



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES
CUAUTITLÁN

Análisis de identidad transcripcional de células troncales
hematopoyéticas humanas
frescas y generadas *in vitro*.

T E S I S

QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN BIOQUÍMICA DIAGNÓSTICA

P R E S E N T A:
Arian Cuauhtémoc Escajeda González

ASESOR:

Dr. Héctor Mayani Viveros

CO-ASESOR:

Q.F.B Alejandro Gutiérrez García

Cuautitlán Izcalli, Estado de México, a 27 abril del 2021.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN

ASUNTO: **VOTO APROBATORIO**

M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE

ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Titulación
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de Tesis**

Análisis de identidad transcripcional de células troncales hematopoyéticas frescas y generadas in vitro

Que presenta el pasante: **Arian Cuauhtémoc Escajeda González**

Con número de cuenta: **415097316** para obtener el Título de: **Licenciado en Bioquímica Diagnóstica.**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO**.

ATENTAMENTE

"POR MI RAZA HABLARÁ EL ESPÍRITU"

Cuautitlán Izcalli, Méx. a 11 de Marzo de 2021.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	M. en C. Idalia Carmen Avila Miyazawa	
VOCAL	M. en C. Gilberto Atilano Amaya Ventura	
SECRETARIO	QFB. Alejandro Gutiérrez García	
1er. SUPLENTE	M.V.Z. Ángel Germán Martínez Sosa	
2do. SUPLENTE	LBD. Larisa Andrea González Salcedo	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

U. N. A. M.
ASUNTO: VOTO APROBATORIO

**M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE**

**ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Titulación
de la FES Cuautitlán.**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de Tesis**

Análisis de identidad transcripcional de células troncales hematopoyéticas frescas y generadas in vitro

Que presenta el pasante: **Arian Cuauhtémoc Escajeda González**
Con número de cuenta: **415097316** para obtener el Título de: **Licenciado en Bioquímica Diagnóstica.**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO.**

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 11 de Marzo de 2021.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	M. en C. Idalia Carmen Avila Miyazawa	
VOCAL	M. en C. Gilberto Atilano Amaya Ventura	
SECRETARIO	QFB. Alejandro Gutiérrez García	
1er. SUPLENTE	M.V.Z. Ángel Germán Martínez Sosa	
2do. SUPLENTE	LBD. Larisa Andrea González Salcedo	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

U. N. A. M.
ASUNTO: VOTO APROBATORIO

**M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE**

**ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Titulación
de la FES Cuautitlán.**

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de Tesis**

Análisis de identidad transcripcional de células troncales hematopoyéticas frescas y generadas in vitro

Que presenta el pasante: **Arian Cuauhtémoc Escajeda González**
Con número de cuenta: **415097316** para obtener el Título de: **Licenciado en Bioquímica Diagnóstica.**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO.**

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 11 de Marzo de 2021.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	M. en C. Idalia Carmen Avila Miyazawa	
VOCAL	M. en C. Gilberto Atilano Amaya Ventura	
SECRETARIO	QFB. Alejandro Gutiérrez García	
1er. SUPLENTE	M.V.Z. Ángel Germán Martínez Sosa	
2do. SUPLENTE	LBD. Larisa Andrea González Salcedo	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

LMCF/javg



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN

ASUNTO: VOTO APROBATORIO

M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE

ATN: LA. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Titulación
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el Trabajo de Tesis

Análisis de Identidad transcripcional de células troncales hematopoyéticas frescas y generadas in vitro

Que presenta el pasante: Arian Cuauhtémoc Escajeda González
Con número de cuenta: 415097316 para obtener el Título de: Licenciado en Bioquímica Diagnóstica.

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el EXAMEN PROFESIONAL correspondiente, otorgamos nuestro VOTO APROBATORIO.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 11 de Marzo de 2021.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	M. en C. Idalia Carmen Avila Miyazawa	
VOCAL	M. en C. Gilberto Atilano Amaya Ventura	
SECRETARIO	QFB. Alejandro Gutiérrez García	
1er. SUPLENTE	M.V.Z. Ángel Germán Martínez Sosa	<i>Ángel Germán Martínez Sosa</i>
2do. SUPLENTE	LBD. Larisa Andrea González Salcedo	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.



**FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
SECRETARÍA GENERAL
DEPARTAMENTO DE TITULACIÓN**

U.N.A.M.
ASUNTO: **VOTO APROBATORIO**

**M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE**

ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Titulación
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: **Trabajo de Tesis**

Análisis de identidad transcripcional de células troncales hematopoyéticas frescas y generadas in vitro

Que presenta el pasante: **Arian Cuauhtémoc Escajeda González**
Con número de cuenta: **415097316** para obtener el Título de: **Licenciado en Bioquímica Diagnóstica.**

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO.**

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cuautitlán Izcalli, Méx. a 11 de Marzo de 2021.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	M. en C. Idalia Carmen Avila Miyazawa	
VOCAL	M. en C. Gilberto Atilano Amaya Ventura	
SECRETARIO	QFB. Alejandro Gutiérrez García	
1er. SUPLENTE	M.V.Z. Ángel Germán Martínez Sosa	
2do. SUPLENTE	LBD. Larisa Andrea González Salcedo	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

Agradecimientos

Quiero expresar mi más profundo agradecimientos al doctor Héctor Mayani. Sin el este trabajo no sería posible. Su guía, experiencia, pensamiento crítico y analítico, fueron fundamentales para dar forma al presente trabajo. Agradezco en especial las sesiones donde me permitía discutir con el uno a uno los temas de la tesis. Creo que su guía fue de gran importancia no solo para este trabajo sino para mi formación en general tanto en el ámbito académico como en el personal. Por esas razones deseo expresar mi profundo agradecimiento a esta destacada personalidad.

Quiero agradecer a mi Abuelo Juan Escajeda, por su apoyo durante la carrera. El fue mi sustento económico y la persona que trato de motivarme día a día para no rendirme a pesar de las adversidades. Siempre, cuando yo requerí apoyo, en los momentos más difíciles él estuvo para mí incondicionalmente.

Agradecer a Ernesto Paas quien me enseñó diligentemente a programar en R estos conocimientos fueron de gran importancia para la culminación del presente trabajo. El se ha convertido en un amigo entrañable no solo relacionado con la ciencia, sino en la parte mas personal de mi ser.

Expreso mi agradecimiento a la doctora Patricia Flores por enseñarme en profundidad las técnicas experimentales relevantes para comprender las implicaciones biológicas de los resultados en el presente trabajo esta experiencia en el laboratorio fue un pilar para comprender el sistema biológico que se presenta en este trabajo.

A los profesores Gilberto Atilano, Ángel Germán y Alejandro Gutiérrez quienes desde que los conozco han implantado en mi un pensamiento crítico científico en la biología y la matemática. Sin ellos este trabajo jamás pudo siquiera haber empezado.

Dedicatorias

Dedico esta tesis en primer lugar a mi madre Juana González Gutiérrez quien me crio para creer en mi mismo, quien implanto en mí el espíritu que representa, no solo esta tesis, sino toda mi licenciatura: utilizar nuestros conocimientos para ayudar a los demás. Es por ella que decidí estudiar esta carrera y es por ella que logré ver la importancia de este tipo de aproximaciones preclínicas. Por ella tengo la moral que hoy en día tengo. Y sin esa moral el conocimiento no es mas que un conjunto de datos sin sentido humano.

A mi hermana quien desde la edad mas temprana compartió conmigo su visión del mundo y su forma de pensar. Eso dejo una profunda marca en mi que perdurara toda mi vida. El ver su impresionante intelecto siempre me motivo, silenciosamente, a no quedarme detrás de ella. Ella es una motivación destacada en mi vida.

A mis hermanos Juan Augusto Escajeda y Abner Emanuel Escajeda, Quienes me dieron su apoyo emocional y de todo tipo durante toda mi vida. Con quienes compartí problemas, discusiones, éxitos y soluciones. Y quienes siempre serán parte de mi motivación para alcanzar nuevas metas. A mi abuelo a quien no puedo dejar de agradecerle todo lo que ha hecho por mí y mi familia. Como el gran patriarca que es siempre nos cobijo bajo su brazo, para darnos la oportunidad de florecer a pesar de la infértil tierra en la que crecimos.

Por último, quiero dedicar esta tesis al doctor Héctor Mayani. Quien ayudo a mi familia en uno de los momentos más difíciles que jamás hayamos vivido. EL me demostró como la ciencia se traduce en ayudar a las personas, en salvar vidas... Por todo su apoyo en el ámbito personal y académico le dedico este trabajo.

Índice.

Índice de figuras y tablas.	1
Glosario de abreviaturas.	1
Resumen.	6
Introducción.	6
Marco teórico.	8
Hematopoyesis.	
Complejidad en sistemas biológicos.	
Ciencias ómicas.	
Bioinformática.	
Métodos de reducción de la dimensionalidad.	
Análisis de enriquecimiento de categorías génicas.	
Tera de redes en sistemas biológicos.	
Control de los sistemas celulares.	
Problema.	
Problema de estudio.	19
Objetivos.	20
Metodología.	20
Métodos experimentales.	
Métodos bioinformáticos.	
Resultados.	22
Discusión.	25
Conclusiones.	28

Índice de figuras

Figura 1. Organización jerárquica del sistema hematopoyético	12
Figura 2. Esquema del funcionamiento de los microarreglos	20
Figura 3. Análisis de enriquecimiento de categorías génicas	28
Figura 4. Análisis de componentes principales de microarreglos	29
Tabla1. Factores de transcripción pertenecientes a los 3 conjuntos génicos con mayor enriquecimiento	23
Figura 5. Redes de interacción proteica	30
Figura 6. Enriquecimientos de complejos de remodelación de la cromatina	31

Glosario de abreviaturas.

DNA: (*Deoxyribonucleic acid*), es el compuesto químico que forma el material genético de todos los seres vivos.

CD: (*Cluster of Differentiation*). Son un grupo de proteínas de membrana que sirven para identificar diferentes tipos células que no son distinguibles a nivel morfológico. Por lo regular se escribe CD seguido de un número y/o letras para cada una de las diferentes proteínas y se identifican por citometría de flujo.

c-Myc: La nomenclatura proviene de la enfermedad aviar Myelocytomatosis, donde fue descubierto por primera vez como un oncogén y un factor de transcripción. Myc es la familia y existen diversos genes Myc. c-Myc en particular es uno de los cuatro genes canónicos que fueron capaces de reprogramar células diferenciadas a estados pluripotentes.

DPP4: (Dipeptidil peptidasa 4). Es la enzima encargada de degradar a GLP-1 una de las incretinas que genera la primera alza en los niveles de insulina después de la ingesta de los alimentos.

EPC: (*Erythroid Progenitor Cell*). Son progenitores tempranos eritroides utilizados en el presente estudio. Su inmunofenotipo se encuentra en el apartado de procedimientos experimentales.

FDR: (*False Discovery Rate*). Es un estadístico utilizado para medir la cantidad de errores tipo 1 cuando se efectúan múltiples comparaciones para una prueba de hipótesis nula.

fHSC: (*Fresh Hematopoietic Stem Cells*). Células troncales hematopoyéticas obtenidas frescas y sin modificación en el laboratorio posterior a su purificación y selección.

GRN: (*Gene Regulatory Networks*). Redes de regulación transcripcional, es un concepto emergente, pero en general se refiere a las interacciones que tiene los factores de transcripción entre sí y con otros genes para regular su expresión.

GSEA: (*Gene Set Enrichment Analysis*). El enriquecimiento de categorías génicas permite ver cuando un conjunto de genes se está expresando de forma armónica y permite inferir la posibilidad de que una función biológica asociada con ese conjunto pudiera estar activa aún sin medir dicha función.

HIF-alfa: (*Hipoxia Inducible Factor*). El factor inducible de hipoxia se descubrió como el principal actor en la activación de la respuesta celular a bajos niveles de oxígeno. Ahora se conocen como una familia de genes, el canónico es el gen alfa.

HLA: (*Human Leucocyte Antigen*). Estos son los antígenos de histocompatibilidad en seres humanos, se requiere ser compatible en ellos para poder llevar a cabo trasplantes de órganos.

HSC: (*Hematopoietic Stem Cells*). Células troncales hematopoyéticas son las células que pueden sostener al hematopoyético a lo largo de la vida de un individuo. Tienen la capacidad de autorrenovarse y de diferenciarse.

iPSC: (*induced Pluripotent Stem Cells*). Son células con propiedades pluripotentes, muy similares a las células troncales embrionarias, que se generaron mediante expresión ectópica de 4 factores de transcripción a partir de células maduras diferenciadas originalmente fibroblastos de ratón.

ivHSC: (*in vitro generated Hematopoietic Stem cells*). Son células troncales hematopoyéticas generadas *in vitro* a partir de HSC frescas. Las condiciones de cultivo se presentan en el apartado de métodos.

Klf4: (*Kruppel Like Factor 4*). Es un factor de transcripción con estructura de dedo de zinc que regula diversos procesos y cuya familia de genes se encuentra conservada desde el pez cebra hasta el ser humano. Este gen es uno de los cuatro genes canónicos que fueron capaces de reprogramar células diferenciadas a estados pluripotentes.

miRNA: (*micro RNA*). Son RNAs cortos que tiene la capacidad de silenciar la expresión de un gen, una vez que el mRNA de dicho gen haya sido sintetizado.

MPC: (*Myeloid Progenitor Cells*). Células progenitoras Mieloides. El inmunofenotipo utilizado para su purificación se presenta en el apartado de métodos.

mRNA: (*Messenger RNA*). Existen diversos tipos de RNA en las células y tejidos, con diferentes funciones. Este tipo de RNA se encarga de llevar la información necesaria al ribosoma para construir una proteína.

Oct3: (*Octamer binding transcription factor 3*). Este factor de transcripción se ha utilizado como un marcador de pluripotencia *in vivo*. Es además es uno de los cuatro genes canónicos que fueron capaces de reprogramar células diferenciadas a estados pluripotentes.

OP9: Es una línea celular de mesénquima de médula ósea de embrión de ratón, ha demostrado la capacidad de diferenciar a células pluripotentes a linajes hematopoyéticos. Se utilizan principalmente como capa de sostén en el presente estudio durante el protocolo de expansión.

PCA: (*Principal Component Analysis*). Técnica que permite reducir el número de dimensiones para cada muestra de estudio centrándose solo en la máxima variación y reduciendo estas variaciones a máximo 3 dimensiones.

PCR: (*Polymerase Chain Reaction*). La reacción en cadena de la polimerasa se utiliza para amplificar fragmentos de DNA.

RNA: (*Ribonucleic Acid*). Es una de las cinco biomoléculas fundamentales para la vida y cumple funciones catalíticas, estructurales y de transporte de información en la célula y los tejidos.

RNAseq: (*RNA sequencing*). Es una técnica que permite secuenciar los RNA expresados en una muestra biológica. Permite cuantificar la expresión de cada gen mejor que la técnica de microarreglos.

Sox2: (*SRY-box transcription factor 2*). Es un factor de transcripción en el desarrollo embrionario y sexual. Además, es uno de los cuatro genes canónicos que fueron capaces de reprogramar células diferenciadas a estados pluripotentes.

TAC: (*Transcriptome Array Console*). Es un programa que permite llevar a cabo análisis de expresión diferencial, PCA y t-SNE, entre otros. De forma automática sin la necesidad de saber cómo utilizar el lenguaje de programación R.

TF: (Transcripción Factor). Son proteínas que pueden activar la expresión un gen al colocarse en el promotor y entonces reclutar la maquinaria de transcripción.

t-SNE: (t-Distributed Stochastic Neighbor Embedding). Es un método de reducción de la dimensionalidad que permite tener indicadores estadísticos sobre la reducción. Es un método que si se tiene una cantidad cercana a 300 o mayor de muestras es más precisa.

UCB: (Umbilical Cord Blood). Es sangre obtenida del cordón umbilical usualmente de la vena umbilical.

UMAP: (Uniform Manifold Approximation and Projection). Es un método de reducción de la dimensionalidad como t-SNE o PCA. Este método conserva mejor la estructura global inicial, de los datos al analizarlos, comparado con PCA o t-SNE.

VGF: (Vascular Growth Factor). Es un factor que promueve el crecimiento de vasos sanguíneos por lo regular en respuesta a condiciones de hipoxia.

Resumen.

La expansión *in vitro* de células troncales hematopoyéticas (HSC) es una meta clave para la hematología experimental debido a su amplio potencial terapéutico. Se ha demostrado con anterioridad que las HSC generadas *in vitro* y las HSC frescas difieren en términos funcionales y transcriptómicos. En el presente estudio se ha expandido un análisis transcriptómico previo, para encontrar redes de regulación transcripcional que sean de relevancia en el control transcripcional en ambas HSC (generadas *in vitro* y frescas). En el presente estudio se identificó a JUN como uno de los factores de transcripción más conectados de la red y al óxido de metilarsina como un posible fármaco para mejorar el protocolo de expansión actual. También se encontró que las HSC generadas *in vitro* poseen enriquecimiento en siete complejos de remodelación epigenético asociados a activación transcripcional. Es posible tomar ventaja de conocer estas diferencias para llevar a cabo modificaciones en el protocolo de expansión que lleven a las HSC generadas *in vitro* más cerca, en términos biológicos, a sus análogos frescos.

Introducción.

Las células troncales hematopoyéticas (HSC, por sus siglas en inglés) son las encargadas de producir la fracción celular de la sangre [1]. Estas células se denominan troncales por poseer dos propiedades muy particulares: la autorrenovación y la diferenciación. La primera se refiere a la capacidad de llevar a cabo mitosis, dando lugar a dos células con propiedades prácticamente idénticas a las de la primera. La segunda, a poder dividirse dando lugar a una o más células distintas a la célula original [1]. La autorrenovación permite que estas células sean trasplantadas de un individuo a otro sin que se agoten sus capacidades para reconstituir el sistema hematopoyético. El hecho de poder ser utilizadas en trasplantes les da una alta relevancia en el campo de la medicina, ya que pueden ser parte de tratamientos para diversas enfermedades, desde el cáncer hasta el lupus [2], [3]. Existen tres fuentes de HSC: Sangre periférica movilizada, médula ósea y sangre de cordón umbilical (UCB, por sus siglas en inglés), las dos primeras son fuentes adultas, su obtención presenta diversos problemas técnicos y una elevada tasa de enfermedad del injerto contra hospederio que desencadena multitud de complicaciones post trasplante [4]. La UCB tiene muchas ventajas metodológicas entre las que se encuentra su fácil obtención, su almacenamiento en bancos de sangre conectados a lo largo el mundo y su capacidad para tener mayor compatibilidad inmunológica. Esta última permite utilizar donadores aunque éstos no tengan compatibilidad completa en los antígenos HLA. Debido a esto la enfermedad de injerto contra hospederio también se presenta con menor frecuencia al utilizar HSC de UCB [5]. Sin embargo, las HSC de UCB presentan un problema: en una unidad de UCB, los números de HSC son muy inferiores a los obtenidos en fuentes adultas, por lo que en muchos casos no son suficientes para trasplantar adultos. Es por esto que investigadores a lo largo del

mundo se encuentran en una cruzada por encontrar un método para poder aumentar los números de HSC, sin que éstas pierdan sus características funcionales, lo que se denomina expansión [6]. A pesar de la diversidad de protocolos para expandir a las HSC de UBC que se han desarrollado [7], sólo uno ha podido ser utilizado sin emplear una unidad de HSC secundaria [8]. Por lo que aún queda un largo camino por recorrer en la búsqueda de un protocolo de expansión eficiente, barato y seguro.

Para medir la expansión de una población debe tomarse en cuenta al menos un parámetro, la población que se expande debe de aumentar en número de células y mantener el parámetro en estas células semejante a la población original. El parámetro puede ser la expresión de algún marcador de superficie, una capacidad funcional o algún otro tipo de parámetro que se defina. Nuestro grupo de investigación ha logrado llevar a cabo expansión en términos de la expresión del marcador CD34, la no expresión de los marcadores CD38 y marcadores asociados a linaje, utilizando una combinación de siete citocinas: eritropoyetina, trombopoyetina, interleucina 6, interleucina 3, factor de células troncales, factor estimulador de colonias de granulocitos, factor estimulador de colonias de granulocitos/monocitos y células estromales OP9 [6]. Si bien esto último representa un avance en la búsqueda de aumentar los número de HSC de UCB, la población troncal no está bien definida por los marcadores utilizados, entonces es necesario utilizar más marcadores y ensayos funcionales muy específicos como parámetros de expansión. Es por esto que nuestro grupo utilizó el mismo sistema de expansión antes mencionado sobre una población inicial denominada fHSC, pero se determinó la expansión de esta población en términos de la expresión del marcador CD34 y la no expresión de: CD45RA CD71 y los marcadores asociados a linaje. Por otra parte, se llevaron a cabo ensayos funcionales para determinar su capacidad de diferenciación mediante ensayos de formación de colonias en cultivo semisólido y su capacidad de autorrenovación. Esto se midió utilizando ensayos de células iniciadoras en cultivo a largo plazo (LTCIC), después de 7 días bajo las condiciones de expansión antes mencionadas. A esta población se le denominó ivHSC [9]. Lo que se obtuvo comparando la población inicial fHSC y la población después de 7 días ivHSC fue una ligera expansión, 1.3 veces, en términos de los marcadores de superficie. Sin embargo, no se encontró expansión en términos de los LTCIC, de hecho, no se logró obtener ningún LTCIC. En cuanto a los ensayos de formación de colonias estos se vieron alterados en el potencial para diferentes linajes, especialmente el mieloide. Además de las diferencias en estado funcional las ivHSC y las fHSC presentaron diferencias en la expresión génica medida mediante microarreglos (HTA 2.0). Estas diferencias entre la población original y la población generada *in vitro* mediante el protocolo de expansión, muestran que las ivHSC son diferentes a nivel de la expresión génica global, con respecto de las fHSC, presentan un sesgo mieloide y no presentaron capacidad de autorrenovación. Lo que nos lleva a la siguiente pregunta: de cómo se podrían minimizar las diferencias entre ambas poblaciones. La respuesta puede encontrarse en los factores de transcripción, ya que, se conoce como estos controlan en términos funcionales a las poblaciones que se encuentran en el ápice de la jerarquía hematopoyética [10]. La información necesaria para determinar los cambios en la expresión de factores de transcripción (TF), con el fin de acercar transcripcional y funcionalmente a dos diferentes poblaciones celulares se encuentra en el transcriptoma [11].

Desde el descubrimiento de la capacidad que tan solo cuatro TF: Oct3, Sox2, c-Myc y Klf4 tienen para reprogramar células a estados pluripotentes (iPSC), es decir, con capacidad de formar un

organismo completo [12], diversos grupos han tratado en encontrar formas de predecir qué factores son esenciales para adquirir una identidad celular particular [12,15]. La naturaleza del funcionamiento de los factores de transcripción se presta a una descripción por modelos que describen muchos elementos que interactúan entre sí de variadas maneras. La teoría de redes y la teoría de la información son las dos teorías base para la mayoría de estos modelos. [16,17]. El uso de estos enfoques en la biología de células troncales ha arrojado un conglomerado de conceptos que juntos forman una nueva área de la ciencia, denominada ingeniería celular [19]. Si bien las iPSC se consiguieron utilizando expresión ectópica de TFs, se han conseguido también utilizando fármacos (moléculas pequeñas) y la regulación de micro RNAs (miRNAs) [19,20]. Aunado a esto, se conoce ampliamente que diferentes fármacos pueden regular la expresión de diversos TF [21,22]. Por lo tanto, cabe pensar en regular la identidad y el comportamiento celular utilizando fármacos que controlen los TF que las metodologías antes descritas arrojen como fundamentales para el control de la identidad celular. Utilizar fármacos puede reducir los costos de experimentación y evitar las malignidades que pueden surgir al llevar a cabo expresión ectópica.

En el presente trabajo se efectúa un análisis de las redes de regulación transcripcional de los microarreglos del protocolo antes mencionado [9]. Dado que la plataforma con que se llevaron a cabo los micro arreglos, HTA2.0, tiene restricciones para utilizar los métodos antes mencionados, se utilizarán análisis bioinformáticos, como el análisis de enriquecimiento de categorías génicas, sobre las redes de regulación transcripcional o TF ya encontrados por estos métodos [15].

Marco teórico.

Hematopoyesis.

La sangre es un fluido corporal indispensable para la vida de la especie humana. Se encuentra compuesta de una porción celular y otra humoral. Al proceso por el cual se producen las células de la sangre se le denomina hematopoyesis. Entender los mecanismos por los cuales la hematopoyesis tiene lugar es fundamental para entender diferentes enfermedades, desde enfermedades propias de la sangre como la anemia de Fanconi, enfermedades del sistema inmune, hasta diferentes tipos de cánceres [24]. La hematopoyesis es uno de los conceptos más

antiguos en el campo de la medicina y su estudio ha servido de modelo para diferentes áreas de la biomedicina[25].

Uno de los principales objetivos en el estudio de la hematopoyesis es encontrar el origen de todas las células del tejido hemático: A través de experimentos, que medían el uso de timidina marcada radiactivamente por las células de la sangre y experimentos donde se infundían células sanguíneas igualmente marcadas, se concluyó que la mayoría de las células que componen la sangre tienen una vida corta. Por lo que para mantener a tejido hemático se requiere la producción constante de células sanguíneas. Llevo décadas el descubrimiento y caracterización del origen de esta renovación constante de células en el tejido hemático. Los experimentos pioneros de Till y MacCulloch en 1961 hicieron avanzar abruptamente el campo de la hematopoyesis, sus experimentos consistieron en inyectar de manera intravenosa células de la médula ósea a ratones, que fueron antes expuestos a radiación letal, los ratones sobrevivieron más tiempo conforme más células les fueron inyectadas. Por otra parte, un análisis morfológico de los órganos de dichos ratones reveló nódulos en el vaso, el número de nódulos se correspondía de forma lineal con el número de células inyectadas a los ratones. Por lo anterior se dedujo que dichos nódulos eran de origen clonal y que las células iniciadoras de dichos nódulos eran capaces de dar lugar a diferentes células del tejido hemático, a esta última propiedad se le conoce como diferenciación celular. A estas células iniciadoras de nódulos se le dio el acrónimo de CFU-S (por sus siglas en inglés) [26]. Análisis posteriores demostraron que dentro de los nódulos se encontraban también CFU-S, esto indica que dichas células tienen la propiedad de auto renovarse, es decir, de dar lugar a células prácticamente idénticas a ellas mismas.

Las CFU-S fueron un hito en el campo de la investigación en hematología y hematopoyesis debido a que fueron las primeras en ser entendidas como las células troncales hematopoyéticas (HSC por sus siglas en inglés). Sin embargo, su estudio presentaba diversas complicaciones: se demostró que estas células son solo una pequeña fracción de las células de la médula ósea, representando solo el 0.005% del componente celular. Por otra parte su morfología es prácticamente indistinguible de la morfología de un linfocito, por lo que aislarlas era complicado y los intentos por cultivar los componentes celulares de la médula ósea parecían infructíferos. A pesar de estas complicaciones, diversos hallazgos permitieron comprender la fisiología de estas células. En 1966 se implementaron cultivos sólidos, con agar y células de sostén que permitieron obtener colonias de granulocitos y monocitos, además de permitir llevar a cabo estudios cuantitativos. Estos estudios demostraron que dichas colonias eran de origen clonal y que además era separables tanto de la población inmadura con morfología de tipo linfoide, como de las CFU-S. Por lo anterior se denominó a estas células progenitores comprometidos a linaje. Posteriormente se consiguió crecer en cultivos sólidos colonias de eritrocitos y linfocitos, estas también se demuestran de origen clonal, dando lugar al concepto de jerarquización en sistema hematopoyético. Es decir que la célula troncal da lugar a los progenitores y éstos a su vez dan lugar a las células maduras [27].

Hoy en día se entiende a la hematopoyesis como un proceso jerarquizado, que se divide, conceptualmente, en compartimentos. En el ápice de la jerarquía se encuentra el compartimento troncal, en este se encuentran las células troncales, de morfología linfoide, con capacidad de autorrenovarse y diferenciarse. Estas células dan lugar al compartimento de las células progenitoras, también de morfología linfoide, sin embargo, éstas poseen escasa o nula capacidad de autorrenovarse. En este compartimento se encuentran dos tipos de progenitores, los progenitores multipotentes y los progenitores comprometidos a linajes específicos. Este último compartimento da lugar al compartimento de los precursores, este se encuentra compuesto por células que poseen morfologías distintas dependiendo del linaje al que pertenecen. Estas células darán lugar a las células maduras del tejido hemático, las células maduras componen el cuarto y último compartimento. Si sigue un linaje celular a través de la hematopoyesis, se puede notar que a medida que se avanza en el proceso se pierde la capacidad de auto renovación y al mismo tiempo se obtiene un estado mayormente diferenciado, mientras que la ventana de diferenciación se cierra. Es decir que la ventana de posibles linajes a los que la progenie celular puede llegar disminuye [25][28].

El panorama clásico de la hematopoyesis muestra que en el compartimento de células progenitoras se pueden encontrar a los progenitores multipotentes (MPP, por sus siglas en ingles), a los progenitores multipotentes con sesgo linfoide (LMPP, por sus siglas en ingles), a los progenitores mieloides comunes (CMP, por sus siglas en ingles), a los progenitores granulomonocíticos (GMP, por sus siglas en ingles), a los progenitores comunes linfoides (CLP, por sus siglas en ingles) y a los progenitores megacario-eritroides (MkEP, por sus siglas en ingles). La célula troncal hematopoyética da lugar al MPP, este progenitor a su vez da lugar a dos progenitores, por un lado, a los LMPP y por otro al CMP. El CMP dará lugar MkE, mientras que el LMPP dará lugar a GLP y ambos darán lugar al GMP. Sin embargo, la progenie posterior al GMP se encuentra dividida entre neutrófilos y células dendríticas, cuando el linaje parte desde LMPP, mientras que cuando el linaje parte desde GMP se darán monocitos y eosinófilos. En el caso CLP su progenie derivará terminalmente en células linfoides tanto linfocitos B como T y NK. Por otra parte, la progenie de MkEP serán megacariocitos y eritrocitos [29]. Para más detalle ver figura 1.

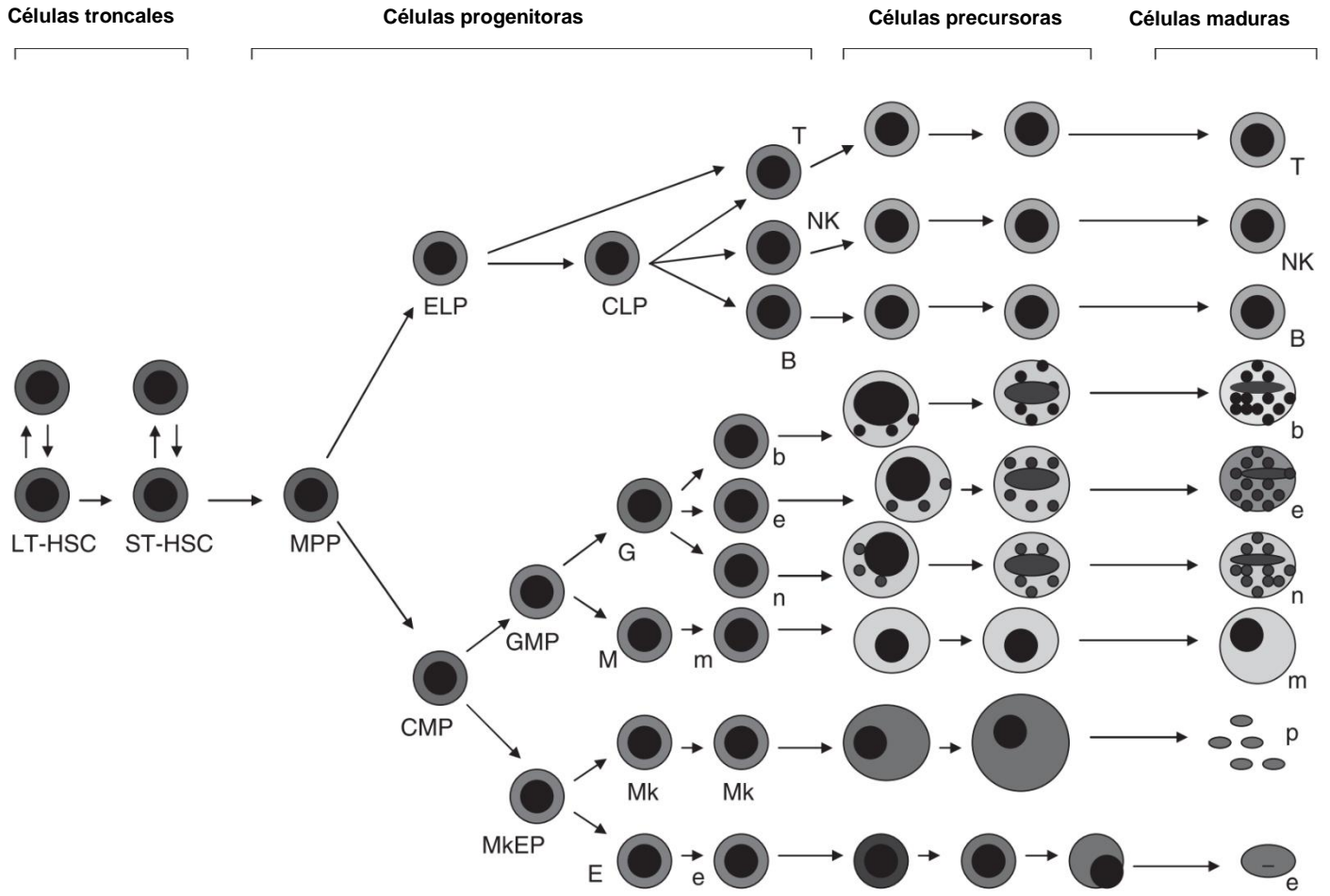


Figura 1. Organización jerárquica del sistema hematopoyético. Al menos 4 compartimentos separados están comprendidos dentro del sistema hematopoyético. El primer compartimento corresponde a las células troncales hematopoyéticas (HSC), que son capaces tanto de autorrenovación como de diferenciación multilinaje. Se han identificado dos subconjuntos diferentes de HSC, capaces de reconstitución a largo plazo (LT) y a corto plazo (ST), respectivamente. Las HSC dan lugar a células progenitoras hematopoyéticas (HPC), incluidas las progenitoras de linaje múltiple y aquellas comprometidas con linajes sanguíneos específicos. Las HPC, a su vez, generan células precursoras morfológicamente reconocibles que, finalmente, dan lugar a células circulantes completamente maduras. Abreviaturas: MPP, progenitor multipotente; ELP, progenitor linfoide temprano; CLP, progenitor linfoide común; CMP, progenitor mieloide común; GMP, progenitor granulocito-macrófago; MKEP, progenitor megacariocito-eritroide; T, célula T; NK, célula asesina natural; B, célula B; b, basófilo; e, eosinófilos; n, neutrófilos; m, monocito; Mk, megacariocito; p, plaquetas; e, eritrocito. Tomado de [29].

Este modelo clásico sobre la hematopoyesis se desarrollado a lo largo de varias décadas utilizando estudios *in vitro* e *in vivo* [24]. Dentro de los estudios *in vitro* existen estudios que permiten medir el potencial de diferenciación, este es al caso de los cultivos semisólidos en metilcelulosa que permiten observar la formación de colonias de uno o varios linajes permitiendo saber que tan amplia es la ventana de diferenciación de un progenitor. Por otra parte, diferentes ensayos se han llevado a cabo para poder encontrar proteínas de superficie, que funcionen como marcadores, para caracterizar a las diferentes poblaciones de células hematopoyéticas. A este conjunto de proteínas de superficie se le denomina inmunofenotipo. Con este inmunofenotipo es posible aislar poblaciones particulares, mediante sort celular, para estudiar su comportamiento en modelo murino. Uno de los ensayos más comunes para estudiar el comportamiento de las HSC humanas *in vivo* es utilizando ratones inmunodeficientes. Éstos son sometidos a radiación letal y luego se les inyecta HSC humanas, obtenida por citometría de flujo acoplada a sort celular, posteriormente se evaluar el comportamiento de dichas células mediante citometría de flujo. Técnicas más avanzadas han permitido utilizar ratones humanizados que comparten características moleculares con el ser humano [30]. La capacidad que esa población celular tenga para recuperar el tejido hemático del ratón irradiado y el tiempo que lo pueda mantener hablara mucho de sus capacidades de autorrenovación. Por otra parte, utilizando citometría de flujo se puede evaluar qué porcentaje de las células maduras proviene del trasplante humano y de esta forma evaluar su potencial de diferenciación *in vivo*.

El sistema hematopoyético está regulado por diferentes factores. Dentro de los primeros en ser descubiertos se encuentran los factores solubles, entre éstos se encuentran proteínas de bajo peso molecular que pueden regular la función celular, a éstas se les denomina citocinas. Pueden promover proliferación celular, quiescencia, activación de alguna propiedad biológica particular o la expresión de ciertos receptores de superficie. La eritropoyetina (EPO, por sus siglas en inglés) es la citocina encargada de regular positivamente la producción de eritrocitos durante la hematopoyesis, a este proceso se le conoce como eritropoyesis. Cuando el organismo se encuentra en bajos niveles de oxígeno el riñón responde con un aumento la producción de esta citocina. Esta citocina ejerce su acción a nivel de los progenitores eritroides tardíos, inhibe su apoptosis e incrementa su proliferación como consecuencia ultima se genera una expansión de células rojas [31]. Una misma citocina puede actuar en diferentes niveles de la jerarquía hematopoyética, tal es el caso de la trombopoyetina (TPO, por sus siglas en ingles), que, si bien actúa a nivel de las unidades formadoras de colonias de megacariocitos, para amplificar la población de plaquetas, también se ha encontrado que puede afectar positivamente la expansión y mantenimiento de las HSC [32]. Otra citocina con la capacidad de actuar en la población troncal es el factor de células troncales (SF, por sus siglas en inglés), este factor posee la capacidad de expandir a las poblaciones tanto troncal como progenitora en el sistema hematopoyético [33]. Las HSC no son solo blancos de diferentes citocinas sino también son productoras de ellas. El ligando de Flt-3 (FL, por sus siglas en inglés), es un claro ejemplo de ello, esta citocina producida por las HSC y tiene escasos efectos en esa misma población, a menos que ejerza sus efectos en paralelo con otras citosinas como son SF, el factor estimulador de colonias de granulocitos y monocitos (GM-CSF, por sus siglas en ingles), el factor estimulador de colonias de granulocitos (G-CSF, por sus siglas en ingles), IL-3 y IL-6. En estos casos el efecto es sinérgico, generando un incremento de las HSC y células progenitoras hematopoyéticas [34]. Esta misma combinación sinérgica tiene efecto proliferativo sobre los progenitores mieloides y precursores de linfocitos B.

Las citocinas antes mencionadas son un ejemplo de factores extrínsecos que regulan el comportamiento de las HSC. Así mismo el microambiente en el que la HSC se encuentre, es un factor extrínseco, que puede ser determinante para controlar su comportamiento. Existen microambientes especializados donde la HSC pueden habitar, a éstos se les conoce como nichos. Al momento se conocen tres nichos: el osteoblástico, el vascular y el reticular [35]. Los osteoblastos del primer nicho promueven quiescencia en las HSC a través de señales como Flamingo, Frizzled-8 y osteopontina [35]. En nicho vascular, constituido por células que recubren las sinusoides, parece tener una acción inductora de la autorrenovación y por último el nicho reticular parece ser el lugar donde se inician los procesos de diferenciación temprana. En este nicho se produce CXCL12, que fundamentalmente permite la adhesión de HSC al nicho, y es producida principalmente por células estromales. El nicho cambia de ubicación conforme el desarrollo del individuo procede. Se sabe que la hematopoyesis en la etapa fetal se desarrolla en el hígado para posteriormente ocurrir en la médula ósea. Además del cambio de ubicaciones el sitio donde se lleva a cabo la hematopoyesis atraviesa un proceso de envejecimiento donde el tejido medular va desapareciendo en favor de tejido adiposo, las HSC por su parte comienzan a tender a producir mayor extirpe mieloide, se empieza a observar un sesgo a dicho linaje [36].

Los efectos finales que las citocinas y otros factores externos son capaces de generar, en las HSC, son en realidad llevados a cabo por factores de transcripción y factores epigenéticos. Un concepto fundamental para entender dicha regulación de la hematopoyesis es la polifonía de linaje: Bajo el esquema clásico de la hematopoyesis se puede encontrar una correspondencia clara entre los factores de transcripción asociados a linaje y los marcadores de superficie que conforman el inmunofenotipo de los linajes que conforman la jerarquía hematopoyética. Sin embargo, esto no ocurre en el ápice de la jerarquía hematopoyética. Donde se puede observar la expresión de reguladores de la transcripción, como TF y potenciadores asociados a diferentes linajes, expresados de forma basal en estas poblaciones, a esto se le conoce como polifonía de linaje. Se piensa en la HSC, no, como una hoja en blanco donde se escribirán los linajes a los que se comprometerá, sino, como una batalla entre los diferentes linajes por expresar su voz. Ir suprimiendo a los demás linajes hasta que uno prevalezca. Un ejemplo claro de este fenómeno es la expresión de los TFs, con ID, GATA-1 y PU.1. El primero es un factor de transcripción promueve la diferenciación eritroide, mientras que el segundo promueve la diferenciación mieloide. Utilizando el modelo de pez cebrado se demostró que los dos TFs antagonizan mutuamente de forma física. Al inhibir GATA-1 utilizando morfolinós en MPP se genera que su diferenciación sea mieloide y mientras que al inhibir a PU.1 se observa la formación de linaje eritroide [37], [38]. Este mismo caso antagónico se repite en los TFs EKLf y FLI-1, para promover linaje eritroide o megacareocítico, respectivamente [39].

En cuanto a autorrenovación se han encontrado varios TFs asociados con esta, entre éstos se encuentran ZFP90, HOXA9, HOXB4, SOX8, SOX18, MYC, RUNX1, JUN Y NFIB [40]–[46], entre otros. Esto puede indicar que existen diferentes formas de promover o evitar la autorrenovación. Los sistemas biológicos en general han demostrado ser dependientes del contexto, por lo que cabe pensar, que dependiendo de las diferentes condiciones en las que estos experimentos se han efectuado, las HSC respondan diferentes factores, ya que la condición particular, en la que se encuentran las hace competentes a dicha respuesta, algunos de estos actores comparten vías de señalización [47]. Si bien los TFs han demostrado ser claves en la regulación del curso natural

del comportamiento de las HSC. Artificialmente se ha conseguido que una célula, ya comprometida a linaje, obtenga una identidad de HSC. A este proceso le conoce como desdiferenciación celular. Esto se logró utilizando la expresión ectopía de 8 TFs en progenitores mieloides y linfoides [45], [48].

Algunos de los TFs antes mencionados interactúan con complejos que son capaces de remodelar la cromatina y con complejos que permiten colocar marcas epigenéticas como son acetilaciones en las histonas y metilaciones en el DNA: En el ratón se ha demostrado la importancia de los genes, con ID, GATA y FOG para el desarrollo de del linaje eritroide y megacareocítico. Los productos de ambos genes forman un complejo que se une al complejo de remodelado de la cromatina NuRD, esta asociación parece permitir el acceso de GATA a sus promotores blanco [49]. Otros ejemplos son el del producto del gen con ID IKAROS, que de igual manera forma un complejo con NuRD en este caso para generar heterocromatina en otras palabras cromatina en estado cerrado, en genes relacionado a linajes diferentes al linfoide [50]. El producto del gen, con ID EKLF, se asocia físicamente con el complejo de remodelación de la cromatina Swi/Snf, cuando esta asociación se rompe en ratones con una mutación que impide la asociación entre los factores antes mencionado esto lleva a la muerte durante la etapa embrionaria [50]. En lo referente a marcas de acetilación o metilación, se sabe que el TF NEF2 se asocia con el MLL2, una enzima capaz de metilar en la lisina 4 de la histona 3, esto es crucial para la expresión de la beta globina en los eritrocitos y precursores de dicho linaje [52]. El represor de dedos de zinc, Gfi, se asocia con complejo de desmetilación LSD1, el cual es capaz de reprimir la expresión de genes asociados a los linajes eritroide, megacareocítico y mieloides para controlar la diferenciación [53]. Se ha encontrado que la acetilación tiene una gran importancia para regular la autorrenovación. Cuando se utiliza el ácido valproico, un inhibidor de la acetilación de las histonas, se observa un aumento significativo de la autorrenovación de células humanas CD34+, esto se probó mediante ensayos de repoblación en ratones inmunodeficientes [54], [55]. Por lo anterior los mecanismos de control epigenéticos parecen tener un rol primordial, junto con los TFs, para controlar el proceso de diferenciación y autorrenovación en el sistema hematopoyético.

Complejidad en sistemas biológicos.

Históricamente la ciencia ha tratado de explicar los fenómenos naturales mediante la reducción de éstos a unas pocas variables y a las relaciones entre ellas. Desde el advenimiento de la

mecánica clásica, con obras como “Principios Matemáticos de la Filosofía Natural” de Issac Newton, la ciencia en general, y no solo la física, se ha regido con dicho enfoque [56]. Tomemos un ejemplo en biomedicina: la mayoría de las enfermedades infecciosas pueden ser explicadas tan solo pensando en el agente infeccioso, erradicando dicho agente la cura ésta encontrada, reducimos todo un fenómeno patológico a una única variable: el agente infeccioso. Consecuentemente se debe encontrar un veneno que pueda matar de forma selectiva a dicho agente infeccioso, en el caso de bacterias un antibiótico, en el caso de hongos un antimicótico, en el caso de un virus un antiviral y en el caso de un parasito un antiparasitario y buenas medidas de higiene. Por supuesto existen agentes infecciosos que exceden esta norma y cuyo control no puede ser alcanzado con medicamentos y medidas de higiene. A pesar de esto, en la biomedicina un enfoque reduccionista ha sido de gran utilidad durante los último 100 años. Sin embargo, durante el siglo XX comenzó a aumentar el número de pacientes con enfermedades que se han resistido a estos enfoques reduccionistas, aunque se conocen estas enfermedades desde hace siglos no hemos sido capaces de curarlas [57]. Tomemos como ejemplo la diabetes mellitus tipo 2, esta enfermedad está caracterizada por una insensibilidad a la hormona insulina por parte de diferentes células del cuerpo, principalmente el musculo esquelético. Esto genera que los niveles de glucosa se eleven descontroladamente y se propone a la gluco-lipo toxicidad como el agente causante de las manifestaciones patológicas de la enfermedad [58]. Luego entonces se espera que, al disminuir los niveles de glucosa mediante diferentes fármacos con alta eficacia para disminuir la glucosa en sangre, como: secretagogos (Diamicron) o inhibidores de la DPP4 (Exenatide), se lograra evitar dicho fenómeno patológico, como en el caso de las enfermedades infecciosas. Sin embargo, estos fármacos tan efectivos en el control glucémico no confieren ninguna protección contra las manifestaciones patológicas más mortales: las complicaciones. Mientras que otros fármacos que proporcionan un control glucémico no tan eficaz si logran una alta protección contra complicaciones (Dapaglifozina) [59]. Entonces se puede pensar que reducir esta enfermedad al aumento de glucosa en sangre nunca será suficiente, para poder describirla y entonces curarla. Mas aun no es posible predecir cuando alguien va a desarrollar diabetes con certeza, sabemos que para que alguien se contagie de hepatitis tiene que haber tenido contacto con alguien infectado o con algo infectado, pero para la diabetes solo contamos con factores de riesgo, no tenemos un agente causal. Parece que este tipo de enfermedades son complejas, ya que, cuentan con muchos elementos que controlan su comportamiento [60]. Así mismo puede que sean o estén al borde del caos: es decir que con pequeñas variaciones en las condiciones iniciales se obtienen resultados muy distintos en el comportamiento de un sistema. Este punto de vista sobre los sistemas biológicos puede dar explicación a fenómenos, que se podrían pensar en un principio como no relacionados: La probabilidad de desarrollar diabetes y eficiencia en el proceso de desdiferenciación celular pueden parecer fenómenos sumamente lejanos. Sin embargo, tomar a ambos procesos como eventos complejos al borde del caos permite explicar porque solo uno de dos gemelos con estilos de vida muy parecidos desarrolle diabetes tipo 2 y el otro no [61], [62] y al mismo tiempo que la eficiencia para desdiferenciar a células a estados pluripotentes sea tan baja, porque todas esas células, al desdiferenciar, poseen variaciones muy sutiles en sus condiciones iniciales antes de la expresión ectópica [63], al igual que los gemelos en sus primeros años de vida. Decir que los sistemas biológicos son caóticos no se refiere a que son desordenados sino a que existe en ellos, una gran sensibilidad a ciertos tipos de variaciones, por pequeñas que éstas puedan parecer, y decir que son complejos, se refiere, a que en realidad no podemos reducir y entender su comportamiento fijándonos en unas pocas variables, debemos tomar en cuenta la mayor cantidad de variables y detalles posibles, para lograr dicho objetivo.

Dichas enfermedades y muchos fenómenos en biomedicina presentan ciertas propiedades únicas de los sistemas complejos, debido a esto, diversos grupos de investigación han intentado conceptualizar problemas y fenómenos biomédicos bajo un enfoque de ciencias de la complejidad [64]. Los sistemas complejos presentan ciertas propiedades que no pueden ser comprendidas a partir de conocer las propiedades de los elementos que conforman el sistema, a estas propiedades se les conoce como emergentes [65]. En el sistema hematopoyético, un ejemplo es la anemia falciforme, donde un cambio en la glutamina 6 de la hemoglobina por una valina, acarrea un cambio morfológico en los eritrocitos y esto un cambio a nivel orgánico, más aun, confiere una ventaja contra la malaria [66]. Este fenotipo no es predecible partir de conocer la mutación y es un efecto colectivo de los elementos, en este caso moléculas de hemoglobina. Una de las propiedades más destacables de los sistemas complejos es que se encuentran entre lo aleatorio y lo determinista [61]. A pesar de presentar un comportamiento regular y predecible, un pequeño cambio en ciertas condiciones pueden llevarlo a un comportamiento aleatorio. Ejemplo de esto es el largo debate sobre la generación de diversidad que ocurre en el compartimento troncal hematopoyético. En donde las características como: autorrenovación, diferenciación y cinética de repoblación en modelo murino, son heterogéneas [67]. Diferentes modelos han surgido para tratar de explicar dicho fenómeno: el modelo intrínseco plantea dos vertientes, la determinista y la estocástica. Para ambos modelos existen evidencias y es probable que ambos modelos sean correctos, simplemente cambios pequeños en ciertas variables pueden dar un comportamiento en dirección determinística o en dirección estocástica. Las dificultades experimentales para diseccionar estos fenómenos con cierta precisión pueden ser la causa de no conocer las variables que nos permiten obtener un resultado estocástico o determinista: Los sistemas complejos se encuentran al borde del caos. A pesar de que los sistemas complejos presentan aleatoriedad en su regulación y propiedades emergentes no triviales en su predicción, también presentan robustez. Es decir que, a pesar de los a dichos estímulos, los sistemas biológicos continuarán comportándose en la forma previa a los estímulos. En el sistema hematopoyético podemos observar esta robustez en diferentes fenómenos, tomemos dos como ejemplo: cuando una persona sufre una hemorragia y su sistema hematopoyético es capaz de reconstituir la sangre perdida, claro que depende de la cantidad de sangre perdida que tanto podrá el sistema hematopoyético restablecer las condiciones normales. Siempre será una lucha entre qué tan robusto sea el sistema y qué tan determinante sea el estímulo.

Un sistema caótico es un sistema que no es aleatorio *per se*, pero se percibe como uno a largo plazo, debido a la limitada precisión de las mediciones que se pueden efectuar. Por otra parte, un sistema complejo es aquel que se encuentra compuesto de una cantidad enorme de partes y que puede o no comportarse de forma caótica, por lo regular los sistemas biológicos son sistemas complejos que se encuentran al borde del caos. Sin embargo, sus peculiaridades están aún descubriéndose.

Ciencias ómicas.

Si se desea comprender a las entidades biológicas en términos de las ciencias de la complejidad, y con esto salir del enfoque reduccionista, es necesario obtener datos holísticos. De forma canónica se ha obtenido información funcional sobre moléculas específicas adjudicándole a estas la importancia mayoritaria en cierto proceso. Un ejemplo de esto es el uso que se le dio a la eritropoyetina para medir la falta de oxígeno durante los experimentos que condujeron al entendimiento del mecanismo de respuesta a hipoxia, galardonado con el premio Nobel de medicina y fisiología 2019 [68]. En este proceso es claro que una postura reduccionista es de suma utilidad para entender el fenómeno. Sin embargo, aunque los actores principales de la respuesta hipoxia son los genes: HIF-alfa y VEGF, en realidad existen más de 300 genes lo que se activan durante este proceso. Estos pueden variar en los diferentes tejidos y el efecto colectivo de estos genes puede que sea imposible de estudiar bajo un enfoque reduccionista. Es de hecho gracias a las tecnologías ómicas que tenemos conocimiento del cambio en la expresión de estos 300 genes.

Los sistemas biológicos, en su enorme mayoría, están compuestos principalmente de: lípidos, carbohidratos, proteínas y ácidos nucleicos. Normalmente se estudia una proteína, un lípido, un carbohidrato, un ácido nucleico RNA o DNA y se asocia con algún fenómeno o entidad biológica en particular: la expresión de la proteína de membrana CD34 fue asociada con la población troncal hematopoyética [69]. Así mismo el estudio de poblaciones específicas, en la ciudad de San Luis en los alrededores de Maracaibo Venezuela, en donde existe una alta tasa de pacientes con corea de Huntington, permitió encontrar un gen y asociarlo con dicha enfermedad y así explicar su carácter familiar [70]. De esta forma, son innumerables los logros que tuvo y continúa teniendo el enfoque reduccionista en la biomedicina. Sin embargo, tener información de multitud de moléculas acercándose a la totalidad de los componentes celulares, nos permite tener información no solo asociativa sino mecanística, causal y explicativa, que sería difícil de obtener mediante experimentos convencionales. Gracias a diversas metodologías bioinformáticas, es posible inferir relaciones, no triviales, entre diversas biomoléculas y éstas pueden ser explicativas de fenómenos biológicos e incluso de fenómenos emergentes [71].

La genómica es la ciencia que estudia multitud de genes, en principio todos ellos, que en conjunto se conocen como genoma y no solo un gen o un par como lo hace la genética, de esta manera, existe también la proteómica que estudia la totalidad de proteínas y no sólo algunas de ellas [72], la metabolómica que estudia la totalidad de metabolitos [73], la epigenómica que mide en conjunto marcas en el DNA y las histonas [74], la transcriptómica que estudia la totalidad de transcritos de RNA o por lo menos eso se intenta [75]. Cada una de estas ciencias en conjunto, conocidas como ciencias ómicas, son posibles gracias a los avances en la capacidad que se tiene para medir cada una de las biomoléculas de manera colectiva, de manera masiva y son un regalo poderoso en sobre manera, que los químicos le han brindado a los biólogos. Cada una de estas ómicas presenta innovaciones en técnicas químicas y bioquímicas. Sin embargo, el colectivo inmenso de datos que estas técnicas arrojan carece de sentido, de utilidad y de practicidad, si no se utilizan métodos que puedan extraer la información esencial de ese colectivo enorme de datos y esta tarea no es trivial, es de hecho uno de los cuellos de botella más importantes en la biología actual, que impide extraer el máximo significado de datos ómicos [76].

Aunado a lo anterior tratar de sobre explotar estas tecnologías y dejar de lado el diseño experimental, es otro de los grandes problemas que esta área del conocimiento presenta y que impide utilizarlas de la forma más inteligente y oportuna. Experimentos pobremente diseñados sosteniéndose únicamente sobre este tipo de técnicas no arrojan conclusiones sólidas que permitan el avance de la ciencia [77].

El presente trabajo se limita a dar una breve explicación del funcionamiento de la tecnología de microarreglos. Dado que medir el transcriptoma es de vital interés, pero los costos de secuenciación eran muy elevados hasta el año 2015, se creó un conjunto de sondas (fragmentos de DNA) cada una de las sondas corresponde al mRNA de un gen en particular, al fragmento necesario para diferenciarlo de los demás genes. De esta forma se tiene un conjunto de sondas que representan una cantidad considerable del genoma. Cada una de estas sondas se adhiere a la misma plataforma de silicón, pero, en lugares específicos, como si de un tablero de ajedrez se tratara cada sonda tiene una posición clara y definida por un conjunto cardinal, por ejemplo: sabemos que la sonda para el gen HER2 se encuentra en la posición 2,3 de la plataforma. A este arreglo microscópico de sondas se le denominó microarreglo. Para utilizar los microarreglos el primer paso es extraer el mRNA de la muestra analizar, posteriormente este mRNA se amplifica mediante PCR. Esta PCR tiene dos características destacables: la primera es que en lugar de uracilo tiene timina de esta forma la amplificación va de RNA a DNA, conocido como cDNA, y por otra parte incorpora fluoróforos al producto de amplificación. Una vez amplificado este material se pasa por la plataforma que contiene las sondas. Debido a la complementariedad de secuencia entre los productos de la amplificación y las sondas de la plataforma estos productos se quedan adheridos. Una vez las sondas están adheridas se les excita con un láser y éstas emiten luz, dependiendo de la intensidad de la luz se estima el número relativo de productos de amplificación del gen adherido en cada posición respectiva de la plataforma. Si para un determinado gen en su posición dentro de la plataforma solo hay una sonda, ese será el número máximo de productos de amplificación que este microarreglo podrá medir para ese gen. Lo anterior genera un sesgo estadístico significativo, ya que, dependiendo del número de sondas que cada gen posea en un tipo de microarreglo determinado (habiendo gran variedad en las clases y tipos de microarreglos) podrá encontrarse más o menos representado dentro del conjunto total de genes. Su funcionamiento se basa en la amplificación del RNA procedente de la muestra a estudiar, pero esta amplificación se lleva a cabo con fluoróforos que se quedan unidos a las nuevas copias de DNA. Posteriormente este DNA marcado, procedente del RNA muestra, se vierte sobre el microarreglo y se une por afinidad de secuencia con sus respectivas sondas, una vez unido el DNA marcado con la sonda, ésta se excita con un láser y un lector registra si hay señal del fluoróforo, dependiendo de la intensidad de esta señal se estima la cantidad de RNA, de ese gen en particular, que se encontraba en la muestra. Información más detallada puede encontrarse en [78], [79] Figura 2.

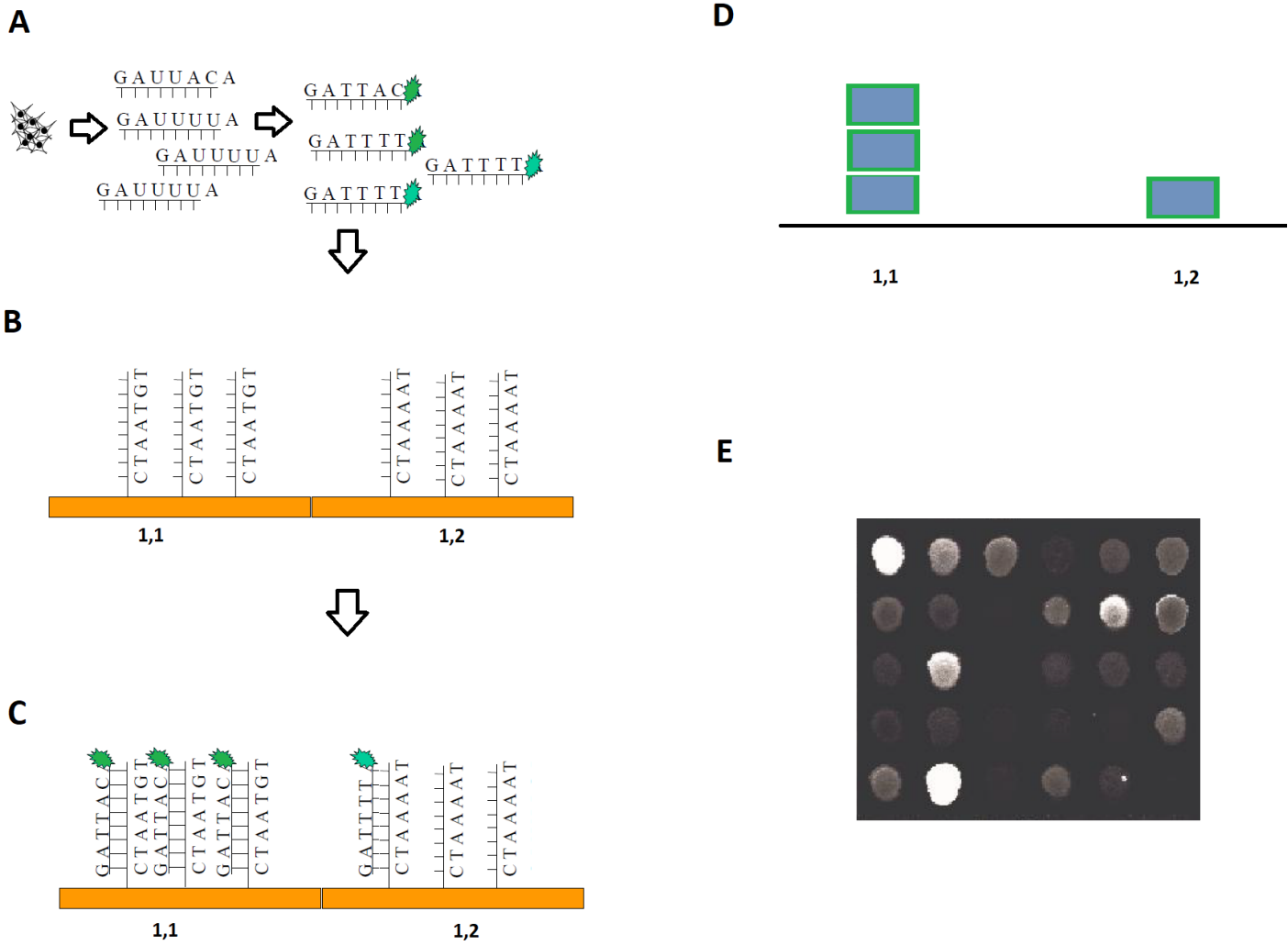


Figura 2. Esquema del funcionamiento de los microarreglos. A) Se muestra cómo se toma el mRNA de una muestra de tejido y esta se transforma hacia cDNA marcado con fluoróforos. B) Se muestra un microarreglo, de forma esquemática se muestran solo 2 casillas con sondas para un gen cada una. C) Se muestra la hibridación de los fragmentos de cDNA, generado a partir del mRNA proveniente del tejido, con el DNA complementario en el microarreglo. D) Las cajas rojas muestran la cantidad de luminosidad medida para la casilla 1.2 del microarreglo y la 1,2 la luminosidad medida para la casilla 1,2, éstas corresponden con el número original de mRNA mensajeros provenientes del tejido. E) Imagen real de cómo se ve un microarreglo bajo el microscopio, cada punto representa una casilla y cada casilla contiene sondas para un gen, se estima la expresión de cada gen dependiendo de qué tan iluminado ésta. Modificado de [78].

Por supuesto que existen diversos inconvenientes con esta tecnología en comparación con la tecnología de RNAseq como: que no todos los polimorfismos de un gen se encontrarán representados, las mediciones son relativas a la cantidad de sondas que cada gen posea y existe un sesgo de medición dado que todo gen tiene un número limitado de sondas y esto varía en las diferentes plataformas de microarreglos. Los microarreglos que poseen una enorme cantidad de sondas para tener una medición amplia de la mayoría de los genes en el genoma se

denominan cuantitativos. Por otra parte, los microarreglos que tienen un número muy limitado de sondas para cada gen se denominan de medición relativa, ya que, es la comparación entre las mediciones de los genes la que nos da una idea de la expresión en general [78],[79].

El Human Transcriptome Array 2.0 (HTA) fue el microarreglo de elección para el estudio que se discute en la presente tesis, este transcriptoma es de mediciones relativas y no cuantitativas, por otra parte, contiene 18,000 genes codificantes específicos de *Homo sapiens* repartidos en 250,000 variantes de splicing y 40,000 micro RNAs (miRNA). Este chip fue diseñado para investigación biomédica básica y traslacional.

Bioinformática.

El término bioinformática hace referencia a la ciencia que estudia datos biológicos de manera global. Gracias a diferentes organizaciones se han creado bases de datos para almacenar diferentes tipos de información biológica, desde secuencias de proteínas en Uniprot, estructuras cristalográficas de las mismas en PDB, hasta transcriptomas completos en GEO. Entonces estos datos pueden ser analizados, no solo por las personas que fueron capaces de obtener la información, sino por cualquier persona con una computadora y acceso a internet. Cabe destacar que esta computadora debe tener cierto poder de cómputo para que pueda llevar a cabo los análisis adecuadamente [80].

Los datos transcripcionales son enormes en cuanto a cantidad de variables se refiere y su análisis no se puede llevar a cabo de forma trivial con métodos estadísticos convencionales, como el análisis de varianza. Se debe proceder con ingenio y aprovechar datos, ya existentes en la literatura, para cada uno de los genes que componen al transcriptoma. Una primera aproximación es relacionar la expresión de ciertos genes con una característica funcional. Tómese como ejemplo la glucólisis, para que esta ruta metabólica se active, es necesario la presencia de todas las enzimas que la componen y es sobre este principio que el análisis de enriquecimiento de categorías génicas o GSEA, por sus siglas en inglés, trabaja. Este análisis toma en cuenta la expresión de manera coordinada de un grupo de genes, una categoría génica, cuya función sea conocida *a priori*, de esta forma, si su expresión es coordinada se asume que esta función biológica puede estar activa. El objetivo final es proponer qué vías y qué funciones biológicas se encuentran activas en una muestra mediante el análisis de la expresión génica. El instituto Broad contiene una base de datos con categorías génicas de todo tipo: desde vías metabólicas a complejos de remodelación de la cromatina. Cabe destacar que el GSEA no está limitado a las categorías que se encuentran en la base de datos antes mencionada, en realidad cada quien puede definir su propia categoría génica la veracidad de esta recaerá en la evidencia previa que esta tenga sobre su existencia y su relevancia en el sistema a estudiar. GSEA será más confiable conforme más transcriptomas se analicen de la misma entidad biológica bajo las mismas condiciones experimentales [81].

En ocasiones se tiene transcriptomas de células diferentes y se quiere saber qué tan cercanos son estos transcriptomas, dado que se podría tratar de entender la cercanía en identidad de un número n de células diferentes mediante diversos ensayos funcionales o análisis morfológico, una mirada al transcriptoma podría arrojar información relevante sobre la similitud de diferentes tipos celulares. Podría mirarse la expresión de un gen en particular y comparar esa entre 3 diferentes tipos celulares, para saber cuál es la relación de expresión entre dichos tipos celulares. Por otra parte, se podría tomar la expresión de 3 genes y graficar a los tipos celulares según su correspondiente, cada gen representaría una coordenada en un plano cartesiano de tres dimensiones.

Si se quiere conocer de manera global qué tan semejantes son entre sí un grupo de transcriptomas, pertenecientes a diferentes tipos celulares, no sería posible graficar la expresión de cada uno de los 30,000 genes, que componen nuestro genoma, ya que estamos limitados a 3 dimensiones y por ende no podríamos graficar 30,000 dimensiones. Tómese como ejemplo el caso de un análisis por citometría de flujo que permita la medición de 21 fluorocromos, como lo permite el FACSSinphony de BD, habrá 21 dimensiones que graficar, Claro está que esa cantidad de dimensiones es imposible de graficar, hasta 30 dimensiones es posible sub-graficar, es decir tomar elementos de tres dimensiones filtrarlos en función de su alta o baja presencia de una de las variables y graficar ahora en función de tres variables distintas y esto se continúa haciendo hasta analizar todas las dimensiones. Sin embargo, cuando la cantidad de variables a graficar ronda los miles, sub-graficar se convierte en una tarea titánica y poco informativa. Debido a este problema se han desarrollado diversos métodos de reducción de la dimensionalidad, estos métodos buscan contraer esas 30,000 dimensiones a solo 3 e incluso 2. Entre estos métodos se encuentra el PCA o análisis de componentes principales, el t-SNE y el UMAP, etc [82], [83]. En este trabajo se abordará de forma simplificada el PCA, debido a que la matemática detrás de este método exige un nivel de rigor fuera de la intención de esta tesis y porque éste fue, de entre todos los métodos, el de elección para el análisis de los datos aquí presentados. De manera superficial el PCA genera una dimensión a partir de la maximización de la varianza entre todas las variables, en este caso, genes de las diferentes muestras, en este caso transcriptomas, a esta dimensión se la denomina primer componente principal, después se busca generar una dimensión ortogonal, es decir, perpendicular la primer componente principal, que de la misma forma maximice la varianza, a esta segunda dimensión se le denomina segundo componente principal. Posteriormente se genera una tercer dimensión con la maximización de varianza de forma ortogonal tanto al primer como al segundo componente principal. Es así como se termina de construir un plano cartesiano tridimensional en el que cada punto es un transcriptoma y conforme más alejados se encuentren dichos transcriptomas más diferentes son de manera global y entre más cercanos se encuentren en el plano más parecida es su expresión global [84].

Los sistemas complejos no presentan una conceptualización que permita un formalismo matemático convencional, por lo que, diversos métodos han sido utilizados para tratar de predecir su comportamiento. Uno de los métodos más ampliamente utilizados es la teoría de redes, una rama de la teoría de grafos, que se encarga de aplicar los principios de la teoría de grafos a problemas prácticos. Las redes se representan por puntos y líneas que conectan dichos puntos entre sí, este esquema permite generar una abstracción sobre las interacciones entre elementos biológicos, por ejemplo: interacción entre proteínas dentro de la célula, interacción entre TF y sus blancos, interacción entre células en un tejido, interacción entre órganos en un sistema, etc. Las

interacciones pueden ser dirigidas o no dirigidas, cuando las interacciones de un TF con su gen blanco no son mutuas, es decir que el TF regula al gen A pero el gen A no regula al TF, entonces puede hablarse de una interacción dirigida y en la red puede ser representada como una flecha que parte del TF a él gen A. Por lo tanto, podemos hablar de redes dirigidas y redes no dirigidas. Estas redes pueden representarse de forma dinámica o de forma estática, cuando son representadas de forma estática se pueden utilizar diferentes métricas para obtener información de la estructura de dichas redes. Una de las métricas más fundamentales es el grado de un nodo. El grado es la cantidad de interacciones que dicho nodo tiene con otros nodos, en el caso de una red no dirigida. Cuando analizamos el grado de los nodos en una red dirigida, el grado puede distinguirse entre el grado de salida y el grado de entrada, cada uno representa la cantidad de flechas que salen del nodo y la cantidad de flechas que apuntan al nodo respectivamente. A partir del grado de cada nodo, en una red podemos generar la distribución de la red, esto es la cantidad de nodos con un cierto grado, en un conjunto de redes generadas de forma aleatoria, por un programa de ordenador, la mayoría de ellas tiene una distribución de grado de Poisson. Sin embargo, la mayoría de redes biológicas siguen una distribución de grado que asemeja una distribución de ley de poder, en donde hay muchos nodos poco conectados y muy pocos nodos altamente conectados. Esta propiedad de la estructura de las redes biológicas parece conveniente para explicar la robustez de los sistemas biológicos, ya que, el sistema de manera global se ve poco afectado al quitar un nodo con poca conectividad. Por otra parte, el tener nodos muy conectados puede explicar fenómenos como que, a partir de la expresión ectópica de tan solo 4 factores de transcripción, las células somáticas se conviertan en células troncales pluripotentes. Otro ejemplo son los eritrocitos estos se cuentan por millones en el tejido hemático. Sin embargo, no representan una fuente importante de regulación para este tejido, en cambio las HSC son muy pocas, pero son las encargadas de la regulación global de dicho tejido. Desde el advenimiento de las ciencias ómicas, es posible facilitar la construcción de redes biológicas a partir de las grandes cantidades de datos que estas técnicas arrojan. Así mismo estas técnicas, especialmente las que permiten la medición del transcriptoma, permiten reflejar el estado de la célula a nivel fisiológico y a nivel de identidad [85]–[88].

Algunas herramientas computacionales para generar redes a partir del estado del arte respecto genes y fármacos son: STRING que es una plataforma web que permite construir una red de interacciones proteicas a partir de un conjunto de genes dados, a partir de información literaria, bioquímica y asociaciones genómicas y STITCH que es una herramienta web que predice redes de interacción proteica tomando en cuenta la interacción de las proteínas de la red con fármacos [89], [90].

Control de los sistemas celulares.

Entender en profundidad los mecanismos que las diferentes entidades biológicas utilizan para regular su comportamiento es sumamente importante, dado que a través de ese conocimiento se planea tratar de controlarlos con diferentes fines, en el caso de la biomedicina el objetivo es curar diversos padecimientos. La palabra clave es “controlar,” esta es una meta a la que cada

vez se acerca más el ser humano. Sin embargo, las entidades biológicas son tan complejas que la mayoría de conocimientos que tenemos sobre cómo controlar sus comportamientos vienen dados por la prueba y el error, es decir, por conocimiento empírico, que podría socavar un poco la frase “entender en profundidad” y a pesar de que se han tratado de generar marcos conceptuales teóricos para tratar de predecir el comportamiento y entonces “controlar”, estos no han tenido tanto éxito. Sin embargo, la utilización de datos ómicos aunado a técnicas de inteligencia artificial y teoría de la información han traído nuevos marcos conceptuales que están teniendo un impacto nunca antes visto en nuestro entendimiento a nivel teórico del control celular y de la identidad celular [91]. Que tanto el transcriptoma es un reflejo de la identidad celular y del estado biológico de la célula es algo que aún se debate [92]. Sin embargo, diversos métodos han logrado avances significativos en el establecimiento de TFs específicos de identidad a partir de datos transcripcionales. Tres métodos han tomado la delantera como los principales en cuanto a la predicción de los factores de transcripción específicos de identidad celular: CellNet: es un método que tiene como objetivo el identificar las redes de regulación transcripcional asociadas a tipos celulares específicos, una vez establecidas dichas redes de regulación se utilizan para analizar el transcriptoma de una célula o población celular alterada de forma *in vitro*, y de esta forma ver cómo cambian en el establecimiento de sus redes de regulación transcripcional. CellNet infiere las redes de regulación transcripcional para un tipo celular dado, por ejemplo, macrófagos, mediante el análisis de una cantidad considerable de transcriptomas, originalmente medido mediante microarreglos: específicamente la plataforma Human Genome U133 Plus 2.0. Los macrófagos de los cuales se obtendrá RNA para hacer la medición del transcriptoma se encuentran en al menos 10 condiciones distintas y el transcriptoma se mide en al menos 60 muestras, esto permite obtener los genes que se mantienen a pesar del cambio en las condiciones. El algoritmo llamado “Probabilidad de relación en contexto”, que tiene como base un árbol aleatorio, es la base de CellNet y permite estimar relaciones entre expresión de genes tanto lineales como no lineales, por lo que su precisión es mayor a la de otros métodos que estiman correlaciones estadísticas como los métodos de correlación de Pearson [11]. Método de entropía: A partir de datos transcriptómicos, obtenidos con la misma plataforma que CellNet utiliza, este método asigna a cada factor de transcripción puntajes que representan que tan específico dicho factor es para un tipo celular dado, es decir, un factor A tendrá varios puntajes para diferentes tipos celulares. De esta forma se genera un atlas de factores de transcripción asignados a diferentes tipos celulares. La forma en la que se calcula el puntaje es mediante un ranqueo basado en entropía no paramétrica de dos propiedades que un factor debe cumplir para ser considerado específico de identidad: que se exprese de manera alta en un tipo celular comparado con otros tipos celulares y que su expresión en otros tipos celulares sea muy baja. De esta forma se construye un atlas que contiene más de 200 tipos celulares y los TF asociados con los puntajes más altos asociados a cada tipo celular [17].

Problema de estudio.

Las HSC de UCB han demostrado ser superiores a las HSC de otras fuentes, esto en términos de las funciones biológicas que se cree son de interés clínico para llevar a cabo trasplantes hematopoyéticos efectivos. Aunado a esto son de todas las fuentes las más fáciles de obtener, ya que se recolectan durante el parto, son también la fuente que se almacena con mayor facilidad en el mundo y las que permiten mayor flexibilidad en cuanto a la compatibilidad inmunológica para el trasplante hematopoyético [5]. Sin embargo, los resultados clínicos de esta fuente de HSC han sido inferiores a lo óptimo, esto es en gran medida adjudicado a la baja cantidad de HSC que cada unidad de SCU posee. Diversas aproximaciones se han llevado a cabo para tratar de aumentar la cantidad de HSC en una unidad de SCU, sin embargo, solo una aproximación ha conseguido llevar a cabo trasplantes únicamente con unidades que fueron expandidas de forma *ex vivo* [8]. El análisis de los transcriptomas obtenidos mediante la comparación de poblaciones frescas y expandidas *in vitro* pueden darnos una visión profunda de los mecanismos que se están viendo alterados entre estas poblaciones y que pueden ser objeto de estudio más detallado o incluso proponerse como blancos para mejorar el proceso de expansión *ex vivo*. En este trabajo presentamos dicho análisis.

Objetivo.

Realizar un análisis del transcriptoma de células troncales hematopoyéticas (HSC), progenitores eritroides y mieloides, frescos y generados *in vitro*, mediante bioinformática, redes de regulación transcripcional, bases de datos de fármacos y factores epigenéticos, para poder proponer fármacos que permitan acercar la identidad celular de las HSCs generadas *in vitro* a la identidad de las HSCs frescas.

Metodología.

Métodos experimentales

Los métodos experimentales mediante los cuales se obtuvieron los microarreglos, que en el presente trabajo se analizan, se encuentran en [9]. Las poblaciones celulares con las cuales se trabaja en este estudio son todas obtenidas de UCB: células troncales hematopoyéticas con inmunofenotipo CD34+ CD38- CD45RA- CD71- Lin- (fHSC), células progenitoras mieloides con inmunofenotipo CD34+ CD38+ CD45RA+ CD71- Lin- (MPC), células progenitoras eritroides con inmunofenotipo CD34+ CD38+ CD45RA- CD71+ Lin- y células troncales hematopoyéticas generadas *in vitro*, mediante un protocolo de expansión en cual las fHSC fueron cultivadas en medio StemSpam, a una densidad de 23,000 células por mililitro, sobre una capa adherente de células OP9, con trombopoyetina, ligando de Ftl-3, interleucina 3, interleucina 6, factor estimulador de colonias de granulocitos y monocitos, factor estimulador de colonias de granulocitos. Se mantuvieron en estas condiciones de cultivo por 7 días, después se seleccionaron por sorteo celular con el mismo inmunofenotipo de inicio del cultivo y se obtuvo el microarreglo, a estas células las llamaremos (ivHSC).

Métodos bioinformáticos.

Se llevo a cabo un análisis de enriquecimiento de categorías génicas, tomando como categorías a los genes pertenecientes a diferentes tipos celulares, reportados en el atlas de factores de transcripción específicos de identidad, se tomaron los 25 primeros genes con mayor puntaje de especificidad. El enriquecimiento se midió entre dos poblaciones MPC y EPC con el objetivo de probar si los factores de transcripción asociados a linaje hematopoyético coincidían con nuestro sistema. De los conjuntos génicos enriquecidos se filtraron los conjuntos génicos pertenecientes a linaje hematopoyético, se consideraron enriquecidos al presentar un $p\text{-val} < 0.05$ y un $FDR < 0.05$, el análisis se llevó a cabo utilizando permutación de los conjuntos génicos, por lo que a pesar de ser estadísticamente significativos debe tomarse dichos resultados con cautela, los resultados se presentan en la figura 3A.

Posteriormente se efectuó un análisis de enriquecimiento de categorías génicas, con las condiciones descritas en el párrafo anterior, entre las poblaciones ivHSC y fHSC, de igual forma los resultados se filtraron con las condiciones descritas antes mencionadas, los resultados se presentan en la figura 1B. Para corroborar los resultados arrojados por el análisis de enriquecimiento de categorías génicas se efectuó un análisis de componentes principales entre las siguientes poblaciones: fHSC, ivHSC y MPC. Utilizando el programa TAC de Thermo Fisher Scitific, con las siguientes especificaciones que viene por defecto en el programa, los resultados

se presentan en la figura 4A, posteriormente se calculó el agrupamiento utilizando el promedio posicional y las distancias entre grupos, lo cual se muestra en la figura 4B. Utilizando los conjuntos génicos mayoritariamente enriquecidos, entre las ivHSC y las fhSC, se generó una lista que se presenta en la tabla 1. Dicha lista fue subida a la plataforma STRING con el objetivo de generar la red de interacciones entre los factores de transcripción que la componen, dicha lista se presenta en la figura 5A. Posteriormente se filtraron los nodos con mayor conectividad también llamados concentradores de la red y se cargaron en STITCH con el objetivo de encontrar fármacos que pudieran interactuar con dichos concentradores reduciendo su expresión o aumentando su expresión dependiendo de su expresión diferencial entre las ivHSC y las fhSC, los resultados se presentan en la figura 5B. Posteriormente utilizando complejos proteicos asociados a eventos de apertura o cierre de la cromatina, reportados en la ontología génica y en la base de datos EpiFactors, se efectuó un GSEA entre las poblaciones ivHSC y fhSC, estos resultados se presentan en la figura 6.

Resultados.

Utilizando los conjuntos génicos reportados como específicos de linaje en [15]. Se efectuó un enriquecimiento de categorías génicas sobre los transcriptomas de progenitores eritroides (EPC) y progenitores mieloides (MPC). Esto con el objetivo de corroborar la especificidad de los conjuntos génicos utilizados. Como se puede observar en la figura 3B el conjunto génico asociado a linaje eritroide ésta exclusivamente enriquecido en el transcriptoma de EPC. El conjunto génico asociado a linaje mioide ésta igualmente enriquecido de forma exclusiva en los MPC. Sin embargo, estos progenitores presentaron un ligero enriquecimiento del conjunto génico asociado a linaje linfoide, esto se explica debido a que el inmunofenotipo, tomado para definir a los MPC, engloba una pequeña población linfoide. Por lo anterior es claro que los conjuntos

génicos reportados en [15] son específicos para las poblaciones del sistema utilizado en el presente análisis. Posteriormente se comparó el enriquecimiento de los mismos conjuntos génicos, pero entre ivHSC y fHSC. Como se puede observar en la figura 3A, la población ivHSC presenta desviaciones en el enriquecimiento de categorías génicas, asociadas a los 3 principales linajes hematopoyéticos, comparada con la población fHSC. Estas desviaciones no se corresponden con lo observado a nivel experimental en los ensayos de formación de colonias efectuados sobre estas poblaciones. Mientras que se observa un sub enriquecimiento del conjunto génico asociado a linaje mieloide en las ivHSC, éstas presentaron en los ensayos de formación de colonias un mayor porcentaje de colonias mieloides que las fHSC.

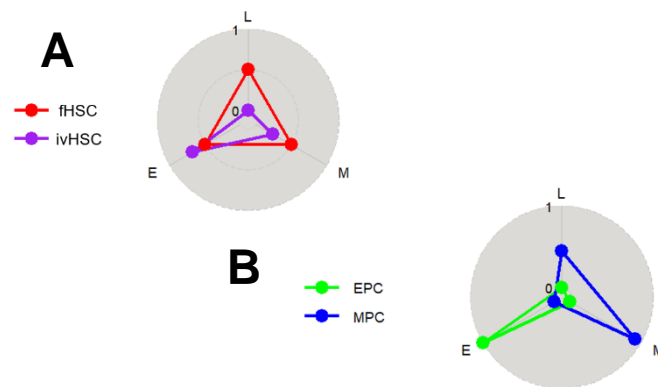


Figura 3. Análisis de enriquecimiento de categorías génicas asociadas con linaje eritroide (E), linfoide (L) y mieloide (M), efectuado sobre microarreglos. Se consideró el mayor puntaje de enriquecimiento normalizado como 1 y en proporción a los demás puntajes. A) Comparación de los puntajes de enriquecimiento entre células troncales hematopoyéticas frescas fHSC y generadas *in vitro* ivHSC. B) Comparación entre células progenitoras eritroides EPC y células progenitoras mieloides MPC.

Para verificar los resultados anteriores, se efectuó un análisis de componentes principales entre las poblaciones ivHSC, fHSC y MPC. Este análisis define el estado de cualquier entidad que este constituida por múltiples variables y permite saber que tan parecidas son 2 entidades a partir de sus variables, sin importar si se trata de cientos o miles de variables. En este caso las entidades a describir son las poblaciones antes mencionadas, mientras que las variables son las expresiones de cada uno de los genes. Se efectuó dicho análisis utilizando el programa TAC, los resultados de distancia entre los puntos nos indica que tan parecidas son las diferentes poblaciones entre sí, como se puede observar en la figura 4 A. Sin embargo, al ser demasiados puntos se procedió a obtener un promedio posicional para medir posteriormente las distancias

entre dichos promedios, estos resultados se pueden observar en la figura 4B. En donde podemos ver que la distancia entre ivHSC y MPC es más grande que la distancia entre fHSC y MPC. Esto corrobora los resultados obtenidos con el GSEA.

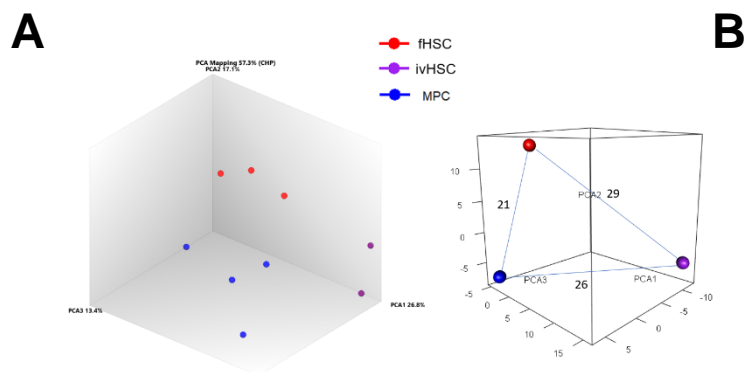


Figura 4. Análisis de componentes principales de microarreglos. A) PCA efectuado con TAC entre poblaciones de células progenitoras mieloides (MPC), células troncales hematopoyéticas frescas (fHSC) y células troncales hematopoyéticas generadas *in vitro* (ivHSC). B) Distancias entre los promedios posicionales de los diferentes tipos de poblaciones celulares.

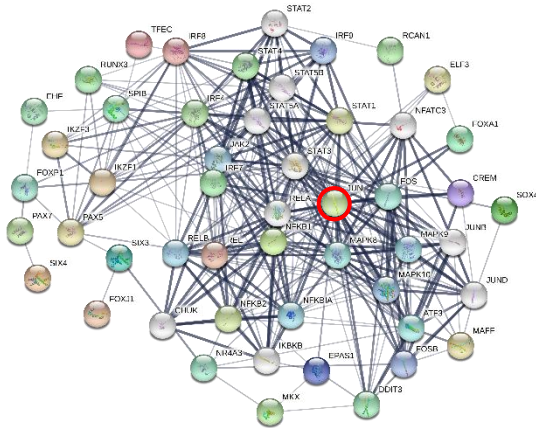
Posteriormente se procedió a analizar de forma más profunda los genes que componen a los conjuntos génicos alterados en las ivHSC. Se seleccionaron los genes que tuvieran más de la mitad del máximo enriquecimiento, estos 20 genes se presentan en la tabla 1.

Tabla 1. Factores de transcripción pertenecientes a los 3 conjuntos génicos con mayor enriquecimiento normalizado en la población fHSC, utilizados para generar la red de interacciones presentadas en la figura 3A.

GFI1B	TAL1	HOXC13	HOXA1
HES6	E2F8	FOXM1	TP63
NFE2	E2F2	GABPB1	MYB
KLF1	MYBL2	TP53	PTTG1
GATA1	IRF6	SMAD3	JUN
TFDP1	BNC1	HNF4A	NFKB1

Los factores de transcripción presentados en la tabla 1 fueron introducidos en la plataforma STRING con el objetivo de determinar cuál es la topología de interacciones entre ellos, con otros factores de transcripción y proteínas. Siguiendo el principio de la distribución de poder [86], se buscó el nodo más conectado de la red obtenida en STRING, este fue el gen JUN como se puede observar en la figura 5 A. Este gen que es un factor de transcripción con tres variantes c-Jun, JUNB y JUND. Se buscaron las 3 variantes en el análisis de expresión diferencial para estimar si se encontraban alteradas en las ivHSC respecto a las fhSC. Solo se encontró el gen JUNB y este se encuentra sub expresado en las ivHSC, mientras que las otras dos formas no se encuentran alteradas. El gen JUNB se introdujo en STITCH, una plataforma que permite predecir interacciones farmacológicas entre proteínas y moléculas pequeñas a partir de diversas bases de datos con un algoritmo bastante preciso [89]. Como se puede observar en la figura 5B el fármaco que es capaz de aumentar la expresión de JUNB es el óxido de metilarsina, con un puntaje de significación de 7.

A



B

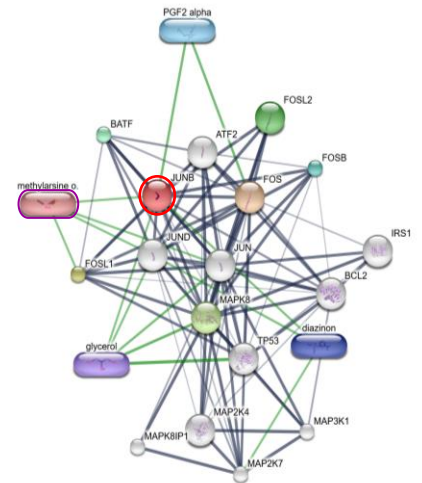


Figura 5. Redes de interacción proteica generada a partir de los factores de transcripción reportados en la tabla 1. A) La red fue generada a través de la plataforma STRING, fue enriquecida buscando interacciones con otras proteínas utilizando la misma plataforma, el círculo rojo enmarca a JUN el nodo con mayor conectividad de la red. B) Red de interacción entre fármacos y proteínas generada mediante la plataforma STITCH. El círculo púrpura remarca al fármaco con el mayor puntaje de asociación para regular la expresión de JUN positivamente, que se encuentra remarcado por un círculo rojo.

Como se presentó en el marco teórico los factores epigenéticos pueden controlar el comportamiento celular a nivel de identidad y de autorrenovación. En ocasiones puede no ser suficiente el controlar la expresión de los factores de transcripción para controlar la identidad celular. Por lo que en el presente estudio se indagó sobre el estado epigenético de las poblaciones celulares de interés. Se ocupó nuevamente la técnica GSEA, en esta ocasión se utilizaron conjuntos génicos formados por genes que se sabe componen complejos de remodelación epigenética. Los conjuntos génicos se obtuvieron de la base de datos EpiFactors [93]. Estos complejos de remodelación se dividieron en dos grupos: unos asociados con apertura de cromatina y otros asociados con cierre de cromatina. En la figura 6 se puede observar círculos, cada uno de estos representa un complejo de remodelación de la cromatina distinto, asociado a activación, y el área de cada círculo representa su enriquecimiento normalizado. Conforme más grande sea el círculo mayor enriquecimiento normalizado tuvo. Si el círculo es azul quiere decir que ese complejo ésta enriquecido en las ivHSC y cuando el círculo es naranja quiere decir que se enriqueció ese complejo en las fHSC. Ningún complejo asociado a cierre de la cromatina se enriqueció, ni en ivHSC ni en fHSC. Por lo anterior se puede destacar que las ivHSC tiene bastante más complejos asociado a cromatina abierta que las fHSC, dado que estas últimas solo presentan un complejo asociado cromatina abierta, siendo éste el de menor enriquecimiento normalizado.

ivHSC/fHSC

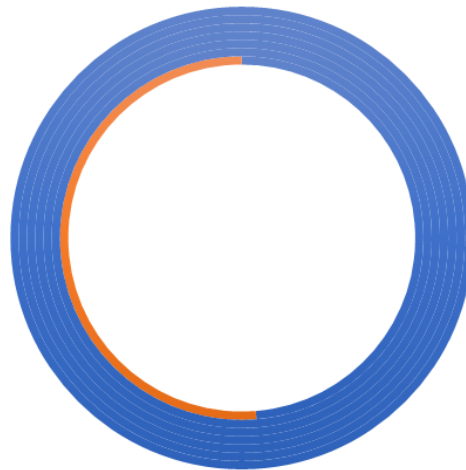


Figura 6. Enriquecimientos de complejos de remodelación de la cromatina, asociados canónicamente con activación de la transcripción, en la población ivHSC (azul) y en la población fHSC (naranja), el área de cada círculo representa el puntaje de enriquecimiento normalizado y cada círculo representa un complejo de remodelación distinto.

Discusión.

Una de las preguntas más importantes a tratar de resolver, en el campo de la terapias celular, es qué tan cercanas son las ivHSC comparadas con las fHSC [30]. Esta cercanía se trató de encontrar en términos funcionales y transcriptómicos en [9] y es claramente contestada en ambos ámbitos. Sin embargo, el objetivo final es poder controlar a las células generadas *in vitro* para que se asemejen más a las células no manipuladas o frescas. Y es por eso que dar una mirada a los grupos de TFs que se sabe *a priori* controlan dicha identidad es fundamental.

En la figura 3B se observa cual es el enriquecimiento obtenido por el método GSEA para los conjuntos génicos de factores transcripción asociados a linajes mieloides, linfoides y eritroides. En dicha figura se puede constatar que para los transcriptomas de células eritroides se obtienen perfectamente el enriquecimiento, mientras que para las mieloides se obtiene un enriquecimiento muy marcado para el conjunto génico de factores mieloides, pero con un sutil enriquecimiento hacia linaje linfoides. Este enriquecimiento era esperado debido al inmunofenotipo con el que se seleccionaron a los progenitores mieloides, que se sabe toma una pequeña población linfoides. Esta primera figura nos permite ver que los factores de transcripción predichos para linaje eritroide y mieloides parecen encajar con las células que se definieron en nuestro estudio como eritroide y mieloides.

Ahora bien, utilizando los mismos grupos de factores de transcripción se midió su enriquecimiento entre las fHSC y las ivHSC. En la figura 3A se puede observar que el mayor enriquecimiento en las fHSC lo tiene el conjunto génico asociado a linaje linfoides, no se llevaron a cabo ensayos de colonias para linaje linfoides en [9] por lo que no se puede comparar este resultado con ensayos funcionales. Sin embargo, la baja en linaje linfoides ya se ha asociado con expansión *in vitro* de HSC de UCB [94]. Por otra parte, en la misma figura se pueden observar que las fHSC se encuentran enriquecidas en linaje mieloides, esto último contradice los resultados experimentales del mismo estudio, ya que, en él se encuentran las ivHSC con un mayor número de colonias mieloides en los ensayos funcionales respecto a las fHSC. Tres explicaciones parecen posibles:

La primera es que debido a la cantidad de repeticiones se genere un sesgo que impida el adecuado funcionamiento de la técnica. La segunda es que el GSEA funciona bajo el principio de expresión en moda constante, a lo largo de los diferentes transcriptomas, de los genes que componen a un conjunto génico definido. En otras palabras, que los genes pertenecientes al conjunto génico en este caso los TF asociados a linaje mieloides, eritroide y linfoides deben expresarse de forma coordinada y constante, si bien, esto tiene mucho sentido para tratar de describir diferentes procesos celulares, como el metabolismo. Sin embargo, puede que en el caso de la diferenciación dentro del sistema hematopoyético esta coordinación no se dé, por lo que,

la técnica GSEA puede no ser capaz de medir un enriquecimiento adecuado. Esta posibilidad se ve apoyada por diferentes evidencias en las que se destaca como la regulación de la diferenciación puede orquestarse de formas diversas a lo largo del paisaje de diferenciación [95],[96].

La tercera opción parece ser la más probable, como se muestra con los resultados en el siguiente párrafo. Esta es que los TF asociados a linaje mielóide si están siendo detectados adecuadamente por GSEA. Sin embargo, el fenotipo mielóide encontrado en los ensayos de colonias puede no ser normal sino uno alterado; es decir que, aunque morfológicamente estas células se van de linaje mielóide puede que en realidad a nivel funcional estén alteradas lo suficiente para que transcripcionalmente no estén enriquecidas en los TF asociados a linaje mielóide. Esta posibilidad podría corroborarse de diferentes formas entre ellas un segundo análisis transcripcional para saber qué tan cercanas son transcripcionalmente las ivHSC a las MPC comparadas con las fHSC.

Se procedió con un análisis de componentes principales (PCA). Como se puede observar en la figura 4A éste tuvo una cobertura de más del 50% por lo que abarca suficiente variación para ser utilizado y al mismo tiempo se observa poca variación entre los transcriptomas de cada grupo. En la figura 4B se presentan los promedios posicionales de los transcriptomas y se obtiene las distancias entre ellos. Es sumamente destacable que la distancia entre las fHSC y los progenitores mieloides MPC es menor que la que existe entre estos mismos progenitores y las ivHSC, esto es congruente con el dato más controversial del párrafo anterior.

Se presentan en la tabla 1 los factores de transcripción que más aportaron, con su expresión, al enriquecimiento diferencial de los conjuntos génicos asociados a linaje; que se observó entre las fHSC y las ivHSC. Con estos factores de transcripción se construyó la red presentada en la figura 5A, mediante la plataforma STRING. En esta red se observa la interacción que tienen dichos factores de transcripción con todo tipo de proteínas. Basados en la teoría de redes biológicas, utilizamos el grado de conectividad como un criterio para encontrar un elemento en la red que fuera importante, en términos de su conectividad con los demás TFs, este fue JUN, como se muestra en la figura 5A este fue el TF con mayor conectividad. JUN tiene diferentes formas: c-Jun, JUNB y JUND, éstas son parte de la vía de señalización JNK y se han reportado con diferentes propiedades biológicas, en diferentes sistemas, incluyendo diferentes tipos de células troncales [97]. Por lo que se obtuvo la expresión diferencial, de todas las formas antes mencionadas, entre fHSC e ivHSC y se encontró únicamente a JUNB sub expresado en las fHSC, con una diferencia de 5.23.

La tercera explicación para el hallazgo más controversial presentado en el segundo párrafo de la presente discusión parece adquirir fuerza si se toma en cuenta los resultados del análisis de redes y el PCA aunados a los siguientes datos: Por una parte, se sabe que la deficiencia de JUNB en sistema murino degenera en síndrome mieloproliferativo [98] y que su sobreexpresión protege contra malignidades mieloides [99]. Fuera del sistema murino, en humanos se ha asociado su deficiencia con diferentes malignidades mieloides y en particular con leucemia mielóide aguda [100][101]. Una de las características de estas patologías es la exacerbación la capacidad proliferativa de los MPC y estos fenotipos mieloides están controlados por GRN diferentes a las de las células en estado sano. Por lo anterior es posible pensar que el fenotipo

mieloide observado en los ensayos de colonia sobre las ivHSC, obtenidos en [9], pueda haberse producido por un conjunto transcripcional alternativo al que en el presente estudio definimos como mieloide y por dicha razón no pudo ser detectado en el GSEA reportado en la figura 3 A. De igual manera esto podría explicar porque, en la figura 4B, se observa a las ivHSC más distantes de los progenitores mieloides. Esta posibilidad tiene soporte extra dado que la sub expresión de JUNB puede alterar las redes de regulación transcripcional [99]. Cabe aclarar que esto no significa necesariamente que las ivHSC estén generando un fenotipo mieloide maligno. Ya que se efectuaron GSEA para conjuntos génicos asociados a diferentes malignidades hematológicas estos resultados fueron nulos: no habiendo enriquecimiento para ninguna de las poblaciones celulares aquí presentadas por lo que no se muestran figuras. Es entonces que los cambios internos que se generan en estas poblaciones obtenidas *in vitro* se pueden alterar su expresión génica y llevar a morfologías mieloides que no necesariamente sean funcionales o malignas, sino algo más debido al protocolo de expansión y esto se ve reflejado en el PCA, GSEA y análisis de redes.

En diferentes estudios se ha relacionado la sub expresión de JUNB con deficiencias significativas en la capacidad de injerto en modelo murino [98], [99]. Lo más destacable es que la vía JNK, a la cual pertenece JUNB, ya ha sido utilizada para expandir a las HSC *in vitro*, disminuyendo la expresión de c-Jun, pero sin afectar negativamente la expresión de JUNB. Consecuentemente se introdujo a JUNB en la plataforma STITCH obteniendo de esta forma diferentes fármacos, moléculas pequeñas, que podían regular su expresión. El fármaco con mayor puntaje de confiabilidad (0.7) y que pudiera regular positivamente la expresión de JUNB fue el óxido de metilarsina, por lo que se le propone como un elemento que pudiese potenciar la expansión *in vitro* de las HSCs, minimizando las diferencias que el protocolo de expansión genera en las ivHSC. Este resultado es interesante tomando en cuenta que este dicho químico ha sido utilizado para tratar leucemias mieloides y se ha demostrado que es citotóxico contra dichas células malignas, *in vitro*. Sin embargo, éste no ha tenido efecto citotóxico sobre las HSC, ni los progenitores multipotentes [102].

Por otra parte, se trató de encontrar un atisbo de información sobre las diferencias en el estado epigenético que presentan las ivHSC comparadas con las fHSC, estas diferencias se presentan en la figura 4. Las diferencias se obtuvieron utilizando GSEA. Cada círculo en la figura 4 representa un complejo de remodelación de la cromatina, siendo los azules complejos enriquecidos en las ivHSC y siendo los naranjas enriquecidos en las fHSC. Estos complejos se tomaron como asociados a activación de forma canónica, ya que existe controversia en cuanto a estas asociaciones [103]. Se puede observar que las ivHSC presentan mayor número de complejos asociados con activación 7, en comparación con los complejos de activación asociados a fHSC solo 1. Además, se puede observar que el área de los círculos azules es mayor que la del único círculo naranja, por lo que los 7 complejos de activación asociados a las ivHSC tienen un mayor enriquecimiento normalizado que el único complejo asociado a fHSC. Por lo que existe la posibilidad de encontrar el estado de la cromatina más abierto en las ivHSC que el de las fHSC [104], [105]. Es posible que este enriquecimiento mayoritario, de complejos de remodelación de la cromatina asociados a activación transcripcional, que se observó en las ivHSC se deba a que al proliferar requieran estar transcripcionalmente más activas que las fHSC. Se sabe que una de las propiedades de las HSC es la quiescencia, esto concordaría con la mínima cantidad de LTCIC que fueron encontrados en las ivHSC [9]. Si bien estos datos no

pueden hablar por sí mismo del estado epigenético son una buena guía, para futuras perspectivas en cuanto al control epigenético que se presenta en las ivHSC. Sin embargo, algunos de los complejos de remodelación de la cromatina que en el presente estudio fueron utilizados, ya han sido propuestos como blancos terapéuticos de epifarmacos (fármacos dirigidos a factores moduladores de la epigenética) [107]. Se requiere saber con mayor precisión que tipo de proteínas están controlando cada región de la cromatina para poder proponer un epifarmaco que pudiera ayudar a mantener la diferenciación suspendida durante el proceso de expansión *in vitro* [106]. Sin embargo, tomando en cuenta la literatura anterior el ácido valproico parece una opción importante.

Si bien todos los resultados, a excepción del PCA y el análisis de redes, son estadísticamente significativos, tiene su base en transcriptomas cuyo número de repeticiones no es el adecuado, es decir que no se cuentan con suficientes transcriptomas por condición biológica. Debido a que el transcriptoma no es tan estable como el genoma [108], en un futuro cercano sería conveniente ampliar el número de muestras, para continuar con estos estudios.

Conclusión.

En conclusión, se propone al óxido de metilarsina como un posible agente potenciador de la expansión *in vitro*, debido a su función como activador de JUN, específicamente para el protocolo presentado en [9]. Por otra parte, parece crucial el obtener un número mayor de transcriptomas repetir entonces el análisis transcripcional. Un análisis del estado de la cromatina por técnicas como Hi-c [110], aunado a un buen análisis bioinformático podrían permitir proponer fármacos en concreto que modulen el estado epigenético para favorecer la expansión de las HSC. Sin embargo, el ácido valproico parece ser la mejor opción, basados en la información que se posee hasta el momento.

Bibliografía.

- [1] H. Mayani, "Hematopoyesis," *Cancerología*, vol. 2, pp. 95–107, 2007.
- [2] N. Esiashvili and M. A. Pulsipher, "Hematopoietic stem cell transplantation," *Pediatr. Oncol.*, vol. 3, no. 97, pp. 301–311, 2018.
- [3] R. K. Burt *et al.*, "Nonmyeloablative hematopoietic stem cell transplantation for systemic lupus erythematosus," *J. Am. Med. Assoc.*, vol. 295, no. 5, pp. 527–535, 2006.
- [4] J. R. Passweg *et al.*, "Hematopoietic stem cell transplantation: A review and recommendations for follow-up care for the general practitioner," *Swiss Med. Wkly.*, vol. 142, no. October, pp. 1–15, 2012.
- [5] H. Mayani, J. E. Wagner, and H. E. Broxmeyer, "Cord blood research, banking, and transplantation: achievements, challenges, and perspectives," *Bone Marrow Transplant.*, 2019.
- [6] V. Fernandez-Snchez *et al.*, "In vitro effects of stromal cells expressing different levels of Jagged-1 and Delta-1 on the growth of primitive and intermediate CD34 + cell subsets from human cord blood," *Blood Cells, Mol. Dis.*, vol. 47, no. 4, pp. 205–213, 2011.
- [7] H. Takizawa, U. Schanz, and M. G. Manz, "Ex vivo expansion of hematopoietic stem cells: Mission accomplished?," *Swiss Med. Wkly.*, vol. 141, no. DECEMBER, pp. 1–9, 2011.
- [8] M. E. Horwitz, S. Wease, B. Blackwell, D. Valcarcel, and F. Frassoni, "Phase I / II Study of Stem-Cell Transplantation Using a Single Cord Blood Unit Expanded Ex Vivo With Nicotinamide abstract," vol. 37, no. 5.
- [9] R. Dircio-Maldonado *et al.*, "Functional Integrity and Gene Expression Profiles of Human Cord Blood-Derived Hematopoietic Stem and Progenitor Cells Generated In Vitro," *Stem Cells Transl. Med.*, vol. 7, no. 8, pp. 602–614, 2018.
- [10] N. K. Wilson *et al.*, "Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators," *Cell Stem Cell*, vol. 7, no. 4, pp. 532–544, 2010.
- [11] P. Cahan, H. Li, S. A. Morris, E. Lummertz Da Rocha, G. Q. Daley, and J. J. Collins, "CellNet: Network biology applied to stem cell engineering," *Cell*, vol. 158, no. 4, pp. 903–915, 2014.
- [12] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.
- [13] A. H. Radley, R. M. Schwab, Y. Tan, J. Kim, E. K. W. Lo, and P. Cahan, "Assessment of engineered cells using CellNet and RNA-seq," *Nat. Protoc.*, vol. 12, no. 5, pp. 1089–1102, 2017.
- [14] O. J. L. Rackham *et al.*, "A predictive computational framework for direct reprogramming between human cell types," *Nat. Genet.*, vol. 48, no. 3, pp. 331–335, 2016.
- [15] A. C. D'Alessio *et al.*, "A systematic approach to identify candidate transcription factors that control cell identity," *Stem Cell Reports*, vol. 5, no. 5, pp. 763–775, 2015.
- [16] M. A. Kinney *et al.*, "A systems biology pipeline identifies regulatory networks for stem cell engineering," *Nat. Biotechnol.*, vol. 37, no. July, 2019.
- [17] D. Marbach *et al.*, "Wisdom of crowds for robust gene network inference," *Nat. Methods*, vol. 9, no. 8, pp. 796–804, 2012.
- [18] P. E. Cholley *et al.*, "Modeling gene-regulatory networks to describe cell fate transitions and predict master regulators," *npj Syst. Biol. Appl.*, vol. 4, no. 1, 2018.
- [19] A. Ma'Ayan and Q. Duan, "A blueprint of cell identity," *Nat. Biotechnol.*, vol. 32, no. 10, pp. 1007–1008, 2014.
- [20] P. Hou *et al.*, "Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds," *Science (80-.)*, vol. 341, no. 6146, pp. 651–654, 2013.
- [21] X. Ma, L. Kong, and S. Zhu, "Reprogramming cell fates by small molecules," *Protein Cell*, vol. 8, no. 5, pp. 328–348, 2017.

- [22] M. Lambert, S. Jambon, S. Depauw, and M. H. David-Cordonnier, "Targeting transcription factors for cancer treatment," *Molecules*, vol. 23, no. 6, pp. 1–51, 2018.
- [23] L. Garcia-Alonso *et al.*, "Transcription factor activities enhance markers of drug sensitivity in cancer," *Cancer Res.*, vol. 78, no. 3, pp. 769–780, 2018.
- [24] B. F. Rodak, G. A. Fritsma, and K. Doig, *Hematology : clinical principles and applications*. Saunders Elsevier, 2007.
- [25] S. H. Orkin and L. I. Zon, "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology," *Cell*, vol. 132, no. 4, pp. 631–644, 2008.
- [26] J. E. Till and E. A. McCulloch, "A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells 1," *Radiat. Res.*, vol. 178, no. 2, pp. AV3–AV7, 2012.
- [27] T. R. Bradley and D. Metcalf, "The growth of mouse bone marrow cells in vitro.," *Aust. J. Exp. Biol. Med. Sci.*, vol. 44, no. 3, pp. 287–299, 1966.
- [28] H. Mayani, "Biological Differences Between Neonatal and Adult Human Hematopoietic Stem/Progenitor Cells," *Stem Cells Dev.*, vol. 19, no. 3, pp. 285–298, 2009.
- [29] A. Görgens *et al.*, "Revision of the Human Hematopoietic Tree: Granulocyte Subtypes Derive from Distinct Hematopoietic Lineages," *Cell Rep.*, vol. 3, no. 5, pp. 1539–1552, 2013.
- [30] A. Tanner, S. E. Taylor, W. Decottignies, and B. K. Berges, "Humanized mice as a model to study human hematopoietic stem cell transplantation," *Stem Cells Dev.*, vol. 23, no. 1, pp. 76–82, 2014.
- [31] A. Grover *et al.*, "Erythropoietin guides multipotent hematopoietic progenitor cells toward an erythroid fate," *J. Exp. Med.*, vol. 211, no. 2, pp. 181–188, 2014.
- [32] K. Kaushansky, "Thrombopoietin and its receptor in normal and neoplastic hematopoiesis," *Thromb. J.*, vol. 14, no. Suppl 1, 2016.
- [33] J. Lennartsson and L. Rönstrand, "Stem cell factor receptor/c-Kit: From basic Science to clinical implications," *Physiol. Rev.*, vol. 92, no. 4, pp. 1619–1649, 2012.
- [34] P. Tsapogas, C. J. Mooney, G. Brown, and A. Rolink, "The cytokine Flt3-ligand in normal and malignant hematopoiesis," *Int. J. Mol. Sci.*, vol. 18, no. 6, 2017.
- [35] S. Pinho and P. S. Frenette, "Haematopoietic stem cell activity and interactions with the niche," *Nat. Rev. Mol. Cell Biol.*, vol. 20, no. 5, pp. 303–320, 2019.
- [36] A. Ivanovs, S. Rybtsov, E. S. Ng, E. G. Stanley, A. G. Elefanty, and A. Medvinsky, "Human haematopoietic stem cell development: From the embryo to the dish," *Dev.*, vol. 144, no. 13, pp. 2323–2337, 2017.
- [37] J. Rhodes *et al.*, "Interplay of pu.1 and Gata1 determines myelo-erythroid progenitor cell fate in zebrafish," *Dev. Cell*, vol. 8, no. 1, pp. 97–108, 2005.
- [38] J. L. Galloway, R. A. Wingert, C. Thisse, B. Thisse, and L. I. Zon, "Loss of Gata1 but not Gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos," *Dev. Cell*, vol. 8, no. 1, pp. 109–116, 2005.
- [39] J. Starck *et al.*, "Functional Cross-Antagonism between Transcription Factors FLI-1 and EKLF," *Mol. Cell. Biol.*, vol. 23, no. 4, pp. 1390–1402, 2003.
- [40] B. King *et al.*, "The ubiquitin ligase Huwe1 regulates the maintenance and lymphoid commitment of hematopoietic stem cells," *Nat. Immunol.*, vol. 17, no. 11, pp. 1312–1321, 2016.
- [41] L. L. Smith *et al.*, "Functional crosstalk between Bmi1 and MLL/Hoxa9 axis in establishment of normal hematopoietic and leukemic stem cells," *Cell Stem Cell*, vol. 8, no. 6, pp. 649–662, 2011.
- [42] E. de Pater *et al.*, "Gata2 is required for HSC generation and survival," *J. Exp. Med.*, vol. 210, no. 13, pp. 2843–2850, 2013.
- [43] T. Liu *et al.*, "The transcription factor Zfp90 regulates the self-renewal and differentiation of hematopoietic stem cells article," *Cell Death Dis.*, vol. 9, no. 6, 2018.
- [44] N. Buza-Vidas, S. Duarte, S. Luc, T. Bouriez-Jones, P. S. Woll, and S. E. W. Jacobsen,

- “GATA3 is redundant for maintenance and self-renewal of hematopoietic stem cells,” *Blood*, vol. 118, no. 5, pp. 1291–1293, 2011.
- [45] J. Riddell *et al.*, “Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors,” *Cell*, vol. 157, no. 3, pp. 549–564, 2014.
- [46] B. R. Webber, M. Iacovino, S. H. Choi, J. Tolar, M. Kyba, and B. R. Blazar, “DNA methylation of Runx1 regulatory regions correlates with transition from primitive to definitive hematopoietic potential in vitro and in vivo,” *Blood*, vol. 122, no. 17, pp. 2978–2986, 2013.
- [47] F. K. Hamey, S. Nestorowa, S. J. Kinston, D. G. Kent, N. K. Wilson, and B. Gottgens, “Reconstructing blood stem cell regulatory network models from single-cell molecular profiles,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 23, pp. 5822–5829, 2017.
- [48] C. Pereira, I. R. Lemischka, and K. Moore, “‘From blood to blood’: de-differentiation of hematopoietic progenitors to stem cells,” *EMBO J.*, vol. 33, no. 14, pp. 1511–1513, 2014.
- [49] S. Sharma and G. Gurudutta, “Epigenetic regulation of hematopoietic stem cells,” *Int. J. Stem Cells*, vol. 9, no. 1, pp. 36–43, 2016.
- [50] J. Kim *et al.*, “Ikars DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes,” *Immunity*, vol. 10, no. 3, pp. 345–355, 1999.
- [51] R. C. Brown *et al.*, “Distinct Domains of Erythroid Krüppel-Like Factor Modulate Chromatin Remodeling and Transactivation at the Endogenous β -Globin Gene Promoter,” *Mol. Cell. Biol.*, vol. 22, no. 1, pp. 161–170, 2002.
- [52] C. Demers *et al.*, “Activator-Mediated Recruitment of the MLL2 Methyltransferase Complex to the β -Globin Locus,” *Mol. Cell*, vol. 27, no. 4, pp. 573–584, 2007.
- [53] S. Saleque, J. Kim, H. M. Rooke, and S. H. Orkin, “Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1,” *Mol. Cell*, vol. 27, no. 4, pp. 562–572, 2007.
- [54] G. Bug *et al.*, “Valproic acid stimulates proliferation and self-renewal of hematopoietic stem cells,” *Cancer Res.*, vol. 65, no. 7, pp. 2537–2541, 2005.
- [55] L. De Felice *et al.*, “Histone deacetylase inhibitor valproic acid enhances the cytokine-induced expansion of human hematopoietic stem cells,” *Cancer Res.*, vol. 65, no. 4, pp. 1505–1513, 2005.
- [56] E. McMullin, “The impact of Newton’s Principia on the philosophy of science,” *Philos. Sci.*, vol. 68, no. 3, pp. 279–310, 2001.
- [57] I. Schäfer *et al.*, “Multimorbidity patterns in the elderly: A new approach of disease clustering identifies complex interrelations between chronic conditions,” *PLoS One*, vol. 5, no. 12, 2010.
- [58] R. J. Mahler and M. L. Adler, “Clinical review 102: Type 2 diabetes mellitus: Update on diagnosis, pathophysiology, and treatment,” *J. Clin. Endocrinol. Metab.*, vol. 84, no. 4, pp. 1165–1171, 1999.
- [59] S. D. Wiviott *et al.*, “Dapagliflozin and cardiovascular outcomes in type 2 diabetes,” *N. Engl. J. Med.*, vol. 380, no. 4, pp. 347–357, 2019.
- [60] F. Mazzocchi, “Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory,” *EMBO Rep.*, vol. 9, no. Science and Society, pp. 10–14., 2008.
- [61] V. B. Betelin, V. M. Eskov, V. A. Galkin, and T. V. Gavrilenko, “Stochastic volatility in the dynamics of complex homeostatic systems,” *Dokl. Math.*, vol. 95, no. 1, pp. 92–94, 2017.
- [62] A. G. Tabák, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimäki, “Prediabetes: A high-risk state for diabetes development,” *Lancet*, vol. 379, no. 9833, pp. 2279–2290, 2012.
- [63] S. A. Telpalo-Carpio, J. M. Aguilar-Yañez, M. T. Gonzalez-Garza, D. E. Cruz-Vega, and J. E. Moreno-Cuevas, “ips cells generation: An overview of techniques and methods,” *J. Stem Cells Regen. Med.*, vol. 9, no. 1, pp. 2–8, 2013.
- [64] S. Jayasinghe, “Complexity science to conceptualize health and disease: Is it relevant to

- clinical medicine?," *Mayo Clin. Proc.*, vol. 87, no. 4, pp. 314–319, 2012.
- [65] C. Systems, "Emergent Properties of Feedback Regulation and Stem Cell Behavior in a Granulopoiesis Model as a Complex System," *Cell*, vol. 14, pp. 45–61, 2003.
- [66] J. M. A. Mauritz, A. Esposito, H. Ginsburg, C. F. Kaminski, T. Tiffert, and V. L. Lew, "The Homeostasis of Plasmodium falciparum-Infected Red Blood Cells," *PLoS Comput. Biol.*, vol. 5, no. 4, pp. 11–14, 2009.
- [67] H. B. Sieburg, "The GOD of Hematopoietic Stem Cells A Clonal Diversity Model of the Stem Cell Compartment," *Cell Cycle*, no. February, pp. 394–398, 2006.
- [68] K. Guillemin and M. A. Krasnow, "The hypoxic response: Huffing and HIFing," *Cell*, vol. 89, no. 1, pp. 9–12, 1997.
- [69] L. E. Sidney, M. J. Branch, S. E. Dunphy, H. S. Dua, and A. Hopkinson, "Concise review: Evidence for CD34 as a common marker for diverse progenitors," *Stem Cells*, vol. 32, no. 6, pp. 1380–1389, 2014.
- [70] F. Saudou and S. Humbert, "The Biology of Huntingtin," *Neuron*, vol. 89, no. 5, pp. 910–926, 2016.
- [71] K. Kamisoglu *et al.*, "Understanding physiology in the continuum: Integration of information from multiple -omics levels," *Front. Pharmacol.*, vol. 8, no. FEB, pp. 1–19, 2017.
- [72] A. F. M. Altelaar, J. Munoz, and A. J. R. Heck, "Next-generation proteomics: Towards an integrative view of proteome dynamics," *Nat. Rev. Genet.*, vol. 14, no. 1, pp. 35–48, 2013.
- [73] C. B. Clish, "Metabolomics: an emerging but powerful tool for precision medicine," *Mol. Case Stud.*, vol. 1, no. 1, p. a000588, 2015.
- [74] K. C. Wang and H. Y. Chang, "Epigenomics technologies and applications," *Circ. Res.*, vol. 122, no. 9, pp. 1191–1199, 2018.
- [75] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLoS Comput. Biol.*, vol. 13, no. 5, pp. 1–23, 2017.
- [76] A. Bayat, "Clinical review Science, medicine, and the future Bioinformatics," *Br. Med. J.*, vol. 324, no. April, pp. 1018–1022, 2002.
- [77] C. D. Stern, "The 'Omics Revolution: How an Obsession with Compiling Lists Is Threatening the Ancient Art of Experimental Design," *BioEssays*, vol. 1900168, pp. 1–4, 2019.
- [78] M. B. Miller and Y. W. Tang, "Basic concepts of microarrays and potential applications in clinical microbiology," *Clin. Microbiol. Rev.*, vol. 22, no. 4, pp. 611–633, 2009.
- [79] I. Barbulovic-Nad, M. Lucente, Y. Sun, M. Zhang, A. R. Wheeler, and M. Bussmann, "Bio-microarray fabrication techniques - A review," *Crit. Rev. Biotechnol.*, vol. 26, no. 4, pp. 237–259, 2006.
- [80] S. Mangul *et al.*, "How bioinformatics and open data can boost basic science in countries and universities with limited resources," *Nat. Biotechnol.*, vol. 37, no. 3, pp. 324–326, 2019.
- [81] S. Balagopalan, S. Ashok, and K. P. Mohandas, "Gene set enrichment analysis: A knowledge- base approach for interpreting genome-wide expression profiles," *PANAS*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [82] A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel, "UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts," *PLoS Genet.*, vol. 15, no. 11, pp. 1–24, 2019.
- [83] D. Kobak and P. Berens, "The art of using t-SNE for single-cell transcriptomics," *Nat. Commun.*, vol. 10, no. 1, 2019.
- [84] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016.
- [85] T. Werner, "Bioinformatics applications for pathway analysis of microarray data," *Curr. Opin. Biotechnol.*, vol. 19, no. 1, pp. 50–54, 2008.

- [86] S. Zhang, X. S. Zhang, and L. Chen, "Biomolecular network querying: A promising approach in systems biology," *BMC Syst. Biol.*, vol. 2, pp. 1–6, 2008.
- [87] H. Janwa, S. E. Massey, J. Velev, and B. Mishra, "On the origin of biomolecular networks," *Front. Genet.*, vol. 10, no. APR, 2019.
- [88] A. Ma'ayan, "Insights into the organization of biochemical regulatory networks using graph theory analyses," *J. Biol. Chem.*, vol. 284, no. 9, pp. 5451–5455, 2009.
- [89] D. Szklarczyk, A. Santos, C. Von Mering, L. J. Jensen, P. Bork, and M. Kuhn, "STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data," *Nucleic Acids Res.*, vol. 44, no. 1, pp. 380–384, 2016.
- [90] D. Szklarczyk *et al.*, "The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible," *Nucleic Acids Res.*, vol. 45, no. 1, pp. 362–368, 2017.
- [91] R. Gam, M. Sung, and A. Prasad Pandurangan, "Experimental and Computational Approaches to Direct Cell Reprogramming: Recent Advancement and Future Challenges," *Cells*, vol. 8, no. 10, p. 1189, 2019.
- [92] L. M. Jakt, S. Moriwaki, and S. Nishikawa, "A continuum of transcriptional identities visualized by combinatorial fluorescent in situ hybridization," *Dev.*, vol. 140, no. 1, pp. 216–225, 2013.
- [93] Y. A. Medvedeva *et al.*, "EpiFactors: A comprehensive database of human epigenetic factors and complexes," *Database*, vol. 2015, pp. 1–10, 2015.
- [94] P. Van Galen *et al.*, "Reduced lymphoid lineage priming promotes human hematopoietic stem cell expansion," *Cell Stem Cell*, vol. 14, no. 1, pp. 94–106, 2014.
- [95] S. Scala and A. Aiuti, "In vivo dynamics of human hematopoietic stem cells : novel concepts and future directions," *Blood Adv.*, vol. 3, no. 12, 2019.
- [96] E. Laurenti and B. Göttgens, "From haematopoietic stem cells to complex differentiation landscapes," *Nature*, vol. 553, no. 7689, pp. 418–426, 2018.
- [97] T. Semba, R. Sammons, X. Wang, X. Xie, K. N. Dalby, and N. T. Ueno, "Concise review: JNK signaling in stem cell self-renewal and differentiation," *Int. J. Mol. Sci.*, vol. 21, no. 7, pp. 1–19, 2020.
- [98] E. Passegué, E. F. Wagner, and I. L. Weissman, "JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells," *Cell*, vol. 119, no. 3, pp. 431–443, 2004.
- [99] M. Santaguida *et al.*, "JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal," *Cancer Cell*, vol. 15, no. 4, pp. 341–352, 2010.
- [100] M. Y. Yang, T. C. Liu, J. G. Chang, P. M. Lin, and S. F. Lin, "JunB gene expression is inactivated by methylation in chronic myeloid leukemia," *Blood*, vol. 101, no. 8, pp. 3205–3211, 2003.
- [101] U. Steidl *et al.*, "Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells," *Nat. Genet.*, vol. 38, no. 11, pp. 1269–1277, 2006.
- [102] G. Q. Chen *et al.*, "Methylated metabolites of arsenic trioxide are more potent than arsenic trioxide as apoptotic but not differentiation inducers in leukemia and lymphoma cells," *Cancer Res.*, vol. 63, no. 8, pp. 1853–1859, 2003.
- [103] C. R. Clapier, J. Iwasa, B. R. Cairns, and C. L. Peterson, "Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes," *Nat. Publ. Gr.*, vol. 18, no. 7, pp. 407–422, 2017.
- [104] P. B. Becker and J. L. Workman, "Nucleosome Remodeling and Epigenetics," *Cold Spring Harb Perspect Biol*, 2013.
- [105] P. Zhang, K. Torres, X. Liu, C. Liu, and R. E. Pollock, "An Overview of Chromatin-Regulating Proteins in Cells," *Curr Protein Pept Sci*, vol. 17, no. 5, pp. 401–410, 2016.
- [106] C. Libardi *et al.*, "Epidrugs : targeting epigenetic marks in cancer treatment," *Epigenetics*,

vol. 0, no. 00, pp. 1–13, 2019.

- [107] M. Cheray and L. Bretaudeau, “Epigenetic protein complexes : the adequate candidates for the use of a new generation of epidrugs in personalized and precision medicine in cancer,” *Epigenomics*, 2019.
- [108] A. Conesa *et al.*, “A survey of best practices for RNA-seq data analysis,” *Genome Biol.*, vol. 17, no. 1, pp. 1–19, 2016.
- [109] V. Amrhein, F. Korner-Nievergelt, and T. Roth, “The earth is flat ($p > 0:05$): Significance thresholds and the crisis of unreplicable research,” *PeerJ*, vol. 2017, no. 7, pp. 1–40, 2017.
- [110] J. M. Belton, R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan, and J. Dekker, “Hi-C: A comprehensive technique to capture the conformation of genomes,” *Methods*, vol. 58, no. 3, pp. 268–276, 2012.