

## Análisis del transcriptoma mediante RNA-seq: Aplicaciones actuales y perspectivas



Oscar Ortega-Recalde, MD, PhD Profesor Asistente | Facultad de Medicina Universidad Nacional de Colombia



1



### Conflictos de Interés

No existe ningún conflicto de interés para participar en este evento.







### Contenido

- Introducción
- Flujos de análisis de RNA-seq
- Ejemplos de aplicaciones actuales
- Perspectivas

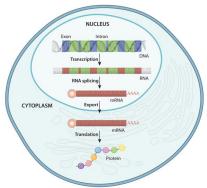


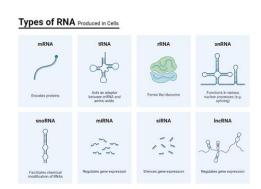
3



### Introducción

El transcriptoma constituye el **conjunto total de moléculas de ARN** presentes en una célula o grupo de células en un momento determinado.







An overview of the flow of information from DNA to protein in a eukaryote, Nature; Jabbar et al, Molecular Biology Reports 2024

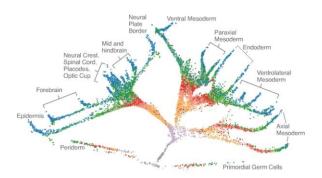
# Simposio Internacional de Genética Humana Simposio di Genómica y Medicina de Precisión en la recepción de la r

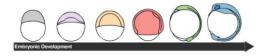
## Introducción



A diferencia del genoma, el transcriptoma es **variable** en **lugar, tiempo** y respuesta a **estímulos**.









Farrell et al, Science, 2018

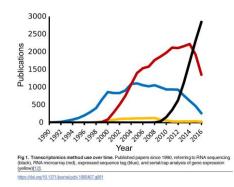
5



### Introducción

Existen diversas técnicas para el análisis de expresión génica / transcritos.

SAGE/CAGE



Method	RNA-Seq	Microarray	
Throughput	High [10]	Higher [10]	
nput RNA amount	Low ~ 1 ng total RNA [25]	High ~ 1 µg mRNA [26]	
Labour intensity	High (sample preparation and data analysis) [10][23]	Low [10][23]	
Prior knowledge	None required, though genome sequence useful [23]	Reference transcripts required for probes [23]	
Quantitation accuracy	~90% (limited by sequence coverage) [27]	>90% (limited by fluorescence detection accuracy) [27]	
Sequence resolution	Can detect <u>SNPs</u> and splice variants (limited by sequencing accuracy of ~99%) [27]	g Dedicated arrays can detect splice variants (limited by probe desig and cross-hybridisation) [27]	
Sensitivity	10 <sup>-6</sup> (limited by sequence coverage) [27]	10 <sup>-3</sup> (limited by fluorescence detection) [27]	
Dynamic range	>10 <sup>5</sup> (limited by sequence coverage) [28]	103-104 (limited by fluorescence saturation) [28]	
Technical reproducibility	>99% [29][30]	>99% [31][32]	
RNA-Seq, RNA Sequ	encing		

RNA-seq

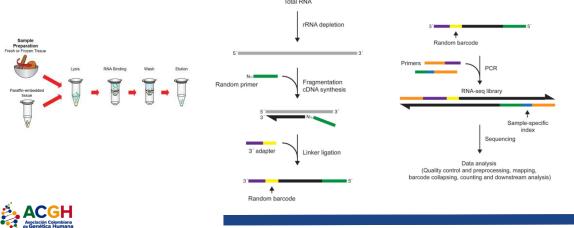


Lowe et al, PLOS Comp Biol, 2016



### Introducción

### Preparación de librerías para RNA-seq



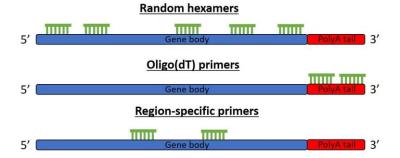
Poulsen et al, Curr Protoc Nucleic Acid Chem, 2018

7



### Introducción

### Preparación de librerías para RNA-seq





https://toptipbio.com/cdna-synthesis-primers/



### Introducción

### Secuenciación de siguiente generación y llamado de bases



IMPRITATION DISCOSSIONAL CONTROLLA DISCOSSIONAL DISCOSSIO

Plataformas de NGS

Archivos FASTQ



9



## Flujos de análisis de RNA-seq

- No existe un pipeline óptimo para todas las diferentes aplicaciones y escenarios de análisis de RNA-seq.
- Los experimentos y las estrategias de análisis deben ser adaptados dependiendo del organismo a ser estudiado y los objetivos del estudio.
- RNA-seq puede ser utilizado de manera aislada o en combinación con otros métodos de genómica funcional para analizar expresión génica.



Conesa et al, Genome Biology, 2016



### **RNA-seq: Basic Bioinformatics Analysis**

Fei Ji1,2 and Ruslan I. Sadreyev1,3,4

<sup>1</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston,

<sup>2</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts <sup>3</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School,

<sup>4</sup>Corresponding author: sadreyev@molbio.mgh.harvard.edu



Conesa et al. Genome Biology (2016) 17:13 DOI 10.1186/s13059-016-0881-8

Genome Biology

### Open Access

### A survey of best practices for RNA-seq data analysis



Ana Conesa<sup>1,28</sup>, Pedro Madrigal<sup>8,4\*</sup>, Sonia Tarazona<sup>2,5</sup>, David Gomez-Cabrero<sup>6,7,89</sup>, Alejandra Cervera<sup>10</sup>, Andrew McPherson<sup>11</sup>, Michał Wojciech Szcześniak<sup>1,2</sup>, Daniel J. Gaffney<sup>3</sup>, Laura L. Elo<sup>13</sup>, Xuegong Zhang<sup>14,15</sup> and Ali Mortazavi<sup>16,1,9\*</sup>



11

### Simposio Internacional de **Genética Humana**

## Flujos de análisis de RNA-seq

#### Preanálisis

### Core-análisis

### Análisis avanzados

- Diseño experimental
- Diseño de la secuenciación
- Control de calidad
- Perfil transcriptómico
- Expresión diferencial
- Perfil funcional
- Visualización
- Integración de datos
- Otros análisis basados en RNA-seq



Conesa et al, Genome Biology, 2016



#### Preanálisis

- Las fases de diseño experimental y de secuenciación son claves para evitar dificultades en los análisis posteriores.
- Puntos para tener en cuenta:
  - o Tipo de librería (single-end vs paired end)
  - Longitud de las lecturas
  - o Numero de replicas y numero de lecturas.
  - o Spike-in
  - Aleatorización



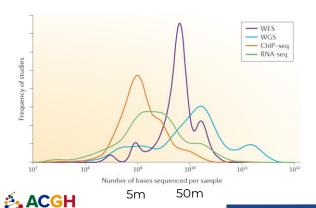
Conesa et al, Genome Biology, 2016

13



## Flujos de análisis de RNA-seq

### Número de lecturas



Los requerimientos de secuenciación dependen del tamaño, complejidad del transcriptoma y objetivo de investigación.

**30-50m** = Permite la cuantificación de genes con expresión > 10 FPKM (80% con 36m)

- > 80m = Alta profundidad, cuantificación de genes con niveles de expresión más bajas.
- > 200m = Detección del rango completo de transcritos

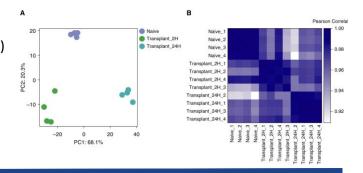
Sims et al, Nat Rev Genet, 2014



### Control de calidad

Se recomienda realizarse en los diferentes niveles de procesamiento

- Lecturas crudas
- Alineamiento de lecturas
- Cuantificación
- Reproducibilidad(e.g. R2 > 0.9)





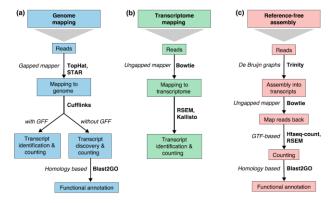
Koch et al, Translational Review, 2018

15

# Simposio Internacional de Genética Humana Simposio de Genómica y Medicina de Precisión en la Región Andrea "Laboratorio-Investigación-Clinica"

### Flujos de análisis de RNA-seq

### Estrategias de mapeo e identificación de transcritos





Conesa et al, Genome Biology, 2016??



### Cuantificación de transcritos

#### Matriz de conteo

Gene Name	Rep1 Counts	Rep2 Counts	Rep3 Counts
A (2kb)	10	12	30
B (4kb)	20	25	60
C (1kb)	5	8	15
D (10kb)	0	0	1

RPKM (Reads per kilo base per million mapped reads)

Gene Name	Rep1 RPKM	Rep2 RPKM	Rep3 RPKM
A (2kb)	1.43	1.33	1.42
B (4kb)	1.43	1.39	1.42
C (1kb)	1.43	1.78	1.42
D (10kb)	0	0	0.009

### RPM (Reads per million mapped reads)

Gene Name	Rep1 RPM	Rep2 RPM	Rep3 RPM
A (2kb)	2.86	2.67	2.83
B (4kb)	5.71	5.56	5.66
C (1kb)	1.43	1.78	1.42
D (10kb)	0	0	0.09

FPKM (Fragments per kilo base per million mapped reads) – PE

https://www.youtube.com/watch?v=TTUrtCY2k-w



17

## Flujos de análisis de RNA-seq



### Análisis de expresión diferencial

### ${\bf Table\,4.\,\,RNA\text{-}Seq\,differential\,gene\,expression\,software.}$

Software	Environment	Specialisation
Cuffdiff2 [111]	Unix-based	Transcript analysis at isoform-level
EdgeR [112]	R/Bioconductor	Any count-based genomic data
DEseq2 [113]	R/Bioconductor	Flexible data types, low replication
Limma/Voom [114]	R/Bioconductor	Microarray or RNA-Seq data, isoform analysis, flexible experiment design

RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t004

La normalización es crítica en el proceso de análisis bioinformático





### Análisis de expresión diferencial

Love et al. Genome Biology (2014) 15:550 DOI 10.1186/s13059-014-0550-8



METHOD

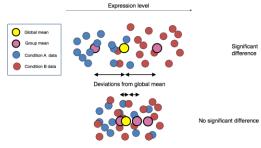
Open Access

## Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2

Michael I Love<sup>1,2,3</sup>, Wolfgang Huber<sup>2</sup> and Simon Anders<sup>2\*</sup>

#### Abstract

In comparative high-throughput sequencing assays, a fundamental task is the analysis of count data, such as read counts per gene in RNA-seq, for evidence of systematic changes across experimental conditions. Small replicate numbers, discreteness, large dynamic range and the presence of outliers require a suitable statistical approach. We present DESeq2, a method for differential analysis of count data, using shrinkage estimation for dispersions and folid changes to improve stability and interpretability of estimates. This enables a more quantitative analysis focused on the strength rather than the mere presence of differential expression. The DESeq2 package is available at http://www.bioconductor.org/packages/release/bioc/htm/DESeq2.html.



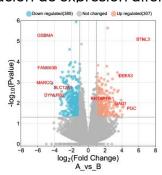


19

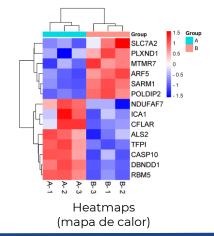
## Flujos de análisis de RNA-seq



### Visualización de expresión diferencial



Volcano plots



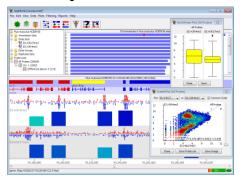
ACGH

Asociación Colombiana de Genética Humana

https://www.bioinformatics.com.cn/plot\_basic\_3\_color\_volcano\_plot\_086\_en



### Visualización y análisis





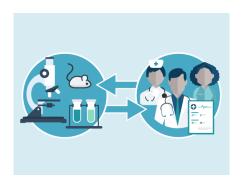
SeqMonk

IGV



21





¿Qué utilidad tienen estas técnicas en investigación biomédica y contexto clínico?





## Ejemplos de aplicaciones actuales

Análisis basados en RNA-seq son actualmente utilizados en el **contexto biomédico y clínico** para mejorar nuestra comprensión de los fenotipos.

### "Fenotipos transcriptómicos"

- Detección de expresión diferencial
- Detección de SNVs
- Detección de fusiones génicas
- Detección de expresión aberrante
- Expresión monoalelica
- Detección de splicing aberrante



Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022

23



## Ejemplos de aplicaciones actuales

GG25\_Arr02\_Montgomery ARjus.ds January 17,20

Received: 1 Airy 2021 | Revised: 15 May 2022 | Accepted: 25 May 2022

DOI: 10.10002/harms.24416

SPECIAL ARTICLE

Guidelines for clinical interpretation of variant pathogenicity
using RNA phenotypes

Dmitrii Smirnov<sup>1,2</sup> | Lea D. Schlieben<sup>1,2</sup> | Fatemeh Peymani<sup>1,2</sup> |

Annual Review of Genomics and Human Genetics RNA Sequencing in Disease Diagnosis

Craig Smail<sup>1</sup> and Stephen B. Montgomery<sup>2</sup>

<sup>1</sup>Genomic Medicine Center, Children's Mercy Research Institute, Children's Mercy Kanus Cray,
Kanus Gray, Missoni, USA; emil comil@mis.edu

<sup>1</sup>Department of Benedical Dana Science, Peparament of Genetics, and Department of Pepalology, Stanford University School of Medicins, Stanford, Chilfornia, USA;

multi-musenger@missned.edu.

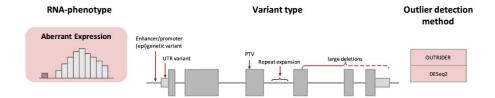


Riccardo Berutti<sup>1,2</sup> | Holger Prokisch<sup>1,2</sup> (0)



## Detección de expresión aberrante

Expresión génica por fuera del rango fisiológico, a menudo asociada con niveles bajos de expresión génica.



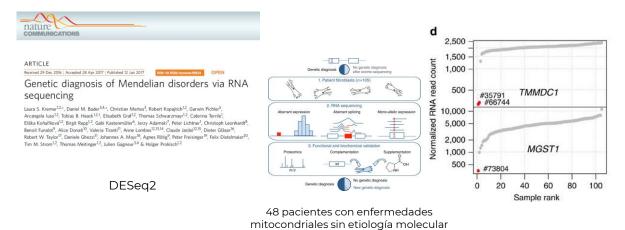


Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022

25



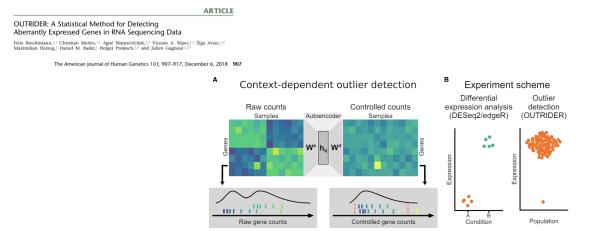
## Detección de expresión aberrante



ACGH
Asociación Colombiana
de Genética Humana



## Detección de expresión aberrante



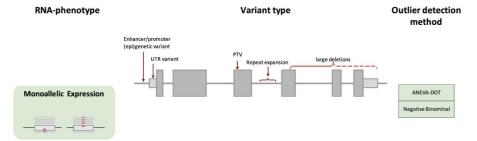


27



### Expresión monoalélica

Condición en la cual solo uno de los alelos es expresado principalmente (>80%). Puede ser resultado de silenciamiento epigenético, variantes en promotores o que un alelo sea degradado.

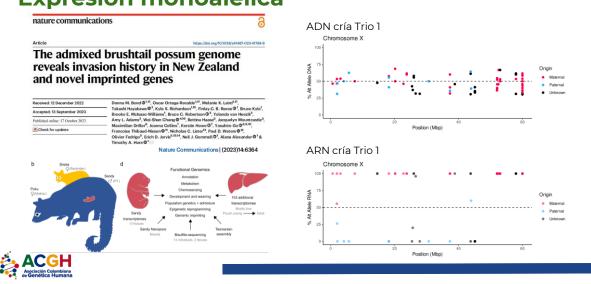




Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022



## Expresión monoalélica



29

# Simposio Internacional de Genética Humana Simposio de Genética y Medicina de Precisión en la Simposio de Carlo de Carlo

## Expresión monoalélica

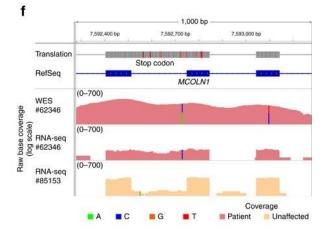
ARTICLE

\*\*Received 29 Dec 20% | Accepted 28 Apr 2007 | Published 12 Jan 2007

\*\*Genetic diagnosis of Mendelian disorders via RNA sequencing

Laura S. Kiemer<sup>1,2+</sup>, Daniel M. Bader<sup>1,4+</sup>, Christian Mertes<sup>2</sup>, Robert Kopajitch<sup>1,2</sup>, Garwin Fichler<sup>5</sup>, Acceptate lauxo<sup>2</sup>, Tolkas B. Hadack<sup>1,2+</sup>, Elisabeth Graft<sup>2</sup>, Thomas Schwarzmay<sup>1,2</sup>, Caterina Terrile<sup>1</sup>, Elisabeth Scholarikoval<sup>2</sup>, Bright Ropel<sup>3</sup>, Caterina Christian Mertes<sup>3</sup>, Robert Kopajitch<sup>1,2</sup>, Garwin Fichler<sup>5</sup>, Recognita Isack<sup>1,2</sup>, Alex Daniel<sup>3</sup>, Visite in Terrile<sup>1</sup>, Elisabeth Graft<sup>2,3</sup>, Tolkas B. Hadack<sup>1,2,3</sup>, Detection College (College Proprint College Proprin

Test de distribución binomial negativa



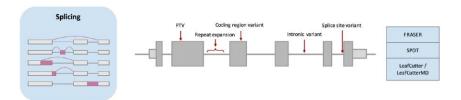




## Detección de splicing aberrante

Las alteraciones de splicing son una causa conocida e importante de patologías humanas hereditarias (15%).

Técnicas basadas en RNA-seq pueden permitir la identificación y cuantificación del defecto subyacente.



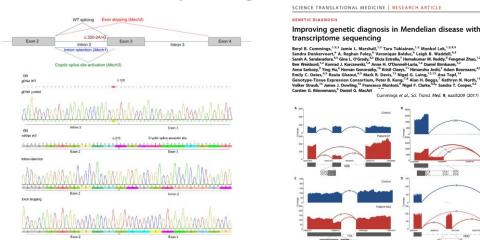


Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022

31

#### Simposio Internacional de Genética Humana Simposio de Genérica y Medicina de Precisión es la Región Ardine: "aboratorio fineta"

## Detección de splicing aberrante

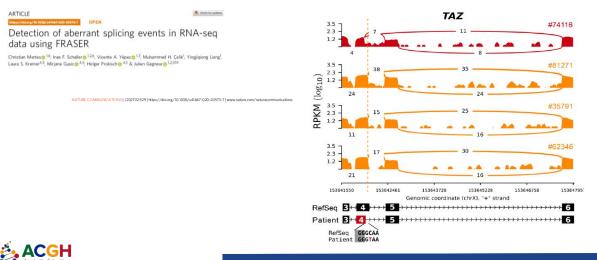




Ortega-Recalde et al, Clinical Dermatology, 2016



## Detección de splicing aberrante



33



## **Perspectivas**

El desarrollo de **nuevos métodos y aplicaciones** basadas en RNA-seq es un área activa de investigación biomédica.

- Single cell RNA-seq
- Transcriptómica espacial
- Secuenciación de ARN nativo
- Secuenciación de ARN no codificante
- Integración a ciencias multiómicas

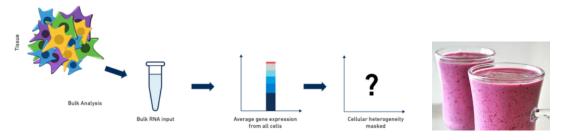


34



## Single-cell RNA-seq

¿Por qué es útil la secuenciación de células únicas?





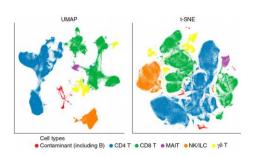
https://www.10xgenomics.com/blog/single-cell-rna-seq-an-introductory-overview-and-tools-for-getting-started and the sequence of the sequence

35

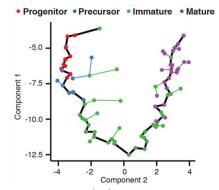


## Single-cell RNA-seq

### Análisis de datos



Reducción de la dimensionalidad (e.g. UMAP, t-SNE)



Pseudotiempos (análisis de trayectorias celulares)



https://cole-trapnell-lab.github.io/projects/sc-trajectories/



## Single-cell RNA-seq

nature reviews genetics

Perspective

Transitioning single-cell genomics into the clinic

Jennifer Lim 

→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→ Lim 
→

Nature Reviews Genetics | Volume 24 | August 2023 | 573-584

#### nature medicine

Perspective

tne://doi.org/10.1038/e41591-022-02104-7

### Impact of the Human Cell Atlas on medicine

Received: 24 August 2022

Accepted: 24 October 2022
Published online: 8 December 2022

Check for undates

Jennifer E. Rood **©** <sup>1,4</sup>, Aidan Maartens<sup>2,4</sup>, Anna Hupalowska<sup>1</sup>, Sarah A. Teichmann **©** <sup>2,3</sup> ⊠ & Aviv Regev **©** <sup>1</sup> ⊠

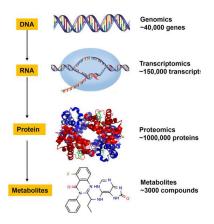
Single-cell atlases promise to provide a 'missing link' between genes, diseases and therapies. By identifying the specific cell types, states, programs and contexts where disease-implicated genes act, we will understand the mechanisms of disease at the cellular and tissue levels and can use this understanding to develop powerful disease diagnostics; identify promising new drug targets; predict their efficacy, toxicity and resistance mechanisms; and empower new kinds of therapies, from cancer therapies to regenerative medicine. Here, we lay out a vision for the potential of cell atlases to impact the future of medicine, and describe how advances over the past decade have begun to realize this potential in common complex diseases, infectious diseases (including COVID-19), rare diseases and cancer.

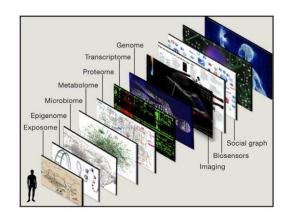


37



## Integración a ciencias multiomicas



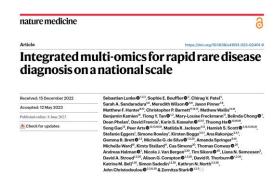




Yu et al, Oncotarget, 2017; Topol, Cell, 2014;



## Integración a ciencias multiomicas



290 pacientes reclutados WGS 153 sin Dx -> 19 nuevos Dx (12,9%)



1204 pacientes reclutados metilación DNA + panel DNA-seq + RNA-seq



39



## Mensajes principales

- Las técnicas de secuenciación de ARN por métodos de siguiente generación (RNA-seq) son metodologías robustas y sensibles para evaluar el transcriptoma.
- No existe un flujo de trabajo bioinformático único para todos los escenarios, es necesario establecer dicho flujo y adoptar mejores prácticas de acuerdo con los objetivos del estudio.
- RNA-seq es una tecnología madura que puede ser utilizada en diferentes escenarios clínicos y de investigación biomédica.
- Existen **numerosas áreas de investigación y desarrollo** dentro de este grupo de técnicas. Es de esperar que su implementación sea mayor en el futuro.





## **Agradecimientos**



### Universidad Nacional de Colombia

Mauricio Rey Buitrago William Usaquen María Paula Meléndez-Flórez Erika Sofia Torres-Narvaez



#### Instituto Nacional de Cancerología

Rafael Parra-Medina Julian Riaño-Moreno William Torres-Jara



#### Universidad del Rosario

Dora Janeth Fonseca-Mendoza Nora Constanza Contreras Bravo Adrien Morel Carlos M. Restrepo



#### **University of Otago**

Timothy A. Hore Alana Alexander Melanie K. Laird Donna M. Bond

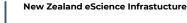


Universidad Pedagógica y Tecnológica de Colombia

Milena Rondón-Lagos



Vertebrate Genome Project





41







# iGRACIAS!

Correo: oortegar@unal.edu.co



Asociación Colombiana de Genética Humana - ACGH





(©) @asociacion\_acgh



**ACGH** 



@asociacion\_acgh