

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

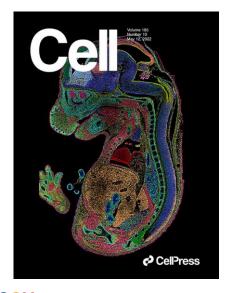


No existe ningún conflicto de interés para participar en este evento.









"Computers do not just scale up biology... they bring with them completely new tools and questions" Stevens Hallam



Chen et al, Cell, 2022

3



Contenido

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

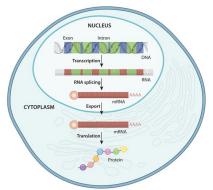


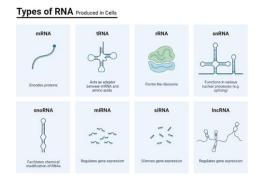
4



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

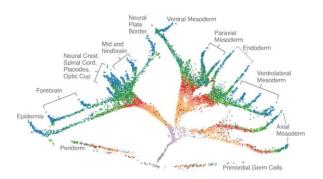
5

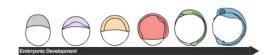
Introducción



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









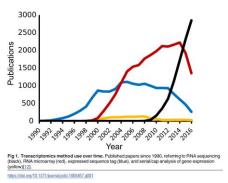
Farrell et al, Science, 2018

Simposio Internacional de **Genética Humana**



Introducción

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



Method	RNA-Seq	Microarray
Throughput	High [10]	Higher [10]
Input 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]	Dedicated arrays can detect splice variants (limited by probe design and cross-hybridisation) [27]
Sensitivity	10 ⁻⁶ (limited by sequence coverage) [27]	10 ⁻³ (limited by fluorescence detection) [27]
Dynamic range	>10 ⁵ (limited by sequence coverage) [28]	103-104 (limited by fluorescence saturation) [28]
Technical reproducibility	>99% [29][30]	>99% [31][32]

EST

RNA microarrays







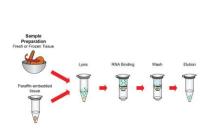
Lowe et al, PLOS Comp Biol, 2016

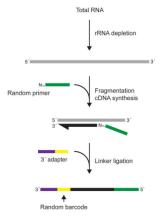
7

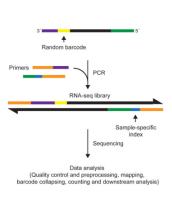
Introducción

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

Preparación de librerías para RNA-seq







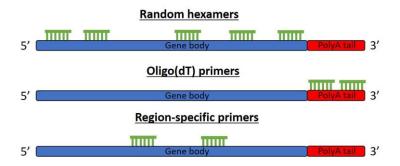


Poulsen et al, Curr Protoc Nucleic Acid Chem, 2018

Simposio Internacional de Genética Humana Simposio de Genética y Médicia de Puecidio es la Región Andre: "Laboratorio hevetigación Clinica"

Introducción

Preparación de librerías para RNA-seq





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

9



Introducción

Secuenciación de siguiente generación y llamado de bases



gillard 11:19 1000000000 CONTROL TITLE 11:181:1931:1792 11:181:133

SIBBORT PERBOCOCCOCCO HENCE CONTROL THAT COLUMN TO CONTROL THAT COLUMN TO CONTROL THAT COLUMN TO C

Plataformas de NGS







- 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

11



Flujos de análisis de RNA-seq

RNA-seq: Basic Bioinformatics Analysis

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

¹Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts

²Department of Genetics, Harvard Medical School, Boston, Massachusetts ³Department of Pathology, Massachusetts General Hospital and Harvard Medical School,

⁴Corresponding author: sadreyev@molbio.mgh.harvard.edu



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

Genome Biology

REVIEW

Open Access

A survey of best practices for RNA-seq data analysis



Ana Conesa^{1,2*}, Pedro Madrigal^{8,4*}, Sonia Tarazona^{2,5}, David Gomez-Cabrero^{6,7,8}, Alejandra Cervera¹⁰, Andrew McPherson¹¹, Michał Wojciech Szcześniak^{1,2}, Daniel J. Gaffney³, Laura L. Elo¹³, Xuegong Zhang^{14,15} and Ali Mortazavi^{16,1,*}





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

13



Flujos de análisis de RNA-seq

Preanálisis

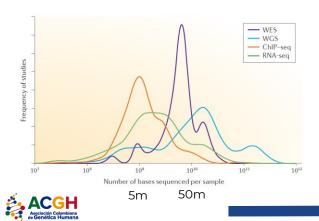
- Las fases de diseño experimental y de secuenciación son claves para evitar dificultades en los análisis posteriores.
- Puntos para tener en cuenta:
 - Tipo de librería (single-end vs paired end)
 - Longitud de las lecturas
 - o Numero de replicas y numero de lecturas.
 - Spike-in
 - Aleatorización



Conesa et al, Genome Biology, 2016



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

15

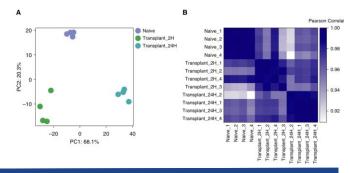


Flujos de análisis de RNA-seq

Control de calidad

Se recomienda realizarse en los diferentes niveles de procesamiento

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

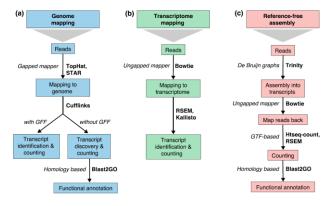




Koch et al, Translational Review, 2018



Estrategias de mapeo e identificación de transcritos





Conesa et al, Genome Biology, 2016??

17



Flujos de análisis de RNA-seq

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

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

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

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

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





Análisis de expresión diferencial

Table 4. RNA-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



19



Flujos de análisis de RNA-seq

Análisis de expresión diferencial

DOI 10.1186/s13059-014-0550-8



METHOD

Open Access

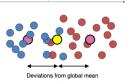
Moderated estimation of fold change and dispersion for RNA-seg data with DESeg2

Michael I Love^{1,2,3}, Wolfgang Huber² and Simon Anders^{2*}

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 fold 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/htm1/DESeq2.html.





Expression level

Significant difference

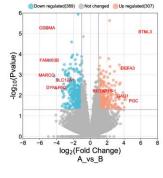


No significant difference

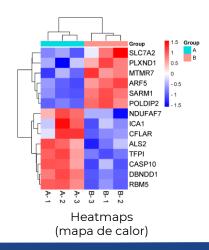








Volcano plots





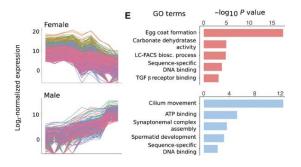
https://www.bioinformatics.com.cn/plot_basic_3_color_volcano_plot_086_en

21

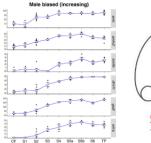
Flujos de análisis de RNA-seq

Simposio Internacional de Genética Humana Simposio de Genómica y Medicina de Precisión en la Simposio de Genómica y Medicina de Precisión en la Simposio Administración Cinicia (1980)

Perfiles funcionales



Co-expresión y Enriquecimiento funcional



Análisis de vías

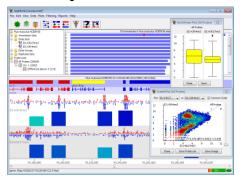
moleculares



Todd et al, Science Advances, 2019



Visualización y análisis





SeqMonk

IGV



23





¿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

25



Ejemplos de aplicaciones actuales

Received: 1. July 2021 | Revised: 15 May 2022 | Accepted: 25 May 2022 |
DOI: 10.1002/harm.24416

SPECIAL ARTICLE

Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes

Dmitrii Smirnov^{1,2} | Lea D. Schlieben^{1,2} | Fatemeh Peymani^{1,2} |

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

Craig Smail¹ and Stephen B. Montgomery²

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

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

mmil strongarous Memoral Code In

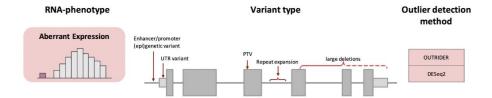


Riccardo Berutti^{1,2} | Holger Prokisch^{1,2} (0)



Detección de expresión aberrante

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



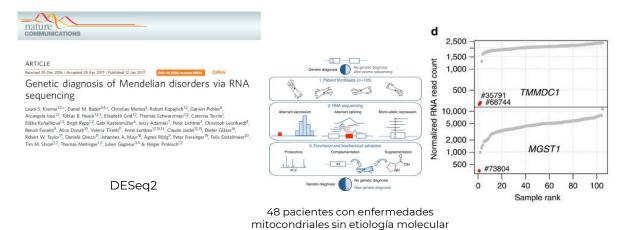


Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022

27



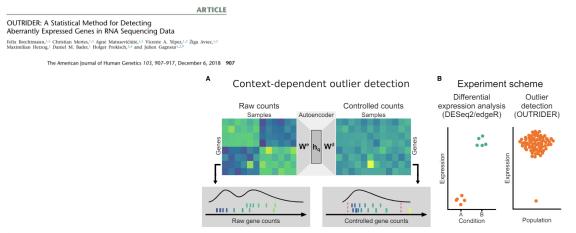
Detección de expresión aberrante



ACGH
Asociación Colombiana
de Genética Humana



Detección de expresión aberrante



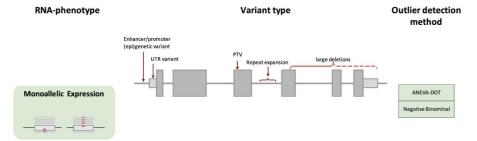


29



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.

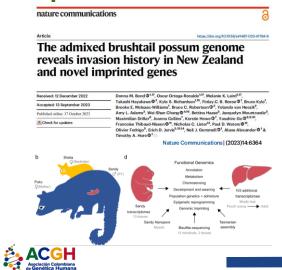


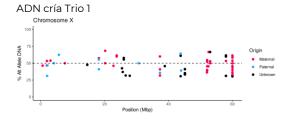


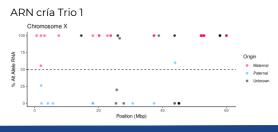
Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022



Expresión monoalélica







31

Simposio Internacional de Genética Humana

Expresión monoalélica

ARTICLE

Bearent 27 Dec 2006. Accepted 29 Apr 2007 (Paldobed 27 as 2007)

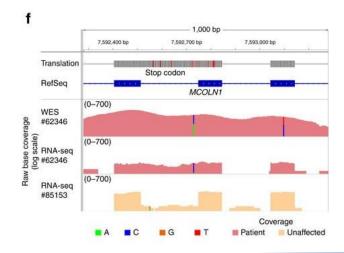
Genetic diagnosis of Mendelian disorders via RNA sequencing

Lau 5. Korema^{1,5}. Daniel M. Bader^{1,4}. Christian Mortes². Robert Kopglich^{1,5}. Carvini Pichler³.

Bibla Kohalimola³. Brigt Rept³. Galt Kattermüler³. Perox Ademail³. Perox Lobert Constanting C

Test de distribución binomial negativa



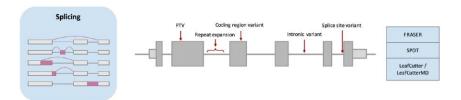




Detección de splicing aberrante

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.



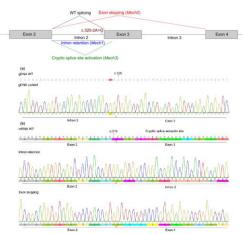


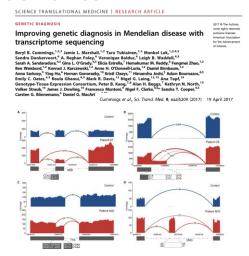
Peymani et al, Pediat Investig, 2022; Smirnov et al, Human Mutation, 2022

33

Simposio Internacional de Genética Humana Simposio de Genérica y Medicina de Precióle ne la Región Ardine: "aboratorie finestra"

Detección de splicing aberrante



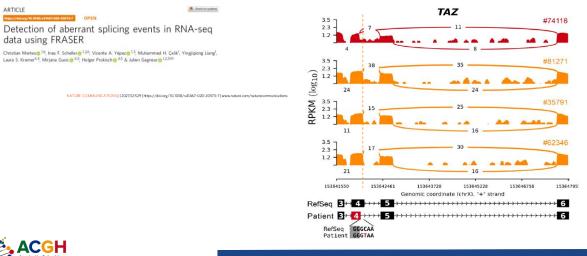




Ortega-Recalde et al, Clinical Dermatology, 2016



Detección de splicing aberrante





35



Perspectivas

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

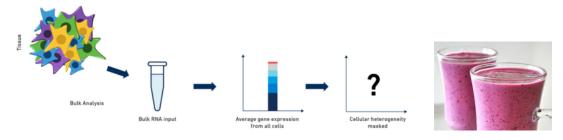
- Single cell RNA-seq
- Transcriptómica espacial
- Secuenciación de ARN nativo
- Secuenciación de ARN no codificante
- Integración a ciencias multiómicas





Single-cell RNA-seq

¿Por qué es útil la secuenciación de células únicas?





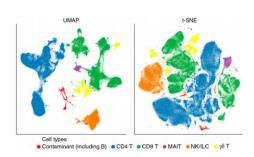
https://www.10xgenomics.com/blog/single-cell-rna-seq-an-introductory-overview-and-tools-for-getting-started and the sequence of the sequence

37

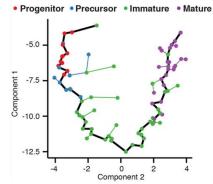


Single-cell RNA-seq

Análisis de datos



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



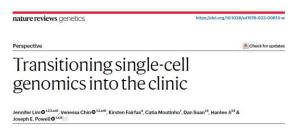
Pseudotiempos (análisis de trayectorias celulares)



 $https:\!/\!/cole\text{-trapnell-lab.github.io/projects/sc-trajectories/}$



Single-cell RNA-seq



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

Jennifer E. Rood © ¹⁴, Aidan Maartens²⁴, Anna Hupalowska¹, Sarah A. Teichmann © ²³ ⊠ & Aviv Regev © ¹ ⊠

Published online: 8 December 2022

Check for updates

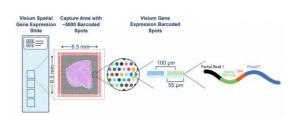
Single-cell atlases promise to provide a 'missing link' between genes, diseases and therapies. By identifying the specific cell types, states, programs and contexts where disease-implicated genes act, we will understand the mechanisms of disease at the cellular and tissue levels and can use this understanding 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.



39

Simposio Internacional de Genética Humana

Transcriptómica espacial



Spatial multi-omic map of human myocardial infarction

https://doi.org/10.1038/s41586-022-05060-x Received: 30 November 2020 Accepted: 29 June 2022 Published online: 10 August 2022 Christoph Steppel": Ricorde C. Barriers From "1" " (Bland I.1" Steppel")

Ricorde T. Kermer "1, "San Leit D. Missel"; A brown "1" (bland I.1" Steppel")

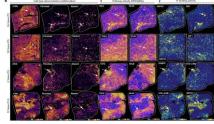
Ricorde T. Kermer "1, "San Leit D. Missel"; A brown "1" (brown Travel")

Ricorde T. Kermer "1, "San Leit D. Missel"; A brown "1" (brown Travel")

Ricorde T. Missel Steppel "1, "March T. Missel"; A brown T. Missel Steppel "1, "Missel Steppel"; A brown T. Gigler I. Missel Steppel "1, "Missel Steppel"; A brown T. Missel Missel"; Turpey Stermal "1, "Yacolan Xil,"

Ricorde T. Missel"; A brown "1, "Missel Missel"; Turpey Stermal "1, "Missel Missel"; Missel Missel"; Turpey Stermal "1, "Missel Missel"; Missel Missel"; A brown T. Missel Missel Missel"; James H. Missel Missel Steppel Missel Mis

766 | Nature | Vol 608 | 25 August 2022

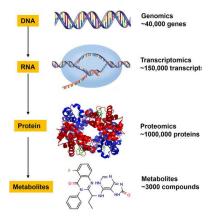


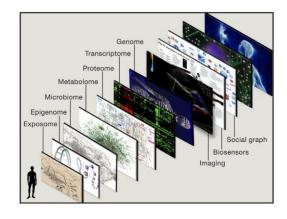


Kuppe et al, Nature, 2020



Integración a ciencias multiomicas





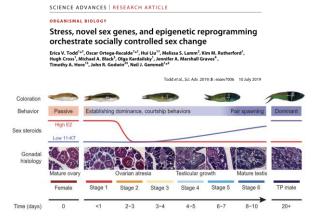


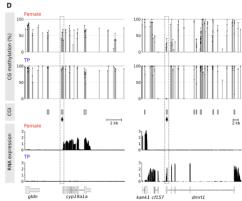
Yu et al, Oncotarget, 2017; Topol, Cell, 2014;

41

Simposio Internacional de Genética Humana Simposio de Genérica y Medicina de Pección en la Resión Andre "Uboración e Mesción a M

Integración a ciencias multiomicas

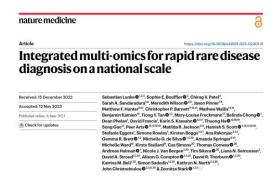








Integración a ciencias multiomicas



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



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



43



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





45







iGRACIAS!

Correo: oortegar@unal.edu.co



Asociación Colombiana de Genética Humana - ACGH





(©) @asociacion_acgh



ACGH



@asociacion_acgh