando a un embrión. D) El genoma del embrión contiene toda la información necesaria para producir todas las células y tejidos presentes en un organismo.

Existe una gran variedad de niveles de regulación epigenéticos (Allis & Jenuwein, 2016), como la metilación en el carbono cinco de las citocinas del DNA (Bestor, 2000; Meehan et al., 1989), las modificaciones postraduccionales en las histonas (Rando, 2012),

las variantes de histonas (Henikoff & Smith, 2015), la actividad de los RNAs largos no codificantes (IncRNAs) (Khalil et al., 2009), la estructura tridimensional de la cromatina (Tang et al., 2015) y la regulación que estos elementos ejercen sobre los enhancers transcripcionales (Andersson et al., 2014).

5.2 Metilación del DNA

De todos los procesos epigenéticos, la metilación de las citocinas es posiblemente el que más se ha estudiado. La metilación del DNA juega un papel importante en el desarrollo temprano de los mamíferos, mediante el silenciamiento de genes que impiden la diferenciación celular (Smith & Meissner, 2013). Se han descrito varios mecanismos por los cuales influye en la expresión génica, generalmente asociados con represión transcripcional (Baylin, 2005; Roadmap Epigenomics Consortium et al., 2015), como el reclutamiento de proteínas represoras de unión a DNA metilado (MECP2), que impiden el reconocimiento de secuencias blanco por factores transcripcionales y promueven la acción de complejos remodeladores de la cromatina (Clemens et al., 2020). Esta marca es dinámica: se ha descrito la familia de enzimas que catalizan su deposición en el DNA de mamíferos (DNMTs) (Bestor & Ingram, 1983), así como la familia de enzimas que median la oxidación de la marca para su posterior remoción (TETs) (Tahiliani et al., 2009). La metilación del DNA influencia otros mecanismos epigenéticos como las modificaciones de histonas (J. Du et al., 2015).

5.3 Modificaciones post-traducionales en las histonas

El DNA se encuentra empaquetado en el núcleo eucarionte dando casi dos vueltas (~147 pares de bases) a multímeros de histonas, conformados por dos histonas H3, dos histonas H4 y dos dímeros de histonas H2A-H2B, es decir, un octámero (Richmond et al., 1984). Se han reportado una gran cantidad de modificaciones postraduccionales localizadas principalmente en los extremos amino-terminales de estas histonas y se ha descrito que estas modificaciones regulan la expresión génica (Strahl & Allis, 2000). Las modificaciones más estudiadas son la acetilación y metilación en residuos de lisina, y la fosforilación de serinas. Estas marcas son dinámicas y su regulación está controlada por la actividad antagónica de enzimas que las adicionan o las remueven (Husmann & Gozani, 2019).

Las metilaciones en la histona H3 participan en una gran variedad de procesos biológicos. Por ejemplo, la tri-metilación de la lisina 4 de la histona H3 (H3K4me3) está enriquecida en los sitios de inicio de la transcripción (TSS) y podría estar involucrada en el reclutamiento de la RNA Polimerasa II, teniendo un impacto positivo en la trascripción (Vermeulen et al., 2007), mientras que las formas menos metiladas de este mismo residuo (me2/me1) se encuentran distribuidas en regiones alejadas a los promotores, y se sabe que están asociadas a elementos reguladores como enhancers, principalmente la H3K4me1 (Birney et al., 2007).

La metilación en lisinas también puede estar implicada en la represión transcripcional, principalmente en el caso de H3K27me3 y H3K9me2/3 (Black et al., 2012). La H3K9me2/3 tiene una distribución discreta en células poco diferenciadas, como las células troncales embrionarias (ESCs) (~4% del genoma), mientras que puede ocupar hasta el ~40% del genoma en tejidos diferenciados. Además, las regiones enriquecidas con esta modificación se encuentran inactivas transcripcionalmente, por lo que se hipotetiza que su función es reprimir definitivamente genes asociados al desarrollo y formar "heterocromatina constitutiva" (Wen et al., 2009). La H3K27me3 está asociada a represión por medio del complejo Polycomb, el cual cuenta con dos subunidades: PRC1 y PRC2; PRC2 cataliza la H3K27me3 y esta modificación es "leída" por PRC1, la cual se asocia a remodeladores de la cromatina silenciadores (Grossniklaus & Paro, 2014) y desacetilasas de histonas (HDAC) (Huang et al., 2002).

La acetilación ocurre en diversos residuos de lisina a lo largo de las histonas, se asocia a activación genica (Allis & Jenuwein, 2016) y está regulada por acetiltransferasas de histonas (HATs) y desacetilasas de histonas (HDACs) (Bannister & Kouzarides, 2011). La acetilación en la lisina 27 de la histona H3 (H3K27ac) es una de las modificaciones más ampliamente estudiadas y entendidas. Se sabe que participa activamente en la expresión de RNA por medio de distintos mecanismos, por ejemplo, interactuando con proteínas facilitadoras de la elongación (Dey et al., 2003).

La integración a nivel genómico de las modificaciones postraduccionales de histonas nos ha permitido definir "estados de la cromatina", como promotores activos/inactivos, regiones silenciadas y enhancers activos/inactivos, incrementando nuestro nivel de asociación y compresión entre estas marcas y otros elementos de la cromatina (Hoffman et al., 2013).

5.4 Regiones reguladoras del genoma del genoma

El genoma humano mide alrededor de tres mil millones de pares de bases (3 Gpb) y contiene aproximadamente 20,000 genes codificantes para proteínas, lo que constituye solo el 3% del su longitud total (I. H. G. S. Consortium, 2004). También se han descrito mas de 20,000 genes no codificantes, clasificados en una gran variedad de RNAs no codificantes (Cech & Steitz, 2014), y finalmente, el resto esta enriquecido con secuencias reguladoras como promotores y enhancers (T. E. P. Consortium, 2012). El control transcripcional ejercido por estos elementos es otro nivel de regulación epigenética (Misteli, 2007).

Las secuencias reguladoras mas ampliamente estudiadas son los promotores core, definidos como la región circundante al sitio de inicio de la transcripción (TSS), la cual contiene motivos de unión a factores transcripcionales (TFs) capaces de reclutar a la RNA-Pol II (Sandelin et al., 2007). Existen métodos que permiten identificar sitios de inicio de la transcripción a escala genómica, como la secuenciaicón masiva de RNA dependiente del Cap, también llamada CAGE-Seq (*Cap analysys of gene expression*), la cual se basa en la biotinilación de todas las moléculas de RNA que contengan la caperuza 7-metilguanosina 5' (**Fig. 2J**) para su posterior captura (**Fig. 2K**) (Kodzius et al., 2006). Esto ha permitido mapear con gran precisión todos los TSS del genoma humano y los promotores core asociados (FANTOM Consortium & Riken Pmi, 2014). Los datos de CAGE también permiten la localización de otros elementos como enhancers (**Fig. 2G**).

5.5 Enhancers transcripcionales

El término enhancer fue acuñado en 1981, cuando Julian Banerji y sus colaboradores describieron el efecto que tenía un repetido de 72 pb correspondiente al virus SV40 sobre la expresión de la β-globina de conejo en células HeLa. La expresión de la construcción con el enhancer era 200 veces mayor que sin la región viral. Los investigadores además concluyeron que el efecto del enhancer era independiente de la orientación de la secuencia y de la distancia al gen (Banerji et al., 1981).

Los enhancers transcripcionales son secuencias genéticas con funciones de regulación distal, pues a diferencia de los promotores core, son capaces de controlar la expresión de genes ubicados a miles o hasta millones de pares de bases (pb) (Hafner & Boettiger, 2023). Las regiones enhancer están conservadas evolutivamente y son ricas en motivos de unión para TFs (Kwak, Fuda, Core, & Lis, 2013). Se ha descrito que estos

elementos regulan la expresión de varios genes simultáneamente y que cada gen está influenciado por varios enhncers a la vez (Lorberbaum et al., 2016). Sin embargo, al no tener una ubicación definida, determinar la localización de estos elementos y sus posibles genes blanco es metodológicamente desafiante (Chen et al., 2023).

5.6 Accesibilidad de la cromatina en enhancers

El nivel de compactación del DNA impacta en los procesos transcripcionales (Mansisidor & Risca, 2022). Los nucleosomas pueden regular el nivel de permisibilidad espacial de la cromatina al modular su periodicidad de aparición, la cual es en promedio de ~200 pb (Gupta et al., 2008). Las regiones reguladores del genoma, como promotores y enhancers, están libres de nucleosomas (nucleosome-free) (Boyle et al., 2008; Thurman et al., 2012). Existe evidencia de que el nivel de accesibilidad de los enhancers esta directamente relacionado con el nivel de actividad de estos elementos (Bozek & Gompel, 2020; Forrest et al., 2017), aunque es importante recalcar que, en ocasiones, los enhancers inactivos pueden tener grados de accesibilidad detectable o que algunos enhancers fuertemente activos pueden tener niveles bajos de accesibilidad de cromatina (Arnold et al., 2013).

Existen métodos para medir la accesibilidad de la cromatina de manera global y esto facilita la localización de enhancers putativos (**Fig. 2A**) (Koh et al., 2016). Inicialmente se describió que algunas regiones genómicas eran más susceptibles a la acción de endonucleasas, como la DNAsa I, y se identificó a estas regiones como posibles sitios de regulación, altamente susceptibles a la unión de factores transcripcionales y remodeladores de la cromatina (**Fig. 2B**). Actualmente, la hipersensibilidad a la DNAsa I es considerado el método menos preciso y menos eficiente para determinar accesibilidad de la cromatina (**Gross & Garrard**, 1988).

Una medida indirecta de accesibilidad de la cromatina es el posicionamiento de los nucleosomas, el cual puede evaluarse por medio de MNAse-Seq (**Fig. 2C**) (Schones et al., 2008). La MNAsa es una nucleasa micrococal con la capacidad de cortar el DNA que hay entre nucleosomas, también llamado DNA *linker*, pues éste está "desprotegido" (Yuan et al., 2005). Los mapas de MNAse-Seq permiten encontrar regiones del genoma en las que los nucleosomas tienen una disposición bien definida o en "fase", como promotores y enhancers; también existen regiones en las que los nucleosomas están "difusos" y a menudo corresponden a regiones genómicas sin potencial regulatorio (C. Jiang & Pugh, 2009).

La siguiente técnica desarrollada para evaluar accesibilidad de la cromatina a escala genómica fue el aislamiento de regiones reguladoras asistido por folmaldehído seguido de seceunciación o FAIRE-Seq (*Formaldehide-Asisted Isolation of Regulatory Elements*) (**Fig. 2D**). En esta técnica se fija la cromatina mediante formaldehído, generando enlaces covalentes entre DNA e histonas en regiones compactas (**Fig. 2E**), pero dejando DNA "desnudo" en regiones accesibles; ese DNA libre es aislado, secuenciado y mapeado (**Fig. 2F**) (Giresi et al., 2007; Song et al., 2011).

El ATAC-Seg es uno de los métodos más recientes para medir accesibilidad. Se basa en la actividad de la transposasa Tn5. Los transposones de la familia Tn5 son elementos genómicos "móviles" que contienen la secuencia codificante para la transposasa, así como sitios de reconocimiento para que la transposición ocurra. La transposasa Tn5 tiene la capacidad de cortar el DNA y simultáneamente catalizar la adición de secuencias a los fragmentos cortados (N. Li et al., 2020). En el ATAC-Seq, la cromatina es digerida con Tn5 previamente cargada con adaptadores de secuenciación. Durante la incubación, la enzima interúctua con las regiones accesibles del DNA y cataliza un proceso llamado tagmentación, del ingles tag, que significa etiquetar, y fragmentation (fragmentar) (Fig. 2G). Posteriormente se genera y secuencia una biblioteca genómica, la cual contiene los sitios accesibles del genoma (Buenrostro et al., 2013). Actualmente, el ATAC-seq es el método más sensible y preciso para medir accesibilidad de la cromatina, además de ser el que requiere menor número de células, pudiendo incluso escalarse al nivel de célula única (K. Zhang et al., 2021). La transposasa Tn5 se ha utilizado en otras técnicas que permiten el estudio de enhancers a nivel genómico, como el CUT&Tag (Cleavage Under Targets and *Tagmentation*) (**Fig. 2J**) (Skene & Henikoff, 2017).

5.7 Modificaciones post-traduccionales en las histonas y otras proteínas asociadas a regiones enhancer

En la actualidad, la técnica mas usada para identificar modificaciones de histonas en regiones del genoma es la inmunoprecipitación de la cromatina seguida de secuenciación (ChIP-Seq) (**Fig. 2K**) aunque tiene algunas limitantes, como requerir una gran cantidad de células como material inicial (>1,000,000) (S. Jiang & Mortazavi, 2018). En esta técnica, la cromatina se fija, extrae y se fragmenta mediante sonicación (200-500 pb). Posteriomente se incuba con un anticuerpo específico para una modificación postraduccional de histona, luego se hace una precipitación de las regiones unidas al anticuerpo (con perlas magnéticas

o de agarosa), se purifica el DNA resultante y éste puede ser usado para preparar una biblioteca apta para secuenciación (Park, 2009).

Los enhancer activos están enriquecidas con acetilación de histonas, como la H3K27ac (Visel et al., 2009). El incremento de H3K27ac en enhancers es suficiente para aumentar la actividad de estos elementos genómicos (Chen et al., 2019). Se ha propuesto que la acetilación podria servir como andamio para el reclutamiento de TFs (Winter et al., 2017). También se ha descrito que los enhancers están marcados por la unión de la acetilasa P300 y la H3K4me1, a diferencia de los promotores, los cuales están enriquecidos de la versión trimetilada (H3K4me3) (Heintzman, et al., 2007). A pesar de la presencia de la H3K4me1, sin enriquecimiento de H3K27ac, los enhancers suelen tener un estado relativamente inactivo también llamado "*poised*" (Creyghton et al., 2010). Los mapas de ChIP-Seq de modificaciones postraduccionales de histonas y otras proteínas de la cromatina como acetilasas de histonas, facilitan la identificación de enhancers activos.

A pesar de que la mayoría de los datos de proteínas de la cromatina están generados por medio de ChIP-Seq, en la actualidad existe un abanico de técnicas basadas en enzimas capaces de mapear modificaciones de histonas y factores transcripcionales de manera más simple y económica que el ChIP-Seq: el CUT&RUN y el CUT&Tag (Akdogan-Ozdilek et al., 2022; Kaya-Okur et al., 2019; Skene & Henikoff, 2017).

El CUT&RUN (*Cleavage Under Targets and Release Using Nuclease*) (**Fig. 2I**) permite mapear con alta resolución la ubicación de proteínas unidas al genoma mediante el acoplamiento de una nucleasa micrococal (MNAsa) a un anticuerpo específico para alguna proteína del DNA, como TFs o histonas. La enzima corta el DNA cercano a la proteína genómica, lo que permite que dicho fragmento de DNA se libere. El DNA libre se purifica y se secuencia, permitiendo determinar la ubicación a través del genoma (Skene & Henikoff, 2017b).

El CUT&Tag (**Fig. 2J**) es la siguiente técnica desarrollada por Henikoff y sus colaboradores. Similar al CUT&RUN, permite mapear con alta resolución la ubicación de proteínas de la cromatina, pero utiliza a la trasnposasa Tn5, en lugar de la MNasa. Se usan anticuerpos específicos para la proteína de interés y se genera un complejo con la transposasa, que corta y agrega adaptadores de secuenciación al DNA cerca de la proteína, lo que libera el DNA "etiquedado", el cual es puricado y secuenciado permitiendo determinar la ubicación de la proteína en el genoma. El CUT&Tag tiene varias ventajas sobre el CUT&RUN, incluyendo una mayor eficiencia de fragmentación y una menor variabilidad técnica entre experimentos. Además, el CUT&Tag también es más rápido y menos costoso

que el CUT&RUN. Sin embargo, no existen tantos anticuerpos validados para utilizarse en CUT&Tag (Kaya-Okur et al., 2019).

A pesar de que el ChIP-seq sea la técnica mas utilizada y establecida para estudiar la distribución de las proteínas de la cromatina, las técnicas enzimáticas (CUT&RUN y CUT &Tag) son más rápidas y eficientes: requieren menos pasos y menos células, además de requerir una menor profundidad de secuenciación y producir menos ruido de fondo (Tao et al., 2020).

5.8 Trascripción para la detección de enhancers

Los enhancer activos se transcriben generando moléculas llamadas enhancer RNAs (eRNAs), un tipo de RNA no codificante que contiene Cap 5' pero carece de cola poli-A, por lo que son más inestables que otros tipos de RNA como el RNA mensajero (mRNA) (de Lara et al., 2019). Se sabe que juegan un papel importante en la función de los enhancers (Sigova et al., 2015) y se han usado como marcadores para la identificación de enhancers activos (Carullo et al., 2020). Como se mencionó anteriormente, el CAGE ha sido ampliamente utilizado para mapear moléculas a partir del Cap 5', y existen estudios en los que se ha utilizado esta técnica para determinar el repertorio de enhancers activos en diversos tejidos humanos. Los datos apuntan a que el CAGE-Seq es actualmente la técnica más precisa para predecir la actividad de elementos enhancer de forma masiva (Andersson et al., 2014).

5.9 Estructura tridimensional de la cromatina y enhancers transcripcionales

La compartimentalización de la cromatina en el núcleo está altamente regulada (Bonev & Cavalli, 2016). Se sabe que los cromosomas ocupan espacios discretos y bien definidos, los cuales se denominan territorios cromosómicos (Bolzer et al., 2005). A su vez, los cromosomas están subdivididos en compartimientos A y compartimientos B, definidos por tener una alta frecuencia de contactos en su interior y una baja frecuencia de contactos con compartimentos, respectivamente (Lieberman-Aiden et al.. 2009). otros Los compartimentos A se caracterizan por ser regiones laxas (eucromatina), tener tiempos de replicación tempranos, presentar marcas de histonas asociadas a activación y estar transcripcionalmente activos, mientras que los compartimentos B son regiones compactas (heterocromatina), poseen tiempos de replicación tardíos y están silenciados (Rao et al., 2014). Los compartimentos están segregados en otros dominios de menor tamaño

conocidos como dominios topológicamente asociados (TADs), las regiones al interior de un TAD interactúan frecuentemente entre sí y están aisladas de TADs contiguos (Nora et al., 2012; Sexton et al., 2012).



Figura 2. Métodos para determinar la localización y funcionalidad de un enhancer. A) Métodos dependientes de accesibilidad de la cromatina para la localización de enhancers activos. B) Ilustración de la DNAsa interactuando con el DNA desnudo. C) Ilustración de la MNasa interactuando con el DNA desnudo. D) Ilustración general del FAIRE. E) Paso de entrecruzamiento de la cromatina para FAIRE-Seq. F) Representación gráfica de la distribución diferencial entre el DNA fijado y el DNA libre en un microtubo. H) Métodos dependientes de proteínas asociadas a la cromatina para la localización de enhancers activos. I) Ilustración del CUT&RUN, en donde interviene la actividad de la MNasa dirigida por un anticuerpo específico para una proteína de la cromatina. J) Ilustración del CUT&Tag, en donde interviene la actividad de la Tn5 dirigida por un anticuerpo específico para una proteína de la cromatina. K) Ilustración de una proteína de la cromatina siendo reconocida por anticuerpo, el cual está acoplado a una perla magnética para el protocolo de ChIP-Seq K) CAGE-Seq, un método dependiente de la trascripción para la localización de enhancers activos. M) Ilustración de los eRNAs, un tipo de RNAs no codificantes con Cap 5'. N) Proceso de transcripción de los RNAs a partir de un enhancer activo. O) Biotinilación del Cap 5' de un eRNA. P) Captura de los eRNAs biotinilados para su posterior secuenciación masiva.

Se ha demostrado que la estructura tridimensional del genoma es dinámica durante la diferenciación celular y el desarrollo de los organismos (Amiri et al., 2018). Se ha probado que de esta manera se propicia la reestructuración de las interacciones enhancer-promotor tipo celular específicas, lo que repercute directamente en la actividad transcripcional (Kazakevych, Sayols, Messner, Krienke, & Soshnikova, 2017; Siersbæk et al., 2017). La remodelación de las estructuras cromatínicas, como los compartimentos A y B, y los TADs, asociada a procesos como la diferenciación celular, son mediadas por un grupo de proteínas conocidas como proteínas arquitectónicas de la cromatina, como YY1, Cohesina (Weintraub et al., 2017) y la proteína multifuncional CTCF (Kang et al., 2015).

La formación de asas de cromatina y su relevancia para la función de los enhancers ha sido extensivamente estudiada y ha permitido generar modelos que explican la relación entre la estructura tridimensional de la cromatina y la trascripción. Uno de estos modelos visualiza que las interacciones genómicas de largo alcance ocurren por medio de "asas" o *loops*, los cuales tienen en su base proteínas como CTCF y Cohesina (Kang et al., 2015). En este panorama, los elementos contenidos al interior de los *loops*, como promotores y enhancers, pueden interactuar libremente (Carullo & Day, 2019). De esta manera, estudiar la estructura tridimensional de la cromatina puede usarse para inferir los posibles genes blanco de regiones reguladoras distales.

En la actualidad, es posible determinar las interacciones tridimensionales entre dos elementos del genoma mediante la técnica de 3C (del inglés *Chromosome Conformation Capture*) y sus derivados (Simonis et al., 2007). Esta técnica consiste en fijar la cromatina, digerir el DNA con una enzima de restricción y ligar enzimáticamente los extremos de las moléculas que se encuentren en proximidad espacial. Después, se realiza una PCR cuantitativa para determinar la frecuencia de interacción entre ambas regiones (Splinter et al., 2004).

El 4C y el 5C (del inglés *Circular Chromosome Conformation Capture* y *Chromosome Conformation Capture Carbon Copy* respectivamente) son variantes del 3C, las cuales añaden pasos adicionales como PCR inversa en el caso del 4C y PCR multiplex en el 5C. Estas técnicas requieren secuenciación masiva y permiten analizar las interacciones de múltiples regiones del genoma a la vez con una alta resolución (Simonis et al., 2007). Finalmente, el Hi-C (*High-throughput Chromosome Conformation Capture*) es la variante más poderosa del 3C, pues permite determinar todas las interacciones del genoma a la vez mediante la secuenciación de todos los fragmentos ligados por proximidad (Lieberman-Aiden et al., 2009).

Los datos de topología han permitido ampliar nuestro conocimiento sobre las regiones reguladoras en distintos contextos. Por ejemplo, existe una base de datos llamada GeneHancer que ha definido cientos de miles de posibles regiones reguladoras combinando datos de Ensemble, ENCODE, FANTOM y VISTA, y ha inferido los posibles blancos de estos enhancers putativos mediante datos de CHi-C, una variante del Hi-C en la que se usa un anticuerpo especifico contra una proteína de la cromatina para capturar únicamente las interacciones genómicas mediadas por dicho elemento genómico, como CTCF. (Fishilevich et al., 2017).

A pesar de la existencia de métodos masivos para inferir la localización de enhancers, es necesario hacer una validación funcional de los elementos reguladores identificados por estos métodos.

5.10 Enhancers transcripcionales y enfermedades

Las alteraciones de la estructura tridimensional de la cromatina dadas por eliminaciones o mutaciones en sitios de unión de proteínas arquitectónicas de la cromatina como CTCF, pueden propiciar contactos promiscuos entre enhancers y promotores, resultando en enfermedades del desarrollo en humanos (Ibn-Salem et al., 2014).

Además, se ha reportado que las secuencias reguladoras no-codificantes del genoma pueden contener polimorfismos de un solo nucleótido (SNPs) con relevancia en la aparición de enfermedades (T. E. P. Consortium, 2012; Hon et al., 2017; Pierce et al., 2018). De hecho, alrededor del 88% de los SNPs de riesgo están localizados en regiones no codificantes, y para poder priorizar su estudio, es necesaria la anotación de todos los elementos funcionales del genoma participando en los procesos biológicos y fenotípicos de cada tipo celular humano en distintas etapas del desarrollo y la vida adulta (Edwards et al., 2013).

5.11 Enhancers transcripcionales y enfermedades del sistema nervioso central

Se han hecho muchos esfuerzos para determinar regiones reguladoras no codificantes en neuronas y otras poblaciones celulares del cerebro. Sin embargo, el encéfalo es un órgano complejo y heterogéneo, lo que dificulta la obtención de firmas moleculares específicas de un solo tipo celular. Además, existe evidencia que sugiere que las neuronas y los precursores neuronales son los tipos celulares con más secuencias reguladoras específicas (Andersson et al., 2014). Se ha reportado que alrededor del 60% de los sitios de accesibilidad genómica en distintas regiones cerebrales humanas (isocorteza, estriado,

hipocampo y *sustantia nigra*) son tipo celular especificas y la disfunción genética o epigenética de estos elementos reguladores se ha asociado con la predisposición a condiciones neurológicas complejas, como la esquizofrenia (SCZ), la enfermedad de Alzheimer (AD) y la enfermedad de Parkinson (PD) (Corces et al., 2020).

Se ha determinado mediante estudios de asociación de genoma completo (GWAS), que los enhancers activos específicos de subpoblaciones cerebrales (microglia, oligodendrocitos, astrocitos y neuronas) tienen enriquecimiento diferencial de SNPs asociados a enfermedades. Por ejemplo, los enhancers de microglía tienen una marcada abundancia de SNPs asociados con AD (Edwards et al., 2013).

La generación y aplicación de los GWAS a bases de datos de enhancers putativos ha permitido encontrar variantes de riesgo con relevancia medica para una gran cantidad de desórdenes neurales, como el insomnio (Ding et al., 2018), el trastorno de inteligencia (Y. Du et al., 2018), la hipotonia, ataxia, síndrome del retraso del desarrollo y autismo (Padhi et al., 2021), déficit de atención e hiperactividad (Cheng et al., 2019; M. Ma et al., 2018), esquizofrenia (Yang et al., 2019) y los ya mencionados AD y PD (M. Li et al., 2018).

5.12 Enhancers transcripcionales y desarrollo del sistema nervioso central

Además de ser importantes en el mantenimiento del fenotipo y las funciones neuronales, los enhancers son cruciales para los procesos de desarrollo y la diferenciación celular (Furlong & Levine, 2018). Existen ejemplos de enhancers cuya actividad está restringida a un solo tejido u órgano durante una ventana muy específica del desarrollo, como el enhancer ZRS de sonic hedgehog (Lettice et al., 2014), el cual es funcional únicamente en los primordios de extremidad durante el desarrollo.

Estudios recientes han caracterizado el panorama regulatorio en la corteza cerebral humana en distintas etapas del desarrollo usando transcriptómica y accesibilidad de la cromatina a nivel de célula única, teniendo hallazgos relevantes para el campo. Por ejemplo, la glia radial y los precursores neurales tienen los perfiles de accesibilidad de la cromatina más específicos, resaltando la importancia de los procesos de regulación mediada por enhancers en etapas tempranas del desarrollo cerebral (Zhu et al., 2023). Las plataformas de estudio de la corteza humana han permitido validar funcionalmente de manera masiva miles de regiones con potencial regulatorio a través de todos los tipos celulares de esta región del cerebro. Incluso se ha medido el efecto de introducir variantes en la secuencia de estos elementos reguladores (Deng et al., 2024). Esto ha permitido una

compresión sistemática de cómo los SNPs asociados a enfermedades neurológicas impactan la transcripción diferencialmente en distintos tipos celulares cerebrales, como neuronas, astrocitos, oligodendrocitos y microglía.

Otro trabajo generó un atlas de accesibilidad de la cromatina a partir de distintas regiones del cerebro humano en la etapa adulta (Y. E. Li et al., 2023). Los autores determinaron, con base en las regiones del genoma accesibles, que existen al menos 107 tipos celulares distintos a lo largo de las regiones que analizaron (amígdala, cerebelo mígdala, corteza, meséencefalo, núcleos basales, puente de Varolio y el tálamo). 47% de las regiones que reportaron no estaban en la base de datos de ENCONDE, demostrando que los linajes neuronales tienen regiones reguladoras altamente específicas. Sin embargo, anotar elementos reguladores durante el desarrollo de subpoblaciones neuronales específicas es metodológicamente desafiante por la inaccesibilidad y la heterogeneidad de las muestras. Una de las alternativas más prometedoras es la generación de tipos celulares específicos *in vitro* a partir de células troncales (SCs; por sus siglas en inglés).

5.13 Células troncales

Las SCs son células indiferenciadas con capacidad de autorrenovación y habilidad para diferenciarse hacia distintos tipos celulares (Cogle et al., 2003; Thomson et al., 1995). Existen distintos tipos de SCs y se clasifican de acuerdo con la cantidad de tejidos que son capaces de dar lugar. Las SCs Embrionarias (ESCs) son pluripotentes, es decir, tienen la capacidad para diferenciarse hacia derivados de las tres capas germinales embrionarias (endodermo, mesodermo y ectodermo). Las primeras cinco líneas de células troncales embrionarias humanas (hESCs) (H1, H13, H14, H7 y H9) fueron derivadas y descritas por Thomson et al., (1998). Se obtuvieron a partir del cultivo de la masa celular interna de embriones humanos producidos por medio de fertilización *in vitro*.

Las ESCs poseen características moleculares que las definen, como la expresión de factores reguladores muy importantes para el mantenimiento de la pluripotencia, como OCT4, SOX2 y NANOG (Nichols et al., 1998; Niwa et al., 2000; Mitsui et al., 2003). También expresan marcadores moleculares de superficie como SSEA3 y SSEA 4, TRA-1-60, CD9, entre otros. Se caracterizan por tener alta actividad de la fosfatasa alcalina tejido inespecífica (TNAP) y de la telomerasa, que impide el acortamiento de los telómeros dependiente de la división celular (Thomson et al. 1998; De Miguel et al., 2010).

Se han desarrollado protocolos para inducir la diferenciación dirigida de hESCs a varios tipos celulares, incluidas las neuronas dopaminérgicas del mesencéfalo (mNDa) (Kriks et al., 2011), un subtipo celular de gran importancia médica por ser el principal tipo neuronal afectado en PD (Rodriguez-Oroz et al., 2009). Las neuronas generadas *in vitro* representan una fuente accesible para el estudio de la inducción dopaminérgica y de la red de enhancers transcripcionales responsable de su diferenciación.

6.- Antecedentes

En busca de evidencia de remodelación tridimensional en la estructura nuclear durante la diferenciación dopaminérgica (**Fig. 3A**), evaluamos los niveles proteicos de CTCF mediante *Western blot* en tres tiempos de la diferenciación: día 0, estado pluripotente (D0), día 14, precursores neurales (D14) y día 28, neuronas inmaduras (D28). Observamos una disminución en los niveles de CTCF para el D14 con respecto a D0 y D28. Posteriormente, hicimos RNA-Seq en los mismos estadios de la diferenciación (D0, D14 y D28) (**Fig. 3B**). Encontramos cambios dramáticos a nivel transcripcional a lo largo de este proceso, principalmente a nivel de RNAs largos no codificantes (**Fig. 3C**). (**Fig. 3D**). Después, realizamos ATAC-Seq en los mismos estadios de la diferenciación para conocer el estado de apertura de la cromatina (**Fig. 3E**) e identificar *enhancers* putativos (**Fig. 3F**). Realizamos una búsqueda de motivos de unión de TFs en las regiones accesibles de cada etapa (**Fig. 3G**). Encontramos abundancia del motivo de unión de 3 TFs (RFX4, NHLH1 y EBF1) para las regiones accesibles del día 28. Este incremento estuvo acompañado de una inducción en los niveles transcripcionales de estos genes (**Anexo 1**).



Figura 3. Antecedentes del proyecto. A) Representación de la diferenciación dopaminérgica *in vitro* a partir de hESC. B) Ilustración de la técnica de RNA-Seq. C) Representación de un mapa de calor mostrando perfiles transcripcionales. D) Ilustración de IncRNAs. E) Ilustración de la técnica de ATAC-Seq. F) Representación del contacto promotor-enhancer, mediado por un asa de cromatina y CTCF. G) Imagen representativa de una matriz de probabilidades para el motivo de unión de un factor transcripcional.

7.- Planteamiento del problema:

El cerebro humano cuenta con más de 100 tipos celulares, todos generados y mantenidos mediante mecanismos muy finos de control de la expresión génica durante el desarrollo, la vida adulta y en respuesta a estímulos ambientales. Uno de esos tipos celulares son las neuronas dopaminérgicas del mesencéfalo; un subtipo neuronal muy importante por tratarse del principal tipo celular afectado en PD. Sin embargo, no existen datos epigenómicos suficientes para determinar los elementos enhancers específicos de la inducción dopaminérgica en el mesencéfalo ventral humano.

8.- Hipótesis

Las mNDA tienen *enhancers* activos específicos durante su desarrollo y el factor transcripcional RFX4 participa en la actividad de dichos enhancers.

9.- Objetivo general

Identificar y validar *enhancers* específicos de la diferenciación dopaminérgica, así como encontrar evidencia de la participación de RFX4 en la actividad de dichos elementos.

10.- Objetivos particulares:

1.- Generar mapas genómicos para las modificaciones de histonas H3K27ac y H3K4me1 en dos estadios de diferenciación dopaminérgica: células pluripotenciales y neuronas diferenciadas a partir de éstas.

2.- Integrar los datos de modificaciones de histonas con datos de ATAC-Seq y RNA-Seq en mNDAs para generar *enhancers* putativos específicos.

3.- Generar mapas genómicos para la unión del factor transcripcional RFX4 durante la diferenciación dopaminérgica.

11.- Resultados:

11.1 Ensayos de ChIP-Seq en neuronas dopaminérgicas

Diferenciamos ESCs (D0) a neuronas dopaminérgicas (D28) (**Fig. 4A**) siguiendo un protocolo previamente reportado y validado (Kriks et al., 2011; Meléndez-Ramírez et al., 2021), e hicimos ChIP-Seq para las modificaciones postraduccionales de histonas H3K27ac y H3K4me1 (**Fig. 4B**), las cuales están asociadas a regiones activas (eucromatina) y a enhancers, respectivamente. Evaluamos las bibliotecas de ChIP-Seq mediante Bioanalyzer (**Fig. 4C**). Secuenciamos las muestras con una configuración de 75 pb *single-end* (**Fig. 4D**) y tras el filtrado de las lecturas usamos MACS2 para el llamado de picos (**Fig. 4E**).



Figura 4. Descripción grafica del procedimiento para obtener y secuenciar las bibliotecas de ChIP-Seq.
A) Poblaciones celulares a partir de las que extrajimos cromatina: D0 (H9 GFP) para hESC y D28 para mNDA.
B) Ilustración del procedimiento de ChIP. C) Ilustración que describe el procedimiento a seguir para la generación de la biblioteca y una imagen representativa de su análisis por medio de Bioanalyzer. D) Representación de la secuenciación. E) Análisis bioinformático para la generación de regiones enriquecidas.

11.2 Definición de enhancers putativos en células troncales embrionarias y neuronas dopaminérgicas del mescencefalo

Identificamos 298,291 picos de H3K4me1 y 75,165 de H3K27ac en D0; 52,269 regiones tienen ambas marcas (**Fig. 5A**). En D28 encontramos 32,323 picos de H3K4me1 y 30,020 de H3K27ac; 13,102 regiones tienen ambas marcas (**Fig. 5D**). 9,401 regiones tienen ambas marcas en D0 y D28 (datos no mostrados). La distribución de picos esta enriquecida alrededor de los TSS (**Fig. 5B-C y E-F**). Analizamos la distribución de picos en intrones, exones, promotores, sitios de término de la transcripción (TTS) y regiones intergénicas e intrónicas (**Fig. 5G**) La H3K27ac está más enriquecida en promotores (9.2%) que la H3K4me1 (5.3%) (**Fig. 5G**). 30% de las regiones en D0 para la intersección de ambas marcas se encuentran reportadas en GeneHancer, así como el 55% de las regiones con ambas marcas más accesibilidad de la cromatina (**Fig. 5H**). 34% de las regiones enriquecidas con ambas marcas en D28 están reportadas en GeneHancer, así como 49.3% de las regiones que contienen ambas marcas y accesibilidad de la cromatina (**Fig. 5H**). Definimos la intersección de ambas marcas de histona como enhancers putativos.

11.3 Identificación de enhancers putativos en *loci* de genes relevantes para la pluripotencia y para la inducción dopaminérgica

Algunos enhancers putativos específicos de D0 (EE D0) están asociados a genes importantes para las ESCs, como *NANOG*, *ESRG* (**Fig. 6A y B**) y *LINC01108* (datos no mostrados), los cuales están sobre expresados en D0 (Meléndez-Ramírez et al., 2021) (**Fig. 6F**). *ESRG* (*Embryonic Stem Cell Related Gene*) y *LINC01108*, son IncRNAs necesarios para la autorenovación y el mantenimiento de la pluripotencia (Li et al., 2023; Ng et al., 2012), así como la hormona peptídica codificada por el gen *APELA* (datos no mostrados) (Ho et al., 2015). De igual forma, encontramos genes sobre expresados en D28 (**Fig. 6F**) con enhancers putativos específicos del D28 (EE D28) en sus *loci* (**Fig. 6C-E**): *RFX4*, *FOXA1*, *ZIC1* y *ZIC4*; los cuáles están implicados en el desarrollo del CNS (Grinberg & Millen, 2005; S. Li et al., 2006; Pristerà et al., 2015; D. Zhang et al., 2007).



Figura 5. Análisis de distribución de los picos de H3K27ac y H3K4me1. A) Traslape entre las regiones enriquecidas con las modificaciones de histona H3K4me1 y la H3K27ac en el D0. B) Distribución de los picos de H3K4me1 en D0 con respecto a los sitios de inicio de la transcripción (TSS). C) Distribución de los picos de H3K27ac en D0 con respecto a los TSS. D) Traslape entre las regiones enriquecidas con las modificaciones de histona H3K4me1 y la H3K27ac en D0 con respecto a los TSS. D) Traslape entre las regiones enriquecidas con las modificaciones de histona H3K4me1 y la H3K27ac en el D28. E) Distribución de los picos de H3K4me1 en D28 con respecto a los TSS. F) Distribución de los picos de H3K27ac en D28 con respecto a los TSS. G) Anotación genómica de las

regiones enriquecidas en H3K4me1 (izquierda) y H3K27ac (derecha). TTS=Sitio de término de la transcripción **H)** (Izquierda) Número de regiones reportadas en GeneHancer en D0 (izquierda): regiones que comparten ambas marcas (K4me1/K27ac), y regiones con ambas marcas y accesibilidad de la cromatina (K4me1/K27ac/ATAC). (Derecha) Número de regiones en D28 (derecha) reportadas en GeneHancer. Regiones que comparten ambas marcas (H3K4me1/H3K27ac), y regiones con ambas marcas y accesibilidad de la cromatina (H3K4me1/H3K27ac/ATAC).



Figura 6. Loci asociados a genes relevantes con enhancers putativos. A) Locus de *NANOG.* **B)** Locus de *ESRG.* **C)** Locus de *RFX4.* **D)** Locus de *FOXA1.* **E)** Locus de *ZIC4* y *ZIC1.* **F)** Niveles de expresión en FPKMs de genes cuyos *loci* está asociado a un EE D0 (izquierda) o un EE D28 (derecha). Las flechas indican la dirección de la transcripción. Las escalas de cada carril del genoma humano están indicadas al inicio: de 0 a 14 para las histonas y de 0 a 3 para la señal de ATAC-Seq.

11.4 Analisis de enhancers putativos candidatos a ser reguladores de la inducción dopaminérgica

31 de los 50 EE D28 con mayor enriquecimiento de la modificación de histona H3K27ac están en el locus (<10 kb) de un gen anotado (**Fig. 7A-F**); 20 de estos genes (61%) aumentan su expresión al menos moderadamente del D0 al D28 (**Fig. 8A**); Cuatro son TFs: *TAF1* (**Fig. 7A**) –cuyas variantes raras pudieran estar implicadas en la aparición temprana de PD en pacientes masculinos (Zeng et al., 2022), *KAT2B* (**Fig. 7B**), *PRDM16* (**Fig. 8A**) *y MGA1*. Siete están asociados a procesos neurales, como sinapsis y desarrollo del sistema nervioso central (SNC): *TENM4* (Hor et al., 2015) (**Fig. 7C**), *SYNDIG1* (Kalashnikova et al., 2010) (**Fig. 7D**), *MACO1* (Arellano-Carbajal et al., 2011) (**Fig. 7E**), *ITSN1* (Deloulme et al., 2015), *TIAM2* (Matsuo et al., 2002) (**Fig. 7F**), *TANC1, y NEK1*. *TAGLN3* es un gen que codifica para una proteína asociada al citoesqueleto, cuya expresión es específica del SNC (W.-Z. Ren et al., 1994).



Figura 7. Enhancers putativos específicos del D28 con mayor enriquecimiento de modificaciones de histonas. A) Locus de *TAF1* en el que se aprecia la señal de ambas modificaciones de histona (H3K4me1 y H3K27ac) para cada día de diferenciación (D0 y D28), así como si existe un pico de ATAC-Seq o GeneHancer en esa región. B) Locus de *SYNDIG1*, C) locus de *TENM4*, D) locus de *KAT2B*, E) locus de *MACO1* y F) locus de *TIAM2* con la misma disposición de *tracks*. En el *track* de ATAC-Seq se indica en que días de diferenciación está presente el pico (D0, D14 o D28) si el pico. Se indica la escala para todos los *tracks* de histonas: 0-14.

11.5 El locus de TAGLN3 contiene un enhancer especifico de la inducción neural

TAGLN3 es uno de los genes más sobre expresados en D28 con respecto a D0 (Log2 FC= 8.2) (**Fig. 8A y C**) y se encuentra a ~6.5 kb de una región enriquecida de H3K27ac y accesible en D28, pero no en D0 (**Fig. 8B y F**). Datos de GeneHancer confirman que existe una asociación regulatoria entre esta región y el gen *TAGLN3*. Además, encontramos el motivo de unión de cuatro TFs en el enhancer putativo de *TAGLN3*: NR2C1, ATF2, EGR1, NR2F1. EGR1 y NR2F1 tienen un aumento significativo en su expresión de D0 a D28 (**Fig. 8D**). Usamos datos públicos para visualizar los niveles de H3K27ac de esta región en otros tipos celulares. Los progenitores neurales (PN) tienen un nivel de H3K27ac similares a nuestras muestras D28, mientras que esta modificación transcripcional está ausente en neuronas glutamatérgicas (GluN) y neuronas motoras (MN) (**Fig. 8F**).



Figura 8. La expresión de *TAGLN3* correlaciona con las marcas de un enhancer en D28 G) -Log10 del FDR y cambios transcripcionales (D0 y D28) en Log2 FC. Azul: genes neurales. Naranja: TFs. Negro: otro tipo de genes. Arriba de la línea punteada gris igual a qvalue<0.05. H) Locus de *TAGLN3*. I) Expresión en FPKMs de *TAGLN3* durante la diferenciación dopaminergica. J) Expresión en FPKMs de NR2C1, ATF2, EGR1, NR2F1.
K) Enhancer putativo de *TAGLN3* con datos de H3K27ac de otros tipos celulares. PN: progenitores neurales, GluN: neuronas glutamatérgicas, MN: neuronas motoras. Las barras grises en el *track* de GeneHancer indican que el elemento esta reportado. Escala indicada por carril: 0-14 para histonas y 0-3 para la señal de ATAC-Seq.

11.6 Enriquecimiento diferencial de las histonas H3K27ac y H3K4me1 entre celulas troncales pluripotentes y neuronas dopaminérgicas del mescencefalo

Generamos picos consenso (presentes en todas las muestras) y comparamos la cantidad de lecturas alineadas para cada marca y día de diferenciación. Encontramos 635 regiones diferencialmente enriquecidas en las histonas H3K27ac y H3K4me1 (log2FC >1.5 o >-1.5) entre D0 y D28 (**Fig. 9A**). Algunas de las regiones diferencialmente enriquecidas están reportadas en GeneHancer, por lo que conocemos sus genes blanco. Usamos la lista de genes blanco para hacer un análisis de enriquecidas en D28 están asociados a cambios transcripcionales, remodelación del citoesqueleto y neurogénesis (**Fig. 9B**). En D0 encontramos procesos relevantes para ESCs, como regulación de la proliferación mediada por SOX2, OCT4 y NANOG (**Fig. 8C**). Alrededor del 20% de los genes contactados por una región diferencialmente enriquecida están diferencialmente expresados (**Fig. 9D**), como *SRGAP1* (**Fig. 9E y F**), que codifica para un activador de GTPasa relevante en la migración neuronal (Kit et al., 2001).



Figura 9. Regiones con ganancia o pérdida de ambas modificaciones de histonas a través de la diferenciación dopaminérgica. A) Mapa de calor mostrando las 635 regiones que ganan ambas modificaciones de histonas (FC >1.5) o que pierden (FC < -1.5) a través de la diferenciación dopaminérgica (D0 vs D28). **B)** Procesos celulares enriquecidos en la lista de genes asociados a las regiones que ganan H3K27ac y H3K4me1 del D0 a D28. Las barras negras indican el -Log10 del FDR (eje izquierdo) y las líneas y puntos amarillos indican el número de genes para cada proceso (eje derecho). **C)** Procesos celulares enriquecidos en la lista de genes asociados a las regiones que ganan H3K27ac y H3K4me1 del D0 a D28. Las barras negras indican el -Log10 del FDR (eje izquierdo) y las líneas y puntos amarillos indican el número de genes para cada proceso (eje derecho). **C)** Procesos celulares enriquecidos en la lista de genes asociados a las regiones que pierden H3K27ac y H3K4me1 del D0 a D28. Las barras negras indican el -Log10 del FDR (eje izquierdo) y las líneas y puntos amarillos indican el número de genes para cada proceso (eje derecho). **C)** Procesos celulares enriquecidos en la lista de genes asociados a las regiones que pierden H3K27ac y H3K4me1 del D0 a D28. Las barras negras indican el -Log10 del FDR (eje izquierdo) y las líneas y puntos amarillos indican el número de genes para cada proceso (eje derecho). **D)** Porcentaje de genes con cambios transcripcionales (>1.5 o >-1.5 Log2 FC) asociados a una región enriquecida en ambas modificaciones de histonas tanto para D0 como para D28. **E)** Locus de *SRGAP1*, en donde se aprecia una región enriquecida para H3K27ac y H3K4me1 en el D28. Se muestran
también datos de accesibilidad de la cromatina (ATAC) para D0 y 28, así como datos de H3K27ac para otros tipos de células neurales. PN: Progenitores neurales, GluN: neuronas glutamatérgicas, MN: Neuronas motoras. La escala de la señal está indicada al inicio de cada carril. De 0 a 14 para las histonas y de 0 a 3 para el ATAC-Seq. **F)** Niveles de expresión en FPKMs de *SRGAP1* a través de la diferenciación dopaminergica D0, D14 y D28.

11.7 Análisis de regiones con enriquecimiento de marcadores de enhancers en neuronas dopaminérgicas del mescencefalo humanas

Las modificaciones postraduccionales de histonas son un buen marcador para la actividad de elementos reguladores. Sin embargo, se ha reportado que no son suficientes para garantizar la presencia de un enhancer activo. Con el objetivo aumentar el nivel de confianza de nuestros enhancers putativos, añadimos información de accesibilidad de la cromatina al análisis de enriquecimiento. Encontramos 1,155 regiones enriquecidas en la señal promedio de las modificaciones de histonas (H3K27ac y H3K4me1) y ATAC-Seq para D28 (Log2FC>1.5) (Fig. 10A). Estas regiones se agrupan en clústeres de acuerdo con su comportamiento, revelando cuatro clases de enhancers putativos (Fig. 10A). El clúster uno (C1) está compuesto por 487 regiones con una marcada ganancia de ambas modificaciones de histonas de D0 a D28 y accesibilidad de la cromatina variable, en ocasiones con disminución en D28 con respecto a D0 (Fig. 10A). C1 contiene algunas regiones asociadas a genes que aumentan su expresión de D0 a D28, destacando DCLK2 (Fig. 10G), el cual codifica para una proteína asociada a la maduración terminal de axones y dendritas (Edelman et al., 2005); la señal de H3K27ac en esta región es considerablemente más alta en nuestras muestras que en neuronas glutamatérgicas (GluN) y neuronas motoras (MN) (Fig. 10B). Este clúster está compuesto principalmente de regiones no reportadas previamente en GeneHancer (71%) (Fig. 10F) y sus elementos están mayoritariamente anotados como regiones intergénicas e intrones (Fig. 10H). El clúster dos (C2) está compuesto por 315 regiones con ganancia robusta de ambas modificaciones de histonas en D28 con respecto a D0 y ganancia moderada de accesibilidad de la cromatina (Fig. 10A). C2 contiene una región asociada al gen CITED2 (Fig. 10C), el cual aumenta su expresión de D0 a D28 (Fig. 10G) y codifica para un factor transcripcional que regula la expresión de PITX2 (Franco et al., 2017), un regulador importante en el desarrollo de neuronas del mesencéfalo (Martin et al., 2004); la señal de H3K27ac en esta región se puede observar también en neuronas glutamatérgicas (Fig. 10C). La mayoría de las regiones en C2 ya están reportadas en GeneHancer (66%) (Fig. 10F) y al igual que C1, se trata principalmente de regiones intergénicas e intrónicas (Fig. 10H).



Figura 9. Existen cuatro clases de enhancers putativos enriquecidos en H3K27ac, H3K4me1 y accesibilidad de la cromatina en D28. A) Mapa de calor mostrando las regiones con ganancia (Log2FC>1.5) de los tres tipos de señales asociadas a enhancers activos (H3K4me1, H3K27ac, accesibilidad de la cromatina)

en D28. Estas pueden agruparse en 4 clusters: C1, C2, C3 y C4. **B**) Locus de *DCLK*2, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. **C**) Locus de *CITED*2, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. **D**) Locus de *POLR3B*, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. **D**) Locus de *POLR3B*, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. **E**) Locus de *RFX4*, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. **E**) Locus de *RFX4*, en donde se aprecia una región enriquecida para H3K27ac, H3K4me1 y accesibilidad de la cromatina en el D28. En cada locus se muestran datos de H3K27ac para otros tipos de células neurales. PN: Progenitores neurales, GluN: neuronas glutamatérgicas, MN: Neuronas motoras. La escala de la señal está indicada al inicio de cada carril. De 0 a 14 para las histonas y de 0 a 3 para el ATAC-Seq. **F**) Graficas de barras que muestran el número de regiones enriquecidas para las tres señales en D28 previamente reportadas en GeneHancer (negro/abajo) o no reportadas (verde/arriba). **G**) Niveles de expresión en FPKMs de *DCLK2, CITED2* y *RFX4* a través de la diferenciación dopaminergica D0, D14 y D28. **H**) Porcentaje de anotación genómica de las regiones en cada cluster. De izquierda a derecha: C1, C2, C3 y C4. TTS=Transcription starting site. TTS=Transcription termination site.

El clúster tres (C3) está compuesto por 152 regiones con robusta ganancia de accesibilidad de la cromatina en D28 que no presentan cambios en la H3K4me1 (siendo consistentemente baja en D0 y D28). También aumentan la señal de H3K27ac con la diferenciación (Fig. 8A). C3 contiene una región anotada en GeneHancer ubicada en el locus del gen POLR3B (Fig. 8D), la cual esta anotada como un enhancer del gen RFX4, el cual aumenta su expresión durante la diferenciación dopaminérgica y se ha reportado como un importante regulador del desarrollo neuronal (Choi et al., 2024) (Fig. 8G). La expresión de POLR3B no cambia a través de la diferenciación (datos no mostrados). La señal de H3K27ac en esta región parece ser especifica de nuestras muestras al compararse con otros tipos neuronales (Fig. 8D). Todas las regiones de este clúster están reportadas en GeneHancer (Fig. 8F) y la mayoría de sus elementos están anotados como promotores (62%) (Fig. 8H). El clúster cuatro (C4) está compuesto por 194 regiones con robusta ganancia de accesibilidad de la cromatina y H3K27ac en D28, acompañadas de un aumento moderado en H3K4me1 (Fig. 8A). C4 contiene una región en el locus gen RFX4 (Fig. 8E), la cual esta anotada en GeneHancer como un enhancer del mismo (Fig. 8G). La señal de H3K27ac en esta región parece ser especifica de linajes neurales ventrales, como mNDa y MN. (Fig. 8E). Casi todas las regiones de este clúster están reportadas en GeneHancer (94.3%) (Fig. 8F) la mayoría caen en regiones intrónicas e intergénicas (Fig. 8G).

11.8 Análisis de la distribución genómica de RFX4 a través de la diferenciación dopaminérgica

La expresión de RFX4 aumenta durante la diferenciación dopaminérgica (Fig. 9C). Su motivo de unión está sobre representado de manera significativa en los picos de accesibilidad de la cromatina de D28, y existen al menos dos enhancers neurales controlando su expresión en D28. Uno de ellos (E2-RFX4) está altamente conservado en vertebrados, desde humano hasta pez cebra (Fig. 9A). Hipotetizamos que este TF puede tener un rol indispensable en la inducción y el desarrollo dopaminérgicos. Con la finalidad de establecer la posible función de RFX4, realizamos un ensayo de CUT&RUN para mapear los sitios de unión de este TF en tres puntos de nuestro protocolo de diferenciación dopaminérgica: D0, D14 y D28. Hicimos el llamado de picos y descartamos todos aquellos que se encontraban en solo una réplica por condición. Tras el filtrado obtuvimos 2 picos en D0, 241 en D14 y 933 en D28 (Fig. 9B), patrón que parece estar en concordancia con los niveles de mRNA de *RFX4* (Fig. 9C). Ninguno de los picos en D0 coincide con una región accesible de la cromatina en D0, ni los picos de D14 con regiones accesibles en D14; ninguno de los picos para D0 y D14 fueron reportados como sitios de unión de RFX4 mediante ChIP-Seg realizado en una línea celular que sobre-expresa este TF (Choi et al., 2024) (Fig. 9B). El 47% de los picos de RFX4 en D28 coinciden con la base de datos de GeneHancer (datos no mostrados). 428 picos de RFX4 en D28 coinciden con una región accesible en D28 (ATAC), de los cuales, 291 fueron reportados como sitios de unión de RFX4 mediante ChIP-Seg realizado en una línea celular que sobre-expresa este TF (ATAC + sitio reportado de RFX4, Choi et al. 2024) (Fig. 9B). Con el fin de conocer los posibles blancos de regulación de RFX4 en neuronas dopaminérgicas, analizamos los 291 picos de RFX4 presentes en cromatina accesible de D28 y previamente reportados y vimos los cambios en la expresión del gen más cercano a éstos, siempre y cuando estuviera a ¿menos de? 10 kb de distancia. De los 148 genes que encontramos con estas características, solo DBP disminuyó su expresión de D0 a D28 (Log2FC = -2.3 y qValue < 0.05) (Fig. 9D) 47 tuvieron un aumento significativo en su expresión (Log2FC > 1.5 y qValue < 0.05) y el resto no tuvieron cambios, aunque la mayoría de ellos tienen una tendencia hacia la sobreexpresión (66 de 100) (Fig. 9D). Los cinco genes con mayores cambios fueron FAM121A, WNT1, CCDC184, CRB2 y NLRP1. En la lista de genes que aumentan su expresión significativamente aparece RFX2, otro miembro de la familia RFX. Estos resultados indican que RFX4 podría participando en la regulación transcripcional positiva de varios genes durante la diferenciación de neuronas dopaminérgicas del mesencéfalo.



Figura 9. Análisis de unión de RFX4 a lo largo de la diferenciación dopaminergica. A) Locus de *RFX4* con información de ambas modificaciones de histonas para D0 y D28, así como la señal de accesibilidad de la

cromatina (ATAC) para ambos días. H3K27ac para otros tipos celulares neurales; PN: precursores neurales, GluN: neuronas glutamatérgicas, MN: neuronas motoras. Se muestran los niveles de conservación a través de los vertebrados con una escala del -5 al 100. La escala de cada carril se encuentra indicada del lado derecho. De 0 a 14 para las marcas de histonas, a excepción de la H3K27ac en MN (0 a 25); de 0 a 3 para la accesibilidad de la cromatina. B) Numero de picos reproducibles (al menos en dos replicas biológicas) de RFX4 para cada día de diferenciación (D0, D14 y D28). Negro: picos que no caen en un sitio de accesibilidad de la cromatina para ese día de diferenciación. Azul claro: picos que si caen en un sitio de accesibilidad de la cromatina para ese día de diferenciación. Azul marino: picos que caen en un sitio de accesibilidad de la cromatina para ese día de diferenciación y adicionalmente son sitios reportados de unión de RFX4. C) Niveles de expresión en FPKMs de RFX4 a lo largo de la diferenciación dopaminergica. D) Volcano plot en donde se observan los patrones de expresión de los genes localizados a <10.000 pb de distancia de un pico de RFX4 en D28. El eje de las x indica el Log2 FC entre D0 y D28. El eje de las y indica el -LOG10 del q Value. En color negro se indican los transcritos que no fueron diferencialmente significativos. En color azul se indican los transcritos (solo uno) que tuvieron una disminución significativa en su expresión de D0 a D28 (qvalue < 0.05) y un Log2FC < 1.5. En color rojo se indican los transcritos (47) que tuvieron un aumento significativo en su expresión (qvalue < 0.05) y un Log2FC > 1.5.

12.-Discusión:

En este trabajo analizamos los cambios de histonas asociados a la inducción dopaminérgica. Identificamos 298,291 picos de H3K4me1 y 75,165 de H3K27ac en hESC D0. Existen 55,269 regiones con ambas modificaciones, las cuales catalogamos como enhancers putativos; en contraste, solo encontramos 32,323 picos de H3K4me1 y 30,020 de H3K27ac, de estos 13,102 picos fueron considerados enhancers putativos, por compartir ambas marcas de histona. Consistentemente con la gran cantidad de picos de H3K4me1 en D0, se ha reportado que las ESC tienen una gran cantidad de enhancers en estado *"poised"*, caracterizados por la presencia de la H3K4me1 y la ausencia de la H3K27ac (Creyghton et al., 2010).

La gran cantidad de enhancers putativos que identificamos en D0 podría explicarse porque se ha reportado que la cromatina de las ESC presenta niveles basales de transcripción de todo el genoma, los cuales están asociados con una heterocromatina menos compacta (Efroni et al., 2008) y mayor accesibilidad de la cromatina global (Meléndez-Ramírez et al., 2021), quizá también acompañada con altos niveles de H3K27ac. Algunos estudios sugieren que las ESC presentan estados de la cromatina "multivalente", es decir, regiones con modificaciones de histonas aparentemente contradictorias, como la cromatina bivalente, caracterizada por la presencia de H3K4me3 (asociada a promotores activos) y H3K27me3 (asociada con promotores reprimidos) (Bernstein et al., 2006), o la cromatina trivalente, caracterizada por contener H3K4me1 (asociada a enhancers), H3K9me3 (asociada con heterocromatina constitutiva) y H3K27ac (asociada con promotores o enhancers activos) (Wapinski et al., 2013).

Ambas marcas de histonas, en ambos tiempos, están principalmente distribuidas en regiones intergénicas (33.4%) e intrónicas (51.7%), lo que está en concordancia con la distribución conocida de elementos enhancer (T. E. P. Consortium, 2012; Fishilevich et al., 2017). La mayoría de los enhancers putativos identificados a través de marcas de histonas ya están catalogados en GeneHancer como elementos reguladores distales, con un 81% en D0 y un 86% en D28. Esto es comprensible dado que GeneHancer recopila elementos reguladores de una amplia variedad de fenotipos y se ha documentado que existe un cierto nivel de traslape en la fracción de enhancers activos entre distintas estirpes celulares. (Andersson et al., 2014).

La localización de las regiones enriquecidas con ambas modificaciones de histona (H3K27ac y H3K4me1) es variable entre ESC y mNDa diferenciadas *in* vitro. Una proporción

de genes (>20%) de los genes proximales a estas regiones diferenciales presentan cambios en su expresión a través de la inducción dopaminérgica. Los genes diferencialmente expresados a la alta o a la baja representan procesos asociados a cada estado de la diferenciación, comprendiendo regulación de la pluripotencia en D0 y neurogénesis en D28. Este fenómeno se ha reportado en muchos otros estudios en los que se ha evaluado la dinámica en la actividad de los enhancers asociados a procesos del desarrollo en distintos tejidos (H. Ma et al., 2022; Zhou et al., 2023) incluido el CNS (Giacoman-Lozano et al., 2022). Nuestro análisis nos permitió encontrar un enhancer putativo en loci de *FOXA1*, un gen indispensable para el desarrollo de las neuronas dopaminérgicas (Pristerà et al., 2015), cuya función también es importante en el mantenimiento de las neuronas dopaminérgicas maduras, pues participa en la regulación transcripcional de proteínas involucradas en la producción y liberación de dopamina.

Mediante el análisis conjunto de los tres tipos de información genómica (H3K27ac/H3K4me1/ATAC) definimos cuatro clústers cuya señal se incrementa del D0 al D28, los cuales podrían representar cuatro clases distintas de enhancers putativos dirigiendo la inducción dopaminérgica. El grupo uno, es el más extenso (487) y a pesar de esto, es el que menos regiones reportadas en GeneHancer contiene (143). Creemos que estas regiones han sido difíciles de anotar mediante otras bases de datos porque la accesibilidad de estas regiones es en general relativamente baja e inconsistente; otros grupos han recalcado que a pesar de que la accesibilidad de la cromatina correlaciona positivamente con la actividad de los enhancers (Bozek & Gompel, 2020), algunos elementos reguladores validados cuya actividad se ha validado experimentalmente, pueden estar relativamente inaccesibles (Arnold et al., 2013). Nuestro análisis multiómico parece tener la resolución necesaria para reportar regiones reguladoras a pesar de su estado de compactación.

Las regiones en el clúster 2 tienen una robusta ganancia en las marcas de histonas y una accesibilidad de la cromatina moderada, por lo que creemos que son elementos reguladores que encajan bien con la definición de enhancer canónica. Concordantemente con esto, la mayoría de estas regiones ya han sido reportadas en GeneHancer (66%). El clúster 3 resalta por tener muy bajos niveles de la H3K4me1 en ambos tiempos, una subida modera de H3K27ac y altos niveles de accesibilidad de la cromatina, lo que coincide con el patrón canónico de promotores activos (Birney et al., 2007; Hoffman et al., 2013). Consistentemente, más del 70% de estas regiones están anotadas como promotores. Un análisis exhaustivo de los genes asociados a las regiones del clúster 3 nos podrían revelar más acerca del papel de la H3K27ac en el aumento de la actividad de los promotores.

El clúster 4 esta caracterizado por una ganancia en los tres tipos de información epigenómica de D0 a D28 y regiones principalmente intergénicas e intrónicas, por lo que seguramente se trata de regiones enhancer canónicas (Heinz et al., 2010; Visel et al., 2009). Consistentemente, la mayoría de estas regiones (177 de 190) ya han sido reportadas en GeneHancer. De las 13 regiones no reportadas, 8 están en la proximidad de un gen (<10,000 bp). De éstos, 7 aumentan su expresión de D0 a D28 y 2 están asociados a funciones neuronales: *ARX* (Kitamura et al., 2002) y *EPB41L1*, que codifica para una proteína necesaria para la estabilización de los receptores de dopamina D1 y D2 en la membrana neuronal (Binda et al., 2002). La actividad de estos 13 enhancers putativos podría ser especifica de la inducción dopaminérgica.

RFX4 parece tener un papel muy importante en la inducción dopaminérgica. No solo está altamente expresado en D28, sino que su abundante presencia en los sitios de cromatina accesible de este mismo estadio de diferenciación apunta a una alta actividad de regulación transcripcional (Meléndez-Ramírez et al., 2021). A pesar de que el motivo de unión de RFX4 tenga una alta abundancia en nuestros picos de D28, sólo encontramos 933 picos que estuvieran presentes en al menos dos réplicas de CUT&RUN. Esto puede deberse a que existen otros miembros de la familia RFX (RFX1, RFX2, RFX3, RFX5, RFX7 y RFX8) con motivos de unión a DNA muy similares (Emery et al., 1996). De estos, únicamente RFX2 y RFX3 aumentan su expresión del D0 a D28 (**cita o data**).

Los picos de RFX4 que reportamos aquí no se encuentran en regiones previamente reportadas por GeneHancer, algo que no esperábamos, puesto que se ha reportado que RFX1, otro miembro de la familia RFX, se une específicamente a enhancers activos en precursores neurales de mosca (Creyghton et al., 2010). En ese mismo trabajo se planteó la posibilidad de que los miembros de la familia RFX estén involucrados en el modelado de la actividad de enhancers durante la neurogénesis. Solo el 47% de nuestros picos de RFX4 caen en una región previamente reportada en GeneHaner, por lo que nuestra hipótesis es que las regiones restantes (53%) son enhancers descritos por primera vez, y algunos de éstos podrían ser específicos de la inducción dopaminérgica en el mesencéfalo ventral humano.

13.-Perspectivas:

Es necesario realizar la validación funcional de los enhancers putativos identificados durante este trabajo. Para esto se planea usar el plásmido reportero *pGL3 EGFP*. El cual contiene a la proteína verde fluorescente EGFP bajo el control de un promotor mínimo. La expresión de la proteína únicamente se lleva a cabo si se le clona un elemento regulador inmediatamente rio arriba del promotor mínimo. Tras clonar la secuencia de los enhancers putativos más interesantes dentro de este plásmido, evaluaremos la actividad de enhancer en distintos contextos celulares, como ESC y mNDA.

Además, se plantea realizar un knockout (KO) del factor de transcripción RFX4 mediante CRISPR-Cas9 en ESC. Posteriormente, estas ESC serán diferenciadas hacia mNDA para evaluar el papel de RFX4 en la inducción dopaminérgica. Los resultados se evaluarán mediante inmunofluorescencias para marcadores de la diferenciación (TH, FOXA2, FOXA1, LMX1A y LMX1B), y la expresión de los blancos putativos de RFX4 (*RFX2, WNT1, NLRP1, CRB2, CCDC184 y FAM181A*). Estos experimentos permitirán confirmar la contribución de RFX4 y de los enhancers en la inducción dopaminérgica, proporcionando una comprensión más profunda de los mecanismos moleculares implicados en la especificación de las mNDA

14.- Conclusión:

Este estudio proporciona una visión integral de la dinámica de la cromatina durante la diferenciación de ESC hacia mNDA. A través de la integración de datos de RNA-Seq, ATAC-Seq y ChIP-Seq, identificamos enhancers putativos que podrían estar orquestando la inducción dopaminérgica. Existe una asociación importante entre la presencia de estos enhancers putativos en un locus especifico y la trascripción de los genes adyacentes. Además, el factor de transcripción RFX4 fue implicado en la regulación de algunos genes que se sobre expresan durante este proceso de diferenciación, indicando la actividad de enhancers no reportados e implicados en la inducción dopaminérgica y el desarrollo del sistema nervioso central. Estos hallazgos ampliaron nuestro entendimiento de los mecanismos moleculares que gobiernan la diferenciación de las mNDA.

15.- Materiales y métodos:

15.1 Cultivo celular y diferenciación dopaminergica

La diferenciación dopaminérgica se realizó a partir la línea de ESCs H9, modificada para expresar de manera constitutiva a la proteína verde fluorescente (H9GFP) (Ramos-Mejía et al., 2014). Las ESCs fueron sembradas en matraces T75 tratados con Matrigel (BD) y cultivadas con medio Knock Out (Gibco ®) suplementado con 20% de remplazo de suero, 2 mM de GlutaMAX, 1X de aminoácidos no esenciales, 10 mM de β-mercaptoetanol y FGF-2 (10ng/mL Sigma), previamente condicionado por fibroblastos inactivados mitóticamente.

La diferenciación dopaminérgica se llevó a cabo mediante la inducción de la placa de piso por medio una inhibición dual de la vía de las SMADs (Kriks et al., 2011). Brevemente, el cultivo se mantuvo con medio Knock Out suplementado (Gibco ®) por 11 días, gradualmente cambiando el medio KO por medio N2, iniciando en los días 5 y 6 con 25% de N2, aumentando a 50% en los días 7 y 8, hasta el 75% en los días 9 y 10. El medio también fue adicionado con LDN193189 (100 nM, Stemagent; días 0–11), SB431542 (10 mM, Tocris; días 0–5), el SAG (1 mM, Sigma; días 1–9), purmorphamina (2 mM, Stemgent; días 1–9), FGF-8 (100 ng/mL, Prepotech; días 1–7) y CHIR99021 (3 mM, Stemgent; días 3–13). Al día 11 las células fueron cultivadas con medio Neurobasal adicionado con B27 (Invitrogen), suplementado con BDNF (20 ng/mL, Prepotech) ácido ascórbico (0.2 mM, Sigma), GDNF (20 ng/mL, Prepotech), TGF-β3 (1 ng/mL, Prepotech), dibutiril cAMP (0.5 mM; Sigma) y DAPT (10 mM; Sigma) para promover la maduración neural. Al día 22, las células fueron disociadas usando TrypLE Express (Life Technologies) y resembradas en placas tratadas con poly-l-ornithina, fibronectina y laminina. Se continuó la maduración neural hasta el día 28 para aislar cromatina.

15.2 Extracción de cromatina

Tanto al día 0 de diferenciación (D0) como a al día 28 (D28), se realizó un protocolo de extracción de cromatina para realizar los experimentos de ChIP-Seq. Primero, se añadieron 200 µL de buffer de fijación o *cross-linking buffer* (11% de formaldehido, EGTA 0.5 mM (Sigma-Aldrich), NaCl 100 mM (Sigma-Aldrich) y HEPES 50 mM (**MARCA**)) y 1.8 mL de medio de cultivo a cada pozo de una placa de 12 pozos con las células a un 90% de confluencia (aproximadamente 2 millones de células).

Las muestras se fijaron por 10 minutos a temperatura ambiente en un agitador a 40 rpm. Posteriormente se adicionaron 200 μ L de glicina (Sigma-Aldrich) 1.38 M, para detener la reacción de fijación; se incubó por 5 minutos a temperatura ambiente. Después las muestras fueron lavadas tres veces con PBS frio (4°C) por cinco minutos en agitación (40 rpm). Finalmente, se añadieron 200 μ L de buffer de lisis (1% SDS (**MARCA**), 10 mM EDTA (Sigma-Aldrich), and 50 mM Tris-HCI (**MARCA**)) con inhibidores de proteasas a cada pozo. Se incubaron a 4°C por cinco minutos y se usó un *scrapper* para recuperar la cromatina fijada en un tubo fresco y se almaceno a 4°C.

Sonicación de la cromatina

La cromatina se fragmentó usando el sonicador Covaris S200. Brevemente, la cromatina fijada se colocó en un microtubo de 130 μ L y fue sonicada por 80 segundos a 140 de *peak power*, 10.0 de *duty factor* y 200 ciclos, manteniendo la temperatura a 4°C. Para evaluar la eficiencia de la sonicación, se trataron 2-5 μ L de las muestras con proteinasa K y RNAse A (Invitrogen). El DNA se extrajo usando el método de Fenol-Cloroformo. El DNA purificado se corrió mediante electroforesis en gel de agarosa (Sigma-Aldrich).

15.3 Inmunoprecipitación de la cromatina

La inmunoprecipitación de la cromatina fue realizada usando el *One-day ChIP kit* de Diagenode (C01010080) según las instrucciones del fabricante. Se usaron 1-2 millones de células por muestra por ChIP y 200,000 células para el control *input*. Se usó un anticuerpo específico para la H3K27ac (10 μ L, Active Motif numero de catalogo 53160) y un anticuerpo específico para la H3K4me1 (3 μ L, Diagenode, numero de catalogo C15410194). Previo a la preparación de la biblioteca para secuenciación, el DNA inmunoprecipitado fue purificado y concentrado usando el *ChIP Elute kit* (Zymo Cat. No. 634887). El DNA resultante fue

cuantificado usado un *Qubit 2.0 Fluorometer* (Life Technologies) con el *Qubit ssDNA Assay Kit* (Life Technologies, Cat. No. Q10212). Preparación y evaluación de la biblioteca Se usó el *DNA SMART™ ChIP-Seq Kit* de Takara Bio USA, Inc. para la preparación de las bibliotecas, siguiendo las especificaciones del fabricante. Se usaron 2 ng de DNA por reacción y utilizando 18 ciclos en la reacción de PCR para la generación de la biblioteca. Se utilizaron AMPure beads XP (Beckman Coulter) para retirar los oligonucleótidos y seleccionar el tamaño de fragmentos, los cuales deben ser idealmente de 250 a 500 pares de bases. Las bibliotecas fueron cuantificadas por medio de *Qubit 2.0 Fluorometer* con el *Qubit ssDNA Assay Kit* y posteriormente evaluadas con el Bioanalyzer.1 de Agilent.

15.4 Secuenciación de Nueva Generación

Después de la preparación de las bibliotecas, las muestras fueron transportadas al laboratorio Tec BASE en Monterrey, México, donde se utilizó una máquina de secuenciación Nova-Seq para secuenciación de fragmentos de 150 bp en un formato de lectura en pares. Las bibliotecas se agruparon y se secuenciaron en una celda de flujo, diluyéndose para obtener aproximadamente 50 millones de lecturas por muestra. El pipeline de ChIP-Seq utilizado se resume en la figura a continuación.

15.5 Análisis bioinformático del ChIP-Seq

Para el análisis de los datos de ChIP-seq, utilizamos el software MACS2 (Model-based Analysis of ChIP-Seq) para identificar los picos de enriquecimiento en las secuencias de DNA asociadas a proteínas de interés. Primero, las lecturas secuenciadas se alinearon al genoma humano de referencia (hg38) utilizando el alineador de alta precisión: Bowtie2 v2.4.4. Posteriormente, los archivos BAM resultantes se filtraron eliminando los duplicados de PCR identificador con Picard tools v2.27.5. Adicionalmente, se eliminaron los fragmentos con una calidad de alineamiento MAPQ < 30, fragmentos no pareados, fragmentos que se alinearon más de una vez en el genoma, y fragmentos no alineados utilizando las herramientas de Samtools v1.13. Después del filtrado, los archivos BAM se ingresaron a MACS2 v2.2.7.1. Posteriormente, los archivos BAM resultantes se ingresaron a MACS2 para realizar la llamada de picos. Configuramos el programa para usar un control de entrada, en este caso el *input*, para mejorar la especificidad de los picos identificados.

MACS2 fue ejecutado con un valor de q ajustado (q-value) de 0.01 para determinar la significancia estadística de los picos, minimizando así los falsos positivos. Adicionalmente se utilizaron los parámetros --broad --broad-cutoff 0.01 -B --nomodel –SPMR y –extsize con el tamaño de fragmento (d) calculado por MACS2 en una corrida previa en la cual se generó el modelo. Los picos identificados se filtraron asegurando que no hubiera traslape con las regiones de enriquecimiento denominadas "Black List Regions" y reportadas por ENCODE. Los picos identificados se anotaron en relación con las regiones genómicas utilizando la herramienta HOMER v4.11.1.

Para el análisis de enriquecimiento diferencial de ChIP-Seq entre las condiciones D0 y D28 de las marcas H3K27ac y H3K4me1, combinamos los picos de todas las condiciones para generar una lista de regiones consenso utilizando la función merge de bedtools v.2.30.0. Luego, obtenemos los conteos de cada muestra para cada región consenso definida en el archivo BED utilizando la función multicov de bedtools y generando así una matríz de conteos en sitios consenso, los cuales tomamos como "pseudogenes". Esta matriz fue importada a DESeq2 v1.38.3, un paquete de R, para llevar a cabo el análisis de enriquecimiento diferencial. Dentro de DESeq2, normalizamos los datos para evitar sesgos relacionados a la profundidad de secuenciación. Los picos con un valor de p ajustado (q-value) menor a 0.05 y con un Log2FC > 1.5 o < -1.5 fueron considerados significativamente diferenciales. Finalmente, los picos presentes en GeneHancer se anotaron como enhancers putativos del gen con el score de asociación más alto, y el gen más cercano en caso de no estar reportados en GeneHancer.

Posteriormente realizamos el mismo procedimiento de enriquecimiento diferencial de señal añadiendo datos de las lecturas de ATAC-Seq de las mismas condiciones: D0 y D28.

15.6 Ensayo de CUT&RUN

Para los experimentos de Cut&Run se utilizó el kit CUTANA™ ChIC/CUT&RUN Kit (EpiCypher) siguiendo las recomendaciones del fabricante. Brevemente, el primer día se prepararon las soluciones buffer para lavado, permeabilización e incubación de anticuerpos. Se lavaron las perlas de concanavalina A (ConA) y se resuspendieron en 10 µL por reacción en buffer de activación. Una vez activadas las perlas, se recolectaron 500,000 células por reacción, se lavaron con 100 µL/reacción de Buffer de lavado a temperatura ambiente y se resuspendieron. Las células se mezclaron con las perlas ConA activadas y se incubaron a temperatura ambiente para permitir la adsorción de las perlas por las células. Posteriormente, las células unidas a las perlas se resuspendieron en 50 µL/reacción de Buffer de anticuerpos frío y se añadió el anticuerpo de RFX4 (Cat #PA5-62205 – Invitrogen) 1:200. En el caso de los controles negativo y positivo se agregó 1 µl de IgG o H3K4me3 Control Antibody, respectivamente. Las muestras se incubaron a 4°C durante la noche en rotación.

El segundo día, se lavaron las perlas con 200 µL/reacción de Buffer de permeabilización celular frío y se resuspendieron. Se añadieron 2.5 µL/reacción de pAG-MNase y las muestras se incubaron a temperatura ambiente por 10 minutos. Después de la incubación, se lavaron nuevamente con 200 µL/reacción de Buffer de permeabilización celular frío. Las muestras se colocaron en hielo y se añadió 1 µL/reacción de cloruro de calcio para la activación de la enzima; se incubaron en agitación a 4°C durante 2 horas. Posteriormente, se añadió la solución para detener la reacción (33 µL/reacción) y se incubó a 37°C por 10 minutos en un termociclador. Esta última solución contiene E. coli Spike-in DNA, que sirve como control interno para el análisis bioinformático. El sobrenadante se recolectó para la purificación de ADN.

Para la purificación, se añadió 420 μ L de Buffer de unión de ADN a cada reacción y se cargaron en columnas de ADN. Las columnas se utilizaron para los lavados con 200 μ L de Buffer de lavado de ADN y se eluyó en 12 μ L de Buffer de elución de ADN. La cantidad de ADN se cuantificó utilizando el equipo Qubit.

50

15.7 Preparación de las bibliotecas de CUT&RUN

Para la primera parte de la construcción de bibliotecas genómicas, que corresponde a la reparación de extremos, se descongelaron los reactivos del kit almacenados a -20°C y se mantuvieron en hielo durante el experimento. Se mezcló el Buffer de preparación de extremos. Se transfirieron 5 ng de ADN enriquecido de CUT&RUN a tubos de 8 tiras nuevos y se ajustó el volumen a 25 µL con Buffer 0.1X TE.

Subsecuentemente, por cada reacción de preparación de biblioteca, se combinaron 4.2 μ L de Buffer de preparación de extremos y 1.8 μ L de enzima de preparación de extremos. Se vortexeó suavemente para mezclar, se centrifugó rápidamente para recoger el líquido y se regresó al hielo. De esta manera, se agregaron 5 μ L de la solución de preparación de extremos a los 25 μ L de ADN de CUT&RUN en tubos de 8 tiras. Se pipeteó 5 veces, se vortexeó suavemente para mezclar y se centrifugó rápidamente.

Se colocaron las reacciones en un termociclador con la tapa calentada a \geq 75°C con los siguientes parámetros: 20 min a 20 °C, 30 min a 65 °C durante un ciclo. Al terminar se centrifugaron rápidamente los tubos y se colocaron en hielo. Se preparó la mezcla de ligación en un nuevo tubo de 1.5 mL sobre hielo. Por cada reacción, se combinaron 16.5 µL de mezcla de ligación y 0.55 µL de potenciador (enhancer) de ligación. Se vortexó suavemente, se centrifugó rápidamente y se regresó al hielo.

Posteriormente, se agregaron 1.25 μ L de Adaptador para Illumina® y 15.5 μ L de la mezcla de ligación a cada reacción, manteniendo los tubos en hielo. Se vortexearon los tubos a fondo para mezclar, se centrifugaron rápidamente y se regresaron al hielo. Seguidamente se colocaron los tubos en un termociclador sin tapa calentada, con el bloque a 20°C y se incubaron durante 15 minutos.

Al terminar se centrifugaron rápidamente los tubos y se colocaron en un soporte a temperatura ambiente. Se agregaron 1 μ L/reacción de enzima de escisión de U. Se pipeteó 3 veces para limpiar las puntas, se vortexeó suavemente para mezclar y se centrifugó rápidamente. Después, se colocaron los tubos en un termociclador con la tapa calentada a \geq 47°C y el bloque a 37°C y se incubaron durante 15 minutos.

Al finalizar, se centrifugaron rápidamente los tubos y se agregó 1 mL de Etanol (EtOH) al 85% por reacción. Se vortexeó el reactivo SPRIselect (Beckman Coulter, Inc.) para resuspender bien las perlas. Se añadieron lentamente 47.75 µL/reacción, se mezcló bien vortexeando para resuspender y se centrifugó rápidamente para recoger el líquido. Se incubó durante 5 minutos a RT.

Después, se colocaron los tubos en un imán por 2-5 minutos. Se pipeteó para eliminar el sobrenadante, manteniendo los tubos en el imán. Se añadieron 180 µL/reacción de etanol al 85%, se eliminó el sobrenadante y se repitió una vez más.

Posteriormente, se centrifugó rápidamente, se regresaron al imán y se eliminó el Etanol residual. Después, se retiraron los tubos del imán se secaron al aire con las tapas abiertas, durante 2-3 minutos a RT; se añadieron 12 μ L/reacción de Buffer 0.1X TE para eluir el ADN y se vortexeó para resuspender e incubar durante 2 minutos a RT. Luego, se colocaron los tubos en el imán durante 2 minutos a RT, y se transfirieron 10.5 μ L de ADN eluido a nuevos tubos de 8 tiras.

Después, se realizó la reacción de PCR de Indexación. Para ello, se asignó un par único de primers i5 e i7 a cada reacción, como identificador. Se mezclaron los tubos de stock que contenían la mezcla de PCR 2X de inicio caliente y los primers i5 e i7, y se centrifugaron rápidamente. A cada reacción de preparación de biblioteca, se agregaron los siguientes reactivos individualmente y en orden: 1 μ L de primer i7, 1 μ L de primer i5 y 12.5 μ L de Mezcla Maestra de PCR 2X de inicio caliente.

Se mezclaron bien vortexeando los tubos, evitando burbujas, y se centrifugaron rápidamente. Se colocaron los tubos en un termociclador con la tapa calentada a 105°C. Se realizó la PCR utilizando los siguientes parámetros: 1 ciclo a 98 °C por 45 seg para la activación de la enzima, 14 ciclos a 98 °C por 15 seg y a 60 °C por 10 seg para el alineamiento/extensión y 1 ciclo a 72 °C por 60 seg para la extensión final.

Después de la reacción de PCR, se vortexeó el reactivo SPRIselect para resuspender bien las perlas. Se añadieron lentamente 25 µL de reactivo SPRIselect a cada reacción de PCR de indexación. Se procedió con la limpieza de la biblioteca siguiendo los pasos previamente descritos para limpieza. El protocolo generó 10.5 µL de bibliotecas de secuenciación de CUT&RUN purificadas. Adicionalmente, se cuantificaron las bibliotecas usando el fluorómetro Qubit. Se examinó la distribución del tamaño de los fragmentos en Bioanalyzer® y se procedió a la secuenciación con el sistema Illumina.

15.8 Procesamiento de los datos de CUT&RUN

El análisis de los datos de CUT&RUN inició con la evaluación de la calidad de las lecturas secuenciadas mediante FastQC. Posteriormente, las lecturas fueron alineadas al genoma de referencia humano (hg38) y al genoma de *Echerichia coli* (spike in) utilizando Bowtie2. Los archivos BAM obtenidos se filtraron para eliminar lecturas duplicadas y de

baja calidad. Contamos el número total de lecturas alineadas al genoma humano y al genoma de *E. coli* y calculamos la proporción de lecturas de *E. coli* respecto al número total de lecturas humanas en cada muestra. Este cálculo se usó para determinar un factor de normalización que ajuste la cantidad de señal detectada en las muestras humanas.

Una vez normalizadas las lecturas, los picos fueron llamados utilizando MACS2. Para la anotación de estos picos en relación con elementos genómicos se usó el gen más cercano.

16.- Referencias:

- Akdogan-Ozdilek, B., Duval, K. L., Meng, F. W., Murphy, P. J., & Goll, M. G. (2022). Identification of chromatin states during zebrafish gastrulation using CUT&RUN and CUT&Tag. Developmental Dynamics, 251(4), 729–742. https://doi.org/10.1002/DVDY.430
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature Reviews Genetics, 17(8), 487–500. https://doi.org/10.1038/nrg.2016.59
- Amiri, A., Coppola, G., Scuderi, S., Wu, F., Roychowdhury, T., Liu, F., Pochareddy, S., Shin, Y., Safi, A., Song, L., Zhu, Y., Sousa, A. M. M., Gerstein, M., Crawford, G. E., Sestan, N., Abyzov, A., & Vaccarino, F. M. (2018). Transcriptome and epigenome landscape of human cortical development modeled in organoids. Science, 362(6420). https://doi.org/10.1126/science.aat6720
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., ... Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. Nature, 507(7493), 455–461. https://doi.org/10.1038/nature12787
- Arellano-Carbajal, F., Briseño-Roa, L., Couto, A., Cheung, B. H. H., Labouesse, M., & de Bono, M. (2011). Macoilin, a conserved nervous system-specific ER membrane protein that regulates neuronal excitability. PLoS genetics, 7(3). https://doi.org/10.1371/JOURNAL.PGEN.1001341
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genomewide quantitative enhancer activity maps identified by STARR-seq. Science (New York, N.Y.), 339(6123), 1074–1077. https://doi.org/10.1126/SCIENCE.1232542
- Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a j&Globin Gene Is Enhanced by Remote SV40 DNA Sequences. En Cell (Vol. 27).

- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. En Cell Research (Vol. 21, Número 3, pp. 381–395). https://doi.org/10.1038/cr.2011.22
- Baylin, S. B. (2005). DNA methylation and gene silencing in cancer. Nature clinical practice. Oncology, 2 Suppl 1(SUPPL. 1). https://doi.org/10.1038/NCPONC0354
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B.,
 Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.
 L., & Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental
 Genes in Embryonic Stem Cells. Cell, 125(2), 315–326.
 https://doi.org/10.1016/j.cell.2006.02.041
- Bestor, T. H. (2000). The DNA methyltransferases of mammals. Human Molecular Genetics, 9(16), 2395–2402.
- Bestor, T. H., & Ingram, V. M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proceedings of the National Academy of Sciences of the United States of America, 80(18), 5559–5563. http://www.ncbi.nlm.nih.gov/pubmed/6577443
- Binda, A. V, Kabbani, N., Lin, R., & Levenson, R. (2002). D2 and D3 Dopamine Receptor Cell Surface Localization Mediated by Interaction with Protein 4.1N. http://molpharm.aspetjournals.org
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., ... De Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447(7146), 799–816. https://doi.org/10.1038/nature05874
- Black, J. C., van Rechem, C., & Whetstine, J. R. (2012). Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. Molecular Cell, 48(4), 491–507. https://doi.org/10.1016/J.MOLCEL.2012.11.006
- Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M. R., & Cremer, T. (2005). Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. PLoS Biology, 3(5), e157. https://doi.org/10.1371/journal.pbio.0030157
- Bonev, B., & Cavalli, G. (2016). Organization and function of the 3D genome. En Nature Reviews Genetics (Vol. 17, Número 11, pp. 661–678). Nature Publishing Group. https://doi.org/10.1038/nrg.2016.112
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-Resolution Mapping and Characterization of Open

Chromatin across the Genome. Cell, 132(2), 311–322. https://doi.org/10.1016/J.CELL.2007.12.014

- Bozek, M., & Gompel, N. (2020). Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity: Considering Quantitative Accessibility Changes between Different Regulatory States of an Enhancer Deconvolutes the Complex Relationship between Accessibility and Activity. BioEssays, 42(4). https://doi.org/10.1002/BIES.201900188
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods, 10(12), 1213–1218. https://doi.org/10.1038/nmeth.2688
- Carullo, N. V. N., & Day, J. J. (2019). Genomic enhancers in brain health and disease. En Genes (Vol. 10, Número 1). MDPI AG. https://doi.org/10.3390/genes10010043
- Carullo, N. V. N., Phillips Iii, R. A., Simon, R. C., Soto, S. A. R., Hinds, J. E., Salisbury, A. J., Revanna, J. S., Bunner, K. D., Ianov, L., Sultan, F. A., Savell, K. E., Gersbach, C. A., & Day, J. J. (2020). Enhancer RNAs predict enhancer-gene regulatory links and are critical for enhancer function in neuronal systems. Nucleic Acids Research, 48(17), 9550–9570. https://doi.org/10.1093/nar/gkaa671
- Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution Trashing old rules to forge new ones. Cell, 157(1), 77–94. https://doi.org/10.1016/J.CELL.2014.03.008
- Chen, L. F., Lee, J., & Boettiger, A. (2023). Recent progress and challenges in single-cell imaging of enhancer–promoter interaction. Current Opinion in Genetics and Development, 79. https://doi.org/10.1016/J.GDE.2023.102023
- Chen, L. F., Lin, Y. T., Gallegos, D. A., Hazlett, M. F., Gómez-Schiavon, M., Yang, M. G., Kalmeta, B., Zhou, A. S., Holtzman, L., Gersbach, C. A., Grandl, J., Buchler, N. E., & West, A. E. (2019). Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. Cell reports, 26(5), 1174-1188.e5. https://doi.org/10.1016/J.CELREP.2019.01.032
- Cheng, B., Du, Y., Wen, Y., Zhao, Y., He, A., Ding, M., Fan, Q., Li, P., Liu, L., Liang, X., Guo, X., Zhang, F., & Ma, X. (2019). Integrative analysis of genome-wide association study and chromosomal enhancer maps identified brain region related pathways associated with ADHD. Comprehensive psychiatry, 88, 65–69. https://doi.org/10.1016/J.COMPPSYCH.2018.11.006
- Choi, W., Choe, M. S., Kim, S. M., Kim, S. J., Lee, J., Lee, Y., Lee, S. M., Dho, S. H., Lee, M. Y., & Kim, L. K. (2024). RFX4 is an intrinsic factor for neuronal differentiation through induction of proneural genes POU3F2 and NEUROD1. Cellular and Molecular Life Sciences, 81(1). https://doi.org/10.1007/s00018-024-05129-y

- Clemens, A. W., Wu, D. Y., Moore, J. R., Christian, D. L., Zhao, G., & Gabel, H. W. (2020). MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. Molecular cell, 77(2), 279-293.e8. https://doi.org/10.1016/J.MOLCEL.2019.10.033
- Cogle, C. R., Guthrie, S. M., Sanders, R. C., Allen, W. L., Scott, E. W., & Petersen, B. E. (2003). An Overview of Stem Cell Research and Regulatory Issues. Mayo Clinic Proceedings, 78(8), 993–1003. https://doi.org/10.4065/78.8.993
- Consortium, I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. Nature, 431(7011), 931–945. https://doi.org/10.1038/nature03001
- Consortium, T. E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. https://doi.org/10.1038/nature11247
- Corces, M. R., Shcherbina, A., Kundu, S., Gloudemans, M. J., Frésard, L., Granja, J. M., Louie, B. H., Eulalio, T., Shams, S., Bagdatli, S. T., Mumbach, M. R., Liu, B., Montine, K. S., Greenleaf, W. J., Kundaje, A., Montgomery, S. B., Chang, H. Y., & Montine, T. J. (2020). Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nature Genetics, 52(11), 1158–1168. https://doi.org/10.1038/s41588-020-00721-x
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America, 107(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107
- de Lara, J. C. F., Arzate-Mejía, R. G., & Recillas-Targa, F. (2019). Enhancer RNAs: Insights Into Their Biological Role. En Epigenetics Insights (Vol. 12). SAGE Publications Ltd. https://doi.org/10.1177/2516865719846093
- Deloulme, J. C., Gory-Fauré, S., Mauconduit, F., Chauvet, S., Jonckheere, J., Boulan, B., Mire, E., Xue, J., Jany, M., Maucler, C., Deparis, A. A., Montigon, O., Daoust, A., Barbier, E. L., Bosc, C., Deglon, N., Brocard, J., Denarier, E., Le Brun, I., ... Andrieux, A. (2015). Microtubule-associated protein 6 mediates neuronal connectivity through Semaphorin 3E-dependent signalling for axonal growth. Nature Communications, 6. https://doi.org/10.1038/NCOMMS8246
- Deng, C., Whalen, S., Steyert, M., Ziffra, R., Przytycki, P. F., Inoue, F., Pereira, D. A., Capauto, D., Norton, S., Vaccarino, F. M., Pollen, A. A., Nowakowski, T. J., Ahituv, N., & Pollard, K. S. (2024). Massively parallel characterization of regulatory elements in the developing human cortex. Science, 384(6698), 868–876. https://doi.org/10.1126/science.adh0559

- Dey, A., Chitsaz, F., Abbasi, A., Misteli, T., & Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. En National Institutes of Health (Vol. 100, Número 15). www.pnas.orgcgidoi10.1073pnas.1433065100
- Ding, M., Li, P., Wen, Y., Zhao, Y., Cheng, B., Zhang, L., Ma, M., Cheng, S., Liu, L., Du, Y., Liang, X., He, A., Guo, X., & Zhang, F. (2018). Integrative analysis of genomewide association study and brain region related enhancer maps identifies biological pathways for insomnia. Progress in neuro-psychopharmacology & biological psychiatry, 86, 180–185. https://doi.org/10.1016/J.PNPBP.2018.05.026
- Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. https://doi.org/10.1038/nrm4043
- Du, Y., Ning, Y., Wen, Y., Liu, L., Liang, X., Li, P., Ding, M., Zhao, Y., Cheng, B., Ma, M., Zhang, L., Cheng, S., Yu, W., Hu, S., Guo, X., & Zhang, F. (2018). A genome-wide pathway enrichment analysis identifies brain region related biological pathways associated with intelligence. Psychiatry research, 268, 238–242. https://doi.org/10.1016/J.PSYCHRES.2018.07.029
- Edelman, A. M., Kim, W. Y., Higgins, D., Goldstein, E. G., Oberdoerster, M., & Sigurdson, W. (2005). Doublecortin kinase-2, a novel doublecortin-related protein kinase associated with terminal segments of axons and dendrites. Journal of Biological Chemistry, 280(9), 8531–8543. https://doi.org/10.1074/jbc.M411027200
- Edwards, S. L., Beesley, J., French, J. D., & Dunning, M. (2013). Beyond GWASs: Illuminating the dark road from association to function. En American Journal of Human Genetics (Vol. 93, Número 5, pp. 779–797). Cell Press. https://doi.org/10.1016/j.ajhg.2013.10.012
- Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., Gingeras, T. R., Misteli, T., & Meshorer, E. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. Cell Stem Cell, 2(5), 437–447. https://doi.org/10.1016/j.stem.2008.03.021
- Emery, P., Durand, B., Mach, B., & Reith, W. (1996). RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. En Nucleic Acids Research (Vol. 24, Número 5).
- FANTOM Consortium, T., & Riken Pmi, the. (2014). A promoter-level mammalian expression atlas. Nature. https://doi.org/10.1038/nature13182
- Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., & Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database : the journal of biological databases and curation, 2017. https://doi.org/10.1093/database/bax028

- Forrest, M. P., Zhang, H., Moy, W., McGowan, H., Leites, C., Dionisio, L. E., Xu, Z., Shi, J., Sanders, A. R., Greenleaf, W. J., Cowan, C. A., Pang, Z. P., Gejman, P. V., Penzes, P., & Duan, J. (2017). Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. Cell Stem Cell, 21(3), 305-318.e8. https://doi.org/10.1016/j.stem.2017.07.008
- Franco, D., Sedmera, D., & Lozano-Velasco, E. (2017). Multiple roles of pitx2 in cardiac development and disease. En Journal of Cardiovascular Development and Disease (Vol. 4, Número 4). MDPI. https://doi.org/10.3390/jcdd4040016
- Furlong, E. E. M., & Levine, M. (2018). Developmental enhancers and chromosome topology. Science (New York, N.Y.), 361(6409), 1341–1345. https://doi.org/10.1126/SCIENCE.AAU0320
- Giacoman-Lozano, M., Meléndez-Ramírez, C., Martinez-Ledesma, E., Cuevas-Diaz Duran, R., & Velasco, I. (2022). Epigenetics of neural differentiation: Spotlight on enhancers. En Frontiers in Cell and Developmental Biology (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fcell.2022.1001701
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Research, 17(6), 877–885. https://doi.org/10.1101/gr.5533506
- Grinberg, I., & Millen, K. J. (2005). The ZIC gene family in development and disease. Clinical Genetics, 67(4), 290–296. https://doi.org/10.1111/J.1399-0004.2005.00418.X
- Gross, D. S., & Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. Annual Review of Biochemistry, 57, 159–197. https://doi.org/10.1146/annurev.bi.57.070188.001111
- Grossniklaus, U., & Paro, R. (2014). Transcriptional Silencing by Polycomb-Group Proteins. Cold Spring Harbor Perspectives in Biology, 6(11), 1–26. https://doi.org/10.1101/cshperspect.a019331
- Gupta, S., Dennis, J., Thurman, R. E., Kingston, R., Stamatoyannopoulos, J. A., & Noble, W. S. (2008). Predicting human nucleosome occupancy from primary sequence.
 PLoS computational biology, 4(8). https://doi.org/10.1371/JOURNAL.PCBI.1000134
- Hafner, A., & Boettiger, A. (2023). The spatial organization of transcriptional control. En Nature Reviews Genetics (Vol. 24, Número 1, pp. 53–68). Nature Research. https://doi.org/10.1038/s41576-022-00526-0
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of

transcriptional promoters and enhancers in the human genome. https://doi.org/10.1038/ng1966

- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell. https://doi.org/10.1016/j.molcel.2010.05.004
- Henikoff, S., & Smith, M. M. (2015). Histone Variants and Epigenetics. Cold Spring Harbor Perspectives in Biology, 7(1). https://doi.org/10.1101/CSHPERSPECT.A019364
- Ho, L., Tan, S. Y. X., Wee, S., Wu, Y., Tan, S. J. C., Ramakrishna, N. B., Chng, S. C., Nama, S., Szczerbinska, I., Chan, Y. S., Avery, S., Tsuneyoshi, N., Ng, H. H., Gunaratne, J., Dunn, N. R., & Reversade, B. (2015). ELABELA Is an Endogenous Growth Factor that Sustains hESC Self-Renewal via the PI3K/AKT Pathway. Cell Stem Cell, 17(4), 435–447. https://doi.org/10.1016/J.STEM.2015.08.010
- Hoffman, M. M., Ernst, J., Wilder, S. P., Kundaje, A., Harris, R. S., Libbrecht, M., Giardine, B., Ellenbogen, P. M., Bilmes, J. A., Birney, E., Hardison, R. C., Dunham, I., Kellis, M., & Noble, W. S. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Research, 41(2), 827–841. https://doi.org/10.1093/nar/gks1284
- Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A. M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y. A., ... Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. Nature, 543(7644), 199–204. https://doi.org/10.1038/nature21374
- Hor, H., Francescatto, L., Bartesaghi, L., Ortega-Cubero, S., Kousi, M., Lorenzo-Betancor, O., Jiménez-Jiménez, F. J., Gironell, A., Clarimón, J., Drechse, O., Agúndez, J. A. G., Kenzelmann Broz, D., Chiquet-Ehrismann, R., Lleó, A., Coria, F., García-Martin, E., Alonso-Navarro, H., Martí, M. J., Kulisevsky, J., ... Estivill, X. (2015). Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. Human molecular genetics, 24(20), 5677–5686. https://doi.org/10.1093/HMG/DDV281
- Huang, D.-H., Chang, Y.-L., Yang, C.-C., Pan, I.-C., & King, B. (2002). pipsqueak Encodes a Factor Essential for Sequence-Specific Targeting of a Polycomb Group Protein Complex . Molecular and Cellular Biology, 22(17), 6261–6271. https://doi.org/10.1128/mcb.22.17.6261-6271.2002
- Husmann, D., & Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. Nature structural & molecular biology, 26(10), 880–889. https://doi.org/10.1038/S41594-019-0298-7

- Ibn-Salem, J., Köhler, S., Love, M. I., Chung, H.-R., Huang, N., Hurles, M. E., Haendel, M., Washington, N. L., Smedley, D., Mungall, C. J., Lewis, S. E., Ott, C.-E., Bauer, S., Schofield, P. N., Mundlos, S., Spielmann, M., & Robinson, P. N. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. En Genome Biology (Vol. 15). http://genomebiology.com/2014/15/9/423
- Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature reviews. Genetics, 10(3), 161–172. https://doi.org/10.1038/NRG2522
- Jiang, S., & Mortazavi, A. (2018). Integrating ChIP-seq with other functional genomics data. Briefings in Functional Genomics, 17(2), 104–115. https://doi.org/10.1093/BFGP/ELY002
- Kalashnikova, E., Lorca, R. A., Kaur, I., Barisone, G. A., Li, B., Ishimaru, T., Trimmer, J. S., Mohapatra, D. P., & Díaz, E. (2010). SynDIG1: an activity-regulated, AMPA-receptor-interacting transmembrane protein that regulates excitatory synapse development. Neuron, 65(1), 80–93. https://doi.org/10.1016/J.NEURON.2009.12.021
- Kang, J. Y., Song, S. H., Yun, J., Jeon, M. S., Kim, H. P., Han, S. W., & Kim, T. Y. (2015). Disruption of CTCF/cohesin-mediated high-order chromatin structures by DNA methylation downregulates PTGS2 expression. Oncogene, 34(45), 5677–5684. https://doi.org/10.1038/onc.2015.17
- Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nature communications, 10(1). https://doi.org/10.1038/S41467-019-09982-5
- Kazakevych, J., Sayols, S., Messner, B., Krienke, C., & Soshnikova, N. (2017). Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. Nucleic Acids Research, 45(10), 5770–5784. https://doi.org/10.1093/nar/gkx167
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Morales, D. R., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., & Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatinmodifying complexes and affect gene expression. Proceedings of the National Academy of Sciences of the United States of America, 106(28), 11667. https://doi.org/10.1073/PNAS.0904715106
- Kit, W., Xiu-Rong, R., Yang-Zhong, H., Yi, X., Guofa, L., Harumi, S., Hao, T., Leng, W., Susann, B.-K., Lin, M., Jane, W., Wen-Cheng, X., & Yi, R. (2001). Signal Transduction in Neuronal Migration: Roles of GTPase Activating Proteins and the Small GTPase Cdc42 in the Slit-Robo Pathway. Cell, 107, 209–221.

- Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., Matsuo, M., Kamijo, S. I., Kasahara, M., Yoshioka, H., Ogata, T., Fukuda, T., Kondo, I., Kato, M., Dobyns, W. B., ... Morohashi, K. I. (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nature Genetics, 32(3), 359–369. https://doi.org/10.1038/ng1009
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., Hayashizaki, Y., & Carninci, P. (2006). Cage: Cap analysis of gene expression. Nature Methods, 3(3), 211. https://doi.org/10.1038/nmeth0306-211
- Koh, P. W., Sinha, R., Barkal, A. A., Morganti, R. M., Chen, A., Weissman, I. L., Ang, L. T., Kundaje, A., & Loh, K. M. (2016). An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. Scientific data, 3. https://doi.org/10.1038/SDATA.2016.109
- Kriks, S., Shim, J.-W., Piao, J., Ganat, Y. M., Wakeman, D. R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., Yang, L., Beal, M. F., Surmeier, D. J., Kordower, J. H., Tabar, V., & Studer, L. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature, 480(7378), 547–551. https://doi.org/10.1038/nature10648
- Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science, 339(6122), 950–953. https://doi.org/10.1126/science.1229386
- Lettice, L. A., Williamson, I., Devenney, P. S., Kilanowski, F., Dorin, J., & Hill, R. E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. Development (Cambridge), 141(8), 1715–1725. https://doi.org/10.1242/dev.095430
- Li, M., Santpere, G., Kawasawa, Y. I., Evgrafov, O. V., Gulden, F. O., Pochareddy, S., Sunkin, S. M., Li, Z., Shin, Y., Zhu, Y., Sousa, A. M. M., Werling, D. M., Kitchen, R. R., Kang, H. J., Pletikos, M., Choi, J., Muchnik, S., Xu, X., Wang, D., ... Sestan, N. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science (New York, N.Y.), 362(6420). https://doi.org/10.1126/SCIENCE.AAT7615
- Li, N., Jin, K., Bai, Y., Fu, H., Liu, L., & Liu, B. (2020). Tn5 Transposase Applied in Genomics Research. International journal of molecular sciences, 21(21), 1–15. https://doi.org/10.3390/IJMS21218329
- Li, S., Liu, H., Liu, W., Shi, N., Zhao, M., Wanggou, S., Luo, W., Wang, L., Zhu, B., Zuo, X., Xie, W., Zhao, C., Zhou, Y., Luo, L., Gao, X., Jiang, X., & Ren, C. (2023). ESRG is critical to maintain the cell survival and self-renewal/pluripotency of hPSCs by

collaborating with MCM2 to suppress p53 pathway. International Journal of Biological Sciences, 19(3), 916–935. https://doi.org/10.7150/ijbs.79095

- Li, S., Shin, Y., Cho, K. W. Y., & Merzdorf, C. S. (2006). The Xfeb gene is directly upregulated by Zic1 during early neural development. Developmental Dynamics, 235(10), 2817–2827. https://doi.org/10.1002/DVDY.20896
- Li, Y. E., Preissl, S., Miller, M., Johnson, N. D., Wang, Z., Jiao, H., Zhu, C., Wang, Z., Xie, Y., Poirion, O., Kern, C., Pinto-Duarte, A., Tian, W., Siletti, K., Emerson, N., Osteen, J., Lucero, J., Lin, L., Yang, Q., ... Ren, B. (2023). A comparative atlas of single-cell chromatin accessibility in the human brain. Science, 382(6667). https://doi.org/10.1126/science.adf7044
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science, 326(5950), 289–293. https://doi.org/10.1126/science.1181369
- Lorberbaum, D. S., Ramos, A. I., Peterson, K. A., Carpenter, B. S., Parker, D. S., De, S., Hillers, L. E., Blake, V. M., Nishi, Y., McFarlane, M. R., Chiang, A. C. Y., Kassis, J. A., Allen, B. L., McMahon, A. P., & Barolo, S. (2016). An ancient yet flexible cisregulatory architecture allows localized hedgehog tuning by patched/Ptch1. eLife, 5(MAY2016). https://doi.org/10.7554/ELIFE.13550
- Ma, H., de Zwaan, E., Guo, Y. E., Cejas, P., Thiru, P., van de Bunt, M., Jeppesen, J. F., Syamala, S., Dall'Agnese, A., Abraham, B. J., Fu, D., Garrett-Engele, C., Lee, T. I., Long, H. W., Griffith, L. G., Young, R. A., & Jaenisch, R. (2022). The nuclear receptor THRB facilitates differentiation of human PSCs into more mature hepatocytes. Cell Stem Cell, 29(5), 795-809.e11. https://doi.org/10.1016/j.stem.2022.03.015
- Ma, M., Liang, X., Cheng, S., Zhang, L., Cheng, B., Yang, J., Ding, M., Zhao, Y., Li, P., Du, Y., Liu, L., Wen, Y., He, A., Fan, Q. R., Guo, X., & Zhang, F. (2018). Integrating genome-wide association study, chromosomal enhancer maps and element-gene interaction networks detected brain regions related associations between elements and ADHD/IQ. Behavioural brain research, 353, 137–142. https://doi.org/10.1016/J.BBR.2018.07.008
- Mansisidor, A. R., & Risca, V. I. (2022). Chromatin accessibility: methods, mechanisms, and biological insights. Nucleus (Austin, Tex.), 13(1), 236–276. https://doi.org/10.1080/19491034.2022.2143106
- Martin, D. M., Skidmore, J. M., Philips, S. T., Vieira, C., Gage, P. J., Condie, B. G., Raphael, Y., Martinez, S., & Camper, S. A. (2004). PITX2 is required for normal

development of neurons in the mouse subthalamic nucleus and midbrain. Developmental Biology, 267(1), 93–108. https://doi.org/10.1016/j.ydbio.2003.10.035

- Matsuo, N., Hoshino, M., Yoshizawa, M., & Nabeshima, Y. I. (2002). Characterization of STEF, a guanine nucleotide exchange factor for Rac1, required for neurite growth. The Journal of biological chemistry, 277(4), 2860–2868. https://doi.org/10.1074/JBC.M106186200
- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell, 58(3), 499–507. https://doi.org/10.1016/0092-8674(89)90430-3
- Meléndez-Ramírez, C., Cuevas-Diaz Duran, R., Barrios-García, T., Giacoman-Lozano, M., López-Ornelas, A., Herrera-Gamboa, J., Estudillo, E., Soto-Reyes, E., Velasco, I., & Treviño, V. (2021). Dynamic landscape of chromatin accessibility and transcriptomic changes during differentiation of human embryonic stem cells into dopaminergic neurons. Scientific Reports 2021 11:1, 11(1), 1–18. https://doi.org/10.1038/s41598-021-96263-1
- Misteli, T. (2007). Beyond the Sequence: Cellular Organization of Genome Function. Cell, 128(4), 787–800. https://doi.org/10.1016/j.cell.2007.01.028
- Ng, S. Y., Johnson, R., & Stanton, L. W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO Journal, 31(3), 522–533. https://doi.org/10.1038/EMBOJ.2011.459
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the Xinactivation centre. Nature, 485(7398), 381–385. https://doi.org/10.1038/nature11049
- Padhi, E. M., Hayeck, T. J., Cheng, Z., Chatterjee, S., Mannion, B. J., Byrska-Bishop, M., Willems, M., Pinson, L., Redon, S., Benech, C., Uguen, K., Audebert-Bellanger, S., Le Marechal, C., Férec, C., Efthymiou, S., Rahman, F., Maqbool, S., Maroofian, R., Houlden, H., ... Turner, T. N. (2021). Coding and noncoding variants in EBF3 are involved in HADDS and simplex autism. Human genomics, 15(1). https://doi.org/10.1186/S40246-021-00342-3
- Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. En Nature Reviews Genetics (Vol. 10, Número 10, pp. 669–680). https://doi.org/10.1038/nrg2641
- Pierce, S. E., Tyson, T., Booms, A., Prahl, J., & Coetzee, G. A. (2018). Parkinson's disease genetic risk in a midbrain neuronal cell line. Neurobiology of Disease, 114, 53–64. https://doi.org/10.1016/j.nbd.2018.02.007

- Pladevall-Morera, D., & Zylicz, J. J. (2022). Chromatin as a sensor of metabolic changes during early development. Frontiers in cell and developmental biology, 10. https://doi.org/10.3389/FCELL.2022.1014498
- Pristerà, A., Lin, W., Kaufmann, A. K., Brimblecombe, K. R., Threlfell, S., Dodson, P. D., Magill, P. J., Fernandes, C., Cragg, S. J., & Ang, S. L. (2015). Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice. Proceedings of the National Academy of Sciences of the United States of America, 112(35), E4929–E4938. https://doi.org/10.1073/pnas.1503911112
- Ramos-Mejía, V., Navarro-Montero, O., Ver´, V., Aylí, V., Bueno, C., Romero, T., Real, P. J., & Menendez, P. (2014). HOXA9 promotes hematopoietic commitment of human embryonic stem cells. https://doi.org/10.1182/blood-2014-03
- Rando, O. J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. https://doi.org/10.1016/j.gde.2012.02.013
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell, 159(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
- Ren, W.-Z., Ng, G. Y. K., Wang, R.-X., Wu, P. H., O'dowd, B. F., Osmond, D. H., George, S. R., & Liew, C.-C. (1994). The identification of NP25" a novel protein that is differentially expressed by neuronal subpopulations. En Molecular Brain Research (Vol. 22).
- Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. Nature, 311(5986), 532–537. https://doi.org/10.1038/311532a0
- Riveiro, A. R., & Brickman, J. M. (2020). From pluripotency to totipotency: an experimentalist's guide to cellular potency. Development (Cambridge, England), 147(16). https://doi.org/10.1242/DEV.189845
- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., ... Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. Nature, 518(7539), 317. https://doi.org/10.1038/NATURE14248
- Rodriguez-Oroz, M. C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., & Obeso, J. A. (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. En The Lancet Neurology (Vol. 8, Número 12, pp. 1128–1139). https://doi.org/10.1016/S1474-4422(09)70293-5

- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: Insights from genome-wide studies. En Nature Reviews Genetics (Vol. 8, Número 6, pp. 424–436). https://doi.org/10.1038/nrg2026
- Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell, 132(5), 887–898. https://doi.org/10.1016/j.cell.2008.02.022
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell, 148(3), 458–472. https://doi.org/10.1016/j.cell.2012.01.010
- Siersbæk, R., Madsen, J. G. S., Javierre, B. M., Nielsen, R., Bagge, E. K., Cairns, J., Wingett, S. W., Traynor, S., Spivakov, M., Fraser, P., & Mandrup, S. (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Molecular Cell, 66(3), 420-435.e5. https://doi.org/10.1016/j.molcel.2017.04.010
- Sigova, A. A., Abraham, B. J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y. E., Jangi, M., Giallourakis, C. C., Sharp, P. A., & Young, R. A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science, 350(6263), 978–991. https://doi.org/10.1126/SCIENCE.AAD3346/SUPPL_FILE/SIGOVA.SM.REVISION.1. PDF
- Simonis, M., Kooren, J., & de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. Nature methods, 4(11), 895–901. https://doi.org/10.1038/NMETH1114
- Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for highresolution mapping of DNA binding sites. eLife, 6. https://doi.org/10.7554/ELIFE.21856
- Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Reviews Genetics 2013 14:3, 14(3), 204–220. https://doi.org/10.1038/nrg3354
- Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., Sheffield, N. C., Gräf, S., Huss, M., Keefe, D., Liu, Z., London, D., McDaniell, R. M., Shibata, Y., Showers, K. A., Simon, J. M., Vales, T., Wang, T., Winter, D., ... Furey, T. S. (2011). Open chromatin defined by DNasel and FAIRE identifies regulatory elements that shape cell-type identity. Genome Research, 21(10), 1757–1767. https://doi.org/10.1101/gr.121541.111
- Splinter, E., Grosveld, F., & De Laat, W. (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods in enzymology, 375, 493–507. https://doi.org/10.1016/S0076-6879(03)75030-7

- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. Nature, 403(6765), 41–45. https://doi.org/10.1038/47412
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, N.Y.), 324(5929), 930–935. https://doi.org/10.1126/SCIENCE.1170116
- Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., Michalski, P., Piecuch, E., Wang, P., Wang, D., Tian, S. Z., Penrad-Mobayed, M., Sachs, L. M., Ruan, X., Wei, C. L., ... Ruan, Y. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell, 163(7), 1611–1627. https://doi.org/10.1016/J.CELL.2015.11.024
- Tao, X., Feng, S., Zhao, T., & Guan, X. (2020). Efficient chromatin profiling of H3K4me3 modification in cotton using CUT&Tag. Plant methods, 16(1). https://doi.org/10.1186/S13007-020-00664-8
- Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., Becker, R. A., & Hearn, J. P. (1995). Isolation of a primate embryonic stem cell line. Proceedings of the National Academy of Sciences of the United States of America, 92(17), 7844– 7848. http://www.ncbi.nlm.nih.gov/pubmed/7544005
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., ... Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. Nature 2012 489:7414, 489(7414), 75–82. https://doi.org/10.1038/nature11232
- Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., Stunnenberg, H. G., Mann, M., & Timmers, H. T. M. (2007). Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. Cell, 131(1), 58–69. https://doi.org/10.1016/J.CELL.2007.08.016
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature, 457(7231), 854–858. https://doi.org/10.1038/nature07730
- Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., Giresi, P. G., Ng, Y. H., Marro, S., Neff, N. F., Drechsel, D., Martynoga, B., Castro, D. S., Webb, A. E., Südhof, T. C., Brunet, A., Guillemot, F., Chang, H. Y., & Wernig, M. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell, 155(3), 621. https://doi.org/10.1016/j.cell.2013.09.028

- Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell, 171(7), 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008
- Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A., & Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. NATURE GENETICS, 41(2). https://doi.org/10.1038/ng.297
- Winter, G. E., Mayer, A., Buckley, D. L., Erb, M. A., Roderick, J. E., Vittori, S., Reyes, J. M., di Iulio, J., Souza, A., Ott, C. J., Roberts, J. M., Zeid, R., Scott, T. G., Paulk, J., Lachance, K., Olson, C. M., Dastjerdi, S., Bauer, S., Lin, C. Y., ... Bradner, J. E. (2017). BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. Molecular cell, 67(1), 5-18.e19. https://doi.org/10.1016/J.MOLCEL.2017.06.004
- Yang, J., Yan, B., Fan, Y., Yang, L., Zhao, B., Zhu, F., Zheng, J., Wang, W., Bai, L., Zhang, F., & Ma, X. (2019). Identification of schizophrenia related biological pathways across eight brain regions. Behavioural brain research, 360, 1–6. https://doi.org/10.1016/J.BBR.2018.11.011
- Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae.
 Science, 309(5734), 626–630.
 https://doi.org/10.1126/SCIENCE.1112178/SUPPL_FILE/YUAN.SOM.PDF
- Zeng, Q., Pan, H., Zhao, Y., Wang, Y., Xu, Q., Tan, J., Yan, X., Li, J., Tang, B., & Guo, J. (2022). Association Study of TAF1 Variants in Parkinson's Disease. Frontiers in Neuroscience, 16. https://doi.org/10.3389/fnins.2022.846095
- Zhang, D., Zeldin, D. C., & Blackshear, P. J. (2007). Regulatory factor X4 variant 3: a transcription factor involved in brain development and disease. Journal of neuroscience research, 85(16), 3515–3522. https://doi.org/10.1002/JNR.21356
- Zhang, K., Hocker, J. D., Miller, M., Hou, X., Chiou, J., Poirion, O. B., Qiu, Y., Li, Y. E., Gaulton, K. J., Wang, A., Preissl, S., & Ren, B. (2021). A single-cell atlas of chromatin accessibility in the human genome. Cell, 184(24), 5985-6001.e19. https://doi.org/10.1016/J.CELL.2021.10.024
- Zhou, P., VanDusen, N. J., Zhang, Y., Cao, Y., Sethi, I., Hu, R., Zhang, S., Wang, G., Ye, L., Mazumdar, N., Chen, J., Zhang, X., Guo, Y., Li, B., Ma, Q., Lee, J. Y., Gu, W., Yuan, G. C., Ren, B., ... Pu, W. T. (2023). Dynamic changes in P300 enhancers and enhancer-promoter contacts control mouse cardiomyocyte maturation.
 Developmental Cell, 58(10), 898-914.e7. https://doi.org/10.1016/j.devcel.2023.03.020

 Zhu, K., Bendl, J., Rahman, S., Vicari, J. M., Coleman, C., Clarence, T., Latouche, O., Tsankova, N. M., Li, A., Brennand, K. J., Lee, D., Yuan, G. C., Fullard, J. F., & Roussos, P. (2023). Multi-omic profiling of the developing human cerebral cortex at the single-cell level. Science Advances, 9(41). https://doi.org/10.1126/sciadv.adg3754

17.-Anexo:

Dynamic landscape of chromatin accessibility and transcriptomic changes during differentiation of human embryonic stem cells into dopaminergic neurons

scientific reports

OPEN

Check for updates

Dynamic landscape of chromatin accessibility and transcriptomic changes during differentiation of human embryonic stem cells into dopaminergic neurons

César Meléndez-Ramírez^{1,2,6}, Raquel Cuevas-Diaz Duran^{3,6}, Tonatiuh Barrios-García³, Mayela Giacoman-Lozano³, Adolfo López-Ornelas^{1,2,4}, Jessica Herrera-Gamboa³, Enrique Estudillo², Ernesto Soto-Reyes⁵, Iván Velasco^{1,2} & Víctor Treviño³

Chromatin architecture influences transcription by modulating the physical access of regulatory factors to DNA, playing fundamental roles in cell identity. Studies on dopaminergic differentiation have identified coding genes, but the relationship with non-coding genes or chromatin accessibility remains elusive. Using RNA-Seg and ATAC-Seg we profiled differentially expressed transcripts and open chromatin regions during early dopaminergic neuron differentiation. Hierarchical clustering of differentially expressed genes, resulted in 6 groups with unique characteristics. Surprisingly, the abundance of long non-coding RNAs (IncRNAs) was high in the most downregulated transcripts, and depicted positive correlations with target mRNAs. We observed that open chromatin regions decrease upon differentiation. Enrichment analyses of accessibility depict an association between open chromatin regions and specific functional pathways and gene-sets. A bioinformatic search for motifs allowed us to identify transcription factors and structural nuclear proteins that potentially regulate dopaminergic differentiation. Interestingly, we also found changes in protein and mRNA abundance of the CCCTC-binding factor, CTCF, which participates in genome organization and gene expression. Furthermore, assays demonstrated co-localization of CTCF with Polycomb-repressed chromatin marked by H3K27me3 in pluripotent cells, progressively decreasing in neural precursor cells and differentiated neurons. Our work provides a unique resource of transcription factors and regulatory elements, potentially involved in the acquisition of human dopaminergic neuron cell identity.

In eukaryotic cells, DNA is bound to histones forming chromatin, whose structure is dynamic and suffers reversible chemical changes, mainly DNA methylation and histone post-translational modifications. These epigenetic modifications have gained increasing interest since they play fundamental roles in modulating gene expression throughout development, differentiation, and in response to environmental cues¹. Chromatin modifications alter its packing level, resulting in mainly two distinct environments: 'open' accessible regions, also referred to as euchromatin, and 'closed' compact regions or heterochromatin. These regions are enriched or depleted in specific histone modifications. Euchromatin is permissive for gene activation, whereas genes in heterochromatin are mainly silenced. Interestingly, regions bound by certain proteins such as CTCF, have a role in maintaining the boundary between heterochromatin and euchromatin². Studies have revealed the role of CTCF in regulating chromatin loops' formation and, hence, in controlling three-dimensional nuclear landscape to regulate gene expression³⁻⁵.

¹Instituto de Fisiología Celular—Neurociencias, Universidad Nacional Autónoma de México, Mexico City, Mexico. ²Laboratorio de Reprogramación Celular, Instituto Nacional de Neurología y Neurocirugía "Manuel Velasco Suárez", Mexico City, Mexico. ³Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, Mexico. ⁴División de Investigación, Hospital Juárez de México, Mexico City, Mexico. ⁵Universidad Autónoma Metropolitana—Cuajimalpa, Mexico City, Mexico. ⁶These authors contributed equally: César Meléndez-Ramírez and Raquel Cuevas-Diaz Duran. [⊠]email: raquel.cuevas.dd@tec.mx; ivelasco@ifc.unam.mx; vtrevino@tec.mx Chromatin accessibility is highly dynamic and plays essential roles in establishing and maintaining cellular identity⁶. Euchromatin contributes to the pluripotency of embryonic stem cells (ESCs), whereas heterochromatin regions increase upon differentiation^{7,8}. Epigenomic regulation and lineage-specific gene expression orchestrate cellular differentiation; however, the temporal interplay between these mechanisms remains elusive. Several large-scale efforts have characterized transcriptional and epigenetic landscapes of cell lines and tissues^{9,10}. Even though these studies have identified putative regulatory elements, none of them focused on dopaminergic differentiation.

Mesencephalic dopaminergic neurons (mDA) are of great medical interest because they are lost in Parkinson's disease (PD)¹¹. Currently, protocols to obtain mDA from ESCs have been described¹². Differentiated neurons in vitro express several dopaminergic markers, like Tyrosine Hydroxylase (TH), GIRK2, LMX1A, EN1 and FOXA2. These neurons have electrophysiological activities that resemble mDA. Furthermore, ESC derived neurons release dopamine, survive after grafting in parkinsonian non-human mammalian models, and improve the behavioral test performance of these animals, when compared with non-grafted condition¹³. Therefore, differentiation of human ESCs to mDA provides an important model for studying the correlation between the dynamic chromatin accessibility landscape and gene expression changes.

PD is the second most common neurodegenerative disease and its main pathological characteristic is the loss of mDA. Genome-wide association studies (GWAS) have revealed that > 80% of disease-associated SNPs are found within poorly annotated regions¹⁴, making the identification of causal polymorphisms a difficult task. Disease-associated SNPs in psychiatric and neurological diseases reside in open chromatin regions (OCRs) during neural development^{15,16}. Familial PD results from a single mendelian recessive or dominant mutation on one of 11 identified genes, and accounts for roughly 5% of cases¹⁷. Unlike familial PD, sporadic PD is more complex and results from a combination of environmental and genetic predisposition linked to SNPs.

Here we set out to generate a more detailed genome-wide map of the transcriptional changes (RNA-Seq) and chromatin accessibility (ATAC-Seq) of dopaminergic induction. We integrated epigenomic and transcriptomic profiles to infer the activity of transcription factors (TFs) and DNA regulatory regions such as enhancers and long non-coding RNAs (lncRNAs). Our work provides a unique resource of TFs and regulatory elements, potentially orchestrating human mDA specification.

Results

Human ES cells efficiently differentiate into mDA. We differentiated pluripotent human ESC H9-GFP¹⁸ expressing OCT4 and SOX2 into mDA (Fig. 1a), using a reported floor-plate protocol¹². NES-TIN + neural precursors (NPC) were present at D14 (Fig. 1b). Neurons identified by the BETA-3 TUBULIN (TUJ1) antibody, that co-express TH (Fig. 1c) and FOXA2 (Fig. 1d) are present at D28. It is important to note that at D14, TH is not expressed; this enzyme is detected by immunocytochemistry in 15% of neurons at D24, reaching 50% of TH + neurons at D28 and this percentage increases to 80% at D42¹³. Immunoblot experiments in pluripotent cells and NPC show that they do not express TH, and this enzyme shows increasing levels at D21 and D28 (Fig. 1e). To further characterize the transition from pluripotency to neurons, we performed immunoblots for OCT4, SOX2, and TUJ1. The pluripotency-related transcription factor OCT4 was present only at D0. SOX2 was present in pluripotent cells and was down-regulated at D14, appearing again at D28, consistent with its role in neural differentiation. On the other hand, TUJ1 increased at D14 and D28 (Fig. S1). Thus, the times selected for this study allow a clear distinction between pluripotent cells, neural precursors, and early mDA. Extended periods of differentiation yield mDA that fire action potentials and release DA in vitro at D60 and D70, respectively¹³.

RNA-Seq reveals transcriptional profiles associated with mDA differentiation. Dopaminergic differentiation can be monitored by mRNA expression: H9 cells at D0 express high levels of *OCT4* and *SOX2*, but not *LMX1A*, *FOXA2* or *TH*, by RT-qPCR; at D14, *LMX1A* and *FOXA2* were detected, and its expression progressively increased. Transcripts for TH were initially detected at D21 and significantly increased afterwards¹³. Here we performed RNA-Seq at D0, D14 and D28 to determine transcriptional profiles underlying early DA induction. The number and category of DEGs found in each comparison are depicted in Fig. 2a,b. A list of 4009 DEGs (3377 protein-coding genes and 632 lncRNAs) was compiled from all comparisons and used for downstream analysis. Gene expression values (FPKM, normalized counts, and batch effect adjusted counts) and statistics of differential expression are included in Supplementary Tables S1 and S2 respectively.

To identify transcriptional patterns, we performed hierarchical clustering of DEGs, resulting in 6 clusters with distinct temporal profiles (Fig. 2c, Supplementary Table S3). Cluster 1 (825 coding genes and 220 lncRNAs) displays downregulation in D14 and D28. Top enriched gene-sets are related to pluripotency and proliferation, consistent with pluripotency loss as cells differentiate. Cluster 2 (309 coding genes and 13 lncRNAs) exhibits significant downregulation in D14, with upregulation at D28; mainly cell cycle gene-sets are enriched. Cluster 3 (410 coding genes and 47 lncRNAs) displays high expression at D0 and D14 and downregulation in D28, with enriched functions related to the regulation of proliferation and cell adhesion. Cluster 4 (581 protein-coding genes and 150 lncRNAs) is upregulated in D14 and decreases its expression in D28. Enriched gene-sets are related to head and hindbrain development, pattern specification, and regionalization. Cluster 5 (163 coding genes and 33 lncRNAs) depicts an upregulation at D14 with a downregulation at D28. Enriched gene-sets include extracellular structure organization. Cluster 6 (1090 coding genes and 168 lncRNAs) displays low levels at D0 and D14 and becomes upregulated at D28. Enriched gene-sets are mainly related to neurogenesis, synapse, axon, and DA neuron differentiation. A heatmap of selected DEGs related to pluripotency and neural DA differentiation is presented in Fig. S3. This analysis confirms that neural lineages are clearly observable from upregulated genes at D14 and maintained at D28 (Cluster 4) or upregulated at D28 (Cluster 6).











Figure 1. Dopaminergic differentiation of human ESC is progressive, showing three defined stages. (a) Fluorescence images depicting immunostaining of pluripotency markers OCT4 and SOX2 in H9-GFP hESC colonies at D0. Nuclei were labeled with Hoechst. (b) Immunofluorescence at D14 showing the high proportion of cells positive for NESTIN, a marker of neural precursors and GFP in all cells. (c) Staining with the neuronal marker TUJ1 and the dopaminergic marker TH at D28. (d) Neurons also express the dopaminergic marker FOXA2, indicating appropriate differentiation. Scale bars, 100 µm. (e) Representative Western blot of total cell lysate showing the increase in TH (~60 kDa) levels during dopaminergic differentiation. GAPDH was used as a loading control. The graph shows the mean ± SD of two independent experiments.

Scientific Reports | (2021) 11:16977 |


Figure 2. Dopaminergic induction of human ESCs reveals transcriptional dynamic profiles composed of both protein-coding genes and lncRNAs. (a) Venn diagram showing the extent of overlap of DEGs found from pairwise comparisons. (b) Number of DE protein-coding genes and lncRNAs found in each pairwise comparison. (c) Hierarchical clustering analysis displaying temporal expression profiles of all 4009 DEGs compiled from time-point pairwise comparisons. Highly enriched gene-sets and ontologies were identified (hypergeometric test with FDR < 0.05). The enrichment of gene-sets is indicated as – log10(FDR), with the number of genes indicated in parenthesis. Heatmap was built using row z-scores of log2 transformed FPKM gene expression values.

Top significant DEGs are lncRNAs. We used the absolute FC to analyze the top ranked DEGs. Overall, lncRNAs are ~16% of the DEGs between D0 and D28 (Fig. 2b). Among the top 10 DEGs in neural precursors (D0 vs. D14), 6 are lncRNAs, and only 4 are protein-coding genes (Fig. 3a). Interestingly, the number of lncRNAs within the top (10, 20, 30, 40, and 50) ranked genes is statistically more significant than the number of protein-coding genes (Fig. 3a, Supplementary Table S4). A similar pattern is observed when comparing the number of lncRNAs and protein-coding genes between D0 vs. D28. This proportion changed when analyzing the top DEGs between D14 and D28, with a higher significance in the number of protein-coding genes. Almost all lncRNAs within the top 50 DEGs in the comparisons D0 vs. D14 and D0 vs. D28 are downregulated. These results suggest that lncRNAs participate in pluripotency and decrease upon neural differentiation.

We categorized DE lncRNAs according to GENCODE annotation which considers their location relative to protein-coding genes. Using a hypergeometric test, we found that a significant number of DE lncRNAs are classified as processed transcripts and long intergenic non-coding RNA (lincRNA). The high number of lincRNAs suggests potential regulatory functions in *trans* (Fig. 3b). Genes are classified as processed transcripts when their transcripts lack an open reading frame and due to complexity in their structure, they cannot be placed in any other category¹⁹. Interestingly, the number of DE processed transcripts is significant in all pairwise comparisons. Given the predominance of top DE lncRNAs we searched among all pairwise comparisons and found 96 DE lncRNAs (majority antisense) with known regulatory functions. We found that 56% of DE lncRNAs (54 out of 96) had significant gene expression correlation (Pearson correlation *p*-value < 0.05, Supplementary Table S5) with their target protein-coding genes in samples from all time points. From significant correlations, 91% (49 out of 54) were positive and only 9% (5 out of 54) were negative, suggesting that DE lncRNAs may regulate target protein-coding genes in *cis* and in concordant direction (up- or downregulated). For visualization purposes, we contrasted FCs of DE lncRNAs with their related protein-coding genes (Fig. 3c). Examples of positively correlated genes (either down or upregulated) include *SOX1:SOX1-OT*, *EPHA1: EPHA1-AS1*, *FOXD3:FOXD3-AS1* and *DIO3:DIO3OS*. However, a few pairs, like *CACNA1C:CACNA1CA-AS2*, showed a negative correlation.

Stages of DA induction are characterized by differentially accessible chromatin regions. We used ATAC-Seq to identify genomic chromatin accessibility depicting active (open) and inactive (closed) chromatin during DA induction. We identified OCRs in samples from D0, D14, and D28 collected simultaneously from the same experiments used for RNA-Seq. Nucleosome packing and an apparent periodicity of approximately 200 bp in insert size distribution plots were evident (Fig. 4a), consistent with previous results⁶. The majority of reads were located in regions less than 100 bp, indicating OCR. Mononucleosome, dinucleosome, and trinucleosome patterns were also observed.

From ATAC-Seq, we identified a total of 464,783 consensus peaks by merging results from all sample timepoints. By comparing peak region locations, we found categories of peaks that were either specific to, or shared between time-points. Overall, 103,989 (22%) of the consensus peaks were specific to D0, 150,787 (32%) were D14-specific, and 89,235 (19%) were D28-specific. Only 49,126 (11%) of consensus peaks were shared among all time-points (Fig. 4b). By mapping consensus peaks to genomic regions, we found a higher proportion of peaks in intronic and intergenic regions (Fig. 4c). Due to the fact that we have one sample per time-point, we compared our ATAC-Seq peaks with public data derived from H9 ESCs²⁰ to assess the quality of our peak regions. We found that 86% and 93% of ATAC-Seq peak regions in two replicates²⁰ overlapped with our D0-specific peak regions.

To analyze global accessibility, we used +/- 5 kb of identified OCRs. Higher chromatin accessibility in D0, compared to D14 and D28 was observed (Fig. 4d), consistent with published results²¹. Additionally, we compared time-point specific OCRs with enhancer regions predicted by GeneHancer²² whose database incorporated DNase and H3K27ac assays using H9 pluripotent stem cells. We found that 60% (62,250 out of 103,989) of D0 specific OCRs overlapped GeneHancer annotated regions. In contrast, 31% (46,571 out of 150,787) and 42% (37,700 out of 89,235) of D14 and D28 specific OCR regions respectively, overlapped GeneHancer regions.

To find pathways and gene-sets revealed by OCR specific to different time-points, we tested for statistical differences. We found 2548, 107, and 1236 peaks at D0, D14, and D28, respectively, which were mapped to promoters, exons, or introns to perform a gene-set enrichment analysis (FDR<0.05). Significant gene-sets enriched in D0-specific peaks are diverse but related to negative regulation of differentiation, stem cell division, and endodermal/mesodermal/ectodermal differentiation (Fig. 4e). In contrast, D28 presented neuronal-related processes, such as neurogenesis, dendrite and axonal development, and synaptic function (Fig. 4f). The list of differentially enriched gene-sets for time-point specific OCRs is included in Supplementary Table S6.

Temporal patterns of chromatin accessibility are correlated to transcriptomic profiles. To correlate significant differential chromatin accessibility and differential gene expression, we selected consensus ATAC-Seq peaks mapped to promoters, exons, or introns whose *p*-value < 0.05 between D0 and D28. The results show a significant correlation between FC of filtered genes (cor=0.44, *p*-value=5.54e-51, Fig. 5a) consistent with previous work²³.

Coverage plots were used to analyze chromatin accessibility in the promoter regions of DEGs for each cluster depicted in Fig. 2c. A predominant pattern of higher accessibility in D0 compared to D14 and D28 in all clusters was observed (Fig. S4a). Interestingly, we found temporal chromatin accessibility patterns that were similar to transcriptomic profiles. For example, genes in cluster 1 depict a downregulation from D0 to D14 and maintain their lower expression at D28. Genes in cluster 6 display an upregulation from D14 to D28, and a similar pattern is observed in chromatin accessibility. We also compared the chromatin accessibility in promoter regions of DE protein-coding genes and lncRNAs in different clusters. Results show that chromatin accessibility at protein-coding promoter regions is higher at the transcript start site (TSS), whereas slightly shifted at lncRNA



TSS (Fig. S4b), potentially due to inaccuracies in lncRNA annotation²⁴. In contrast, 500 random non-expressed genes lack signals at TSS (Fig. S4c).

Analyzing individual loci, we observed changes in both chromatin accessibility and expression levels of expected genes at D0 and D28. We complemented our analysis by using public histone modification and DNase I hypersensitivity datasets. Pluripotency genes like *PRDM14* (Fig. 5b) and *NANOG* (not shown) showed a simultaneous significant downregulation of RNA expression and a decrease in OCRs at their promoters. Both loci were also marked by an enrichment of promoter histone mark H3K4me3, the activating histone modification

◄ Figure 3. Most significant DEGs found in dopaminergic induction are lncRNAs. (a) Number of protein-coding genes and lncRNAs found within the top 10, 20, 30, 40, and 50 DEGs in each pairwise time-point comparison. Protein-coding genes and lncRNAs are depicted in blue and red, respectively with different shades for upregulated (up) or downregulated (down). An asterisk indicates that the *p*-value corresponding to the number of DE lncRNAs was lower than the *p*-value of DE protein-coding genes, indicating higher statistical significance, according to hypergeometric tests. (b) Distribution of DE lncRNA gene classes: long intergenic RNAs (lincRNA), antisense, sense intronic, to be experimentally confirmed (TEC), processed transcript, and sense overlapping in pairwise comparisons. **p*-value <0.05, ****p*-value <0.001, according to hypergeometric tests. (c) Correlation of expression fold-change (FC) between pairs of DE lncRNAs and their target protein-coding genes. Positive correlations within DE protein-coding and lncRNAs are represented in blue (downregulated) or red (upregulated). Grey circles represent pairs with negative correlation, where pairs of protein-coding and lncRNA move in opposite directions. Filled circles indicate that both lncRNA and protein-coding target gene are DE.</p>

H3K27ac, and DNase I hypersensitivity in H9 undifferentiated hESCs. We also identified a decrease in chromatin accessibility at an upstream region coincident with enhancer histone modification H3K4me1, histone activation modification H3K27ac, and DNase I hypersensitivity. Interestingly, H3K27ac is absent in the same location in hESC-generated NPC. This putative enhancer region is already annotated in GeneHancer database²²; our results suggest that this region is a pluripotency-related enhancer (Fig. 5b). In contrast, *CORIN* and *SLIT2*, floor-plate markers involved in dopaminergic differentiation²⁵ (Fig. 5c,d), depicted a significant increase in gene expression and gained OCRs at both promoter regions and nearby putative enhancer regions, determined by H3K4me1 and H3K4me3 histone modifications in neurons (N) as well as H3K27ac in adult *substantia nigra* (SN). Some of these regions have already been annotated in GeneHancer database.

Differentially accessible open chromatin regions of D28 coincide with PD-related SNPs. Previous studies have demonstrated that certain disease-associated genetic variants are found within non-coding genomic regions^{26,27}. Hence, we compared both consensus and time-point specific OCRs in our experiments with PD-associated single nucleotide polymorphisms (SNPs) derived from GWAS analysis ²⁸. Interestingly, significant overlaps were found between PD-related SNPs²⁸ and OCRs in consensus peaks and D0 specific peaks only in promoters. Whereas, OCRs specific to mDA (D28) depicted significant overlap with PD-related SNPs only in intronic and intergenic regions. The significancy of overlaps was determined using a hypergeometric statistical test (Supplementary Table S7).

We also assessed the *SNCA* locus, known to be associated with PD-related SNPs²⁹, for enrichment in OCR in mDA (D28). SNCA expression and accessibility in OCR overlapping TSS, presumably at the promoter, were invariant through differentiation. In contrast, in the same locus there was an increase in chromatin accessibility at intergenic and intronic OCR by D14 and a higher one at D28 (Fig. S5). Some of these OCR are overlapping with enhancer histone modifications in ESC-derived NPC and N, and importantly, in adult SN. Moreover, the OCR regions identified in D28 in SNCA locus are also annotated in GeneHancer database and appear to be enriched in PD SNPs.

DEGs are enriched in TF motifs. From the 3377 DE protein-coding genes, 348 (10.3%) are TFs. We searched for TF binding motifs in promoter regions $(\pm 1 \text{ kb up} - \text{ and downstream})$ of protein-coding and lncRNA DEGs. For motif search, we used FIMO (FDR < 0.1)³⁰ with position weight matrices obtained from ENCODE³¹. Resulting motifs were further filtered to include only those belonging to differentially expressed TFs. We found 65 TFs with significant binding motifs in promoter regions of DE genes, including 48 TFs in protein-coding genes and 49 TFs in lncRNAs. The list of TF motifs found and their metrics are included in Supplementary Table S8. A total of 32 TFs had binding sites in both gene types, whereas motifs of 16 and 17 TFs were found exclusively in protein-coding genes and lncRNAs, respectively. Interestingly, POU5F1, a well-known pluripotency marker depicted binding sites only in protein-coding genes whereas PAX6, a TF involved in the development of neural tissues was specific to lncRNAs. Selected protein-coding and lncRNA-specific TFs with binding motifs are listed in Fig. 6a.

To compare the frequency of TF binding motifs in promoter regions of DE genes from the previously defined expression clusters, we counted the number of genes enriched by each TF binding motif per cluster. This process resulted in a matrix of TFs (rows) and clusters (columns) whose values represent the percentage of genes in each cluster which had a TF binding motif. After filtering out motifs of TFs which were not expressed (FPKM < 1) and TFs whose frequencies remained constant (median absolute deviation < 0.05) in all clusters, we obtained a list of 24 TFs (Fig. 6b, Supplementary Table S8). We observed that TFs such JUN, KLF4, E2F1, and CHD2 have highly occurring motifs in DEGs from all clusters, suggesting global transcriptional regulation. However, there are TFs such as ZBTB7A, EGR3 and NFKB1 with high motif enrichment in specific clusters. In another example, LMO2, and NANOG are TFs with fair enrichment in DEGs of one cluster only. Interestingly, NANOG is upregulated in D0 and depicts binding motifs in DEGs from cluster 3 which have expression upregulation in D0 and D14. Even though not all TFs are DE they might regulate transcription indirectly by interacting with other factors. These results suggest that sets of TFs potentially regulate the coordinated transcriptional changes of genes in clusters while others may have regulatory mechanisms in numerous genes at different time points.

Motif enrichment analysis in open chromatin regions highlights a preferential location of specific TFs in promoters. To investigate the potential binding of TFs in OCR, we performed a motif search in consensus peak regions. We filtered out motifs with FDR>0.1. Next, for each TF motif, we calculated its fre-



quency in peaks defined as promoter-TSS and enhancer regions (intergenic, intron, exon, 3' UTR, and 5' UTR). Based on ENCODE's motif matrix, TFs may have more than one motif, thus we selected the most enriched motif representative of a TF. This strategy is intended to identify potentially important regulators of dopaminergic differentiation. We obtained a list of 171 TFs, which were further filtered to contain only TFs in the list of 4009 previously defined DE genes. We found 23 and 18 DE TFs with a higher frequency of binding motifs in promoter or enhancer peaks, respectively (Fig. 6c). The mean frequency of motifs in promoter regions was higher ($6.4 \pm 13\%$; range: 0-46%) than in enhancer regions ($1.5 \pm 13\%$; range: 0-11.8%). From the TFs with the highest frequency

◄Figure 4. Analysis of chromatin accessibility during dopaminergic induction using ATAC-Seq. (a) Fragment size distribution of a representative ATAC-Seq peak sample (D0) provides genome-wide information on chromatin compaction. Fragments smaller than 100 bp represent sequence reads in open chromatin. Reads that span one nucleosome (mononucleosomes) are represented with the peak at 200 bp. Regions of more compact chromatin are indicated by peaks at 440 bps (dinucleosomes) and 600 bps (trinucleosomes). (b) Venn diagram representing the number of ATAC-Seq consensus peaks common and unique to open chromatin regions of each time-point. (c) Number of consensus ATAC-seq peaks depending on their genomic location. (d) Chromatin accessibility at consensus peaks or open chromatin region (OCR) in samples from different induction time-points. Each row represents one consensus peak. Color represents the intensity of normalized chromatin accessibility. Peaks are grouped using k-means clustering. Chromatin accessibility is assessed at regions centered at OCR ± 5 kb. (e,f) Gene-set enrichment analysis using OCR regions differentially accessible at D0 and D28, respectively. Only OCR regions annotated as promoter-TSS, exon, or intron were used. Selected highly enriched (FDR < 0.05) gene-sets are shown. Black bars depict gene-set enrichment as −log10(FDR) and yellow scale represents number of genes.</p>

in promoters, 6 were upregulated in D0, 5 in D14, and 12 in D28. Similarly, from the DE TFs with the highest frequency in enhancers, 6 were upregulated in D0, 5 in D14, and 7 in D28. ESRRA and EBF1, upregulated in D0 and D28, respectively were TFs with the highest frequencies in both promoter and enhancer peaks. Even though CTCF is marginally differentially expressed (FC = -1.65, 1.38, and -1.16 in D0/D14, D14/D28, and D0/D28, respectively), we added it into the plot because it is highly represented at D0. Supplementary Table S9 lists TFs and their frequencies in promoter and enhancer regions.

Gene expression correlation between selected TFs and DEGs. The top 8 TFs with motif frequencies annotated to promoter regions were ESRRA, CTCF, EBF1, POU2F2, BHLHE40, HEY1, SP8, and HIC1. Motifs of ESRRA, CTCF, EBF1, and POU2F2 were also the most enriched in enhancer regions. We calculated the gene expression Pearson correlation of these TFs with all 4009 DEGs considering significant correlations at FDR <0.05. ESRRA, whose expression decreases from D0 to D28, is positively (18%) and negatively (23%) correlated with DEGs, (Fig. 6d). Conversely, EBF1 increases from D0 to D28, and is positively correlated with 33% of DEGs and negatively with 31% (Fig. 6d). Other significant TFs are included in Fig. S6.

CTCF displays a variable expression during mDA induction. We identified changes in the expression of important proteins controlling chromatin structure and nuclear architecture (Fig. 7a). The highest FC (9.8) was found in ACTL6B, which codes for BAF53B, a subunit of the BAF remodeling complex. Notably, most of the genes associated with 3D chromatin structure show a consistent downregulation from D0 to D14 and an upregulation from D14 to D28. To further analyze chromatin states, we selected CTCF, a major component of the nuclear organization machinery that shows a dynamic expression pattern during DA induction (normalized RNA-Seq counts, Fig. 7b). CTCF protein levels are significantly decreased from D0 to D14, increasing again at D28 (Fig. 7c). A similar CTCF abundance pattern was observed through immunostaining, with a significant decrease between D0 and D14 (Fig. 7d). We then used confocal microscopy to detect the coincidence of CTCF with heterochromatin regions identified by high Hoechst staining (Fig. S7). CTCF is excluded from dense Hoechst regions at D0, and this spatial co-localization is significantly increased at D14 and D28 (Fig. S7). A further analysis was performed to determine the co-localization of CTCF with the Polycomb-repressed chromatin mark H3K27me3, by confocal microscopy. The co-localization of CTCF with this epigenetic modification, which identifies facultative heterochromatin, was assessed by image analysis (Fig. 7e). The correlation coefficient was maximal at D0, significantly decreasing at D14, and more notably at D28 (Fig. 7f). Three-dimensional reconstruction of nuclei of cells labeled with CTCF and H3K27me3 antibodies are shown for D0 (Video S1), D14 (Video S2) and D28 (Video S3).

We noted a variable expression of CTCF isoforms (aliases and transcript IDs in Supplementary Table S10) during DA induction (Fig. 7h). Isoform A has an expression pattern similar to the average of transcripts (genelevel) and protein, depicting a transient decrease in D14 and the highest expression in D28. Isoform B decreases expression with no detectable levels at D28. In contrast, isoforms C and D are upregulated in D14. Isoform E is only expressed in D0 (Fig. 7g, Supplementary Table S10). CTCF promoter shows similar levels of chromatin accessibility in our ATAC-Seq samples, to those of H3K27ac reported in pluripotent H9, but also to ESC-derived NPC and N. The latter suggests that CTCF has ubiquitous activity. However, we detected two intronic OCRs within CTCF, previously annotated by GeneHancer. Both regions are markedly accessible at D0, the stage in which CTCF showed the highest transcript and protein levels (Fig. 7b,c), but only one of them overlapped with the H3K27ac mark in H9 hESCs. The region that lacks visible levels of H3K27ac in ESC-derived N. D14, the stage with the lowest CTCF expression, presented only one OCR within the promoter region.

Discussion

Differentiation of pluripotent human cells to mDA involves morphological changes but also modifications in chromatin structure and RNA expression. Here we adopted a protocol mimicking mDA development¹² with conditions inducing floor plate precursors prior to dopaminergic neuron specification. We observed the expected downregulation of pluripotent-associated transcripts and the induction of *FOXA2*, *LMX1A*, *EN1*, *EN2*, *DDC*, *NR4A2*, *NEUROG2*, *NEUROD1*, *ASCL1*, *NKX6-1*, *LMX1B*, and *TH*. Gene-set enrichment analyses were consistent with these processes.



Our characterization of transcriptional changes allowed us to generate a repertoire of DE lncRNAs with potential regulatory functions in pluripotency and mDA commitment. Unexpectedly, there was a high proportion of DE lncRNAs. The majority of DE lncRNAs found were lincRNAs, or antisense relative to the closest protein-coding gene. Notably, most were lincRNAs showing downregulation. Loss-of-function assays in mouse ESCs showed that lincRNAs have key roles in maintaining pluripotency and repressing differentiation programs³². Furthermore, lincRNAs identified in ESCs primarily regulate gene expression in *trans*, binding to pluripotency-related TFs, or interacting with chromatin regulatory proteins^{32–34}. We found 54 pairs of DE lncRNAs with known

Figure 5. Temporal patterns of chromatin accessibility are correlated to transcriptomic profiles. (a) Correlation between gene expression and chromatin accessibility. Consensus peaks annotated as promoter-TSS, exon, and intron were correlated with gene expression. Log2 transformed FC of differential peak accessibility and log2 FC of normalized gene expression counts in the comparison between D0 and D28 were used. (b) ATAC-Seq derived chromatin accessibility tracks and RNA-Seq expression changes at the PRMD14 locus during dopaminergic induction. ENCODE tracks of H9 pluripotent hESC DNAse I, H3K4me3, H3K4me1 and H3K27ac, H1 hESC CTCF, and ESC-derived neural precursors (NPC) H3K27ac are included for comparison. Known and predicted enhancer regions obtained from GeneHancer database are shown. Orange shadow highlights PRMD14 promoter region and blue region indicates a putative enhancer. Tracks data ranges are shown in the right. (c,d) Chromatin accessibility and gene expression tracks within CORIN and SLIT2 loci at different dopaminergic induction time-points. ENCODE tracks of ESC-derived neurons (N) H3K27me3, H3K9me3, H3K4me3 and H3K4me1, as well as substantia nigra (SN) H3K27ac are included for comparison. Predicted and known enhancer regions derived from GeneHancer database are depicted. Orange zones indicate CORIN and SLIT2 promoter regions. Blue shades highlight putative enhancers. Track data ranges are shown in the right. ATACseq tracks labelled by asterisks depict statistical significance obtained by Mann Whitney tests (*p-value < 0.05, **p-value < 0.01, ***p-value < 0.001) using the number of normalized reads in open chromatin regions of gene locus and promoter regions.

regulatory targets which were significantly correlated in all time points. Interestingly, 91% (49 out of 54) showed a positive correlation, suggesting that these lncRNAs participate in positive or negative feedback loops. However, their roles in dopaminergic differentiation remain elusive. LncRNAs such as linc-ROR (lincRNA-ST8SIA3) can reprogram somatic cells to pluripotency³⁵. Further studies with simultaneous analysis of TFs, lncRNAs, miRNAs, and epigenetic marks are required to better understand regulatory networks controlling the transitioning from pluripotent stem cells to NPC and mDA.

Previous studies profiled gene expression by microarrays with the same differentiation protocol¹², and also a comprehensive approach was performed in human DA development with single-cell resolution³⁶. However, gene expression is not sufficient for elucidating regulators of cell identity. The integration of ATAC-Seq and the greater resolution of RNA-Seq is a powerful resource to address regulatory heterogeneity between different cellular populations⁶. We found higher levels of global chromatin accessibility in D0 when compared with D14 and D28. These results confirm that ESC have a more open or "loose" chromatin correlating with a more permissive transcriptional state⁷. Differentiation is associated with heterochromatinization and global reorganization of heterochromatin structure, density and position⁸. A clear relationship between chromatin accessibility and gene expression has been demonstrated³⁷. We observed a moderate correlation when comparing these parameters, suggesting the participation of other elements in the regulation of gene transcription such as TF binding, DNA methylation, histone tail modifications, miRNAs, and lncRNAs^{38,39}.

We identified known and novel TFs potentially orchestrating neurogenesis and dopaminergic specification. We found a higher density of TF motifs in promoter regions compared to enhancer regions, consistent with previous data⁴⁰. By studying DA differentiation in mice, researchers concluded that expression of lncRNAs and accessibility of enhancer regions represent a more accurate molecular signature associated with cell specificity than the expression of protein-coding genes⁴¹. We found a higher overlap of ATAC-Seq D0-specific OCRs with GeneHancer regulatory elements than with D14 and D28 OCRs suggesting that cell type-specific enhancer regions are potentially involved in dopaminergic differentiation. D14 and D28 OCRs not annotated by GeneHancer are interesting candidates for experimental validation.

Overall, 6 out of 8 DE TFs, whose motifs were most frequently found in OCR, have expression patterns significantly correlated with a considerable number of DE genes. Interestingly, these TFs have diverse profiles with their highest expression in D0, D14, or D28, denoting their roles in pluripotency, neural precursor commitment, or mDA specification. ESRRA and SP8, upregulated in D0, might participate in pluripotency maintenance. ESRRA is a nuclear receptor, but its participation in self-renewal and pluripotency has not been elucidated yet; however, another family member, ESRRB is a crucial regulator of pluripotency in mouse ESCs⁴² and contributes to reprogramming fibroblasts to pluripotency⁴³. Recent research revealed a pioneering role of ESRRB in binding silenced enhancers to promote a permissive chromatin state for pluripotent core TF recruitment⁴⁴. Expression profiling studies of nuclear receptors in mouse and human ESCs showed that ESRRB is unique to mice. In contrast, ESRRA is present in both mice and humans⁴⁵, suggesting that ESRRA might subserve biological functions in hESCs similar to those of murine ESRRB.

Motifs of ESRRA, CTCF, and EBF1 were enriched in enhancer regions, suggesting their role in distal regulatory mechanisms. Some of the TFs identified in this study have not been previously implicated in dopaminergic differentiation, but they have a role in neurogenesis. Early B cell factor (EBF1), highly expressed in D28, has been shown to bind to the promoter⁴⁶ and enhancer regions⁴⁷ and may interact with p300/CBP regulating gene transcription⁴⁸. EBF1 is expressed in post-mitotic neurons and has an important role in mouse striatal differentiation⁴⁹. Given that computationally predicted motif binding sites do not always indicate physiological binding of TFs⁵⁰, additional ChIP-seq experiments are needed to address the in vivo validation of TF binding in progenitors and mDA⁵¹.

Genome-wide association studies (GWAS) have identified associations between SNPs and human pathologies. The majority of identified SNPs lie within non-coding regions including lncRNAs and enhancer regions^{26,27}, suggesting that these SNPs mark variations that might impact regulatory functions increasing the risk for PD. Concordantly, we observed that OCRs in differentiated DA neurons overlap significantly with SNPs in noncoding regions. Specifically differentiated DA neurons show that OCRs in *SCNA* overlap with reported PD-related SNPs. Further studies are required to pinpoint the functional effect of enhancer region SNPs with TF binding.



We observed marginal changes in gene expression of several subunits of the chromatin remodeling complex BAF, except for the *ACTL6B* transcript that codes for BAF53B subunit and is significantly upregulated from D14 to D28. Interestingly, cell cycle exit in NPCs is accompanied by a subunit exchange from BAF53A to BAF53B⁵², consistent with our results. To elucidate molecular triggers driving chromatin accessibility changes, we queried expression profiles of diverse chromatin remodelers and selected the architectural protein CTCF, given its importance in brain development⁵³. CTCF is ubiquitously expressed in numerous cell types and developmental stages and has been implicated in regulating widespread functions⁵⁴.

Figure 6. Motif enrichment analysis in DEG promoter regions and open chromatin regions reveals a set of potential TFs involved in dopaminergic differentiation. (a) Motifs of TFs found only in lncRNA or in protein-coding gene promoter regions. (b) Hierarchical clustering of the frequency of TF binding motifs in promoters of genes from the 6 expression clusters (K) depicted in Fig. 2. White to purple colors represent the density of TF binding in promoter regions of DEGs assigned to expression clusters. The bottom panel highlights the average regulation of genes in each cluster in specific time points. On the right side the RNA expression is shown for D0, D14 and D28. DEGs are labeled with red. (c) Enrichment of TF binding motifs in consensus peaks classified as promoters or enhancers. Only motifs of DE TFs are depicted, except for CTCF, which is marginally differentially expressed. Colors indicate the time-point of highest TF gene expression correlation in all induction time-points between selected highly enriched TFs (ESRRA and EBF1) and DEGs. Dotted red lines represent FDR < 0.05 of Pearson correlation coefficient. The inset blue plots indicate the corresponding TF gene expression at D0, D14 and D28.</p>

The leading hypothesis denoting CTCF's diverse regulatory functions is by conforming chromatin loops⁵⁵. Notably, we observed a transient decrease at both protein and mRNA levels of CTCF in samples from D14. Beagan et al. found a decreased genome-wide CTCF occupancy in the transition from mouse ES cells to NPCs⁵³. They also showed down-regulation in the protein levels of CTCF when comparing mouse pluripotent cells with NPCs, but they did not analyze differentiated neurons; we observed an increase in CTCF when human NPCs progress to differentiated neurons. Heterochromatin regions can be identified by intense Hoechst staining; however, a more specific epigenetic mark for Polycomb-mediated facultative heterochromatin is H3K27me3⁵⁶. Through further assays, we found the lowest co-localization of CTCF with intense Hoechst staining at D0, increasing at D14 and D28. However, the co-localization of CTCF with H3K27me3 in pluripotent cells is highest. At D14, there was a discrete, although significant decrease in this coincidence. In contrast, there was a significant dissociation of CTCF presence with H3K27me3 deposition at D28. Previous work showed that the majority of CTCF sites present in ESC-derived NPC are already present in pluripotent ESC, suggesting that neural commitment is accompanied by a massive loss of CTCF occupancy instead of an extensive CTCF site acquisition or reshuffling⁵³. Reduced CTCF presence in heterochromatin at D0 may result from the heterochromatin-free nuclei of ESCs. Increased co-localization of CTCF with heterochromatin might be due to the extensive heterochromatinization that takes place through differentiation⁸ or to CTCF acting as an insulator to prevent excessive repression. Further experimental assays will unveil the role of CTCF in chromatin looping and its changes during DA specification. Our results indicate that CTCF relocates to euchromatin regions in differentiated neurons.

Conclusion

In summary, we generated genome-wide maps of transcriptional changes and chromatin accessibility and through their integration, we derived a list of TFs and lncRNAs with potential functions in pluripotency maintenance or DA differentiation. We pinpointed important chromatin remodelers and demonstrated the dynamics of CTCF protein and mRNA. Importantly, we highlight the coordinated functions of chromatin accessibility, chromatin remodelers, TFs and lncRNAs to regulate spatiotemporal gene expression. Our work provides a useful resource of TFs and regulatory elements, potentially orchestrating human mDA specification.

Methods

For detailed protocols, refer to Supplementary Information.

Dopaminergic differentiation of hESCs. H9 human embryonic stem cells (WA09) of female origin were purchased from WiCell (Madison, WI, USA), transduced with lentiviral vectors to express GFP as described¹⁸ and cultured until 75% confluency. These GFP-expressing cells are useful for transplantation studies¹³ and behave as its parental H9 hESC. At day zero (D0), cells were considered pluripotent. We followed a dopaminergic induction protocol¹² with minor modifications⁵⁷. At D14, neural precursors are present, whereas at D28, cells presented a neuronal morphology. Samples were analyzed at D0, D14 and D28, some plates were used for RNA extraction while others for nuclei isolation and chromatin assays.

Immunostaining. Undifferentiated H9-GFP cells were cultured for 2–3 days. Neural progenitors were reseeded on induction day 13 or 21 on 24-well plates and fixed the next day. Cells were permeabilized, washed, blocked, and incubated with primary and secondary antibodies. Incubation with Hoechst 33258 (1 ng/ml) was used for nuclear labeling.

Immunoblot. Standard immunoblot protocols were performed for CTCF and TH. Images were analyzed using ImageJ-Fiji⁵⁸.

Immunofluorescence analysis of CTCF protein levels. Photographs of CTCF- and Hoechst-stained cells, or confocal images were normalized, and analyzed with ImageJ. Co-localization of CTCF with heterochromatin regions stained with Hoechst was determined by the overlapping area from merged images as reported⁵⁶. Also, the co-localization of CTCF with H3K27me3 was determined as previously described⁵⁹.



ATAC and RNA sequencing. The assay for transposase-accessible chromatin (ATAC) libraries were prepared on D0, D14, and D28, following the original method⁶⁰. After cell lysis, tagmentation, and adapter incorporation, transposed DNA was purified and quality was assessed. Paired-end 75 cycle sequencing reads were acquired on the Illumina HiSeq 2500 sequencer. A total of 50, 25, and 40 million reads were obtained from D0, D14, and D28 samples respectively.

Total RNA was extracted using TRIzol following manufacturer's instructions. Samples with RNA integrity number greater than 8 were used. For library construction, Illumina TruSeq RNA Sample Prep Kit (Illumina)

Figure 7. Differential expression of genes that are important for the nuclear landscape and changes in the architectural protein CTCF. (a) FC expression of genes involved in chromatin remodeling, histone modifications, and 3D configuration from D0 to D14 and D14 to D28. The dashed vertical lines represent a FC of 1.5. (b) Quantification of normalized counts of CTCF mRNA levels determined by RNA-Seq during dopaminergic differentiation. (c) Representative experiment of CTCF (~130 kDa) detection by immunoblot, during dopaminergic induction. β-ACTIN was used as a loading control. On the right side, CTCF quantification at D0, D14 and D28 from 3 independent experiments is shown. * p < 0.05. (d) Immunostaining of nuclear CTCF during dopaminergic differentiation and its quantification in 5 independent experiments. *p < 0.05. Scale bars, 100 µm. (e) Confocal microscope fluorescence images showing nuclear co-localization of CTCF and H3K27me3 through dopaminergic differentiation. (f) Correlation coefficient for the co-localization of CTCF with H3K27me3 during differentiation, quantified from confocal images in 30 nuclei per time point. * p < 0.05; *** p < 0.001. (g) Expression of the 5 CTCF isoforms at D0, D14 and D28. (h) ATAC-Seq chromatin accessibility and RNA-Seq tracks within CTCF locus in comparison with ENCODE tracks of H1 CTCF, H9 H3K4me1, H9 H3K4me3, H3K27ac, ESC-derived NPC H3K27ac and ESC-derived N H3K4me1. GeneHancer predicted and known enhancers are included. Five known CTCF isoforms detected by RNA-Seq are also shown. Orange shadow indicates CTCF promoter region. Blue shades indicate putative enhancers. Track data ranges are shown in the right.

were used. Samples were run on MiSeq and HiSeq 2500 Illumina sequencers using paired-end 75 cycle mode. A total of 80.4 million reads were obtained and distributed among 3, 4, and 3 replicates from D0, D14, and D28 time points, respectively.

ATAC-Seq pre-processing. After removing adaptors using cutadapt⁶¹ and trimmomatic⁶² we obtained 36–126 bp paired-end ATAC-Seq reads for each time-point samples. The quality of trimmed sequencing datasets was verified using FASTQC⁶³. We followed the ENCODE ATAC-Seq pipeline (http://www.encodeproject. org/atac-seq/) to process samples. Paired-end reads were aligned to the human reference genome (GRCh38/hg38) using Bowtie2 with parameters '-X 2000 -k 5'⁶⁴. Next, we used SAMTools⁶⁵ to remove unmapped reads, fragments with unmapped mates, non-primary alignment reads, and reads failing platform quality checks. Low quality reads (MAPQ < 30) and reads which mapped to mitochondrial DNA were also excluded. Optical and PCR duplicates were also removed using Picard tools (http://broadinstitute.github.io/picard). To accurately locate the center of each transposon-binding event, the remaining reads were offset by +4 and –5 for positive and negative strands respectively⁶⁶. To evaluate the expected periodicity of DNA winding around nucleosomes, we obtained insert size histograms using Picard tools (Fig. 4a). ATAC-Seq peak regions were called for each sample using MACS2 with parameters -shift 75 –extsize 150 –nomodel –keep-dup all –call-summits⁶⁷. Resulting peaks which overlapped with ENCODE blacklisted regions were filtered out⁶⁸.

Analysis of differential chromatin accessibility. A consensus set of 464,783 unique peaks was generated merging peaks within 100 bp. Exclusive peaks were obtained using BEDTools⁶⁹. Reads in peaks were counted using HTSeq⁷⁰ and tested for differential accessibility using DESeq2⁷¹ considering $|log_2(fold-change)| > 1.5$ and *p*-value < 0.05. Promoter-TSS was defined as 1 kb up- plus 100 bp downstream of the transcript start site (TSS). HOMER annotatePeaks function was used⁷², to associate peaks to genes (GENCODE v29) or genomic regions (intergenic, promoter-TSS, exon, intron, 5' UTR, 3' UTR, and TTS). We used Mann Whitney tests to compare the number of normalized reads found in open chromatin regions (OCR) of specific loci between ATAC-Seq samples (D0 vs D14 and D0 vs D28). Asterisks indicate statistical significance (*p*-value < 0.05) in IGV tracks (Fig. 5b–d).

Gene-set enrichment analysis. The hypergeometric statistical test in R (https://cran.r-project.org/) and a subset of gene-sets from MSigDB⁷³ combined with Wikipathways 2019^{74,75} were used. Gene-sets were considered significantly enriched with FDR < 0.05.

Public datasets. For comparison, we used public datasets for CTCF ChIP-seq, DNase I hypersensitivity, H3K27ac, H3K27me3, H3K4me1, H3Kme3 for H1, H9, NPC, N, LUHMES, and substantia nigra cells (SN) from Gene Expression Omnibus accession numbers indicated in Supplemental information. We used GeneHancer database from GeneCards Suite v4.14²² for comparison of enhancer regions and predicted target genes.

Transcription factor binding analysis. We performed a motif search using position weight matrices from the ENCODE TF ChIP-seq datasets³¹. Promoter regions of differentially expressed (DE) genes (DEGs) (1 kb up- and downstream of the TSS) and open chromatin regions (OCRs, common and differentially accessible) were scanned for motif occurrences using FIMO³⁰ with FDR<0.1. The most frequently occurring DE TFs were correlated with DEGs using the Pearson coefficient. Only correlations with FDR<0.05 were considered significant.

Gene expression analysis. Standard quality procedures were performed using cutadapt⁶¹, trimmomatic⁶², and FASTQC⁶³. Reads were mapped to hg38 (https://www.gencodegenes.org/) using a previously described pipeline⁷⁶. Assembly of mapped reads was performed with Cufflinks v2.2.1⁷⁷ and values of Fragments Per Kilobase of transcript per million Mapped reads (FPKM) were calculated for all annotated genes and transcripts (GENCODE v29, https://www.gencodegenes.org/). We used HTSeq⁷⁰ to calculate read counts for annotated

genes and transcripts. We adjusted batch effect (Fig. S2) and performed pairwise comparison of read counts implementing $DESeq2^{71}$. Genes were labelled as differentially expressed (DEG) if at least one of the replicates in the comparison had FPKM \geq 1, and normalized count FC>4 with an FDR<0.05. Temporal gene expression profiles of DEGs were obtained using hierarchical clustering. DEGs in each cluster were used for gene-set enrichment analysis.

Data availability

Raw sequencing datasets and processed files have been deposited in NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under accession GSE153005.

Received: 29 December 2020; Accepted: 4 August 2021 Published online: 20 August 2021

References

- 1. Berger, S. L. The complex language of chromatin regulation during transcription. Nature 447, 407–412 (2007).
- 2. Ong, C. & Corces, V. G. Interactions: A common evolutionary theme. J. Biol. 2, 8-11 (2009).
- 3. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. Cell Rep. 15, 2038–2049 (2016).
- 4. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. Cell 169, 693-707 (2017).
- Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell 169, 930–944 (2017).
- Buenrostro, J. D. et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 173, 1535–1548 (2018).
- 7. Efroni, S. et al. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2, 437-447 (2008).
- Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell 10, 105–116 (2006).
- 9. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317-329 (2015).
- 10. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489, 75-82 (2012).
- 11. Axelsen, T. M. & Woldbye, D. P. D. Gene therapy for Parkinson's disease, an update. J. Parkinsons. Dis. 8, 195–215 (2018).
- 12. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480**, 547–551 (2011).
- López-Ornelas, A. et al. Human embryonic stem cells-derived dopaminergic neurons transplanted in parkinsonian monkeys recover dopamine levels and motor behavior. *BioRxiv* https://doi.org/10.1101/2020.07.08.192591 (2020).
- Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012 (2019).
- 15. Forrest, M. P. et al. Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. Cell Stem Cell 21, 305–318 (2017).
- Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science 366, 1134–1139 (2019).
- Hatano, T., Kubo, S. I., Sato, S. & Hattori, N. Pathogenesis of familial Parkinson's disease: New insights based on monogenic forms of Parkinson's disease. J. Neurochem. 111, 1075–1093 (2009).
- Domingo-Reines, J. et al. Hoxa9 and EGFP reporter expression in human Embryonic Stem Cells (hESC) as useful tools for studying human development. Stem Cell Res. 25, 286–290 (2017).
- 19. ENCODE. 6 Non-coding RNA characterization. Nature https://doi.org/10.1038/nature28175 (2019).
- Liu, Q. et al. Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. Circ. Res. 121, 376–391 (2017).
- 21. Gaspar-Maia, A. et al. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460, 863-868 (2009).
- 22. Fishilevich, S. et al. GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. Database 2017, 28 (2017).
- de la Torre-Ubieta, L. et al. The dynamic landscape of open chromatin during human cortical neurogenesis. Cell 172, 289-304.e18 (2018).
- 24. Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. & Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **19**, 535–548 (2018).
- Yan, C. H., Levesque, M., Claxton, S., Johnson, R. L. & Ang, S. L. Lmx1a and Lmx1b function cooperatively to regulate proliferation, specification, and differentiation of midbrain dopaminergic progenitors. J. Neurosci. 31, 12413–12425 (2011).
- Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U. S. A. 106, 9362–9367 (2009).
- Ward, L. D. & Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. Nat. Biotechnol. 30, 1095–1106 (2012).
- Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. Lancet Neurol. 18, 1091–1102 (2019).
- Pierce, S. E., Tyson, T., Booms, A., Prahl, J. & Coetzee, G. A. Parkinson's disease genetic risk in a midbrain neuronal cell line. *Neurobiol. Dis.* 114, 53-64 (2018).
- 30. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018 (2011).
- Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res. 42, 2976–2987 (2014).
- 32. Guttman, M. et al. LincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300 (2011).
- Nagano, T. et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322, 1717– 1720 (2008).
- 34. Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. 106, 11667–11672 (2009).
- 35. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117 (2010).
- 36. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. Cell 167, 566-580 (2016).
- Frank, C. L. et al. Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. Nat. Neurosci. 18, 647–656 (2015).
- Heintzman, N. D. & Ren, B. The gateway to transcription: Identifying, characterizing and understanding promoters in the eukaryotic genome. Cell. Mol. Life Sci. 64, 386–400 (2007).
- Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E. & Ohler, U. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* 22, 1711–1722 (2012).

- Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat. Genet. 49, 1073–1081 (2017).
- Gendron, J. et al. Long non-coding RNA repertoire and open chromatin regions constitute midbrain dopaminergic neuron—specific molecular signatures. Sci. Rep. 9, 1–16 (2019).
- 42. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* 11, 491–504 (2012).
- Feng, B. *et al.* Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat. Cell Biol.* 11, 197–203 (2009).
- 44. Adachi, K. et al. Esrrb unlocks silenced enhancers for reprogramming to naive pluripotency. Cell Stem Cell 23, 900-904 (2018).
- 45. Xie, C. Q. *et al.* Expression profiling of nuclear receptors in human and mouse embryonic stem cells. *Mol. Endocrinol.* 23, 724–733 (2009).
- Treiber, T. *et al.* Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity* 32, 714–725 (2010).
- Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* 11, 635–643 (2010).
- Zhao, F., McCarrick-Walmsley, R., Åkerblad, P., Sigvardsson, M. & Kadesch, T. Inhibition of p300/CBP by early B-Cell factor. *Mol. Cell. Biol.* 23, 3837–3846 (2003).
- Garel, S., Marín, F., Grosschedl, R. & Charnay, P. Ebf1 controls early cell differentiation in the embryonic striatum. *Development* 126, 5285–5294 (1999).
- Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. Nat. Rev. Genet. 5, 276–287 (2004).
- Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. Science 316, 1497–1502 (2007).
- 52. Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* 460, 642–646 (2009).
- 53. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* 27, 1139–1152 (2017).
- Ong, C. T. & Corces, V. G. CTCF: An architectural protein bridging genome topology and function. Nat. Rev. Genet. 15, 234–246 (2014).
- 55. Phillips, J. E. & Corces, V. G. C. T. C. F. Master weaver of the genome. Cell 137, 1194-1211 (2009).
- Linhoff, M. W., Garg, S. K. & Mandel, G. A high-resolution imaging approach to investigate chromatin architecture in complex tissues. Cell 163, 246–255 (2015).
- 57. Carballo-Molina, O. A. et al. Semaphorin 3C released from a biocompatible hydrogel guides and promotes axonal growth of rodent and human dopaminergic neurons. Tissue Eng. Part A 22, 850–861 (2016).
- 58. Schindelin, J. et al. Fiji: An open-source platform for biological-image analysis. Nat. Methods 9, 676-682 (2012).
- Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. Am. J. Physiol. Cell Physiol. 300, C723–C742 (2011).
- Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 1–9 (2015).
- 61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10-12 (2011).
- 62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- 63. Andrews, S. A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
- 64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357-359 (2012).
- 65. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218 (2013).
- 67. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
- Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* 9, 9354 (2019).
- 69. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- 70. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- 71. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014).
- 72. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550 (2005).
- 74. Kuleshov, M. V. et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97 (2016).
- Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* 14, 128 (2013).
 Cuevas-Diaz Duran, R. *et al.* The systematic analysis of coding and long non-coding RNAs in the sub-chronic and chronic stages
- of spinal cord injury. Sci. Rep. 7, 41008 (2017). 77 Trannell C et al Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching
- Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).

Acknowledgements

We are grateful for the technical assistance of Dr. Itzel Escobedo-Avila and B. Sc. Georgina Guerrero. Authors also thank the kind donation of CTCF antibodies from Dr. Félix Recillas-Targa.

Author contributions

V.T. and I.V. initially conceived the project. Experiments were performed by C.M.-R., T.B.-G., A.L.-O., J.H.-G. and E.E. The data was analyzed by C.M.-R., R.C.-D.D. and M.G.-L. Funding was secured by V.T. and I.V. The original draft was prepared by C.M.-R., R.C.-D.D. and M.G-L. Revisions and editing of the manuscript were made by C.M.-R., R.C.-D.D., M.G.-L., E.S.-R., V.T. and I.V. All authors approved the final manuscript.

Funding

CONACyT Ciencia Básica 255747 (V.T.) and 284748 (E.S-R.), CONACyT Salud-2016–272815 (I.V.), UNAM-PAPIIT (IN213719 to I.V.) and CONACyT 300461. CM-R and MG-L received a graduate fellowship from CONA-CyT. TB-G received a CONACyT postdoctoral fellowship.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-021-96263-1.

Correspondence and requests for materials should be addressed to R.C.-D., I.V. or V.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021