

PRINCIPIOS BÁSICOS DE LAS CIENCIAS ÓMICAS Y SUS TECNOLOGÍAS

Ruth Mélida Sánchez Mora



UNIVERSIDAD COLEGIO
MAYOR DE CUNDINAMARCA
SELLO EDITORIAL

PRINCIPIOS BÁSICOS
DE LAS CIENCIAS ÓMICAS
Y SUS TECNOLOGÍAS

RUTH MÉLIDA SÁNCHEZ MORA

Grupo Biotecnología y Genética UCMC
Bacterióloga y Laboratorista Clínica
Maestría en Genética Humana
Doctorado en Biotecnología
E-mail: rmsanchezm@universidadmayor.edu.co
Universidad Colegio Mayor de Cundinamarca

PRINCIPIOS BÁSICOS DE LAS CIENCIAS ÓMICAS Y SUS TECNOLOGÍAS



**UNIVERSIDAD COLEGIO
MAYOR DE CUNDINAMARCA**

SELLO EDITORIAL

Catalogación en la publicación – Biblioteca Nacional de Colombia

Sánchez Mora, Ruth Mélida, autora

Principios básicos de las ciencias ómicas y sus tecnologías / Ruth Mélida Sánchez Mora. -- Bogotá: Sello Editorial Universidad Colegio Mayor de Cundinamarca, 2025. 224 páginas 216.

Incluye glosario -- Incluye referencias bibliográficas.

ISBN 978-958-5198-46-3 -- 978-958-5198-47-0 (e-book)

1. Bioinformática - Fundamentos 2. Biología computacional - Investigaciones - Estudio de casos 3. Epidemiología clínica - Innovaciones tecnológicas

CDD: 570.285 ed. 23

CO-BoBN– a1156521

© RUTH MÉLIDA SÁNCHEZ MORA

© UNIVERSIDAD COLEGIO MAYOR DE CUNDINAMARCA

Sello Editorial Universidad Colegio Mayor de Cundinamarca

Carrera 13 No. 38- 29, Edificio San Juan, noveno piso

selloeditorial@unicolmayor.edu.co

www.unicolmayor.edu.co

ISBN: 978-958-5198-46-3

ISBN e-book: 978-958-5198-47-0

Queda prohibida la reproducción parcial o total de este libro por cualquier proceso reprográfico o fónico, especialmente por fotocopia, microfilme, offset o mimeógrafo.

Ley 23 de 1982

Corrección de estilo: Delio David Arango Navarro

Diagramación electrónica: Yaneth Guarín A.

Ilustración de portada: Tomada de Freepik

Diseño de portada: Vanessa Peña A.

Impresión: GRUPO EDITORIAL IBÁÑEZ

CONTENIDO

PREFACIO.....	17
---------------	----

CAPÍTULO 1. CIENCIAS ÓMICAS

1. INTRODUCCIÓN.....	19
2. GENERALIDADES DE LAS CIENCIAS ÓMICAS	19
2.1. GENÓMICA.....	20
2.2. TRANSCRIPTÓMICA.....	22
2.3. PROTEÓMICA.....	23
2.4. METABOLÓMICA.....	24
2.5. OTRAS ÓMICAS.....	26
2.5.1. <i>EPIGENÓMICA</i>	26
2.5.2. <i>METAGENÓMICA</i>	27
2.5.3. <i>FARMACOGENÓMICA</i>	28
3. HISTORIA DE LAS CIENCIAS ÓMICAS.....	30
3.1. PRIMEROS AVANCES EN LAS CIENCIAS ÓMICAS	30
3.1.1. <i>MAPAS GENÉTICOS BASADOS EN EL ANÁLISIS DE LIGAMIENTO</i> ...	31
3.1.2. <i>MAPEO FÍSICO</i>	32
3.2. ERA PREÓMICA.....	33
3.3. ERA ÓMICA.....	35
3.3.1. <i>PROYECTO GENOMA HUMANO</i>	35
3.3.2. <i>PROYECTO ENCODE</i>	39
3.3.3. <i>PROYECTO GENCODE (GENOME CODING)</i>	45
3.3.4. <i>PROYECTO HAPMAP</i>	48
3.3.5. <i>PROYECTO DE LOS 1.000 GENOMAS</i>	52

3.3.6.	<i>PROYECTO PROTEOMA HUMANO (HPP)</i>	55
3.3.7.	<i>PROYECTO PANGENOMAS</i>	57
3.4.	NUEVOS PROYECTOS EMERGENTES EN CIENCIAS ÓMICAS	60
3.4.1.	<i>PROYECTOS EN METAGENÓMICA</i>	60
3.4.2.	<i>PROYECTOS EN EPIGENÓMICA</i>	61
4.	ERA MULTIÓMICA.....	63
4.1.	INTEGRACIÓN DE DATOS MULTIÓMICOS Y APRENDIZAJE AUTOMÁTICO ..	65
4.2.	MINERÍA DE DATOS	69
5.	BIOINFORMÁTICA Y ANÁLISIS DE DATOS MULTIÓMICOS	69
5.1.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN GENÓMICA	70
5.2.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN TRANSCRIPTÓMICA.....	72
5.3.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN PROTEÓMICA.....	75
5.4.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN METABOLÓMICA	78
5.5.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN OTRAS ÓMICAS.....	80
6.	MEDICINA DE PRECISIÓN Y ENFOQUE ÓMICO	81
	CONCLUSIONES.....	83
	REFERENCIAS.....	84

CAPÍTULO 2. PRINCIPALES TÉCNICAS ÓMICAS

1.	INTRODUCCIÓN.....	101
2.	TÉCNICAS EN GENÓMICA	102
2.1.	SECUENCIACIÓN DE PRIMERA GENERACIÓN	104
2.1.1.	<i>TÉCNICA DE MAXAM Y GILBERT</i>	104
2.1.2.	<i>TÉCNICA DE SANGER</i>	105
2.2.	SECUENCIACIÓN DE SEGUNDA GENERACIÓN	107

2.2.1.	<i>PROCEDIMIENTO GENERAL PARA LAS TÉCNICAS DE NGS</i>	108
2.2.2.	<i>PIROSECUENCIACIÓN</i>	109
	2.2.2.1. <i>Preparación de las muestras y bibliotecas</i>	109
	2.2.2.2. <i>Amplificación y secuenciación</i>	109
	2.2.2.3. <i>Salida y análisis de datos</i>	109
2.2.3	<i>SECUENCIACIÓN POR LIGACIÓN Y DETECCIÓN DE OLIGONUCLEÓTIDOS (SOLID)</i>	111
	2.2.3.1. <i>Preparación de las muestras y bibliotecas</i>	111
	2.2.3.2. <i>Amplificación y secuenciación</i>	111
	2.2.3.3. <i>Salida y análisis de datos</i>	112
2.2.4.	<i>SECUENCIACIÓN ILLUMINA</i>	113
	2.2.4.1. <i>Preparación de la muestra y biblioteca</i>	114
	2.2.4.2. <i>Amplificación y secuenciación</i>	114
	2.2.4.3. <i>Salida y análisis de datos</i>	115
2.2.5.	<i>SECUENCIACIÓN DE NANOESFERAS DE ADN</i>	116
	2.2.5.1. <i>Preparación de la muestra y amplificación</i>	116
	2.2.5.2. <i>Amplificación y secuenciación</i>	117
	2.2.5.3. <i>Salida y análisis de datos</i>	117
2.2.6.	<i>PRINCIPALES LIMITACIONES DE LAS TÉCNICAS DE SEGUNDA GENERACIÓN</i>	118
	2.2.6.1. <i>Ion Torrent</i>	120
	2.2.6.2. <i>Limitaciones de la secuenciación de tercera generación</i>	121
2.3.	<i>SECUENCIACIÓN DE TERCERA GENERACIÓN</i>	122
	2.3.1. <i>HELISCOPE</i>	123
	2.3.2. <i>NANOPORE</i>	125
2.4.	<i>SECUENCIACIÓN DE CUARTA GENERACIÓN</i>	128
	2.4.1. <i>SECUENCIACIÓN DE MOLÉCULA ÚNICA EN TIEMPO REAL</i>	128
	2.4.2. <i>LA SECUENCIACIÓN POR CAPTURA DE IMAGEN (ICS, POR SUS SIGLAS EN INGLÉS)</i>	130
	2.4.3. <i>LIMITACIONES DE LAS TÉCNICAS DE SECUENCIACIÓN DE CUARTA GENERACIÓN</i>	131

2.5.	MICROARREGLOS	133
3.	TÉCNICAS EN TRANSCRIPTÓMICA	135
3.1.	ENFOQUES BASADOS EN HIBRIDACIÓN	137
3.1.1.	<i>ARN-FISH</i>	139
3.1.2.	<i>NORTHERN BLOT</i>	139
3.1.3.	<i>ARN FISH USANDO AMPLIFICACIÓN DE SEÑAL DE ADN RAMIFICADO</i>	140
3.1.4.	<i>MICROARREGLOS DE ADN (ADN MICROARREGLOS)</i>	142
3.1.4.1.	<i>Microarreglos de oligonucleótidos</i>	142
3.1.4.2.	<i>Microarreglos de ADN de expresión genética</i>	143
3.1.4.3.	<i>Microarreglos de ADNc</i>	143
3.2.	ENFOQUES BASADOS EN SECUENCIAS.....	146
3.2.1.	<i>ANÁLISIS EN SERIE DE LA EXPRESIÓN GÉNICA (SAGE)</i>	146
3.2.2.	<i>ANÁLISIS CAP DE LA EXPRESIÓN GÉNICA (CAGE)</i>	146
3.2.3.	<i>SECUENCIACIÓN MASIVA DE FIRMAS EN PARALELO (MPSS)</i>	147
3.3	ARN-SEQ.....	149
3.3.1.	<i>TECNOLOGÍA ARN-SEQ</i>	150
3.4.	ETIQUETADO Y MULTIPLEXACIÓN DE UNA SOLA CÉLULA MEDIANTE SECUENCIACIÓN /SLAM-SEQ.....	154
3.5.	scARN-SEQ.....	155
3.6.	TÉCNICAS PARA CUANTIFICACIÓN DE LA TRANSCRIPCIÓN	159
3.6.1.	<i>PROTOCOLOS DE SECUENCIACIÓN DE CÉLULAS INDIVIDUALES</i>	160
3.6.2.	<i>PROTOCOLOS BASADOS EN DISPOSITIVOS DE MICROFLUIDOS</i> ...	160
3.6.3.	<i>PROTOCOLOS BASADOS EN PLACA</i>	161
3.7.	TRANSCRIPTÓMICA ESPACIAL	161
4.	TÉCNICAS UTILIZADAS EN PROTEÓMICA.....	165
4.1.	CROMATOGRAFÍA	168
4.1.1.	<i>CROMATOGRAFÍA EN PAPEL</i>	169
4.1.2.	<i>CROMATOGRAFÍA EN CAPA FINA</i>	170
4.1.3.	<i>CROMATOGRAFÍA EN COLUMNA</i>	170
4.1.4.	<i>CROMATOGRAFÍA DE GASES</i>	170

4.1.5.	<i>CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN</i>	171
4.2.	ELECTROFORESIS.....	171
4.2.1.	<i>EN GEL DE POLIACRILAMIDA (SDS-PAGE)</i>	172
4.2.2.	<i>ELECTROFORESIS CAPILAR</i>	172
4.2.3.	<i>ELECTROFORESIS BIDIMENSIONAL</i>	173
4.3.	WESTERN BLOTTING (INMUNOTRANSFERENCIA).....	174
4.4.	ESPECTROMETRÍA	174
4.4.1.	<i>ESPECTRÓMETRO DE MASAS</i>	175
4.4.2.	<i>ESPECTROMETRÍA DE MASAS BASADA EN TIEMPO DE VUELO (TOF-MS)</i>	176
4.4.3.	<i>ESPECTROMETRÍA DE MASAS EN TÁNDEM (MS/MS O MS2)</i>	178
4.4.4.	<i>ESPECTROMETRÍA DE MASAS DE RELACIÓN ISOTÓPICA (IRMS)</i>	180
4.4.5.	<i>ESPECTROMETRÍA DE MASAS EN TÁNDEM DE CAPTURA DE ELECTRONES (ECD-MS/MS)</i>	180
4.4.6.	<i>CROMATOGRAFÍA LÍQUIDA ACOPLADA A ESPECTROMETRÍA DE MASAS (LC-MS)</i>	180
4.4.7.	<i>ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN (HR-MS)</i>	180
4.5.	PROTEÓMICA CUANTITATIVA	180
4.5.1.	<i>PROTEÓMICA CUANTITATIVA BASADA EN ETIQUETADO</i>	181
4.5.2.	<i>PROTEÓMICA SIN ETIQUETADO (LABEL-FREE)</i>	181
4.6.	MÉTODOS PEPTIDÓMICOS	181
4.6.1.	<i>ELECTROFORESIS CAPILAR Y ESPECTROMETRÍA DE MASAS (CE-MS)</i>	181
5.	TÉCNICAS EN METABOLÓMICA	182
5.1.	MÉTODOS DE SEPARACIÓN Y ESPECTROMETRÍAS	184
5.1.1.	<i>CROMATOGRAFÍA DE GASES-ESPECTROMETRÍA DE MASAS (GC-MS)</i>	184
5.1.2.	<i>ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN</i>	185
5.1.3.	<i>ESPECTROMETRÍA DE MASAS BASADA EN IONIZACIÓN POR ELECTROSPRAY (ESI-MS)</i>	185
5.1.4.	<i>ESPECTROMETRÍA DE MASAS BASADA EN IONIZACIÓN POR DESORCIÓN LÁSER ASISTIDA POR MATRIZ (MALDI-MS)</i>	185

5.1.5.	<i>ESPECTROMETRÍA DE MASAS DE REACCIÓN EN CADENA (MS/MS) ..</i>	185
5.1.6.	<i>ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN EN TÁNDEM (HRMS/MS)</i>	186
5.2.	RESONANCIA MAGNÉTICA NUCLEAR (RMN)	186
5.3.	ESPECTROSCOPIA INFRARROJA POR TRANSFORMADA DE FOURIER.....	190
5.4.	ESPECTROSCOPIA RAMAN	190
6.	HERRAMIENTAS DE OTRAS CIENCIAS ÓMICAS	192
6.1.	LIPIDÓMICA.....	192
6.1.1.	<i>LA CROMATOGRAFÍA DE FASE REVERSA (RPLC).....</i>	193
6.1.2.	<i>ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN (HRMS)...</i>	194
6.1.3.	<i>CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN (HPLC)</i>	195
6.1.4.	<i>ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS).....</i>	195
6.2.	GLICÓMICA.....	196
6.2.1.	<i>CROMATOGRAFÍA DE EXCLUSIÓN MOLECULAR</i>	196
6.2.2.	<i>RMN DE PROTONES</i>	197
6.2.3.	<i>RMN DE CARBONO-13.....</i>	197
6.2.4.	<i>TÉCNICAS COMPLEMENTARIAS</i>	197
6.3.	EPIGENÓMICA.....	198
6.3.1.	<i>BISULFITO DE SODIO (BS).....</i>	198
6.3.2.	<i>CROMATINA DE INMUNOPRECIPITACIÓN (CHIP)</i>	199
6.3.3.	<i>MICROARREGLOS EPIGENÉTICOS.....</i>	201
	CONCLUSIONES.....	202
	REFERENCIAS.....	203
	GLOSARIO.....	213

LISTA DE FIGURAS

Figura 1.	<i>Principales ciencias ómicas</i>	20
Figura 2.	<i>Primer mapa genético de Alfred Sturtevant.....</i>	31
Figura 3.	<i>Mapas genéticos.....</i>	32
Figura 4.	<i>Mapas físicos.....</i>	33
Figura 5.	<i>Era preómica.....</i>	34
Figura 6.	<i>Proyecto genoma humano</i>	36
Figura 7.	<i>Estrategia del consorcio PGH.....</i>	37
Figura 8.	<i>Estrategia de Celera Genomic PGH</i>	38
Figura 9.	<i>Proyecto ENCODE</i>	40
Figura 10.	<i>Alcanzando la estructura del genoma humano.....</i>	41
Figura 11.	<i>Nivel integrador enciclopedia ENCODE</i>	43
Figura 12.	<i>Enciclopedia ENCODE nivel de suelo</i>	43
Figura 13.	<i>Enciclopedia ENCODE SCREEN.....</i>	44
Figura 14.	<i>La canalización GENCODE</i>	46
Figura 15.	<i>SNP, Haplotipos y Tags SNP</i>	49
Figura 16.	<i>Proyecto 1.000 Genomas</i>	53
Figura 17.	<i>Base de datos Proyecto Proteoma Humano.....</i>	56
Figura 18.	<i>Proyecto Pangenomas.....</i>	58
Figura 19.	<i>Integración de datos ómicos y enfoques de la enfermedad</i>	64
Figura 20.	<i>Representación esquemática de las principales estrategias para la integración de datasets multiómicos</i>	67
Figura 21.	<i>Evolución de la medicina</i>	82
Figura 22.	<i>Método de secuenciación Maxam y Gilbert.....</i>	104
Figura 23.	<i>Secuenciación enzimática Sanger</i>	106
Figura 24.	<i>Secuenciación por pirosecuenciación</i>	110

Figura 25.	<i>Secuenciación por ligación y detección de oligonucleótidos (SOLiD)</i>	113
Figura 26.	<i>Secuenciación Illumina.....</i>	115
Figura 27.	<i>Secuenciación por nanoesferas</i>	118
Figura 28.	<i>Secuenciación por Ion Torrent</i>	120
Figura 29.	<i>Secuenciación por Heliscope.....</i>	124
Figura 30.	<i>Secuenciación por nanopore</i>	126
Figura 31.	<i>Secuenciación de molécula única en tiempo real</i>	129
Figura 32.	<i>Secuenciación por captura de imagen</i>	131
Figura 33.	<i>Microarreglos.....</i>	134
Figura 34.	<i>Enfoques basados en hibridación.....</i>	138
Figura 35.	<i>Northern blot</i>	140
Figura 36.	<i>Proceso de Formación del ADN ramificado</i>	141
Figura 37.	<i>Microarreglos de ADNc.....</i>	144
Figura 38.	<i>Transcriptómica: enfoques basados en secuencia</i>	148
Figura 39.	<i>Ensayo de ARN-Seq</i>	151
Figura 40.	<i>scARN-seq</i>	157
Figura 41.	<i>Transcriptómica espacial.....</i>	165
Figura 42.	<i>Proteómica</i>	167
Figura 43.	<i>Cromatografías</i>	170
Figura 44.	<i>Electroforesis.....</i>	173
Figura 45.	<i>Espectrómetro de masas</i>	176
Figura 46.	<i>Espectrometría de masas en tándem (MS/MS o MS2).....</i>	179
Figura 47.	<i>Estados de espín</i>	187
Figura 48.	<i>Estados de espín con la fuerza del campo magnético</i>	187
Figura 49.	<i>Resonancia magnética nuclear.....</i>	188
Figura 50.	<i>Resonancia magnética nuclear.....</i>	191
Figura 51.	<i>Cromatografía en fase inversa</i>	193
Figura 52.	<i>Técnica de bisulfito.....</i>	199
Figura 53.	<i>Cromatina de inmunoprecipitación (ChIP).....</i>	200
Figura 54.	<i>Microarreglos epigenéticos</i>	202

LISTA DE TABLAS

Tabla 1.	<i>Actividades en genómica</i>	22
Tabla 2.	<i>Actividades en transcriptómica</i>	23
Tabla 3.	<i>Actividades en proteómica</i>	24
Tabla 4.	<i>Actividades en metabolómica</i>	26
Tabla 5.	<i>Actividades en epigenómica</i>	27
Tabla 6.	<i>Actividades en metagenómica</i>	28
Tabla 7.	<i>Actividades en farmacogenómica</i>	29
Tabla 8.	<i>Bases de datos y herramientas bioinformáticas usadas en genómica.</i>	72
Tabla 9.	<i>Bases de datos y herramientas bioinformáticas usadas en transcriptómica</i>	75
Tabla 10.	<i>Bases de datos y herramientas bioinformáticas usadas en proteómica</i>	77
Tabla 11.	<i>Bases de datos usadas en metabolómica</i>	80
Tabla 12.	<i>Bases de datos y herramientas bioinformáticas usadas en otras ómicas</i>	81
Tabla 13.	<i>Tecnología de alto rendimiento en genómica</i>	102
Tabla 14.	<i>Técnicas en transcriptómica</i>	136
Tabla 15.	<i>Ventajas clave sobre las tecnologías existentes</i>	152
Tabla 16.	<i>Ventajas y desventajas de las técnicas en transcriptómica</i>	159
Tabla 17.	<i>Tecnologías en proteómica</i>	165
Tabla 18.	<i>Definiciones claves en cromatografía</i>	168
Tabla 19.	<i>Tecnologías de alto rendimiento utilizadas en metabolómica</i>	183

PREFACIO

Para la docente Ruth Mélida Sánchez Mora, líder del grupo de investigación en Biotecnología y Genética de la Universidad Colegio Mayor de Cundinamarca, es un logro presentar esta obra que recopila algunos de los *principios básicos de las ciencias ómicas y sus tecnologías*. Resulta fascinante comprender los conceptos fundamentales de las ómicas y las diferentes tecnologías empleadas en sus análisis. Estos conocimientos permitirán a los lectores adquirir una comprensión más amplia de las aplicaciones de las ciencias ómicas en áreas como la biotecnología y disciplinas afines.

En el primer capítulo, se abordan una amplia variedad de temas fundamentales en estas disciplinas emergentes, desde su introducción hasta la exploración de conceptos clave como la genómica, la proteómica, la transcriptómica y la metabolómica. Se ofrece un detallado análisis de la evolución histórica de las ciencias ómicas, cubriendo desde sus primeros avances hasta las eras preómica y ómica, además de examinar proyectos innovadores en desarrollo. También se profundiza en la era multiómica, en la que se resalta la importancia de la bioinformática y el análisis de datos multiómicos. Finalmente, se abordan la medicina de precisión y el enfoque ómico en las enfermedades, mediante los cuales se exploraron los orígenes y avances de las ciencias ómicas.

El segundo capítulo aborda una amplia gama de herramientas y métodos utilizados en el estudio de disciplinas relacionadas con las ciencias ómicas. Comienza con una introducción que establece el contexto y la importancia de estas técnicas. Luego, se detallan las técnicas en genómica, que incluyen diferentes generaciones de secuenciación y microarreglos. A continuación, se exploran las técnicas en transcriptómica, abarcando desde enfoques basados en hibridación hasta métodos avanzados como ARN-seq y transcriptómica espacial. El capítulo continúa con las técnicas utilizadas en proteómica, como cromatografía, electroforesis y espectrometría, seguidas de las técnicas en metabolómica, que incluyen métodos de separación y espectrometrías, así como técnicas específicas de espectroscopia. Finalmente, se presentan herramientas relacionadas con otras ciencias ómicas, como la lipidómica, la glicómica y la epigenómica, proporcionando una visión completa de las metodologías utilizadas en este campo.

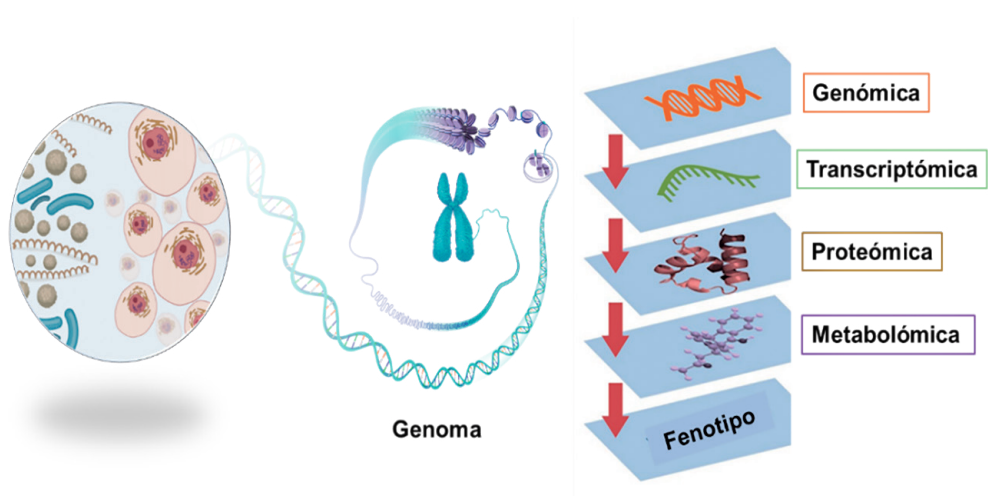
CAPÍTULO 1. CIENCIAS ÓMICAS

1. INTRODUCCIÓN

A finales del siglo XIX se reconoció el papel esencial de los genes en la transmisión de la herencia. Sin embargo, no fue hasta el descubrimiento de la estructura del ácido desoxirribonucleico (ADN) que se abordó la herencia desde una perspectiva molecular. El Proyecto Genoma Humano (PGH), iniciado en 2003, y otros proyectos similares, han brindado una visión integral de la información genética presente en una célula, revolucionando los estudios biológicos. Esto ha dado lugar a la aparición de las ciencias ómicas modernas, con la genómica liderando el camino, al permitir una comprensión completa de los genes en una célula. En este contexto, se presentan los proyectos que respaldan el avance de las ciencias ómicas y la era multiómica, así como los adelantos en bioinformática y en análisis multiómicos a gran escala. Estos desarrollos abren un nuevo panorama en el presente y crean un espacio para enfrentar y aprovechar los diversos campos de acción de las ciencias ómicas, que han tenido un impacto significativo en la medicina de precisión.

2. GENERALIDADES DE LAS CIENCIAS ÓMICAS

El término *ómicas* se utilizó por primera vez en la década de 1980 para describir el estudio, caracterización y cuantificación de conjuntos de datos masivos de moléculas individuales o interrelacionadas, que se involucran en diferentes vías de funcionamiento de un organismo (HOOD y ROWEN, 2013). Las principales ciencias ómicas comprenden el estudio del genoma, del transcriptoma, del proteoma y del metaboloma, con el objetivo de obtener información detallada sobre el estado fisiológico de una célula u organismo en estudio, así como determinar su influencia en la estructura, función y dinámica de los procesos biológicos (KARCZEWSKI y SNYDER, 2018). Además, permite comprender realmente los fenotipos observados y las causas subyacentes a ellos (**figura 1**).

Figura 1. Principales ciencias ómicas

Nota. El genoma de un organismo se hereda de las células progenitoras. Las ciencias ómicas, como la genómica, la transcriptómica, la proteómica y la metabolómica, se enfocan en estudiar cómo estas células están correlacionadas entre sí para determinar los diversos fenotipos.

Existen numerosos campos que emplean el sufijo *ómica* en la actualidad. Sin embargo, es esencial identificar principalmente cuatro áreas de estudio de las ciencias ómicas utilizadas en la investigación básica y clínica: la genómica, la proteómica, la transcriptómica y la metabolómica. Estas disciplinas se centran en el análisis a gran escala de diversas moléculas biológicas, como el ADN, el ácido ribonucleico (ARN), los genes, las proteínas y los metabolitos (SÁEZ, 2022).

Además de estas áreas principales, se han desarrollado numerosas disciplinas ómicas especializadas que abordan la medición de todos o algunos de los componentes biológicos. Algunos ejemplos de estas disciplinas son la epigenómica, la metagenómica, la farmacogenómica, la glicómica y la lipodómica, entre otras (KARCZEWSKI y SNYDER, 2018).

Los avances en herramientas para el análisis molecular y bioinformático han permitido llevar a cabo estudios a gran escala en el campo de la biología (PIÑA *et al.*, 2018). Estos avances han impulsado el surgimiento de nuevas disciplinas ómicas y el desarrollo de ciencias ómicas emergentes, que se explicarán a continuación.

2.1. GENÓMICA

El genoma se refiere al conjunto completo de material genético (ADN Y ARN) presente en el núcleo de un organismo y los orgánulos celulares. Por ejemplo, el

genoma humano está compuesto por el ADN que se empaqueta en los 23 pares de cromosomas que participan durante la división celular, de los cuales 22 son cromosomas somáticos y un par sexual. Además, el genoma humano incluye el ADN presente en las mitocondrias, parte de la célula que cumple la función de producir energía.

Es importante destacar que cada organismo tiene su propio genoma, lo cual significa que la información genética contenida en el ADN de un individuo es específica para él. En cuanto a la variabilidad entre especies, esta se refleja, por ejemplo, en el número y organización de los cromosomas, que puede variar ampliamente entre diferentes especies. Por otro lado, dentro de una misma especie, existen variaciones genómicas entre individuos, estas variaciones contribuyen a la diversidad en características físicas, susceptibilidad a enfermedades y respuestas a tratamientos médicos. Comprender esta diversidad genómica tanto entre como dentro de especies es fundamental para el estudio de la variabilidad biológica (GOSÁLVEZ y HORCAJADAS, 2018).

Dentro de este contexto, la genómica es el estudio sistemático del genoma completo de un organismo y sus funciones. Se divide en genómica estructural y funcional. La primera se enfoca en el estudio de la organización física y espacial del genoma, incluyendo la secuencia de nucleótidos, la estructura tridimensional del ADN y la localización de los genes en los cromosomas. Este campo permite entender cómo están organizados los elementos genéticos y cómo interactúan entre sí dentro del genoma (MICHALSKA y JOACHIMIAK, 2021). Por otro lado, la genómica funcional se centra en asignar funciones a las secuencias genómicas, comprendiendo cómo los genes son regulados y expresados, así como su papel en los procesos biológicos y las enfermedades (BRAICU, 2021). Ambas ramas de la genómica son complementarias y fundamentales para avanzar en la comprensión de la genética y su aplicación en la medicina y otras áreas de la biología molecular.

Por esta razón el estudio del genoma humano es fundamental para comprender la herencia genética, la variabilidad entre individuos y las bases moleculares de las enfermedades (BROWN *et al.*, 2022). En la **tabla 1** se presentan los principales usos de la genómica, así como las actividades, áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 1. *Actividades en genómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La genómica se dedica al estudio y análisis metódico del genoma de los organismos.	Secuenciación del ADN.	Genómica comparativa.	Secuenciación del genoma.
	Estudio de la estructura y organización del genoma.	Genómica estructural y funcional.	Análisis de variaciones genéticas.
	Análisis de variantes genéticas.	Genómica de poblaciones.	Anotación y predicción de genes.
	Análisis de expresión génica.	Epigenómica.	Estudio de la estructura y organización del genoma.
	Anotación y función genómica.	Metagenómica.	
	Análisis de interacciones gen-gen y redes biológicas.		

2.2. TRANSCRIPTÓMICA

La transcripción es el proceso mediante el cual una célula copia la información genética contenida en una secuencia de ADN y la transcribe en una molécula de ARN. Durante la transcripción, la enzima ARN polimerasa sintetiza una cadena de ARN complementaria a una hebra de ADN. Esta molécula de ARN resultante es una copia exacta o una versión transcrita de un gen específico. La transcripción es el primer paso en el proceso de expresión génica, donde la información genética contenida en el ADN se utiliza para producir proteínas y ARN para realizar diversas funciones celulares. Este proceso es esencial para la síntesis de proteínas y la regulación de la actividad génica en los organismos vivos.

Los ARN mensajeros (ARNm) contienen la secuencia que codifica proteínas y son procesados por los ribosomas para su traducción. Además de estos ARNm, existen ARN no codificantes (ARNnc) que tienen funciones regulatorias en la expresión génica y la función proteica. Los ARN no codificantes (ARNnc) incluyen microARN (miARN) y ARN largos no codificantes (lncARN), los cuales desempeñan funciones regulatorias esenciales en la célula. Por ejemplo, los miARN pueden unirse a moléculas de ARNm específicas para inhibir su traducción o promover su degradación, modulando la expresión de genes asociados a procesos como el desarrollo, la diferenciación celular y la apoptosis. Por otro lado, los lncARN pueden interactuar con el ADN, ARN o proteínas para influir en la estructura de la cromatina, la transcripción génica y el empalme

alternativo, afectando funciones celulares como la inactivación del cromosoma X y la regulación de genes por impronta genética (WANG *et al.*, 2009).

En este contexto la transcriptómica describe el conjunto completo de transcritos de ácido ribonucleico (ARN), que incluye ARN mensajero (ARNm), ARN ribosómico (ARNr) y ARN de transferencia (ARNt) que codifican el genoma de un tipo específico de célula o tejido (WANG *et al.*, 2009). En la **tabla 2** se presentan los principales usos de la transcriptómica, así como las actividades y áreas de estudio y trabajo relacionadas con esta disciplina (WANG *et al.*, 2009).

Tabla 2. *Actividades en transcriptómica*

Descripción	Actividades que se desarrollan	Áreas de estudio	Área de trabajo
La transcriptómica se dedica al estudio de la expresión génica para comprender la regulación y función de genes en diferentes condiciones biológicas.	Extracción y secuenciación de ARN, procesamiento y análisis de datos, validación de genes mediante qPCR, análisis de rutas metabólicas y redes génicas, visualización y comunicación de resultados.	Genómica funcional, regulación génica, biología celular, bioinformática, biología de sistemas, genética molecular.	Laboratorio de biología molecular, bioinformática, análisis computacional, redacción científica.

Es importante resaltar que cada tipo de célula, tejido u organismo tiene su propio transcriptoma único que refleja su estado fisiológico y sus respuestas específicas a diferentes condiciones y estímulos (KLEINO *et al.*, 2022). Por ejemplo, estudiar el transcriptoma de células cancerosas ayuda a identificar genes y vías de señalización involucrados en la carcinogénesis, así como a identificar posibles biomarcadores y terapias dirigidas contra el cáncer.

2.3. PROTEÓMICA

Dentro del genoma se encuentran los genes, concebidos como regiones específicas del ADN que contienen la información necesaria para la síntesis del ARN funcional y las proteínas. Estos genes son responsables de codificar las instrucciones para la formación y funcionamiento de los diferentes componentes del organismo (MORAES y GOES, 2016). Así pues, a través de la expresión de los genes, se producen ARN y proteínas que desempeñan funciones vitales en los procesos biológicos.

El proteoma se refiere al conjunto de todas las proteínas expresadas por el genoma en un momento, condición y ubicación específicos, como una célula,

tejido u organismo. El estudio de los proteomas permite obtener una visión más completa de las funciones y procesos biológicos en los diferentes sistemas vivos (LILL *et al.*, 2021). Por ejemplo, el proteoma bacteriano hace referencia al conjunto de proteínas expresadas en una bacteria específica. Cada especie bacteriana tiene su propio proteoma que está relacionado con su metabolismo, patogenicidad y función biológica (FENG *et al.*, 2015a).

En este sentido, la proteómica es un campo de estudio amplio que se enfoca en perfiles de proteínas a gran escala. Comprende la investigación de la composición, estructura y función de las proteínas, así como el estudio de su expresión, modificaciones postraduccionales e interacciones entre diferentes proteínas y otras moléculas (FENG *et al.*, 2015a). Esta ciencia ómica brinda información valiosa sobre el complejo y dinámico mundo de las proteínas y su contribución a los procesos biológicos. En la **tabla 3** se presentan los principales usos de la proteómica, así como las actividades, áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 3. *Actividades en proteómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La proteómica implica el estudio sistemático a gran escala del proteoma completo, que es la suma de todas las proteínas presentes en un organismo, tejido, célula o líquido biológico.	Identificación, caracterización y cuantificación de proteínas. Análisis de modificaciones postraduccionales. Análisis de interacciones proteína-proteína. Perfilado proteómico. Estudio de isoformas y variantes proteicas. Análisis de interacciones proteína-ADN. Proteómica comparativa.	Identificación de biomarcadores. Análisis de modificaciones postraduccionales. Interacciones proteína-proteína. Análisis de subproteomas y localización subcelular. Estudio de variantes proteicas e isoformas. Análisis de estructura y función de proteínas.	Investigación biomédica. Farmacología y desarrollo de medicamentos. Biología de sistemas. Biotecnología y producción de proteínas recombinantes. Agronomía y mejora de cultivos. Medio ambiente y biología ambiental.

2.4. METABOLÓMICA

Los metabolitos son moléculas que se producen, utilizan o son transformadas durante el proceso metabólico en un organismo. Estas sustancias pueden ser el resultado de la descomposición de alimentos, medicamentos, productos químicos o tejidos del propio cuerpo.

Los metabolitos son componentes clave en diversas vías metabólicas y desempeñan roles importantes en la regulación y mantenimiento de las funciones celulares y del organismo en su conjunto. Por ejemplo, participan en la producción de energía, como la glucosa en la respiración celular; actúan en la señalización celular, como el AMP cíclico en rutas de señalización; y son precursores en la síntesis de macromoléculas, como aminoácidos y nucleótidos para proteínas y ácidos nucleicos (AZAD *et al.*, 2023). Los metabolitos son objeto de estudio en diversos campos, incluida la medicina, la biología molecular, y ecología entre otros.

La metabolómica es una ciencia ómica que se centra en el estudio integral de los metabolitos presentes en una muestra, organismo o sistema biológicos en un momento dado. Su objetivo principal es la identificación y cuantificación de metabolitos presentes en el metaboloma, que es el conjunto completo de metabolitos de un sistema (AZAD *et al.*, 2023). Para llevar a cabo este estudio, en la metabolómica se utilizan técnicas analíticas avanzadas como la espectrometría de masas y la resonancia magnética nuclear (RMN) (BAUERMEISTER *et al.*, 2022). Estas técnicas permiten caracterizar y analizar los perfiles metabólicos, lo que proporciona una visión detallada de los cambios metabólicos y las interacciones en los sistemas biológicos.

La metabolómica proporciona valiosa información sobre la fisiología, el estado de salud y las respuestas de los organismos a estímulos externos (AZAD *et al.*, 2023). Por ejemplo, el estudio del metaboloma vegetal ayuda a comprender los procesos metabólicos implicados en el crecimiento, desarrollo, respuesta al estrés ambiental y producción de compuestos beneficiosos para la alimentación humana y la industria (PATTI *et al.*, 2022). Además, resulta interesante estudiar los metabolitos presentes en el suelo y su interacción con los organismos vivos, lo cual contribuye a comprender los procesos metabólicos en el suelo, como la desintegración de la materia orgánica, la disponibilidad de nutrientes y los intercambios químicos entre el suelo y las plantas. En la **tabla 4** se presentan los principales usos de la metabolómica, así como las actividades y áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 4. *Actividades en metabolómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La metabolómica se encarga del estudio sistemático de la identificación y cuantificación de todos los metabolitos que se encuentran en una célula, tejido, órgano o en un organismo en diversas condiciones.	Identificación y cuantificación de metabolitos. Análisis de vías metabólicas. Estudio de perfiles metabólicos. Identificación de biomarcadores. Integración de datos multiómicos. Estudio de la respuesta metabólica. Modelado y análisis estadístico.	Metabolómica clínica, metabolómica nutricional, microbiana, ambiental, farmacológica. Metabolómica de toxicología, enfermedades metabólicas, neurodegenerativas, cáncer.	Investigación biomédica. Farmacología y desarrollo de medicamentos. Nutrición y alimentación. Agricultura y seguridad alimentaria. Ecología y medio ambiente. Biotecnología y bioprocesos.

2.5. OTRAS ÓMICAS

Además de las ciencias ómicas más conocidas, existen otras disciplinas que también desempeñan un papel importante en el estudio integral de los sistemas biológicos. Estas ciencias ómicas adicionales permiten obtener una visión más completa y detallada de los diferentes aspectos de la biología y la salud.

2.5.1. EPIGENÓMICA

La epigenética investiga cómo los factores ambientales y los conductuales pueden modificar la actividad génica sin alterar la secuencia del ADN, impactando así en el desarrollo, la salud y las enfermedades. Implica cambios heredables en la expresión génica que no se reflejan en la secuencia de nucleótidos del ADN y que pueden ser influenciados por elementos externos, como la dieta, el ejercicio y el estrés (Fu *et al.*, 2023). Este campo es fundamental en procesos biológicos como la diferenciación celular, la regulación del desarrollo y la adaptación al entorno. Asimismo, se ha asociado con diversas afecciones médicas, como el cáncer, las enfermedades cardiovasculares y neurológicas.

La epigenómica es un campo multidisciplinario que combina la biología molecular, la genética, la bioinformática y otras disciplinas relacionadas. Sus principales actividades incluyen la identificación y caracterización de modificaciones epigenéticas, el análisis de su influencia en la expresión génica

y el estudio de su relación con enfermedades y procesos biológicos (Fu *et al.*, 2023). En la **tabla 5** se presentan de los principales usos de la epigenética, así como las actividades y áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 5. *Actividades en epigenómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La epigenética se dedica al estudio sistemático de los cambios en la expresión génica global debido a procesos epigenéticos y no por cambios en la cadena de bases del ADN.	Identificación de modificaciones epigenéticas Caracterización de cambios en la estructura de la cromatina. Análisis de la influencia de factores ambientales en la epigenética. Estudio de la relación entre modificaciones epigenéticas y enfermedades. Desarrollo de técnicas de secuenciación y análisis bioinformático.	Metilación del ADN. Modificaciones de las histonas. Interferencia del ARN. Epigenética del cáncer. Epigenética del envejecimiento.	Investigación biomédica. Medicina y salud humana. Investigación del desarrollo. Genómica comparativa. Farmacología y terapia génica.

2.5.2. METAGENÓMICA

El metagenoma se refiere al conjunto completo de material genético obtenido directamente de una muestra ambiental, que puede contener material genético de diversos organismos presentes en ese entorno. Esta técnica permite estudiar la diversidad genética y funcional de una comunidad microbiana sin la necesidad de aislar y cultivar cada organismo por separado (CABELLO-YEVES *et al.*, 2021). La metatranscriptómica, por otro lado, analiza el conjunto completo de ARN mensajero transcrito en esa comunidad y proporciona información sobre los genes que están siendo expresados activamente.

El análisis del metagenoma y la metatranscriptómica son fundamentales en campos como la microbiología ambiental, la medicina, la agricultura y la biotecnología, ya que proporcionan información sobre la composición genética, las funciones metabólicas y la actividad funcional de las comunidades microbianas en diversos hábitats. Además, el metagenoma y la metatranscriptómica pueden revelar la presencia de microorganismos patógenos, la diversidad de genes relacionados con la resistencia a antibióticos y la capacidad de producir compuestos útiles, entre otros aspectos importantes para la comprensión y aplicación en diversos campos científicos y prácticos (CABELLO-YEVES *et al.*, 2021).

La metagenómica es un área de estudio que se enfoca en el análisis genómico y funcional de comunidades microbianas presentes en diferentes muestras. Las principales actividades que se desarrollan en la metagenómica incluyen la secuenciación de muestras ambientales, el análisis de metagenomas y la caracterización de comunidades microbianas (CABELLO-YEVES *et al.*, 2021). En la **tabla 6** se presentan de los principales usos de la metagenómica, así como las actividades y áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 6. *Actividades en metagenómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La metagenómica se dedica al estudio sistemático de los genomas de conjuntos de organismos presentes en el suelo, el agua, el tracto gastrointestinal humano, etc.	Secuenciación de muestras ambientales.	Diversidad microbiana.	Ecología y medio ambiente.
	Análisis de metagenomas.	Funcionalidad del microbioma.	Investigación en microbiología.
	Caracterización de comunidades microbianas.	Interacciones microbianas	Salud humana y animal.
	Identificación de nuevos microorganismos.	Biodegradación y biotecnología.	Agricultura y alimentación.
	Estudio de la evolución y adaptación de los microorganismos.	Genómica comparativa.	Descubrimiento de fármacos.

2.5.3. FARMACOGENÓMICA

La farmacogenética es el estudio de cómo la variabilidad genética de un individuo influye en su respuesta a los medicamentos. Se centra en identificar cómo los genes de una persona pueden afectar la eficacia y los posibles efectos secundarios de los tratamientos farmacológicos. Esta disciplina busca personalizar los tratamientos médicos según las características genéticas únicas de cada paciente, lo que puede mejorar la eficacia de los medicamentos y reducir los riesgos de efectos adversos. La farmacogenética también ayuda a entender por qué algunas personas responden de manera diferente a los mismos medicamentos, lo que puede tener implicaciones importantes en la práctica clínica y en la optimización de la terapia farmacológica.

La farmacogenómica es un campo de estudio que investiga la influencia del genoma completo en la respuesta de los individuos a los fármacos (ZHANG y DOLAN, 2010). A diferencia de la farmacogenética, que se centra en el análisis de genes individuales y cómo sus variantes afectan la respuesta a un medicamento específico, la farmacogenómica considera la interacción de múltiples genes y

variantes genéticas en conjunto, lo que permite una comprensión más amplia y detallada de los mecanismos genéticos que influyen en la eficacia y seguridad de los medicamentos. Las ventajas de la farmacogenómica sobre la farmacogenética incluyen la capacidad de identificar patrones genéticos complejos que afectan la respuesta a los medicamentos, algo no posible al estudiar genes de forma aislada; desarrollar terapias más efectivas y personalizadas al considerar la influencia combinada de múltiples variantes genéticas; mejorar la seguridad de los medicamentos al predecir de manera más precisa las reacciones adversas basadas en el perfil genómico completo del individuo; y avanzar en la medicina de precisión al integrar datos genómicos para optimizar el tratamiento y reducir la variabilidad en la respuesta farmacológica (NAITHANI *et al.*, 2021).

En este contexto, la farmacogenómica ofrece una perspectiva más integral que permite abordar la complejidad de las interacciones genético-farmacológicas. En la **tabla 7** se presentan los principales usos de la farmacogenómica, así como las actividades y áreas de estudio y trabajo relacionadas con esta disciplina.

Tabla 7. *Actividades en farmacogenómica*

Descripción	Actividades que se desarrollan	Área de estudio	Área de trabajo
La farmacogenómica es la rama de la genética que se enfoca en la diversidad genética relacionada con las respuestas individuales a los medicamentos y el tratamiento de estos.	<p>Estudio de la influencia de los genes en la respuesta a los fármacos.</p> <p>Identificación de biomarcadores genéticos.</p> <p>Análisis de interacciones gen-fármaco.</p> <p>Estudio de los mecanismos moleculares de la acción de los fármacos.</p> <p>Aplicación de la genómica en el diseño de nuevos fármacos.</p>	<p>Variabilidad genética.</p> <p>Farmacocinética y farmacodinamia.</p> <p>Farmacogenética y farmacogenómica.</p> <p>Farmacología molecular.</p> <p>Genética de las enfermedades.</p>	<p>Investigación farmacéutica.</p> <p>Medicina de precisión.</p> <p>Desarrollo de terapias personalizadas.</p> <p>Farmacovigilancia.</p> <p>Mejora de la eficacia y seguridad de los medicamentos.</p>

Además de estas ciencias ómicas adicionales, existen otras que amplían el espectro del conocimiento y proporcionan una visión más completa de los sistemas biológicos. Estas disciplinas permiten una comprensión más profunda de la complejidad de la biología y su relación con la salud y las enfermedades. Algunos ejemplos de estas disciplinas son la glicómica, la lipidómica, la ionómica, la microbiómica, entre otras (FRANCESCHI *et al.*, 2018; Singh *et al.*, 2021; KUMAR, 2021).

La microbiómica es el campo de estudio que se enfoca en el análisis integral de los microbiomas, es decir, las comunidades de microorganismos (bacterias, arqueas, hongos, virus y otros microbios) que habitan en un ambiente específico, como el cuerpo humano, el suelo o los océanos. La microbiómica no solo analiza la composición taxonómica de estas comunidades, sino también sus funciones, interacciones y cómo afectan al entorno o al organismo huésped.

3. HISTORIA DE LAS CIENCIAS ÓMICAS

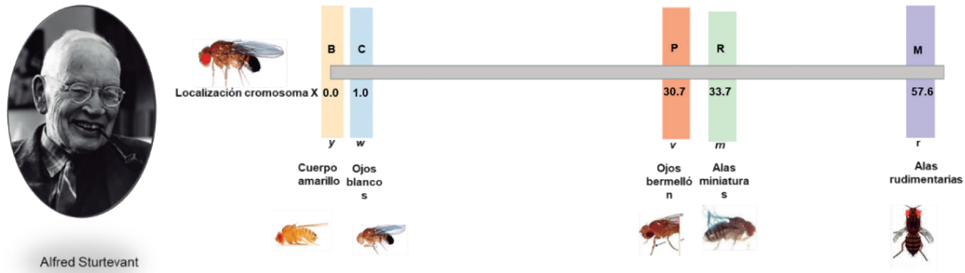
La historia de las ciencias ómicas ha revolucionado la comprensión de los sistemas biológicos a lo largo de los años. Estas disciplinas emergieron en la década de 1980 con el objetivo de explorar y analizar conjuntos masivos de datos moleculares interrelacionados, lo cual ha abierto nuevas puertas para comprender la complejidad de la biología (HOOD y ROWEN, 2013). En este apartado, se abordarán los principales proyectos que desempeñaron un papel clave en el surgimiento de las ciencias ómicas.

3.1. PRIMEROS AVANCES EN LAS CIENCIAS ÓMICAS

El mapeo genético consiste en el proceso de determinar la ubicación de los genes en los cromosomas. En los mapas genéticos iniciales, se llevó a cabo un análisis de ligamiento, este establece que, a medida que dos genes se encuentran cerca en un cromosoma, aumenta la probabilidad de heredarse juntos (CLAUSSNITZER *et al.*, 2020). De esta manera, al seguir los esquemas de herencia, se puede establecer la posición relativa de los genes.

En 1911, el genetista Alfred Sturtevant realizó un avance significativo al desarrollar una técnica para mapear la ubicación de genes específicos en los cromosomas de la mosca de la fruta, *Drosophila melanogaster*. Logró determinar las frecuencias de recombinación y las relacionó con la distancia entre dos genes en el cromosoma. Observó que, a medida que aumentaba la distancia entre dos genes, también aumentaba la probabilidad de que se separaran durante la recombinación (CLAUSSNITZER *et al.*, 2020). A través de sus experimentos, este investigador consiguió construir el primer mapa genético de un cromosoma en 1913.

En este mapa ubicó cinco características (rasgos) a lo largo del cromosoma de manera lineal, basándose en la distancia relativa entre ellos. Algunos de los rasgos incluidos en este mapa fueron el color del cuerpo, los ojos blancos, los ojos rojos, las alas pequeñas y las alas rudimentarias (**figura 2**).

Figura 2. Primer mapa genético de Alfred Sturtevant

Nota. Se muestran cinco rasgos que están dispuestos a lo largo de un cromosoma lineal de acuerdo con la distancia relativa de cada uno. Las características incluyen el cuerpo amarillo (b), los ojos blancos (c), los ojos bermellón (p), las alas en miniatura (r) y las alas rudimentarias (m).

El trabajo pionero de Sturtevant sentó las bases para la construcción de mapas genéticos más detallados y precisos (RUSHTO, 2022). Su aporte al utilizar la recombinación genética como una medida de distancia entre genes sigue siendo una herramienta fundamental en los estudios de genética y en los avances de las ciencias ómicas.

El mapeo genético juega un papel fundamental en la genómica pues permite comprender la ubicación y organización de los genes en los cromosomas, lo que es esencial para identificar genes relacionados con enfermedades y otras características específicas (CLAUSSNITZER *et al.*, 2020), lo que propicia un mejor entendimiento de cómo los genes contribuyen a diferentes características y la predisposición genética a ciertas enfermedades y trastornos hereditarios.

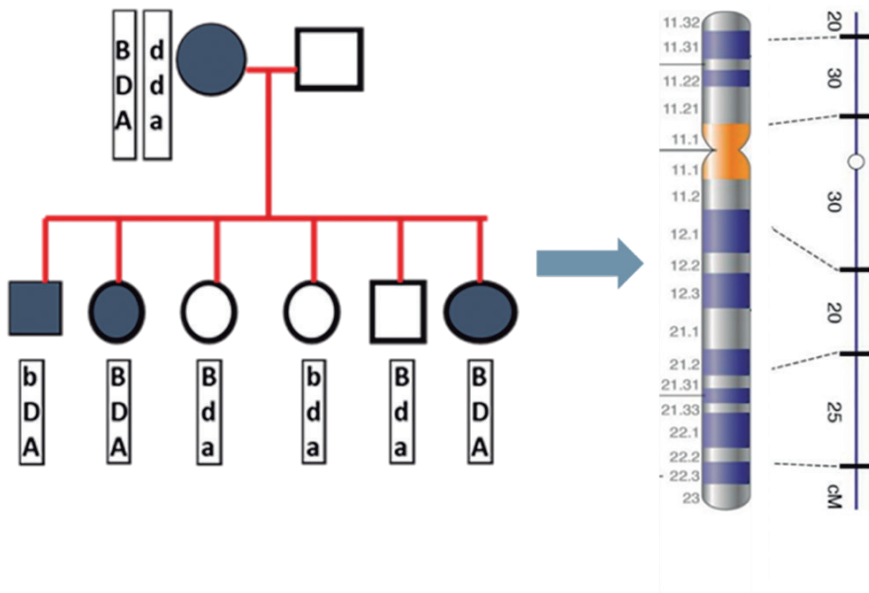
Existen dos enfoques principales en el campo del mapeo genético. El primero se centra en mapas genéticos basados en el análisis de ligamiento, mientras que el segundo aborda la asociación entre características fenotípicas y marcadores genéticos por mapeo físico. Estos enfoques son fundamentales para comprender la herencia y la variabilidad genética en diferentes organismos.

3.1.1. MAPAS GENÉTICOS BASADOS EN EL ANÁLISIS DE LIGAMIENTO

Este mapeo se basa en el análisis de genealogías y en el cálculo de las frecuencias de recombinación entre los alelos (formas alternativas del gen) de dos *loci* genéticos que están ligados y se encuentran en el mismo cromosoma. Estas frecuencias predicen que cuanto más juntos están dos genes entre sí en un

cromosoma, es más probable que se hereden de la misma manera. Las mediciones se realizan en CentiMorgan (cM), que representan la distancia en la cual se observa una recombinación del 1 % durante la meiosis, lo cual equivale a la distancia genética (CLAUSSNITZER *et al.*, 2020). Los datos se obtienen mediante la comparación de los fenotipos normales y los mutantes asociados a enfermedades genéticas específicas, así como a través del análisis de genealogías. Idealmente, se busca contar con la indagación de varias generaciones y un gran número de individuos para realizar un análisis más preciso (figura 3).

Figura 3. Mapas genéticos



Nota. Analiza los patrones de herencia en las genealogías para determinar las posiciones relativas de los genes. La figura muestra un árbol genealógico (mujeres y hombres sanos y enfermos), además un esquema cromosómico que representa la herencia de ciertos alelos en la familia. Se muestra una posible ubicación en un cromosoma específico.

3.1.2. MAPEO FÍSICO

El mapeo físico se basa en el uso de marcadores genéticos, como los polimorfismos de nucleótido único (SNP) o microsatélites, que son regiones del ADN que presentan variaciones entre individuos. Estos marcadores se utilizan como puntos de referencia para determinar la posición relativa de los genes en el genoma (RUSHTO, 2022). Existen dos herramientas utilizadas en el mapeo físico:

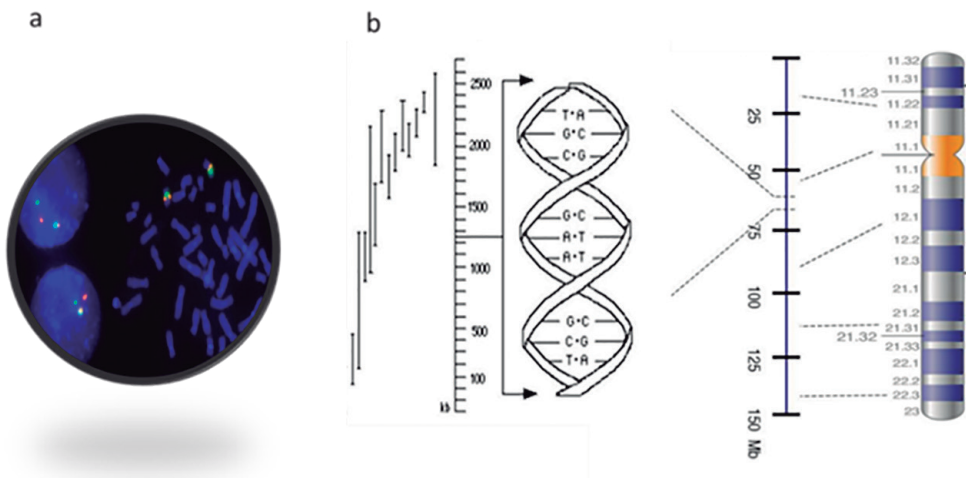
- **Mapeo físico basado en hibridación**

Este enfoque utiliza sondas de ADN para identificar regiones específicas del genoma. Las sondas se unen a las secuencias objetivo y se detectan mediante técnicas de hibridación, lo que permite identificar la ubicación y la organización de los genes (**figura 4a**).

- **Mapeo físico basado en restricción**

Este utiliza enzimas de restricción para cortar el ADN en fragmentos más pequeños. Tales fragmentos se separan utilizando técnicas de electroforesis y se crean mapas físicos utilizando la información de las distancias relativas entre los fragmentos de ADN (RUSHTO, 2022), como se puede ver en la **figura 4b**.

Figura 4. Mapas físicos



Nota. Se muestra un mapeo genético, se examina la estructura y posición de secuencias de ADN en un cromosoma. a) Muestra la identificación de genes en células usando técnicas como FISH. (b) Análisis con enzimas de restricción para determinar las posiciones relativas de los genes en el ADN y su ubicación cromosómica.

3.2. ERA PREÓMICA

Durante la era preómica, que abarca desde 1944 hasta la década de 1970, se originaron importantes avances en el estudio molecular del genoma humano.

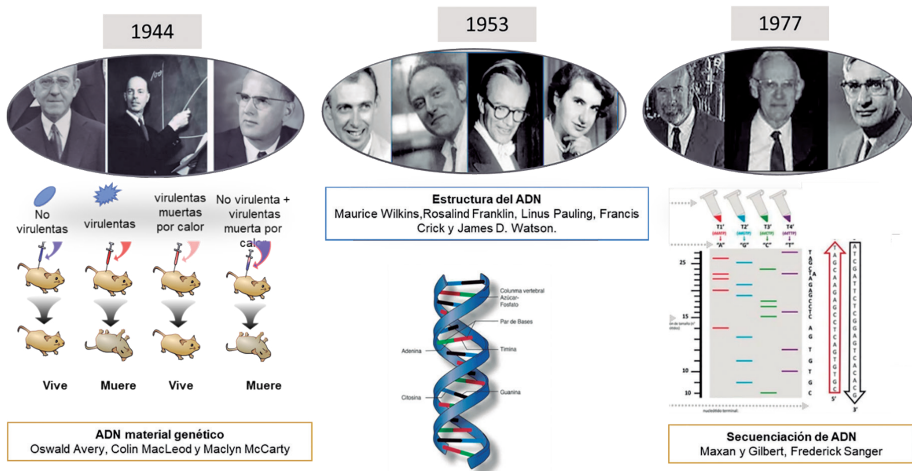
En 1944, Oswald Avery, Colin MacLeod y Maclyn McCarty revelaron que el ADN era el factor transformante clave en la transmisión de la información genética y no las proteínas, como se pensaba anteriormente (National Human Genome Research Institute, 1944). Posteriormente, en 1952, Alfred Hershey y Martha Chase señalaron experimentalmente que el ADN era el material genético responsable de la herencia en los bacteriófagos (O'FARRELL, 2008).

Uno de los hallazgos más significativos ocurrió en 1953, cuando James Watson y Francis Crick describieron la estructura del ADN en forma de una doble hélice (WATSON y CRICK, 1953; Lindahl *et al.*, 2016). Este descubrimiento de la estructura del ADN sentó las bases para futuros avances en la genómica. Actualmente, la estructura del ADN se reconoce como un logro de Maurice Wilkins, Rosalind Franklin, Linus Pauling, Francis Crick y James D. Watson (DEMARIA, 2003; WATSON y CRICK, 2007).

En la década de 1970, Frederick Sanger logró secuenciar por primera vez un fragmento de ADN utilizando la secuenciación enzimática de ADN (uso de una taq polimerasa) basada en la terminación de cadena con el uso de los dideoxi-nucleótidos (SANGER *et al.*, 1977). Este logro marcó un hito concluyente en la capacidad de estudiar y comprender la secuencia de bases del ADN, conocida como secuenciación de primera generación (SANGER, 2004). El surgimiento de la secuenciación allanó el camino para el desarrollo de tecnologías de secuenciación genómica más avanzadas en las décadas posteriores (**figura 5**). Algunas de estas técnicas serán abordadas en el siguiente capítulo.

Figura 5. Era preómica

Era preómica de las ciencias ómicas



Nota. Se presentan los principales hallazgos que generaron un avance significativo en el desarrollo de las ciencias ómicas.

Otras técnicas moleculares importantes que surgieron en ese período incluyen el trabajo con polimorfismos de longitud de fragmentos de restricción (FRLP), la reacción en cadena de la polimerasa (PCR), los cromosomas artificiales bacterianos (BAC), los cromosomas artificiales de levaduras (YAC) y la electroforesis de campo pulsante (ANJUM *et al.*, 2017), técnicas que facilitaron la manipulación y análisis de fragmentos de ADN, así como la generación de mapas físicos y genéticos más detallados.

Estos avances tecnológicos y la capacidad de mapear y secuenciar los genomas de organismos modelo de tamaño más modesto sentaron las bases para la secuenciación a gran escala del genoma humano.

3.3. ERA ÓMICA

Desde 2001, los avances científicos, tecnológicos y bioinformáticos han fortalecido las técnicas de secuenciación del genoma humano. Esto ha brindado una base sólida para investigaciones más profundas sobre la genética de las enfermedades, la variabilidad genética y la identificación de blancos terapéuticos. Además, ha impulsado avances significativos en la medicina personalizada y en la comprensión de los mecanismos moleculares de diversas enfermedades. Estos avances, liderados por el Proyecto Genoma Humano (PGH), han dado lugar a otros proyectos, como el Proyecto HapMap, los Mil Genomas, ENCODE, HPP, HPRC y GENCODE, por mencionar algunos, sentando las bases para la aparición de las ciencias ómicas (GOSÁLVEZ y HORCAJADAS, 2018).

3.3.1. PROYECTO GENOMA HUMANO

El PGH se inició en 1986 a partir de una propuesta presentada por el Departamento de Energía de los Estados Unidos (DOE). Su objetivo principal

Objetivo. El principal objetivo del PGH fue establecer la secuencia de las bases químicas que componen el ADN de todo el genoma humano, así como identificar y cartografiar los genes del genoma desde una perspectiva física y funcional.

era investigar los posibles efectos de las radiaciones en el ADN y recopilar datos sobre la protección del genoma contra los efectos mutagénicos. En 1988, el Congreso de los Estados Unidos asignó financiamiento al DOE y al Instituto Nacional de

Salud (NIH), que son las agencias gubernamentales encargadas de coordinar las investigaciones y actividades técnicas relacionadas con el PGH (MORAES *et al.*, 2016; BIRNEY, 2021).

En 1990, se inició el proyecto público para secuenciar el genoma humano, y en ese momento se designó a James Watson como director del NIH, institución que

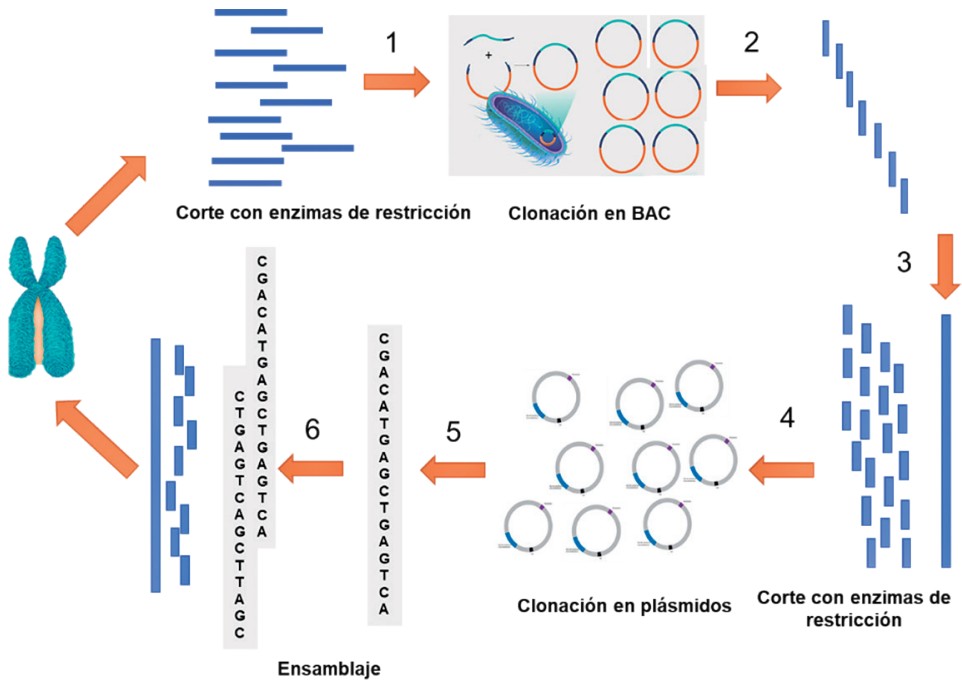
posteriormente fue nombrada Oficina de Investigaciones del Genoma Humano. Después del inicio del PGH, en 1996 se establecieron varios centros entre ellos el Centro para la Investigación de Enfermedades Hereditarias (CIDR, por sus siglas en inglés) con el propósito de estudiar la genética de enfermedades complejas (MORAES *et al.*, 2016). Posteriormente, en 1997, se creó el Instituto Nacional de Investigación del Genoma Humano (NHGRI, por sus siglas en inglés) y la División de Investigación Intramural (DIR), la cual se encarga de desarrollar y utilizar tecnologías genómicas para el estudio de enfermedades específicas. Esta división forma parte del Instituto Nacional de Alergias y Enfermedades Infecciosas (NIAID, por sus siglas en inglés) y ha estado liderando la investigación en enfermedades inmunológicas, alérgicas e infecciosas durante más de 65 años (figura 6) (BIRNEY, 2021; APELLANIZ-RUIZ *et al.*, 2016).

Figura 6. Proyecto genoma humano



Nota. Se muestra la página oficial del National Human Genome Research Institute. Principales centros de investigación que se formaron durante el PGH. Disponible en <https://www.genome.gov/es/genetics-glossary/Proyecto-Genoma-Humano>.

Los procedimientos utilizados inicialmente, en las décadas de los 80 y los 90, para secuenciar el genoma humano (GH) fueron las técnicas de aislamiento de ADN. El consorcio PGH aprovechó el uso de enzimas de restricción, cortó el ADN de cada cromosoma y cada fragmento lo clonó en vectores de clonación BAC, colocándolos en bacterias para crear clones bacterianos estables, que almacenaron aproximadamente 250 KB de ADN. Cada clon contenía diferentes regiones individuales del genoma humano en un BAC (GOSÁLVEZ Y HORCAJADAS, 2018), estas colecciones se denominaron genotecas. Utilizando marcadores moleculares tipo STS se crearon *contig* o conjuntos de clones solapantes que cubren una región del genoma, cada uno de estos BAC fueron segmentados en fragmentos más pequeño y clonados en vectores tipo plasmídicos, que fueron secuenciados y ensamblados, las secuencias permitieron reconstruir la secuencia de un cromosoma completo (figura 7) (BIRNEY, 2021).

Figura 7. Estrategia del consorcio PGH

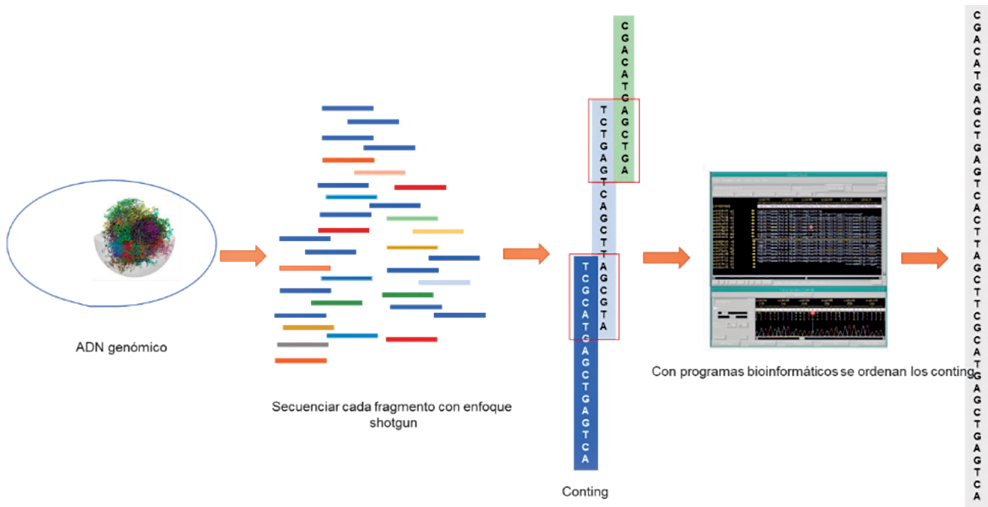
Nota. 1) El ADN genómico se divide en trozos de aproximadamente 150 Mb y se insertan en vectores BAC. 2) Los fragmentos de BAC se aíslan y se mapean para determinar el orden de cada fragmento clonado de 150 Mb. 3-4) Luego, cada BAC se corta al azar en fragmentos más pequeños, los cuales se clonan en un plásmido y se secuencian. 5) Posteriormente las secuencias resultantes se alinean, permitiendo identificar las secuencias que se superponen entre sí. 6) Finalmente, estos fragmentos contiguos se ensamblan, lo que permite obtener la secuencia completa.

El PGH de las dos entidades públicas estuvo conformado por aproximadamente 50 bibliotecas de este tipo. La BAC RP11 es una de las fuentes de información más comunes y corresponde a un individuo afroamericano. Sin embargo, se reporta que incluye una mayor diversidad ancestral y presenta un sesgo hacia la ascendencia europea (MORAES *et al.*, 2016; GOSÁLVEZ y HORCAJADAS, 2018).

En 1998, el bioquímico estadounidense Craig Venter, quien era investigador en el Instituto NIH, cuestionó el funcionamiento del proyecto público genoma, argumentando que era costoso, lento y empleaba una tecnología poco eficiente. Fue en este contexto en el que su empresa, Celera Genomics, se unió al proyecto y desarrolló un método más rápido para identificar fragmentos del genoma humano, utilizando la técnica conocida como secuenciación al azar o *Shotgun* (MORAES *et al.*, 2016).

La técnica consiste en fragmentar el genoma en pedazos más pequeños, secuenciarlos y luego ensamblarlos para obtener la secuencia completa. Este método resultó en un proceso más eficiente y aceleró significativamente el ritmo de secuenciación del genoma humano. Craig Venter y Celera Genomics jugaron un papel importante en el avance de la secuenciación del genoma humano al introducir esta técnica innovadora, que permitió una secuenciación más rápida y efectiva de los fragmentos de ADN. La **figura 8** ilustra la técnica de Celera Genomic.

Figura 8. Estrategia de Celera Genomic PGH



Nota. El ADN se divide en fragmentos pequeños y se realiza la secuenciación completa de cada uno de ellos. Estos fragmentos se organizan en contigs, basándose en la superposición de las secuencias.

Finalmente, el 12 de febrero de 2001, las revistas Nature y Science anunciaron la publicación de la secuencia del genoma humano. Esta secuencia representó aproximadamente el 90 por ciento de los casi tres mil millones de pares de bases que conforman el genoma humano y se obtuvo mediante los esfuerzos combinados de las entidades públicas y privadas (MORAES *et al.*, 2016; DONG *et al.*, 2021).

CONCLUSIONES RELEVANTES

Algunas de las conclusiones más importantes alcanzadas a partir del PGH fueron:

- El PGH logró la secuenciación completa del genoma humano, identificando los aproximadamente 20.000 a 25.000 genes que lo componen.

- Se descubrió que el genoma humano está estructuralmente organizado, con regiones codificantes (exones) que generan proteínas y regiones no codificantes (intrones) que desempeñan un papel en la regulación génica y otras funciones.
- Se reveló una amplia variabilidad genética entre individuos, con millones de variantes genéticas, como SNPs, inserciones, deleciones y duplicaciones de ADN. Esta variabilidad tiene implicaciones en la salud y la enfermedad, y ha contribuido al avance de la medicina genómica en el diagnóstico y tratamiento de enfermedades.
- La secuenciación del genoma humano permitió la genómica comparativa, que ha proporcionado información sobre la evolución y la relación genética entre especies, identificando regiones conservadas y genes compartidos entre humanos y otros organismos.
- El proyecto ha facilitado la identificación de genes relacionados con enfermedades genéticas hereditarias y variantes genéticas asociadas a enfermedades comunes, lo que ha impulsado el avance de la medicina genómica y mejorado el diagnóstico y tratamiento de enfermedades.

Si bien el PGH fue un avance significativo en la comprensión del genoma y su complejidad, también dejó interrogantes, especialmente en relación con el llamado *ADN basura* y cómo se organiza y funciona la información genética en el genoma humano.

3.3.2. PROYECTO ENCODE

Después de la finalización del Proyecto Genoma Humano (PGH), se inició el análisis del transcriptoma, que estudia la expresión del genoma a nivel de transcripción. Este análisis reveló

Objetivo. Identificar, mapear y catalogar todos los elementos funcionales en el genoma humano y de ratón. Incluye la anotación de todos los genes que codifican proteínas, pseudogenes y ARN, así como los loci transcritos no codificantes del genoma y caracterizar los productos de transcripción.

la compleja red de regulación de los procesos biológicos y proporcionó información para la predicción y prevención de enfermedades.

En septiembre de 2003, se presentó el proyecto ENCODE (Consortio Enciclopedia de Elementos de ADN) como continuación del PGH. El propósito de ENCODE fue caracterizar e identificar elementos funcionales en el genoma humano, como promotores, potenciadores, silenciadores y ARN no codificante, presentes en las regiones no codificantes del genoma (GERSTEIN *et al.*, 2007). El proyecto se centró en comprender la regulación y coordinación de los genes y otros elementos durante

la expresión génica, así como en identificar regiones genómicas asociadas con funciones biológicas específicas. El Consorcio ENCODE se encarga de producir datos de alta calidad y llevar a cabo un análisis integral de los mismos (SIGGENS y EKWALL, 2014). Esta enciclopedia organiza los resultados más destacados de los análisis en anotaciones y proporciona herramientas para buscar y visualizar dichos datos (**figura 9**).

Durante el desarrollo del proyecto ENCODE, se aplicaron diversos métodos para encontrar elementos funcionales en el genoma humano. La anotación del ADN de un genoma implica la identificación estructural de genes, regiones codificantes y motivos, así como la reproducción de funciones a estas regiones (**figura 10**). Consiste en el proceso de identificar y etiquetar todas las características relevantes en una secuencia del genoma.

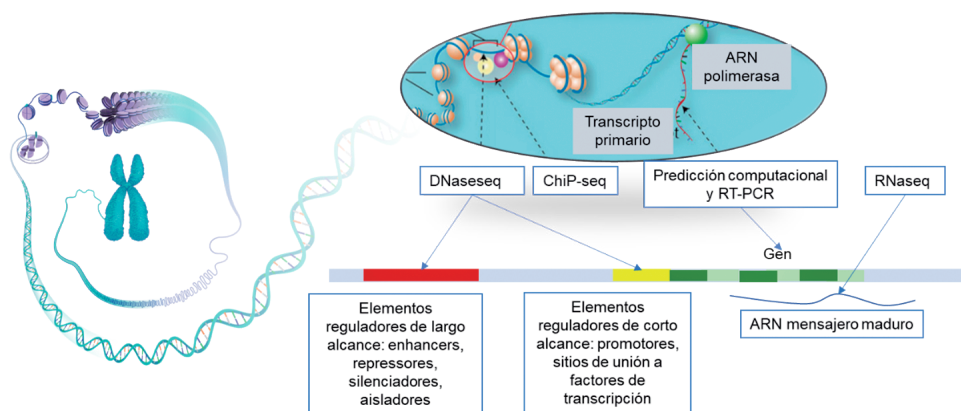
Figura 9. Proyecto ENCODE



Nota. Página principal del proyecto ENCODE. Principales áreas de investigación que se formaron durante el proyecto. Disponible en <https://www.encodeproject.org/>.

Para descubrir y anotar elementos genéticos, se emplea principalmente la técnica de secuenciación de ARN (RNAseq) procedente de diversas fuentes, así como la genómica comparativa y métodos bioinformáticos integrativos. Por otro lado, los elementos reguladores de largo y corto alcance se analizan a través de pruebas de hipersensibilidad del ADN, de metilación del ADN e inmunoprecipitación (IP) de proteínas que interactúan con el ADN y el ARN (SIGGENS y EKWALL, 2014). Estas proteínas incluyen histonas modificadas, factores de transcripción, reguladores de la cromatina y proteínas de unión al ARN (MOORE *et al.*, 2020), posteriormente, se realiza la RNAseq para obtener información detallada de estas interacciones (**figura 10**).

Figura 10. Alcanzando la estructura del genoma humano



Nota. Se ilustran las principales metodologías de ENCODE para el estudio de elementos funcionales del genoma. DNase-seq: secuenciación de sitios de hipersensibilidad a DNasa, ChIP-seq: secuenciación de inmunoprecipitación de cromatina, RT-PCR: reacción en cadena de la polimerasa con transcriptasa inversa, RNA-seq: secuenciación de ARN.

El Consorcio ENCODE proporciona anotaciones de elementos reguladores en cis. Estas anotaciones son clasificadas y organizadas según los marcadores específicos de cada tipo celular, lo que permite un análisis más detallado de la regulación génica en diferentes contextos celulares (MOORE *et al.*, 2020; Siggen y Ekwall, 2014). El proyecto ENCODE inicialmente se llevó a cabo en dos fases:

- *Primera fase piloto.* Tuvo como objetivo analizar el ADN en las regiones funcionales o codificadoras de proteínas, que representan aproximadamente el 1,5 % del genoma. Se llevaron a cabo pruebas de inmunoprecipitación de cromatina seguida de secuenciación, las pruebas permitieron analizar la maquinaria y algunos factores de transcripción. Mediante estudios

comparativos, se identificaron elementos evolutivamente conservados en otras especies, lo que permitió inferir las funciones biológicas.

- *Segunda fase.* Se analizó el 98,5 % restante del genoma humano. Se realizaron estudios en diferentes líneas celulares, se cuantificó el ARN en las células enteras y en diferentes compartimentos celulares, se mapearon las regiones codificadoras de proteínas, se analizaron las modificaciones de histonas y los sitios de unión de los factores de transcripción, así como el mapeo de los sitios de metilación. Como resultado, se llegó a la conclusión de que el 80,4 % del genoma es funcional en al menos una línea celular.

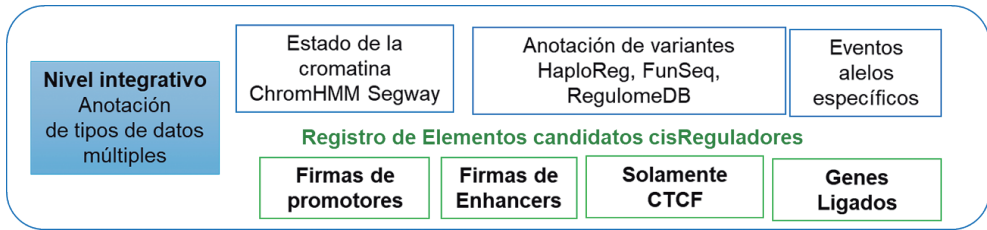
Las fases posteriores se han encargado de organizar e integrar la información de manera que sea fácilmente accesible y visualizable a través de la Enciclopedia ENCODE, que cuenta con dos niveles de anotaciones.

En primer lugar, las **anotaciones de nivel integrador** combinan diversos tipos de datos experimentales y anotaciones obtenidas en campo. Estas anotaciones tienen como objetivo integrar información de diferentes fuentes y proporcionar una visión más completa y amplia (Gerstein *et al.*, 2007).

El núcleo del nivel integrador de la Enciclopedia ENCODE es el registro de candidatos a elementos de reguladores en cis (cCRE, por sus siglas en inglés). Este nivel se basa en datos de alta calidad obtenidos a través de técnicas como DNase-seq y ChIP-seq. Los cCRE son sitios del genoma altamente sensibles a ADNasas que exhiben diversas firmas epigenéticas, lo que los clasifica en tres grupos de cCRE (SIGGENS y EKWALL, 2014; MOORE *et al.*, 2020) (**figura 11**).

- Tipo promotor, que se caracterizan por la modificación de histonas H3K4me3.
- Tipo potenciador, que presentan modificaciones de histonas H3K27ac.
- Otros relacionados con la unión del factor de transcripción CTCF, que desempeña un papel importante en la estructura de la cromatina y la regulación de la transcripción, entre otros procesos.

Figura 11. Nivel integrador enciclopedia ENCODE

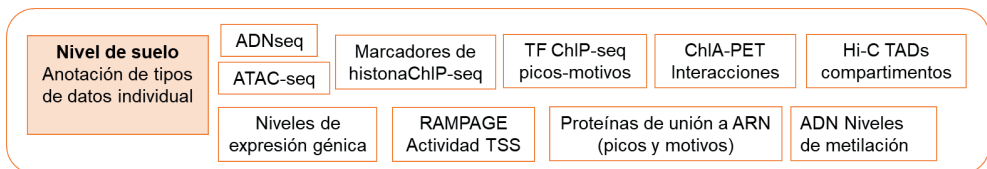


Nota. Utiliza datos múltiples procesados por las canalizaciones de procesamiento uniforme. ChromHMM: Modelo Oculito de Markov para la cromatina, Segway: anotación de segmentos genómico, HaploReg: herramienta de exploración de haplotipos, FunSeq: secuenciación funcional, RegulomeDB: base de datos reguladora, CTCF: factor de unión a CCCTC. Tomada y modificada de Moore et al. (2020).

El nivel integrador utiliza una herramienta que permite explorar y analizar los cCREs, así como su relación con otras anotaciones en la Enciclopedia. Además, se utilizan métodos semiautomatizados para asignar estados de la cromatina basados en datos epigenómicos, en diferentes tipos celulares y tejidos. Esto proporciona una mayor comprensión de la regulación génica y la organización de la cromatina en el genoma (SIGGENS y EKWALL, 2014).

En segundo lugar, las anotaciones a nivel del suelo se generan directamente a partir de los datos experimentales y suelen ser el resultado de procesos uniformes de procesamiento. Estas anotaciones se obtuvieron de manera más directa y están vinculadas de manera más estrecha con los datos experimentales producidos. Se hacen anotaciones de cromatina abierta, enriquecimiento de la marca de histona, unión al factor de transcripción, expresión génica generación de perfiles de actividad del sitio de inicio de transcripción, ocupación de proteínas de unión al ARN, metilación del ADN entre otras (GERSTEIN *et al.*, 2007) (**figura 12**).

Figura 12. Enciclopedia ENCODE nivel de suelo

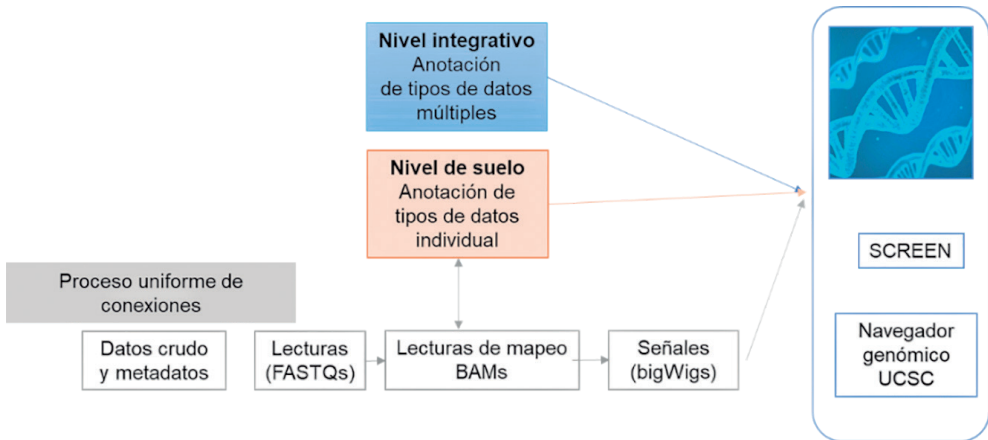


Nota. Anotaciones de experimentos individuales producidos por el proyecto ENCODE. ATAC-seq: ensayo de cromatina accesible mediante transposasa y secuenciación, ChIP-seq: secuenciación de inmunoprecipitación de cromatina, TF ChIP-seq: secuenciación de inmunoprecipitación de cromatina para factores de

transcripción, ChIA-PET: análisis de interacciones de cromatina mediante secuenciación de etiquetas de extremos emparejados, Hi-C: captura de conformación de cromosomas de alta resolución. Técnica que permite mapear interacciones tridimensionales en la estructura del ADN dentro del núcleo celular, TADs: dominios de asociación topológica, RAMPAGE: anotación y mapeo de ARN de promotores para el análisis de expresión génica. Tomada y modificada de Moore et al. (2020).

Finalmente, la herramienta SCREEN (Search Candidate Regulatory Elements by ENCODE) permite buscar y visualizar los cCRE dentro del Registro de ENCODE (**figura 13**). Esta herramienta integra los datos de expresión génica asociados a los cCRE, así como la intersección entre ellos y los SNP de GWAS (estudio de asociación de genoma completo). Para agilizar y facilitar la visualización de las diversas anotaciones epigenéticas, se utiliza el navegador del genoma UCSC. Esto convierte a SCREEN en una poderosa herramienta para la investigación en biología (GERSTEIN *et al.*, 2007; MOORE *et al.*, 2020).

Figura 13. Enciclopedia ENCODE SCREEN



Nota. Integra todos los niveles de anotaciones y datos brutos y permite a los usuarios visualizarlos en el navegador de genomas ESCREEN de la UCSC. FASTQ: Formato FASTQ, BAM: Archivo de Mapeo Binario. Tomada y modificada de Moore (2020).

CONCLUSIONES RELEVANTES

En conclusión, el consorcio ENCODE que lleva cuatro fases, proporciona anotaciones de elementos cisreguladores candidatos en su registro, basándose en datos de alta calidad obtenidos de diferentes técnicas. Estas anotaciones son clasificadas y organizadas según los marcadores específicos de cada tipo celular, lo que permite un análisis más detallado de la regulación génica en diferentes contextos celulares de humano y ratón. Las conclusiones más relevantes son:

- Anotación de elementos funcionales que desempeñan un papel clave en la regulación de la expresión génica. Se ha llevado a cabo una extensa anotación de elementos funcionales, incluyendo promotores, enhancers, sitios de unión de factores de transcripción, elementos reguladores y regiones de cromatina abierta.
- El proyecto ha proporcionado una visión detallada de cómo los elementos funcionales interactúan entre sí y con los genes para regular la expresión génica. Se ha descubierto que muchos elementos reguladores pueden estar lejos de los genes que regulan, y que la interacción tridimensional del genoma juega un papel importante en la regulación génica.
- El proyecto ENCODE ha demostrado que la regulación génica varía significativamente entre individuos y células. Se han identificado variantes genéticas asociadas con diferencias en la regulación génica y se ha demostrado que las diferencias en la expresión génica pueden tener implicaciones en la susceptibilidad a enfermedades y en las características fenotípicas.
- El proyecto ha proporcionado información valiosa para la anotación de variantes genéticas identificadas en GWAS. Esto ha ayudado a comprender mejor cómo las variantes genéticas pueden afectar la función de los elementos reguladores y contribuir a la predisposición a enfermedades.

3.3.3. PROYECTO GENCODE (GENOME CODING)

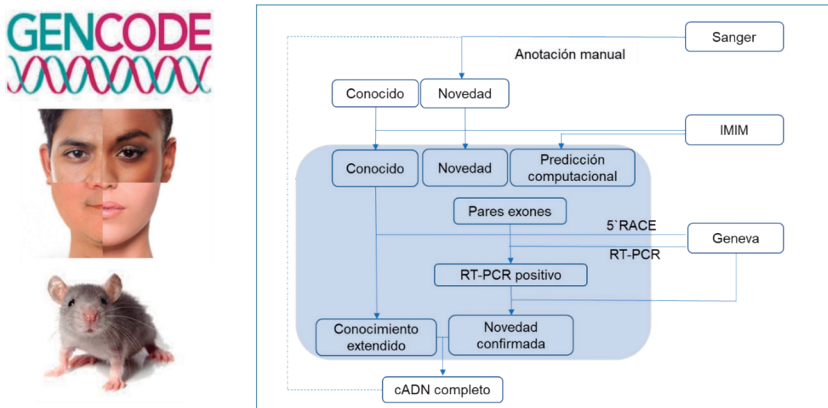
Objetivo. Proporcionar una anotación exhaustiva y precisa de los genes codificantes para proteínas y otros elementos funcionales en el genoma humano y de ratón. Se centra en la identificación y clasificación de genes, así como en la caracterización de sus isoformas y variantes de empalme alternativo. Además de los genes codificantes para proteínas, GENCODE también anota ARN no codificantes y pseudogenes.

El proyecto GENCODE tiene como objetivo generar un catálogo de regiones codificantes de proteínas, ARN no codificantes y pseudogenes en el genoma humano (PEI *et al.*, 2012). Forma parte del proyecto ENCODE y contribuye a comprender la estructura y función del genoma, incluyendo elementos reguladores que controlan la actividad génica en diferentes circunstancias.

Se enfoca en la regulación y coordinación de genes y elementos en la expresión génica, así como en la identificación de regiones genómicas asociadas con funciones biológicas específicas. En la fase actual, se busca mejorar la cobertura y precisión de los conjuntos de genes humanos y de ratón en GENCODE, mediante la anotación de características genéticas basadas en evidencia, como variantes de empalmes alternativos (DERRIEN *et al.*, 2012).

El proyecto GENCODE buscó proporcionar una anotación detallada de los elementos funcionales en el genoma humano. Esto se logró mediante la recopilación de datos de diversas fuentes, incluyendo secuencias de ADN y datos de expresión génica. Expertos en genómica de diferentes partes del mundo entre ellos IMIM (Instituto de Investigaciones Médicas del Hospital del Mar en Barcelona, España) y Ginebra, llevaron a cabo una anotación manual, identificando genes codificantes para proteínas, ARN no codificantes, pseudogenes y otros elementos funcionales. Además, se utilizó anotación automatizada con herramientas computacionales (ver **figura 14**). Los datos anotados se integraron y se utilizaron rigurosos controles de calidad para asegurar la precisión de la anotación. El resultado fue un catálogo completo de elementos funcionales en el genoma humano, que sirve como referencia valiosa investigaciones genómicas (HARROW *et al.*, 2012; PEI *et al.*, 2012).

Figura 14. La canalización GENCODE



Nota. Se muestra el flujo de datos entre los tres grupos involucrados en el consorcio GENCODE (IMIM y Ginebra) para producir una anotación verificada experimentalmente de la región ENCODE. Elaborada con elementos de GENCODE (2025).

El proyecto GENCODE ha permitido identificar los exones, intrones, sitios de inicio y finalización de la transcripción, y las diferentes isoformas génicas generadas por un mismo gen. Además, se descubrió que una gran proporción del genoma humano se transcribe en ARN, pero no codifica proteínas. Estos ARN no codificantes desempeñan funciones importantes en la regulación génica y otros procesos biológicos. El proyecto también reveló una gran variabilidad en la estructura génica entre individuos, incluyendo variantes genéticas que afectan a los límites de los genes y la estructura de los ARN mensajeros, lo que tiene implicaciones en la expresión génica y en la susceptibilidad a enfermedades (PEI *et al.*, 2012).

El proyecto GENCODE estimó que aproximadamente el 2,94 % de todo el genoma es codificante para proteínas, lo que destaca la importancia de estas secuencias en la función biológica. Además, se descubrió que el 80,4 % de las secuencias del genoma tienen un papel en la regulación de los genes, lo que resalta la complejidad de los mecanismos de control génico. Por otro lado, se mapearon los pseudogenes, duplicados de genes que parecen estar inactivos a nivel transcripcional, encontrándose dispersos en todo el genoma, lo que sugiere su posible papel en la evolución y la diversidad genética, aunque no están activamente transcritos (HARROW *et al.*, 2012; PEI *et al.*, 2012).

CONCLUSIONES RELEVANTES

El proyecto GENCODE ha proporcionado una serie de conclusiones que han contribuido significativamente a nuestra comprensión del genoma humano y su complejidad. Algunas de las conclusiones más destacadas incluyen.

- GENCODE ha sido fundamental para la identificación de nuevos genes en el genoma humano y otros genomas modelados.
- El proyecto ha permitido la caracterización detallada de los transcriptomas, es decir, el conjunto de ARN transcritos a partir de los genes. Esto ha revelado la existencia de múltiples variantes de transcripción para muchos genes, lo que sugiere una mayor complejidad en la regulación génica.
- GENCODE ha contribuido significativamente a la anotación de elementos reguladores en el genoma, como promotores, potenciadores y sitios de unión de factores de transcripción. Estos elementos desempeñan un papel crucial en la regulación precisa de la expresión génica.
- El proyecto ha destacado la presencia y relevancia de ARN no codificantes, que no se traducen en proteínas, pero pueden desempeñar funciones regulatorias y estructurales importantes en la célula.
- GENCODE ha trabajado en estrecha colaboración con otros proyectos y bases de datos genómicas para integrar y compartir datos de manera eficiente. Esto ha promovido la colaboración y el avance de la investigación en genómica.

Con el avance en estos proyectos han surgido proyectos específicos destinados a comprender las variaciones genéticas entre diferentes individuos y poblaciones. Uno de estos proyectos fue el HapMap, y en 2008 se inició el Proyecto 1000 Genomas.

3.3.4. PROYECTO HAPMAP

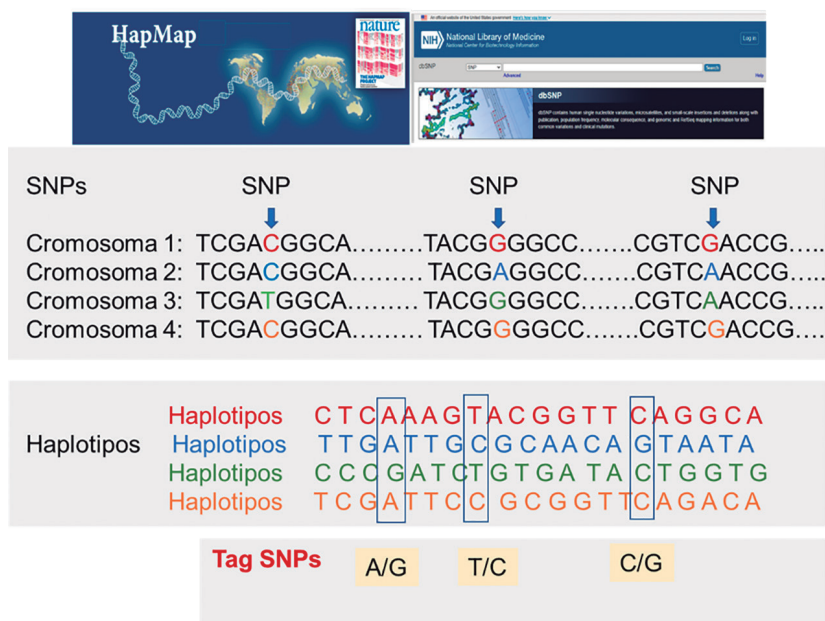
Objetivo. Identificar y clasificar las similitudes y diferencias genéticas entre los individuos representativos de cuatro poblaciones humanas principales. Además, buscó analizar todos los SNP conocidos que presentaron una frecuencia deseada de, por lo menos, el 5 % en cada muestra.

El proyecto internacional HapMap (Mapa de Haplotipos Humanos) fue una iniciativa colaborativa que involucró a varios países, como el Reino Unido, Cana-

dadá, Japón, China, Estados Unidos y Nigeria. Realizado entre 2002 y 2009, su objetivo principal fue identificar y mapear la variación genética en el genoma humano (GABRIEL *et al.*, 2002).

Este proyecto desempeñó un papel fundamental en el descubrimiento y análisis de millones de polimorfismos de nucleótido simple (SNP) en el genoma humano. Los SNP son variaciones genéticas comunes en las cuales un solo nucleótido difiere entre los individuos (ver **figura 15**). Estos SNP han demostrado ser marcadores valiosos para estudios de asociación, ya que permiten identificar genes implicados en enfermedades genéticas comunes y complejas, como el cáncer, la hipertensión y la diabetes, que tienen un componente genético y ambiental.

El descubrimiento de los bloques haplotípicos, que se caracterizan por tener una baja tasa de recombinación y están separados por segmentos con una alta tasa de recombinación, ha sido un adelanto significativo en el área de la genómica. Esto ha permitido la identificación de los llamados SNP etiqueta (Tag SNP). Gracias a este descubrimiento, se ha logrado reducir el número inicial de millones de SNP a una cantidad más manejable, lo que facilita y optimiza el estudio de los polimorfismos genéticos (International HapMap Consortium, 2005). Esta estrategia ha demostrado ser más accesible y efectiva para analizar y comprender la diversidad genética (**figura 15**).

Figura 15. *SNP, Haplotipos y Tags SNP*

Nota. Un SNP es una variación en la cadena de ADN que afecta a una sola base. Un haplotipo es un conjunto de variaciones de ADN a lo largo de un cromosoma que se heredarán juntas debido a su proximidad y la falta de recombinación entre ellas. Tag SNP, son variaciones derivadas de los haplotipos.

El primer polimorfismo fue descrito por Whyman y White en 1980. Este polimorfismo en particular era bialélico, lo que implica que los individuos pueden tener las variantes AA, AT o TT en ese *locus* específico. Desde entonces, con los avances en el PGH y otros estudios genéticos, se han logrado identificar y catalogar millones de SNP en el genoma humano. Estos datos están disponibles en diversas bases de datos públicas y privadas, lo que ha permitido un acceso amplio y la utilización de esta información para la investigación y los estudios genéticos en la población.

La presencia de SNP en el genoma humano, que ocurre aproximadamente cada 300 pares de bases, ha sido aprovechada para identificar genes implicados en diversas enfermedades. Se realizan estudios de asociación en los que se comparan las secuencias de ADN de individuos afectados por una enfermedad específica con las de individuos sanos (controles). Estos estudios buscan identificar SNP que frecuentemente están presentes en los pacientes con la enfermedad en cuestión. Por ejemplo, al analizar el ADN de personas con y sin artritis, se puede detectar un SNP que se encuentra con mayor frecuencia en los pacientes con artritis. Esto

proporciona pistas importantes para la localización de un gen específico que puede estar involucrado en el desarrollo o la susceptibilidad a la artritis.

El proyecto buscó crear un catálogo de variantes genéticas comunes y simples, y examinar su distribución en las distintas poblaciones. Además, se investigaron los bloques haplotípicos y los Tags SNP que definen la variación en la población, con el objetivo de identificar los genes implicados en enfermedades complejas y comprender las respuestas individuales a los fármacos (TOURDOT *et al.*, 2021; INTERNATIONAL HAPMAP CONSORTIUM, 2005).

Se realizó un análisis en un total de 269 muestras que representaban tres grupos poblacionales distintos africanos (yorubas de Ibadán, Nigeria), asiáticos del este (japoneses de Tokio y chinos Han de Pekín) y europeos (residentes de Utah con ascendencia del norte y oeste de Europa). El proyecto HapMap se llevó a cabo en dos fases principales (International HapMap Consortium, 2005).

- **Fase I.** En esta etapa, se procedió a genotipar al menos un SNP común, con una frecuencia esperada de aproximadamente 0,05 o superior, cada 5 Kb del genoma en las 269 muestras. Esto implicó el análisis de más de un millón de SNP en cada muestra. Se seleccionaron 10 regiones de 500 Kb cada una, basadas en el proyecto ENCODE. Estas regiones fueron secuenciadas en 48 individuos y todos los SNP descubiertos en el análisis, así como aquellos ya existentes en la base de datos públicos de SNP (dbSNP). El proyecto HapMap permitió confirmar la existencia de puntos calientes de recombinación junto con segmentos largos que mostraron un fuerte desequilibrio de ligamiento y diversidad haplotípica (International HapMap Consortium, 2005).
- **Fase II.** En esta segunda fase del proyecto, se analizaron adicionalmente 4,6 millones de SNP en cada muestra analizada en el proyecto HapMap. Esta etapa amplió significativamente el número de SNP investigados y proporcionó una visión más detallada de la genética presente en las poblaciones estudiadas (International HapMap Consortium, 2005).

Muchos investigadores de diversos países se han unido al proyecto HapMap con un enfoque inicial en la teoría, los SNP y la asociación, sin tener en cuenta el mapeo completo del genoma. Estos estudios revelaron que, a pesar de contar con muestras pequeñas, el HapMap proporciona mapas mayormente confiables de haplotipos y desequilibrio de ligamiento (LD), lo cual ayuda a interpretar las diferencias entre poblaciones y mejorar el mapeo de asociaciones genéticas (TOURDOT *et al.*, 2021).

Las muestras de HapMap han sido ampliamente empleadas en investigaciones relacionadas con la farmacogenómica y otras áreas de estudio. Por ejemplo, se

ha utilizado la secuenciación del genoma completo de muestras de HapMap para identificar regiones genéticas asociadas con la respuesta a la quimioterapia y la citotoxicidad (TOURDOT *et al.*, 2021). El objetivo de estos estudios es descubrir marcadores genéticos que puedan aplicarse en la evaluación clínica y en la personalización de tratamientos médicos.

CONCLUSIONES RELEVANTES

Algunas de las conclusiones relevantes alcanzadas a partir de este proyecto incluyen.

- El proyecto HapMap permitió identificar haplotipos, que son secuencias de variantes genéticas cercanas que tienden a transmitirse juntas. Estos haplotipos pueden utilizarse para inferir información sobre variantes genéticas no secuenciadas y ayudan a comprender la estructura genética de diferentes poblaciones. Además, el proyecto demostró que la variación genética no se distribuye de manera uniforme en el genoma, sino que tiende a agruparse en regiones conocidas como bloques de desequilibrio de ligamiento (*linkage disequilibrium blocks*).
- El proyecto identificó millones de polimorfismos de nucleótido único (SNP) en el genoma humano. Estos SNP son variantes genéticas comunes que incorporan la mayoría de la variación genética en la población humana. El mapeo de SNP en el proyecto HapMap ha sido de gran ventaja para identificar genes asociados con enfermedades, así como para estudiar la relación entre la variación genética y rasgos complejos.
- El proyecto ayudó a definir y caracterizar la estructura genética de diferentes poblaciones humanas alrededor del mundo. Se identificaron diferencias genéticas significativas entre poblaciones y se demostró que las poblaciones humanas no son homogéneas desde el punto de vista genético. Esta información ha sido crucial para el estudio de las bases genéticas de enfermedades y para comprender las migraciones humanas y la historia evolutiva.
- El proyecto proporcionó datos valiosos para la identificación de variantes genéticas asociadas con enfermedades comunes, como enfermedades cardiovasculares, diabetes, cáncer y enfermedades neurológicas. Además, el conocimiento obtenido a través del proyecto HapMap ha sido fundamental para el campo de la farmacogenómica, que busca entender cómo las variaciones genéticas influyen en la respuesta de los individuos a los medicamentos.

- Gracias al uso de *haploview*, se pueden diseñar adecuadamente los estudios y seleccionar los SNP (Tag SNP) que definen el bloque. Esta información es esencial para identificar SNP causales de enfermedades con componentes genéticos. Lo que permite mejorar y actualizar la elección de estos SNP, en base a una comprensión más completa y actualizada de la estructura de los bloques haplotípicos.

3.3.5. PROYECTO DE LOS 1.000 GENOMAS

Objetivo. Secuenciar el genoma completo de al menos mil individuos de diversas poblaciones alrededor del mundo, con el fin de obtener una comprensión más completa de la diversidad genética humana y su relevancia para la salud y la enfermedad. Esto incluyó el estudio de variaciones genéticas comunes y raras, así como mutaciones genéticas asociadas con enfermedades genéticas hereditarias y cáncer.

El proyecto 1.000 Genomas fue una colaboración internacional que tuvo como objetivo crear un mapa detallado de variación genética en la población humana. El proyecto fue anunciado en 2008 y se completó en 2015.

Sin embargo, a pesar de la cobertura de datos relativamente alta proporcionada por HapMap, el proyecto 1.000 Genomas generará un mapa más detallado de la variación genética humana.

Esto permitirá un análisis más profundo de cómo la variación genética contribuye a la respuesta a los medicamentos y otros rasgos complejos, como la expresión génica (ZHANG y DOLAN, 2010). Los estudios futuros que utilicen este nuevo recurso de datos podrán mejorar significativamente nuestra comprensión de las bases genéticas de la respuesta a los medicamentos, lo que a su vez ayudará a avanzar en el campo de la medicina personalizada y en el desarrollo de nuevas terapias.

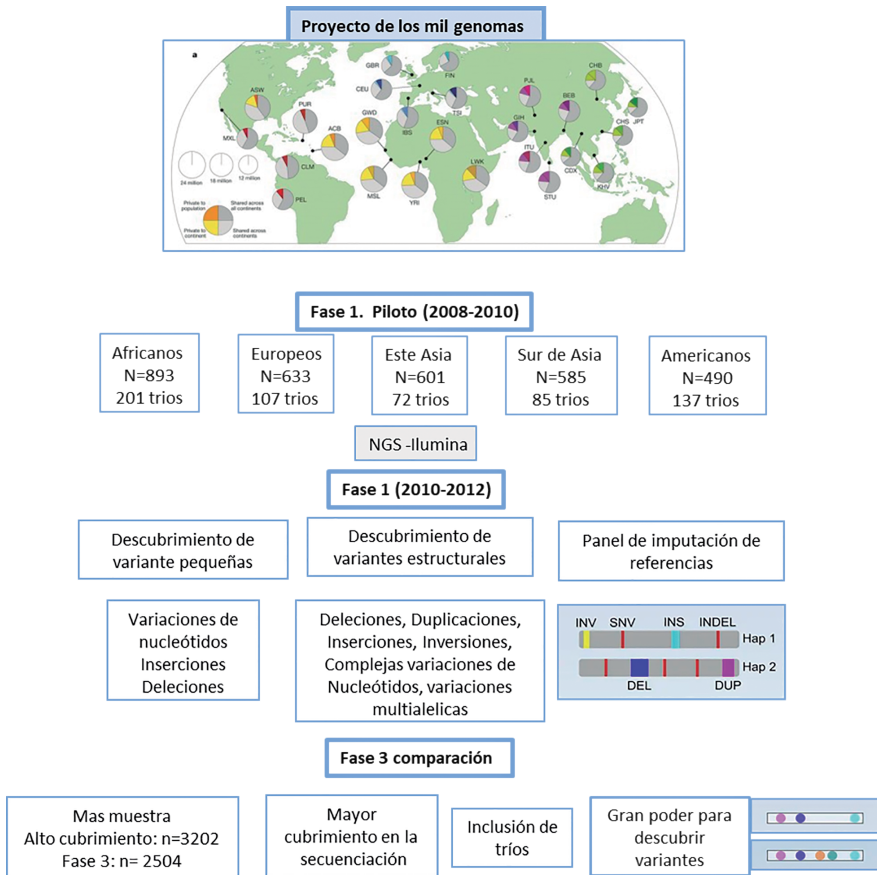
Para seleccionar a los individuos participantes en el proyecto, se tuvieron en cuenta factores como la diversidad étnica, la disponibilidad de muestras de ADN de alta calidad y la representatividad de diferentes poblaciones en todo el mundo. Se obtuvieron muestras de sangre, saliva u otros tejidos de miles de individuos voluntarios, a los cuales se les extrajo el ADN para su secuenciación. Así mismo, una de las fuentes de muestras utilizadas en este proyecto son las muestras de International HapMap, que consisten en líneas de células linfoblastoides derivadas de individuos pertenecientes a diferentes poblaciones mundiales (ZHANG y DOLAN, 2010; BYRSKA-BISHOP *et al.*, 2022).

Se utilizaron tecnologías de secuenciación de siguiente generación (NGS, por sus siglas en inglés) (ver sección Capítulo 2, sección 2.2). Estas técnicas han permitido secuenciar grandes cantidades de ADN de manera más rápida y

a un costo más bajo que los métodos anteriores. Además, permitieron obtener información detallada sobre la secuencia del ADN de cada individuo (**figura 16**). (BYRSKA-BISHOP *et al.*, 2022).

Después de la secuenciación, se utilizó *software* especializado para identificar variantes genéticas en las secuencias de ADN de los individuos. Esto incluyó la identificación de polimorfismos de nucleótido único (SNP), inserciones y deleciones de ADN, y otras variaciones genéticas. Estas variantes se compararon con las secuencias de referencia del genoma humano para determinar las diferencias y los patrones de variación genética.

Figura 16. Proyecto 1.000 Genomas



Nota. Se muestra la cantidad de individuos participantes y las diferentes variaciones que se estudiaron durante el desarrollo del proyecto. Tomado y modificado de Byrska-Bishop *et al.* (2022).

Se llevaron a cabo análisis bioinformáticos para caracterizar y clasificar las variantes genéticas identificadas. Esto incluyó la anotación de las variantes en relación con los genes, la identificación de variantes comunes y raras, y la evaluación de su posible impacto en la función génica y la salud humana (BYRSKA-BISHOP *et al.*, 2022)

A medida que se completaron las secuenciaciones, los datos se hicieron públicos y se pusieron a disposición de la comunidad científica para su análisis y uso en la investigación genómica. Esto permitió a otros investigadores utilizar los datos del proyecto de los 1.000 Genomas para realizar estudios adicionales sobre la genética humana (BYRSKA-BISHOP *et al.*, 2022). Entre ellos, se llevó cabo el análisis para investigar la estructura genética de las poblaciones estudiadas y las relaciones entre ellas. Esto incluyó la estimación de la diversidad genética, el análisis de ascendencia ancestral y la detección de migraciones y mezclas genéticas.

CONCLUSIONES RELEVANTES

El proyecto de los 1.000 Genomas proporcionó una serie de conclusiones y descubrimientos importantes sobre la variación genética humana. Aquí hay algunas de las principales conclusiones del proyecto.

- El proyecto reveló una amplia diversidad genética en las poblaciones humanas. Se encontraron millones de variantes genéticas, incluyendo tanto las comunes como las raras, lo que resalta la complejidad de la genética humana.
- El estudio permitió identificar muchas variantes genéticas asociadas con enfermedades. Estas variantes proporcionan información importante para la comprensión de las bases genéticas de enfermedades comunes, como el cáncer, las enfermedades cardiovasculares, la diabetes y enfermedades neurológicas.
- Se descubrieron miles de mutaciones genéticas raras en los genomas secuenciados. Estas mutaciones pueden tener un impacto significativo en la salud y pueden estar asociadas con enfermedades genéticas raras y síndromes hereditarios.
- El proyecto permitió rastrear el flujo de genes entre diferentes poblaciones y proporcionó información sobre los ancestros comunes de diferentes grupos étnicos. Esto ayudó a comprender mejor la historia evolutiva de la humanidad y las migraciones humanas a lo largo del tiempo.
- Los datos del proyecto sentaron las bases para la genómica personalizada, que busca utilizar la información genética individual para predecir el riesgo de enfermedades y personalizar los tratamientos médicos en función de la genética de cada persona.

3.3.6. PROYECTO PROTEOMA HUMANO (HPP)

Objetivo. mapear y caracterizar el proteoma completo del cuerpo humano, busca identificar y describir todas las proteínas presentes en el cuerpo humano, incluyendo sus isoformas, modificaciones postraduccionales y variantes. También tiene como objetivo proporcionar herramientas y recursos valiosos para la investigación en proteómica.

La Organización del Proteoma Humano (HUPO, por sus siglas en inglés) es una entidad internacional que promueve la investigación en el campo del proteoma humano. Fundada en 2001, reúne a científicos

y profesionales de diferentes disciplinas relacionadas con la proteómica. Su principal objetivo es mapear y comprender los componentes proteicos del cuerpo humano. Para lograr esto, HUPO lanzó el Proyecto del Proteoma Humano (HPP) en 2010, una iniciativa internacional que se enfoca en el mapeo del proteoma humano (OMENN *et al.*, 2018). El HPP busca entender el proteoma en el contexto de redes de redes, colaborar con consorcios complementarios, desarrollar estrategias futuras y servir como punto focal para investigadores en mecanismos, diagnósticos y terapias. Sus metas son mapear el proteoma humano, identificar variantes y modificaciones postraduccionales, y comprender las interacciones proteicas en redes y vías biológicas. Además, busca aplicar los conocimientos de proteómica en el campo clínico y biomédico, incluyendo el desarrollo de biomarcadores y dianas terapéuticas (OMENN *et al.*, 2023).

Para llevar a cabo el Proyecto del Proteoma Humano, se recolectan muestras de tejidos, células y fluidos biológicos, como sangre, orina y líquido cefalorraquídeo, de individuos sanos y afectados por diversas enfermedades. Con tecnologías de alto impacto se realiza la extracción de proteínas utilizando métodos de lisis celular y fraccionamiento para separar las distintas fracciones proteicas.

Se utilizan técnicas de separación de proteínas, como electroforesis bidimensional, cromatografía líquida de alta resolución (HPLC, por sus siglas en inglés) y espectrometría de masas, para identificar y cuantificar las proteínas presentes en las muestras (**figura 17**). Estas técnicas se abordarán en el próximo capítulo. Se generan grandes volúmenes de datos proteómicos que se utilizan para construir bases de datos y mapas de interacciones proteína-proteína. Además, se buscan aplicaciones clínicas y biomédicas, como el desarrollo de biomarcadores y terapias más efectivas (OMENN *et al.*, 2023).

Figura 17. Base de datos Proyecto Proteoma Humano



Nota. Se muestra el portal de HUPPO y las secuencias de actividades desarrolladas en el proyecto. Disponible en <https://hupo.org/human-proteome-project>.

Finalmente, el HPP establece metas y estrategias para el mapeo y la caracterización del proteoma humano, abordando desafíos técnicos y científicos.

CONCLUSIONES RELEVANTES

El HPP es un proyecto en curso y se espera que se realicen más descubrimientos y avances, sin embargo, estas son algunas de las conclusiones a la fecha.

- El HPP ha contribuido a la identificación y caracterización de un gran número de proteínas humanas previamente desconocidas.
- Se ha descubierto que las proteínas humanas pueden tener múltiples isoformas y variantes debido a modificaciones postraduccionales, lo que tiene implicaciones en la función y la regulación de las proteínas.
- Se han realizado esfuerzos para anotar y clasificar las proteínas humanas según su función, localización subcelular y participación en vías biológicas específicas.
- El HPP ha trabajado en el establecimiento de estándares de calidad y mejores prácticas para la identificación y cuantificación de proteínas, lo que mejora la reproducibilidad y la comparabilidad de los datos proteómicos generados en diferentes laboratorios.
- Se han desarrollado herramientas y recursos informáticos para la integración y el análisis de datos proteómicos, lo que permite una comprensión más completa de las interacciones proteína-proteína y las redes biológicas.

Los avances logrados a través del HPP tienen el potencial de tener un impacto significativo en el campo clínico y biomédico. Esto incluye el desarrollo de biomarcadores para el diagnóstico temprano de enfermedades, la identificación de nuevas dianas terapéuticas y el diseño de terapias personalizadas.

3.3.7. PROYECTO PANGENOMAS

Objetivo. Comprender mejor la diversidad genómica dentro de una especie y cómo esta influirá en las características y la adaptación de los organismos. Busca impulsar la investigación médica y el desarrollo de medicamentos. Al resaltar la diversidad humana, se espera que ayude a comprender cómo las variaciones genéticas afectan la salud y conduzca a una medicina más favorable para todos.

El término pangenoma se refiere al conjunto completo de genes de todas las cepas que forman un organismo que comparten un ancestro común filogenético. En otras palabras, un pangenoma es la combinación de todos los genomas de los organismos que pertenecen al mismo *clado* (**figura 18**). Hace aproximadamente 15 años, Tettelin y su equipo de colaboradores

idearon el concepto de pangenoma con el propósito de explicar y representar la totalidad genómica de un grupo taxonómico específico, como una especie, serovar, filo o reino. Desde entonces, la terminología asociada a este concepto se ha expandido considerablemente.

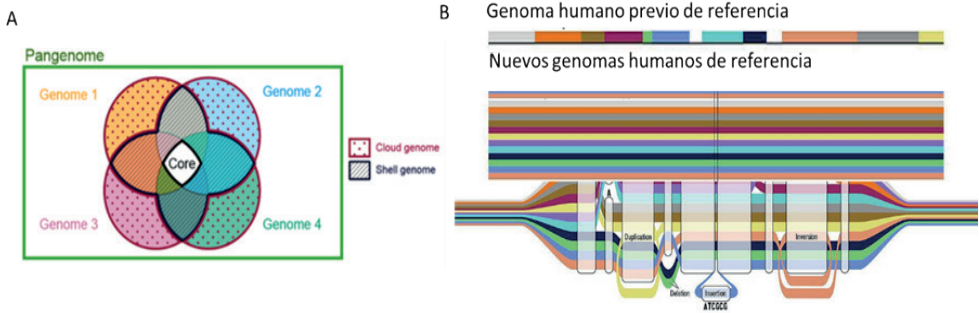
El concepto de pangenoma, que incluye tanto los genes esenciales (o *core* genes) que están presentes en todos los genomas de una especie, como los genes dispensables (o genes de la nube y genes de la concha) que pueden variar entre cepas. Según la clasificación de Tettelin *et al.* (2005), el pangenoma está compuesto por genes esenciales, genes dispensables y genes específicos de cepa, que son únicos en un genoma en particular.

Además, el pangenoma incluye el supragenoma o genoma extendido, que comprende los genes distribuidos y genes únicos (LAPIERRE y GOGARTEN, 2009), representados en la figura como el genoma de la nube y el genoma de la concha. Estos genes aportan diversidad genética y funcionalidad adicional en diferentes contextos ambientales. Por último, las regiones flexibles (RODRÍGUEZ-VALERA Y USSERY, 2012) son segmentos del genoma que contienen estas variaciones y que pueden incluir duplicaciones, deleciones, inserciones e inversiones, como se muestra en la segunda parte de la figura.

En conclusión, según la definición original, el genoma central describe el conjunto de secuencias compartidas por todos los miembros del grupo taxonómico de interés, el genoma dispensable abarca un subconjunto de secuencias compartidas por algunos miembros del grupo (influyendo la diversidad del grupo, como vías bioquímicas alternativas, adaptación a nichos, resistencia a

antibióticos, etc.), mientras que el pangenoma representa simplemente la unión de los genomas central y dispensable (describiendo la totalidad de los taxones en términos de conjuntos de secuencias).

Figura 18. *Proyecto Pangenomas*



Nota. Se muestra en el pangenoma de la figura (A) tiene tres zonas de genomas (genoma core, shell y cloud). El core comprende los genes presentes en la totalidad de los genomas analizados. El shell o genoma de concha incluye genes compartidos por la mayoría de los genomas, pero no en su totalidad, lo que refleja genes que son comunes, pero no universales. Finalmente, el genoma de nube (cloud) comprende los genes presentes en un solo genoma o que tienen una ocurrencia muy baja (menos del 10 %) En (B) se muestra la representación gráfica del pangenoma humano realizada por Darryl Leja (NHGRI) (Institutos Nacionales de Salud [NIH], s.f.). Tomado de <https://www.nih.gov/news-events/news-releases/scientists-release-new-human-pangenome-reference>.

Este proyecto es una iniciativa de investigación genómica que se centra en el estudio y la comparación de los genomas de una especie en particular, en lugar de analizar solo un genoma de referencia que es el enfoque tradicional, el concepto de pangenoma reconoce que las especies pueden tener una diversidad genómica significativa (WANG *et al.*, 2022). Al analizar múltiples genomas de una especie, se pueden identificar genes presentes en algunos individuos, pero ausentes en otros, así como variantes genéticas que pueden estar asociadas con rasgos específicos.

El proyecto utiliza técnicas de secuenciación de nueva generación (NGS) para secuenciar y analizar múltiples genomas de una especie. Estos genomas se comparan entre sí para identificar genes comunes y genes específicos de subpoblaciones o linajes particulares. Además, se investiga cómo los pangenomas pueden influir en la evolución, la adaptación y la respuesta a factores ambientales o enfermedades (WANG *et al.*, 2022).

El Proyecto Pangenomas ha revelado que la diversidad genómica dentro de una especie puede ser mucho mayor de lo que se pensaba inicialmente, lo que

tiene indicaciones importantes para la comprensión de la genética, la evolución y la biología de las especies. También proporciona una base de datos más completa y precisa para el estudio de la genética y la biología de una especie en particular (LAPIERRE Y GOGARTEN 2009).

PanSeq es una base de datos de pangenomas que almacena secuencias de genomas completos y permite la comparación y análisis de genes y regiones genómicas comunes y específicas en diferentes cepas o linajes de una especie. Esta herramienta permite identificar las regiones principales y accesorios en un conjunto de secuencias genómicas, se pueden extraer las regiones exclusivas de un genoma o grupo de genomas, así como identificar SNP dentro de las regiones genómicas compartidas. Además, genera archivos que pueden ser utilizados en programas de filogenia, tanto en la presencia/ausencia de regiones accesorias como en SNP dentro de las regiones centrales (RODRÍGUEZ-VALERA Y USSERY, 2012).

El Consortium Reference de Pangenoma Humano (HPRC) ha publicado en la revista Nature el borrador del genoma como resultado de un proyecto internacional iniciado en 2019 (VERNIKOS, 2020). El objetivo principal de este proyecto es mapear toda la variación genética humana para establecer una referencia completa que permita comparar secuencias genéticas en investigaciones futuras. El borrador del pangenoma, que abarca una colección de secuencias obtenidas de 47 personas de diversas partes del mundo, ha sido publicado después de la divulgación en 2022 de la primera secuencia completa del genoma humano. A diferencia del PGH original y su sucesor, que se basaron principalmente en el ADN de un solo individuo, el borrador del pangenoma incluye una representación más precisa y diversa de la variación genómica global (WANG *et al.*, 2022).

El HPRC tiene como objetivo construir el pangenoma humano de la más alta calidad posible, aprovechando las innovaciones en tecnología, diseño de estudios y asociaciones globales. Este proyecto busca mejorar la representación de datos y optimizar los análisis para permitir el ensamblaje rutinario de genomas diploides completos (VERNIKOS, 2020).

Con una atención especial a los marcos éticos, la referencia del pangenoma humano no solo mejorará los estudios de asociación gen-enfermedad en diferentes poblaciones, sino que también expandirá el alcance de la investigación genómica a las regiones más repetitivas y polimórficas del genoma. Además, esta referencia servirá como un recurso genético invaluable para futuras investigaciones biomédicas y medicina de precisión (NAITHANI *et al.*, 2021).

CONCLUSIONES RELEVANTES

Algunas de las conclusiones más destacadas son las siguientes.

- El estudio del pangenoma ha revelado una mayor diversidad genética en comparación con el genoma de referencia. Esto demuestra la importancia de tener en cuenta la escasez genética en diferentes poblaciones para comprender mejor cómo las variaciones genéticas influyen en la salud y las enfermedades.
- Mediante el análisis del pangenoma, se han identificado regiones genómicas exclusivas presentes en ciertas cepas o grupos de genomas. Estas regiones pueden ser de gran importancia para comprender las características únicas de ciertas cepas y su relación con la salud y las enfermedades.
- El pangenoma ha permitido identificar variantes de un solo nucleótido (SNP) dentro de las regiones genómicas centrales compartidas entre diferentes cepas. Estas variaciones pueden tener un impacto significativo en la susceptibilidad a enfermedades y en la respuesta a tratamientos médicos.
- El pangenoma proporciona información valiosa para la construcción de árboles filogenéticos basados tanto en la presencia/ausencia de regiones genómicas accesorias como en SNP dentro de las regiones genómicas centrales. Además, esta información puede contribuir al desarrollo de una medicina más personalizada y precisa, teniendo en cuenta la diversidad genética de los individuos.

3.4. NUEVOS PROYECTOS EMERGENTES EN CIENCIAS ÓMICAS

El surgimiento de las ciencias ómicas ha sido impulsado por una combinación de avances tecnológicos y proyectos de investigación. A continuación, se mencionan algunos de los proyectos que han surgido como resultado de los avances en las principales ciencias ómicas como el proyecto en metagenómica, epigenómica entre otros. Algunos de estos temas se tratarán en los capítulos siguientes del libro.

3.4.1. PROYECTOS EN METAGENÓMICA

Los proyectos de metagenómica tienen como objetivo principal caracterizar y comprender la diversidad y función de los microorganismos en su entorno natural (CHOPRA *et al.*, 2020). Este proyecto incluye otros proyectos, a saber:

- Proyecto Microbioma Humano (HMP). El proyecto lanzado en 2007 busca caracterizar el microbioma humano en diferentes partes del cuerpo, utilizando técnicas de NGS. Se ha demostrado que el microbioma desempeña un papel crucial en funciones metabólicas, fisiológicas y en la protección

contra patógenos. Desequilibrios en el microbioma se han asociado con diversas enfermedades, como trastornos gastrointestinales, obesidad, enfermedades metabólicas, autoinmunidad y trastornos neuropsiquiátricos (Integrative HMP Research Network Consortium, 2019).

- Proyecto Microbioma Intestinal Humano. Se centra específicamente en el microbioma intestinal humano. El objetivo es mapear la diversidad y la función de las comunidades microbianas presentes en el tracto gastrointestinal y su implicación en la salud y las enfermedades (CANI, 2018).
- Proyecto Microbioma Terrestre. Este proyecto tiene como objetivo catalogar y comprender la diversidad microbiana presente en diferentes ecosistemas terrestres, incluyendo suelos, agua dulce, océanos y el aire. El objetivo es obtener una visión global de los microorganismos en la Tierra y su papel en los ciclos biogeoquímicos (GILBERT *et al.*, 2018).
- Proyecto Tara Océanos. Este proyecto recopiló muestras de agua de océanos de todo el mundo para analizar el microbioma marino y su impacto en los ecosistemas oceánicos. Se obtuvieron datos genómicos de una amplia gama de microorganismos, incluyendo bacterias, arqueas y virus (PIRELLA *et al.*, 2020).

3.4.2. PROYECTOS EN EPIGENÓMICA

La epigenómica es el estudio de las modificaciones químicas y estructurales en el ADN y las histonas que regulan la expresión génica sin cambiar la secuencia del ADN. El Consorcio del Epigenoma Humano es una colaboración público-privada que tiene como objetivo identificar y clasificar posiciones variables de metilación (MVP) en el genoma humano (FU *et al.*, 2023). Los proyectos de epigenómica se enfocan en investigar cómo estas modificaciones epigenéticas influyen el desarrollo, la enfermedad y otros procesos biológicos.

El Proyecto ENCODE, mencionado anteriormente, incluye un componente de epigenómica. Su objetivo es identificar y caracterizar todos los elementos funcionales en el genoma humano, incluyendo las regiones epigenéticas que regulan la expresión génica (SIGGENS y EKWAL, 2014). A continuación, se presentan otros proyectos emergentes en el campo de la epigenómica.

- Mapa de ruta de la epigenómica. Este proyecto fue una iniciativa colaborativa que mapeó el epigenoma humano en diferentes tipos de células y tejidos. El objetivo era proporcionar un recurso completo de datos epigenéticos para comprender la regulación génica y su papel en el desarrollo humano y las enfermedades, como por ejemplo la epigenética asociada con estímulos ambientales y el riesgo de cáncer (HERCEG *et al.*, 2018).

- BLUEPRINT. Este proyecto se centró en el mapeo de la epigenómica en diferentes tipos de células sanguíneas humanas. Su objetivo era proporcionar una visión detallada de las modificaciones epigenéticas en relación con la hematopoyesis, las enfermedades de la sangre y otros procesos biológicos entre ellos la inmunidad (EISSENBERG, 2014).
- BLUEPRINT-EPIC. Se centra en el estudio de la epigenómica en células del sistema inmunológico humano. Su objetivo es comprender cómo las modificaciones epigenéticas influyen en la respuesta inmunológica y su papel en enfermedades autoinmunes y trastornos relacionados (EISSENBERG, 2014).
- Consorcio Internacional del Epigenoma Humano (IHEC). Es una colaboración global que tiene como objetivo mapear el epigenoma humano en una amplia variedad de tejidos y células. El IHEC busca comprender cómo las modificaciones epigenéticas contribuyen a la diversidad humana, la enfermedad y el envejecimiento (KANAI y ARAI 2014).
- PsychENCODE. el proyecto tiene como objetivo mapear y caracterizar el epigenoma, transcriptoma y proteoma del cerebro humano en diferentes etapas del desarrollo y en diferentes regiones cerebrales. Este proyecto se centra en el estudio de las modificaciones epigenéticas en el cerebro humano. Su objetivo es comprender cómo las modificaciones epigenéticas contribuyen a la diversidad neuronal, al desarrollo cerebral y a las enfermedades neuropsiquiátricas (PsychENCODE Consortium, 2015).
- Proyecto DEEP (Consorcio para el Estudio del Impacto de los Factores Ambientales en el Epigenoma y el Proteoma). Este proyecto se enfoca en investigar los efectos de los factores ambientales en el epigenoma humano. Busca comprender cómo la exposición a diferentes factores ambientales puede afectar las modificaciones epigenéticas y contribuir al desarrollo de enfermedades.
- Iniciativa Cerebro del Proyecto Proteoma Humano (HPP-BI). Además del enfoque en el proteoma, existe un componente esencial dedicado al epigenoma. Esta parte del proyecto se centra en mapear y caracterizar las modificaciones epigenéticas en el cerebro humano, tales como la metilación del ADN y las modificaciones de histonas en diferentes regiones cerebrales y etapas de desarrollo. Estas modificaciones epigenéticas influyen en la expresión y función de las proteínas, proporcionando una capa adicional de regulación que afecta la actividad de los genes sin alterar la secuencia de ADN (Su et al., 2021). El estudio del epigenoma en el contexto del HPP-BI es crucial porque estas modificaciones pueden variar en función de factores como el desarrollo, la experiencia y el entorno. Estas variaciones

epigenéticas permiten entender mejor cómo se regula la expresión proteica en el cerebro y cómo esta regulación está relacionada con funciones cerebrales específicas y con enfermedades neurológicas y psiquiátricas. Así, el enfoque en el epigenoma complementa el mapeo del proteoma al revelar cómo los cambios epigenéticos influyen en las redes de proteínas y en los procesos biológicos en el cerebro humano (SU *et al.*, 2021).

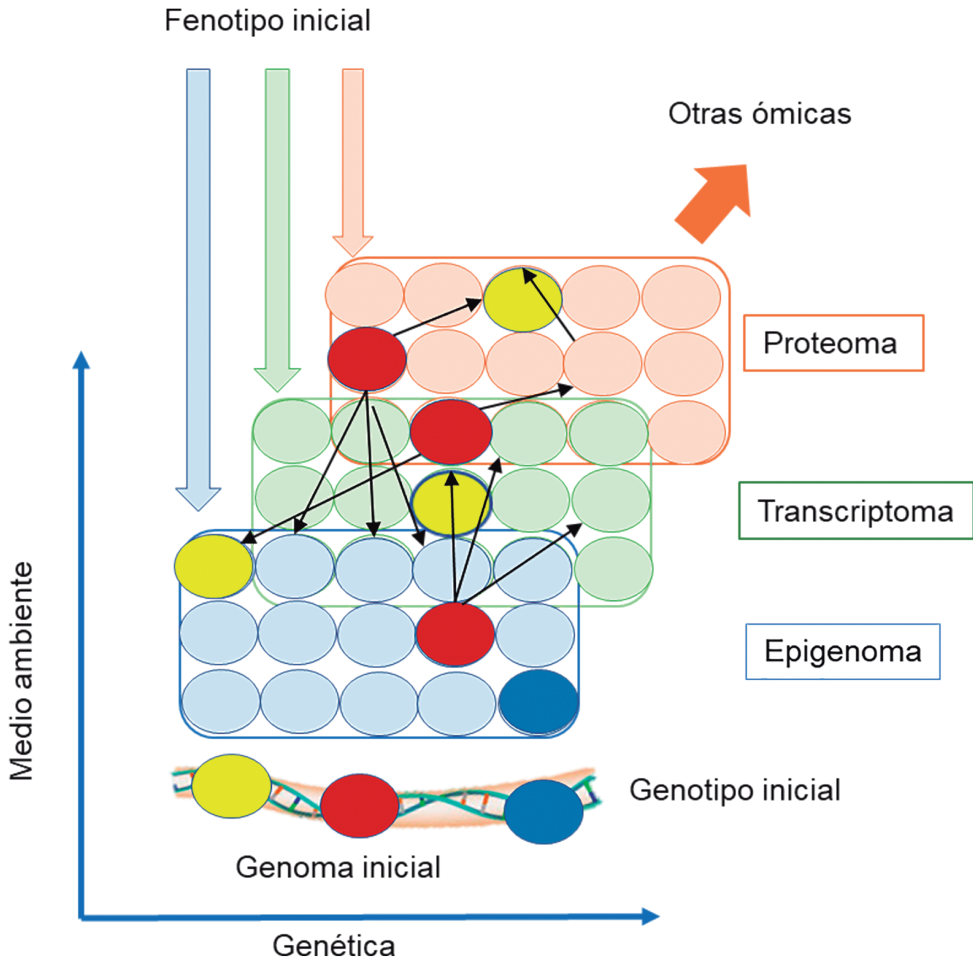
- Proyecto de red biológica pulmonar del Atlas Celular Humano (LBHN). el proyecto se centra en el estudio del atlas celular del pulmón humano. Utiliza enfoques de transcriptómica de una célula y proteómica para identificar y caracterizar los diferentes tipos de células presentes en el tejido pulmonar, lo que permite una mejor comprensión de las enfermedades pulmonares y el desarrollo de terapias más efectivas (ADAMS *et al.*, 2023).

4. ERA MULTIÓMICA

Los avances en las ciencias ómicas han generado una gran cantidad de datos detallados sobre los componentes moleculares de los organismos y sus interacciones. Sin embargo, estos datos suelen ser analizados de manera individual, lo que dificulta su interpretación completa. Para comprender realmente los fenotipos observados y las causas subyacentes a ellos, es necesario combinar los datos ómicos de diferentes disciplinas (PAMMI *et al.*, 2023; HASIN *et al.*, 2017). Por ejemplo, el microbioma tiene influencia en la expresión de genes y proteínas, que a su vez afectan al metaboloma. Todos estos procesos interactúan y se regulan mutuamente (**figura 19**).

La era multiómica consiste en un nuevo enfoque de investigación que integra múltiples tecnologías ómicas (genómica, proteómica, transcriptómica y metabolómica) con el fin de estudiar y comprender de manera integral los componentes moleculares de un organismo y sus interacciones. Esto permite obtener una visión más completa y detallada de los sistemas biológicos, así como comprender la relación entre los componentes moleculares y los fenotipos observados, para analizar y comprender la complejidad de los organismos vivos (HASIN *et al.*, 2017).

Figura 19. Integración de datos ómicos y enfoques de la enfermedad



Nota. Representa diferentes tipos de datos ómicos en capas, los círculos representan las moléculas de las que se recopilan los datos. La regulación genética y el medio ambiente influyen en las moléculas en cada capa, excepto en la capa del genoma. Las flechas finas indican posibles interacciones o correlaciones entre moléculas en diferentes capas. Las flechas gruesas en la parte superior señalan puntos de partida para consolidar múltiples datos ómicos y comprender los sistemas biológicos y la patogénesis de la enfermedad. Los enfoques pueden comenzar desde el genoma, el fenotipo u otras capas ómicas. Tomado y modificado de HASIN (2017).

Los análisis multiómicos utilizan caminos integrados para generar y analizar datos de múltiples niveles ómicos. Esto implica la recopilación de datos de

expresión génica, perfiles de proteínas y metabolitos en un mismo conjunto de muestras. Posteriormente, se aplican métodos bioinformáticos y estadísticos para integrar y analizar conjuntamente estos datos (JEHAN *et al.*, 2020).

Es así como la era multiómica ha abierto nuevas oportunidades para la integración de datos provenientes de diferentes disciplinas ómicas, lo que ha permitido comprender la interacción entre las diferentes ómicas, el descubrimiento de biomarcadores y nuevas dianas terapéuticas, así como el desarrollo de terapias más precisas y efectivas para las enfermedades, fortaleciendo así la medicina de precisión (PAMMI *et al.*, 2023).

4.1. INTEGRACIÓN DE DATOS MULTIÓMICOS Y APRENDIZAJE AUTOMÁTICO

El aprendizaje automático o *Machine Learning*, es un subcampo de la inteligencia artificial que utiliza algoritmos y modelos para que las máquinas puedan aprender (LI *et al.*, 2018). En el contexto de la medicina, el análisis multiómico y el aprendizaje automático son herramientas poderosas para descubrir patrones, identificar biomarcadores y comprender los mecanismos biológicos de las enfermedades (KARCZEWSKI y SNYDER, 2018). Esta combinación permite un enfoque integral y preciso en la medicina personalizada, facilitando la identificación de nuevas terapias.

Aunque se han logrado avances significativos en la clasificación de tipos de cáncer y predicción de respuestas a medicamentos, aún se requiere desarrollar nuevas metodologías de aprendizaje automático para la integración de datos en el descubrimiento de fármacos y tratamientos personalizados (KARCZEWSKI Y SNYDER, 2018). La era multiómica ha abierto nuevas oportunidades en la medicina de precisión al permitir la integración de datos de diferentes disciplinas, lo que ha llevado al descubrimiento de biomarcadores, nuevas dianas terapéuticas y terapias más precisas (NAITHANI *et al.*, 2021).

El análisis de datos multiómicos respalda el avance de la medicina de precisión al desentrañar la biología de sistemas y obtener una comprensión integral de los sistemas biológicos, gracias al uso de herramientas de minería de datos y aprendizaje automático (PICARD *et al.*, 2021; NAITHANI *et al.*, 2021).

El perfeccionamiento de nuevas plataformas de medición de alto rendimiento y análisis de los datos arrojados por los análisis de las diferentes fuentes ómicas permiten descubrir el funcionamiento de la biología de sistemas y obtener información sobre los sistemas biológicos dándole un enfoque integrador (PICARD *et al.*, 2021). Es así como surge el progreso de la medicina de precisión, respaldado por herramientas que permiten el análisis de minería de datos y enfoques de aprendizaje automático.

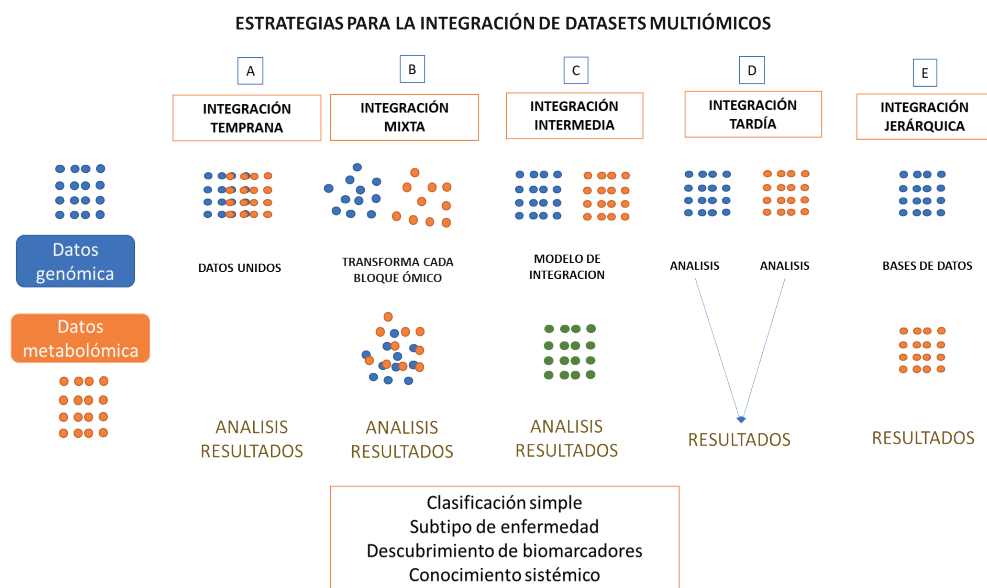
La mayor disponibilidad de tecnologías de alto rendimiento ha generado un número cada vez mayor de datos ómicos que buscan retratar muchas ciencias biológicas diferentes pero complementarias, incluidas la genómica, la epigenómica, la transcriptómica, la proteómica y la metabolómica. Se ha obtenido una nueva visión de estos datos mediante algoritmos de aprendizaje automático que han producido biomarcadores de diagnóstico y clasificación (NAITHANI *et al.*, 2021). Sin embargo, la mayoría de los biomarcadores obtenidos hasta la fecha solo incluyen una medición ómica a la vez y, por lo tanto, no aprovechan al máximo los recientes experimentos multiómicos que ahora capturan toda la complejidad de los sistemas biológicos (MAGHSOUDI *et al.*, 2022).

Para combinar el conocimiento complementario proporcionado por cada ciencia ómica, es necesario utilizar estrategias de integración de datos multiómicos. Estas estrategias ofrecen enfoques diferentes para combinar y aprovechar el conocimiento complementario proporcionado por cada capa ómica, permitiendo así una comprensión más completa y precisa de los procesos biológicos (PICARD *et al.*, 2021; MAGHSOUDI *et al.*, 2022). Se han distinguido dos tipos de integración: la integración horizontal que se centra en el estudio de las mismas capas ómicas en diferentes grupos de muestras y la integración vertical que permite analizar cómo las variables ómicas se comportan en distintos contextos biológicos o condiciones, lo que puede revelar patrones y asociaciones relevantes. Sin embargo, los métodos de integración de datos más recientes se clasifican en cinco estrategias de integración diferentes, temprana, mixta, intermedia, tardía y jerárquica (**figura 20**) (REEL *et al.*, 2021).

- Estrategia temprana: integra datos en etapas tempranas del análisis, antes de realizar análisis específicos en cada ciencia ómica por separado. Esta aproximación permite identificar rápidamente las relaciones entre los diferentes datos ómicos y encontrar patrones o asociaciones emergentes.
- Estrategia mixta: combina datos ómicos en ciertos puntos específicos del análisis, por ejemplo, al combinar datos genómicos y transcriptómicos para identificar genes que muestran una correlación entre niveles de expresión y variantes genéticas.
- Estrategia intermedia: se enfoca en la integración de datos en etapas intermedias del análisis ómico. Esto implica realizar análisis independientes para cada capa ómica y luego combinar los resultados para obtener una visión más completa de los procesos biológicos.
- Estrategia tardía: realiza análisis por separado para cada capa ómica y luego integrar los resultados finales. Esta aproximación permite un mayor control sobre el análisis individual de cada ciencia ómica antes de combinar los datos.

- Estrategia jerárquica: implica una integración secuencial de datos ómicos, donde se utilizan ciencias ómicas previamente analizadas como base para la comprensión de ciencias posteriores. Esto permite aprovechar la información previamente obtenida para un análisis más detallado y exhaustivo.

Figura 20. Representación esquemática de las principales estrategias para la integración de datasets multiómicos



Nota. Se muestra en (A) la integración temprana concatena todos los conjuntos de datos ómicos en una sola matriz en la que se puede aplicar el modelo de aprendizaje automático. (B) La integración mixta primero transforma o mapea independientemente cada bloque ómico en una nueva representación antes de combinarlos para el análisis posterior. (C) La integración intermedia transforma simultáneamente los conjuntos de datos originales en representaciones comunes y específicas de ómicas. (D) La integración tardía analiza cada ómica por separado y combina sus predicciones finales. (E) La integración jerárquica basa la integración de conjuntos de datos en relaciones regulatorias previas entre capas ómicas. Tomado y modificado de Reel (2021).

La base del análisis de datos multiómicos es el aprendizaje automático, sobre la base del cual se han desarrollado muchas herramientas (LÓPEZ *et al.*, 2019). Los algoritmos de aprendizaje automático están entrenados para modelar patrones complejos que no pueden ser capturados con precisión por modelos matemáticos tradicionales en datos de alta dimensión, como en el caso del cáncer (NAKAGAWA y FUJITA, 2018).

En estudios de cáncer, los datos multiómicos incluyen información de diversas capas moleculares, como el genoma (variantes genéticas), el epigenoma (modificaciones químicas del ADN), el transcriptoma (expresión de ARN), el proteoma (niveles de proteínas) y el metaboloma (compuestos metabólicos). Cada uno de estos datos proporciona información única, pero cuando se combinan permiten una visión integral de los procesos biológicos complejos en las células cancerígenas (NAKAGAWA y FUJITA, 2018).

Sin embargo, debido a la alta dimensión y la complejidad de los datos multiómicos, los métodos tradicionales de análisis matemático son insuficientes para identificar patrones significativos. Aquí es donde los algoritmos de aprendizaje automático son particularmente útiles. Por ejemplo, los algoritmos de aprendizaje supervisado, como las máquinas de vectores de soporte (SVM) y los árboles de decisión, pueden ser entrenados con datos etiquetados para clasificar tipos de cáncer según patrones moleculares específicos. Estos modelos aprenden a identificar combinaciones de características moleculares que diferencian a los tumores según su origen, tipo o grado de agresividad (NAKAGAWA y FUJITA, 2018).

El aprendizaje automático en el análisis multiómico se refiere a la aplicación de técnicas y algoritmos de inteligencia artificial para extraer información y conocimiento de conjuntos de datos complejos y heterogéneos que abarcan múltiples capas ómicas (PICARD *et al.*, 2021). La combinación de datos multiómicos y aprendizaje automático permite un enfoque más integral y preciso para comprender los mecanismos biológicos subyacentes a las enfermedades, facilitando así la identificación de nuevas terapias y enfoques personalizados en la medicina (REEL *et al.*, 2021).

Un ejemplo de esto es la aplicación de cinco enfoques de aprendizaje automático utilizando datos de la Enciclopedia de Líneas de Cáncer. En este estudio, se informa la precisión en la clasificación del tipo de cáncer y se evalúa el error absoluto medio en la predicción de la respuesta a los medicamentos, al mismo tiempo que se considera la eficiencia del tiempo de ejecución (NAKAGAWA Y FUJITA, 2018).

Sin embargo, es necesario desarrollar nuevas metodologías de aprendizaje automático para la integración de datos, que serán esenciales para el descubrimiento de fármacos, el diseño de ensayos clínicos y los tratamientos personalizados. *Multi-omics (X-omics)* es un neologismo que brinda una gran oportunidad de mejora para la medicina de precisión (REEL *et al.*, 2021; ZHANG *et al.*, 2021).

La información multiómica es esencialmente valiosa para el desarrollo de fármacos si indica nuevos objetivos farmacológicos o firmas de enfermedades que se asocian con el resultado de la enfermedad y/o la respuesta al tratamiento aprendizaje automático.

4.2. MINERÍA DE DATOS

La minería de datos se aplica en diversos campos y sectores para descubrir patrones, tendencias y conocimientos útiles a partir de grandes conjuntos de datos. La minería de datos en ciencias ómicas describe el proceso de descubrir modelos, relaciones y conocimientos útiles, con base en múltiples conjuntos de datos ómicos (WU *et al.*, 2021; Reel *et al.*, 2021). Las ciencias ómicas, como la genómica, transcriptómica, proteómica y metabolómica, generan una gran cantidad de datos complejos y heterogéneos que contienen información detallada sobre los componentes biológicos y procesos en un organismo (NAKAGAWA y FUJITA, 2018).

La minería de datos utiliza técnicas y algoritmos de análisis de datos para explorar, visualizar y extraer información relevante de estos conjuntos de datos. Estas técnicas pueden incluir métodos estadísticos, aprendizaje automático, análisis de redes y otras herramientas de análisis de datos. El objetivo es descubrir patrones significativos, identificar biomarcadores y obtener conocimientos que puedan contribuir al avance de la investigación en biología y medicina (ZIA *et al.*, 2022).

En las ciencias ómicas, la minería de datos juega un papel fundamental en el descubrimiento de nuevas asociaciones entre genes, proteínas, metabolitos y fenotipos, lo que puede conducir a la identificación de nuevos biomarcadores, terapias más efectivas y una comprensión más profunda de los procesos biológicos. Además, esta disciplina permite el análisis de grandes volúmenes de datos de manera eficiente, lo que ayuda a acelerar la investigación y el desarrollo de la medicina personalizada (ZIA *et al.*, 2022).

5. BIOINFORMÁTICA Y ANÁLISIS DE DATOS MULTIÓMICOS

La bioinformática, una disciplina que combina biología e informática, ha revolucionado el análisis de datos biológicos mediante el uso de bases de datos especializadas. Estas bases de datos son fundamentales para almacenar y organizar grandes cantidades de información biológica (BREITWIESER *et al.*, 2019). La bioinformática utiliza diversas técnicas, como el alineamiento de secuencias, el ensamblaje de genomas y el análisis de expresión génica, para extraer información de los datos generados (SUBRAMANIAN *et al.*, 2020). Además, se emplean herramientas bioinformáticas para el análisis de interacciones proteína-proteína, metilación del ADN, microarreglos y secuenciación masiva, entre otros (DINIZ y CANDURI, 2017; ZHANG *et al.*, 2021)

La combinación de datos ómicos y herramientas bioinformáticas permite comprender mejor los mecanismos subyacentes y las vías biológicas. Las bases de datos de bioinformática albergan una amplia gama de información biológica,

como secuencias de ADN, ARN y proteínas, estructuras tridimensionales y patrones de expresión génica. Los investigadores pueden acceder a estas bases de datos a través de interfaces de búsqueda y consulta, y muchas de ellas también ofrecen herramientas de análisis y visualización integradas (SUBRAMANIAN *et al.*, 2020).

Cabe destacar que el campo de las ciencias ómicas y las herramientas bioinformáticas están en constante evolución, por lo que es importante mantenerse actualizado sobre las últimas herramientas disponibles.

5.1. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN GENÓMICA

Todas las bases de datos y herramientas bioinformáticas desempeñan un papel fundamental en el análisis y la interpretación de los datos generados a partir de secuencias de ADN. Estas permiten realizar una variedad de tareas, como el ensamblaje y anotación de genomas, la identificación de variantes genéticas, el análisis de expresión génica y la predicción de funciones de genes. Cada herramienta tiene su propio propósito y se utiliza para una variedad de análisis, desde la anotación y el alineamiento de secuencias hasta el análisis funcional y la visualización de datos genómicos (DINIZ *et al.*, 2017). Algunas de las **bases de datos** y programas bioinformáticos más usadas son:

- GenBank: es una base de datos de secuencias genéticas. Contiene información sobre secuencias de ADN y ARN de una variedad de organismos de todas las categorías, incluyendo bacterias, virus, plantas, animales y otros grupos taxonómicos. Es una de las bases de datos más grandes de secuencias biológicas en el mundo, permite a los investigadores acceder a información genética, como secuencias de genes y genomas para análisis y comparación, es mantenida por el NIH (Langmead y Salzberg, 2012; Altschul *et al.*, 1990).
- Ensembl: es una base de datos que proporciona anotaciones genómicas para una amplia variedad de especies, incluyendo humanos, ratones, moscas de la fruta, levaduras y muchas otras. Ensembl integra datos genómicos de diversas fuentes y proporciona acceso a secuencias, anotaciones funcionales, variaciones genéticas y otros datos relacionados (SEEMANN, 2014; CAMPBELL *et al.*, 2014).
- Gene Ontology (GO): es una base de datos y conjunto de herramientas para el análisis funcional de genes y proteínas. Se utiliza para asignar funciones y categorías biológicas a genes basados en anotaciones (KANEHISA y GOTO, 2000).

- dbSNP: es una base de datos que almacena información sobre variaciones genéticas, como polimorfismos de un solo nucleótido (SNP) y mutaciones genéticas asociadas a enfermedades. dbSNP proporciona identificadores únicos para las variantes genéticas y es ampliamente utilizada en estudios de asociación genética y medicina personalizada (MCKENNA *et al.*, 2010; DANECEK *et al.*, 2021).

En cuanto a Herramientas bioinformáticas disponibles, están:

- RefSeq: mantenida por el Centro Nacional de Información Biotecnológica (NCBI), es una base de datos que proporciona secuencias de referencia y anotaciones para genes, transcritos y proteínas de diferentes organismos. Se actualiza regularmente y ofrece una amplia cobertura de genomas secuenciados (PRUITT *et al.*, 2007; BRAY *et al.*, 2016).
- UCSC Genome Browser: es una plataforma interactiva en línea que permite visualizar y explorar genomas completos y sus anotaciones. Proporciona acceso a una amplia gama de datos genómicos, incluyendo secuencias, genes, variantes genéticas, regiones reguladoras y datos epigenéticos (ROBINSON *et al.*, 2011; KENT *et al.*, 2002).
- BLAST (Basic Local Alignment Search Tool): se utiliza para buscar similitudes entre secuencias de ADN o proteínas en una base de datos para identificar homologías y anotar genes (LANGMEAD y SALZBERG, 2012; ALTSCHUL *et al.*, 1990).
- Picard: es un conjunto de herramientas de bioinformática ampliamente utilizadas en el análisis de datos de secuencias de ADN. Picard está diseñado principalmente para trabajar con archivos BAM (Binary Alignment/Map) y SAM (Sequence Alignment/Map), que son formatos estándar para almacenar datos de secuencias alineadas (BROAD INSTITUTE, s. f.).

En la **tabla 8** se mencionan algunas de las principales bases de datos y herramientas bioinformáticas utilizadas en genómica.

Tabla 8. Bases de datos y herramientas bioinformáticas usadas en genómica

Bases de datos	Herramientas de análisis	Usos	Referencias
NCBI, Ensembl	Prokka, MAKER, Augustus, GeneMark, Blast2GO	Anotación de genomas	SEEMANN, 2014; CAMPBELL <i>et al.</i> , 2014
dbSNP, 1000 Genomes, ClinVar	GATK (Genome Analysis Toolkit), SAMtools, FreeBayes, VarScan, Bcftools, BRAKER	Identificación de variantes genéticas	MCKENNA <i>et al.</i> , 2010; DANECEK <i>et al.</i> , 2021
KEGG, Gene Ontology, InterPro	InterProScan, Pfam, Gene Ontology (GO) enrichment analysis, KEGG pathway analysis	Predicción de funciones de genes	JONES <i>et al.</i> , 2014; KANEHISA y GOTO, 2000
SILVA, TreeBase, NCBI	MEGA, PhyML, RAxML, BEAST	Análisis de filogenia	KUMAR <i>et al.</i> , 2018; GUINDON <i>et al.</i> , 2010
UCSC Genome Browser, Ensembl, NCBI	IGV, UCSC Genome Browser, JBrowse, Artemis, Circos	Visualización de datos genómicos	ROBINSON <i>et al.</i> , 2011; Kent <i>et al.</i> , 2002
GenBank, Ensembl	Bowtie, BWA, BLAST, Clustal Omega, MAFFT	Alineamientos	LANGMEAD y SALZBERG, 2012; Altschul <i>et al.</i> , 1990

5.2. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN TRANSCRIPTÓMICA

En la transcriptómica, las bases de datos y herramientas bioinformáticas son esenciales para el análisis de los datos generados a partir de experimentos de expresión génica y RNA-Seq. Estas herramientas permiten el procesamiento, análisis y visualización de los datos de transcripción, así como la identificación de genes diferencialmente expresados y el descubrimiento de nuevos ARN no codificantes. A continuación, se describen algunas de las principales herramientas bioinformáticas utilizadas en transcriptómica. En lo que refiere a **bases datos**, se tiene:

- Gene Expression Omnibus (GEO): es una de las principales bases de datos de expresión génica, mantenida por el NCBI. Almacena datos de microarreglos y secuenciación de ARN de diversos organismos y ofrece herramientas para buscar y analizar datos de expresión génica (CLOUGH y BARRETT., 2016).

- ArrayExpress: administrada por el European Bioinformatics Institute (EBI), esta base de datos también contiene datos de microarreglos y secuenciación de ARN. Permite a los investigadores acceder a datos de transcriptómica para su análisis y reutilización (BRAZMA *et al.*, 2003).
- ENCODE (Encyclopedia of DNA Elements): aunque su enfoque es la anotación funcional del genoma humano, contiene datos transcriptómicos sobre la expresión y regulación de genes. Es útil para estudiar la regulación génica y su impacto en la expresión (The ENCODE Project Consortium, 2012).
- Human Protein Atlas: aunque su nombre sugiere solo proteínas, también ofrece datos sobre la expresión de ARN en varios tejidos y tipos de células humanos, proporcionando una visión de la transcriptómica a nivel de tejidos específicos (UHLEN *et al.*, 2015).
- TCGA (The Cancer Genome Atlas): contiene datos transcriptómicos de muestras de cáncer humano. Esta base de datos permite estudios detallados de la expresión génica en diferentes tipos de cáncer (WEINSTEIN *et al.*, 2013; SILVA *et al.*, 2016).
- Expression Atlas: ofrece datos de expresión génica en una amplia gama de organismos y condiciones experimentales, lo que facilita los estudios de comparación de expresión entre diferentes tejidos, condiciones y estados de desarrollo (PAPATHEODOROU *et al.*, 2020).
- RNACentral: esta base de datos contiene información sobre ARN no codificante de varias especies, proporcionando una visión específica de elementos de ARN que no traducen proteínas, pero juegan roles importantes en la regulación génica (RNACentral Consortium, 2021).
- GTEx (Genotype-Tissue Expression Project): ofrece datos sobre la expresión de ARN en tejidos humanos sanos, proporciona una referencia importante sobre la expresión génica en diferentes tejidos y poblaciones (CARITHERS *et al.*, 2015).
- Single Cell Expression Atlas: enfocada en datos de expresión de células individuales, permite la exploración de la variabilidad de expresión en diferentes tipos celulares y estados celulares, lo cual es crucial para estudios de transcriptómica a nivel de célula única (PAPATHEODOROU *et al.*, 2020).

En cuanto a herramientas bioinformáticas, están:

- DESeq2: es una herramienta de análisis de expresión diferencial ampliamente utilizada en transcriptómica. Se utiliza para comparar los niveles de expresión génica entre diferentes muestras y determinar

qué genes están diferencialmente expresados. DESeq2 utiliza modelos estadísticos para identificar los genes que muestran cambios significativos en su expresión y ayuda a comprender los patrones de expresión diferencial (LOVE *et al.*, 2014).

- edgeR: es otra herramienta popular para el análisis de expresión diferencial en transcriptómica. Al igual que DESeq2, edgeR utiliza modelos estadísticos para detectar genes diferencialmente expresados entre diferentes condiciones o grupos de muestras. edgeR tiene en cuenta la variabilidad biológica y técnica en los datos de expresión y proporciona estadísticas y puntuaciones de significancia para identificar genes con cambios significativos (LOVE *et al.*, 2014).
- Gene Ontology (GO): es una herramienta que se utiliza para asignar funciones biológicas a los genes y comprender las características funcionales de los conjuntos de genes identificados en los estudios de transcriptómica. Utiliza una estructura de ontología para categorizar y agrupar genes en términos de sus funciones biológicas, procesos y componentes celulares (KANEHISA y GOTO, 2000).
- KEGG Pathway Analysis. KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Analysis: se utiliza para identificar las vías metabólicas y las redes de interacción de genes relacionadas con los conjuntos de genes diferencialmente expresados. Proporciona una visión más amplia de cómo los genes interactúan y se regulan entre sí dentro de un contexto funcional y de vías metabólicas específicas (KANEHISA y GOTO, 2000).

Es importante destacar que también se emplean herramientas de visualización como Heatmap y Volcano Plot que son utilizadas en el análisis de expresión diferencial en transcriptómica. Los *Heatmaps* se utilizan para visualizar patrones de expresión de genes en diferentes muestras o condiciones, donde se representa la intensidad de expresión mediante colores. Los *Volcano Plots* representan gráficamente los valores de p-value en función del *fold-change* para resaltar los genes diferencialmente expresados.

En la **tabla 9** se mencionan algunas de las principales herramientas bioinformáticas utilizadas específicamente en transcriptómica.

Tabla 9. Bases de datos y herramientas bioinformáticas usadas en transcriptómica

Bases de datos	Herramientas de análisis	Usos	Referencias
GEO (Gene Expression Omnibus)	GEO2R, DESeq2, edgeR, limma	Análisis de expresión génica diferencial	CLOUGH y BARRETT, 2016; Love <i>et al.</i> , 2014
ArrayExpress	Bioconductor, limma, pheatmap	Análisis de expresión génica diferencial	BRAZMA <i>et al.</i> , 2003
TCGA (The Cancer Genome Atlas)	cBioPortal, TCGAblinks	Estudios de expresión génica en cáncer	WEINSTEIN <i>et al.</i> , 2013; Silva <i>et al.</i> , 2016
ENCODE (Encyclopedia of DNA Elements)	DESeq2, STAR, HTSeq	Análisis de expresión génica y regulación	The ENCODE Project Consortium, 2012
RefSeq	Kallisto, Salmon, RSEM	Cuantificación de transcritos	PRUITT <i>et al.</i> , 2007; BRAY <i>et al.</i> , 2016
GTEX (Genotype-Tissue Expression)	DESeq2, edgeR, STAR, HISAT2	Análisis de expresión en distintos tejidos	CARITHERS <i>et al.</i> , 2015
Human Protein Atlas (para datos transcriptómicos)	DESeq2, edgeR, limma	Análisis de transcriptómica en tejido	UHLEN <i>et al.</i> , 2015
SRA (Sequence Read Archive)	FastQC, Trimmomatic, STAR, HISAT2, Cufflinks	Procesamiento y análisis de datos RNA-Seq	LEINONEN <i>et al.</i> , 2011
miRBase	miRDeep2, miranda, TargetScan	Análisis de microARNs	GRIFFITHS-JONES <i>et al.</i> , 2006

5.3. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN PROTEÓMICA

En la proteómica, las bases de datos y herramientas bioinformáticas permiten la identificación, caracterización y análisis de las proteínas presentes en una muestra. Estas herramientas permiten realizar análisis de datos, identificar proteínas, predecir funciones y realizar análisis estadísticos. A continuación, se presentan algunas de las principales herramientas bioinformáticas utilizadas en proteómica (CHEN *et al.*, 2017). En cuanto a bases de datos están:

- UniProt: base de datos de secuencias y funciones de proteínas (UniProt Consortium, 2019).

- PDB (Protein Data Bank): base de datos de estructuras tridimensionales de proteínas y ácidos nucleicos (MICHALSKA y JOACHIMIAK, 2021).
- PRIDE: base de datos de espectrometría de masas para identificaciones proteicas (PÉREZ-RIVEROL *et al.*, 2019).
- PeptideAtlas: mapa de péptidos basado en datos de espectrometría de masas (DESIERE *et al.*, 2006).
- ProteomicsDB: base de datos de expresión y cuantificación de proteínas (UHLEN *et al.*, 2015).
- Human Protein Atlas: base de datos de expresión de proteínas en tejidos humanos (UHLEN *et al.*, 2015).
- PhosphoSitePlus: base de datos de modificaciones postraduccionales en proteínas (LEE *et al.*, 2006).
- BioGRID: Base de datos de interacciones proteína-proteína (STARK *et al.*, 2006).

En lo que atañe a herramientas bioinformáticas, se pueden mencionar:

- Mascot (Modular Approach to Software Construction Operation and Test): es una herramienta de búsqueda de identificación de proteínas ampliamente utilizada en proteómica. Se utiliza para comparar los espectros de masas generados por espectrometría de masas con bases de datos de secuencias proteicas para identificar las proteínas presentes en una muestra. Mascot utiliza algoritmos de puntuación para determinar la probabilidad de que una coincidencia entre un espectro y una secuencia proteica sea genuina (PERKINS *et al.*, 1999).
- MaxQuant: es una herramienta de análisis de proteómica cuantitativa que permite el análisis de datos de espectrometría de masas basados en etiquetado de isótopos estables o etiquetado químico. MaxQuant realiza la identificación y cuantificación de proteínas, así como el análisis de expresión diferencial, análisis de enriquecimiento de funciones y visualización de datos (PEREZ-RIVEROL *et al.*, 2019).
- Scaffold: es una herramienta de análisis y visualización de datos proteómicos. Se utiliza para organizar, visualizar y analizar los resultados de los experimentos de proteómica. Scaffold permite la identificación y cuantificación de proteínas, la comparación de muestras y la realización de análisis estadísticos para identificar proteínas diferencialmente expresadas (SEARLE, 2010).

- **STRING**: es utilizada para analizar las interacciones proteína-proteína y las redes de interacción en proteómica, integra información de diferentes bases de datos para predecir las interacciones entre proteínas y generar redes de interacción. Además, proporciona información sobre las funciones y vías biológicas asociadas con las proteínas de interés (UHLEN *et al.*, 2015).
- **Perseus**: es una herramienta de análisis estadístico utilizada en proteómica. Permite el análisis de expresión diferencial, análisis de agrupamiento, análisis de enriquecimiento de vías y otros análisis estadísticos para identificar patrones y diferencias (WANG *et al.*, 2015).

En la **tabla 10** se mencionan algunas de las principales bases de datos y herramientas bioinformáticas utilizadas específicamente en esta disciplina.

Tabla 10. Bases de datos y herramientas bioinformáticas usadas en proteómica

Bases de datos	Herramientas de análisis	Usos	Referencias
UniProt	BLAST, InterProScan	Anotación de proteínas	UniProt Consortium, 2019
PRIDE (Proteomics Identification Database)	MaxQuant, Proteome Discoverer, OpenMS	Análisis de espectrometría de masas, identificación de proteínas	PEREZ-RIVEROL <i>et al.</i> , 2019
PeptideAtlas	Skyline, OpenMS	Cuantificación de proteínas, validación de espectros	DESIERE <i>et al.</i> , 2006
Human Protein Atlas	Cytoscape, STRING	Análisis de expresión proteica en tejidos y células	UHLEN <i>et al.</i> , 2015
SWISS-2DPAGE	Melanie, Delta2D	Análisis de electroforesis en gel bidimensional (2D)	HOOGGLAND <i>et al.</i> , 2004
dbPTM	ModPred, PhosphoSitePlus	Análisis de modificaciones postraduccionales	LEE <i>et al.</i> , 2006
PROSITE	ScanProsite	Identificación de dominios y sitios funcionales en proteínas	SIGRIST <i>et al.</i> , 2013
PaxDb	Perseus, Cytoscape	Análisis de abundancia proteica en diferentes especies	WANG <i>et al.</i> , 2015

5.4. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN METABOLÓMICA

Las herramientas bioinformáticas desempeñan un papel clave en el análisis e interpretación de los datos generados a partir de perfiles de metabolitos. Estas herramientas permiten el procesamiento, la identificación y la anotación de los metabolitos, así como el análisis estadístico y la visualización de los datos. A continuación, se mencionan algunas de las principales bases de datos y herramientas bioinformáticas utilizadas en metabolómica. Las bases de datos son.

- LIPID MAPS (Lipid Metabolites and Pathways Strategy): es un recurso integral de acceso abierto que proporciona herramientas y bases de datos para la investigación de lípidos. Su objetivo principal es facilitar el estudio de la estructura, función y dinámica de los lípidos en diversos contextos biológicos (FAHY *et al.*, 2005).
- El Metabolomics Workbench es un repositorio internacional que ofrece acceso a datos y metadatos de estudios metabolómicos. Desarrollado como parte del programa Common Fund del Instituto Nacional de Salud (NIH) de EE. UU., su objetivo es facilitar la recopilación, el análisis y la difusión de datos metabolómicos para la comunidad científica (SUD *et al.*, 2016).
- MetaCyc es una base de datos curada que recopila información sobre rutas metabólicas y enzimas de todos los dominios de la vida. Su objetivo principal es catalogar de manera exhaustiva las rutas metabólicas experimentales conocidas, abarcando tanto el metabolismo primario como el secundario (CASPI *et al.*, 2018).
- Reactome es una base de datos de vías biológicas que proporciona información detallada sobre rutas de señalización y metabolismo en humanos y otros organismos. Se utiliza ampliamente en investigación bioinformática y biología molecular para explorar las interacciones moleculares y los procesos bioquímicos en células y organismos. Reactome permite a los investigadores visualizar y analizar rutas complejas que incluyen metabolismo, señalización celular, transducción de señales, regulación de la expresión génica, y procesos de desarrollo celular (GILLESPIE *et al.*, 2022).

Las herramientas bioinformáticas disponibles son:

- MetScape y KEGG se utilizan para analizar los datos de metabolómica y comprender las vías metabólicas y las redes de interacción implicadas.

Estas herramientas utilizan bases de datos de vías metabólicas y redes de interacción para asignar los metabolitos detectados a vías específicas y proporcionar una visión más amplia de los cambios metabólicos asociados con las condiciones experimentales (XIA *et al.*, 2009).

- MetaboAnalyst es una herramienta ampliamente utilizada en metabolómica y permite realizar análisis de expresión diferencial, análisis de agrupamiento, análisis de enriquecimiento de vías, análisis de correlación y visualización de datos. MetaboAnalyst integra varias técnicas estadísticas y algoritmos para identificar patrones y diferencias significativas en los perfiles metabólicos y proporciona una interfaz intuitiva para la exploración de los resultados (CHONG *et al.*, 2019; XIA *et al.*, 2009).
- MetFrag y GNPS (Global Natural Products Social Molecular Networking), se utilizan para asignar identidades químicas a los metabolitos detectados en los datos de metabolómica. Estas herramientas comparan los espectros de masas experimentales con bases de datos de referencia y utilizan algoritmos de anotación para identificar los metabolitos y proporcionar información sobre su estructura química y posibles vías metabólicas asociadas (TSUGAWA, 2018).
- MS-DIAL. MS-DIAL: es una herramienta especializada para el análisis de datos de espectrometría de masas en metabolómica no dirigida. Permite realizar el procesamiento de datos, identificar y cuantificar metabolitos y realizar análisis estadísticos. MS-DIAL utiliza algoritmos de agrupamiento y filtrado para identificar características de masa y retención únicas y proporciona una interfaz fácil de usar para el análisis de datos de metabolómica (TSUGAWA *et al.*, 2020).

Las herramientas de visualización, como MetaboAnalyst, Cytoscape y Heatmaps, se utilizan para visualizar y explorar los datos de metabolómica. Estas herramientas permiten visualizar perfiles metabólicos, representar las relaciones entre metabolitos y realizar análisis de correlación y agrupamiento. La visualización de los datos de metabolómica es fundamental para identificar patrones y tendencias en los perfiles metabólicos y comprender las interacciones entre los metabolitos (CHEN *et al.*, 2017).

En la **tabla 11** se mencionan algunas de las principales bases de datos y herramientas bioinformáticas utilizadas específicamente en esta disciplina.

Tabla 11. Bases de datos usadas en metabolómica

Bases de datos	Herramientas de análisis	Usos	Referencias
Metabolomics Workbench	MetaboAnalyst, XCMS	Análisis de perfiles metabolómicos	SUD <i>et al.</i> , 2016
HMDB (Human Metabolome Database)	MAVEN, MetaboAnalyst	Identificación y cuantificación de metabolitos	WISHART <i>et al.</i> , 2018
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Pathway Analysis, Cytoscape	Análisis de rutas metabólicas	KANEHISA y GOTO, 2000
LipidMaps	LipidXplorer, LipidSearch	Identificación y cuantificación de lípidos	FAHY <i>et al.</i> , 2007
MetaCyc	Pathway Tools, Cytoscape	Análisis de rutas metabólicas específicas de organismos	CASPI <i>et al.</i> , 2018
MassBank	mzMine, OpenMS	Identificación de compuestos por espectrometría de masas	HORAI <i>et al.</i> , 2010
ChEBI (Chemical Entities of Biological Interest)	OpenMS, MetFrag	Identificación de compuestos y anotación de metabolitos	HASTINGS <i>et al.</i> , 2013
GNPS (Global Natural Products Social Molecular Networking)	QIIME, Cytoscape	Análisis de productos naturales y metabolitos	WANG <i>et al.</i> , 2016

5.5. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICA USADAS EN OTRAS ÓMICAS

La bioinformática y las ciencias ómicas emergentes como la lipidómica, glicómica, metagenómica, epigenómica, entre otras, son campos interdisciplinarios en constante evolución que están revolucionando la investigación biológica y médica. Desde la identificación de nuevos genes y proteínas hasta el estudio de las modificaciones epigenéticas y la caracterización de las vías metabólicas, estas disciplinas están abriendo nuevas puertas hacia un mayor entendimiento de la biología y su aplicación en la medicina y la biotecnología (DINIZ y CANDURI,

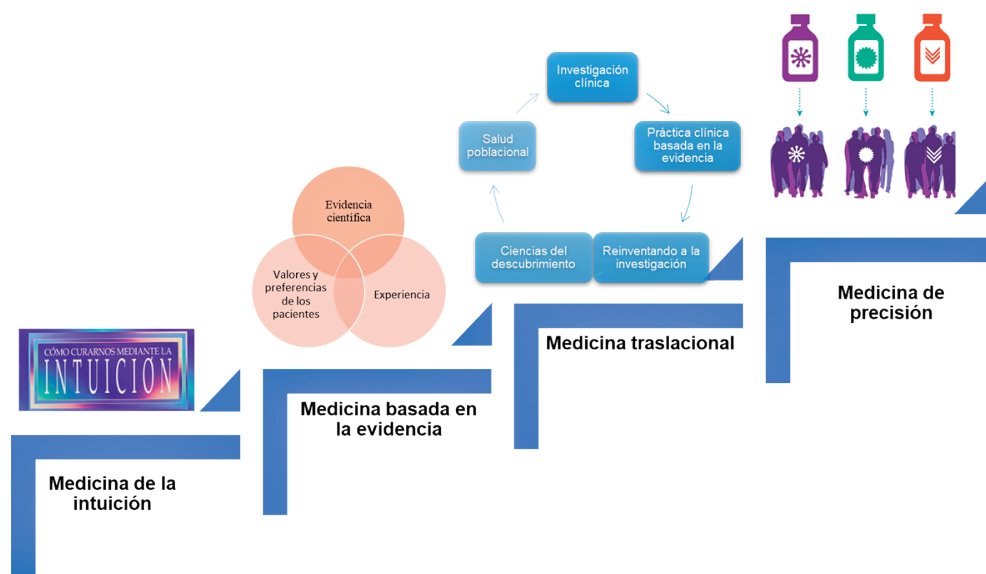
2017). En la **tabla 12** se mencionan algunas de las principales herramientas y bases de datos bioinformáticas utilizadas específicamente en estas disciplinas.

Tabla 12. Bases de datos y herramientas bioinformáticas usadas en otras ómicas

Área	Bases de datos	Herramientas de análisis	Usos	Bibliografía
Lipidómica	LipidMaps, LipidBlast	LipidXplorer, LipiDex, MS-DIAL	Identificación y cuantificación de lípidos	FAHY <i>et al.</i> , 2007; KIND <i>et al.</i> , 2013
Glicómica	GlyTouCan, UniCarbKB, GlyGen	GlycoWorkbench, GlycoMod, GlySpace	Identificación y anotación de glicanos	AOKI <i>et al.</i> , 2016; CAMPBELL <i>et al.</i> , 2014; York <i>et al.</i> , 2020
Metagenómica	MG-RAST, IMG/M (Integrated Microbial Genomes)	QIIME, MetaPhlAn, Kraken2	Análisis de comunidades microbianas	MEYER <i>et al.</i> , 2008; SCHULTZ <i>et al.</i> , 2015
Epigenómica	Roadmap Epigenomics, ENCODE	Bismark, MethGo, MeDIP-seq	Mapeo de modificaciones epigenéticas	Roadmap Epigenomics Consortium, 2015
Farmacogenómica	PharmGKB, DrugBank	CPIC (Clinical Pharmacogenetics Implementation Consortium) guidelines, PharmCAT	Predicción de respuestas a fármacos según el perfil genético	WHIRL-CARRILLO <i>et al.</i> , 2012; WISHART <i>et al.</i> , 2018

6. MEDICINA DE PRECISIÓN Y ENFOQUE ÓMICO

En el pasado, la medicina se practicaba mediante la evaluación de signos y síntomas y se basaba únicamente en el conocimiento del médico individual y era medicina de la intuición. Actualmente, la medicina se basa en la investigación científica, incluidos los ensayos clínicos, y es una medicina basada en la evidencia. En las últimas décadas, la medicina basada en la evidencia dio surgimiento a la medicina traslacional, que trata de combinar la investigación fundamental con la clínica experimental, además comprende otros niveles de necesidades de base social más que individual (**figura 21**) (PAMMI *et al.*, 2023).

Figura 21. Evolución de la medicina

Nota. Se muestran los cambios que ha sufrido la medicina gracias al avance de las ciencias ómicas.

Durante los últimos años, las ciencias ómicas han revolucionado la medicina traslacional con su aplicación integrada y con las tecnologías clínicas que amplían el conocimiento fisiopatológico de las enfermedades humanas. Es así como en un futuro la medicina se practicará mediante algoritmos basados en el fenotipo, el genoma, el epigenoma u otros datos ómicos del paciente que individualizarán el tratamiento, lo que constituye la medicina de precisión con un enfoque *ómico* que implica un conjunto de valoración integral de moléculas (NAITHANI *et al.*, 2021).

La medicina de precisión ofrece una forma de cambiar los enfoques clínicos que brindan opciones precisas de prevención, diagnóstico y tratamiento. Con el desarrollo de las tecnologías de NGS la medicina de precisión se está volviendo atractiva y práctica, lo que representa una gran promesa para el futuro de la medicina (NAITHAN *et al.*, 2021). La combinación de estudios de metabolómica, genómica, transcriptómica, proteómica y epigenómica conducirá a una mejor comprensión de la fisiopatología de la enfermedad como por ejemplo en el cáncer (NAKAGAWA y FUJITA, 2018). Por lo tanto, este punto de vista se considera un paso clave hacia la medicina de precisión.

Las ciencias ómicas, describe tecnologías experimentales de alto rendimiento que proporcionan las herramientas para monitorear ampliamente

el desarrollo de la enfermedad a nivel molecular que se enfoca en grandes datos (PAMMI *et al.*, 2023).

La medicina genómica está en el corazón de la medicina de precisión. Genome Wide Association Studies (GWAS) y Whole Exome Sequencing (WES) son nuevas herramientas en el campo de la genómica para comprender variantes asociadas de enfermedades multifactoriales comunes (NAITHAN *et al.*, 2021). Los avances genómicos tienen el potencial de revolucionar la medicina y la salud pública. El valor de la multiómica para estudiar diversos aspectos en las enfermedades obliga a integrar la multiómica utilizando herramientas de aprendizaje automático (PAMMI *et al.*, 2023).

CONCLUSIONES

En el capítulo se indagó por el panorama general de las ciencias ómicas, abarcando áreas como la genómica, la proteómica, la transcriptómica, la metabolómica y otras ómicas, proporcionando una comprensión integral de estas disciplinas. Se resaltó el papel transformador de las ciencias ómicas en la investigación biomédica y la biotecnología, brindando nuevas herramientas y enfoques para comprender la complejidad de los sistemas biológicos a niveles molecular y sistémico.

La historia de las ciencias ómicas se presentó de manera precisa, desde los primeros avances hasta los proyectos más recientes, como el Proyecto Genoma Humano, el Proyecto ENCODE y otros proyectos emergentes en el campo.

El capítulo también abordó la importancia de la era multiómica y la integración de datos multiómicos, junto con el papel fundamental de la bioinformática y el análisis de datos en estas áreas. Además, se planteó la necesidad continua de investigación y desarrollo en el campo de las ciencias ómicas, con el objetivo de superar desafíos técnicos, así como de aprovechar todo su potencial para beneficio de la sociedad.

Se ha destacado el impacto de la medicina de precisión y el enfoque ómico en la investigación, subrayando el potencial de estas tecnologías para la mejora de la salud humana. Se enfatizó la importancia de la colaboración interdisciplinaria entre científicos, bioinformáticos, médicos y otros profesionales de la salud para maximizar el potencial de las ciencias ómicas en la mejora de la medicina personalizada y la salud pública.

REFERENCIAS

ADAMS, T. S., MARLIER, A. and KAMINSKI, N. (2023). Lung Cell Atlases in Health and Disease. *Annual Review of Physiology*, 85, 47-69. <https://acortar.link/ZFbPCd>.

- AKBARIAN, S., LIU, C., KNOWLES, J. A., VACCARINO, F. M., FARNHAM, P. J., CRAWFORD, G. E., JAFFÉ, A. E., PINTO, D., DRACHEVA, S., GESCHWIND, D. H., MILL, J., NAIRN, A. C., ABYZOV, A., POCHAREDDY, S., PRABHAKAR, S., WEISSMAN, S., SULLIVAN, P. F., STATE, M. W., WENG, Z.,... SESTAN, N. (2015). The PsychENCODE project. *Nature Neuroscience*, 18(12), 1707-1712. <https://doi.org/10.1038/nn.4156>.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. and LIPMAN, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. <https://acortar.link/Dm5n4G>.
- AOKI-KINOSHITA, K., AGRAVAT, S., AOKI, N. P., ARPINAR, S., CUMMINGS, R. D., FUJITA, A., FUJITA, N., HART, G. M., HASLAM, S. M., KAWASAKI, T., MATSUBARA, M., MOREMAN, K. W., OKUDA, S., PIERCE, M., RANZINGER, R., SHIKANAI, T., SHINMACHI, D., SOLOVIEVA, E., SUZUKI, Y.,... NARIMATSU, H. (2016). GlyTouCan 1.0—The international glycan structure repository. *Nucleic Acids Research*, 44(D1), 1237-1242. <https://doi.org/10.1093/nar/gkv1041>
- APELLANIZ-RUIZ, M., GALLEGRO, C., RUIZ-PINTO, S., CARRACEDO, A. and RODRÍGUEZ-ANTONA, C. (2016). Human genetics: International projects and personalized medicine. *Drug Metabolism and Personalized Therapy*, 31(1), 3-8. <https://doi.org/10.1515/dmpt-2015-0032>
- AZAD, A. K., HAKIM, A., SOHAG, Md. M. H. and RAHMAN, M. (2023). Metabolomics in clinical diagnosis, prognosis, and treatment of infectious diseases. In M. Rahman (Ed.), *Metabolomics* (pp. 71-119). Academic Press. <https://acortar.link/2f7A1p>
- BAUERMEISTER, A., MANNOCHIO-RUSSO, H., COSTA-LOTUFO, L. V., JARMUSCH, A. K. and DORRESTEIN, P. C. (2022). Mass spectrometry-based metabolomics in microbiome investigations. *Nature Reviews Microbiology*, 20(3), 143-160. <https://acortar.link/Tw1ZZ8>
- BINO, R. J., HALL, R. D., FIEHN, O., KOPKA, J., SAITO, K., DRAPER, J., NIKOLAU, B. J., MENDES, P., ROESSNER-TUNALI, U., BEALE, M. H., TRETHERWEY, R. N., LANGE, B. M., WURTELE, E. S. and SUMNER, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, 9(9), 418-425. <https://doi.org/10.1016/j.tplants.2004.07.004>
- BIRNEY, E. (2021). The International Human Genome Project. *Human Molecular Genetics*, 30(R2), 161-163. <https://doi.org/10.1093/hmg/ddab198>
- BRAICU, C. (2021). Functional Genomics in Health and Disease. *International Journal of Molecular Sciences*, 22(23). <https://doi.org/10.3390/ijms222312944>
- BRAY, N. L., PIMENTEL, H., MELSTED, P. and PACTER, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5), 525-527. <https://doi.org/10.1038/nbt.3519>

- BRAZMA, A., PARKINSON, H., SARKANS, U., SHOJATALAB, M., VILO, J., Abeygunawardena, N., Holloway, E., KAPUSHESKY, M., KEMMEREN, P., LARA, G. G., OEZCIMEN, A., ROCCA-SERRA, P. and SANSONE, S. A. (2003). ArrayExpress—A public repository for microarray gene expression data at the EBI. *Nucleic Acids Research*, 31(1), 68-71. <https://acortar.link/Vtv8qw>
- BREITWIESER, F. P., Lu, J. and SALZBERG, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4), 1125-1136. <https://doi.org/10.1093/bib/bbx120>
- Broad Institute. (n.d.). *Picard Toolkit* [Repository in GitHub]. <https://broadinstitute.github.io/picard/>
- BROWN, H. L., Sherburn, I. A., GAFF, C., TAYLOR, N. and BEST, S. (2022). Structured approaches to implementation of clinical genomics: A scoping review. *Genetics in Medicine. Official Journal of the American College of Medical Genetics*, 24(7), 1415-1424. <https://doi.org/10.1016/j.gim.2022.03.017>
- BYRSKA-BISHOP, M., EVANI, U. S., ZHAO, X., BASILE, A. O., ABEL, H. J., REGIER, A. A., CORVELO, A., CLARKE, W. E., MUSUNURI, R., NAGULAPALLI, K., FAIRLEY, S., RUNNELS, A., WINTERKORN, L., LOWY, E., Human Genome Structural Variation Consortium, Flicek, P., Germer, S., Brand, H., Hall, I. M.,... Zody, M. C. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*, 185(18), 3426-3440. <https://doi.org/10.1016/j.cell.2022.08.004>
- CABELLO-YEVES, P. J., CALLIERI, C., PICAZO, A., MEHRSHAD, M., HARO-MORENO, J. M., RODA-GARCIA, J. J., DZHEMBKOVA, N., SLABAKOVA, V., SLABAKOVA, N., MONCHEVA, S. and Rodriguez-Valera, F. (2021). The microbiome of the Black Sea water column analyzed by shotgun and genome centric metagenomics. *Environmental Microbiome*, 16. <https://doi.org/10.1186/s40793-021-00374-1>
- CAMPBELL, M. P., HAYES, C. A., STRUWE, W. B., WILKINS, M. R., AOKI-KINOSHITA, K. F., HARVEY, D. J., RUDD, P. M., KOLARICH, D., LISACEK, F., KARLSSON, N. G. and PACKER, N. H. (2011). UniCarbKB: putting the pieces together for glycomics research. *Proteomics*, 11(21), 4117-4121. <https://doi.org/10.1002/pmic.201100302>
- CAMPBELL, M. S., HOLT, C., MOORE, B. and YANDELL, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48. <https://doi.org/10.1002/0471250953.bi0411s48>
- Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, 45(10), 1113-1120. <https://doi.org/10.1038/ng.2764>

- CANI, P. D. (2018). Human gut microbiome: hopes, threats and promises. *Gut*, 67(9), 1716-1725. <https://doi.org/10.1136/gutjnl-2018-316723>
- CARITHERS, L. J., ARDLIE, K., BARCUS, M., BRANTON, P. A., BRITTON, A., BUIA, S. A., Compton, C. C., DeLuca, D. S., Peter-Demchok, J., Gelfand, E. T., Guan, P., Korzeniewski, G. E., Lockhart, N. C., Rabiner, C. A., Rao, A. K., Robinson, K. L., Roche, N. V., Sawyer, S. J., Segrè, A. V.,... GTEx Consortium. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking*, 13(5), 311-319. <https://doi.org/10.1089/bio.2015.0032>
- CASPI, R., BILLINGTON, R., KESELER, I. M., KOTHARI, A., KRUMMENACKER, M., MIDFORD, P. E., Ong, W. K., Paley, S., Subhraveti, P. and Karp, P. D. (2020). The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Research*, 48(D1), 445-453. <https://doi.org/10.1093/nar/gkz862>
- CHEN, C., HUANG, H. and WU, C. H. (2017). Protein Bioinformatics Databases and Resources. In C. Wu, C. Arighi and K. Ross (Eds.), *Protein Bioinformatics. Methods in molecular biology* (Vol. 1558, pp. 3-39). Humana Press. <https://acortar.link/yNHI2H>
- CHEN, I. A., CHU, K., PALANIAPPAN, K., PILLAY, M., RATNER, A., HUANG, J., HUNTEMANN, M., VARGHESE, N., WHITE, J. R., SESHADRI, R., SMIRNOVA, T., KIRTON, E., JUNGBLUTH, S. P., WOYKE, T., ELOE-FADROSH, E. A., IVANOVA, N. N. and KYRPIDES, N. C. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, 47(D1), 666-677. <https://acortar.link/py79vw>
- CHONG, J., WISHART, D. S. and Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics*, 68(1). <https://doi.org/10.1002/cpbi.86>
- CHOPRA, R. S., CHOPRA, C. and SHARMA, N. R. (Eds.). (2020). *Metagenomics: Techniques, Applications, Challenges and Opportunities*. Springer Singapore. <https://doi.org/10.1007/978-981-15-6529-8>
- CLAUSSNITZER, M., CHO, J. H., COLLINS, R., COX, N. J., DERMITZAKIS, E. T., HURLES, M. E., KATHIRESAN, S., KENNY, E. E., LINDGREN, C. M., MACARTHUR, D. G., NORTH, K. N., Plon, S. E., REHM, H. L., RISCH, N., ROTIMI, C. N., SHENDURE, J., SORANZO, N. and MCCARTHY, M. I. (2020). A brief history of human disease genetics. *Nature*, 577(7789), 179-189.
- CLOUGH, E. and BARRETT, T. (2016). The Gene Expression Omnibus Database. In E. Mathé and S. Davis (Eds.), *Statistical Genomics. Methods in Molecular Biology* (Vol. 1418, pp. 93-110). Humana Press. https://doi.org/10.1007/978-1-4939-3578-9_5

- Consultor Salud. (1 de septiembre de 2022). Proyecto 1000 genomas amplía su oferta de recursos de acceso abierto. <https://acortar.link/blI2qQ>.
- DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M. and LI, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- DEMARIA, A. N. (2003). A structure for deoxyribose nucleic acid. *Journal of the American College of Cardiology*, 42(2), 373-374. [https://doi.org/10.1016/s0735-1097\(03\)00800-3](https://doi.org/10.1016/s0735-1097(03)00800-3)
- ERRIEN, T., JOHNSON, R., BUSSOTTI, G., TANZER, A., DJEBALI, S., TILGNER, H., GUERNEC, G., MARTIN, D., MERKEL, A., KNOWLES, D. G., LAGARDE, J., VEERAVALLI, L., RUAN, X., RUAN, Y., LASSMANN, T., CARNINCI, P., BROWN, J. B., LIPOVICH, L., GONZALEZ, J. M.,... GUIGÓ, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9), 1775-1789. <https://acortar.link/jWekQ9>
- DESIERE, F., DEUTSCH, E. W., KING, N. L., NESVIZHSHII, A. I., MALLICK, P., ENG, J., CHEN, S., Eddes, J., Loevenich, S. N. and Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Research*, 34, 655-658. <https://doi.org/10.1093/nar/gkj040>
- DINIZ, W. J. and CANDURI, F. (2017). Review-Article. Bioinformatics: an overview and its applications. *Genetics and Molecular Research*, 16(1), 1-21. <https://doi.org/10.4238/gmrl6019645>
- DONG, W., QIANG, B. and YANG, H. (2021). The international Human Genome Project. A milestone for life sciences and humanity —The three stages and three major impacts of the HGP, and three contributions by China. *Quantitative Biology*, 9(3). <https://doi.org/10.15302/J-QB-021-0265>
- EISSENBERG, J. C. (2014). Epigenetics: modifying the genetic blueprint. *Missouri Medicine*, 111(5), 428-433.
- Encode Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74. <https://doi.org/10.1038/nature11247>
- ENCODE. (n.d.). *Encode portal*. <https://www.encodeproject.org/>
- Epigenómica. (n.d.). *National Human Genome Research Institute*. <https://www.genome.gov/es/about-genomics/fact-sheets/Epigenomica>
- FAHY, E., SUBRAMANIAM, S., MURPHY, R. C., NISHIJIMA, M., RAETZ, C. R., SHIMIZU, T., SPENER, F., van Meer, G., Wakelam, M. J. and Dennis, E. A.

- (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50, 9-14. <https://doi.org/10.1194/jlr.R800095-JLR200>
- FENG, Y., ZHANG, Y., YING, C., WANG, D. and DU, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, 13(1), 4-16. <https://doi.org/10.1016/j.gpb.2015.01.009>
- FRANCESCHI, C., GARAGNANI, P., PARINI, P., GIULIANI, C. and SANTORO, A. (2018). Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nature Reviews Endocrinology*, 14(10), 576-590. <https://doi.org/10.1038/s41574-018-0059-4>
- FU, M. P., MERRILL, S. M., SHARMA, M., GIBSON, W. T., TURVEY, S. E. and KOBOR, M. S. (2023). Rare diseases of epigenetic origin: Challenges and opportunities. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1113086>
- GABRIEL, S. B., SCHAFFNER, S. F., NGUYEN, H., MOORE, J. M., ROY, J., BLUMENSTIEL, B., HIGGINS, J., DEFELICE, M., LOCHNER, A., FAGGART, M., LIU-CORDERO, S. N., ROTIMI, C., ADEYEMO, A., COOPER, R., WARD, R., LANDER, E. S., DALY, M. J. and ALTSHULER, D. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-2229.
- GENCODE. (n.d.). *Gencode portal*. <https://www.genencodegenes.org/>
- Genoma Humano. (n.d.). *National Human Genome Research Institute*. <https://www.genome.gov/es/genetics-glossary/Proyecto-Genoma-Humano>
- GERSTEIN, M. B., BRUCE, C., ROZOWSKY, J. S., ZHENG, D., DU, J., KORBEL, J. O., EMANUELSSON, O., ZHANG, Z. D., WEISSMAN, S. and SNYDER, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research*, 17(6), 669-681. <https://doi.org/10.1101/gr.6339607>
- GHURYE, J. S., CEPEDA-ESPINOZA, V. and POP, M. (2016). Metagenomic Assembly: Overview, Challenges and Applications. *The Yale Journal of Biology and Medicine*, 89(3), 353-362.
- GILBERT, J. A., JANSSON, J. K. and KNIGHT, R. (2018). Earth Microbiome Project and Global Systems Biology. *MSystems*, 3(3). <https://doi.org/10.1128/mSystems.00217-17>
- GILLESPIE, M., JASSAL, B., STEPHAN, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., SHAMOVSKY, V., WEISER, J., BRUNSON, T., SANATI, N., . . . D'EUSTACHIO, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1), 687-692. <https://doi.org/10.1093/nar/gkab1028>

- GOSÁLVEZ, J. and HORCAJADAS, J. A. (2018). Introduction. Human genome projects. The omics starting point. In J. Horcajadas and J. Gosálvez (Eds.), *Reproductomics. The -Omics Revolution and Its Impact on Human Reproductive Medicine* (pp. 7-29). Academic Press. <https://doi.org/10.1016/B978-0-12-812571-7.00001-0>
- GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. and ENRIGHT, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, *34*, 140-144. <https://doi.org/10.1093/nar/gkj112>
- GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. and GASCUEL, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307-321. <https://doi.org/10.1093/sysbio/syq010>
- HapMap. Mapa de Haplotipos. (n.d.). *National Human Genome Research Institute*. <https://www.genome.gov/es/genetics-glossary/HapMap-mapa-de-haplotipos>
- HARROW, J., FRANKISH, A., GONZALEZ, J. M., TAPANARI, E., DIEKHANS, M., KOKOCINSKI, F., AKEN, B. L., BARRELL, D., ZADISSA, A., SEARLE, S., BARNES, I., BIGNELL, A., BOYCHENKO, V., HUNT, T., KAY, M., MUKHERJEE, G., RAJAN, J., DESPACIO-REYES, G., Saunders, G.,... Hubbard, T. J. (2012). Gencode: the reference human genome annotation for The ENCODE Project. *Genome Research*, *22*(9), 1760-1774. <https://doi.org/10.1101/gr.135350.111>
- HASIN, Y., SELDIN, M. and LUSIS, A. (2017). Multi-omics approaches to disease. *Genome Biology*, *18*(1). <https://doi.org/10.1186/s13059-017-1215-1>
- HASTINGS, J., DE MATOS, P., DEKKER, A., ENNIS, M., HARSHA, B., KALE, N., MUTHUKRISHNAN, V., OWEN, G., TURNER, S., WILLIAMS, M. and STEINBECK, C. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research*, *41*, 456-463. <https://doi.org/10.1093/nar/gks1146>
- HERCEG, Z., GHANTOUS, A., WILD, C. P., SKLIAS, A., CASATI, L., DUTHIE, S. J., FRY, R., ISSA, J. P., KELLERMAYER, R., KOTURBASH, I., KONDO, Y., LEPEULE, J., LIMA, S. C. S., MARSIT, C. J., RAKYAN, V., Saffery, R., Taylor, J. A., Teschendorff, A. E., Ushijima, T.,... Hernandez-Vargas, H. (2018). Roadmap for investigating epigenome deregulation and environmental origins of cancer. *International Journal of Cancer*, *142*(5), 874-882. <https://doi.org/10.1002/ijc.31014>
- HOOD, L. and ROWEN, L. (2013). The Human Genome Project: big science transforms biology and medicine. *Genome Medicine*, *5*(9). <https://doi.org/10.1186/gm483>

- HOOGLAND, C., MOSTAGUIR, K., SANCHEZ, J. C., HOCHSTRASSER, D. F. and APPEL, R. D. (2004). SWISS-2DPAGE, ten years later. *Proteomics*, 4(8), 2352-2356. <https://acortar.link/fEoNGk>
- HORAI, H., ARITA, M., KANAYA, S., NIHEI, Y., IKEDA, T., SUWA, K., OJIMA, Y., TANAKA, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M. Y., Nakanishi, H., Ikeda, K.,... Nishioka, T. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703-714. <https://doi.org/10.1002/jms.1777>
- Human Proteome Organization. (n.d.). *Human Proteome Organization (HUPO) portal*. <https://www.hupo.org/>
- Integrative HMP (iHMP) Research Network Consortium. (2019). The Integrative Human Microbiome Project. *Nature*, 569(7758), 641-648. <https://acortar.link/IO7rRa>
- International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320. <https://doi.org/10.1038/nature04226>
- JEHAN, F., SAZAWAL, S., Baqui, A. H., Nisar, M. I., Dhingra, U., Khanam, R., Ilyas, M., Dutta, A., Mitra, D. K., Mehmood, U., Deb, S., Mahmud, A., Hotwani, A., Ali, S. M., Rahman, S., Nizar, A., Ame, S. M., Moin, M. I., Muhammad, S.,... Prematurity Research Center at Stanford University. (2020). Multiomics Characterization of Preterm Birth in Low- and Middle-Income Countries. *JAMA Network Open*, 3(12). <https://acortar.link/FYg1OL>
- Jones, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., McANULLA, C., McWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236-1240. <https://acortar.link/AF1WPb>
- KANAI, Y. and ARAI, E. (2014). Multilayer-omics analyses of human cancers: exploration of biomarkers and drug targets based on the activities of the International Human Epigenome Consortium. *Frontiers in Genetics*, 5. <https://acortar.link/cjutzE>
- KANEHISA, M. and GOTO, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27-30. <https://doi.org/10.1093/nar/28.1.27>
- KARCZEWSKI, K. J. and SNYDER, M. P. (2018). Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5), 299-310. <https://doi.org/10.1038/nrg.2018.4>

- KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., Pringle, T. H., ZAHLER, A. M. and HAUSSLER, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996-1006. <https://doi.org/10.1101/gr.229102>
- KHORKOVA, O., MYERS, A. J., HSIAO, J. and WAHLESTEDT, C. (2014). Natural antisense transcripts. *Human Molecular Genetics*, 23(R1), 54-63. <https://doi.org/10.1093/hmg/ddu207>
- KIND, T., LIU, K. H., LEE, D. Y., DEFELICE, B., MEISSEN, J. K. and FIEHN, O. (2013). LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nature Methods*, 10(8), 755-758. <https://doi.org/10.1038/nmeth.2551>
- KLEINO, I., FROLOVAITÈ, P., SUOMI, T. and ELO, L. L. (2022). Computational solutions for spatial transcriptomics. *Computational and Structural Biotechnology Journal*, 20, 4870-4884. <https://doi.org/10.1016/j.csbj.2022.08.043>
- KOSTIDIS, S., SÁNCHEZ-LÓPEZ, E. and GIERA, M. (2023). Lipidomics analysis in drug discovery and development. *Current Opinion in Chemical Biology*, 72. <https://acortar.link/ULjGmh>
- KUMAR, P. S. (2021). Microbiomics: Were we all wrong before? *Periodontology 2000*, 85(1), 8-11. <https://doi.org/10.1111/prd.12373>
- KUMAR, S., STECHER, G., LI, M., KNYAZ, C. and TAMURA, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>
- LANGMEAD, B. and SALZBERG, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. <https://doi.org/10.1038/nmeth.1923>
- LAPIERRE, P. and GOGARTEN, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in Genetics*, 25(3), 107-110. <https://doi.org/10.1016/j.tig.2008.12.004>
- LEE, T. Y., HUANG, H. D., HUNG, J. H., HUANG, H. Y., YANG, Y. S. and WANG, T. H. (2006). DbPTM: an information repository of protein post-translational modification. *Nucleic Acids Research*, 34, 622-627. <https://doi.org/10.1093/nar/gkj083>
- LEINONEN, R., SUGAWARA, H. and SHUMWAY, M. (2011). The sequence read archive. *Nucleic Acids Research*, 39, 19-21. <https://doi.org/10.1093/nar/gkq1019>
- LI, Y., WU, F. X. and NGOM, A. (2018). A review on machine learning principles for multi-view biological data integration. *Briefings in Bioinformatics* 19(2). <https://acortar.link/RzyX6V>

- LILL, J. R., MATHEWS, W. R., ROSE, C. M. and SCHIRLE, M. (2021). Proteomics in the pharmaceutical and biotechnology industry: a look to the next decade. *Expert Review of Proteomics*, 18(7), 503-526. <https://doi.org/10.1080/14789450.2021.1962300>
- LINDAHL, T., MODRICH, P. and SANCAR, A. (2016). The 2015 Nobel Prize in Chemistry The Discovery of Essential Mechanisms that Repair DNA Damage. *Journal of the Association of Genetic Technologists*, 42(1), 37-41.
- LÓPEZ DE MATURANA, E., ALONSO, L., ALARCÓN, P., MARTÍN-ANTONIANO, I. A., PINEDA, S., PIORNO, L., CALLE, M. L. and MALATS, N. (2019). Challenges in the Integration of Omics and Non-Omics Data. *Genes*, 10(3). <https://doi.org/10.3390/genes10030238>
- LOVE, M. I., HUBER, W. and ANDERS, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- MAGHSOUDI, Z., NGUYEN, H., TAVAKKOLI, A. and NGUYEN, T. (2022). A comprehensive survey of the approaches for pathway analysis using multi-omics data integration. *Briefings in Bioinformatics*, 23(6). <https://doi.org/10.1093/bib/bbac435>
- MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. and DEPRISTO, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. <https://doi.org/10.1101/gr.107524.110>
- Metagenómica. (n.d.). *National Human Genome Research Institute*. <https://www.genome.gov/es/genetics-glossary/Metagenomica>
- MEYER, F., PAARMANN, D., D'SOUZA, M., OLSON, R., GLASS, E. M., KUBAL, M., PACZIAN, T., RODRIGUEZ, A., STEVENS, R., WILKE, A., WILKENING, J. and EDWARDS, R. A. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9. <https://doi.org/10.1186/1471-2105-9-386>
- MICHALSKA, K. and JOACHIMIAK, A. (2021). Structural genomics and the Protein Data Bank. *The Journal of Biological Chemistry*, 296. <https://doi.org/10.1016/j.jbc.2021.100747>
- MOORE, J. E., PURCARO, M. J., PRATT, H. E., EPSTEIN, C. B., SHORESH, N., ADRIAN, J., KAWLI, T., DAVIS, C., DOBIN, A., KAUL, R., HALOW, J., VAN NOSTRAND, E., FREESE, P., GORKIN, D., SHEN, Y., HE, Y., MACKIEWICZ, M., PAULI-BEHN, F., WILLIAMS, B.,... WENG, Z. (2020). Enciclopedias ampliadas de elementos de ADN en los genomas humanos y de ratón. *Nature*, 583(7818), 699-710.

- MORAES, F. and GÓES, A. (2016). A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochemistry and Molecular Biology Education*, 44(3), 215-223. <https://doi.org/10.1002/bmb.20952>
- NAITHANI, N., SINHA, S., MISRA, P., VASUDEVAN, B. and SAHU, R. (2021). Precision medicine: Concept and tools. *Medical Journal Armed Forces India*, 77(3), 249-257. <https://doi.org/10.1016/j.mjafi.2021.06.021>
- NAKAGAWA, H. and FUJITA, M. (2018). Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Science*, 109(3), 513-522. <https://acortar.link/uzBCYz>
- National Human Genome Research Institute. (n.d.). 1944: DNA is “Transforming Principle”. Last updated April 23, 2013. <https://acortar.link/noRhSu>
- National Institutes of Health. (10 may, 2023). *Scientists release a new human “pangenome” reference*. <https://acortar.link/9Oaa7H>
- O’FARRELL, P. H. (2008). The pre-omics era: the early days of two-dimensional gels. *Proteomics*, 8(23-24), 4842-4852. <https://doi.org/10.1002/pmic.200800719>
- OMENN, G. S., LANE, L., OVERALL, C. M., CORRALES, F. J., SCHWENK, J. M., PAIK, Y. K., Van Eyk, J. E., Liu, S., Snyder, M., Baker, M. S. and Deutsch, E. W. (2018). Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *Journal of Proteome Research*, 17(12), 4031-4041. <https://acortar.link/AJuSIT>
- OMENN, G. S., LANE, L., OVERALL, C. M., PINEAU, C., PACKER, N. H., CRISTEA, I. M., Lindskog, C., Weintraub, S. T., Orchard, S., Roehrl, M. H. A., Nice, E., Liu, S., Bandeira, N., Chen, Y. J., Guo, T., Aebersold, R., Moritz, R. L. and Deutsch, E. W. (2023). The 2022 Report on the Human Proteome from the HUPO Human Proteome Project. *Journal of Proteome Research*, 22(4), 1024-1042. <https://doi.org/10.1021/acs.jproteome.2c00498>
- PAMMI, M., Aghaeepour, N. and Neu, J. (2023). Multiomics, artificial intelligence, and precision medicine in perinatology. *Pediatric Research*, 93(2), 308-315. <https://doi.org/10.1038/s41390-022-02181-x>
- PANGENOMAS. (n.d.). *National Institutes of Health*. <https://www.nih.gov/news-events/news-releases/scientists-release-new-human-pangenome-reference>.
- PAPATHEODOROU, I., MORENO, P., MANNING, J., FUENTES, A. M., GEORGE, N., FEXOVA, S., FONSECA, N. A., FÜLLGRABE, A., GREEN, M., HUANG, N., HUERTA, L., IQBAL, H., JIANU, M., MOHAMMED, S., ZHAO, L., JARNUCZAK, A. F., JUPP, S., MARIONI, J., MEYER, K.,... BRAZMA, A. (2020). Expression Atlas update: from tissues to single cells. *Nucleic Acids Research*, 48(D1), 77-83. <https://doi.org/10.1093/nar/gkz947>

- PATTI, G. J., YANES, O. and SIUZDAK, G. (2012). Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4), 263-269. <https://acortar.link/Yw2nai>
- PEI, B., SISU, C., FRANKISH, A., HOWALD, C., HABEGGER, L., MU, X. J., HARTE, R., BALASUBRAMANIAN, S., TANZER, A., DIEKHANS, M., REYMOND, A., HUBBARD, T. J., HARROW, J. and GERSTEIN, M. B. (2012). The GENCODE pseudogene resource. *Genome Biology*, 13(9). <https://doi.org/10.1186/gb-2012-13-9-r51>
- PEREZ-RIVEROL, Y., CSORDAS, A., Bai, J., BERNAL-LLINARES, M., HEWAPATHIRANA, S., KUNDU, D. J., INUGANTI, A., GRISS, J., MAYER, G., EISENACHER, M., PÉREZ, E., USZKOREIT, J., PFEUFFER, J., SACHSENBERG, T., YILMAZ, S., TIWARY, S., COX, J., AUDAIN, E., WALZER, M.,... VIZCAÍNO, J. A. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, 47(D1), 442-450. <https://doi.org/10.1093/nar/gky1106>
- PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. and COTTRELL, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551-3567. <https://acortar.link/ILd1EI>
- PICARD, M., SCOTT-BOYER, M. P., BODEIN, A., PÉRIN, O. and DROIT, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal*, 19, 3735-3746. <https://doi.org/10.1016/j.csbj.2021.06.030>
- PIRELLA KARLUSICH, J. J., IBARBALZ, F. M. and BOWLER, C. (2020). Phytoplankton in the Tara Ocean. *Annual Review of Marine Science*, 12, 233-265. <https://acortar.link/NwLs5x>
- PIÑA, B., RALDÚA, D., BARATA, C., PORTUGAL, J., NAVARRO-MARTÍN, L., MARTÍNEZ, R., FUERTES, I. and CASADO, M. (2018). Functional Data Analysis. Omics for Environmental Risk Assessment. In J. Jaumot, C. Bedia and R. Tauler (Eds.), *Comprehensive Analytical Chemistry* (Vol. 82, pp. 583-611). Elsevier. <https://acortar.link/X0MPwj>
- PRUITT, K. D., TATUSOVA, T. and MAGLOTT, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33, 501-D504. <https://doi.org/10.1093/nar/gki025>
- REEL, P. S., REEL, S., PEARSON, E., TRUCCO, E. and JEFFERSON, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances*, 49. <https://doi.org/10.1016/j.biotechadv.2021.107739>

- RNAcentral Consortium. (2021). RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Research*, 49(D1), 212-220. <https://doi.org/10.1093/nar/gkaa921>
- ROADMAP EPIGENOMICS CONSORTIUM, KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L.,... KELLIS, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539), 317-330. <https://doi.org/10.1038/nature14248>
- ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24-26. <https://doi.org/10.1038/nbt.1754>
- RODRÍGUEZ-VALERA, F. and USSERY, D. W. (2012). Is the pan-genome also a pan-selectome? *F1000 Research*, 1, 1-8. <https://doi.org/10.12688/f1000research.1-16.v1>
- RUSHTON, A. R. (2022). Cambridge geneticists and the chromosome theory of inheritance: William Bateson, Leonard Doncaster and Reginald Punnett 1879-1940. *Annals of Science*, 79(4), 468-496. <https://doi.org/10.1080/00033790.2022.2113141>
- SÁEZ, M. E. (2022). Omics in Clinical Practice: How Far Are We? *Diagnostics*, 12(7). <https://doi.org/10.3390/diagnostics12071692>
- SANGER, F. (2004). Determination of nucleotide sequences in DNA. *Bioscience Reports*, 24(4-5), 237-253. <https://doi.org/10.1007/s10540-005-2733-8>
- SANGER, F., NICKLEN, S. and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467. <https://doi.org/10.1073/pnas.74.12.5463>
- SATTERLEE, J. S., CHADWICK, L. H., TYSON, F. L., McALLISTER, K., BEAVER, J., BIRNBAUM, L., VOLKOW, N. D., WILDER, E. L., ANDERSON, J. M. and Roy, A. L. (2019). The NIH Common Fund/Roadmap Epigenomics Program: Successes of a comprehensive consortium. *Science Advances*, 5(7). <https://doi.org/10.1126/sciadv.aaw6507>
- SEARLE, B. C. (2010). Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*, 10(6), 1265-1269. <https://doi.org/10.1002/pmic.200900437>
- SEEMANN, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>

- SIGGENS, L. and EKWALL, K. (2014). Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *Journal of Internal Medicine*, 276(3), 201-214. <https://doi.org/10.1111/joim.12231>
- SIGRIST, C. J., de CASTRO, E., CERUTTI, L., CUCHE, B. A., HULO, N., BRIDGE, A., Bougueleret, L. and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41, 344-347. <https://doi.org/10.1093/nar/gks1067>
- SILVA, T. C., COLAPRICO, A., OLSEN, C., D'ANGELO, F., BONTEMPI, G., CECCARELLI, M. and NOUSHMEHR, H. (2016). TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000 Research*, 5, 1-51. <https://acortar.link/lwGn4H>
- SINGH, D., CHAUDHARY, P., TAUNK, J., SINGH, C. K., SINGH, D., TOMAR, R. S. S., ASKI, M., KONJENBAM, N. S., RAJE, R. S., SINGH, S., SENGAR, R. S., YADAV, R. K. and PAL, M. (2021). Fab Advances in Fabaceae for Abiotic Stress Resilience: From 'Omics' to Artificial Intelligence. *International Journal of Molecular Sciences*, 22(19). <https://acortar.link/dhoQJr>
- SU, M., ZHANG, Z., ZHOU, L., HAN, C., HUANG, C. and NICE, E. C. (2021). Proteomics, Personalized Medicine and Cancer. *Cancers*, 13(11). <https://doi.org/10.3390/cancers13112512>
- SUBRAMANIAN, I., VERMA, S., KUMAR, S., JERE, A. and ANAMIKA, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*, 14. <https://doi.org/10.1177/1177932219899051>
- SUD, M., FAHY, E., COTTER, D., AZAM, K., VADIVELU, I., BURANT, C., EDISON, A., FIEHN, O., HIGASHI, R., NAIR, K. S., Sumner, S. and Subramaniam, S. (2016). Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Research*, 44(D1), 463-470. <https://doi.org/10.1093/nar/gkv1042>
- TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., Ward, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., MARGARIT y ROS, I., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C.,... FRASER, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial pan-genome. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13950-13955. <https://doi.org/10.1073/pnas.0506758102>
- TOURDOT, R. W., BRUNETTE, G. J., PINTO, R. A. and ZHANG, C. Z. (2021). Determination of complete chromosomal haplotypes by bulk DNA sequencing.

- Genome Biology*, 22(1). <https://doi.org/10.1186/s13059-021-02330-1>
- TSUGAWA, H. (2018). Advances in computational metabolomics and databases deepen the understanding of metabolisms. *Current Opinion in Biotechnology*, 54, 10-17. <https://doi.org/10.1016/j.copbio.2018.01.008>
- TSUGAWA, H., IKEDA, K., TAKAHASHI, M., SATOH, A., MORI, Y., UCHINO, H., OKAHASHI, N., YAMADA, Y., TADA, I., BONINI, P., HIGASHI, Y., OKAZAKI, Y., ZHOU, Z., ZHU, Z. J., KOELMEL, J., CAJKA, T., FIEHN, O., SAITO, K., ARITA, M. and ARITA, M. (2020). A lipidome atlas in MS-DIAL 4. *Nature Biotechnology*, 38(10), 1159-1163. <https://doi.org/10.1038/s41587-020-0531-2>
- UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPE, C., SJÖSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S.,... PONTÉN, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220). <https://doi.org/10.1126/science.1260419>
- UNIProt CONSORTIUM. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1), 506-515. <https://doi.org/10.1093/nar/gky1049>
- VERNIKOS, G. S. (2020). A Review of Pangenome Tools and Recent Studies. In H. Tettelin and D. Medini (Eds.), *The Pangenome. Diversity, Dynamics and Evolution of Genomes* (pp. 89-112). Springer.
- WANG, M., CARVER, J. J., PHELAN, V. V., SANCHEZ, L. M., GARG, N., PENG, Y., NGUYEN, D. D., WATROUS, J., KAPONO, C. A., LUZZATTO-KNAAN, T., PORTO, C., BOUSLIMANI, A., MELNIK, A. V., MEEHAN, M. J., Liu, W. T., CRÜSEMANN, M., BOUDREAU, P. D., ESQUENAZI, E., SANDOVAL-CALDERÓN, M.,... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828-837. <https://doi.org/10.1038/nbt.3597>
- WANG, M., WEISS, M., SIMONOVIC, M., HAERTINGER, G., SCHRIMPF, S. P., HENGARTNER, M. O. and von MERING, C. (2012). PaxDb, a database of protein abundance averages across all three domains of life. *Molecular and Cellular Proteomics*, 11(8), 492-500. <https://doi.org/10.1074/mcp.O111.014704>
- WANG, T., Antonacci-Fulton, L., HOWE, K., LAWSON, H. A., LUCAS, J. K., Phillippy, A. M., POPEJOY, A. B., ASRI, M., CARSON, C., CHAISSON, M. J. P., CHANG, X., COOK-DEEGAN, R., FELSENFELD, A. L., FULTON, R. S., GARRISON, E. P., GARRISON, N. A., GRAVES-LINDSAY, T. A., Ji, H., KENNY, E. E.,... HAUSSLER, D. (2022). The Human Pangenome Project: a global resource to map genomic diversity. *Nature*, 604, 437-446. <https://doi.org/10.1038/s41586-022-04601-8>

- WANG, Z., GERSTEIN, M. and SNYDER, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>
- WATSON, J. D. and CRICK, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738. <https://acortar.link/roTfBj>
- WATSON, J. D. and CRICK, F. H. (2007). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Clinical Orthopaedics and Related Research*, 462, 3-5. <https://doi.org/10.1097/BLO.0b013e31814b9304>
- WHIRL-CARRILLO, M., HUDDART, R., GONG, L., SANGKUHL, K., THORN, C. F., Whaley, R. and Klein, T. E. (2021). An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology and Therapeutics*, 110(3), 563-572. <https://doi.org/10.1002/cpt.2350>
- WISHART, D. S., FEUNANG, Y. D., MARCU, A., GUO, A. C., LIANG, K., VÁZQUEZ-FRESNO, R., SAJED, T., JOHNSON, D., Li, C., KARU, N., SAYEEDA, Z., LO, E., ASSEMPOUR, N., BERJANSKII, M., SINGHAL, S., ARNDT, D., LIANG, Y., BADRAN, H., GRANT, J.,... SCALBERT, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, 46(D1), 608-617. <https://doi.org/10.1093/nar/gkx1089ç>
- WU, W. T., Li, Y. J., FENG, A. Z., LI, L., HUANG, T., XU, A. D. and LYU, J. (2021). Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Medical Research*, 8. <https://doi.org/10.1186/s40779-021-00338-z>
- XIA, J., PSYCHOGIOS, N., YOUNG, N. and WISHART, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, 37, 652-660. <https://doi.org/10.1093/nar/gkp356>
- YORK, W. S., MAZUMDER, R., RANZINGER, R., EDWARDS, N., KAHSAY, R., AO-KI-KINOSHITA, K. F., CAMPBELL, M. P., CUMMINGS, R. D., FEIZI, T., MARTIN, M., NATALE, D. A., PACKER, N. H., WOODS, R. J., AGARWAL, G., ARPINAR, S., BHAT, S., BLAKE, J., CASTRO, L. J. G., FOCHTMAN, B.,... ZHANG, W. (2020). GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology*, 30(2), 72-73. <https://doi.org/10.1093/glycob/cwz080>
- ZHANG, H.-W., Lv, C., ZHANG, L.-J., GUO, X., SHEN, Y.-W., NAGLE, D. G., ZHOU, Y.-D., Liu, S.-H., ZHANG, W.-D. and LUAN, X. (2021). Application of omics- and multi-omics-based techniques for natural product target discovery. *Biomedicine and Pharmacotherapy*, 141. <https://doi.org/10.1016/j.biopha.2021.111833>

- ZHANG, W. and DOLAN, M. E. (2010). Impact of the 1000 genomes project on the next wave of pharmacogenomic discovery. *Pharmacogenomics*, 11(2), 249-256. <https://doi.org/10.2217/pgs.09.173>
- ZIA, A., AZIZ, M., POPA, I., KHAN, S. A., HAMEDANI, A. F. and ASIF, A. R. (2022). Artificial Intelligence-Based Medical Data Mining. *Journal of Personalized Medicine*, 12(9). <https://doi.org/10.3390/jpm12091359>.

CAPÍTULO 2. PRINCIPALES TÉCNICAS ÓMICAS

1. INTRODUCCIÓN

La revolución en las ciencias biológicas ha sido impulsada en gran medida por la incursión de las ciencias ómicas en el campo de la investigación. Estas disciplinas, que abarcan la genómica, la transcriptómica, la proteómica y la metabolómica, representan un conjunto de herramientas analíticas de vanguardia que han desvelado las intrincadas redes moleculares y los secretos celulares que subyacen en los sistemas biológicos. En este capítulo se exploran las principales técnicas ómicas que han transformado el entendimiento de la biología, descubriendo las incógnitas del genoma, la expresión génica, las proteínas y los metabolitos, así como su impacto en la investigación científica actual.

Las principales técnicas utilizadas en las ciencias ómicas emplean herramientas analíticas de alto rendimiento que permiten generar y estudiar grandes cantidades de datos, revelando así la complejidad y la interconexión de los sistemas biológicos. Estos datos forman parte de una vía de información biológica diversa integrada, por lo tanto, es importante tener en cuenta que la salida de datos depende de la entrada y de los mecanismos de regulación que se producen entre ellos.

Conocer los fundamentos de las diferentes técnicas de secuenciación NGS, los microarreglos o, incluso, la espectrometría de masas, la espectroscopía de Raman y las imágenes de resonancia magnética –entre otras– brindarán una mejor comprensión de la arquitectura genética, la expresión génica, la estructura proteica y los metabolitos en los diferentes sistemas biológicos.

2. TÉCNICAS EN GENÓMICA

La genómica actual se fundamenta principalmente en el uso de técnicas y métodos analíticos de alto rendimiento, los cuales han demostrado ser precisos y eficientes. Cada una de estas técnicas tiene aplicaciones y beneficios propios, que permiten analizar y comprender diferentes aspectos de la genética y la biología de los organismos vivos (REUTER *et al.*, 2015). Gracias a estas técnicas, se ha logrado una mejor comprensión en la genética de enfermedades comunes

y huérfanas, así como el estudio de genomas completos, variaciones genéticas, expresión génica y otros aspectos de la genómica.

En la **tabla 13** se presentan las tecnologías de alto rendimiento utilizadas en la genómica, las cuales permiten realizar estudios más profundos y detallados de los organismos vivos.

Tabla 13. *Tecnología de alto rendimiento en genómica*

Tecnología de alto rendimiento en genómica	Datos obtenidos	Bases de datos	Usos principales	Referencia
NGS (Secuenciación de Nueva Generación)	SNP, variantes VNTR	DDBJ, GenBank, ENA	Análisis estructural y funcional del genoma	GOODWIN <i>et al.</i> (2016)
Secuenciación de genoma completo (WGS)	Secuencias de todo el genoma	DDBJ, GenBank, ENA	Mutaciones en enfermedades complejas, estudios de poblaciones	M A R D I S (2017)
Secuenciación de exoma completo (WES)	Variantes en regiones codificantes	DDBJ, GenBank, ENA	Diagnóstico de enfermedades hereditarias	BAMSHAD <i>et al.</i> (2011)
Secuenciación de una célula	Variabilidad en células individuales	ArrayExpress, NCBI SRA	Heterogeneidad genética, estudios en cáncer	W A N G y NAVIN (2015)
Secuenciación dirigida	Genes específicos o regiones	UCSC Genome, NCBI SRA	Mutaciones específicas, farmacogenómica	MELDRUM <i>et al.</i> (2011)
Microarreglos	CNV, SNP	GEO, ArrayExpress	Variación en el número de copias, perfiles de variación genética	LAFRAMBOISE (2009)
Genotipificación por secuenciación (GBS)	SNPs en poblaciones	NCBI dbSNP	Genética de poblaciones, mejoramiento genético	ELSHIRE <i>et al.</i> (2011)

Nota. Resalta tecnologías clave y sus aplicaciones en el análisis genómico, considerando bases de datos relevantes para almacenamiento y consulta de datos genómicos.

A lo largo de la historia, la secuenciación del ADN ha experimentado un progreso significativo y ha generado una revolución en la comprensión de la genética y abriendo nuevas perspectivas en la investigación científica. Desde el

histórico descubrimiento de la estructura del ADN en 1953 por Watson, Crick y Franklin, se han logrado importantes hitos en esta área. En 1965, Robert Holley llevó a cabo la secuenciación del primer ARNt, y tan solo siete años después, en 1972, Walter Fiers se convirtió en el primero en secuenciar un gen completo (PIERREL, 2012; BARCISZEWSKA *et al.*, 2016). En particular, Fiers secuenció el gen MS2 que codifica la proteína de la cubierta del bacteriófago. Para lograr este avance se utilizaron enzimas ARNasas para digerir el ARN del virus, aislando luego los oligonucleótidos que conforman el gen, y finalmente los separaron mediante técnicas de electroforesis y cromatografía (PIERREL, 2012).

Los sistemas de electroforesis también han impulsado el desarrollo de las técnicas de NGS (BEHJATI *et al.*, 2013; HU *et al.*, 2021). El primer sistema de electroforesis en gel de agarosa automática fue desarrollado por SMITH y sus colegas en el año 1962. Las principales diferencias entre este método y los convencionales radica en el marcaje y el número de reacciones (SMITH, 1991). En lugar de utilizar un solo fluorocromo, se etiquetaron los cuatro nucleótidos con fluorocromos de diferentes longitudes de onda de emisión fluorescente, lo que permitió reducir las cuatro reacciones a una sola. Este enfoque demostró una optimización en el tamaño del gel (como el diámetro y la longitud), las condiciones de la electroforesis (como la corriente y la temperatura), así como en los aspectos ópticos y electrónicos usados en la adquisición de datos y el *software* utilizado para el análisis de datos (SMITH, 1991).

Por otro lado, Cohen y sus colaboradores demostraron el uso de capilares rellenos de gel de poliacrilamida para lograr la separación de un solo nucleótido marcado de los oligonucleótidos de ADN mediante la detección por UV (COHEN *et al.*, 1988). Posteriormente, Swerdlow y Gesteland mejoraron esta técnica mediante el uso de fluorescencia ultrasensible, lo que resultó en una mejora notable en la velocidad, resolución y eficiencia en comparación con los métodos anteriores (SWERDLOW y GESTELAND, 1990).

Además, HUANG y su equipo desarrollaron la secuenciación del ADN utilizando la electroforesis de matriz capilar, la cual permite una separación paralela rápida seguida de la detección en columna mediante un escáner de fluorescencia confocal multicolor. Esta técnica ha demostrado ser altamente eficiente y ha contribuido significativamente al avance de la secuenciación del ADN (HUANG *et al.*, 2023).

En el campo de la genómica las diversas técnicas de alto rendimiento permiten analizar y secuenciar grandes cantidades de información genética. A continuación, se mencionan algunas de las técnicas, partiendo de los primeros métodos de secuenciación de primera generación hasta las modernas técnicas de segunda, tercera y cuarta generación.

2.1. SECUENCIACIÓN DE PRIMERA GENERACIÓN

Los métodos de secuenciación de primera generación se fundamentaron en la fragmentación química o terminación de cadena con didesoxinucleótidos (ddNTP) combinados con la separación por tamaño mediante electroforesis en gel.

2.1.1. TÉCNICA DE MAXAM Y GILBERT

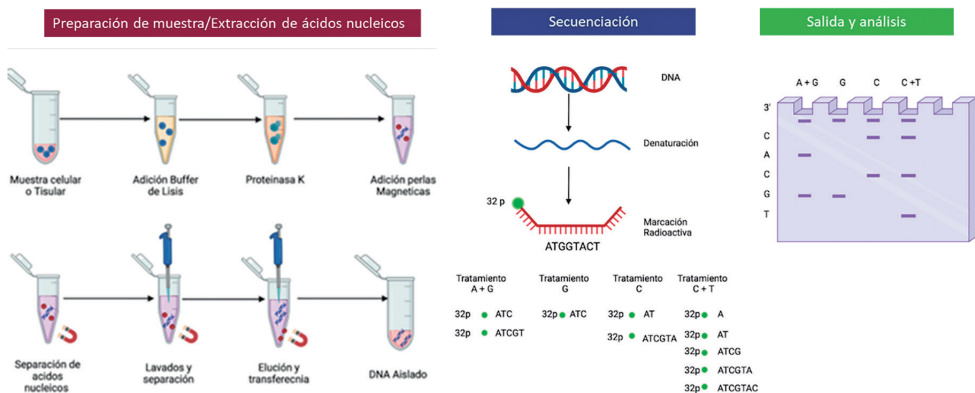
Fundamento: se basó en la capacidad de ciertos reactivos químicos para romper de manera controlada el ADN en posiciones específicas.

Las técnicas de secuenciación de primera generación se originaron con Allan Maxam y Walter Gilbert en 1970. Una vez se extraía el ADN mediante técnicas convencionales, este se radiomarcaba y se trataba con

varios productos químicos que rompían la cadena de ADN en bases específicas (MARTÍN *et al.*, 2020).

Posteriormente, estos fragmentos se separaban mediante la electroforesis en geles de poliacrilamida, con el objetivo de determinar la posición de cada nucleótido, mediante el análisis de las bandas resultantes se determinaba la secuencia del fragmento. La visualización de las bandas se lograba mediante el uso de marcadores radioactivos. Aunque esta técnica fue un hito en la genómica, presentaba limitaciones en términos de escalabilidad y costo (**figura 22**).

Figura 22. Método de secuenciación Maxam y Gilbert



2.1.2. TÉCNICA DE SANGER

Fundamento: la técnica de Sanger se fundamenta en la síntesis controlada de fragmentos de ADN utilizando ADN polimerasa y dideoxinucleótidos (ddNTPs) como terminadores de la cadena de ADN durante la replicación.

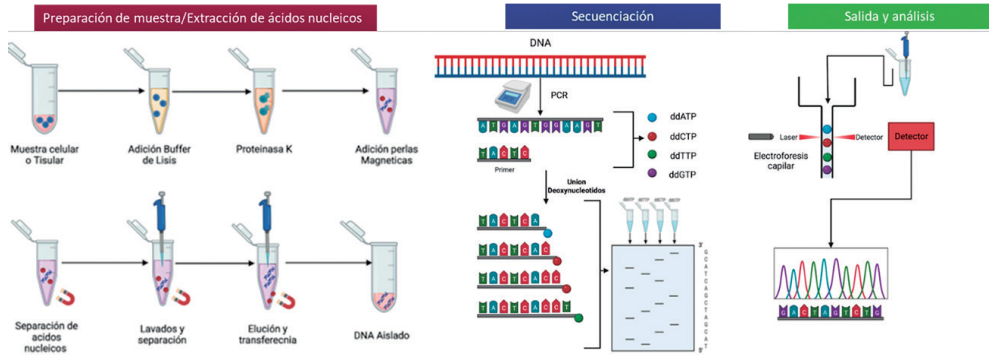
Esta técnica fue desarrollada en la década de 1970 por Frederick Sanger y permitió la secuenciación efectiva del ADN. Sanger introdujo una técnica revolucionaria basada en la síntesis del ADN mediante la actividad enzimática de la ADN polimerasa.

También utilizó análogos químicos de los desoxinucleótidos (dNTP) llamados dideoxinucleótidos (ddNTP), este método implicaba el uso de un cebador de ADN marcado, etiquetado con fluorescencia radiación (PAREEK *et al.*, 2011).

La principal innovación de esta técnica radicó en la incorporación selectiva de ddNTPs, los cuales actuaban como finalizadores de la cadena de ADN durante la síntesis *in vitro*. En cada reacción, se incorporaba un ddNTP que actuaba como un inhibidor, deteniendo la elongación de la cadena por parte de la enzima. Además, inicialmente estos ddNTPs se marcaban con radioisótopos de p32 para su detección y análisis (MARTÍN *et al.*, 2020).

Cada una de las reacciones en el método de Sanger se realiza de manera individual. En cada reacción se utiliza un dideoxinucleótido trifosfato marcado con un isótopo radiactivo, el cual actúa como un agente de terminación para la reacción. Posteriormente, las reacciones se separan en un gel de poliacrilamida y mediante una autorradiografía, se pueden visualizar los distintos fragmentos de ADN y así determinar la secuencia.

Este enfoque de secuenciación permitió obtener fragmentos de ADN de diferentes tamaños y marcados con precisión, los cuales luego podían ser separados y analizados en geles de electroforesis para determinar la secuencia de nucleótidos de la cadena original (PAREEK *et al.*, 2011). Después de separar y visualizar los fragmentos de ADN, la secuencia se leía manualmente a partir del patrón de las cuatro carreras paralelas para cada uno de los nucleótidos (**figura 23**).

Figura 23. *Secuenciación enzimática Sanger*

Nota. La figura ilustra el proceso de secuenciación de Sanger en tres etapas: primero, se extrae el ADN de una muestra mediante lisis celular y se purifica utilizando perlas magnéticas. Luego, se amplifica el ADN en una PCR con nucleótidos terminadores fluorescentes (ddNTPs) que generan fragmentos de diferentes longitudes. Finalmente, los fragmentos se separan por electroforesis capilar y un detector láser identifica la secuencia de bases, generando un cromatograma para su análisis.

En 1987, Leroy Hood y Michael Hunkapiller desarrollaron el ABI 370, un instrumento que automatiza el proceso de secuenciación de Sanger. Su logro innovador más importante fue el etiquetado automático de fragmentos de ADN con colorantes fluorescentes en lugar de moléculas radiactivas (PERVEZ *et al.*, 2022a). Este cambio no solo hizo que el método fuera más seguro, sino que también permitió que las computadoras analizaran los datos adquiridos.

El método de Sanger sigue siendo utilizado actualmente con mejoras, utilizando ddNTP marcados con fluorocromos y la electroforesis capilar (CE) y análisis con *software* especializado. Estas mejoras han allanado el camino para nuevas tecnologías de secuenciación del ADN, NGS (MARTÍN *et al.*, 2020).

Existen diversas limitaciones asociadas a las técnicas de secuenciación de primera generación.

- Con el método de Sanger solo se pueden secuenciar fragmentos cortos de ADN de aproximadamente 300 a 1.000 pares de bases (pb).
- La calidad de una secuencia de Sanger a menudo no es muy buena en las primeras 15 a 40 bases, porque aquí es donde se unen los cebadores.
- La secuenciación se degrada después de 700 a 900 bases.
- En primer lugar, el principal inconveniente radica en la generación y análisis de solo una secuencia por carril de electroforesis o tubo capilar. Esto limita

la capacidad de obtener múltiples secuencias simultáneamente, lo que resulta en un proceso más lento y menos eficiente (MARTÍN *et al.*, 2020).

- El rendimiento de las técnicas es generalmente más bajo en comparación con las tecnologías de NGS, lo que significa que se obtiene menos información en cada experimento. El alto costo asociado con estas técnicas es otro factor limitante, es más costosa que la secuenciación de segunda o tercera generación por base secuenciada.

En conclusión, la era de la genómica inició a finales de la década de 1980, cuando se realizaron modificaciones al método de Sanger para permitir la secuenciación automatizada y masiva, así como su integración con *software* de análisis incorporado en computadoras. Durante ese tiempo, se lograron avances significativos en la automatización de la secuenciación del ADN, con el desarrollo de secuenciadores automáticos equipados con láseres y detectores que registraban la emisión de luz generada durante la síntesis de ADN. Asimismo, se crearon programas de análisis que facilitaban la interpretación y procesamiento de los datos generados, permitiendo la identificación de secuencias de nucleótidos y su comparación con bases de datos genómicas existentes. Estos avances sentaron las bases para la genómica moderna y abrieron nuevas posibilidades en la investigación científica y médica.

Desde entonces, la tecnología de secuenciación ha seguido evolucionando, dando lugar a la aparición de nuevas técnicas de NGS, que permiten la secuenciación masiva, rápida y asequible de genomas completos (SATAM *et al.*, 2023). Estos avances continúan impulsando la genómica y su aplicación en la investigación científica y la medicina.

2.2. SECUENCIACIÓN DE SEGUNDA GENERACIÓN

Las técnicas de segunda generación se conocen como NGS, estas técnicas surgieron a finales de la década de 2000 y revolucionaron la genómica al permitir secuenciar grandes cantidades de ADN de manera simultánea debido a sus plataformas de alto rendimiento de secuenciación paralelas y rápidas que promueven el grado de cobertura de secuencia y precisión de lecturas individuales.

Estas técnicas se fundamentan en la amplificación, la síntesis secuencial y la detección de señales fluorescentes como medio para obtener millones de secuencias de ADN de manera simultánea, utilizan plataformas que permiten llevar a cabo simultáneamente miles de millones de procesos de secuenciación en paralelo sobre una superficie deslizante, como vidrio o perlas (SATAM *et al.*, 2023).

Este enfoque multiplexado (capacidad de secuenciar múltiples muestras de ADN en paralelo en un solo experimento) mejora significativamente el rendimiento y reduce el costo de la secuenciación, ya que se pueden obtener grandes cantidades de datos de secuenciación en un solo experimento. La evolución de los diferentes sistemas de electroforesis, marcajes químicos, mayor automatización en el laboratorio, el uso de computadoras y la miniaturización de las reacciones han permitido el avance en esta tecnología (SATAM *et al.*, 2023). Estas técnicas utilizan diferentes plataformas, como Illumina, Ion Torrent y PacBio entre otras, que generan millones de secuencias cortas o largas de ADN en paralelo (PERVEZ *et al.*, 2022a).

2.2.1. PROCEDIMIENTO GENERAL PARA LAS TÉCNICAS DE NGS

A pesar de la amplia gama de tecnologías NGS disponibles, los pasos básicos para secuenciar ADN y ARN son similares y están basados en los fundamentos de la secuenciación de Sanger con algunas modificaciones importantes (SATAM *et al.*, 2023).

- Preparación de las muestras y bibliotecas: como primer paso, se realiza la extracción del ADN/ARN, estas moléculas son fragmentadas y clonadas para formar una biblioteca de ADN. Los fragmentos se unen con moléculas adaptadoras que actúan en la hibridación de los fragmentos de la biblioteca con la matriz y proporcionan un sitio de cebado.
- Amplificación y secuenciación: la biblioteca se convierte en moléculas monocatenarias por desnaturalización. Durante la amplificación por la reacción en cadena de la polimerasa (PCR) se crean grupos de moléculas de ADN. Todos los grupos realizan reacciones individuales durante una sola ejecución de secuenciación.
- Salida y análisis de datos: la salida de un experimento de secuenciación depende de la tecnología y la química de secuenciación. Algunos secuenciadores generan señales de fluorescencia que se almacenan en archivos de salida específicos y otros pueden generar señales eléctricas que se almacenan en los formatos de archivo correspondientes. Generalmente, la cantidad de datos generados, los datos sin procesar, es muy grande. Dichos datos requieren un procesamiento por diferentes plataformas de bioinformática (SATAM *et al.*, 2023).

A continuación, se describen algunas de las tecnologías de segunda generación más utilizadas.

2.2.2. PIROSECUENCIACIÓN

Fundamento: se fundamenta en la medición de la síntesis de ADN por la detección de la luminiscencia generada por la liberación de pirofosfato (PPi) durante la incorporación de nucleótidos en la secuenciación de ADN.

En 1996, Mostafa Ronaghi, Mathias Uhlen y Pål Nyren introdujeron una nueva técnica de secuenciación por síntesis de ADN llamada pirosecuenciación, esta nueva técnica uti-

liza el método luminiscente para medir la síntesis de pirofosfato (HARRINGTON *et al.*, 2013). Este método marco un hito en la secuenciación del ADN y fue pionero en las técnicas de segunda generación. A continuación, se describe brevemente.

2.2.2.1. Preparación de las muestras y bibliotecas

El ADN extraído de la muestra que se quiere secuenciar se divide en fragmentos más cortos, generalmente de unos pocos cientos de pares de bases de longitud. Estos fragmentos se unen a adaptadores específicos en sus extremos, que contienen secuencias de reconocimiento para la amplificación.

2.2.2.2. Amplificación y secuenciación

En la pirosecuenciación, el proceso de amplificación del ADN tiene lugar dentro de microgotas de agua que se dispersan en una solución de aceite, técnica conocida como PCR en emulsión (emPCR). Cada gota contiene una sola plantilla de ADN y las enzimas necesarias para la amplificación por PCR. La molécula de ADN se une a una perla recubierta de cebador, donde se producen múltiples copias del fragmento de ADN original mediante ciclos repetidos de amplificación, lo que da como resultado la formación de colonias clonales (**figura 24**). Esta amplificación clonal permite la amplificación de plantillas de ADN individuales, lo que permite la secuenciación de múltiples moléculas de ADN en paralelo (SLATKO *et al.*, 2018).

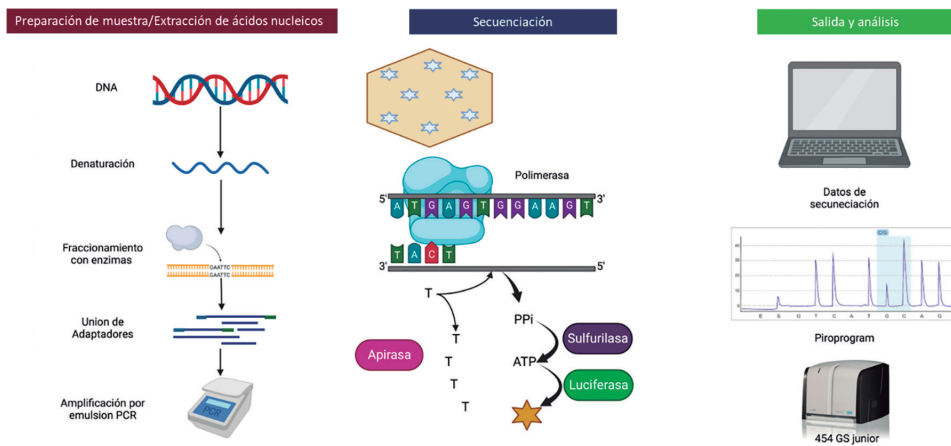
La máquina de secuenciación utilizada en la pirosecuenciación consta de numerosos pocillos con un volumen de picolitros, que contiene una sola perla y las enzimas para secuenciación. Cada microgota se fusiona con una esfera de captura que contiene enzimas de secuenciación y nucleótidos marcados con fluoróforos específicos para cada nucleótido (HARRINGTON *et al.*, 2013).

2.2.2.3. Salida y análisis de datos

El proceso de secuenciación implica la adición de nucleótidos individuales a la plantilla de ADN. A medida que se agrega cada nucleótido durante la

secuenciación, se libera PPi que se transforma en ATP que es detectado por la enzima luciferasa (oxiluciferina se transforma en luciferina), y genera una señal luminosa. La máquina de secuenciación detecta la luz emitida y la intensidad de la luz corresponde a la identidad del nucleótido agregado. Al detectar la luz emitida durante la secuenciación, la pirosecuenciación genera lecturas de secuencias que brindan información sobre la secuencia de ADN. Estas secuencias se ensamblan posteriormente utilizando algoritmos bioinformáticos para generar la secuencia completa del ADN original (**figura 24**).

Figura 24. Secuenciación por pirosecuenciación



Nota. La figura representa el proceso de pirosecuenciación en tres fases. Primero, el ADN se desnaturaliza, se fragmenta con enzimas, se le añaden adaptadores y se amplifica mediante PCR en emulsión. Luego, en la etapa de secuenciación, una polimerasa incorpora nucleótidos, generando una reacción en cascada que involucra a las enzimas apirasa, sulfurilasa y luciferasa, produciendo una señal luminosa cada vez que se añade un nucleótido. Finalmente, los datos se analizan en un pirograma, visualizando la secuencia de ADN en un sistema de secuenciación, como el 454 GS junior.

Los instrumentos desarrollados para llevar a cabo esta técnica han permitido un paralelismo masivo de reacciones de secuenciación. Esto significa que se pueden realizar simultáneamente miles de millones de reacciones de secuenciación en una sola ejecución. Esta capacidad de secuenciación masiva revolucionó el campo de la secuenciación del ADN al aumentar de manera significativa la cantidad de ADN que se puede secuenciar en un solo experimento (SLATKO *et al.*, 2018).

El secuenciador Roche GS-FLX 454 desarrollado en 2005 fue el primer secuenciador en automatizar completamente el proceso de pirosecuenciación en una sola máquina automatizada. Aunque la pirosecuenciación ha sido

reemplazada en gran medida por tecnologías de secuenciación más nuevas, desempeñó un papel importante en el avance del campo de la secuenciación de ADN y proporcionó información valiosa sobre el análisis de ADN. Su desarrollo y utilización allanaron el camino para la innovación y el progreso continuos en el campo de la secuenciación de alto rendimiento.

2.2.3. SECUENCIACIÓN POR LIGACIÓN Y DETECCIÓN DE OLIGONUCLEÓTIDOS (SOLiD)

La técnica de secuenciación SOLiD fue desarrollada por Applied Biosystems, genera fragmentos de ADN y se unen a adaptadores específicos mediante una ligación enzimática (CARRILLO, 2010). Después de la amplificación, se realiza una ligadura de oligonucleótidos cortos, conocidos como *oligos de color* que permiten la detección del ADN secuenciado. A continuación, se describen los pasos.

2.2.3.1. Preparación de las muestras y bibliotecas

Se inicia con una preparación de la muestra en la que se realiza una fragmentación del ADN (digestión con enzimas de restricción y/o sonicación) y la adición de dos tipos de adaptadores universales que se deben ligar a los fragmentos de ADN (p1 y p2). Estos adaptadores contienen secuencias conocidas en ambos extremos, lo que permite la amplificación y la fijación de los fragmentos de ADN en una matriz sólida, como un dispositivo en cristal o una microesfera (YEGNASUBRAMANIAN, 2013).

2.2.3.2. Amplificación y secuenciación

Luego, se realiza una amplificación por PCR para generar millones de copias de los fragmentos de ADN unidos a los adaptadores. La amplificación comienza con la hibridación de la secuencia con las nanoesferas, que tienen secuencias que van a hibridar con las secuencias del adaptador p1, luego se hace un proceso de emulsión, en el que se colocan los reactivos de la PCR con aceite.

Posteriormente, se realiza la desnaturalización, anillamiento de un fragmento de ADN con la nanoesfera, la polimerasa amplifica dicho fragmento que será el oligo *forward*, este fragmento se vuelve a desnaturalizar, el fragmento de la nueva síntesis hibrida con otro fragmento de la nanoesfera por su extremo p1, estos dos fragmentos se replican aumentando la cantidad de fragmentos que hibridaron con P1. Las esferas vacías se retiran colocando nanoesferas de poliestireno que hibridan con secuencias de p2. Una vez las nanoesferas de interés esta separadas se colocan en una placa de vidrio para fijarlas (YEGNASUBRAMANIAN, 2013).

El proceso de secuenciación implica el ADN a secuenciar, la ligasa, el primer p1, y sondas octámeras, cada una de las cuales consta de dos bases específicas (extremo 3'). Las bases específicas de la sonda pueden ser cualquiera de las 16 combinaciones posibles de dos bases. Dependiendo de las bases será el color del marcaje de la sonda. A continuación, hay tres bases universales se pueden unir a todas las bases y la tres bases con marcador fluorescentes al final que impide que la ligasa actúe.

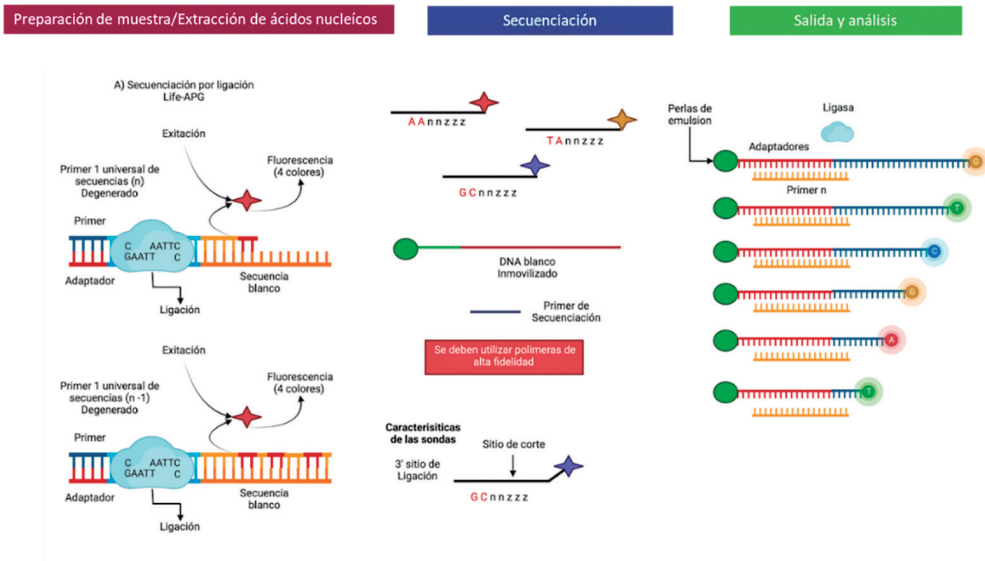
La ligación comienza cuando el primer p1 se une a la secuencia de ADN, luego se une la sonda. La plantilla es interrogada por sondas que representan las 16 combinaciones de dos bases. Por ejemplo, si las bases específicas complementarias al molde son AT, se utiliza la sonda correspondiente. Después del recocido y ligadura de la sonda, se registra la fluorescencia (YEGNASUBRAMANIAN, 2013). Luego de la unión se retira el marcaje de la sonda con láser y se elimina la fluorescencia que fue detectada por el equipo. Al quedar la sonda con el extremo 5' libre este se fosforila para permitir la unión de otra sonda marcada con fluorescencia y se repite el ciclo.

A continuación, la siguiente sonda se templea y se liga, y el cebador (n) se extiende por completo a lo largo de siete ciclos de ligadura. El producto extendido del cebador (n) se desnaturaliza del adaptador/plantilla y la segunda ronda de secuenciación se realiza mediante cebador (n-1).

2.2.3.3. Salida y análisis de datos

Mediante el uso de cebadores desplazados progresivamente, las bases adaptadoras se secuencian y esta secuencia conocida, junto con la codificación del espacio de color, se utiliza para determinar la secuencia de la plantilla a través de la de convolución.

El análisis de datos se realiza con el sistema *2 base encoding* que se basa en una carta de colores que determina las secuencia de la primera y segunda base. Se analiza cada nanoesfera para determinar el último nucleótido de esta y el color de la fluorescencia para determinar la base a la que corresponde (YEGNASUBRAMANIAN, 2013). En esta tecnología, cada base de plantilla se interroga dos veces para garantizar la precisión en la secuenciación (**figura 25**).

Figura 25. Secuenciación por ligación y detección de oligonucleótidos (SOLiD)

Nota. La figura describe la secuenciación por ligación en tres pasos: primero, el ADN se prepara mediante la adición de adaptadores específicos y se inmoviliza en una superficie. Luego, en la etapa de secuenciación, sondas fluorescentes, marcadas con colores correspondientes a cada base (A, T, C, G), se ligan a secuencias específicas en el ADN inmovilizado, emitiendo señales de fluorescencia al unirse. Finalmente, estas señales se detectan y analizan para obtener la secuencia completa, utilizando polimerasas de alta fidelidad para asegurar precisión en la identificación de cada base.

La secuenciación SOLiD ofreció alta precisión y se empleó para aplicaciones como la secuenciación del genoma completo, la resecuenciación y el análisis de metilación del ADN.

2.2.4. SECUENCIACIÓN ILLUMINA

Fundamento: se apoya en la amplificación de fragmentos de ADN, su anclaje en una superficie sólida, la síntesis de cadenas complementarias mediante la incorporación de bases fluorescentes y la detección de la señal luminosa generada en cada incorporación.

En 1998, Shankar Balasubramanian y David Klenerman desarrollaron y adaptaron el proceso de secuenciación por síntesis para un nuevo método que utiliza colorantes fluorescentes en la empresa Solexa.

La tecnología de Solexa también forma la base de los secuenciadores de Illumina, que dominan el mercado actual. El método se basa en la secuenciación por síntesis (SBS) (KUMAR *et al.*, 2019).

En este método, se utiliza una mezcla de ADN polimerasa, cebadores y cuatro nucleótidos deoxirribonucleótidos trifosfato (dNTP) con marcadores fluorescentes específicos para cada base. La adición de una base ocurre de forma secuencial gracias al grupo hidroxilo en el extremo 3'-OH de los dNTP. Al finalizar la reacción de secuenciación, se eliminan la ADN polimerasa y los dNTP no utilizados, luego, se agrega una solución tampón para la excitación de la fluorescencia, y la señal fluorescente es registrada por un equipo óptico. Este proceso repetitivo permite la lectura secuencial de millones de fragmentos de ADN, lo que resulta en la secuencia completa de la muestra original (KUMAR *et al.*, 2019).

2.2.4.1. Preparación de la muestra y biblioteca

Para realizar la secuenciación de ADN o ARN, primero se extrae el material genético de la muestra biológica y se cuantifica su cantidad y se evalúa su calidad. A continuación, el ADN o ARN se fragmenta en trozos más pequeños y se añaden adaptadores, que son secuencias de ADN necesarias para su unión a la plataforma de secuenciación de Illumina y para los pasos posteriores de amplificación y secuenciación.

2.2.4.2. Amplificación y secuenciación

El primer paso en la secuenciación de las plataformas Illumina es la amplificación clonal e inmovilización de cada fragmento de ADN. La amplificación clonal es necesaria para generar una señal lo suficientemente fuerte para su detección. Durante este proceso se crea un racimo o perla que contiene aproximadamente 1.000 copias idénticas de una única molécula de ADN madre, que están físicamente aisladas de otras moléculas.

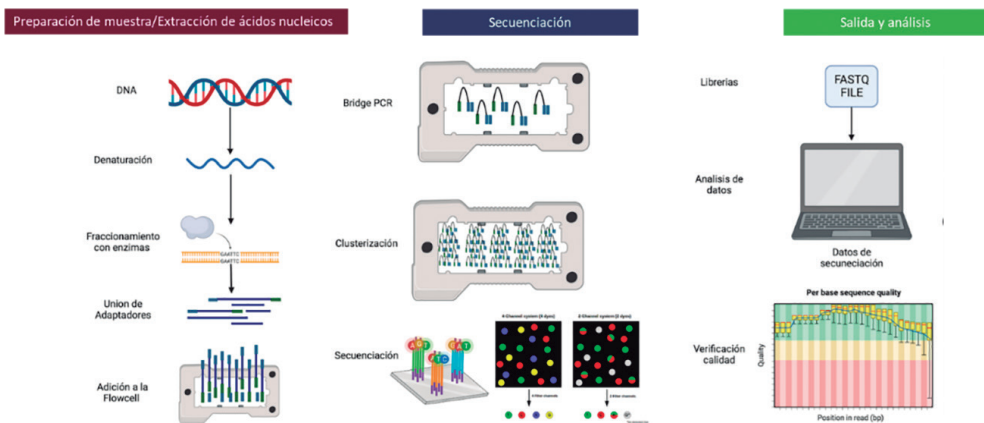
En la secuenciación de Illumina, las moléculas de ADN se unen a cebadores en un portaobjetos y se someten a un proceso de amplificación llamado amplificación de puente. Durante esta etapa, las moléculas de ADN se amplifican para formar grupos donde se encuentran múltiples copias del mismo fragmento de ADN en una pequeña área del portaobjetos (KUMAR *et al.*, 2019).

El método de amplificación de puente se basa en la marcación de los fragmentos de ADN con adaptadores. Estos fragmentos marcados se unen a una matriz de oligonucleótidos complementarios en una celda de flujo. Utilizando una reacción en cadena de la polimerasa (PCR) en fase sólida, se generan clones a partir de cada hebra de ADN unida a la celda original (KUMAR *et al.*, 2019). Esta amplificación clonal e inmovilización de los fragmentos de ADN prepara la muestra para el proceso de secuenciación (**figura 26**).

Los secuenciadores Illumina utilizan una celda de flujo que contiene perlas con secuencias complementarias aparte del adaptador presente en los fragmentos de ADN. La concentración adecuada de ADN de entrada es esencial para garantizar que solo se una un fragmento de ADN por perla y para que los fragmentos estén espaciados correctamente en la celda de flujo.

Durante la secuenciación, se agregan los cuatro nucleótidos marcados con fluorescencia, compitiendo por un lugar en la secuencia. El nucleótido marcado complementario se une, pero un bloqueador evita la adición de más de un nucleótido en cada ronda (química de terminación reversible). Los nucleótidos no unidos se eliminan mediante lavado. La excitación con láser produce una emisión fluorescente que se registra simultáneamente para cada grupo de fragmentos de ADN (KUMAR *et al.*, 2019).

Figura 26. *Secuenciación Illumina*



Nota. La figura muestra el proceso de secuenciación por síntesis de Illumina en tres pasos: en la primera etapa, el ADN se desnaturaliza, se fragmenta con enzimas, se le añaden adaptadores y se carga en una flowcell. En la etapa de secuenciación, se realiza una PCR en puente (Bridge PCR) para amplificar el ADN, formando grupos o clusters de secuencias idénticas en la flowcell. Luego, cada base se incorpora secuencialmente, emitiendo una señal fluorescente que es detectada para identificar la secuencia. Finalmente, en la fase de salida y análisis, los datos de secuenciación se registran en archivos FASTQ y se someten a verificación de calidad, visualizando la precisión de la lectura en un gráfico por posición de base.

2.2.4.3. Salida y análisis de datos

Después de obtener las imágenes de los nucleótidos marcados, se realiza el análisis para determinar la secuencia de los fragmentos de ADN. La intensidad

de la señal de fluorescencia en cada posición indica el nucleótido incorporado, y la secuencia se determina según el orden de las adiciones de nucleótidos (**figura 26**).

Cabe destacar que los detalles específicos y los *kits* utilizados pueden variar dependiendo del tipo de secuenciación (genómica, transcriptómica, etc.) y del protocolo utilizado. Es importante seguir las instrucciones proporcionadas por Illumina o adaptar los protocolos estándar a las necesidades experimentales específicas (KUMAR *et al.*, 2019).

En el método de secuenciación de Illumina, el tamaño de las secuencias o *reads* obtenidos varía según la plataforma utilizada. Generalmente, los tamaños pueden ser: 50 a 300 pares de bases (bp) para la mayoría de las plataformas Illumina, como el MiSeq y el NextSeq. En plataformas de alta capacidad como NovaSeq, los tamaños pueden alcanzar hasta 250-300 bp en modos de lectura simple o de extremos emparejados (*paired-end*). La elección del tamaño de lectura depende del tipo de experimento y de la plataforma específica de Illumina.

2.2.5. SECUENCIACIÓN DE NANOESFERAS DE ADN

Fundamento: se refiere a una tecnología de secuenciación que utiliza microesferas, generalmente de tamaño nanométrico, para realizar la secuenciación del ADN.

La empresa Complete Genomics ha desarrollado una tecnología emergente de secuenciación conocida como secuenciación de nano esferas de ADN (DNB), basada en la secuenciación

por hibridación y ligadura. En este método, cada microesfera contiene un fragmento de ADN que actúa como una plantilla para la síntesis de una cadena complementaria. Esta tecnología ofrece una alta capacidad de multiplexación y una producción masiva de datos secuenciales.

La DNB es una tecnología de secuenciación de alto rendimiento, se usa para establecer la secuencia completa del genómica de un organismo. Ofrece un enfoque rentable para secuenciar una gran cantidad de nanoesferas de ADN en una sola ejecución. La compañía *MGI Tech*, anexa del Grupo del Instituto de Genómica de Beijing (BGI), lanzó una serie equipos de NGS (BGI-200, BGI-500, MGISEQ-2000 y MGISEQ-T7) basadas en esta tecnología (KUMAR *et al.*, 2019).

2.2.5.1. Preparación de la muestra y amplificación

La secuenciación de nueva generación (DNB) implica inicialmente el aislamiento del ADN que será objeto de secuenciación. A continuación, se procede a fragmentarlo en trozos de 100 a 350 pares de bases (pb), seguido de la ligación de adaptadores a los fragmentos y su circularización.

2.2.5.2. Amplificación y secuenciación

Para amplificar los fragmentos de moléculas de ADN genómico en nanoesferas de ADN, se utiliza un proceso de replicación en círculo rodante. En este proceso, los fragmentos circulares se copian, lo que genera múltiples copias monocatenarias de cada fragmento. Estas copias de ADN se unen en una hebra larga, de la cabeza a la cola, y luego se compactan en nanoesferas de ADN. Los fragmentos de ADN se circularizan y replican varias veces para crear nanoesferas de ADN, que son esferas pequeñas y densas compuestas por muchas copias replicadas del fragmento de ADN original (KUMAR *et al.*, 2019).

En este método, se amplifican fragmentos de ADN de 440-500 pb mediante replicación en círculo rodante para obtener DNB. Estas nanoesferas de ADN se cargan en una celda de flujo, con una nanobola en cada pocillo. Las bases de la plantilla del 1 al 10 se procesan de manera emparejada, similar a la química de llamada exacta en la secuenciación SOLiD. Después de eliminar las secuencias ligadas, se agregan nuevas sondas para diferentes posiciones de interés. El proceso de recocido, lavado, ligadura y lectura de imágenes se repite para todas las posiciones cercanas a un extremo del adaptador (KUMAR *et al.*, 2019), el procedimiento se realiza para los terminales restantes del adaptador. La principal limitación de la secuenciación DNB es su tiempo de ejecución y las longitudes de lectura cortas, sin embargo, su principal ventaja radica en la alta cantidad de DNB generados (KUMAR *et al.*, 2019) (aproximadamente 350 millones).

2.2.5.3. Salida y análisis de datos

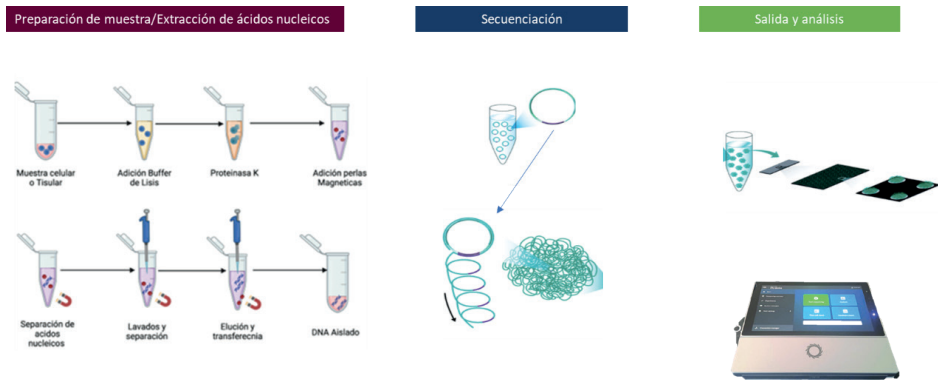
Después de eso, las nanoesferas se colocan en una celda de flujo de secuenciación, donde se realiza la adsorción. La cámara de alta resolución registra el color de fluorescencia en cada posición interrogada. Los datos de fluorescencia se analizan mediante bioinformática para realizar una llamada de base y mapear o cuantificar las lecturas de extremo único o emparejado de 50bp, 100bp o 150bp (KUMAR *et al.*, 2019) (**figura 27**).

La tecnología de secuenciación de nanoesferas de ADN presenta ciertas ventajas en comparación con otras plataformas de secuenciación. En primer lugar, se elimina la presencia de duplicados ópticos, ya que las nanoesferas de ADN permanecen en su lugar en la matriz modelada, sin interferir con las nanoesferas adyacentes. Esto contribuye a una mayor precisión en los resultados de secuenciación, además, se utiliza la ADN polimerasa Phi 29 de alta fidelidad, lo que asegura una amplificación precisa de la plantilla circular y minimiza los errores durante la lectura.

Posteriormente, se implementó el enfoque *Retrovolocity* para generar secuencias de alta calidad de todo el genoma (WG) y del exoma completo

(WE) con una cobertura de 50x en menos de 8 días. Según los informes, se han secuenciado más de 20.000 genomas humanos completos utilizando este instrumento y los procedimientos correspondientes (KUMAR *et al.*, 2019).

Figura 27. *Secuenciación por nanoesferas*



Nota. La figura muestra el proceso de secuenciación en tres pasos: primero, se extrae y purifica el ADN de una muestra mediante lisis y perlas magnéticas; luego, el ADN se prepara para la amplificación en estructuras circulares; finalmente, se coloca en un chip de secuenciación para su análisis en un sistema digital, obteniendo la información de la secuencia en tiempo real.

2.2.6. PRINCIPALES LIMITACIONES DE LAS TÉCNICAS DE SEGUNDA GENERACIÓN

- Las máquinas de secuenciación son costosas y, a menudo, deben compartirse con colegas.
- Los secuenciadores de segunda generación son máquinas estacionarias grandes y no diseñadas para trabajo de campo.
- En general, la secuenciación de segunda generación da como resultado muchos fragmentos de secuenciación cortos (lecturas) que son difíciles de usar para genomas nuevos.
- La calidad del resultado de la secuenciación depende del genoma de referencia.
- Comparadas con las técnicas de secuenciación de tercera generación, las plataformas de segunda generación suelen tener longitudes de lectura más cortas. Esto puede limitar la capacidad de secuenciar regiones genómicas que contienen repeticiones largas o estructuras complejas.

- Las técnicas de secuenciación de segunda generación pueden presentar errores de secuenciación, especialmente al final de las lecturas. Esto puede afectar la precisión de la secuencia obtenida y requerir técnicas adicionales, como el mapeo de lecturas o el uso de secuenciación de tercera generación, para corregir y validar los resultados.
- La mayoría de las técnicas de secuenciación de segunda generación requieren una etapa de amplificación previa de las muestras de ADN o ARN. Esto puede introducir sesgos en la representación de las secuencias y afectar la precisión de los resultados.
- Aunque las técnicas de segunda generación han mejorado significativamente en cuanto a costo y tiempo de procesamiento en comparación con las técnicas tradicionales de secuenciación, aún pueden requerir una inversión considerable y tiempo para la preparación de las muestras y la secuenciación en sí. Esto puede limitar su accesibilidad para algunos proyectos de investigación o aplicaciones clínicas.

La generación de grandes volúmenes de datos de secuenciación conlleva desafíos en el análisis y la interpretación de los resultados. El procesamiento y la interpretación de los datos de NGS requieren habilidades y recursos bioinformáticos adecuados para extraer información relevante y realizar análisis avanzados.

Es importante tener en cuenta que estas desventajas varían según la plataforma de secuenciación específica utilizada y que las tecnologías de NGS continúan evolucionando, abordando gradualmente muchas de estas limitaciones (SATAM *et al.*, 2023).

En conclusión, las técnicas de segunda generación, también conocidas como tecnologías de NGS, han revolucionado la genómica y ha permitido avances en la comprensión de la genética y biología molecular. Estas técnicas ofrecen secuenciación masiva y de alto rendimiento, acelerando la velocidad de secuenciación y posibilitando el estudio eficiente de genomas completos. Además, han reducido los costos de secuenciación, lo que las hace más accesibles para investigadores y amplía sus aplicaciones. También han mejorado la resolución y precisión de la secuenciación, permitiendo la detección detallada de variantes genéticas y mutaciones, su uso abarca diversos campos como la genómica, transcriptómica, epigenómica y metagenómica, y ha sido fundamental en el diagnóstico preciso de enfermedades genéticas, identificación de biomarcadores y estudio de enfermedades complejas. Estas tecnologías continúan evolucionando y desempeñando un papel esencial en la genómica moderna.

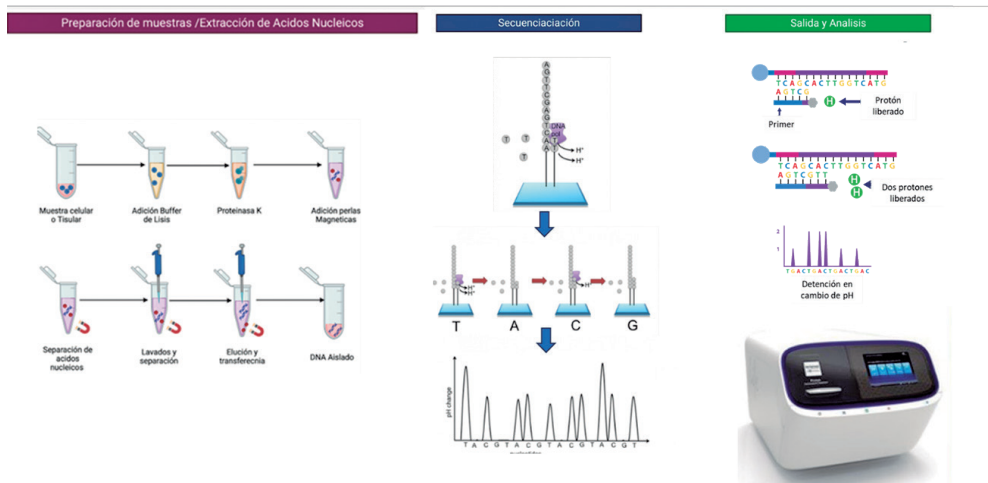
2.2.6.1. Ion Torrent

Fundamento: se fundamenta en la secuenciación por síntesis (SBS) y mide la diferencia de pH causada por la liberación de protones durante la polimerización, lo cual genera secuencias de nucleótidos.

Es una tecnología de secuenciación desarrollada por Life Technologies (ahora parte de Thermo Fisher Scientific). Esta tecnología se basa en la detección de protones liberados durante la síntesis del ADN complementario. Cuando se incorpora un nucleótido a la cadena de ADN en crecimiento, se libera un ion de hidrógeno (protones), la detección de estos protones permite determinar la secuencia de ADN original (SATAM *et al.*, 2023).

El fundamento de la tecnología Ion Torrent de Life Technologies se basa en la detección de iones liberados durante el proceso de secuenciación del ADN. Esta tecnología utiliza una plataforma de secuenciación basada en semiconductores, se coloca una matriz de sensores de silicio que detectan cambios en el pH cuando se liberan iones de hidrógeno durante la incorporación de nucleótidos en la cadena de ADN en crecimiento (**figura 28**).

Figura 28. Secuenciación por Ion Torrent



Nota. La figura resume la secuenciación por Ion Torrent en tres pasos: *i*) extracción y purificación del ADN mediante lisis celular y perlas magnéticas, *ii*) secuenciación mediante detección de protones liberados al añadir nucleótidos, y *iii*) análisis del cambio de pH para obtener la secuencia de ADN en tiempo real.

Durante la secuenciación, se realiza la amplificación de ADN en microesferas llamadas microesferas de emulsión. Estas microesferas contienen una hebra de ADN moldeada y se distribuyen en compartimentos individuales. Luego, se añaden los nucleótidos y las enzimas de secuenciación a cada compartimento, y cuando se incorpora un nucleótido a la cadena de ADN se libera un ion de hidrógeno que provoca un cambio de pH en el compartimento (SATAM *et al.*, 2023).

La matriz de sensores de silicio en la plataforma de Ion Torrent detecta estos cambios de pH y los convierte en una señal eléctrica proporcional a la secuencia del ADN. Esta señal eléctrica se procesa y se traduce en una secuencia de ADN.

La tecnología Ion Torrent es conocida por su velocidad y facilidad de uso, ya que no requiere la utilización de fluoróforos o técnicas de escaneo óptico. Además, ha demostrado ser eficiente en la secuenciación de fragmentos de ADN de longitud corta a mediana, sin embargo, tiene algunas limitaciones en la secuenciación de regiones repetitivas y en la detección de ciertas modificaciones epigenéticas (SATAM *et al.*, 2023).

2.2.6.2. Limitaciones de la secuenciación de tercera generación

- Algunos secuenciadores de tercera generación exhiben tasas de error más altas que los secuenciadores de segunda generación.
- Los reactivos son generalmente más caros que la secuenciación de segunda generación.
- Si bien las técnicas de secuenciación de tercera generación ofrecen muchas ventajas, también presentan algunas desventajas a tener en cuenta. Algunas de las desventajas más comunes de las técnicas de secuenciación de tercera generación son las siguientes:
 - En los costos en comparación con las generaciones anteriores, todavía pueden resultar más costosas que las técnicas de NGS. Esto puede limitar su accesibilidad para algunos laboratorios o proyectos con presupuestos limitados.
 - En términos de precisión, todavía pueden tener una tasa de error más alta en comparación con las técnicas de secuenciación de nueva generación. Esto puede llevar a cabo una generación de errores en la secuencia de ADN y se necesita un análisis adicional para corregirlos.

En conclusión, las técnicas de secuenciación de tercera generación han revolucionado aún más el campo de la genómica al superar las limitaciones de las tecnologías anteriores.

Las tecnologías de tercera generación permiten la secuenciación de fragmentos de ADN mucho más largos en comparación con las técnicas de segunda generación. Esto es especialmente útil para el ensamblaje de genomas complejos y la identificación de estructuras genéticas más grandes, como los elementos repetitivos. Estas técnicas han reducido significativamente el tiempo requerido para completar un experimento de secuenciación. Esto acelera la obtención de resultados y permite un análisis más rápido de las muestras. A diferencia de las técnicas de amplificación clonal utilizadas en la secuenciación de segunda generación, las tecnologías de tercera generación evitan en gran medida el sesgo de amplificación, lo que significa que se obtiene una representación más precisa del genoma original.

Las técnicas de tercera generación permiten una mejor detección de modificaciones epigenéticas, como la metilación del ADN y las modificaciones de histonas. Esto es crucial para comprender la regulación genética y los mecanismos epigenéticos en diferentes procesos biológicos. También han brindado una visión más detallada de la estructura del genoma, incluida la identificación de variantes estructurales, como inversiones, deleciones y traslocaciones. Esto ayuda a comprender mejor la variabilidad genómica y su relación con las enfermedades y los rasgos hereditarios.

2.3. SECUENCIACIÓN DE TERCERA GENERACIÓN

Las técnicas de secuenciación de tercera generación –conocidas como *Single Molecule Sequencing* (secuenciación de una sola molécula) y que también se conocen como enfoque de plantilla única– han surgido como una alternativa a las técnicas de secuenciación de segunda generación (VAN DIJK *et al.*, 2018).

Estas técnicas permiten la secuenciación directa de una sola molécula de ADN o ARN en tiempo real sin la necesidad de amplificación previa. Además, permiten la secuenciación de lectura larga, que describe la capacidad de obtener fragmentos de nucleótidos de longitudes más largas que las que generan los secuenciadores de lectura corta habituales de Illumina (del orden de 75 a 300 pares de bases según el secuenciador). Esto es especialmente importante para el ensamblaje de nuevos genomas sin un genoma de referencia disponible. En combinación con secuenciadores portátiles, que son de tamaño pequeño y no requieren máquinas más complejas para la química, la secuenciación ahora está lista para el campo y puede usarse incluso lejos de las instalaciones de laboratorio para recolectar muestras (KUMAR *et al.*, 2019).

2.3.1. HELISCOPE

Fundamento: se basa en una tecnología llamada secuenciación por síntesis de una sola molécula (SMS, por sus siglas en inglés) a través de la síntesis y la detección de nucleótidos de manera secuencial. Esto permite la secuenciación de fragmentos de ADN individualmente, evita la amplificación clonal y reduce la necesidad de manipulación previa de la muestra.

Fue desarrollado por la empresa Helicos Biosciences se basa en un enfoque de secuenciación llamado secuenciación por síntesis de una sola molécula. En esta tecnología, se utilizan cientos

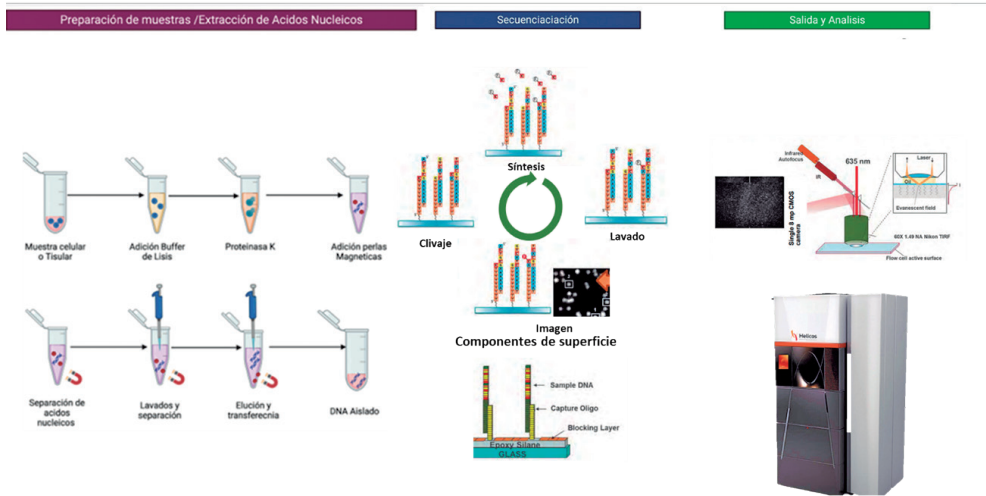
de miles de pequeñas cámaras de flujo para secuenciar directamente el ADN de una sola molécula. Cada cámara contiene una sola molécula de ADN y una enzima que sintetiza el nuevo ADN complementario. La secuenciación se realiza detectando la incorporación de nucleótidos marcados con fluoróforos.

En la secuenciación Heliscope, una molécula de ADN se desnaturaliza y se une a una cubierta en una superficie plana. Luego, se incorpora un cebador iniciador y se marca con una molécula fluorescente. A través de una reacción de polimerización, se sintetiza una cadena complementaria a partir de la molécula de ADN original, durante la síntesis, se agregan dNTP marcados con diferentes fluoróforos para identificar cada nucleótido incorporado (KUMAR *et al.*, 2019).

Después de cada ciclo de síntesis, se realiza una lectura de fluorescencia para determinar el nucleótido añadido en esa posición. Los fluoróforos se eliminan químicamente y el proceso se repite iterativamente hasta obtener la secuencia completa (**figura 29**).

El proceso de secuenciación en Helioscope implica una secuenciación basada en extensión con lavados cíclicos de la celda de flujo utilizando nucleótidos marcados con fluorescencia. Durante cada ciclo, se incorpora un nucleótido marcado en la cadena de ADN en crecimiento y se detecta y registra la señal de fluorescencia.

Las lecturas, o secuencias de los fragmentos de ADN, son generadas por el secuenciador Helioscope. Esta plataforma de secuenciación está diseñada para manejar los requisitos específicos de la secuenciación de una sola molécula de Helioscope, lo que garantiza una generación de datos precisa y confiable.

Figura 29. *Secuenciación por Helioscope*

Nota. La figura describe el proceso de secuenciación Helioscope: primero, se extrae y purifica el ADN mediante lisis celular y perlas magnéticas. Luego, en la etapa de secuenciación, el ADN se fija a una superficie de vidrio y se secuencia mediante síntesis con nucleótidos fluorescentes que se detectan en cada ciclo. Finalmente, en la fase de salida y análisis, un sistema de láser y detección óptica registra las señales de fluorescencia para generar la secuencia de ADN.

Una característica de la secuenciación Helioscope es que las lecturas son relativamente cortas, normalmente hasta 55 bases por ejecución, sin embargo, las mejoras recientes en la metodología han permitido lecturas más precisas de homopolímeros, que son tramos del mismo nucleótido en la secuencia de ADN. Además, la secuenciación de Helioscope se ha adaptado para la secuenciación de ARN, lo que permite el estudio de la expresión génica y la transcriptómica (SATAM *et al.*, 2023).

La longitud de lectura corta en la secuenciación de Helioscope puede presentar desafíos cuando se trata del ensamblaje y análisis del genoma, ya que generalmente se prefieren lecturas más largas para investigaciones genómicas integrales. No obstante, la secuenciación Helioscope ofrece ventajas en términos de precisión y la capacidad de capturar características genómicas específicas, como homopolímeros y transcripciones de ARN.

La secuenciación de una sola molécula con helioscopio ha encontrado aplicaciones en diversas áreas de investigación, incluidas la genómica, la transcriptómica y la secuenciación dirigida. Su capacidad para generar lecturas

precisas y proporcionar información sobre regiones genómicas específicas o perfiles de expresión de ARN lo convierte en una herramienta valiosa en la investigación de biología molecular. Es una tecnología de secuenciación que utiliza fragmentos de ADN con adaptadores de cola poliA añadidos. Estos fragmentos de ADN se adhieren a la superficie de una celda de flujo, que sirve como plataforma para la secuenciación (SATAM *et al.*, 2023).

Se espera que los continuos avances y refinamientos en la tecnología de secuenciación Helioscope mejoren aún más sus capacidades y amplíen sus aplicaciones en el futuro.

2.3.2. NANOPORE

Fundamento: se basa en el paso de fragmentos de ácido nucleico genómico monocatenario, a través de poros nanoporosos, se mide la corriente eléctrica generada por los nucleótidos individuales. Cada base de nucleótido individual bloquea físicamente la corriente, lo cual se puede medir mediante técnicas electrofisiológicas estándar.

La secuenciación por nanoporos es una técnica de secuenciación de ADN o ARN que se basa en el uso de nanoporos para determinar la secuencia de nucleótidos de una molécula. Un nanoporo es un poro extremadamente pequeño, del orden de nanómetros, que se inserta en una membrana (FENG *et al.*, 2015). En

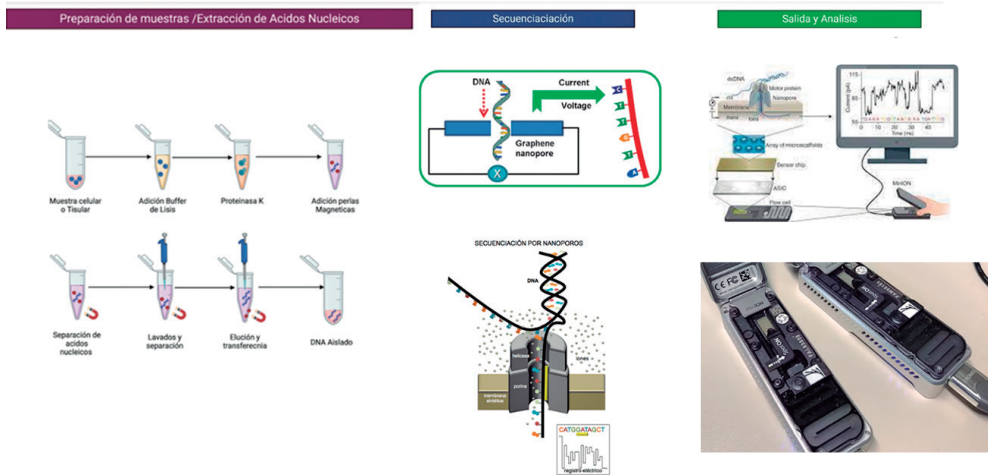
esta secuenciación la molécula de ADN o ARN se pasa a través del nanoporo y se registra la señal eléctrica generada por la interacción de los nucleótidos con el nanoporo (LIN *et al.*, 2021).

En este enfoque, se aplica un campo eléctrico que obliga a las moléculas individuales a atravesar un nanoporo con un diámetro de 2 nm. Debido al tamaño reducido del nanoporo, las moléculas monocatenarias pasan a través de él de manera ordenada y lineal. A medida que la molécula de ADN atraviesa el nanoporo, se generan señales eléctricas distintivas. La tecnología de nanoporos de *Oxford* es ampliamente reconocida en este campo y se destaca por su solidez, utiliza nanoporos, tanto biológicos como de estado sólido, que están incrustados en una membrana por la cual fluye una corriente iónica. Permite secuenciar genomas completos con lecturas largas de hasta un millón de pares de bases y diagnosticar enfermedades de manera eficiente y a bajo costo (WANG *et al.*, 2021).

El proceso de secuenciación por esta técnica comienza con la preparación de genotecas de muestra de ADN o ARN. La molécula se desenrolla y se pasa a través del nanoporo utilizando una corriente eléctrica. A medida que los nucleótidos individuales atraviesan el nanoporo, generan cambios en la corriente eléctrica que pueden ser registrados y analizados (WANG *et al.*, 2021). Cada nucleótido produce un patrón característico de señales eléctricas, lo que permi-

te determinar su identidad y, en última instancia, la secuencia completa de la molécula (figura 30).

Figura 30. *Secuenciación por nanopore*



Nota. La figura muestra el proceso de secuenciación por nanoporos. Primero, el ADN se extrae y purifica mediante lisis celular y perlas magnéticas. En la etapa de secuenciación, el ADN se pasa a través de un nanoporo impulsado por un voltaje, y cada base genera una señal eléctrica distinta mientras atraviesa el poro. Finalmente, en la fase de salida y análisis, un dispositivo detecta las variaciones de corriente y las traduce en la secuencia de ADN en tiempo real.

A diferencia de otros enfoques de NGS, la secuenciación de molécula única mediante nanoporos ofrece dos características distintivas. En primer lugar, permite la observación de la distribución espacial de las lecturas de ADN en la muestra, lo que proporciona información valiosa para destacar la heterogeneidad del tejido en función de los marcadores conocidos (WANG *et al.*, 2021). Esta capacidad espacial es especialmente relevante en aplicaciones de análisis de tejidos y muestras complejas. La segunda diferencia radica en la capacidad de analizar un gran número de células de manera simultánea. Por ejemplo, se han desarrollado enfoques robustos de secuenciación de ARN de células individuales que son rentables y capaces de secuenciar múltiples células utilizando una cantidad muy pequeña de material de partida (WANG *et al.*, 2021). La secuenciación de nanoporos se clasifica en tres categorías.

Secuenciación 1D. Se secuencian el ADN monocatenario, es decir, una sola cadena de ADN.

Secuenciación 2D. Se utilizan dos cadenas de ADN unidas por una estructura similar a una horquilla. Se obtiene la secuencia de una de las cadenas de ADN y luego se secuencia la segunda cadena. De esta manera, la secuenciación se realiza dos veces para aumentar la calidad de la llamada de bases.

Secuenciación 1D2. Se asemeja mucho a la secuenciación 2D, pero en este caso no es necesario mantener las dos hebras de ADN conectadas mediante la estructura de horquilla.

El principal desafío de esta técnica es el hecho de que el tejido biológico está compuesto por miles de células, por lo que secuenciar células individuales no es un proceso fácil desde el punto de vista técnico y computacional. Sin embargo, se prevé que la secuenciación *in situ* se utilizará para extraer información clínicamente relevante de los datos generados por los enfoques de NGS convencionales (WANG *et al.*, 2021).

En la secuenciación de nanoporos, existen varias plataformas populares en este campo. Por ejemplo, la *plataforma GS* tiene la capacidad de identificar moléculas, como proteínas, ADN y ARN, mientras pasan a través de pequeños agujeros arraigados en una membrana delgada. Esta tecnología propone una forma novedosa de secuenciación que constituye varios enfoques que dependen de la fusión de la preparación de la plantilla, la determinación del orden de las bases, la alineación de las secuencias y el ensamblaje del genoma (WANG *et al.*, 2021).

La secuenciación por nanoporos ofrece varias ventajas. En primer lugar, es una técnica de secuenciación de una sola molécula, lo que significa que no requiere amplificación previa del ADN o ARN. Esto evita problemas asociados con la amplificación, como la introducción de errores y la pérdida de información sobre modificaciones químicas. Además, la secuenciación por nanoporos tiene la capacidad de detectar modificaciones químicas en la secuencia, lo que es importante en la comprensión de la epigenética y otras modificaciones que pueden influir en la función del ADN o ARN (WANG *et al.*, 2021).

Sin embargo, la secuenciación por nanoporos también presenta desafíos. La lectura precisa de las señales eléctricas generadas por los nucleótidos puede ser complicada debido a la variabilidad de las señales y el ruido. Además, la velocidad de secuenciación actual de los dispositivos de nanoporos puede ser limitada en comparación con otras técnicas de secuenciación masiva paralela (WANG *et al.*, 2021).

Oxford Nanopore Technologies presentó GridION en 2012. GridION y sus sucesores MinIO y Flongle son secuenciadores portátiles para la secuenciación de ADN y ARN que producen lecturas de más de 2 Mb. En particular, dicho dispositivo de secuenciación incluso cabe en una sola mano humana. Los secuenciadores Oxford Nanopore observan los cambios en la corriente eléctrica

que se producen cuando los ácidos nucleicos pasan a través de los nanoporos, para identificar la secuencia de nucleótido (WANG *et al.*, 2021).

La secuenciación *in situ* dirigida puede aplicarse para detectar biomarcadores validados directamente en las muestras, mientras que la técnica no dirigida puede ser útil para desarrollar perfiles moleculares de las muestras y clasificar enfermedades a nivel molecular o atender las necesidades específicas de los pacientes. La integración de la secuenciación *in situ* en los métodos convencionales de NGS aceleraría el desarrollo de estas técnicas y, eventualmente, se convertirían en herramientas esenciales para la medicina personalizada (WANG *et al.*, 2021).

2.4. SECUENCIACIÓN DE CUARTA GENERACIÓN

La secuenciación de cuarta generación ha incorporado la tecnología de nanoporos en la secuenciación de molécula única, también conocida como tecnología de secuenciación *in situ*. Esta nueva tecnología ha revolucionado la secuenciación de ADN al permitir la identificación del orden de los nucleótidos en células y tejidos fijados de manera masiva, rápida y rentable. Estas tecnologías han revolucionado el campo de la genómica al permitir la secuenciación de genomas completos y facilitar el estudio de la variabilidad genética en diferentes organismos y enfermedades. La tecnología de secuenciación nanopor permite la secuenciación en tiempo real, proporcionando secuencias largas y la detección de modificaciones epigenéticas (KE *et al.*, 2016; MIGNARDI y NILSSON, 2014).

2.4.1. SECUENCIACIÓN DE MOLÉCULA ÚNICA EN TIEMPO REAL

Fundamento: es una tecnología que permite la secuenciación de una sola molécula de ADN o ARN sin la necesidad de amplificación previa y en tiempo real.

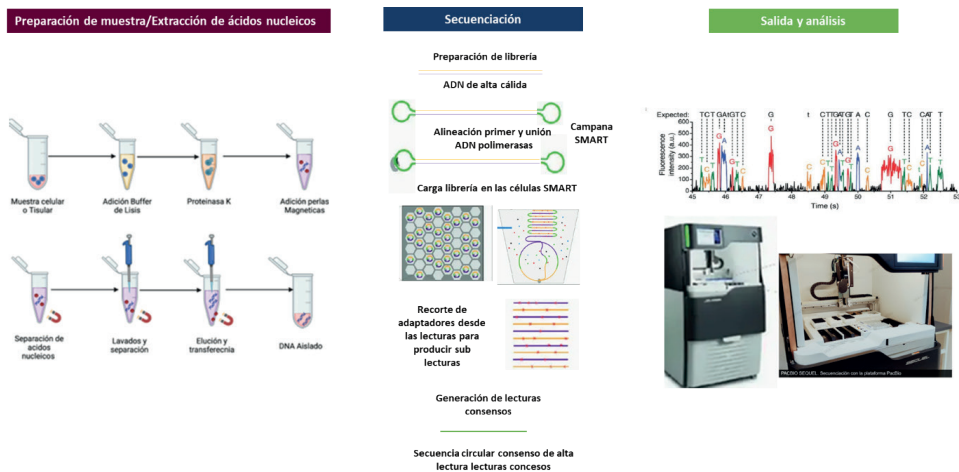
La secuenciación de molécula única en tiempo real (*Single Molecule Real-Time Sequencing, SMRT*) es una tecnología revolucionaria de secuenciación de ADN desarrollada por Pacific Biosciences. Se basa en el concepto de secuenciación por síntesis y ofrece capacidades únicas para la secuenciación de alto rendimiento y lectura larga. La SMRT, también conocida como secuenciación ARNP. utiliza la actividad de la ARN polimerasa para secuenciar moléculas de ADN en tiempo real (ARDUI *et al.*, 2018; VAN DIJK *et al.*, 2023).

Pacific Biosciences (PacBio) introdujo la secuenciación de guía de onda de modo cero (ZMW) en 2010, que utiliza los llamados nanoagujeros que contienen una sola ADN polimerasa. Esto permite que la incorporación de cualquier nucleótido único sea observada directamente por detectores conectados debajo de los nanoagujeros. Cada tipo de nucleótido está marcado con un colorante

fluorescente específico que emite señales fluorescentes durante el proceso de incorporación, que posteriormente se miden como lectura de secuencia, las lecturas obtenidas de los secuenciadores PacBio suelen ser de 8 a 15 kilobases (kb), con posibilidades de hasta 70 kb (VAN DIJK *et al.*, 2023).

En esta técnica, una única molécula de ADN o ARN se fija en una superficie y se le agrega una enzima polimerasa y nucleótidos marcados con fluoróforos. A medida que la polimerasa incorpora los nucleótidos en la molécula en crecimiento, se registra la señal de fluorescencia generada por cada nucleótido incorporado en tiempo real. Esto permite obtener la secuencia de la molécula original de manera directa y sin amplificación. La tecnología de secuenciación de molécula única en tiempo real proporciona lecturas largas y ofrece la capacidad de detectar modificaciones epigenéticas, como la metilación del ADN, durante el proceso de secuenciación (**figura 31**) (ARDUI *et al.*, 2018; VAN DIJK *et al.*, 2023). Esta técnica ha sido utilizada en diversas aplicaciones en el campo de la genómica y la biología molecular, y ha contribuido a la comprensión de la complejidad genética y la variabilidad de las moléculas de ADN y ARN (MANTERE *et al.*, 2019).

Figura 31. Secuenciación de molécula única en tiempo real



Nota. La figura representa el proceso de secuenciación de molécula única en tiempo real (SMRT) en tres etapas. Primero, el ADN se extrae y purifica mediante lisis celular y perlas magnéticas. En la fase de secuenciación, el ADN se fija en una matriz donde cada molécula se lee de manera individual y en tiempo real, utilizando nucleótidos marcados con fluoróforos que emiten señales cuando son incorporados en la cadena. Finalmente, en la etapa de salida y análisis, se detectan estas señales de fluorescencia de cada base incorporada, generando una secuencia precisa de la molécula de ADN en tiempo real.

La secuenciación comienza con la fabricación de una biblioteca de ADN de doble cadena ultra larga. A continuación, los adaptadores, la ADN polimerasa y los cebadores se unen al ADN de doble cadena para formar la campana SMRT, que luego se insertará en la célula SMRT. La biblioteca se distribuye aleatoriamente dentro de la celda SMRT del instrumento secuenciador; en circunstancias óptimas, un tercio de los ZMW se cargarán con una campana SMRT. En cada ZMW, la ADN polimerasa y una campana SMRT están unidas a la parte inferior del ZMW. La secuenciación SMRT utiliza la plantilla de ADN circular para generar una lectura larga continua en cada cámara ZMW (VAN DIJK *et al.*, 2023). Después de quitar los adaptadores de esta lectura larga, las lecturas superpuestas se pueden combinar para formar una lectura HiFi, que es una secuencia de consenso de alta calidad (ARDUI *et al.*, 2018).

La tecnología de secuenciación en tiempo real de una sola molécula (SMRT) utiliza guías de onda de modo cero (ZMW) para sintetizar y secuenciar moléculas de ADN o ARN en forma individual. Cada ZMW contiene una polimerasa de ADN y nucleótidos marcados con fluorescencia, permitiendo la detección en tiempo real de la incorporación de nucleótidos en una plantilla de secuencia complementaria (ARDUI *et al.*, 2018; VAN DIJK *et al.*, 2023). Esta tecnología ofrece lecturas extremadamente largas de hasta 10 kilobases (kb), lo que proporciona una alta resolución y cobertura para la caracterización de regiones genéticas complejas y la detección de variantes genéticas. Además, la secuenciación SMRT permite la observación de modificaciones naturales del ADN, como la metilación, durante el proceso de secuenciación. La secuenciación SMRT ha contribuido significativamente a nuestra comprensión de la genómica y la biología molecular al desbloquear regiones inaccesibles del genoma y revelar variaciones genéticas complejas (ARDUI *et al.*, 2018). Esta tecnología es valiosa para el ensamblaje de genomas complejos, la resolución de variaciones estructurales y el estudio de modificaciones epigenéticas. A medida que la tecnología continúa evolucionando, se espera que su desarrollo y aplicaciones se amplíen en investigación genómica, medicina personalizada y agricultura de precisión (ARDUI *et al.*, 2018; VAN DIJK *et al.*, 2023).

2.4.2. LA SECUENCIACIÓN POR CAPTURA DE IMAGEN (ICS, POR SUS SIGLAS EN INGLÉS)

Fundamento: técnica de secuenciación que se basa en la captura y detección de imágenes para obtener información sobre la secuencia de ADN.

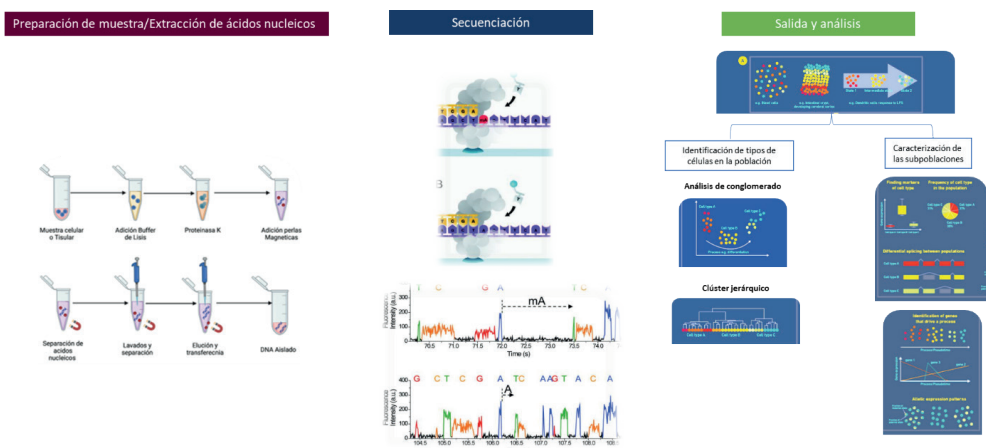
La secuenciación por captura de imagen es una tecnología que utiliza microscopía de alta resolución y amplificación localizada para secuenciar el ADN objetivo. En esta técnica, los fragmentos de ADN objetivo se unen a sondas

de ADN complementarias en una superficie sólida (SATAM *et al.*, 2023).

Luego, se realiza una amplificación localizada mediante PCR o amplificación isotérmica para generar múltiples copias de los fragmentos de ADN. Durante la secuenciación, se añaden nucleótidos marcados con fluoróforos específicos en cada ciclo, y las polimerasas de ADN catalizan la incorporación de estos nucleótidos en las cadenas de ácido nucleico complementarias.

La microscopía de alta resolución captura imágenes de la síntesis de ADN en tiempo real, y los tiempos de llegada y las duraciones de los pulsos de fluorescencia proporcionan información sobre la cinética de la polimerasa. Esto permite la detección directa de nucleótidos modificados en la plantilla de ADN. La secuenciación de molécula única en tiempo real proporciona una determinación precisa de la secuencia primaria de ADN, al tiempo que permite la detección de modificaciones epigenéticas como N6-metiladenosina, 5-metilcitosina y 5-hidroximetilcitosina (SATAM *et al.*, 2023) (figura 32).

Figura 32. Secuenciación por captura de imagen



Nota. La figura muestra la secuenciación por captura de imagen en tres pasos: extracción y purificación del ADN, incorporación de nucleótidos fluorescentes capturados en imágenes para registrar la secuencia, y análisis de datos para caracterizar tipos y subpoblaciones celulares y patrones de expresión genética.

2.4.3. LIMITACIONES DE LAS TÉCNICAS DE SECUENCIACIÓN DE CUARTA GENERACIÓN

Las técnicas de secuenciación de cuarta generación han revolucionado el campo de la genómica y han superado muchas de las limitaciones de las técnicas

de secuenciación de generaciones anteriores. Sin embargo, también presentan algunas limitaciones propias:

- Aunque las tecnologías de NGS pueden generar grandes cantidades de datos secuenciales, la longitud de las lecturas individuales es limitada en comparación con las técnicas de secuenciación de lectura larga. Esto puede dificultar el ensamblaje de genomas complejos y la identificación precisa de variantes estructurales (SATAM *et al.*, 2023).
- Las tecnologías de NGS están sujetas a errores de secuenciación, especialmente en regiones repetitivas o de baja calidad. Estos errores pueden afectar la precisión de los resultados y requerir análisis y correcciones adicionales.
- La mayoría de las tecnologías de NGS requieren la amplificación previa de las muestras de ADN o ARN para obtener suficiente cantidad de material para la secuenciación. La amplificación puede introducir sesgos y errores, lo que puede afectar la representatividad y la precisión de los resultados.
- Aunque los costos de la secuenciación de NGS han disminuido significativamente en los últimos años, sigue siendo una técnica costosa en comparación con otras técnicas de diagnóstico o análisis genético. Además, la implementación y el análisis de los datos generados por las tecnologías de NGS pueden requerir experiencia técnica y bioinformática avanzada (SATAM *et al.*, 2023).
- El procesamiento y análisis de los datos de NGS pueden llevar tiempo debido a la gran cantidad de información generada. Esto puede retrasar los resultados y limitar su utilidad en aplicaciones que requieren respuestas rápidas, como el diagnóstico de enfermedades.

En conclusión, la secuenciación por nanoporos ha demostrado ser una herramienta poderosa en la investigación genómica y la medicina. La tecnología continúa avanzando, lo que puede conducir a mejoras en la velocidad, precisión y costo de la secuenciación por nanoporos, ampliando así su aplicación en diferentes áreas de estudio. Las técnicas de secuenciación de cuarta generación permiten la generación de secuencias a gran escala en un tiempo más corto y a un menor costo por base secuenciada en comparación con las técnicas de secuenciación tradicionales.

Estas técnicas permiten la secuenciación simultánea de millones de fragmentos de ADN, lo que facilita el análisis de genomas completos, identificación de variantes genéticas y estudios de expresión génica a gran escala además, tienen diversas aplicaciones en campos como la medicina, la investigación genómica, la agricultura, la biotecnología y más. Permiten el estudio de enfermedades

genéticas, la identificación de biomarcadores, el análisis de variantes genéticas y la comprensión de la regulación génica, entre otros.

Estas técnicas ofrecen la posibilidad de un diagnóstico más preciso y temprano de enfermedades genéticas, lo que puede tener un impacto significativo en la atención médica y el desarrollo de terapias personalizadas. Sin embargo, el acceso a grandes volúmenes de información genética plantea desafíos éticos y legales, como la privacidad de los datos y la discriminación genéticos. Es necesario abordar estos problemas y establecer regulaciones adecuadas para proteger los derechos de las personas.

2.5. MICROARREGLOS

Fundamento: los microarreglos se basan en la hibridación específica de secuencias de ADN o ARN presentes en la muestra con sondas de ADN o ARN complementarias unidas a una superficie sólida, como un portaobjetos o una placa.

La tecnología de microarreglos, también conocida como micromatrices de ADN o chips de ADN, ha revolucionado el campo de los análisis genéticos al permitir la evaluación simultánea de miles de genes.

Estos microarreglos representan

una herramienta poderosa que logra la miniaturización de procesos biológicos. Inicialmente, se utilizó para medir la expresión génica de manera simultánea en miles de genes (APARNA y TETALA, 2023). Sin embargo, con el avance tecnológico, se ha mejorado la calidad y se ha ampliado el espectro de aplicaciones de los microarreglos, consolidándolos como herramientas indispensables en la investigación genética y con diversas aplicaciones en el ámbito de la medicina (ZHANG *et al.*, 2021).

El funcionamiento de los microarreglos de expresión se basa en el fenómeno de hibridación entre moléculas complementarias de ADN. En primer lugar, se depositan pequeñas cantidades de ADN correspondientes a los genes de interés cuya expresión se desea medir en una base de cristal (VAN STEENSEL, 2005). Para lograr esto, se utilizan robots de precisión que emplean agujas especiales para extraer las moléculas de ADN de sus recipientes y depositarlas en las coordenadas adecuadas en la matriz. A estas muestras de ADN depositadas en la matriz se les conoce como dianas (SATAM *et al.*, 2023).

El proceso de análisis mediante microarreglos implica varias etapas: Primero se diseñan sondas o secuencias de ADN o ARN específicas para cada gen o región de interés. Estas sondas son generalmente fragmentos cortos de ADN o ARN que son complementarios a la secuencia objetivo. La muestra de ADN o ARN de interés se extrae y se marca con una etiqueta fluorescente o un marcador detectable para permitir su posterior visualización (APARNA y TETALA, 2023).

La muestra preparada se aplica al microarreglo, donde las sondas están dispuestas en forma de una matriz ordenada. Las secuencias de ADN o ARN presentes en la muestra se hibridan específicamente con las sondas complementarias correspondientes en el microarreglo. Después de la hibridación, se realiza un escaneo del microarreglo para detectar la señal fluorescente o el marcador utilizado en la muestra. La intensidad de la señal en cada punto del microarreglo se correlaciona con la cantidad de secuencia de ADN o ARN presente en la muestra (VAN STEENSEL, 2005).

Los datos generados por el escaneo del microarreglo se analizan mediante métodos bioinformáticos para identificar los genes o secuencias de interés y realizar comparaciones entre muestras.

En un microarreglo típico, una superficie de 2 x 2 cm está diseñada para contener más de 10,000 dianas, que se disponen como pequeños puntos separados de manera precisa. Para medir la expresión de las células que deseamos analizar, comenzamos obteniendo una muestra de ARN, que se convierte en ADN complementario (ADNc) y se marca con una molécula fluorescente (APARNA y TETALA, 2023), a esta muestra marcada se le conoce como sonda, la cual se coloca en contacto con las dianas presentes en el microarreglo.

Cada molécula de ADNc marcada en la sonda se desplaza por difusión hacia la diana que contiene su molécula complementaria, se hibrida con ella y queda fijada en ese lugar específico. Después de un tiempo suficiente para permitir la hibridación de la mayoría de las cadenas complementarias, se procede a lavar el microarreglo y se realiza una medición relativa de la cantidad de ADN de la sonda que ha quedado fijada en cada diana (**figura 33**).

Figura 33. Microarreglos

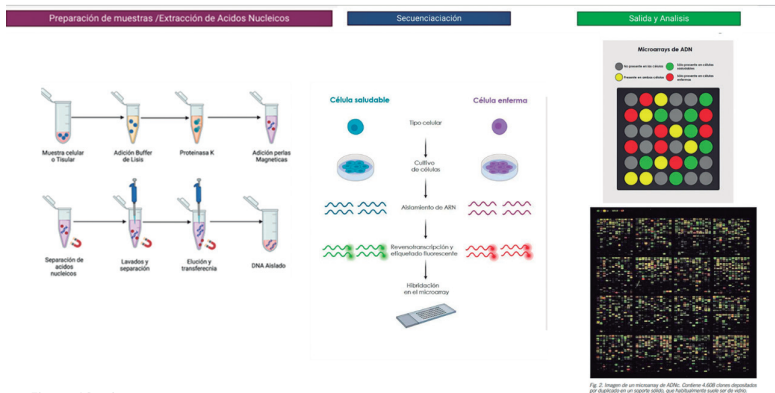


Figura 12 micro arrays

Fig. 2. Imagen de un microarreglo de ADN. Contiene 4.000 genes representados por fluorescencia en un arreglo más de 100 micrómetros por lado.

Nota. La figura muestra la técnica de microarrays de ADN en tres pasos. Primero, se extrae y purifica el ADN o ARN de células saludables y enfermas mediante lisis

celular y purificación con perlas magnéticas. Luego, en la etapa de secuenciación, el ARN extraído de ambos tipos de células se convierte en ADN complementario (cADN), se etiqueta con fluoróforos de distintos colores (verde para células saludables y rojo para células enfermas) y se hibrida en un microarray. Finalmente, en la fase de salida y análisis, se captura la imagen del microarray para analizar la expresión génica comparativa, donde el color y la intensidad de las señales indican la presencia y el nivel de expresión de genes en las diferentes muestras.

Existe una tecnología adicional que utiliza oligonucleótidos, que son secuencias cortas de ADN de alrededor de 15 a 30 bases. En lugar de depositar estos oligonucleótidos en el soporte mediante un robot, se sintetizan directamente sobre el soporte utilizando una técnica llamada fotolitografía. Esta técnica es similar a la utilizada en la fabricación de circuitos microelectrónicos de silicio. Sin embargo, es importante destacar que esta tecnología requiere una infraestructura altamente sofisticada y actualmente está limitada a unas pocas empresas especializadas, destacando *Affymetrix* (ZHANG *et al.*, 2021).

Para detectar la expresión de un gen, se emplea una amplia serie de oligonucleótidos, aproximadamente 30. Esto permite que estos microarreglos contengan muchas más dianas en comparación con otras tecnologías. La fotolitografía posibilita alcanzar una mayor densidad de oligonucleótidos en el soporte, lo que contribuye a la capacidad de detección mejorada de estos microarreglos (APARNA y TETALA, 2023).

Una vez amplificadas adecuadamente, una muestra de ADN obtenida de un tejido o fluido se puede analizar para detectar cambios en genes de interés o variantes genéticas, como polimorfismos de un solo nucleótido (SNP). Esta metodología tiene usos potenciales en la identificación de riesgos o susceptibilidades a enfermedades. Además, existen variantes de estos microarreglos que permiten la secuenciación de genes específicos conociendo su secuencia (APARNA y TETALA, 2023).

En el campo de la investigación genómica, se han desarrollado microarreglos similares a la técnica de hibridación genómica comparada (CGH) para detectar ganancias o pérdidas alélicas en miles de secuencias. Esto posibilita obtener mapas cromosómicos mucho más detallados que los obtenidos con la CGH tradicional. Estas técnicas son de gran interés para el estudio del pronóstico de tumores, ya que el nivel de daño genómico está asociado a dicho pronóstico, además, también pueden ser útiles para detectar nuevos oncogenes y genes supresores de tumores (ZHANG *et al.*, 2021).

3. TÉCNICAS EN TRANSCRIPTÓMICA

El transcriptoma se refiere al conjunto total de transcritos presentes en una célula en particular durante una etapa de desarrollo o una condición fisiológica

específica. Es fundamental comprender el transcriptoma para realizar una interpretación correcta de los elementos funcionales del genoma y descubrir los componentes moleculares de células y tejidos, así como comprender el desarrollo y la evolución de las enfermedades. Al medir los niveles de ARNm, podemos evaluar cómo las células remodelan su transcriptoma para adaptarse al entorno existente (por ejemplo, salud y enfermedad) (**tabla 14**).

Conocer la secuencia de ADN de un organismo y las posiciones de sus elementos reguladores nos dice muy poco sobre las operaciones dinámicas en tiempo real de una célula. Por ejemplo, al combinar diferentes sitios de empalme de ARNm y exones del mismo precursor de ARNm, un gen puede codificar múltiples proteínas. Este evento de empalme alternativo ocurre naturalmente y se ve comúnmente en eucariotas, sin embargo, una variante podría dar como resultado una enzima no funcional y un estado patológico inducido. Aquí es donde entra en juego la secuenciación de ARN (ARN-Seq). ARN-Seq sigue en gran medida los protocolos de secuenciación de ADN, pero incluye un paso de transcripción inversa en el que se sintetiza ADN complementario (ADNc) a partir de la plantilla de ARN (GONDANE y ITKONEN, 2023). La secuenciación del ARN permite obtener instantáneas de células, tejidos u organismos en el momento de la secuenciación en forma de perfiles de expresión de genes

La transcriptómica tiene como objetivos principales catalogar todas las especies de transcritos, que incluyen ARNm, ARN no codificantes y ARN pequeños; determinar la estructura de transcripción de los genes, incluyendo los sitios de inicio, los extremos 5' y 3', los patrones de corte y empalme, así como otras modificaciones postranscripcionales; y cuantificar los niveles de expresión cambiantes de cada transcrito durante el desarrollo y en diversas condiciones (**tabla 14**).

Tabla 14. *Técnicas en transcriptómica*

Técnicas en transcriptómica	Datos arrojados	Tecnología de alto rendimiento	Bases de datos	Enlace	Usos
mARN	Perfiles de expresión de Marn	ARN-Seq	GEO, ArrayExpress	GEO	Análisis de expresión génica
miARN	Perfiles de expresión de miARN	Microarreglos, ARN-Seq	miRBase	miRBase	Regulación génica
lncARN	Perfiles de expresión de lncARN	ARN-Seq	LNCipedia, NONCODE	LNCipedia	Regulación de expresión

Técnicas en transcriptómica	Datos arrojados	Tecnología de alto rendimiento	Bases de datos	Enlace	Usos
ARN de célula única	Expresión a nivel de célula única	scRNA-Seq	Single Cell Atlas	Atlas	Heterogeneidad celular
ARNt	Identificación y cuantificación	ARN-Seq	GtRNADB	GtRNADB	Perfiles de ARNt
ARNr	Perfiles de ARNr	Secuenciación dirigida, ARN-Seq	SILVA	SILVA	Análisis de comunidad microbiana

Nota. Esta tabla simplificada proporciona un resumen de las técnicas en transcriptómica, las tecnologías y bases de datos clave, así como sus aplicaciones. Elaboración propia a partir de Gondane e Itkonen (2023).

Se han desarrollado diversas tecnologías para inferir y cuantificar el transcriptoma, entre las cuales se incluyen enfoques basados en hibridación, secuenciación, hasta las últimas técnicas de transcriptomas espaciales.

La secuenciación de ARN moderna permite un muestreo imparcial de transcritos en contraste con, por ejemplo, ensayos basados en micromatrices o RT-qPCR, que requiere un diseño de sonda para dirigirse específicamente a las regiones de interés. Los perfiles de expresión génica obtenidos permiten además la detección de isoformas génicas, fusiones génicas, variantes de un solo nucleótido y muchas otras propiedades interesantes. La secuenciación de ARN moderna no está limitada por el conocimiento previo y permite la captura de características conocidas y novedosas, lo que da como resultados conjuntos de datos ricos que se pueden usar para el análisis de datos exploratorios (ITKONEN, 2023).

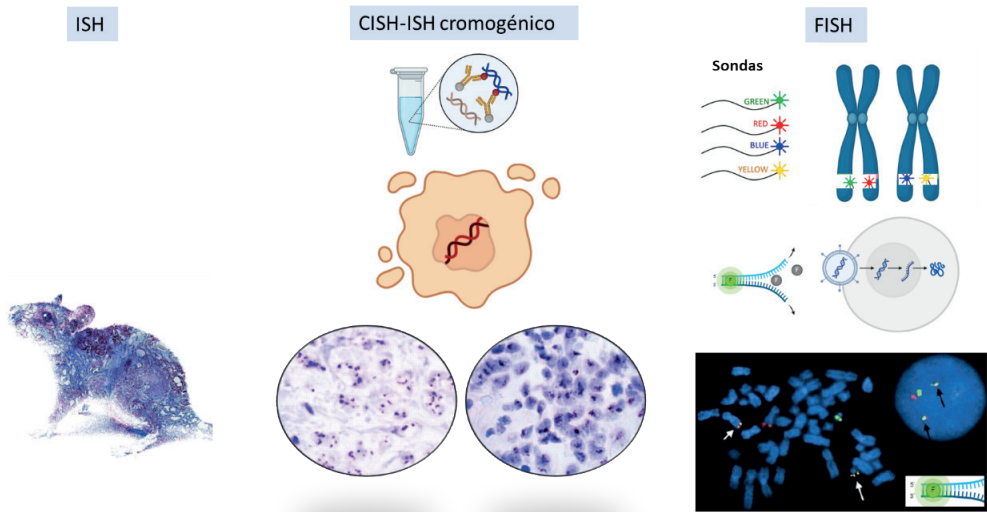
La secuenciación del ARN se puede llevar a cabo principalmente de dos formas: mediante la secuenciación del ARN mixto de la fuente de interés entre las células (secuenciación masiva) o mediante la secuenciación de los transcriptomas de las células individualmente (secuenciación unicelular).

3.1. ENFOQUES BASADOS EN HIBRIDACIÓN

La hibridación es una técnica muy utilizada para estudiar el ARN. Permite detectar y cuantificar la presencia de moléculas de ARN específicas en una muestra. A continuación, se presentan algunas de las técnicas de hibridación comunes utilizadas para el estudio del ARN (**figura 34**).

Hibridación *in situ* (ISH): esta técnica permite detectar la localización espacial de moléculas de ARN dentro de células o tejidos. Se utilizan sondas de ARN marcadas con fluorocromos o enzimas para hibridar con el ARN diana en la muestra. La detección de la señal se realiza mediante microscopía de fluorescencia o métodos histoquímicos.

Figura 34. Enfoques basados en hibridación



Nota. La figura presenta tres técnicas de hibridación *in situ*: ISH, para detectar secuencias específicas de ADN o ARN en tejidos, visualizando la expresión genética; CISH o ISH cromogénico, que utiliza marcadores visibles en microscopía óptica para localizar secuencias en el núcleo; y FISH, que emplea sondas fluorescentes para identificar secuencias en cromosomas y estudiar alteraciones cromosómicas. Estas técnicas permiten análisis detallados de expresión génica y estructura cromosómica en muestras biológicas.

Existen dos formas básicas de visualizar el ARN. Por hibridación *in situ*: la detección por fluorescencia (FISH) y la detección cromogénica (CISH). Ambos utilizan una sonda específica etiquetada que se hibridiza con la muestra, pero el instrumental utilizado para visualizar las muestras es diferente para cada método. En CISH se visualiza utilizando microscopía de campo claro, lo que permite ver la señal CISH y la morfología del tejido simultáneamente, esta tecnología es especialmente útil en el diagnóstico de patología molecular (ITKONEN, 2023).

3.1.1. *ARN-FISH*

Fundamento: es una técnica que utiliza sondas fluorescentes complementarias a secuencias específicas de ARN para detectar y visualizar la ubicación de ARN en células o tejidos.

Es una técnica de hibridación *in situ* que utiliza sondas de ARN marcadas con fluorocromos para detectar y localizar moléculas de ARN específicas dentro de células o tejidos, es útil para estudiar la

expresión génica. La hibridación se visualiza mediante microscopía de fluorescencia, lo que permite la observación directa de la localización espacial del ARN. También se puede visualizar utilizando citometría de flujo y citometría de alta capacidad (ITKONEN, 2023).

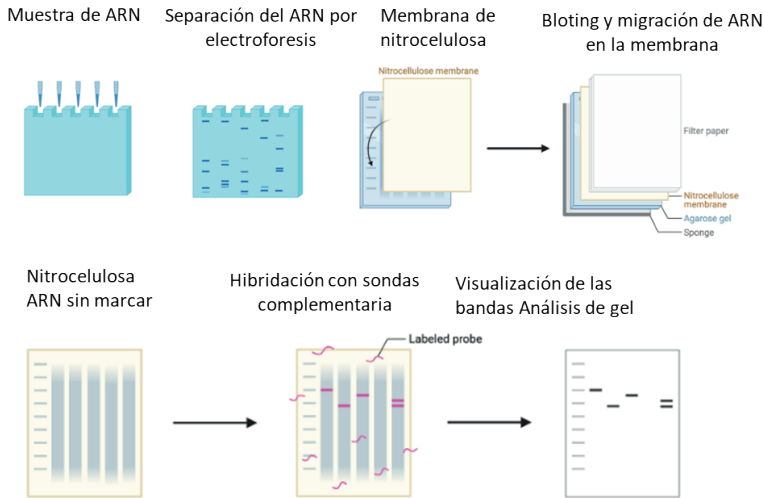
3.1.2. *NORTHERN BLOT*

Fundamento: es una técnica utilizada para detectar y analizar la presencia, cantidad y tamaño de ARN específico en una muestra biológica.

Esta técnica, desarrollada en la década de 1970, se utiliza para detectar y cuantificar la presencia de ARN específico en una muestra. Consiste en la separación de las moléculas de ARN por electroforesis en gel, la transferencia a una membrana y la hibridación con

sondas de ADN o ARN complementarias marcadas que se unen al ARN de interés (**figura 35**).

Es una técnica que permite la detección y cuantificación de ARN específicos en una muestra. Consiste en la separación de los ARN mediante electroforesis en gel de agarosa y su transferencia a una membrana. Luego, se realiza la hibridación con sondas de ARN específicas marcadas con radioisótopos o enzimas para detectarla (ITKONEN, 2023).

Figura 35. Northern blot

Nota. La figura ilustra la técnica de Northern Blot para detectar ARN. El ARN se separa en un gel, se transfiere a una membrana y se hibrida con una sonda marcada para visualizar ARN específico, proporcionando información sobre su tamaño y abundancia.

3.1.3. ARN FISH USANDO AMPLIFICACIÓN DE SEÑAL DE ADN RAMIFICADO

Fundamento: es una técnica que amplifica la señal de ARN específico en una muestra biológica. Utiliza sondas de ARN y sondas de ADN ramificado para detectar y amplificar la señal de ARN, mejorando la sensibilidad y la visualización del objetivo de ARN en la muestra.

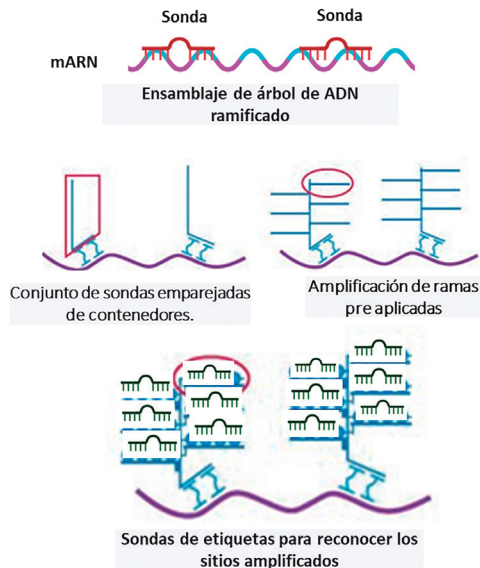
Las técnicas tradicionales de FISH utilizan secuencias de oligonucleótidos grandes marcadas con uno a cinco fluoróforos y están limitadas por un fondo alto y una sensibilidad baja debido a la unión no específica y la amplificación de señal insuficiente. Con la tecnología de ADN_b, un conjunto de sondas específicas del objetivo de aproximadamente 20 pares de oligonucleótidos cortos se hibrida con las regiones específicas en la secuencia de ARN objetivo de interés (ITKONEN, 2023).

Los pasos posteriores de amplificación de la señal requieren que cada par de oligonucleótidos se una al ARN objetivo en posiciones adyacentes, proporcionando especificidad al objetivo, así como un sitio de acoplamiento disponible para que otros oligonucleótidos se hibriden. Los pasos de hibridación adicionales finalmente dan como resultado un *árbol de hibridación* marcado con fluorescencia que proporciona detección de moléculas de ARN individuales mediante imágenes o citometría de flujo (ITKONEN, 2023).

Actualmente hay disponibles cuatro tipos de juegos de sondas fluorescentes para la detección de ARN de ViewARN Cell o PrimeFlow, utilizando varios colorantes Alexa Fluor con distintos espectros de emisión. Al multiplexar más de un objetivo de ARN en una sola muestra, cada conjunto de sondas debe ser de un tipo único para diferenciar su señal de las demás. Para los ensayos de tejido ViewARN ISH, solo hay dos tipos de conjuntos de sondas disponibles que son detectables por fluorescencia o quimioluminiscencia utilizando sustratos Fast Red o Fast Blue (ITKONEN, 2023).

La amplificación de la señal mediante la tecnología de ADN_b se logra a través de una serie de pasos de hibridación secuencial, que dan como resultado la formación de una estructura similar a un árbol. Las moléculas preamplificadoras se hibridan con su respectivo par de sondas de oligonucleótidos unidas para formar el tronco del *árbol*. Múltiples moléculas amplificadoras se hibridan con su respectivo preamplificador para crear las *ramas*. Finalmente, varias sondas de etiquetas se hibridan con los amplificadores para formar las *hojas* del árbol. Un árbol de amplificación de señal completamente ensamblado contiene 400 sitios de unión de sonda de etiqueta. Si todos los oligonucleótidos específicos del objetivo en un conjunto de sondas de 20 pares de oligonucleótidos se unen al transcrito de ARN objetivo, se puede lograr una amplificación de 8000 veces (**figura 36**).

Figura 36. Proceso de formación del ADN ramificado



Nota. La figura muestra el proceso de ARN FISH con amplificación de señal mediante ADN ramificado. Primero, se ensamblan árboles de ADN ramificado en la secuencia

objetivo de ARN mediante sondas específicas. Luego, se aplican sondas contenedoras emparejadas para construir ramas adicionales, amplificando así las señales en los sitios de unión. Finalmente, las sondas etiquetadas se unen a los sitios amplificados, lo que permite una señal más intensa y detectable en el análisis de hibridación *in situ*, mejorando la sensibilidad para la detección de ARN.

3.1.4. MICROARREGLOS DE ADN (ADN MICROARREGLOS)

Fundamento: también conocidos como chips de expresión génica, se basan en la hibridación de ARN para detectar y cuantificar la expresión de miles de genes simultáneamente.

Aunque también se utilizan en genómica, los microarreglos de ADN son una herramienta importante en transcriptómica. Permiten medir simultáneamente la expresión de miles de genes y evaluar los niveles de ARNm en una muestra. Los microarreglos contienen sondas de ADN que se hibridan con el ARNm presente en la muestra y se detectan mediante fluorescencia (APARNA y TETALA, 2023).

Las técnicas de microarreglos son una poderosa herramienta para estudiar el ARN y analizar simultáneamente la expresión de miles de genes en una muestra. Está basada en la hibridación y generalmente implican la incubación de ADNc marcado con fluorescencia en microarreglos personalizados o en oligomicroarreglos comerciales de alta densidad. Estas técnicas permiten detectar y cuantificar la abundancia relativa de ARN específicos en una muestra (ZHANG *et al.*, 2021).

En 1995, se reportaron métodos para cuantificar los niveles de expresión génica mediante un pequeño chip: el microarreglo. Por primera vez, fue posible analizar la expresión génica en todo el genoma. Sin embargo, esta técnica podría usarse solo para genes diana ya conocidos. Otra limitación de la micromatriz es la falta de información a nivel de exón. Esto condujo al desarrollo de micromatrices de exón, proporcionando así una visión más profunda de la biología subyacente (APARNA y TETALA, 2023). Algunas de las técnicas de microarreglos más utilizadas para el estudio del ARN se presentan a continuación.

3.1.4.1. Microarreglos de oligonucleótidos

En esta técnica, se imprimen miles de sondas de oligonucleótidos en un soporte sólido, como una diapositiva de vidrio o una matriz de silicio. Estas sondas están incluidas para ser complementarias a secuencias de ARN específicas. Luego, se extrae y etiqueta el ARN de interés de la muestra de estudio, y se hibrida con los oligonucleótidos en el microarreglo. La detección de la señal se realiza mediante la identificación de la etiqueta, que puede ser un fluorocromo o una enzima. La intensidad de la señal detectada proporciona información sobre la abundancia relativa del ARN en la muestra.

3.1.4.2. *Microarreglos de ADN de expresión génica*

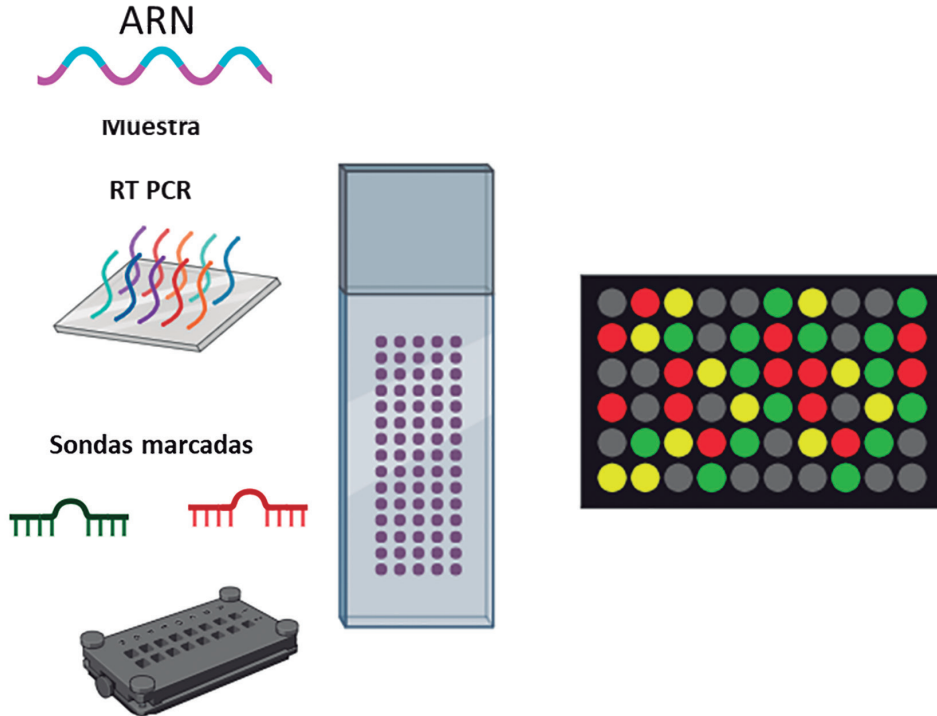
En esta técnica se utilizan microarreglos en los que se imprime una colección de fragmentos de ADN que representan genes específicos o regiones genómicas. Estos fragmentos de ADN actúan como sondas para hibridar con ARN de la muestra. El ARN extraído de la muestra se marca y se hibrida con los fragmentos de ADN en el microarreglo. La detección de la señal se realiza de manera similar a los microarreglos de oligonucleótidos. La intensidad de la señal detectada proporciona información sobre la expresión relativa de los genes en la muestra (ZHANG *et al.*, 2021).

Estas técnicas de microarreglos permiten el estudio de la expresión génica, la identificación de biomarcadores, la caracterización de enfermedades y la investigación de vías biológicas. Además, han sido ampliamente utilizados en la investigación genómica y en el descubrimiento de nuevos genes relacionados con enfermedades y procesos biológicos.

También, se han creado matrices especializadas, como aquellas con sondas que cubren uniones de exones, lo que permite detectar y cuantificar diferentes isoformas generadas mediante el proceso de empalme. Se han desarrollado microarreglos de mosaico genómico que presentan una alta densidad y permiten un mapeo de regiones transcritas con una resolución muy precisa, desde unos pocos pares de bases hasta aproximadamente 100 pares de bases. Los enfoques basados en la hibridación son altamente eficientes y relativamente económicos, a excepción de las matrices de mosaico de alta resolución que se utilizan para analizar genomas extensos (ZHANG *et al.*, 2021).

3.1.4.3. *Microarreglos de ADNc*

La metodología utilizada para detectar genes aumentados en un tejido se describe brevemente de la siguiente manera: en primer lugar, se extrae el ARN de la muestra. Dado que el ARN es altamente inestable y se degrada rápidamente, es crucial que los tejidos se mantengan frescos o sean congelados inmediatamente después de su obtención. A continuación, el ARN se convierte en ADN complementario (ADNc) utilizando una transcriptasa inversa. Durante este proceso, el ADNc se marca con un fluorocromo, que es una molécula capaz de emitir luz cuando se excita por un láser adecuado (**figura 37**).

Figura 37. *Microarreglos de ADNc*

Nota. La figura muestra el proceso de microarreglos de ADNc. Primero, se prepara ARN de muestra y se convierte en ADNc marcado con sondas de colores específicos. Luego, el ADNc se hibrida en una matriz de microarreglos, donde las sondas marcadas se unen a secuencias complementarias. Finalmente, la matriz se escanea para detectar las señales fluorescentes, con cada color indicando la presencia y el nivel de expresión de genes específicos en la muestra.

La sonda de ADNc marcada, que contiene una muestra de los genes expresados en el tejido original, se hibrida con las secuencias diana depositadas en el microarreglo. Estas secuencias diana son complementarias a las secuencias génicas expresadas. Sin embargo, debido a que la eficiencia del marcado puede variar entre diferentes genes y la cantidad de ADN diana puede diferir entre microarreglos, se suelen llevar a cabo experimentos de competición. Estos experimentos implican la comparación de dos muestras de ADN de tejidos diferentes en el mismo microarreglo, siendo cada muestra marcada con un fluorocromo distinto. Ambas muestras se mezclan y se hibridan simultáneamente en el microarreglo (ZHANG *et al.*, 2021).

Este diseño experimental presenta varias ventajas significativas. En primer lugar, al realizar las comparaciones entre las dos muestras en un mismo microarreglo, se reducen las variaciones y las posibles fuentes de error en comparación con las comparaciones entre muestras de diferentes microarreglos. Esto proporciona resultados más precisos y confiables. Además, este enfoque permite reducir a la mitad el número de microarreglos necesarios para realizar las comparaciones. Esto no solo ahorra tiempo y recursos, sino que también facilita la logística experimental al minimizar la manipulación de múltiples microarreglos (ZHANG *et al.*, 2021).

Una vez completada la hibridación, el microarreglo se somete a un escaneo utilizando un escáner láser, generando dos imágenes distintas, una para cada fluorocromo utilizado. Estas imágenes muestran puntos de luz cuyas intensidades varían según el nivel de hibridación alcanzado en cada diana del microarreglo. Posteriormente, estas imágenes se procesan mediante un *software* especializado que cuantifica la señal de cada punto (diana) para cada fluorocromo (muestra), generando una base de datos que se analizará utilizando técnicas estadísticas.

Frecuentemente, se crea una imagen superpuesta en forma de pseudocolor, donde una de las imágenes se representa en rojo y la otra en verde. De esta manera, los puntos que aparecen en color amarillo indican una señal alta en ambas muestras, en cantidades similares. Esta representación visual facilita la identificación rápida de regiones de interés que muestran cambios en la expresión génica entre las dos muestras analizadas (APARNA y TETALA, 2023).

En la actualidad, se reconoce que los experimentos basados en un único microarreglo tienen limitaciones en términos de validez. Esto se debe a la considerable variabilidad de los datos, la cual debe ser controlada mediante el uso de un número adecuado de réplicas y tamaños de muestra.

Los métodos de hibridación y microarreglos presentan diversas limitaciones que es importante tener en cuenta. En primer lugar, se basan en el conocimiento existente sobre la secuencia del genoma, lo que implica cierta dependencia de la información previa. Además, pueden presentar altos niveles de fondo debido a la hibridación cruzada, lo que puede interferir con la detección precisa de las señales de interés. Asimismo, el rango dinámico de detección está limitado tanto por el fondo como por la saturación de las señales (APARNA y TETALA, 2023).

Por otro lado, la comparación de niveles de expresión entre diferentes experimentos suele ser un desafío, y a menudo requiere métodos de normalización complejos para obtener resultados confiables y comparables. Es importante tener en cuenta estas limitaciones al interpretar y analizar los datos obtenidos mediante hibridación y microarreglos, y considerar enfoques complementarios y validaciones adicionales para obtener una imagen más completa y precisa de los perfiles de expresión génica.

3.2. ENFOQUES BASADOS EN SECUENCIAS

Los enfoques basados en secuencias proporcionan información directa sobre la secuencia de ADNc. En un principio, la secuenciación de Sanger se utilizó para secuenciar ADNc o bibliotecas de EST (secuencias expresadas). Estos métodos, basados en etiquetas, permitieron superar algunas limitaciones de los enfoques de hibridación, como la presencia y cantidad de ARN objetivo. Algunos ejemplos de estos métodos incluyen el análisis en serie de la expresión génica (SAGE), el análisis Cap de la expresión génica (CAGE) y la secuenciación masiva de firmas en paralelo (MPSS) (ZHANG *et al.*, 2021).

3.2.1. ANÁLISIS EN SERIE DE LA EXPRESIÓN GÉNICA (SAGE)

Fundamento: se basa en la generación de etiquetas únicas de secuencias cortas de ARN de la muestra, que luego se amplifican y secuencian.

Esta técnica es utilizada para obtener información sobre la expresión de genes en una muestra específica, permite obtener un perfil detallado de la expresión génica en una muestra determinada, incluyendo la identidad y la abundancia relativa de cada ARNm presente.

En primer lugar, se aísla el ARNm de la muestra de interés y se generan fragmentos de ADN complementarios (ADNc) utilizando cebadores biotinilados. Estos cebadores se unen al ARNm y generan fragmentos de ADNc complementarios basados en las secuencias de ARNm presentes. A continuación, los fragmentos de ADNc generados se amplifican mediante técnicas de PCR y se someten a secuenciación de alto rendimiento para obtener información sobre las secuencias de los genes expresados en la muestra. Los fragmentos de ADNc secuenciados contienen *Tags* que representan secuencias cortas de ADN que se originan de los ARNm presentes en la muestra. Estos *Tags* se analizan y se utilizan para identificar y cuantificar los genes expresados en la muestra.

Es importante tener en cuenta que esta técnica tiene algunos desafíos, como la identificación de *Tags* cortos y desconocidos de genes y su posible falta de especificidad, sin embargo, también tiene ventajas, como la capacidad de identificar genes desconocidos y proporcionar medidas cuantitativas de la expresión génica sin necesidad de información previa sobre las secuencias de los genes (ZHANG *et al.*, 2021).

3.2.2. ANÁLISIS CAP DE LA EXPRESIÓN GÉNICA (CAGE)

La técnica de análisis del extremo 5'(CAGE), permite obtener la población instantánea de ARNm en una muestra biológica. Esta técnica permite identificar los sitios de inicio de la transcripción y cuantificar la abundancia de los transcritos de ARN.

Fundamento: se basa en la secuencia llamada *cap* al extremo 5' de los ARN transcritos. Luego, se aíslan y secuencian los fragmentos de ARN marcados con el *cap*, lo que permite identificar los sitios de inicio de la transcripción y determinar los niveles de expresión génica.

Los resultados de CAGE consisten en un conjunto de secuencias nucleotídicas cortas (llamadas etiquetas) junto con su recuento de ocurrencias. Los fragmentos pequeños de ARNm, que tienen una longitud de 27 nucleótidos se extraen y se someten a una RT-PCR para obtener ADNc, se amplifican por

PCR y se secuencian (ZHANG *et al.*, 2021; APARNA y TETALA, 2023).

Los números de copias de las etiquetas de CAGE proporcionan una cuantificación digital de las abundancias de los transcritos de ARN en las muestras biológicas. Utilizando un genoma de referencia, se puede determinar, con cierta confianza, el ARNm original (y, por lo tanto, el gen) del cual se extrajo la etiqueta.

CAGE se utiliza principalmente para localizar los sitios exactos de inicio de la transcripción en el genoma, este conocimiento, a su vez, permite investigar la estructura del promotor necesaria para la expresión génica.

Las etiquetas de CAGE tienden a comenzar con una guanina adicional (G) que no está codificada en el genoma. Esta guanina adicional se atribuye a la extensión de 5' sin plantilla durante la síntesis de ADNc de cadena simple o a la reversión transcripción de la tapa misma. Cuando no se corrige, esto puede llevar a un mapeo erróneo de las etiquetas de CAGE, por ejemplo, a pseudogenes no transcritos. Sin embargo, la adición de estas guaninas también se ha utilizado como una señal para filtrar los picos de sitios de inicio de la transcripción más confiables (ZHANG *et al.*, 2021).

A lo largo de los años, se han desarrollado variantes y mejoras de la técnica original de CAGE. Algunas de estas variantes incluyen el uso de cebadores de reversión transcripción aleatorios, secuenciación de alto rendimiento y el uso de métodos alternativos para capturar los extremos 5' del ARN. Estas mejoras han ampliado las aplicaciones y la eficiencia de CAGE en el análisis de la expresión génica.

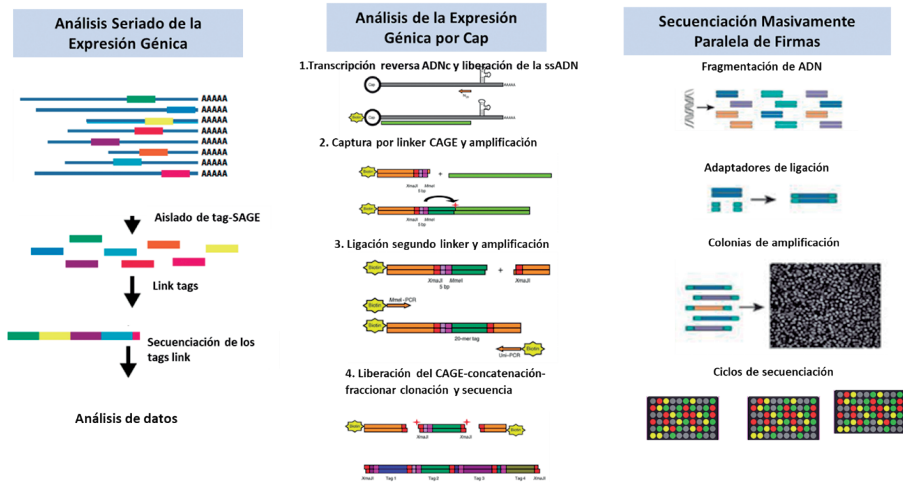
3.2.3. SECUENCIACIÓN MASIVA DE FIRMAS EN PARALELO (MPSS)

Fundamento: se basa en la generación de pequeños fragmentos de ARN conocidos como firmas, que se unen a oligonucleótidos en una matriz sólida. Luego, las firmas se amplifican y se someten a secuenciación masiva, lo que permite determinar la abundancia relativa de cada firma y, por lo tanto, de los genes correspondientes.

La técnica de secuenciación masiva de firmas en paralelo (MPSS, por sus siglas en inglés) es un método utilizado para identificar y cuantificar los transcritos de ARNm mediante la generación de secuencias de firma de 17-20 pb cerca del extremo 3' del sitio de restricción designado (ZHANG *et al.*, 2021; APARNA y TETALA, 2023).

En esta técnica se parte de una muestra de ARNm que se convierte en ADNc utilizando transcriptasa inversa. Los ADNc se fusionan con un pequeño oligonucleótido llamado *etiqueta* que permite amplificar los ADNc mediante PCR y luego unirlos a microperlas. Después de varias rondas de determinación de secuencias utilizando hibridación de sondas marcadas con fluorescencia, se determina una secuencia de firma de aproximadamente 16-20 pares de bases (pb) en cada perla. La imagen de fluorescencia captura la señal de todas las perlas, que están fijadas en una superficie bidimensional, lo que permite determinar las secuencias de ADN de todas las perlas en paralelo. En cada experimento se obtienen aproximadamente 1.000.000 de secuencias. El recuento de transcritos indica el nivel de expresión, y se calcula dividiendo el número de transcritos de un gen específico por el número total de transcritos presentes en el conjunto de datos (**figura 38**).

Figura 38. Transcriptómica: enfoques basados en secuenciación



Nota. La figura describe tres técnicas de transcriptómica basadas en secuenciación: RNA-Seq, que permite cuantificar la expresión génica global y detectar isoformas; scRNA-Seq, que analiza la expresión a nivel de célula única para estudiar la heterogeneidad celular; y Ribo-Seq, que secuencia solo ARN asociado a ribosomas para estudiar la traducción activa. Estas técnicas ofrecen una visión completa de la expresión y regulación génica en diferentes contextos.

La técnica MPSS tiene varias ventajas, como su alta sensibilidad, ya que puede detectar solo unas pocas moléculas de ARNm por célula, y su capacidad para generar conjuntos de datos en formato digital que facilitan el análisis y la gestión de los datos. Además, los conjuntos de datos MPSS son aditivos, lo que

significa que se pueden combinar los datos de múltiples análisis con la misma muestra inicial de ARNm. Sin embargo, también tiene algunas limitaciones, como la imposibilidad de asignar firmas a un pequeño porcentaje de genes y transcriptos, lo que impide medir sus niveles de expresión.

Estos enfoques de secuenciación basados en etiquetas son altamente eficientes y pueden proporcionar niveles de expresión génica digitales precisos. Sin embargo, la mayoría de ellos se basan en la tecnología de secuenciación de Sanger, lo cual conlleva algunas limitaciones. Una parte significativa de las etiquetas cortas generadas no puede ser mapeada de manera única en el genoma de referencia, lo que dificulta su asignación precisa. Además, estos enfoques suelen analizar solo una fracción de la transcripción, lo que implica que las isoformas no se pueden distinguir entre sí. Estas desventajas restringen el uso de la tecnología de secuenciación tradicional para una anotación exhaustiva de la estructura de los transcriptomas (APARNA y TETALA, 2023).

3.3. ARN-SEQ

Fundamento: se basa en la conversión del ARN en ADN complementario (cADN) y su posterior secuenciación masiva. En el proceso de ARN-seq, el ARN total o el ARN mensajero (ARNm) se extraen de la muestra y se convierte en cADN mediante la acción enzimática de la transcriptasa inversa.

En los últimos años, han surgido nuevos métodos de secuenciación de ADN de alto rendimiento que han abierto nuevas posibilidades en el mapeo y cuantificación de transcriptomas. Uno de estos métodos, conocido como ARN-Seq (secuenciación de ARN),

presenta claras ventajas en comparación con los enfoques existentes y se espera que revolucione la forma en que se analizan los transcriptomas en organismos eucariotas. El ARN-Seq ha sido aplicado con éxito en diferentes organismos, como *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, ratones y células humanas. Esta tecnología ha demostrado su potencial para obtener una visión más completa y detallada de los transcriptomas, brindando información invaluable sobre la expresión génica y la regulación de genes en diversos contextos biológicos (WAN *et al.*, 2009).

El ARN-Seq es un enfoque avanzado en el análisis de transcriptomas que utiliza tecnologías de secuenciación de alta profundidad. Este método ha revolucionado nuestra comprensión del alcance y la complejidad de los transcriptomas en organismos eucariotas. Al emplear el ARN-Seq, se logra una medición mucho más precisa de los niveles de expresión de los transcritos y sus isoformas, proporcionando una visión completa y detallada de la regulación génica. Como resultado, el ARN-Seq ha transformado la forma en que

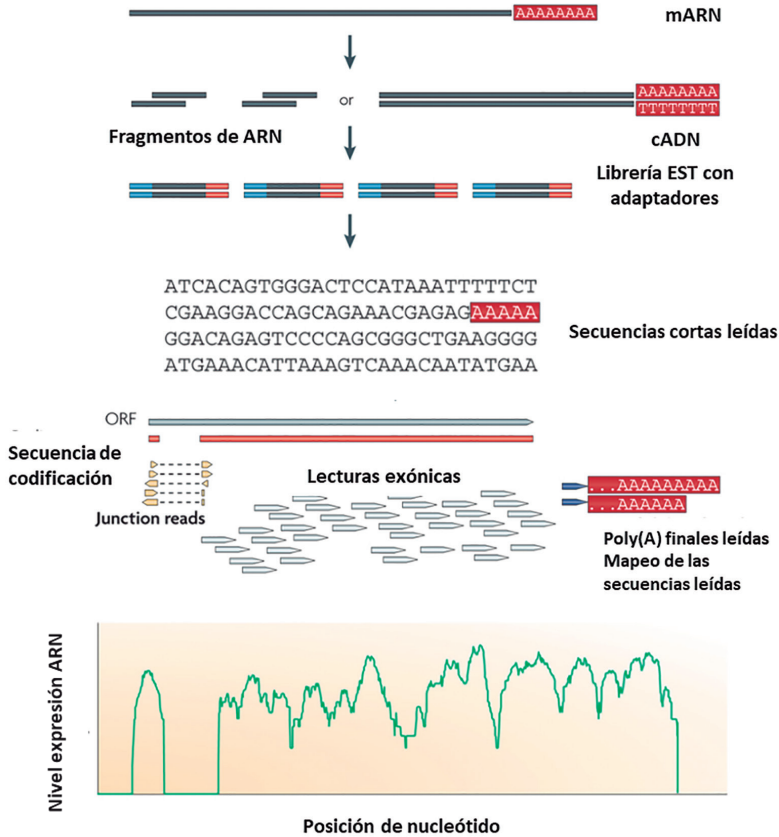
investigamos y comprendemos los transcriptomas, abriendo nuevas perspectivas en la investigación biológica y médica (KHOZYAINOVA *et al.*, 2023).

3.3.1. TECNOLOGÍA ARN-SEQ

ARN-Seq utiliza tecnologías de secuenciación profunda desarrolladas recientemente. En general, para realizar el ARN-Seq, una población de ARN, ya sea total o fraccionado (como poli(A)+), se convierte en una biblioteca de fragmentos de ADNc mediante la unión de adaptadores a uno o ambos extremos. Luego, cada molécula, con o sin amplificación, se somete a secuenciación de alto rendimiento para obtener secuencias cortas de uno o ambos extremos. La longitud de las lecturas suele variar entre 30 y 400 pb, dependiendo de la tecnología de secuenciación de ADN utilizada. Se pueden emplear diversas tecnologías de secuenciación de alto rendimiento, como Applied Biosystems SOLiD y Roche 454 Life Science, que ya se han utilizado en estudios de ARN-Seq. Además, el sistema tSMS de Helicos Biosciences, aunque no se ha utilizado en estudios de ARN-Seq publicados, también es adecuado y presenta la ventaja adicional de evitar la amplificación del ADNc objetivo (WAN *et al.*, 2009; KHOZYAINOVA *et al.*, 2023).

Después de la secuenciación, las lecturas obtenidas se alinean con un genoma de referencia o con transcripciones de referencia, o se ensamblan de novo sin la secuencia genómica para generar un mapa de transcripción a escala del genoma. Este mapa proporciona información tanto sobre la estructura transcripcional como sobre el nivel de expresión de cada gen, permitiendo un análisis completo de los transcriptomas (**figura 39**).

En síntesis, en el proceso de ARN-Seq, los ARN largos se convierten en una biblioteca de fragmentos de ADNc mediante fragmentación de ARN o fragmentación de ADN. Luego, se agregan adaptadores de secuenciación a cada fragmento de ADNc, y se realiza la secuenciación de alto rendimiento para obtener secuencias cortas de cada ADNc. Las secuencias resultantes se alinean con el genoma o transcriptoma de referencia y se clasifican en tres tipos: lecturas exónicas, lecturas de unión y lecturas finales poli(A). Estos tipos de lecturas se utilizan para generar un perfil de expresión de resolución de base para cada gen.

Figura 39. Ensayo de ARN-Seq

Nota. La figura muestra el proceso de RNA-Seq, donde el ARN mensajero se fragmenta, convierte en cADN, y se secuencian en pequeñas lecturas. Estas lecturas se alinean con el genoma para mapear regiones exónicas y extremos poli(A), generando un gráfico que indica la cobertura y niveles de expresión de los transcritos. Tomada y modificada de Wang et al. (2009).

Respecto a las *ventajas de ARN-Seq*, se puede decir que aunque el ARN-Seq es todavía una tecnología en desarrollo activo, ofrece diversas ventajas significativas en comparación con las tecnologías existentes (**tabla 15**). En primer lugar, a diferencia de los enfoques basados en hibridación, el ARN-Seq no se limita a detectar transcripciones que se correspondan con la secuencia genómica existente. Por ejemplo, se ha utilizado el ARN-Seq basado en 454 para secuenciar el transcriptoma de la mariposa fritillary de Glanville, lo cual lo

hace particularmente atractivo para el estudio de organismos no modelo cuyas secuencias genómicas aún no se han determinado. Además, el ARN-Seq puede revelar con precisión los límites de transcripción a nivel de una sola base, lo que brinda información detallada sobre la estructura de los transcritos (KHOZYAINOVA *et al.*, 2023).

Las lecturas cortas de 30 pb en el ARN-Seq proporcionan información sobre la conexión de dos exones, mientras que las lecturas más largas o las lecturas cortas en pares permiten revelar la conectividad entre varios exones. Estas características hacen que el ARN-Seq sea una herramienta invaluable para el estudio de transcriptomas complejos. Además, el ARN-Seq tiene la capacidad de revelar variaciones de secuencia, como polimorfismos de un solo nucleótido (SNP), en las regiones transcritas, lo que amplía su utilidad en la detección de variaciones genéticas. Estas técnicas presentan algunas ventajas (KHOZYAINOVA *et al.*, 2023) (tabla 15).

Tabla 15. *Ventajas clave sobre las tecnologías existentes*

Especificaciones tecnológicas	Microarreglo de mosaico	Secuenciación de ADNc o EST	ARN-seq
Principio	Hibridación	secuenciación de Sanger	Secuenciación de alto rendimiento
Resolución	De varios a 100 pb	Base única	Base única
Rendimiento	Alto	Bajo	Alto
Confianza en la secuencia genómica	Sí	No	En algunos casos
Ruido de fondo	Alto	Bajo	Bajo
Solicitud			
Mapear simultáneamente las regiones transcritas y la expresión génica	Sí	Limitado para la expresión génica.	Sí
Rango dinámico para cuantificar el nivel de expresión génica	Hasta unos cientos de veces	No práctico	>8,000 veces
Habilidad para distinguir diferentes isoformas	Limitado	Sí	Sí
Habilidad para distinguir la expresión alélica	Limitado	Sí	Sí

Temas prácticos			
Cantidad requerida de ARN	Alto	Alto	Bajo
Costo para mapear transcripciones de genomas grandes	Alto	Alto	Relativamente bajo

Tomando en consideración todas estas ventajas, el ARN-Seq se posiciona como el primer método basado en secuenciación que permite examinar de manera cuantitativa y con alto rendimiento todo el transcriptoma. Este enfoque ofrece una resolución a nivel de base única para la anotación de genes y niveles de expresión génica digitales a escala del genoma. Además, el ARN-Seq suele tener un costo significativamente más bajo en comparación con las matrices de mosaico o la secuenciación Sanger EST a gran escala (KHOZYAINOVA *et al.*, 2023).

El método óptimo para el análisis del transcriptoma debería ser capaz de identificar y cuantificar directamente todos los tipos de ARN, ya sean pequeños o grandes. Aunque el ARN-Seq ofrece ventajas significativas, es importante tener en cuenta que implica varias etapas de manipulación durante la generación de las bibliotecas de ADNc. Estas etapas pueden presentar desafíos y complicaciones en el perfilado exhaustivo de todos los tipos de transcritos. Sin embargo, a medida que la tecnología avanza, se espera que se superen estos obstáculos y que el ARN-Seq siga evolucionando como una herramienta cada vez más poderosa y confiable para el análisis completo del transcriptoma.

Otras ventajas de en comparación con los microarreglos de ADN. En primer lugar, el ARN-Seq tiene una señal de fondo extremadamente baja o incluso inexistente, debido a que las secuencias de ADN se pueden mapear de manera unívoca en regiones únicas del genoma. Esto contrasta con los microarreglos de ADN, que pueden presentar un fondo de señal más alto y ambiguo. Además, el ARN-Seq no tiene un límite superior para la cuantificación, ya que está correlacionado con el número de secuencias obtenidas. Esto resulta en un amplio rango dinámico de niveles de expresión que se pueden detectar. Por ejemplo, estudios han estimado un rango de más de 9.000 veces en *Saccharomyces cerevisiae* mediante el análisis de 16 millones de lecturas mapeadas, y un rango que abarca cinco órdenes de magnitud estimado para 40 millones de lecturas de secuenciación en ratón (KHOZYAINOVA *et al.*, 2023).

En contraste, los microarreglos de ADN tienen una menor sensibilidad para genes expresados en niveles bajos o muy altos, lo que limita su rango dinámico a cientos o pocas centenas. Además, el ARN-Seq ha demostrado ser altamente preciso en la cuantificación de los niveles de expresión, como se ha confirmado mediante técnicas como PCR cuantitativa (qPCR) y controles de ARN con

concentración conocida. Los resultados del ARN-Seq también han demostrado una alta reproducibilidad tanto en réplicas técnicas como biológicas. Por último, debido a que el ARN-Seq no requiere pasos de clonación y la tecnología Heliscope no implica amplificación, se requiere una menor cantidad de muestra de ARN en comparación con otros métodos.

El ARN-Seq se destaca por su baja señal de fondo, su amplio rango dinámico de detección, su precisión en la cuantificación de expresión génica y su alta reproducibilidad. Estas características hacen que el ARN-Seq sea una herramienta valiosa en la investigación transcriptómica, con beneficios significativos sobre los microarreglos de ADN y otros enfoques tradicionales (KHOZYAINOVA *et al.*, 2023).

Teniendo en cuenta todas estas ventajas, ARN-Seq es el primer método basado en la secuenciación que permite examinar todo el transcriptoma de una manera cuantitativa y de muy alto rendimiento. Este método ofrece una resolución de base única para la anotación y niveles de expresión génica *digitales* a escala del genoma, a menudo a un costo mucho menor que las matrices de mosaico o la secuenciación Sanger EST a gran escala (KHOZYAINOVA *et al.*, 2023).

Las ARN-Seq son ampliamente usadas y tiene claras ventajas sobre los métodos transcriptómicos desarrollados anteriormente. El próximo gran desafío para ARN-Seq es apuntar a transcriptomas más complejos para identificar y rastrear los cambios de expresión de isoformas de ARN raras de todos los genes. Las tecnologías que impulsarán el logro de este objetivo son la secuenciación de extremos de pares, la secuenciación específica de cadena y el uso de lecturas más largas para aumentar la cobertura y la profundidad. A medida que el costo de la secuenciación continúa cayendo, se espera que ARN-Seq reemplace a los microarreglos para muchas aplicaciones que involucran la determinación de la estructura y la dinámica del transcriptoma (KHOZYAINOVA *et al.*, 2023).

3.4. ETIQUETADO Y MULTIPLEXACIÓN DE UNA SOLA CÉLULA MEDIANTE SECUENCIACIÓN /SLAM-SEQ

Fundamento: es una técnica que combina la marca metabólica de ARN con la secuenciación de ARN para obtener perfiles de expresión génica a nivel de células individuales. Consiste en marcar el ARN recién sintetizado con nucleótidos modificados durante un período de tiempo definido, seguido de la secuenciación del ARN marcado para identificar los genes expresados.

La técnica SLAM-seq es una técnica novedosa que se utiliza para estudiar la dinámica de la transcripción a nivel de células individuales, permite analizar la síntesis de ARN en células individuales y rastrear la actividad transcripcional a lo largo del tiempo (GONDANE E ITKONEN, 2023).

La técnica se basa en la incorporación de nucleótidos etiquetados de forma no convencional en el ARN recién sintetizado en las células. Estos nucleótidos etiquetados contienen un grupo reactivo que permite la unión de una etiqueta de oligonucleótido específico. Luego, las células se someten a una lisis celular y el ARN total se extrae de las células.

A continuación, los fragmentos de ARN se someten a una reacción de transcripción inversa para generar ADNc a partir del ARN. Durante esta etapa, los oligonucleótidos etiquetados se unen al ADNc correspondiente a las transcripciones recién sintetizadas. Estas etiquetas permiten distinguir las transcripciones sintetizadas en diferentes momentos (KHOZYAINOVA *et al.*, 2023).

Después de etiquetar y generar el ADNc, se realiza una amplificación por PCR para aumentar la cantidad de ADNc y se prepara una biblioteca de secuenciación. La biblioteca se somete a una NGS para determinar qué transcripciones se sintetizaron en cada célula y en qué momento.

La secuenciación de la biblioteca permite identificar las secuencias de ARN que se producen en diferentes momentos y rastrear la dinámica de la transcripción a nivel de células individuales. Esto proporciona información valiosa sobre la actividad transcripcional en diferentes tipos de células y bajo diferentes condiciones experimentales (KHOZYAINOVA *et al.*, 2023).

Para finalizar, la técnica SLAM-seq combina la incorporación de nucleótidos etiquetados durante la síntesis de ARN, la generación de ADNc etiquetado y la NGS para rastrear la dinámica de la transcripción a nivel de células individuales. Esta técnica ha demostrado ser una herramienta poderosa para el estudio de la biología celular y la regulación génica.

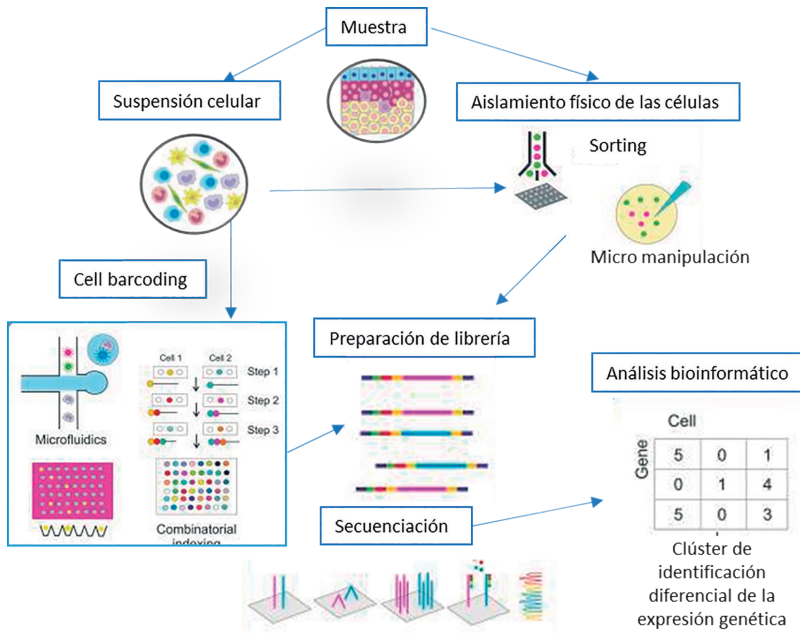
3.5. scARN-SEQ

Fundamento: Consiste en la captura y amplificación de los ARN de células individuales, seguido de la preparación de librerías y la secuenciación masiva.

La secuenciación de ARN unicelular (scARN-seq) es una herramienta revolucionaria para estudiar la fisiología de tejidos normales y patológicamente alterados. Este enfoque proporciona información sobre las características moleculares (expresión génica, mutaciones, accesibilidad a la cromatina, etc.) de las células, abre la posibilidad de analizar las trayectorias/filogenia de la diferenciación celular y las interacciones célula-célula, y ayuda a descubrir nuevos tipos de células y procesos previamente inexplorados (HEUMOS *et al.*, 2023; ALJANAHI *et al.*, 2018).

La secuenciación de ARN de una sola célula (scARN-seq) es una técnica verdaderamente revolucionaria que ha ampliado significativamente nuestra comprensión de la heterogeneidad y la dinámica de los transcriptomas en las células de varios organismos. Este método se utilizó por primera vez en 2009 para investigar blastómeros de ratón en la etapa de segunda división. Aunque se demostró que la secuenciación unicelular era superior a la tecnología de microarreglos en la evaluación cuantitativa de la expresión génica, la principal limitación de scARN-seq en ese momento era la imposibilidad de multiplexación, ya que las bibliotecas para células individuales tenían que crearse manualmente en tubos separados. Sin embargo, ya en 2011, se desarrolló el primer protocolo para el multiplex scARN-seq, y en 2014, se introdujo en el mercado la primera plataforma comercial para la preparación automatizada de bibliotecas unicelulares, Fluidigm C1. En la actualidad, existen varias plataformas disponibles para scARN-seq, incluidas Fluidigm C1/Smart-seq, BD Rhapsody (BD Biosciences, EE. UU.), Chromium (10x Genomics, EE. UU.), etc., que brindan la posibilidad de análisis de alto rendimiento (HEUMOS et al., 2023).

El flujo de trabajo para scARN-seq se presenta en **figura 40**. La suspensión celular se prepara mediante la homogeneización de la muestra, y las células se separan mediante métodos físicos, como clasificación celular o micromanipulación o mediante códigos de barras utilizando placas con oligonucleótidos o microfluídica y combinatoria. Las muestras de sangre y los cultivos celulares pueden someterse a clasificación o micromanipulación sin una preparación preliminar de la suspensión. Las células obtenidas se utilizan para la preparación de bibliotecas y posterior secuenciación, y los datos se procesan mediante técnicas bioinformáticas.

Figura 40. *scARN-seq*

Nota. La figura muestra el proceso de secuenciación de ARN de célula única. Primero, se obtiene una muestra y se prepara una suspensión celular. Las células se aíslan físicamente y se etiquetan mediante cell barcoding (codificación de células) para identificar cada célula individualmente. Luego, se utiliza microfluídica y técnicas de indexación combinatoria para crear una librería con códigos específicos. Esta librería se secuencia para identificar los perfiles de expresión génica de cada célula. Finalmente, el análisis bioinformático permite interpretar los datos, generando matrices de expresión que detallan el número de genes expresados en cada célula individual. Tomado y modificado de Heumos et al. (2023).

Los avances en las tecnologías *scARN-seq* han permitido caracterizar los principales mecanismos celulares y moleculares involucrados en el desarrollo de enfermedades cardiovasculares, neurodegenerativas, oncológicas y otras, para determinar la taxonomía celular de organismos modelo ampliamente utilizados como el berro (*Arabidopsis thaliana*), la mosca de la fruta (*Drosophila*) y el pez cebra (*Danio rerio*), y descifrar la heterogeneidad de las células del sistema inmunológico en estados normales y patológicos, en particular, enfermedades oncológicas (ALJANAHI et al., 2018).

Los estudios realizados al comienzo del desarrollo de la secuenciación unicelular se habían centrado principalmente en la investigación de las

poblaciones celulares de las muestras analizadas, el análisis de la expresión génica diferencial, el examen del ciclo celular y la identificación de tipos de células. Sin embargo, era obvio incluso en ese momento que las posibilidades de scARN-seq son mucho más amplias, y el desarrollo de enfoques bioinformáticos para integrar y transformar lecturas de nucleótidos en información multimodal que describe diferentes estados celulares era solo cuestión de tiempo (ALJANAHI *et al.*, 2018).

Esta técnica emergente permite analizar el transcriptoma de células individuales, lo que proporciona una visión más detallada y precisa de la expresión génica en diferentes tipos celulares y estados de desarrollo. Utiliza técnicas de NGS y metodologías de amplificación del ARN para estudiar el transcriptoma a nivel de una sola célula.

La scARN-seq ha revolucionado el estudio de la fisiología de tejidos normales y alterados patológicamente. Esta poderosa herramienta proporciona información detallada sobre las características moleculares de las células, como la expresión génica, las modificaciones y la accesibilidad a la cromatina. Además, permite analizar las trayectorias de diferenciación celular y las interacciones entre células, descubrir nuevos tipos celulares y explorar procesos previamente desconocidos (ALJANAHI *et al.*, 2018).

Desde una perspectiva clínica, el scARN-seq permite un análisis profundo de los mecanismos moleculares de las enfermedades, sentando las bases para el desarrollo de estrategias preventivas, diagnósticas y terapéuticas innovadoras

Esta revisión examina diversos enfoques para el análisis de datos scARN-seq, evalúa las ventajas y desventajas de las herramientas bioinformáticas disponibles, brinda recomendaciones y ejemplos de su uso exitoso, y sugiere posibles direcciones de mejora (KHOZYAINOVA *et al.*, 2023).

Asimismo, se subraya la importancia de desarrollar nuevos protocolos, como los multiómicos, para la preparación de bibliotecas de ADN/ARN de células individuales, con el fin de obtener una comprensión más completa de las mismas. Estos prometen una visión más profunda y enriquecedora de la biología celular, impulsando así el progreso en la investigación y la medicina (HEUMOS *et al.*, 2023).

Las siguientes tecnologías de secuenciación tienen ventajas y desventajas; seleccionar el enfoque correcto depende de la configuración experimental (**tabla 16**).

Tabla 16. *Ventajas y desventajas de las técnicas en transcriptómica*

Técnica	Descripción	Ventajas	Limitaciones
Secuencia de ARN estándar	Cuantifica los niveles de ARN de una muestra biológica en un momento dado.	No se limita a los genes conocidos. Permite el análisis de empalmes alternativos.	Se necesita una gran cantidad de material de partida (para la preparación típica de bibliotecas). Enmascara la heterogeneidad de la muestra (por ejemplo, en el caso de la biopsia).
Secuencia de ARN naciente	Técnicas basadas en análogos de nucleótidos utilizadas para evaluar la síntesis de ARN. Los ARN nacientes incorporan el análogo de nucleótido y se enriquecen utilizando técnicas basadas en afinidad o se decodifican computacionalmente.	Mide la transcripción activa. Causalidad directa entre el regulador y el objetivo. Permite la evaluación de la vida media del ARNm y de la degradación del ARNm.	Laborioso en comparación con ARN-seq. Seleccionar el tiempo de etiquetado adecuado es fundamental: si es demasiado corto, no proporciona una instantánea de la síntesis de ARNm; si es demasiado largo, no se puede distinguir el ARN naciente y el ARN total.
scARN-seq	Mide los niveles de expresión génica con resolución de una sola célula.	Permite evaluar la heterogeneidad de la muestra. Permite el seguimiento (potencial) de transiciones de destino celular/pasos de desarrollo.	Laborioso en comparación con ARN-seq. Conseguir que las células lleguen a una solución de célula única es un desafío y elegir la técnica correcta para hacerlo es crucial.

3.6. TÉCNICAS PARA CUANTIFICACIÓN DE LA TRANSCRIPCIÓN

La cuantificación de transcritos es el proceso de contar las coincidencias de los transcritos secuenciados con las secuencias de genes. Estos hits contados eventualmente llegan a la tabla de conteo. Más detalles sobre este proceso computacional se describirán más adelante. Hay dos enfoques principales para la cuantificación de la transcripción: de longitud completa y basada en etiquetas. Los protocolos de longitud completa intentan cubrir toda la transcripción de manera uniforme con lecturas de secuenciación, mientras que los protocolos basados en etiquetas solo capturan los extremos 5' o 3' (LOWE *et al.*, 2017). El método de cuantificación de la transcripción tiene fuertes implicaciones en los genes capturados y, por lo tanto, los analistas deben conocer el proceso de cuantificación utilizado.

3.6.1. *PROTOCOLOS DE SECUENCIACIÓN DE CÉLULAS INDIVIDUALES*

Actualmente, existen tres tipos de protocolos de secuenciación de una sola célula, que se agrupan principalmente por sus protocolos de aislamiento celular: estrategias basadas en dispositivos de microfluidos donde las células se encapsulan en gotas de hidrogel; protocolos basados en placa de pocillos donde las células se separan físicamente en pocillos y, por último, la solución comercial basada en un chip microfluídico (Fluidigm C1) que carga y separa las células en pequeñas cámaras de reacción. Estos tres enfoques difieren en su capacidad para recuperar transcritos, el número de células secuenciadas y muchos otros aspectos (LOWE *et al.*, 2017). En las siguientes subsecciones, analizaremos brevemente cómo funcionan, sus fortalezas y debilidades, y los posibles sesgos que los analistas de datos deben tener en cuenta con respecto a los protocolos respectivos (KHOZYAINOVA *et al.*, 2023).

3.6.2. *PROTOCOLOS BASADOS EN DISPOSITIVOS DE MICROFLUIDOS*

Las estrategias unicelulares basadas en dispositivos de microfluidos atrapan células dentro de gotas de hidrogel, lo que permite la compartimentación en cámaras de reacción unicelulares. Los protocolos más utilizados en Drop, Drop-seq y el 10x Genomics Chromium disponible comercialmente son capaces de generar tales gotitas varios miles de veces por segundo. Este proceso masivamente paralelo genera una gran cantidad de gotas a un costo relativamente bajo. Aunque los tres protocolos difieren en los detalles, las gotas del tamaño de un nanolitro que contienen células encapsuladas siempre están diseñadas para capturar perlas y células simultáneamente (LOWE *et al.*, 2017).

El proceso de encapsulación se lleva a cabo con microperlas especializadas con cebadores en perlas que contienen un identificador de PCR, un código de barras celular y un identificador molecular único (UMI, ver más abajo) de 4-8b pb de longitud y una cola poli-T. Tras la lisis, el ARNm de la célula se libera instantáneamente y es capturado por los oligonucleótidos con código de barras que se adjuntan a las perlas. A continuación, las gotas se recogen y se rompen para liberar transcriptomas unicelulares unidos a micropartículas (STAMP). A esto le sigue la PCR y la transcripción inversa para capturar y amplificar las transcripciones.

Finalmente, se lleva a cabo la etiquetación donde las transcripciones se cortan al azar y se adjuntan adaptadores de secuenciación. Este proceso da como resultado bibliotecas que están listas para la secuenciación como se describe anteriormente. En los protocolos basados en microfluidos, solo se recupera alrededor del 10 % de las transcripciones de la célula. En particular, esta secuenciación baja es suficiente para una identificación sólida de los tipos de células (LOWE *et al.*, 2017).

3.6.3. PROTOCOLOS BASADOS EN PLACA

Los protocolos basados en placas normalmente separan las células físicamente en placas de micropocillos. El primer paso implica la clasificación de células, por ejemplo, clasificación de células activadas por fluorescencia (FACS), donde las células se clasifican de acuerdo con marcadores de superficie celular específicos; o por micropipeteo. A continuación, las células seleccionadas se colocan en pocillos individuales que contienen tampones de lisis celular, donde posteriormente se lleva a cabo la transcripción inversa. Esto permite analizar varios cientos de células en un solo experimento con 5.000 a 10.000 genes capturados cada una. Los protocolos de secuenciación basados en placas incluyen, entre otros, SMART-seq2, MARS-seq, QUARTZ-seq y SRCB-seq (KHOZYAINOVA *et al.*, 2023).

En términos generales, los protocolos difieren en su capacidad de multiplexación. Por ejemplo, MARS-seq permite tres niveles de código de barras, a saber, etiquetas a nivel molecular, celular y de placa, para capacidades robustas de multiplexación. SMART-seq2, por el contrario, no permite la multiplexación temprana limitando el número de celdas. Una comparación sistemática de protocolos realizada por Mereu *et al.* en 2020 reveló que QUARTZ-seq2 puede capturar más genes que SMART-seq2, MARS-seq o SRCB-seq por célula, lo que significa que QUARTZ-seq2 es capaz de capturar bien los genes marcadores específicos del tipo de célula, lo que permite una anotación segura del tipo de célula (KHOZYAINOVA *et al.*, 2023).

Los avances recientes en las tecnologías unicelulares han permitido la elaboración de perfiles moleculares de células de alto rendimiento en todas las modalidades y ubicaciones. Los datos transcriptómicos de una sola célula ahora se pueden complementar con la accesibilidad de la cromatina, la expresión de proteínas de superficie, el perfilado del repertorio de receptores inmunitarios adaptativos y la información espacial. La creciente disponibilidad de datos de una sola célula en todas las modalidades ha motivado el desarrollo de nuevos métodos computacionales para ayudar a los analistas a obtener conocimientos biológicos. A medida que crece el campo, se vuelve cada vez más difícil navegar por el vasto panorama de herramientas y pasos de análisis (LOWE *et al.*, 2017; MOFFITT *et al.*, 2018).

3.7. TRANSCRIPTÓMICA ESPACIAL

Fundamento: métodos que preservan la ubicación de las moléculas de ARN en las células y tejidos, lo que proporciona información detallada sobre la localización y distribución de los transcritos.

El perfil transcriptómico espacial proporciona información genómica de células individuales, las cuales se encuentran intrincadamente organizadas en su entorno

tisular nativo. Aunque técnicas como la secuenciación de células individuales permiten conocer la composición celular, estas metodologías aíslan las células en gotitas y no preservan la estructura tisular, la cual es un componente esencial de los organismos biológicos (KLEINO *et al.*, 2022).

Los tejidos son el resultado de interacciones complejas entre las células, que se comunican de diversas formas, como a través de interacciones ligando-receptor, secreción de proteínas y moléculas pequeñas, e incluso fuerzas mecánicas. Por lo tanto, aunque la secuenciación de una sola célula puede proporcionar información sobre los tipos celulares presentes en un tejido, entender la composición celular de manera espacial directa es fundamental para comprender por qué se observan ciertos tipos celulares y cómo las variaciones en el estado celular pueden ser resultado del microambiente único presente en el tejido (ZAPPIA y THEIS, 2021).

El perfil espacial de las células en un tejido nos proporciona una valiosa herramienta para comprender la organización y la interacción de diferentes tipos celulares en su contexto tridimensional. Esta visión detallada nos permite obtener una comprensión más completa de la biología de los tejidos y nos ayuda a identificar células clave en procesos específicos. Además, el perfil espacial revela la distribución de las células en diferentes regiones del tejido y cómo su ubicación puede influir en su función y comportamiento (KLEINO *et al.*, 2022).

Al analizar el perfil espacial, podemos observar cómo las células se agrupan y organizan en estructuras específicas dentro del tejido, como folículos, criptas o capas celulares. Esto proporciona información crucial sobre cómo las células interactúan y se comunican entre sí en su entorno natural. Además, el perfil espacial nos ayuda a comprender cómo la proximidad a otras células o a componentes del microambiente puede influir en la expresión génica, la diferenciación celular y la función especializada.

La transcriptómica espacial es una herramienta poderosa que nos permite medir la complejidad de los sistemas biológicos a lo largo de los ejes relevantes para su función. Por un lado, la lectura transcriptómica altamente multiplexada nos revela la complejidad derivada de la gran cantidad de genes presentes en el genoma. Esto nos permite obtener información detallada sobre los diferentes transcritos y sus niveles de expresión en una muestra biológica (ZAPPIA y THEIS, 2021).

Por otro lado, la alta resolución espacial en la transcriptómica espacial permite capturar los sitios exactos donde se expresa cada transcrito dentro de un tejido o una estructura biológica, lo que resulta fundamental para comprender la organización espacial de las células y cómo se comunican entre sí en su microambiente. Al obtener información sobre la localización precisa de los transcritos, podemos identificar regiones específicas del tejido donde se produce la expresión génica y comprender cómo esto contribuye a la función y la biología del sistema.

En conjunto, la transcriptómica espacial combina la riqueza de la información transcriptómica con una resolución espacial fina, lo que permite obtener una visión completa y detallada de la expresión génica en el contexto espacial adecuado (ZAPPIA y THEIS, 2021). Esta técnica tiene aplicaciones importantes en la investigación biológica y médica, ya que ayuda a comprender cómo los genes se regulan y se expresan en diferentes tejidos y en respuesta a estímulos ambientales o patológicos (ZAPPIA y THEIS, 2021). La transcriptómica espacial brinda una perspectiva única para desentrañar la complejidad de los sistemas biológicos y avanzar en nuestro conocimiento de la biología y la salud.

La tecnología MERFISH ha evolucionado a partir de diversas técnicas FISH en los últimos años, permitiendo lograr avances significativos en la resolución espacial y la capacidad de detección de transcripciones de ARN objetivo. Un método destacado es el FISH de molécula única (smFISH), el cual utiliza múltiples sondas y un microscopio fluorescente para lograr una sensibilidad extrema en la detección y cuantificación de transcripciones de ARN específicas. Este enfoque poderoso nos permite no solo contar, sino también visualizar y ubicar con precisión los ácidos nucleicos de interés en células individuales (KLEINO *et al.*, 2022).

El smFISH ofrece una herramienta invaluable para estudiar la abundancia y la distribución espacial de genes específicos en células individuales. Aunque se limita a observar unos pocos genes a la vez, su capacidad de cuantificación directa y alta resolución espacial brinda información detallada sobre la expresión génica a nivel individual. Esto nos permite comprender mejor la variabilidad en la expresión génica entre células y cómo esto se relaciona con la función celular y los procesos biológicos (MOFFITT *et al.*, 2018).

Además de su capacidad para detectar transcripciones específicas, el smFISH también se puede combinar con otras técnicas para estudiar la colocalización de diferentes transcritos o proteínas en el mismo tejido o célula. Esta capacidad de multiplexación nos permite obtener información más completa y detallada sobre las interacciones y las redes moleculares dentro de las células.

MERFISH es una técnica altamente multiplexada y avanzada derivada del smFISH, que fue desarrollada en el 2015. Una característica destacada de MERFISH es su capacidad para utilizar códigos de barras combinatorios robustos a errores. Esto permite medir miles de especies de ARN en miles de células simultáneamente y con una alta precisión. A pesar de su capacidad multiplexada, esta técnica mantiene la sensibilidad y la naturaleza cuantitativa del smFISH mediante el uso de múltiples sondas por gen (MOFFITT *et al.*, 2018).

En los experimentos de MERFISH, se capturan imágenes de moléculas individuales y luego se procesan para revelar las variaciones en los transcriptomas

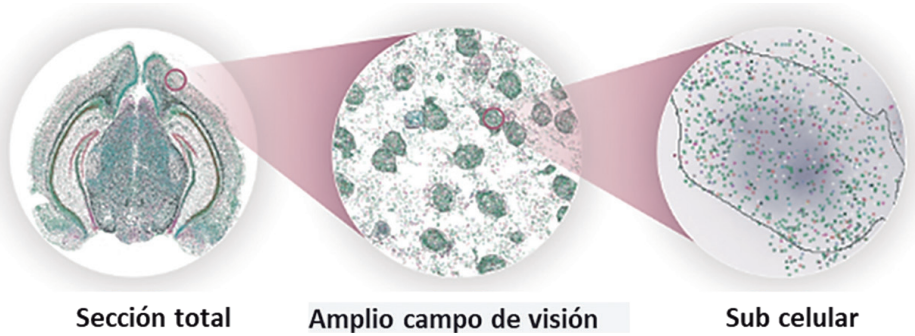
de las células en la muestra, preservando el contexto espacial en el análisis. Esto proporciona perfiles espaciales de alta resolución de la expresión de ARN en células individuales. El uso de esta técnica ha brindado información detallada sobre el tipo de célula, su estado, organización, interacciones y función en cultivos celulares y tejidos (KLEINO *et al.*, 2022).

La capacidad de obtener perfiles espaciales de expresión de ARN a nivel de células individuales ha abierto nuevas perspectivas en la investigación biológica y médica. MERFISH ha demostrado ser una herramienta poderosa para comprender la heterogeneidad celular, la dinámica de la expresión génica y las interacciones moleculares en el contexto espacial.

En el artículo publicado en Science en 2018, MOFFIT *et al.* demostraron la capacidad de MERFISH para crear un atlas espacial detallado de la composición celular y molecular en la región MPOA (preóptica medial) del cerebro de ratón. Los resultados obtenidos mediante este procedimiento fueron consistentes con los identificados previamente mediante la secuenciación de ARN de célula única (scRNA-seq), pero esta técnica permitió una diferenciación más precisa de los subtipos neuronales debido a su mayor eficiencia de detección. Este estudio también utilizó MERFISH en combinación con la identificación de genes tempranos inmediatos para comprender qué neuronas se activan durante diferentes comportamientos en esta región cerebral (KLEINO *et al.*, 2022).

Más recientemente, en un artículo publicado en 2020, muestran cómo varios grupos han expandido esta línea de investigación al realizar un atlas de la corteza motora primaria en combinación con el trazado retrógrado y el perfil de expresión del transcriptoma a lo largo de los axones y dendritas en neuronas cultivadas. Estas investigaciones proporcionan una visión detallada de la conectividad neuronal y las características de expresión génica en diferentes regiones del cerebro (KLEINO *et al.*, 2022).

Los estudios destacan el poder y la versatilidad de MERFISH como una herramienta para el mapeo espacial de tejidos y la caracterización de la heterogeneidad celular a nivel molecular. La combinación de MERFISH con otras técnicas complementarias permite obtener una comprensión más completa de la organización y la función de los circuitos neuronales en el cerebro. Estos avances en la transcriptómica espacial están impulsando nuestro conocimiento sobre la diversidad celular y las interacciones en el tejido cerebral, y tienen el potencial de abrir nuevas oportunidades en el campo de la neurociencia y la medicina (**figura 41**).

Figura 41. *Transcriptómica espacial*

Nota. La figura representa un análisis de transcriptómica espacial, en el que se visualizan perfiles de expresión génica dentro de un tejido completo. Comienza con una vista amplia de la estructura tisular y se va enfocando en áreas específicas, permitiendo observar la distribución espacial de los transcritos a nivel celular. Este enfoque facilita la comprensión de la organización espacial y la expresión génica en el contexto de la estructura del tejido. Tomado y modificado de Kleino et al. (2022).

4. TÉCNICAS UTILIZADAS EN PROTEÓMICA

La proteómica se describe al estudio completo de las proteínas presentes en una célula o un organismo en un momento determinado. Su objetivo es comprender la estructura, función, expresión y relaciones entre las proteínas que conforman el proteoma. En la **tabla 17** se muestran las principales tecnologías de alto rendimiento y las bases de datos utilizadas para su análisis en proteómica (DONG *et al.*, 2021).

Tabla 17. *Tecnologías en proteómica*

Datos arrojados	Tecnología de alto rendimiento	Bases de datos	Link	Usos	Referencias
Expresión de proteínas	Matriz de proteínas de fase inversa (RPPA) Cromatografía líquida Espectrometría de masas	HPA, PDB, Pfam, UniProt	RCSB PDB, Pfam, UniProt	Análisis de secuencias y estructuras de proteínas, identificación de biomarcadores, estudios de expresión.	(WILHELM <i>et al.</i> , 2014)

Datos arrojados	Tecnología de alto rendimiento	Bases de datos	Link	Usos	Referencias
Interacciones proteína-proteína	Espectrometría de masas acoplada a inmunoprecipitación (IP-MS), Detección de proximidad (PLA)	BioGRID, IntAct, STRING	BioGRID , IntAct , STRING	Mapeo de redes de interacción proteína-proteína, análisis de vías de señalización.	(SZKLARCZYK <i>et al.</i> , 2019), (Stark <i>et al.</i> , 2006)
Modificaciones postraduccionales	Espectrometría de masas (MS), Cromatografía líquida de alto rendimiento (HPLC)	PhosphoSite-Plus, PTM-code	PhosphoSite-Plus , PTM-code	Identificación de modificaciones como fosforilación, ubiquitinación y acetilación en proteínas.	(HORNBECK <i>et al.</i> , 2015)
Perfil de proteínas en tejidos	Inmunohistoquímica, Microarreglos de proteínas, RPPA	HPA, ProteomicsDB	HPA , ProteomicsDB	Identificación de proteínas específicas en tejidos y células, validación de biomarcadores.	(UHLEN <i>et al.</i> , 2015)

La proteómica también se centra en comprender el conjunto de interacciones entre proteínas que conforman la red de interacciones y que son fundamentales para el funcionamiento de los organismos vivos. En conclusión, la proteómica es el estudio del proteoma, es decir, el conjunto completo de proteínas presentes en un organismo o en una muestra biológica.

La proteómica fue definida por primera vez en 1995 como el contenido de proteínas complementarias a un genoma (WILKINS *et al.*, 1996), esta ciencia ómica complementa la genómica y proporciona una mejor comprensión biológica, fenotípica y funcional de toda la fisiología. Este concepto y nuevo campo de la ciencia comenzó después de que el genoma humano fuera casi completamente secuenciado y 24.000 genes codificados fueron informados (MORAES *et al.*, 2016; DONG *et al.*, 2021)

La información del proteoma trae nuevos desafíos en esta área de la ciencia, especialmente implica la medición e identificación de la cantidad de productos génicos, considerando la complejidad y diferentes procesos regulatorios. Además, el paso de expresión génica se puede modular desde la transcripción hasta la modificación postraduccionales de proteínas (PTM) (figura 42). Considerando este complejo mecanismo regulador, el desarrollo de una técnica que sea capaz de

identificar la totalidad proteoma en un solo análisis es muy desafiante (EBRAHIMI-NAJAFABADI *et al.*, 2014).

Figura 42. *Proteómica*



Para el estudio del proteoma, se utilizan técnicas clásicas como la electroforesis, la cromatografía, y la espectrometría de masas. Sin embargo, hay una técnica más innovadora que ha demostrado buenos resultados: los *chips de proteínas*, que son similares a los *chips de ADN*, que permiten analizar simultáneamente miles de proteínas. Además, gracias a la bioinformática, ahora se cuenta con programas que pueden identificar las reacciones metabólicas de una proteína específica con el resto del proteoma. Mediante la simulación por computadora, también es posible modelar diferentes formas mutantes de una proteína e incluso determinar cómo interactúa con sus sustratos correspondientes.

La proteómica busca establecer una conexión entre los genes y sus productos, las proteínas. Para lograrlo, se sigue el siguiente proceso: primero, se realiza la preparación de la muestra utilizando técnicas de electroforesis en 2D y cromatografía. A continuación, se lleva a cabo la determinación parcial de la secuencia de aminoácidos. Luego, se procede a la identificación y cuantificación de las proteínas utilizando técnicas de biocomputación, por último, se realiza el estudio de las interacciones proteicas, la estructura y la localización celular (**figura 42**).

Teniendo en cuenta lo emergente campos ómicos, ya que la proteómica abarca la cuantitativa, funcional, estructural y PTM caracterización de proteínas, nuevas áreas como la interatómica, hacen posible la determinación de las relaciones entre proteínas y aumentan el desarrollo de la biología de sistemas.

4.1. CROMATOGRAFÍA

Fundamento: se basa en la separación de una mezcla de componentes en función de sus diferentes afinidades por dos fases: una fase móvil y una fase estacionaria. La fase móvil es un solvente que fluye a través de la fase estacionaria, que puede ser un gel, un sólido poroso o un líquido inmovilizado en una matriz.

La cromatografía es una técnica que se remonta a finales del siglo XIX, cuando se realizaron los primeros experimentos utilizando una columna de sílice para separar pigmentos vegetales. Estos experimentos dieron origen a la cromatografía en columna, que posteriormente se desarrolló en diversas

técnicas cromatográficas, como la cromatografía en capa fina y la HPLC (OGISO *et al.*, 2008).

La cromatografía es un método analítico ampliamente utilizado en diversas disciplinas científicas para la separación de mezclas complejas. Se basa en el principio de la retención selectiva, que permite cuantificar, identificar y separar los componentes de una mezcla. La retención selectiva se logra al aprovechar el comportamiento diferencial de los componentes de la mezcla en relación con un soporte específico, como papel, gas, líquido o resina, y una fase líquida o gaseosa que fluye a través del soporte (HUANG *et al.*, 2023).

La cromatografía abarca diversas técnicas que se valen de las diferencias en la velocidad de retención de cada componente, lo que permite su separación, identificación y cuantificación. En muchos casos, la adsorción juega un papel importante, ya que las partículas son retenidas en una superficie. La separación y la posterior cuantificación o identificación de los componentes se logran gracias a las disparidades en las velocidades de adsorción y la afinidad de cada componente por el soporte utilizado (HUANG *et al.*, 2023). Es esencial comprender algunos conceptos clave para comprender el funcionamiento de las técnicas cromatográficas (**tabla 18**).

Tabla 18. *Definiciones claves en cromatografía*

Concepto	Definición
Fase estacionaria	En la cromatografía, la fase estacionaria se refiere a una sustancia inmóvil que se encuentra fija en una posición durante el proceso de separación.
Fase móvil	La fase móvil, ya sea líquida o gaseosa, es la sustancia que se desplaza durante la cromatografía. La muestra, que contiene los componentes a analizar, se introduce en la fase móvil.
Analitos	Son las sustancias que se separan, cuantifican o identifican mediante cromatografía.

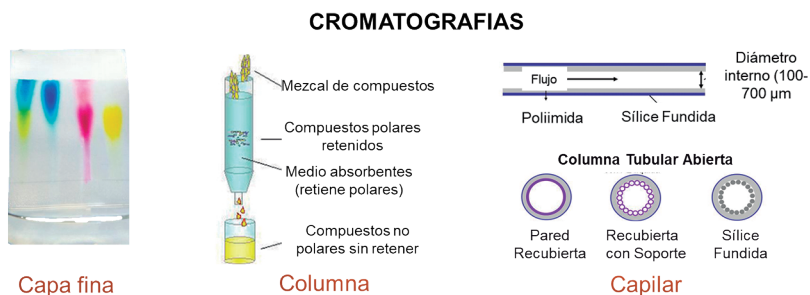
Muestra	Es la mezcla que se analiza, compuesta por uno o varios analitos y otros componentes que pueden no ser de interés y que se separan de los analitos.
Tiempo de retención	El tiempo de retención es el intervalo de tiempo que tarda un analito en pasar desde la columna o sistema donde fluye la fase móvil hasta el detector. Durante este tiempo, el detector registra una señal de detección utilizando alguna propiedad específica del analito.
Cromatograma	Es una representación visual que permite visualizar y analizar los resultados de la separación cromatográfica, y los picos presentes en el cromatograma indican la presencia y las características de los componentes individuales de la mezcla separada.

La cromatografía se utiliza tanto en aplicaciones preparativas de separación, para obtener componentes más puros y utilizables posteriormente, como en aplicaciones analíticas, donde se mide la proporción de los componentes en una mezcla. Estos dos tipos no son excluyentes y se pueden utilizar de manera complementaria en diferentes contextos científicos. Dependiendo de la tecnología utilizada, la naturaleza del soporte (fase estacionaria) y la sustancia móvil (fase móvil), es posible distinguir varios tipos de cromatografías, a continuación, se describen las más utilizadas en los laboratorios en general (HUANG *et al.*, 2023).

4.1.1. CROMATOGRFÍA EN PAPEL

En la década de 1940, Roger Williams utilizó la cromatografía en papel para sugerir la presencia de patrones metabólicos característicos en la orina y saliva asociados con enfermedades como la esquizofrenia. Este fue uno de los primeros intentos de medir los perfiles metabólicos de manera cuantitativa (OGISO *et al.*, 2008).

La fase estacionaria se compone de una tira de papel filtro, donde se deposita la muestra en un extremo mediante pequeñas gotas de la solución y se procede a evaporar el disolvente. A continuación, se utiliza el disolvente como fase móvil y se permite que ascienda por capilaridad a lo largo de la tira de papel. Para completar el proceso, se coloca la tira de papel de forma vertical, con la muestra ubicada en la parte inferior, dentro de un recipiente que contiene fase móvil en el fondo (**figura 43**).

Figura 43. Cromatografías

Nota. La figura muestra tres tipos de cromatografía (capa fina, columna y capilar) para separar compuestos según sus afinidades con fases estacionarias y móviles.

4.1.2. CROMATOGRAFÍA EN CAPA FINA

La cromatografía en capa fina (TLC, por sus siglas en inglés *Thin-Layer Chromatography*) es una técnica de separación ampliamente utilizada para analizar y separar compuestos químicos en una muestra. En esta técnica, se aplica una capa fina de fase estacionaria sobre una placa de vidrio, aluminio o plástico, que actúa como soporte (OGISO *et al.*, 2008).

4.1.3. CROMATOGRAFÍA EN COLUMNA

En la cromatografía de columna, la fase estacionaria se sitúa dentro de una columna, que es utilizada como fase estacionaria y puede estar fabricada con diversos materiales, como vidrio o acero inoxidable. La fase móvil, por otro lado, puede ser líquida o gaseosa (OGISO *et al.*, 2008). En la cromatografía en columna, la muestra se coloca en el extremo superior de la columna y se permite que descienda a través de la columna con la ayuda de la gravedad y la fase móvil.

De esta manera, la cromatografía en columna se clasifica como un método en el cual la separación y el análisis de los componentes se llevan a cabo mediante el proceso de elución en una columna. Este tipo de cromatografía es ampliamente utilizado en diversos campos científicos y permite separar los componentes de una muestra en función de sus diferentes afinidades con la fase estacionaria y la fase móvil.

4.1.4. CROMATOGRAFÍA DE GASES

Es una técnica de separación basada en la volatilidad y la afinidad de los compuestos por una fase estacionaria sólida o líquida. El proceso exacto de la

cromatografía de gases (CG) puede variar según el tipo de cromatógrafo de gases utilizado, la columna cromatográfica y el detector seleccionados, así como la naturaleza de la muestra a analizar (HUANG *et al.*, 2023).

4.1.5. CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN

Utiliza una fase móvil líquida que fluye a través de una fase estacionaria de alta resolución. En el proceso de CG, se realiza la preparación del equipo, que incluye el sistema de inyección de muestra, la columna cromatográfica, el horno de control de temperatura, el detector y el sistema de registro de datos (TANG *et al.*, 2016). Luego, se prepara la muestra de manera adecuada, llevando a cabo la extracción y purificación de los compuestos de interés. Posteriormente, se realiza la inyección de la muestra en el sistema de CG, utilizando técnicas como la inyección divididas o separadas (*Split*), en donde se divide en una pequeña fracción que se introduce en la columna, mientras que el resto se desecha; o la inyección sin división inicial (*splitless*) donde la muestra se introduce en la columna sin división inicial, lo que permite una mayor sensibilidad para detectar componentes minoritarios (HAGGARTY Y BURGESS, 2017).

A continuación, la muestra se volatiliza y se transporta a través de la columna cromatográfica, donde ocurre la separación de los componentes debido a las interacciones con la fase estacionaria. A medida que los compuestos pasan por el detector, se generan señales eléctricas o cambios físicos que se registran y se utilizan para generar un cromatograma, que muestra la intensidad de la señal en función del tiempo de retención (HUANG *et al.*, 2023). Por último, se analizan los datos obtenidos para identificar y cuantificar los compuestos presentes en la muestra, ya sea comparando los tiempos de retención con compuestos de referencia conocidos o utilizando técnicas de espectrometría de masas para una identificación precisa.

4.2. ELECTROFORESIS

Fundamento: radica en la migración diferencial de partículas cargadas eléctricamente bajo la influencia de un campo eléctrico, lo cual permite la separación y el análisis de diversas muestras, incluyendo proteínas.

La electroforesis es una técnica que se basa en la migración de partículas cargadas eléctricamente en un medio conductor bajo la influencia de un campo eléctrico aplicado. En el caso de la electroforesis de proteínas, por ejemplo,

se aprovecha la carga eléctrica neta que poseen las proteínas para separarlas y analizarlas en función de sus propiedades eléctricas (HOLLAND y CASTO-BOGGESS, 2023).

Cuando se aplica un campo eléctrico a un gel o a un medio acuoso en el que se encuentran las proteínas, estas se desplazan a diferentes velocidades según su carga eléctrica. Las proteínas cargadas negativamente migran hacia el ánodo (electrodo positivo), mientras que las proteínas cargadas positivamente se dirigen hacia el cátodo (electrodo negativo).

La velocidad de migración de las proteínas está determinada por su carga, su tamaño y la naturaleza del medio en el que se realiza la electroforesis. Mediante la aplicación de técnicas específicas, como la electroforesis en gel de poliacrilamida o la CE, es posible separar las proteínas en función de su tamaño, carga o incluso interacciones específicas con otros componentes del medio (FILEP y GUTTMAN, 2021).

4.2.1. *EN GEL DE POLIACRILAMIDA (SDS-PAGE)*

Esta técnica clásica, desarrollada en la década de 1960, se utiliza para separar proteínas en función de su tamaño molecular mediante la migración en un gel. Las proteínas se desnaturalizan con SDS (dodecil sulfato de sodio) y se cargan en un gel de poliacrilamida, donde migran hacia el ánodo bajo la influencia de un campo eléctrico (FILEP y GUTTMAN, 2021). Esto permite separar las proteínas según su tamaño (Swerdlow y Gesteland, 1990), después de la separación, se pueden analizar las bandas de proteínas resultantes.

4.2.2. *ELECTROFORESIS CAPILAR*

La EC es similar al de la electroforesis convencional, pero se lleva a cabo en un capilar estrecho y de longitud extendida, generalmente de sílice fundido. Se basa en la migración diferencial de las partículas cargadas eléctricamente, como iones o moléculas biológicas, son separadas en función de su movilidad electroforética a través del capilar, lo que permite la separación y el análisis de diversas muestras de forma eficiente y precisa (GARCÍA *et al.*, 2017). En los estudios peptidómicos o proteómicos, la CE es una técnica alternativa para la cromatografía líquida. Es un método de separación altamente eficiente y rápido que requiere cantidades extremadamente bajas de muestra.

El principio básico de la EC se basa en aplicar un campo eléctrico a lo largo del capilar lleno de un electrolito. Cuando se aplica el campo eléctrico, las especies cargadas en la muestra migran a diferentes velocidades según su carga y tamaño, las especies más pequeñas y cargadas se mueven más rápidamente que las más grandes y menos cargadas (MONTÓN *et al.*, 2005).

Además, durante la EC, también se produce otro fenómeno llamado migración electrosmótica. Este fenómeno se refiere al movimiento del líquido electrolítico

en el capilar debido a la interacción entre el campo eléctrico aplicado y las cargas en la superficie interna del capilar. La migración electrosmótica ayuda a arrastrar las especies cargadas a través del capilar y contribuye a la separación eficiente (GARCÍA *et al.*, 2017).

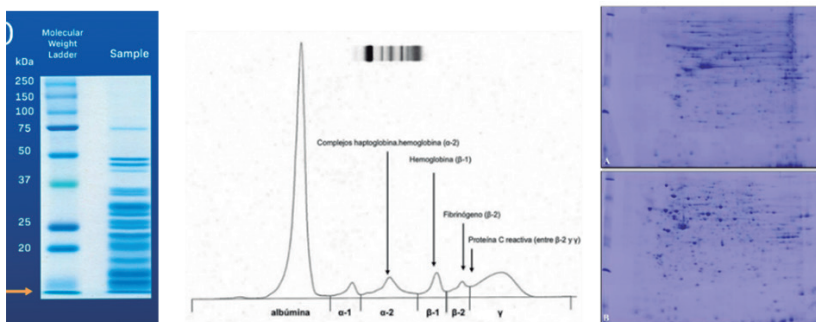
La separación de las especies se logra mediante la detección de las señales eléctricas generadas cuando las diferentes especies pasan por un punto de detección, generalmente al final del capilar. Estas señales son registradas y analizadas para obtener información sobre la composición y concentración de las especies presentes en la muestra.

4.2.3. ELECTROFORESIS BIDIMENSIONAL

La electroforesis bidimensional (2D) es una técnica ampliamente empleada en el análisis de proteínas. Su objetivo principal es separar las proteínas de una mezcla en dos dimensiones en función de sus propiedades moleculares. La separación se realiza en dos etapas: en la primera dimensión, las proteínas se separan según su carga eléctrica utilizando el isoelectroenfoco, mientras que en la segunda dimensión se separan según su peso molecular mediante la electroforesis en poliacrilamida-SDS (WEISS y GÖRG, 2009).

No obstante, uno de los desafíos más significativos de la electroforesis 2D es el análisis y la comparación de mezclas de proteínas. En la actualidad se cuenta con bases de datos de geles bidimensionales que permiten la comparación de patrones (WEISS y GÖRG, 2009). Estos sistemas facilitan la comparación automática de las manchas (*spots*) presentes en los geles, lo cual posibilita un análisis cuantitativo de las proteínas objeto de estudio (figura 44).

Figura 44. Electroforesis



Nota. La figura muestra técnicas de electroforesis para separar proteínas según su tamaño y características. Incluye geles de poliacrilamida, perfiles de proteínas séricas y electroforesis bidimensional, permitiendo un análisis detallado del perfil proteico en estudios de proteómica y diagnóstico.

4.3. WESTERN BLOTTING (INMUNOTRANSFERENCIA)

Fundamento: Se basa en la capacidad de los anticuerpos para reconocer y unirse a proteínas específicas en una mezcla compleja.

También conocido como inmunotransferencia o *immunoblot*. Desarrollada en la década de 1970, esta técnica se utiliza para detectar y analizar proteínas específicas en una muestra. Después de la separación de proteínas por SDS-PAGE, se transfieren a una membrana de nitrocelulosa y se detectan mediante anticuerpos específicos que se unen a la proteína de interés. Los anticuerpos se detectan utilizando métodos de marcaje con enzimas o fluoróforos para su detección (MISHRA *et al.*, 2017).

El *Western blot* comienza con la extracción y separación de las proteínas, generalmente mediante electroforesis en gel. Luego, las proteínas se transfieren a una membrana y se bloquean los sitios no específicos de unión. Después del bloqueo, se incuban la membrana con anticuerpos primarios y luego con anticuerpos secundarios conjugados con enzimas o marcadores fluorescentes. Tras lavados y revelado, se obtienen bandas que representan la presencia de la proteína de interés (MISHRA *et al.*, 2017). Finalmente, se analizan y cuantifican los resultados comparándolos con controles y estándares conocidos para obtener información sobre la proteína detectada.

4.4. ESPECTROMETRÍA

Fundamento: se basa en la interacción entre la radiación electromagnética y la materia para obtener información sobre las propiedades de las sustancias.

La espectrometría es una técnica que surgió en la década de 1980 analítica utilizada para el análisis cualitativo y cuantitativo de sustancias químicas, es el método más ampliamente empleado para estudios proteómicos, peptidómicos y metabolómicos. Se emplea para evaluar el espectro de composición de una muestra y proporcionar información sobre los componentes presentes y su concentración. En la proteómica, la espectrometría de masas se utiliza para identificar, caracterizar y cuantificar proteínas en una muestra. La técnica se emplea para determinar la composición de proteínas, identificar modificaciones postraduccionales y caracterizar interacciones proteína-proteína (HAAG, 2016).

En la espectrometría de masas, las proteínas se ionizan y se separan según su relación masa-carga (m/z). Luego, se detectan y analizan los iones generados para identificar las proteínas presentes.

4.4.1. ESPECTRÓMETRO DE MASAS

El espectrómetro de masas desempeña un papel fundamental en las determinaciones cualitativas y cuantitativas de elementos, así como en la medición del contenido isotópico en compuestos. Este instrumento está específicamente diseñado para separar átomos de distintos elementos o diferentes isótopos de un mismo elemento. Mediante la relación masa/carga (m/e), el espectrómetro de masas permite el aislamiento y análisis de varios elementos presentes en una muestra. De esta manera, es posible identificar y cuantificar los componentes individuales de una mezcla y determinar la composición isotópica de los elementos presentes (HAAG, 2016). Durante la operación de un espectrómetro de masas (**figura 45**), se llevan a cabo los siguientes procesos principales:

Vaporización: el primer y fundamental paso consiste en vaporizar la mezcla de muestra que se va a analizar. Esto se logra mediante la aplicación de una presión atmosférica extremadamente baja a la muestra (HAAG, 2016).

Ionización: en el siguiente paso, las partículas de la muestra se convierten en iones. Este proceso puede realizarse mediante diversos métodos, como el uso de haces de electrones de alta energía o rayos X (HAAG, 2016).

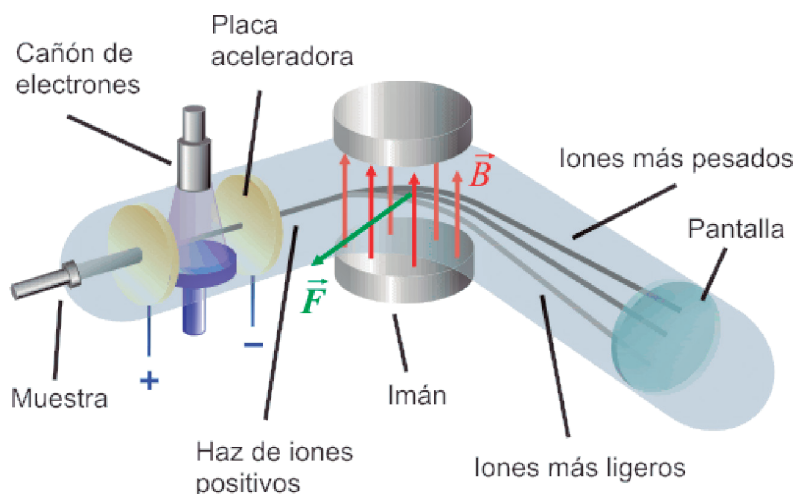
Aceleración de iones: es un proceso en el cual los iones son convertidos en un haz unidireccional y luego se aceleran eléctricamente hacia un campo magnético. Durante este proceso, los iones cargados eléctricamente son manipulados y dirigidos para que sigan una trayectoria específica. Primero, los iones son producidos y extraídos de una fuente ionizante, luego se enfocan y se enfocan en un haz estrecho utilizando campos eléctricos (HAAG, 2016).

Separación de iones: a continuación, se aplican campos eléctricos de aceleración para aumentar la velocidad de los iones y dirigirlos hacia una región donde se encuentra un campo magnético. Este campo magnético desvía los iones en función de su relación masa/carga, lo que permite separar y analizar diferentes especies iónicas. La aceleración de iones es un proceso fundamental en la espectrometría de masas y se utiliza en una variedad de aplicaciones científicas y tecnológicas (HAAG, 2016).

Captura de iones y generación de corriente eléctrica: después de que los iones son separados, estos alcanzan el detector de iones. Al interactuar con las paredes metálicas del detector, los electrones de las capas externas de los átomos metálicos se excitan y llenan los vacíos de electrones en las capas externas de los iones de muestra, neutralizándolos. Como resultado, se genera una corriente eléctrica en el detector de iones metálicos debido al flujo de electrones cargados negativamente. La intensidad de la corriente varía en función del número de iones que se aproximan al detector (HAAG, 2016).

Formación del espectro: la corriente eléctrica generada por los distintos iones se puede medir en función de sus relaciones masa/carga (m/e). Estos datos se utilizan para crear una representación gráfica del análisis composicional de la muestra, mostrando un espectro característico (**figura 45**).

Figura 45. *Espectrómetro de masas*



Nota. La figura muestra un espectrómetro de masas, donde una muestra se ioniza, los iones son acelerados y desviados en un campo magnético según su masa/carga, y se detectan en una pantalla para analizar su composición y masa. Elaboración propia Tomado y modificado de Haag (2016).

Existen diferentes tipos de espectrómetros clasificados según la técnica de ionización utilizada. En proteómica se utilizan los espectrómetros de masa con diversas modificaciones que han permitido el avance de esta ciencia ómica, las más comunes utilizadas se presentan a continuación.

4.4.2. *ESPECTROMETRÍA DE MASAS BASADA EN TIEMPO DE VUELO (TOF-MS)*

Esta técnica mide el tiempo que tardan los iones en viajar desde la fuente de ionización hasta el detector. Los iones se aceleran mediante una diferencia de potencial y su tiempo de vuelo se correlaciona con su masa. TOF-MS es conocida por su alta resolución y capacidad para analizar una amplia gama de masas.

En los años 70, se introdujo el uso del láser en la espectrometría de masas para analizar la desorción de iones. Sin embargo, se observaron espectros de baja

intensidad que contenían fragmentos de las moléculas analizadas. Fue entonces cuando se descubrió que, al mezclar estas moléculas con una matriz de naturaleza orgánica, la luz del láser era absorbida de manera más eficiente, lo que resultaba en espectros de mayor intensidad con una fragmentación mínima. Este enfoque se conoce como MALDI-MS (Espectrometría de Masas por Desorción/Ionización Asistida por Matriz) y permite la detección de moléculas en un amplio rango de masas moleculares (VESTAL y CAMPBELL, 2005).

Tanto la ionización MALDI-MS como la ionización ESI (*Electrospray*) se consideran técnicas de *ionización suave* (*soft-ionization*) debido a que permiten la ionización de biomoléculas, como péptidos y proteínas, sin provocar su ruptura durante el proceso. En el mercado existen varios analizadores acoplados a esta fuente de ionización, si bien el analizador más común es el TOF (tiempo de vuelo).

La muestra se mezcla en exceso con la matriz sobre una superficie de metal de manera que ambas cristalizan juntas durante la evaporación del solvente. Esta preparación se expone a breves pulsos láser en un ambiente de alto vacío, lo que hace que la matriz absorba energía y la transfiera a la muestra, resultando en excitación y transferencia de iones H⁺ (ionización). Esto generalmente produce especies con carga única. El área irradiada, que tiene unas pocas micras de tamaño, se calienta y provoca la desorción de los iones de la fase sólida a la fase gaseosa (VESTAL y CAMPBELL, 2005).

El analizador de tiempo de vuelo (TOF) es típicamente el instrumento que se utiliza en conjunto con la fuente MALDI con mayor frecuencia. Para determinar la masa en un entorno de alto vacío, se lleva a cabo una medición precisa del tiempo transcurrido desde que los iones son acelerados en la fuente hasta que impactan en el detector.

La ionización MALDI se considera un método de *ionización suave* debido a su capacidad para detectar moléculas termolábiles, como las proteínas, de forma intacta. Sin embargo, durante el proceso de aceleración o mientras viajan a través del tubo, puede ocurrir un fenómeno de descomposición metastable conocido como PSD (decaimiento posfuente). El análisis de los iones generados por este fenómeno proporciona información estructural valiosa sobre la molécula original, pero requiere la separación de estos fragmentos (VESTAL Y CAMPBELL, 2005).

En un analizador lineal, esto no es posible, ya que los iones formados por PSD tienen la misma velocidad que el ion original y viajan juntos hasta el detector. Para separarlos, se utilizan analizadores tipo reflector que operan con voltajes variables, esto permite obtener un espectro de fragmentación-PSD único compuesto por múltiples espectros correspondientes a los diferentes voltajes utilizados, que luego se fusionan mediante *software* informático (VESTAL y CAMPBELL, 2005).

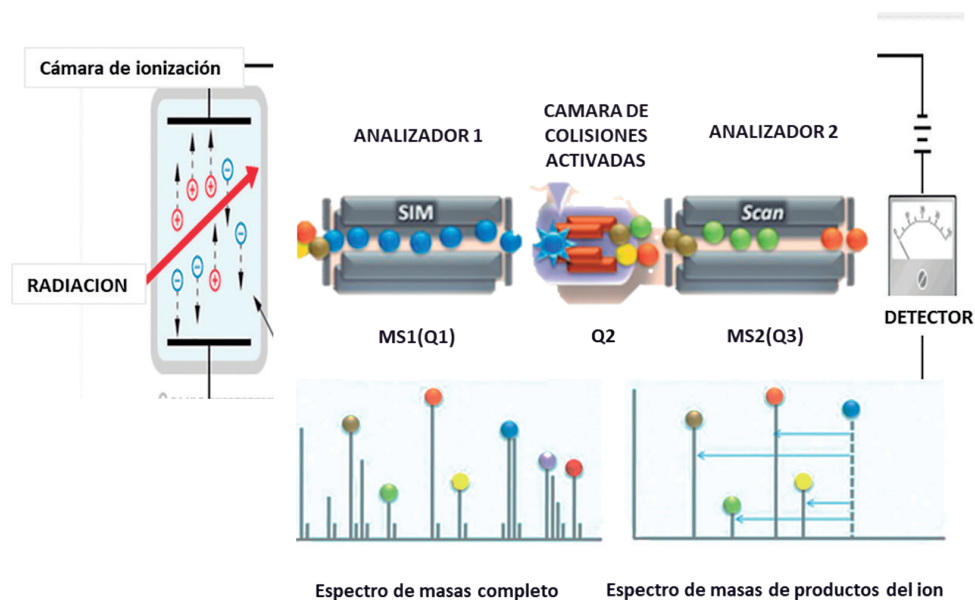
4.4.3. ESPECTROMETRÍA DE MASAS EN TÁNDEM (MS/MS o MS2)

Esta técnica consiste en dos etapas de espectrometría de masas. En la primera etapa, se selecciona un precursor iónico específico y se fragmenta en fragmentos iónicos más pequeños. Luego, estos fragmentos se analizan en la segunda etapa, proporcionando información adicional sobre la secuencia de aminoácidos y la estructura de la proteína.

La espectrometría de masas en tándem, también conocida como MS/MS o MS2, es una técnica de análisis instrumental en la que se acoplan dos o más analizadores de masas utilizando un paso adicional de reacción para aumentar su capacidad de analizar muestras químicas (LUO *et al.*, 2023). A continuación, se presenta de una manera breve cómo se realiza la técnica de espectrometría de masas en tándem.

Primero se realiza una ionización de la muestra para convertirlas en iones. Esto se puede lograr utilizando diferentes técnicas de ionización, como *electrospray* (ESI) o ionización por impacto electrónico (EI). En el primer analizador de masa, los iones se separan según su relación masa/carga (m/z). Este analizador de masa actúa como un filtro y selecciona los iones de interés para su análisis posterior. Los iones seleccionados en el MS1 se someten a una etapa de fragmentación. Esto se logra mediante técnicas como la disociación inducida por colisión (CID), la reacción ion-molécula o la fotodisociación. La fragmentación produce fragmentos iónicos más pequeños y específicos (LUO *et al.*, 2023).

Los fragmentos iónicos generados en la etapa de fragmentación se introducen en el segundo analizador de masa (MS2). Este analizador de masa separa y detecta los fragmentos iónicos en función de su relación m/z . La detección de los fragmentos iónicos proporciona información sobre la composición y estructura de las moléculas presentes en la muestra. Los datos obtenidos de la espectrometría de masas en tándem se analizan utilizando *software* especializado (**figura 46**).

Figura 46. *Espectrometría de masas en tándem (MS/MS o MS2)*

Nota. La figura muestra la espectrometría de masas en tándem (MS/MS), donde las moléculas se ionizan, los iones se seleccionan y fragmentan en dos etapas, y los fragmentos resultantes se analizan para obtener información detallada sobre la estructura y composición de las moléculas.

En el modo de operación de registro de iones-producto (*Product Ion Scan*) del triple cuadrupolo, se lleva a cabo un proceso de filtrado y fragmentación de los iones de interés. En el primer analizador MS1 (Q1), que opera en modo SIM (*Selected Ion Monitoring*), se seleccionan y filtran los iones de interés específicos. Estos iones se dirigen luego a la cámara de colisiones activadas (Q2 o q), donde son activados y fragmentados en iones-producto (LUO *et al.*, 2023).

Los iones-producto resultantes se analizan en el segundo analizador MS2 (Q3), que opera en modo *full scan*, realizando un barrido de masas (m/z) menor que el del ion precursor. Esto significa que se escanean y registran las masas menores a la masa del ion precursor, lo que permite detectar y analizar los iones fragmentados resultantes de la colisión en la cámara de colisiones activadas.

Es importante destacar que existen diferentes configuraciones de instrumentos de espectrometría de masas en tándem, como los espectrómetros de masas en tándem de cuadrupolo-trampa de iones (QqQ), los espectrómetros de masas en tándem de tiempo de vuelo (TOF) y los espectrómetros de masas en tándem de

cuadrupolo-TOF (Q-TOF), entre otros. Cada configuración tiene sus propias ventajas y aplicaciones específicas (LUO *et al.*, 2023).

4.4.4. *ESPECTROMETRÍA DE MASAS DE RELACIÓN ISOTÓPICA (IRMS)*

Esta técnica se utiliza para medir la relación de isótopos estables en proteínas. Al etiquetar las proteínas con isótopos estables (por ejemplo, con nitrógeno-15 o carbono-13), se pueden realizar comparaciones cuantitativas de la abundancia relativa de diferentes proteínas.

4.4.5. *ESPECTROMETRÍA DE MASAS EN TÁNDEM DE CAPTURA DE ELECTRONES (ECD-MS/MS)*

Esta técnica está basada en la captura de electrones por parte de iones precursor y la posterior fragmentación. Se utiliza para analizar modificaciones postraduccionales, como fosforilaciones y glicosilaciones, ya que la captura de electrones puede preservar estas modificaciones durante la fragmentación (LUO *et al.*, 2023).

4.4.6. *CROMATOGRAFÍA LÍQUIDA ACOPLADA A ESPECTROMETRÍA DE MASAS (LC-MS)*

Esta técnica combina la separación cromatográfica de proteínas o péptidos con la espectrometría de masas. Se utiliza para analizar mezclas complejas de proteínas y permite la identificación y cuantificación de proteínas en una muestra. La LC-MS se ha convertido en una herramienta esencial en proteómica debido a su alta sensibilidad y capacidad para analizar grandes conjuntos de datos (LUO *et al.*, 2023).

4.4.7. *ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN (HR-MS)*

Esta técnica, desarrollada a fines de la década de 1990, permite una mayor resolución y precisión en la determinación de masas moleculares y secuencias de proteínas. La HR-MS se basa en la capacidad de detectar diferencias en las masas atómicas con mayor precisión y se utiliza en combinación con otras técnicas de espectrometría de masas para análisis más detallados (LUO *et al.*, 2023).

4.5. PROTEÓMICA CUANTITATIVA

La proteómica cuantitativa busca determinar los cambios relacionados con la abundancia de proteínas dentro de un organismo, estos cambios están relacionados

con procesos biológicos y enfermedades, y que la identificación y cuantificación precisas de estas proteínas pueden proporcionar información valiosa sobre los mecanismos subyacentes.

4.5.1. *PROTEÓMICA CUANTITATIVA BASADA EN ETIQUETADO*

Esta técnica permite cuantificar las diferencias en la expresión de proteínas entre muestras. Se utilizan diferentes métodos de etiquetado, como ICAT (*Isotope-Coded Affinity Tags*), iTRAQ (*Isobaric Tags for Relative and Absolute Quantitation*) y TMT (*Tandem Mass Tagging*), que permiten marcar proteínas con etiquetas isotópicas y luego comparar las intensidades de las señales para determinar las diferencias de expresión. Esto permite la cuantificación relativa o absoluta de proteínas en diferentes condiciones.

4.5.2. *PROTEÓMICA SIN ETIQUETADO (LABEL-FREE)*

Esta técnica, utilizada a partir de la década de 2000, se basa en la comparación directa de las intensidades. En esta técnica, se utiliza la espectrometría de masas para comparar las intensidades de las señales de las proteínas en diferentes muestras sin el uso de etiquetas isotópicas. Permite la identificación y cuantificación relativa de proteínas en diferentes condiciones sin la necesidad de un marcado previo.

4.6. MÉTODOS PEPTIDÓMICOS

Los péptidos juegan un papel crucial en muchas funciones vitales importantes de los organismos vivos. El objetivo de la peptidómica es la identificación de los *peptidomas*, el contenido total de péptidos de un célula, órgano, tejido, líquido corporal u organismo. Las técnicas utilizadas en la peptidómica se basan en la extracción, digestión y análisis de los péptidos presentes en una muestra mediante técnicas como la cromatografía líquida acoplada a la espectrometría de masas.

4.6.1. *ELECTROFORESIS CAPILAR Y ESPECTROMETRÍA DE MASAS (CE-MS)*

La CE-MS ha sido empleada en numerosos estudios relacionados con la determinación de péptidos biomarcadores en diferentes fluidos corporales para varias enfermedades o en la investigación para el análisis de peptidómica de alimentos, identificación de péptidos con actividades biológicas especiales (NESBITT *et al.*, 2008).

Los métodos peptidómicos basados en MS son capaces de detectar alto número de péptidos en un solo experimento e identificar sus PTM. En muestras biológicas, los niveles relativos de distintos péptidos pueden variar en varios órdenes de magnitud. Los péptidos de baja abundancia, en particular, no suelen estar presentes en concentraciones suficientemente altas como para generar señales confiables en el análisis MS/MS necesario para su identificación. Además, los péptidos más abundantes pueden suprimir las señales de los menos abundantes, dificultando aún más su detección. Por esta razón, con frecuencia es necesario aplicar técnicas de pre-separación, como la cromatografía líquida (LC) o la electroforesis capilar (CE), antes de la ionización por electrospray (ESI)-MS o MALDI -MS (NESBITT *et al.*, 2008).

Comúnmente, el sistema CE está acoplado en línea a ESI fuente de iones, y fuera de línea para MALDI-MS. La ventaja del régimen fuera de línea es una mejor tolerancia a las sales y la posibilidad de almacenar muestras separadas durante mucho tiempo. En los enfoques de micropéptidos, la CE se combina comúnmente con la detección espectrometría de masas (MS). También se combina con la ionización por electropulverización MS y, con menor frecuencia, con la desorción/ionización láser asistida por matriz EM (NESBITT *et al.*, 2008).

En contraste con el análisis de la expresión del ARNm, la proteómica indica la existencia real, en lugar de la potencial, de estados funcionales de un sistema biológico. El cuello de botella en proteómica es que no hay un paso de amplificación, como una amplificación de PCR para ADN. La baja abundancia y alta gama de proteínas en muestras biológicas, así como el tiempo de adquisición y análisis de datos, siguen siendo un desafío en esta área, por lo tanto, el enfoque proteómico impulsa el desarrollo continuo de análisis, técnicas y herramientas bioinformáticas con el objetivo de profundizar la comprensión de las funciones biológicas.

La proteómica y la metabolómica son ciencias ómicas complementarias y han contribuido significativamente al avance del conocimiento en biología y biomedicina, especialmente en el estudio de enfermedades y el descubrimiento de biomarcadores (NESBITT *et al.*, 2008).

5. TÉCNICAS EN METABOLÓMICA

La metabolómica es una disciplina reciente que se centra en el estudio sistemático de los metabolitos presentes en una muestra biológica. Los metabolitos son moléculas pequeñas que están involucradas en los procesos químicos y metabólicos de una célula o un organismo (FIEHN, 2008). La metabolómica busca analizar y comprender por medio de diferentes tecnologías los perfiles de metabolitos presentes en la muestra analizada, lo que proporciona información sobre el estado fisiológico, las rutas metabólicas y los cambios asociados con

enfermedades o condiciones específicas (BAUERMEISTER *et al.*, 2022). El campo de la metabolómica está en constante desarrollo, y se están implementando nuevas técnicas y enfoques para mejorar la detección y el análisis de metabolitos (tabla 19).

Tabla 19. *Tecnologías de alto rendimiento utilizadas en metabolómica*

Datos arrojados	Tecnología de alto rendimiento	Bases de datos	Link	Usos	Referencias
Expresión de metabolitos	Espectrometría de masas (MS), Cromatografía líquida (LC), Cromatografía de gases (GC), Cromatografía capilar (CE), HRMS, Espectroscopia de RMN	H M D B , KEGG, METLIN, LipidMaps	H M D B , K E G G , METLIN, LipidMaps	Determinación y cuantificación de metabolitos en rutas metabólicas y perfiles en tejidos y fluidos	(WISHART <i>et al.</i> , 2018), (SMITH <i>et al.</i> , 2005)
Metabolitos de lípidos	Espectrometría de masas acoplada a cromatografía líquida (LC-MS), Espectrometría de masas de alta resolución (HRMS)	LipidMaps, LIPID Metabolites and Pathways Strategy (LIPID MAPS)	LipidMaps	Análisis de perfiles lipídicos, identificación de lípidos en rutas metabólicas específicas	(FAHY <i>et al.</i> , 2005)
Metabolómica de plantas	Espectroscopia de RMN, Cromatografía líquida de alta resolución acoplada a espectrometría de masas (LC-HRMS)	PlantCyc, KEGG	PlantCyc, KEGG	Análisis de metabolitos específicos en plantas, estudio de rutas metabólicas vegetales	(WECKWERTH, 2011)
Metabolitos de carbohidratos	Cromatografía de gases acoplada a espectrometría de masas (GC-MS), Espectroscopia de RMN	H M D B , KEGG	H M D B , KEGG	Perfilación y análisis de metabolitos de carbohidratos, estudio de glucólisis y gluconeogénesis	(BENNETT <i>et al.</i> , 2008)
Perfilación metabólica en humanos	LC-MS, GC-MS, RMN	H M D B , K E G G , METLIN	H M D B , METLIN	Identificación de biomarcadores en enfermedades, análisis de perfiles metabólicos en sangre y orina	(NICHOLSON <i>et al.</i> , 2002)

Nota. Esta tabla ofrece una visión general de las tecnologías de alto rendimiento en metabolómica, las bases de datos asociadas y sus aplicaciones.

Un enfoque integrado en el trabajo metabolómico implica llevar a cabo una serie de pasos que incluyen la preparación, normalización y calibración de las muestras, así como métodos de separación, detección de metabolitos y análisis de datos. Debido a la complejidad de las mezclas que se presentan en las muestras utilizadas en análisis metabolómicos (el plasma, la orina, la saliva, los tejidos y las células), es necesario utilizar métodos de separación para simplificarlas, ya que resulta difícil analizar simultáneamente una amplia gama de metabolitos con diferentes propiedades fisicoquímicas presentes en estas muestras.

5.1. MÉTODOS DE SEPARACIÓN Y ESPECTROMETRÍAS

Los métodos de separación comúnmente utilizados son los mismos utilizados en proteómica: la GC, la cromatografía líquida de alto rendimiento (HPLC) y la electroforesis capilar (CE). Estos métodos de separación se combinan a menudo con técnicas de espectrometría de masas para la detección, como la GC-MS o la LC-MS.

En la metabolómica, la espectrometría de masas se utiliza para analizar y cuantificar pequeñas moléculas, como metabolitos, en una muestra biológica. Se puede emplear para el perfilado metabolómico, la identificación de biomarcadores, la elucidación de vías metabólicas y el estudio de las interacciones entre metabolitos y proteínas.

5.1.1. CROMATOGRAFÍA DE GASES-ESPECTROMETRÍA DE MASAS (GC-MS)

En los años 60 y 70, se desarrolló la técnica de GC-MS, que permitía medir compuestos presentes en la orina y tejidos humanos. Fue utilizada por Horning y sus colaboradores, quienes acuñaron el término *perfil metabólico*. Es una técnica analítica que combina la separación de los metabolitos mediante cromatografía de gases y su detección e identificación mediante espectrometría de masas. Permite la identificación y cuantificación de una amplia gama de metabolitos volátiles y semivolátiles (PACCHIAROTTA *et al.*, 2013). El éxito del análisis de moléculas pequeñas, especialmente en metabolómica, mediante GC/MS depende de la capacidad de identificar químicamente, mediante la anotación de la composición elemental y la caracterización estructural, los compuestos en picos desconocidos (FIEHN, 2008).

La técnica GC-MS es ampliamente empleada para analizar metabolitos primarios. Los equipos disponibles son altamente eficientes y permiten el análisis de grandes volúmenes de muestras. Además, los avances recientes han ampliado la gama de compuestos analizables y se utilizan algoritmos y bases de datos mejoradas para obtener información relevante desde el punto de vista biológico. La mayoría

de los informes sobre perfiles de metabolitos mediante GC-MS se enfocan en aplicaciones biológicas, abarcan desde el análisis exhaustivo de compuestos volátiles hasta la evaluación de flujos metabólicos para la ingeniería biológica.

Los desarrollos en la metodología se centran en mejorar los procedimientos de extracción, evaluar la calidad de los datos obtenidos y explorar diferentes enfoques para el procesamiento de la información (FIEHN, 2008). No obstante, aún existen desafíos pendientes, como la identificación rápida de picos desconocidos y la integración de conocimientos biológicos para una mejor interpretación de los resultados.

5.1.2. *ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN*

La HRMS se desarrolló y se aplicó en metabolómica para mejorar la resolución y precisión en la determinación de la masa de los metabolitos. Esto permitió una identificación más precisa de los metabolitos y un mejor entendimiento de las vías metabólicas. Es una técnica que proporciona una mayor resolución y precisión en la determinación de la masa de los metabolitos.

5.1.3. *ESPECTROMETRÍA DE MASAS BASADA EN IONIZACIÓN POR ELECTROSPRAY (ESI-MS)*

Esta técnica se introdujo y se utilizó ampliamente en metabolómica para analizar metabolitos en fase líquida en solución, y obtener información sobre su estructura y composición. Permite la identificación y cuantificación de una amplia gama de metabolitos, incluyendo aquellos de baja polaridad.

5.1.4. *ESPECTROMETRÍA DE MASAS BASADA EN IONIZACIÓN POR DESORCIÓN LÁSER ASISTIDA POR MATRIZ (MALDI-MS)*

La MALDI-MS es una técnica que se utiliza para analizar metabolitos en estado sólido o en matriz. Permite la identificación y cuantificación de metabolitos en muestras biológicas complejas. Se utiliza en metabolómica para la identificación y cuantificación de metabolitos de masa alta mediante la ionización de una matriz absorbida por los analitos.

5.1.5. *ESPECTROMETRÍA DE MASAS DE REACCIÓN EN CADENA (MS/MS)*

Esta técnica se utiliza para la fragmentación y análisis secundario de metabolitos ionizados en espectrometría de masas. Proporciona información adicional sobre la estructura de los metabolitos.

5.1.6. ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN EN TÁNDEM (HRMS/MS)

Esta técnica combina la espectrometría de masas de alta resolución con la fragmentación controlada de los metabolitos. Permite la identificación y caracterización estructural más precisa de los metabolitos.

Además, existen otras técnicas analíticas que se utilizan habitualmente en la detección de metabolitos, como la RMN, la espectroscopia infrarroja por transformada de FOURIER (FT-IR) y la espectroscopia Raman. Por último, el análisis de los datos metabolómicos requiere el uso de herramientas y *software* sofisticados para realizar una identificación y cuantificación rigurosas de los compuestos, así como para una interpretación precisa de los datos. A continuación, se mencionan algunas de las técnicas más utilizadas actualmente tanto en separación como análisis en metabolómica.

5.2. RESONANCIA MAGNÉTICA NUCLEAR (RMN)

Fundamento: la RMN se basa en las propiedades magnéticas de los núcleos atómicos, que interactúan con un campo magnético externo, generando diferentes niveles de energía.

La espectroscopia se encarga de analizar la cantidad de luz absorbida, dispersada (reflejada) o emitida por un objeto. Su objetivo es descomponer la luz y medir las diversas longitudes de onda, tanto visibles como invisibles. Asimismo, la espectroscopia encuentra aplicaciones en

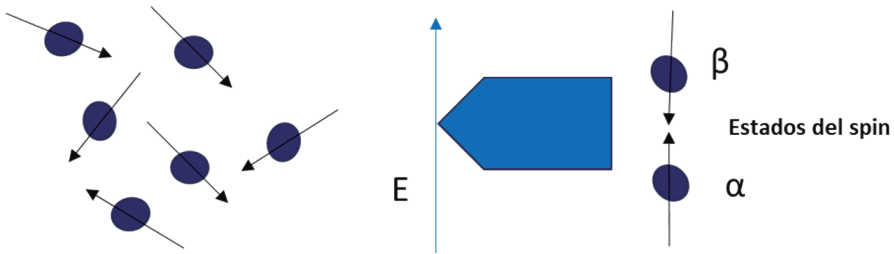
campos como la física, la química y la biología, entre otras disciplinas científicas (HALL, 1964). En medicina, se emplean diferentes técnicas espectroscópicas para examinar tejidos y facilitar diagnósticos.

La espectroscopia de RMN fue desarrollada a finales de la década de 1940 con el propósito de estudiar los núcleos atómicos. En 1951, los químicos descubrieron que esta técnica espectroscópica podía utilizarse para determinar las estructuras de compuestos orgánicos. Sin embargo, la RMN solo es aplicable a núcleos atómicos con un número impar de protones o neutrones, o ambos, lo cual se cumple en átomos como el hidrógeno-1 (^1H), carbono-13 (^{13}C), flúor-19 (^{19}F) y fósforo-31 (^{31}P) (SEMMLER, 2018; HALL, 1964). Estos núcleos se consideran magnéticamente activos debido a su espín, una propiedad similar a la de los electrones. Dado que los núcleos poseen carga positiva y rotan sobre un eje, se comportan como pequeños imanes.

En ausencia de un campo magnético, los espines nucleares se distribuyen aleatoriamente. Sin embargo, cuando se coloca una muestra en presencia de un campo magnético, como se muestra en la **figura 47**, los núcleos con espín positivo se alinean en la misma dirección del campo, alcanzando un estado de energía

mínima conocido como estado de espín α (SEMMLER, 2018). Por otro lado, los núcleos con espín negativo se orientan en dirección opuesta al campo magnético, en un estado de energía más alta llamado estado de espín β (figura 47).

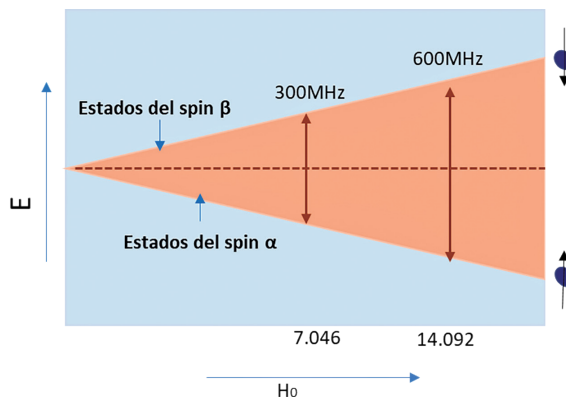
Figura 47. Estados de espín



Nota. La figura ilustra los estados de espín en un campo magnético, mostrando cómo los espines se orientan en direcciones de baja o alta energía cuando se aplica un campo, base de técnicas como la resonancia magnética nuclear (RMN).

La energía diferencial entre los estados de espín α y β está determinada por la intensidad del campo magnético aplicado H_0 . A medida que aumenta la fuerza del campo magnético, la diferencia energética entre los dos estados de espín también aumenta. Esto se ilustra en la figura 48 en la que representa el incremento de la diferencia energética entre los estados de espín en función del aumento de la fuerza del campo magnético (BOESCH,1990).

Figura 48. Estados de espín con la fuerza del campo magnético



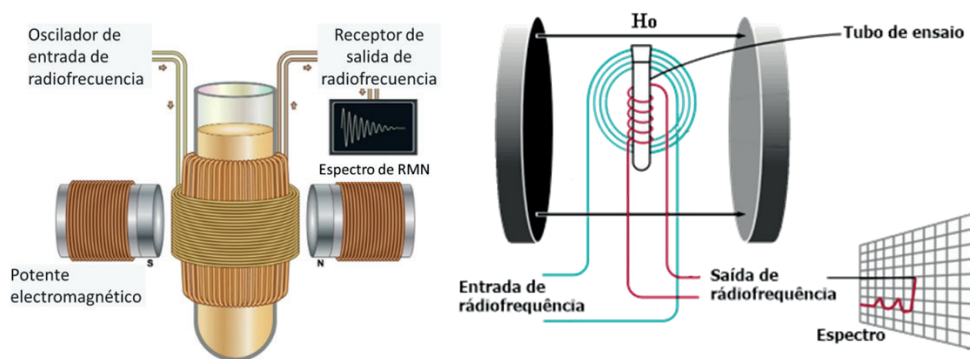
Nota. La figura muestra que, al aumentar la intensidad del campo magnético, la separación energética entre los estados de espín α y β crece, lo que eleva la frecuencia de resonancia necesaria (por ejemplo, de 300 MHz a 600 MHz) en técnicas como la RMN.

Cuando una muestra con un compuesto orgánico es irradiada brevemente con un pulso intenso de radiación de radiofrecuencia (rf), los núcleos en el estado de espín α se excitan al estado de espín β . Al volver al estado inicial, los núcleos emiten señales cuya frecuencia está relacionada con la diferencia de energía (ΔE) entre los estados de espín α y β (Boesch, 1990). Estas señales son detectadas por el espectrómetro de RMN y registradas como un espectro de RMN, que muestra las frecuencias frente a la intensidad. La RMN se refiere al fenómeno en el que los núcleos están en resonancia con la radiación de radiofrecuencia, cambiando de un estado de espín a otro en respuesta a la radiación rf aplicada (SEMMLER, 2018).

Cabe destacar que, para un mismo tipo de núcleo, las frecuencias de resonancia son diferentes debido a las distintas interacciones químicas presentes en el entorno molecular, lo que se conoce como desplazamiento químico (δ).

El espectrómetro de la RMN se compone de cuatro componentes principales representados de manera esquemática. En primer lugar, se encuentra un imán estable con un controlador que genera un campo magnético preciso, a continuación, está el transmisor de radiofrecuencias, que emite frecuencias exactas, asimismo, se cuenta con un detector que mide la absorción de energía de radiofrecuencia de la muestra, por último, se encuentran el ordenador y el registrador, encargados de generar las gráficas que conforman el espectro de RMN (SEMMLER, 2018) (figura 49).

Figura 49. Resonancia magnética nuclear



Nota. La figura representa un espectrómetro de RMN, donde un electroimán aplica un campo magnético constante y un oscilador emite radiofrecuencia hacia la muestra. La señal resultante es captada, procesada y convertida en un espectro de RMN para analizar la estructura molecular.

Para obtener un espectro de la RMN, se utiliza un tubo de vidrio largo que contiene una pequeña cantidad del compuesto orgánico disuelto en aproximadamente medio mililitro de disolvente (SEMMLER, 2018). Este tubo se coloca dentro del campo magnético del aparato y se hace girar alrededor de su eje vertical. En los equipos modernos, el campo magnético se mantiene constante mientras se emite un breve pulso de radiación de radiofrecuencia (RF) que excita simultáneamente a todos los núcleos presentes en la muestra. Cada núcleo absorbe la radiación de la frecuencia necesaria para entrar en resonancia y cambiar su estado de espín (BAUERMEISTER *et al.*, 2022). A medida que los núcleos regresan a su estado inicial, emiten radiación a una frecuencia igual a la diferencia de energía entre los estados de espín. La intensidad de esta radiación disminuye con el tiempo a medida que todos los núcleos vuelven a su estado original (SEMMLER, 2018).

Un espectro de RMN es una representación gráfica que muestra la relación entre la radiofrecuencia aplicada y la absorción de energía por parte de los núcleos atómicos. La posición en la gráfica donde ocurre la absorción se conoce como desplazamiento químico. El desplazamiento químico está determinado por la densidad electrónica alrededor del núcleo. Si un núcleo está rodeado de una alta densidad electrónica, está protegido del campo magnético externo, lo que resulta en un desplazamiento hacia valores superiores en el espectro de RMN (SEMMLER, 2018). Por otro lado, si un núcleo está rodeado de un átomo electronegativo, la densidad electrónica alrededor del núcleo disminuye, lo que provoca un efecto de *desprotección* y desplaza la señal hacia valores inferiores en el espectro de RMN. Además, el espín de los núcleos circundantes también puede afectar las señales observadas en el espectro de RMN a través de un fenómeno llamado *acoplamiento espín-espín*, donde las señales se escinden en varias líneas debido a las interacciones magnéticas entre los núcleos vecinos (SEMMLER, 2018).

La RMN se utiliza para identificar y cuantificar metabolitos en una muestra biológica. Esta técnica permite analizar la estructura y composición química de los metabolitos presentes en una muestra, lo que proporciona información sobre los cambios metabólicos.

La RMN se realiza utilizando un equipo especializado llamado espectrómetro de RMN. El proceso general para realizar una espectroscopia de RMN, es el siguiente:

- La preparación de la muestra es el primer paso en la espectroscopia de RMN. Se selecciona una muestra biológica y se extraen y purifican los metabolitos de interés. Luego, la muestra se coloca en el espectrómetro de RMN, donde se aplica un campo magnético intenso. Se utilizan secuencias de pulsos de radiofrecuencia para excitar los núcleos atómicos de los metabolitos presentes en la muestra, y las señales de radiofrecuencia emitidas por los núcleos se detectan y registran en forma de espectro de RMN (BAUERMEISTER *et al.*, 2022).

- Posteriormente, los datos espectrales se procesan y se lleva a cabo el análisis utilizando software especializado. Se realizan correcciones y ajustes, y se eliminan los artefactos. Luego, se utilizan métodos estadísticos y herramientas bioinformáticas para identificar y cuantificar los metabolitos presentes en la muestra. Los espectros de RMN procesados proporcionan información sobre la composición metabólica de la muestra, y se comparan con bases de datos y referencias para asignar y anotar los metabolitos identificados (BAUERMEISTER *et al.*, 2022).

5.3. ESPECTROSCOPIA INFRARROJA POR TRANSFORMADA DE FOURIER

La espectroscopia infrarroja por transformada de Fourier es una técnica analítica utilizada para estudiar la interacción de la radiación infrarroja con la materia. Mediante el análisis de la radiación absorbida por una muestra, es posible obtener información sobre la estructura y composición de los materiales. Esta técnica se basa en los principios de la transformada de Fourier, que permite convertir una señal temporal en su correspondiente espectro de frecuencia (SONG *et al.*, 2020).

Fundamento: es una técnica utilizada para obtener un espectro infrarrojo de absorción o emisión de un sólido, líquido o gas. Un espectrómetro FTIR recopila simultáneamente datos de alta resolución en un amplio rango espectral.

La espectroscopia infrarroja es una técnica utilizada para analizar la radiación infrarroja absorbida por una muestra en función de la longitud de onda. Esta radiación corresponde a la región del espectro electromagnético que se encuentra entre la radiación visible y la radiación de microondas. La espectroscopia infrarroja por transformada de Fourier es una variante de esta técnica que utiliza un interferómetro para obtener el espectro infrarrojo completo de una muestra (SONG *et al.*, 2020). A partir de este espectro, es posible identificar las vibraciones moleculares características de los compuestos presentes en la muestra, lo que permite su análisis y caracterización.

El ordenador recoge la intensidad de la señal en función del tiempo y la convierte en una representación de la intensidad en función de la frecuencia mediante una técnica llamada transformada de FOURIER (FT-RMN). Con esta transformación, se obtiene un espectro de RMN en tan solo 2 segundos, utilizando menos de 5 mg de muestra (SONG *et al.*, 2020). Esta técnica permite obtener información detallada sobre la estructura y composición de la muestra de manera rápida y eficiente.

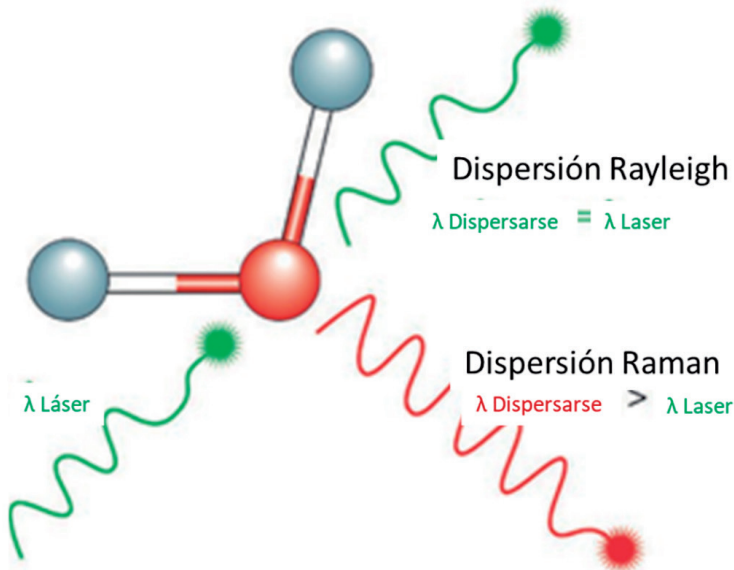
5.4. ESPECTROSCOPIA RAMAN

La espectroscopía Raman es una técnica que se basa en la dispersión inelástica de luz monocromática. En esta técnica, los fotones de la luz interactúan con una

muestra y experimentan un cambio en su frecuencia, lo que se conoce como efecto Raman. Este cambio en la frecuencia proporciona información valiosa sobre las transiciones vibracionales, rotacionales y otras de baja frecuencia en las moléculas analizadas (DODO *et al.*,2022).

La espectroscopía Raman es una herramienta eficiente para determinar la composición química y la estructura molecular de los compuestos. Además, ofrece ventajas significativas, como la capacidad de análisis no destructivo, la falta de necesidad de preparación de la muestra y la rapidez en la obtención de resultados. Estas características hacen que la espectroscopía Raman sea ideal para aplicaciones *in situ*, como en misiones espaciales, y forma parte de la nueva generación de instrumentación utilizada en las exploraciones planetarias (DODO *et al.*,2022) (figura 50).

Figura 50. Resonancia magnética nuclear



Nota. La figura muestra la dispersión Rayleigh (sin cambio en la longitud de onda) y la dispersión Raman (con cambio en la longitud de onda) cuando un láser interactúa con una molécula, siendo este último útil en espectroscopía Raman para estudiar vibraciones moleculares.

6. HERRAMIENTAS DE OTRAS CIENCIAS ÓMICAS

Otras ciencias ómicas, como la lipidómica, la epigenómica y la gliocómica, aplican las técnicas estándar de las principales ómicas. No obstante, se utilizan enfoques especiales, principalmente en los métodos de extracción y separación, que permiten llevar a cabo estos análisis. La metagenómica, junto con sus técnicas, se abordará más adelante.

6.1. LIPIDÓMICA

Los lípidos son biomoléculas importantes en todos los sistemas biológicos y cumplen numerosas funciones celulares esenciales. El análisis global de lípidos complejos es un gran desafío debido a la extrema diversidad en las estructuras de los lípidos (PHAM *et al.*, 2019). La lipidómica es el estudio integral de los lípidos presentes en una muestra biológica, que incluye su identificación, cuantificación y caracterización. Esta disciplina utiliza una variedad de técnicas analíticas para investigar la composición y la función de los lípidos presentes en una muestra.

La variación en los enlaces y las posiciones de las cadenas de acilo graso en la cadena principal de los lípidos, la modificación del grupo funcional y la aparición de especies moleculares como isómeros o isobaras se encuentran entre algunos de los mayores desafíos a resolver en lipidómica (PHAM *et al.*, 2019).

Antes del análisis, los lípidos deben extraerse de la muestra biológica. Las técnicas de extracción más utilizadas en lipidómica incluyen la extracción líquido-líquido con solventes orgánicos, como cloroformo/metanol, y la extracción en fase sólida (SPE) con cartuchos de fase sólida que retienen los lípidos (OGISO *et al.*, 2008). Las cromatografías también son utilizadas: la líquida se utiliza en combinación con la espectrometría de masas para separar y purificar los lípidos antes de su análisis. La cromatografía de fase reversa (RPLC) es ampliamente utilizada en lipidómica debido a su capacidad para retener y separar diferentes clases de lípidos en función de sus propiedades hidrofóbicas (PHAM *et al.*, 2019).

Para la identificación y cuantificación precisa de los lípidos, se utilizan estándares de referencia y bibliotecas de espectros de masas. Estos recursos contienen información sobre las masas, estructuras y propiedades espectrales de los lípidos conocidos, lo que permite la comparación con los espectros de masas generados experimentalmente. La lipidómica genera grandes cantidades de datos complejos, también se necesita análisis bioinformático que son esenciales para procesar, visualizar e interpretar estos datos. Se utilizan *software* especializado para identificar y cuantificar los lípidos, realizar análisis estadísticos y comparaciones entre muestras (OGISO *et al.*, 2008).

6.1.1. LA CROMATOGRAFÍA DE FASE REVERSA (RPLC)

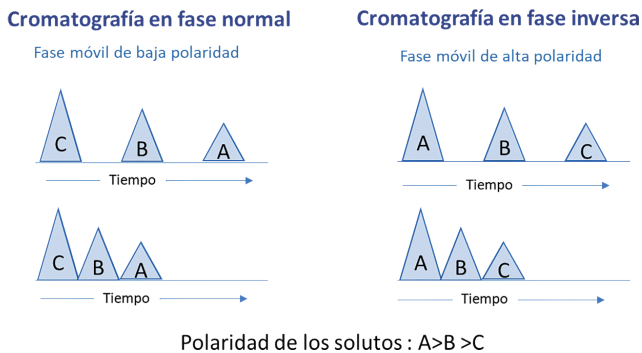
La cromatografía de fase reversa se ha convertido en una técnica de separación muy utilizada en el laboratorio debido a su eficacia y versatilidad en la separación de una amplia gama de compuestos químicos. Además, se han desarrollado diferentes tipos de fases estacionarias y eluyentes para adaptarse a las necesidades específicas de separación de diferentes componentes en las muestras a analizar (LI *et al.*, 2019).

Esta cromatografía está basada en interacciones hidrofóbicas, se fundamenta en las fuerzas de repulsión entre un disolvente relativamente polar, un compuesto relativamente apolar y una fase estacionaria apolar. Estas interacciones son responsables de la separación de los componentes en una mezcla.

A diferencia de la cromatografía convencional, en la cromatografía de interacciones hidrofóbicas, la fase estacionaria consiste en partículas de sílica químicamente modificadas con hidrocarburos saturados, insaturados o aromáticos de diferentes tipos. Esto convierte a la fase estacionaria en una matriz apolar (LI *et al.*, 2019). Por lo tanto, para este tipo de cromatografía, se utilizan mezclas de solventes polares, como agua, acetonitrilo, acetato de etilo, acetona y alcoholes alifáticos (**figura 51**).

La fase móvil utilizada en la cromatografía de fase reversa es una mezcla de agua o solución tampón con eluyentes polares, como metanol, acetonitrilo o tetrahydrofurano. La fase móvil polar actúa como disolvente y ayuda a eluir los compuestos de la muestra a través de la columna cromatográfica. La fase estacionaria es no polar y está compuesta por hidrocarburos de cadena larga unidos a un soporte de sílice o híbrido. Esta fase permite la retención y separación de los compuestos de la muestra en función de su interacción con ella (PHAM *et al.*, 2019).

Figura 51. Cromatografía en fase inversa



Nota. La figura muestra la cromatografía en fase normal (compuestos menos polares eluyen primero) y la cromatografía en fase inversa (compuestos más polares eluyen primero), diferenciándose por la polaridad de las fases móvil y estacionaria para separar compuestos.

En la cromatografía de fase reversa, los compuestos de la muestra se retienen en la columna debido a las interacciones hidrofóbicas que establecen con la fase estacionaria no polar. Estas interacciones hidrofóbicas son en general bastante débiles, pero debido a su gran cantidad, permiten la separación de los compuestos. Para eluir los compuestos retenidos en la columna, se utiliza un gradiente de elución (Li *et al.*, 2019). El gradiente de elución implica cambiar gradualmente la proporción de los eluyentes polares y no polares en la fase móvil. A medida que se incrementa la concentración de los eluyentes no polares en la fase móvil, los compuestos retenidos se desplazan y se eluyen de la columna. A medida que los compuestos eluidos pasan por el detector, se registran las señales correspondientes y se generan cromatogramas (Li *et al.*, 2019). Estos cromatogramas proporcionan información sobre la retención y la cantidad de cada compuesto en la muestra analizada.

En lipidómica se pueden utilizar diferentes enfoques de EM que permite la identificación y cuantificación de los lípidos presentes en una muestra. La espectrometría de masas en tándem (MS/MS) y espectrometría de masas de alta resolución (HRMS), permiten obtener información estructural detallada y mejorar la precisión de la identificación de los lípidos (Li *et al.*, 2019).

6.1.2. ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN (HRMS)

La espectrometría de masas de alta resolución es una técnica química utilizada para determinar la masa de las especies químicas en una muestra con una mayor precisión y exactitud. A continuación, se describe cómo se realiza esta técnica.

La muestra se somete a un proceso de ionización en el cual las moléculas se convierten en iones cargados eléctricamente (ZÜLLIG y KÖFELER, 2021). Esto se puede lograr mediante diferentes técnicas de ionización, como la ionización por impacto electrónico (EI), la ionización por *electrospray* (ESI) o la ionización química. Los iones generados en la etapa de ionización se introducen en el analizador de masa, que es capaz de separar los iones en función de su relación masa/carga (m/z). En el caso de la espectrometría de masas de alta resolución, se utilizan analizadores de masa más avanzados, como el espectrómetro de masas de cuadrupolo-tiempo de vuelo (Q-TOF), que ofrecen una mayor precisión y resolución. Los iones separados se detectan y se mide su relación masa/carga. Los espectrómetros de masas de alta resolución permiten una mayor precisión en la determinación de la masa exacta de los iones presentes en la muestra (ZÜLLIG y KÖFELER, 2021). Esto se logra mediante el uso de analizadores con alta resolución y prestaciones mejoradas, como el cuadrupolo-tiempo de vuelo (Q-TOF). Los datos obtenidos se analizan utilizando *software* especializado que permite interpretar los espectros de masas y realizar la identificación y cuantificación de los compuestos presentes en la muestra. Además, se pueden

comparar los espectros de masas obtenidos con bases de datos para identificar compuestos conocidos y realizar la caracterización estructural.

6.1.3. CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN (HPLC)

El empleo de cromatografía líquida de alta resolución (HPLC) es ampliamente reconocido en el ámbito de la lipidómica para el análisis regular de complejas mezclas de lípidos (TANG *et al.*, 2016). Esta técnica implica la utilización tanto de cromatografía líquida de fase normal (NPLC) como de fase inversa (RPLC). Mientras que la NPLC separa típicamente los lípidos según su grupo funcional polar, la RPLC distingue las especies de lípidos según su composición lipófila. Se ha despertado un interés particular en la cromatografía líquida de interacción hidrófila (HILIC), la cual facilita la separación de clases de lípidos en función de la composición del grupo principal, ofreciendo similitudes con la NPLC (BLUM, 2014). En este enfoque, se emplean las mismas fases móviles de la RPLC para mejorar la eficacia y la reproducibilidad de la ionización. HILIC se puede aplicar como un sistema alternativo o complementario, utilizado de manera mixta con la cromatografía líquida de fase inversa, en el análisis de lípidos complejos.

La espectrometría de masas de alta resolución tiene diversas aplicaciones en diferentes áreas de la química y la investigación científica. Permite la identificación precisa de compuestos, la determinación de masas exactas, el estudio de estructuras moleculares y la caracterización de biomoléculas, entre otras aplicaciones.

La lipidómica también puede emplear otras técnicas analíticas, como la espectroscopia de resonancia magnética nuclear (RMN) y la espectrometría de movilidad iónica (IMS), para estudiar los lípidos en detalle o proporcionar información adicional sobre su estructura y función. Es importante destacar que la lipidómica es un campo en constante evolución y que constantemente se están desarrollando nuevas técnicas y enfoques para mejorar la capacidad de análisis y comprensión de los lípidos en diferentes sistemas biológicos (BLUM, 2014).

6.1.4. ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS)

Es una técnica que se utiliza para separar y caracterizar moléculas en función de su movilidad bajo un campo eléctrico en un gas a baja presión. Se basa en la medida de la velocidad con la que las moléculas viajan a través de un tubo de reacción, donde la movilidad está determinada por su tamaño, forma y carga. Esta técnica permite la identificación de compuestos individuales y la obtención de información sobre su estructura y características (GÖTH y PAGEL, 2017). Es especialmente útil en la detección y cuantificación de contaminantes en muestras complejas.

La espectrometría de movilidad iónica se basa en la separación de las moléculas en un gas mediante un campo eléctrico. Durante el proceso, las moléculas se ionizan y adquieren una carga eléctrica. A medida que se mueven a través del tubo de reacción, las moléculas se separan según su movilidad, con las moléculas más pequeñas y menos cargadas viajando más rápido que las moléculas más grandes y cargadas (GÖTH Y PAGEL, 2017). Esta separación se logra mediante la aplicación de un campo eléctrico en combinación con un gas portador de baja presión. La velocidad de las moléculas se registra utilizando detectores de ionización, lo que permite la generación de un espectro de movilidad iónica que puede ser utilizado para identificar y cuantificar las moléculas presentes en la muestra (WINTER *et al.*, 2019).

6.2. GLICÓMICA

La glicómica es una disciplina científica que estudia los carbohidratos (glúcidos) presentes en una muestra biológica, incluyendo sus estructuras, funciones y perfiles de expresión (LAGEVEEN-KAMMEIJER *et al.*, 2022). También estudia los glicoconjugados, tales como glicoproteínas y glucolípidos, y su papel en diversos procesos biológicos. Estos glicoconjugados desempeñan un papel crucial en la comunicación celular, la adhesión celular y la respuesta inmune. La comprensión de la estructura y función de los glicoconjugados es fundamental para avanzar en la investigación biomédica, el desarrollo de fármacos y la comprensión de enfermedades relacionadas con alteraciones en la glicómica (THAYSEN-ANDERSEN *et al.*, 2021). Por lo tanto, la glicómica tiene una gran relevancia en la investigación científica actual y utiliza una variedad de técnicas analíticas, al igual que el análisis bioinformático que utiliza herramientas y *software* especializados para el procesamiento de datos, la asignación de estructuras de glicanos y la interpretación de los resultados.

Los glicanos pueden ser analizados mediante la utilización de enzimas específicas. Las enzimas glicosidasas se utilizan para degradar y liberar los glicanos de una muestra, lo que permite su posterior análisis y caracterización. Por otro lado, técnicas como la cromatografía líquida, en combinación con diferentes tipos de detectores, se utiliza para separar y purificar los glicanos presentes en una muestra. Además, se pueden emplear diferentes métodos de cromatografía, como la cromatografía de exclusión molecular (SEC) y la cromatografía de interacción hidrofílica (HILIC) para separar glicanos en función de sus propiedades fisicoquímicas y de sus distintos tamaños (BARTH *et al.*, 1994).

6.2.1. CROMATOGRAFÍA DE EXCLUSIÓN MOLECULAR

Es una técnica cromatográfica que se utiliza para separar moléculas según su tamaño y peso molecular. En esta técnica, se utiliza una columna con una

matriz porosa que permite el paso de moléculas más pequeñas, mientras que las moléculas de mayor tamaño se retienen en la matriz (BARTH *et al.*, 1994). De esta manera, se logra separar moléculas según su peso y determinar su distribución de tamaños en una muestra. Se acopla a técnicas claves como la EM, proporciona información sobre las masas moleculares y estructuras de los glicanos, lo que facilita su identificación y caracterización. Se utiliza el enfoque de espectrometría de masas en tándem (MS/MS) y espectrometría de masas de alta resolución (HRMS), para analizar los glucanos presentes en la muestra (THAYSEN-ANDERSEN *et al.*, 2021).

6.2.2. RMN DE PROTONES

La RMN de protones es una técnica de espectroscopia que permite analizar y estudiar la estructura y composición de compuestos orgánicos e inorgánicos en solución. Los protones presentes en las moléculas son sensibles a los campos magnéticos, lo que permite generar señales detectables y obtener información sobre la conectividad y el entorno químico de los átomos de hidrógeno en una muestra (KISHIMOTO *et al.*, 2018). Esta técnica es ampliamente utilizada en la determinación de estructuras moleculares y en el análisis de reacciones químicas, proporcionando información detallada sobre las interacciones y la dinámica molecular.

6.2.3. RMN DE CARBONO-13

Es una técnica de espectroscopia que se utiliza para analizar la estructura y composición de compuestos orgánicos e inorgánicos. A diferencia de la RMN de protones, que se centra en la detección de señales de átomos de hidrógeno, la RMN de carbono-13 se basa en la detección de señales generadas por los átomos de carbono presentes en una muestra (RASMUSSEN y HOFFMAN, 2022). Esta técnica es especialmente útil en el análisis de compuestos orgánicos, ya que proporciona información sobre el tipo y la posición de los átomos de carbono en una molécula.

La RMN de carbono-13 se complementa con la RMN de protones para obtener una caracterización completa de las moléculas en estudio. Se utiliza para analizar los glicanos y determinar la conectividad de los azúcares y los enlaces glicosídicos presentes en sus estructuras (RASMUSSEN y HOFFMAN, 2022).

6.2.4. TÉCNICAS COMPLEMENTARIAS

Además de las técnicas mencionadas anteriormente, la glicómica puede emplear otras técnicas analíticas, como la cromatografía de afinidad, la electroforesis capilar y la microscopía de fuerza atómica (AFM), para estudiar

los glicanos en detalle o proporcionar información adicional sobre su interacción con otras moléculas o su localización en células y tejidos (LAGEVEEN-KAMMEIJER *et al.*, 2022).

6.3. EPIGENÓMICA

En la epigenómica, que es el estudio de las modificaciones químicas en el ADN y las proteínas que influyen en la regulación de la expresión génica, se utilizan varias técnicas analíticas para investigar y comprender los cambios epigenéticos. La epigenómica genera grandes cantidades de datos complejos, por lo que el análisis bioinformático desempeña un papel fundamental (FANTER *ET al.*, 2022). Se utilizan herramientas y algoritmos especializados para procesar y analizar los datos de secuenciación, identificar sitios de modificación epigenética, realizar comparaciones entre muestras y descubrir patrones y asociaciones relevantes.

Dentro de las técnicas la NGS ha revolucionado el campo de la epigenómica al permitir el análisis a gran escala y de alta resolución de los cambios epigenéticos. Se utilizan diferentes enfoques de NGS, como el bisulfito de secuenciación (BS-seq) y el ChIP-seq, para generar datos detallados sobre la metilación del ADN, las modificaciones de las histonas y las interacciones proteína-ADN. A continuación, se describen brevemente estas técnicas analíticas utilizadas actualmente en epigenómica.

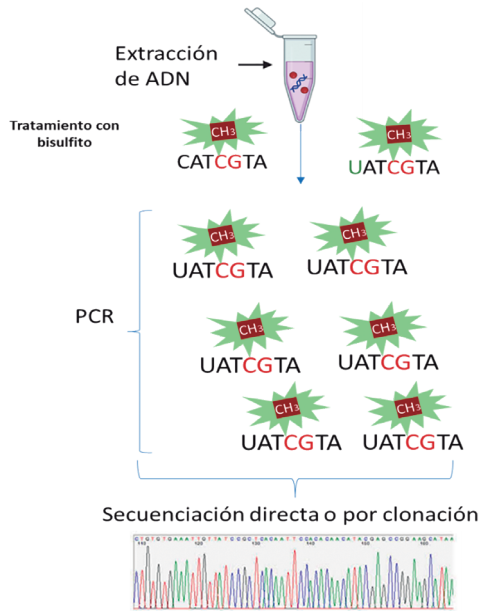
6.3.1. BISULFITO DE SODIO (BS)

El bisulfito de sodio convierte químicamente los residuos de citosina no metilados en uracilo, mientras que los residuos de citosina metilados no sufren cambios (PAJARES *et al.*, 2021). Posteriormente, se pueden utilizar diferentes enfoques, como la reacción en cadena de la polimerasa (PCR) o la NGS, para identificar los patrones de metilación en el ADN. La técnica de bisulfito de sodio es compleja y requiere experiencia en su realización y análisis de los resultados obtenidos. Además, hay variantes y protocolos específicos para diferentes aplicaciones y tipos de muestras.

En esta técnica el tratamiento de la muestra comienza con la aplicación de bisulfito de sodio al ADN genómico, lo cual provoca una reacción selectiva con los residuos de citosina no metilados, convirtiéndolos en uracilo. Por otro lado, los residuos de citosina metilados permanecen sin cambios, posteriormente, se procede a la desnaturalización del ADN tratado mediante calor o alcalinidad, lo cual separa las dos hebras del ADN (PAJARES *et al.*, 2021). A continuación, las hebras desnaturalizadas se someten a una reacción de bisulfito de sodio, donde

los uracilos derivados de las citosinas no metiladas se transforman en timinas, mientras que las citosinas metiladas se mantienen inalteradas (**figura 52**).

Figura 52. Técnica de bisulfito



Nota. La figura muestra la técnica de bisulfito, donde el tratamiento con bisulfito convierte citosinas no metiladas en uracilos, mientras que las metiladas permanecen iguales. Tras la secuenciación, se identifican los patrones de metilación en el ADN.

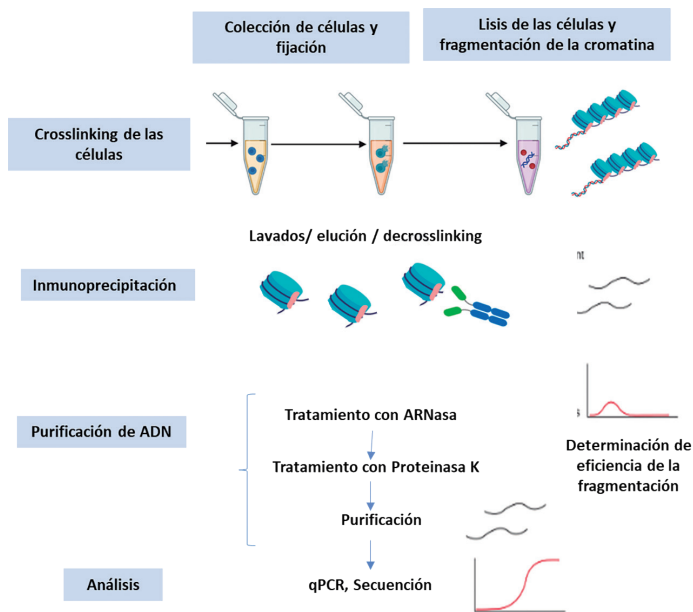
Una vez completado este proceso, se realiza la purificación del ADN tratado con bisulfito de sodio con el fin de eliminar los reactivos residuales y otros contaminantes. Finalmente, el ADN purificado se somete a técnicas de secuenciación o PCR para analizar la metilación de los sitios específicos de citosina. La diferencia en la secuencia de ADN después del tratamiento con bisulfito de sodio permite identificar los sitios de citosina metilados y no metilados (PAJARES *et al.*, 2021).

6.3.2. CROMATINA DE INMUNOPRECIPITACIÓN (CHIP)

Esta técnica permite investigar las interacciones entre las proteínas de unión al ADN, como las histonas modificadas y los factores de transcripción, y las regiones específicas del ADN (DECAPRIO y KOHL, 2020). El ChIP se combina con la NGS (ChIP-seq) para identificar los sitios de unión de las proteínas y comprender cómo afectan la expresión génica.

La técnica de Cromatina de Inmunoprecipitación (ChIP) involucra varios pasos para investigar las interacciones entre las proteínas y la cromatina en el núcleo de las células. En primer lugar, las células se tratan con un agente de fijación para preservar las interacciones proteína-ADN en su estado natural. Luego, se realiza la lisis celular para liberar la cromatina, que consiste en ADN, histonas y proteínas (RODRÍGUEZ-UBREVA Y BALLESTAR; 2014; DECAPRIO Y KOHL, 2020). La cromatina se separa en pequeños fragmentos mediante técnicas como ultrasonografía, enzimas o digestión química. A continuación, se agrega un anticuerpo específico para la proteína de interés, y el complejo anticuerpo-proteína-ADN se captura utilizando una matriz de afinidad, como perlas magnéticas recubiertas de proteína A/G. Después de los lavados para eliminar impurezas, se realiza la elución para separar los complejos inmunoprecipitados y liberar el ADN y las proteínas asociadas (**figura 53**) (RODRÍGUEZ-UBREVA Y BALLESTAR, 2014).

Figura 53. Cromatina de inmunoprecipitación (ChIP)



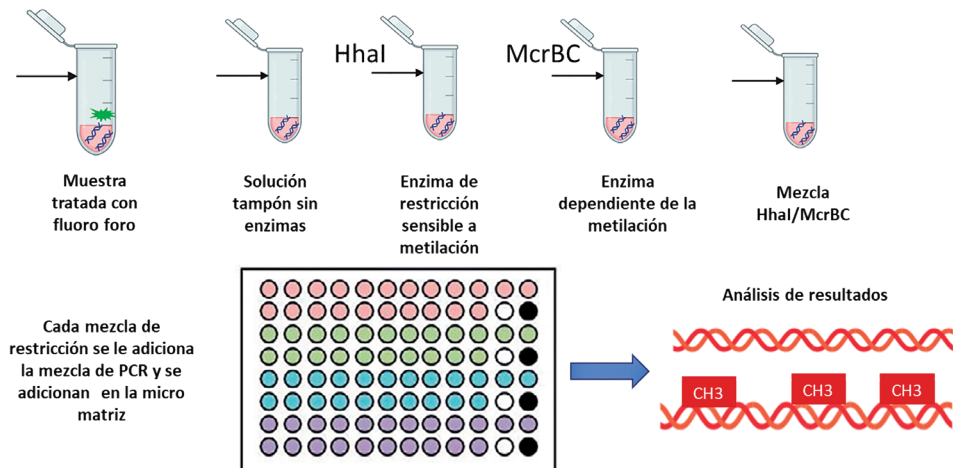
Nota. La figura describe la inmunoprecipitación de cromatina (ChIP), donde se fija la interacción proteína-ADN, se fragmenta la cromatina, se usa un anticuerpo para precipitar las proteínas de interés junto con su ADN asociado, y luego se purifica y analiza el ADN para identificar regiones genómicas que interactúan con esas proteínas.

Luego, se revierten las uniones cruzadas entre proteínas y ADN, permitiendo la liberación del ADN. El ADN recuperado se puede analizar mediante técnicas como PCR, secuenciación o microarreglos para identificar los sitios de unión de la proteína de interés en el genoma (RODRÍGUEZ-UBREVA Y BALLESTAR, 2014). La técnica de ChIP proporciona información valiosa sobre la regulación génica y los procesos epigenéticos en las células.

6.3.3. *MICROARREGLOS EPIGENÉTICOS*

Son herramientas que permiten analizar simultáneamente múltiples regiones del genoma para detectar cambios epigenéticos. Estos microarreglos están diseñados para identificar y cuantificar modificaciones específicas del ADN o las histonas en una muestra, lo que proporciona información sobre los perfiles epigenéticos (FENG Y LOU, 2019). La técnica de microarreglos epigenéticos comienza con la preparación de la muestra, donde se extrae el ADN genómico y se trata con enzimas de restricción o bisulfito de sodio para modificar las bases citosina no metiladas. Esto permite distinguir las regiones metiladas y no metiladas del ADN. A continuación, se generan sondas de ADN o ARN representando secuencias específicas del genoma, que se imprimen en un microarreglo (**figura 54**) (VAN STEENSEL, 2005).

La muestra de ADN tratada se marca con un fluoróforo y se hibrida con el microarreglo, permitiendo la detección de la presencia o ausencia de metilación en regiones específicas del genoma. Posteriormente, el microarreglo se escanea para capturar la fluorescencia de los fluoróforos unidos a las sondas, y los datos se analizan para determinar los niveles de metilación en diferentes regiones del genoma (FENG Y LOU, 2019). Finalmente, los resultados se interpretan comparándolos con una referencia de control, lo que proporciona información sobre los patrones epigenéticos y las modificaciones en la metilación del ADN en la muestra.

Figura 54. *Microarreglos epigenéticos*

Nota. La figura muestra el proceso de análisis mediante microarreglos epigenéticos para estudiar patrones de metilación del ADN. Primero, el ADN se extrae y se trata con bisulfito para convertir las citosinas no metiladas en uracilos, mientras que las citosinas metiladas permanecen sin cambios. Luego, el ADN tratado se amplifica y se aplica a un microarreglo con sondas específicas que permiten identificar los sitios metilados en el genoma. Al final, el análisis revela patrones de metilación (marcados con grupos metilo, CH₃), lo cual es clave en el estudio de la regulación epigenética y su relación con procesos biológicos y enfermedades.

CONCLUSIONES

Las tecnologías utilizadas en las diferentes ciencias ómicas son fundamentales para comprender la complejidad y diversidad de los sistemas biológicos. Estas tecnologías permiten estudiar de manera integral los diferentes niveles de información molecular, como el genoma, el transcriptoma, el proteoma y el metaboloma, lo que proporciona una visión completa de la estructura y función de los organismos vivos.

Las diversas tecnologías ómicas son herramientas vigorosas para el descubrimiento de biomarcadores y la comprensión de enfermedades. Estas tecnologías admiten identificar patrones moleculares asociados con enfermedades, lo que facilita la detección temprana, el diagnóstico preciso y el desarrollo de terapias más eficaces. Además, brindan información detallada sobre las redes de interacción molecular y los mecanismos subyacentes de las enfermedades.

Estas tecnologías ómicas tienen aplicaciones en diversos campos, como la agricultura, la biotecnología y la medicina personalizada. Son tecnologías que permiten mejorar la producción agrícola, desarrollar nuevos productos y terapias basadas en la comprensión de los perfiles moleculares individuales y personalizar los tratamientos médicos de acuerdo con las características genéticas y moleculares de cada individuo.

REFERENCIAS

- ALJANAHI, A. A., DANIELSEN, M. and DUNBAR, C. E. (2018). An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Molecular therapy. Methods and Clinical Development*, 10, 189-196. <https://doi.org/10.1016/j.omtm.2018.07.003>
- APARNA, G. M. and TETALA, K. K. R. (2023). Recent Progress in Development and Application of DNA, Protein, Peptide, Glycan, Antibody, and Aptamer Microarrays. *Biomolecules*, 13(4). <https://doi.org/10.3390/biom13040602>
- ARDUI, S., AMEUR, A., VERMEESCH, J. R. and HESTAND, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research*, 46(5), 2159-2168. <https://doi.org/10.1093/nar/gky066>
- BAMSHAD, M. J., Ng, S. B., BIGHAM, A. W., TABOR, H. K., EMOND, M. J., NICKERSON, D. A. and SHENDURE, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11), 745-755. <https://doi.org/10.1038/nrg3031>
- BARCISZEWSKA, M. Z., PERRIGUE, P. M. and BARCISZEWSKI, J. (2016). tRNA -the golden standard in molecular biology. *Molecular BioSystems*, 12(1), 12-17. <https://acortar.link/FdstRr>
- BARTH, H. G., BOYES, B. E. and JACKSON, C. (1994). Size exclusion chromatography. *Analytical Chemistry*, 66(12), 595-620. <https://doi.org/10.1021/ac00084a022>
- BAUERMEISTER, A., MANNOCHIO-RUSSO, H., COSTA-LOTUFO, L. V., JARMUSCH, A. K. and DORRESTEIN, P. C. (2022). Mass spectrometry-based metabolomics in microbiome investigations. *Nature Reviews Microbiology*, 20(3), 143-160. <https://acortar.link/TwlZZ8>
- BEHJATI, S. and TARPEY, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), 236-238. <https://doi.org/10.1136/archdischild-2013-304340>

- BENNETT, B. D., YUAN, J., KIMBALL, E. H. and RABINOWITZ, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols*, 3(8), 1299-1311. <https://doi.org/10.1038/nprot.2008.107>
- BLUM, F. (2014). High performance liquid chromatography. *British Journal of Hospital Medicine*, 75(2), 18-21. <https://doi.org/10.12968/hmed.2014.75.Sup2.C18>
- BOESCH, C. (1990). Kernspinresonanz [Nuclear magnetic resonance]. *Beitrage zur Infusionstherapie*, 25, 359-373
- CARRILLO SÁNCHEZ, K. (2010). *Secuenciación Nueva Generación SOLiD System*. <https://acortar.link/XSwIf3>
- COHEN, A. S., NAJARIAN, D. R., PAULUS, A., GUTTMAN, A., SMITH, J. A. and KARGER, B. L. (1988). Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America*, 85(24), 9660-9663. <https://doi.org/10.1073/pnas.85.24.9660>
- DECAPRIO, J. and KOHL, T. O. (2020). Chromatin Immunoprecipitation. *Cold Spring Harbor Protocols*. <https://doi.org/10.1101/pdb.prot098665>
- DODO, K., FUJITA, K. and SODEOKA, M. (2022). Raman Spectroscopy for Chemical Biology Research. *Journal of the American Chemical Society*, 144(43), 19651-19667. <https://doi.org/10.1021/jacs.2c05359>
- DONG, W., QIANG, B. and YANG, H. (2021). The international Human Genome Project. A milestone for life sciences and humanity —The three stages and three major impacts of the HGP, and three contributions by China. *Quantitative Biology*, 9(3). <https://doi.org/10.15302/J-QB-021-0265>
- DUARTE, R. M. B. O., BRANDÃO, P. F. and DUARTE, A. C. (2023). Multidimensional chromatography in environmental analysis: Comprehensive two-dimensional liquid versus gas chromatography. *Journal of Chromatography A*, 1706. <https://acortar.link/8jZuRe>
- EBRAHIMI-NAJAFABADI, H., LEARDI, R. and JALALI-HERAVI, M. (2014). Experimental design in analytical chemistry-Part I: Theory. *Journal of AOAC International*, 97(1), 3-11. <https://doi.org/10.5740/jaoacint.sgebrahimil>
- ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S. and MITCHELL, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5). <https://doi.org/10.1371/journal.pone.0019379>
- FAHY, E., SUBRAMANIAM, S., BROWN, H. A., GLASS, C. K., MERRILL Jr., A. H., MURPHY, R. C., RAETZ, C. R., RUSSELL, D. W., SEYAMA, Y., SHAW, W., SHI-

- MIZU, T., SPENER, F., VAN MEER, G., VAN NIEUWENHZE, M. S., WHITE, S. H., WITZTUM, J. L. and DENNIS, E. A. (2005). A comprehensive classification system for lipids. *Journal of Lipid Research*, 46(5), 839-861. <https://doi.org/10.1194/jlr.E400004-JLR200>
- FANTER, C., MADELAIRE, C., GENEREUX, D. P., van BREUKELLEN, F., LEVESQUE, D. and HINDLE, A. (2022). Epigenomics as a paradigm to understand the nuances of phenotypes. *The Journal of Experimental Biology*, 225. <https://doi.org/10.1242/jeb.243411>
- FENG, L. and LOU, J. (2019). DNA Methylation Analysis. *Methods in Molecular Biology*, 1894, 181-227. https://doi.org/10.1007/978-1-4939-8916-4_12
- FENG, Y., ZHANG, Y., YING, C., WANG, D. and DU, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, 13(1), 4-16. <https://doi.org/10.1016/j.gpb.2015.01.009>
- FIEHN, O. (2008). Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends in Analytical Chemistry*, 27(3), 261-269. <https://doi.org/10.1016/j.trac.2008.01.007>
- FILEP, C. and GUTTMAN, A. (2021). Capillary sodium dodecyl sulfate gel electrophoresis of proteins: Introducing the three-dimensional Ferguson method. *Analytica Chimica Acta*, 1183. <https://doi.org/10.1016/j.aca.2021.338958>
- GARCÍA, A., GODZIEN, J., LÓPEZ-GONZÁLVIZ, Á. and BARBAS, C. (2017). Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. *Bioanalysis*, 9(1), 99-130. <https://doi.org/10.4155/bio-2016-0216>
- GONDANE, A. and ITKONEN, H. M. (2023). Revealing the History and Mystery of RNA-Seq. *Current Issues in Molecular Biology*, 45(3), 1860-1874. <https://doi.org/10.3390/cimb45030120>
- GONULALAN, E. M., NEMUTLU, E., BAYAZEID, O., KOÇAK, E., YALÇIN, F. N. and DEMIREZER, L. O. (2020). Metabolomics and proteomics profiles of some medicinal plants and correlation with BDNF activity. *Phytomedicine. International Journal of Phytotherapy and Phytopharmacology*, 74. <https://doi.org/10.1016/j.phymed.2019.152920>
- GOODWIN, S., MCPHERSON, J. D. and MCCOMBIE, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>
- GÖTH, M. and PAGEL, K. (2017). Ion mobility-mass spectrometry as a tool to investigate protein-ligand interactions. *Analytical and Bioanalytical Chemistry*, 409(18), 4305-4310. <https://doi.org/10.1007/s00216-017-0384-9>

- HAAG, A. M. (2016). Mass Analyzers and Mass Spectrometers. *Advances in Experimental Medicine and Biology*, 919, 157-169. https://doi.org/10.1007/978-3-319-41448-5_7
- HAGGARTY, J. and BURGESS, K. E. (2017). Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome. *Current Opinion in Biotechnology*, 43, 77-85. <https://doi.org/10.1016/j.copbio.2016.09.006>
- HALL, L. D. (1964). Nuclear Magnetic Resonance. *Advances in Carbohydrate Chemistry*, 19, 51-93. [https://doi.org/10.1016/s0096-5332\(08\)60279-9](https://doi.org/10.1016/s0096-5332(08)60279-9)
- HAN, C., WANG, Q., YAO, Y., ZHANG, Q., HUANG, J., ZHANG, H. and QU, L. (2023). Thin layer chromatography coupled with surface enhanced Raman scattering for rapid separation and on-site detection of multi-components. *Journal of Chromatography A*, 1706. <https://doi.org/10.1016/j.chroma.2023.464217>
- HARRINGTON, C. T., LIN, E. I., OLSON, M. T. and ESHLEMAN, J. R. (2013). Fundamentals of pyrosequencing. *Archives of Pathology and Laboratory Medicine*, 137(9), 1296-1303. <https://doi.org/10.5858/arpa.2012-0463-RA>
- HEUMOS, L., SCHAAR, A. C., LANCE, C., LITINETSKAYA, A., DROST, F., ZAPPIA, L., LÜCKEN, M. D., STROBL, D. C., HENAO, J., CURION, F., Single-cell Best Practices Consortium, Schiller, H. B. and Theis, F. J. (2023). Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 24, 550-572. <https://doi.org/10.1038/s41576-023-00586-w>
- HOLLAND, L. A. and CASTO-BOGGESS, L. D. (2023). Gels in Microscale Electrophoresis. *Annual Review of Analytical Chemistry*, 16(1), 161-179. <https://acortar.link/ULzREF>
- HORNBECK, P. V., ZHANG, B., MURRAY, B., KORNHAUSER, J. M., LATHAM, V. and SKRZYPEK, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, 43, 512-520. <https://doi.org/10.1093/nar/gku1267>
- HU, T., Chitnis, N., MONOS, D. and DINH, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801-811. <https://doi.org/10.1016/j.humimm.2021.02.012>
- HUANG, J. Y., XIA, L., XIAO, X. H. and LI, G. K. (2023). Advances in microchip electrophoresis for the separation and analysis of biological samples. *Chinese Journal of Chromatography*, 41(8), 641-650. <https://doi.org/10.3724/SP.J.1123.2022.12004>
- KE, R., MIGNARDI, M., HAULING, T. and NILSSON, M. (2016). Fourth Generation of Next-Generation Sequencing Technologies: Promise and Consequences. *Human Mutation*, 37(12),1363-1367. <https://doi.org/10.1002/humu.23051>

- KHOZYAINOVA, A. A., VALYAIEVA, A. A., ARBATSKY, M. S., ISAEV, S. V., IAMSHCHIKOV, P. S., Volchkov, E. V., Sabirov, M. S., Zainullina, V. R., Chechekhin, V. I., Vorobev, R. S., MENYAILO, M. E., TYURIN-KUZMIN, P. A. and DENISOV, E. V. (2023). Complex Analysis of Single-Cell RNA Sequencing Data. *Biochemistry*, 88(2), 231-252. <https://acortar.link/EUj48P>
- KISHIMOTO, S., KRISHNA, M. C., KHRAMTSOV, V. V., UTSUMI, H. and LURIE, D. J. (2018). In Vivo Application of Proton-Electron Double-Resonance Imaging. *Antioxidants and Redox Signaling*, 28(15), 1345-1364. <https://doi.org/10.1089/ars.2017.7341>
- KLEINO, I., FROLOVAITÉ, P., SUOMI, T. and ELO, L. L. (2022). Computational solutions for spatial transcriptomics. *Computational and Structural Biotechnology Journal*, 20, 4870-4884. <https://doi.org/10.1016/j.csbj.2022.08.043>
- KUMAR, K. R., COWLEY, M. J. and DAVIS, R. L. (2019). Next-Generation Sequencing and Emerging Technologies. *Seminars in Thrombosis and Hemostasis*, 45(7), 661-673. <https://doi.org/10.1055/s-0039-1688446>
- LAFRAMBOISE, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research*, 37(13), 4181-4193. <https://doi.org/10.1093/nar/gkp552>
- LAGEVEEN-KAMMEIJER, G. S. M., KUSTER, B., REUSCH, D. and WUHRER, M. (2022). High sensitivity glycomics in biomedicine. *Mass Spectrometry Reviews*, 41(6), 1014-1039. <https://doi.org/10.1002/mas.21730>
- LI, L., FOLEY, J. P. and HELMY, R. (2019). Simultaneous separation of small interfering RNA and lipids using ion-pair reversed-phase liquid chromatography. *Journal of Chromatography A*, 1601, 145-154. <https://doi.org/10.1016/j.chroma.2019.04.061>
- LIN, B., HUI, J. and MAO, H. (2021). Nanopore technology and its applications in gene sequencing. *Biosensors*, 11(7). <https://doi.org/10.3390/bios11070214>
- LOWE, R., SHIRLEY, N., BLEACKLEY, M., DOLAN, S. and SHAFEE, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, 13(5). <https://acortar.link/SHEGdo>
- LUO, R. Y., COMSTOCK, K., DING, C., Wu, A. H. B. and LYNCH, K. L. (2023). Comparison of liquid chromatography-high-resolution tandem mass spectrometry (MS²) and multi-stage mass spectrometry (MS³) for screening toxic natural products. *Journal of Mass Spectrometry and Advances in the Clinical Lab*, 30, 38-44. <https://acortar.link/fnm346>
- MANTERE, T., KERSTEN, S. and HOISCHEN, A. (2019). Long-read sequencing emerging in medical genetics. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00426>

- MARDIS, E. R. (2017). DNA sequencing technologies: 2006-2016. *Nature Protocols*, 12(2), 213-218. <https://doi.org/10.1038/nprot.2016.182>
- MELDRUM, C., DOYLE, M. A. and TOTHILL, R. W. (2011). Next-generation sequencing for cancer diagnostics: a practical perspective. *The Clinical Biochemist Reviews*, 32(4), 177-195.
- MIGNARDI, M. and NILSSON, M. (2014). Fourth-generation sequencing in the cell and the clinic. *Genome Medicine*, 6(4). <https://doi.org/10.1186/gm548>
- MISHRA, M., TIWARI, S. and GOMES, A. V. (2017). Protein purification and analysis: next generation Western blotting techniques. *Expert Review of Proteomics*, 14(11), 1037-1053. <https://doi.org/10.1080/14789450.2017.1388167>
- MOFFITT, J. R., BAMBAH-MUKKU, D., EICHHORN, S. W., VAUGHN, E., SHEKHAR, K., PEREZ, J. D., RUBINSTEIN, N., HAO, J., REGEV, A., DULAC, C. and ZHUANG, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, 362(6416).
- MONTON, M. R., IMAMI, K., NAKANISHI, M., KIM, J. B. and TERABE, S. (2005). Dynamic pH junction technique for on-line preconcentration of peptides in capillary electrophoresis. *Journal of Chromatography A*, 1079(1-2), 266-273. <https://acortar.link/1HFBFa>
- MORAES, F. and GÓES, A. (2016). A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochemistry and Molecular Biology Education*, 44(3), 215-223. <https://doi.org/10.1002/bmb.20952>
- NESBITT, C. A., ZHANG, H. and YEUNG, K. K. (2008). Recent applications of capillary electrophoresis-mass spectrometry (CE-MS): CE performing functions beyond separation. *Analytica Chimica Acta*, 627(1), 3-24. <https://acortar.link/gTiqUv>
- NICHOLSON, J. K., CONNELLY, J., LINDON, J. C. and HOLMES, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery*, 1(2), 153-161. <https://doi.org/10.1038/nrd728>
- Noor, Z., Ahn, S. B., Baker, M. S., Ranganathan, S. and Mohamedali, A. (2021). Mass spectrometry-based protein identification in proteomics -A review. *Briefings in Bioinformatics*, 22(2), 1620-1638. <https://doi.org/10.1093/bib/bbz163>
- OGISO, H., SUZUKI, T. and TAGUCHI, R. (2008). Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Analytical Biochemistry*, 375(1), 124-131. <https://doi.org/10.1016/j.ab.2007.12.027>

- PACCHIAROTTA, T., DERKS, R. J., HURTADO-FERNANDEZ, E., VAN BEZOOIJEN, P., HENNEMAN, A., SCHIEWEK, R., FERNÁNDEZ-GUTIÉRREZ, A., CARRASCO-PANCORBO, A., DEELDER, A. M. and MAYBORODA, O. A. (2013). Online spectral library for GC-atmospheric pressure chemical ionization-ToF MS. *Bioanalysis*, 5(12), 1515-1525. <https://doi.org/10.4155/bio.13.116>
- PAJARES, M. J., PALANCA-BALLESTER, C., URTASUN, R., ALEMANY-COSME, E., LAHOZ, A. and SANDOVAL, J. (2021). Methods for analysis of specific DNA methylation status. *Methods*, 187, 3-12. <https://doi.org/10.1016/j.ymeth.2020.06.021>
- PAREEK, C. S., SMO CZYNSKI, R. and TRETYN, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4), 413-435. <https://acortar.link/BmywV1>
- PERVEZ, M. T., HASNAIN, M. J. U., ABBAS, S. H., MOUSTAFA, M. F., ASLAM, N. and SHAH, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Research International*, (1), 1-12. <https://doi.org/10.1155/2022/3457806>
- PHAM, T. H., ZAEEM, M., FILLIER, T. A., NADEEM, M., VIDAL, N. P., MANFUL, C., CHEEMA, S., CHEEMA, M. and THOMAS, R. H. (2019). Targeting Modified Lipids during Routine Lipidomics Analysis using HILIC, and C30 Reverse Phase Liquid Chromatography coupled to Mass Spectrometry. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-41556-9>
- PIERREL, J. (2012). An RNA Phage Lab: MS2 in Walter Fiers' laboratory of molecular biology in Ghent, from genetic code to gene and genome, 1963-1976. *Journal of the History of Biology*, 45(1), 109-138. <https://doi.org/10.1007/s10739-010-9267-z>
- RASMUSSEN, C. and HOFFMAN, D. W. (2022). Novel Nuclear Magnetic Resonance Method for Position-Specific Carbon Isotope Analysis of Organic Molecules with Significant Impurities. *Analytical Chemistry*, 94(43), 15124-15131. <https://acortar.link/K6bC5F>
- REUTER, J. A., SPACEK, D. V. and SNYDER, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, 58(4), 586-597. <https://doi.org/10.1016/j.molcel.2015.05.004>
- RODRÍGUEZ-UBREVA, J. and BALLESTAR, E. (2014). Chromatin immunoprecipitation. *Methods in Molecular Biology*, 1094, 309-318. <https://acortar.link/eFD9kE>
- SATAM, H., JOSHI, K., MANGROLIA, U., WAGHOO, S., ZAIDI, G., RAWOOL, S., THAKARE, R. P., BANDAY, S., MISHRA, A. K., DAS, G. and MALONIA, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, 12(7). <https://doi.org/10.3390/biology12070997>

- SEMMLER, W. (2018). Die magnetische Kernresonanz: Wegbereiter der Entdeckung [Nuclear Magnetic Resonance: Pioneers of discovery]. *Der Radiologe*, 58(6), 590-594. <https://doi.org/10.1007/s00117-018-0385-5>
- SLATKO, B. E., GARDNER, A. F. and AUSUBEL, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1). <https://doi.org/10.1002/cpmb.59>
- SMITH, C. L. (1991). Separation and analysis of DNA by electrophoresis. *Current Opinion in Biotechnology*, 2(1), 86-91. [https://doi.org/10.1016/0958-1669\(91\)90065-d](https://doi.org/10.1016/0958-1669(91)90065-d)
- SMITH, C. A., O'MAILLE, G., WANT, E. J., QIN, C., TRAUGER, S. A., BRANDON, T. R., Custodio, D. E., Abagyan, R. and Siuzdak, G. (2005). METLIN: a metabolite mass spectral database. *Therapeutic Drug Monitoring*, 27(6), 747-751. <https://acortar.link/RJVfWO>
- SONG, Y., CONG, Y., WANG, B. and ZHANG, N. (2020). Applications of Fourier transform infrared spectroscopy to pharmaceutical preparations. *Expert Opinion on Drug Delivery*, 17(4), 551-571. <https://doi.org/10.1080/17425247.2020.1737671>
- STARK, C., BREITKREUTZ, B. J., REGULY, T., BOUCHER, L., BREITKREUTZ, A. and TYERS, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34, 535-539. <https://doi.org/10.1093/nar/gkj109>
- SWERDLOW, H. and GESTELAND, R. (1990). Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Research*, 18(6), 1415-1419. <https://doi.org/10.1093/nar/18.6.1415>
- SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., JENSEN, L. J. and MERING, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), 607-613. <https://doi.org/10.1093/nar/gky1131>
- TAMARA, S., den BOER, M. A. and HECK, A. J. R. (2022). High-Resolution Native Mass Spectrometry. *Chemical Reviews*, 122(8), 7269-7326. <https://acortar.link/vxt3il>
- TANG, Z., Cao, T., Lin, S., Fu, L., Li, S., Guan, X. Y. and Cai, Z. (2016). Characterization of oncogene-induced metabolic alterations in hepatic cells by using ultrahigh performance liquid chromatography-tandem mass spectrometry. *Talanta*, 152, 119-126.

- THAYSEN-ANDERSEN, M., KOLARICH, D. and PACKER, N. H. (2021). Glycomics and Glycoproteomics: From Analytics to Function. *Molecular Omics*, 17(1), 8-10. <https://acortar.link/bFP8hy>
- THOMAS, K. E., FOTAKI, A., BOTNAR, R. M. and FERREIRA, V. M. (2023). Imaging Methods: Magnetic Resonance Imaging. *Circulation. Cardiovascular Imaging*, 16(1). <https://doi.org/10.1161/CIRCIMAGING.122.014068>
- UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPE, C., SJÖSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S.,... PONTÉN, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220). <https://doi.org/10.1126/science.1260419>
- VALDERRAMA MARTÍN, J. M., ORTIGOSA, F. y CAÑAS, R. A. (2020). Métodos de secuenciación de ácidos nucleicos: primera generación. *Encuentros en la Biología*, 13(173), 19-25.
- VAN DIJK, E. L., JASZCZYSZYN, Y., NAQUIN, D. and THERMES, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, 34(9), 666-681. <https://doi.org/10.1016/j.tig.2018.05.008>
- VAN DIJK, E. L., NAQUIN, D., GORRICHON, K., JASZCZYSZYN, Y., OUAZAHROU, R., THERMES, C. and HERNANDEZ, C. (2023). Genomics in the long-read sequencing era. *Trends in Genetics*, 39(9), 649-671. <https://doi.org/10.1016/j.tig.2023.04.006>
- VAN STEENSEL, B. (2005). Mapping of genetic and epigenetic regulatory networks using microarrays. *Nature Genetics*, 37, 18-24. <https://doi.org/10.1038/ng1559>
- VESTAL, M. L. and CAMPBELL, J. M. (2005). Tandem time-of-flight mass spectrometry. *Methods in Enzymology*, 402, 79-108. [https://doi.org/10.1016/S0076-6879\(05\)02003-3](https://doi.org/10.1016/S0076-6879(05)02003-3)
- WANG, Y. and NAVIN, N. E. (2015). Advances and applications of single-cell sequencing technologies. *Molecular Cell*, 58(4), 598-609. <https://acortar.link/Mlj50x>
- WANG, Y., ZHAO, Y., BOLLAS, A., WANG, Y. and AU, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- WANG, Z., GERSTEIN, M. and SNYDER, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>

- WECKWERTH, W. (2011). Green systems biology—From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *Journal of Proteomics*, 75(1), 284-305. <https://doi.org/10.1016/j.jprot.2011.07.010>
- WEISS, W. and GÖRG, A. (2009). High-resolution two-dimensional electrophoresis. *Methods in Molecular Biology*, 564, 13-32. https://doi.org/10.1007/978-1-60761-157-8_2
- WILHELM, M., SCHLEGL, J., HAHNE, H., GHOLAMI, A. M., LIEBERENZ, M., SAVITSKI, M. M., ZIEGLER, E., BUTZMANN, L., GESSULAT, S., MARX, H., MATHIESON, T., LEMEER, S., SCHNATBAUM, K., REIMER, U., WENSCHUH, H., MOLLENHAUER, M., SLOTTA-HUSPENINA, J., BOESE, J. H., BANTSCHIEFF, M.,... KUSTER, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502), 582-587. <https://doi.org/10.1038/nature13319>
- WILKINS, M. R., SANCHEZ, J. C., GOOLEY, A. A., APPEL, R. D., HUMPHERY-SMITH, I., HOCHSTRASSER, D. F. and WILLIAMS, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews*, 13, 19-50. <https://acortar.link/OcwCZP>
- WINTER, D. L., WILKINS, M. R. and DONALD, W. A. (2019). Differential Ion Mobility-Mass Spectrometry for Detailed Analysis of the Proteome. *Trends in Biotechnology*, 37(2), 198-213. <https://doi.org/10.1016/j.tibtech.2018.07.018>
- WISHART, D. S., FEUNANG, Y. D., MARCU, A., GUO, A. C., LIANG, K., VÁZQUEZ-FRESNO, R., SAJED, T., JOHNSON, D., LI, C., KARU, N., SAYEEDA, Z., LO, E., ASSEMPOUR, N., BERJANSKII, M., SINGHAL, S., ARNDT, D., LIANG, Y., BADRAN, H., GRANT, J.,... SCALBERT, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, 46(D1), 608-617. <https://doi.org/10.1093/nar/gkx1089>
- YEGNASUBRAMANIAN, S. (2013). Preparation of fragment libraries for next-generation sequencing on the applied biosystems SOLiD platform. *Methods in Enzymology*, 529, 185-200. <https://doi.org/10.1016/B978-0-12-418687-3.00015-X>
- ZAPPIA, L. and THEIS, F. J. (2021). Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biology*, 22. <https://doi.org/10.1186/s13059-021-02519-4>
- ZÜLLIG, T. and KÖFELER, H. C. (2021). High Resolution Mass Spectrometry in Lipidomics. *Mass Spectrometry Reviews*, 40(3), 162-176. <https://doi.org/10.1002/mas.21627>

GLOSARIO

5'RACE:	Amplificación rápida de los extremos del ADNc.
ADN:	Ácido desoxirribonucleico.
ADNc	ADN complementario.
ARN:	Ácido ribonucleico.
ARNm:	ARN mensajero.
ARNr:	ARN ribosómico.
ARN-Seq:	Secuenciación de ARN.
ARNt:	ARN de transferencia.
BAC:	Cromosomas artificiales bacterianos.
BGI:	Grupo del Instituto de Genómica de Beijing.
CAGE:	Análisis Cap de la expresión génica.
cCRE:	Elementos de reguladores en Cis candidatos.
CE:	Electroforesis capilar.
CE-MS:	Electroforesis capilar y espectrometría de masas.
CGH:	Hibridación genómica comparada.
ChIP-seq:	Inmunoprecipitación de la cromatina acoplada a secuenciación.
ChromHMM:	<i>Software</i> para caracterizar los estados de la cromatina.
CIDR:	Centro para la Investigación de Enfermedades Hereditarias.
CISH:	Detección cromogénica.
cM:	CentiMorgan.

CTCF:	Represor transcripcional CTCF, proteína de dedo de zinc 11 o factor de unión a CCCTC.
dbSNP:	Base de datos públicos de SNP.
DDBJ:	Banco de datos del Japón.
ddNTP:	Didesoxinucleótidos.
DEEP:	Consortio para el Estudio del Impacto de los Factores Ambientales en el Epigenoma y el Proteoma.
DIR:	División de Investigación Intramural.
DNase-seq:	Método para identificar la ubicación de regiones reguladoras, basado en la secuenciación de todo el genoma de regiones sensibles a la escisión por DNase I.
DNB:	Secuenciación de nanoesferas de ADN.
dNTP	Desoxinucleótidos.
DOE:	Departamento de Energía de los Estados Unidos.
emPCR:	PCR en emulsión.
ENA:	Archivo de Nucleótidos Europeos.
ENCODE:	Consortio Enciclopedia de Elementos de ADN.
FACS:	Clasificación de células activadas por fluorescencia.
FISH:	Fluorescencia hibridación <i>in situ</i> .
FRLP:	Polimorfismos de longitud de los fragmentos de restricción.
FunSeq:	Herramienta especializada en priorizar variantes somáticas de la secuenciación del genoma completo del cáncer.
GENCODE:	Anotación de genes codificadores.
GWAS:	Estudio de asociación de genoma completo.
HaploReg:	Explorar anotaciones del genoma no codificante en variantes en bloques de haplotipos.
HapMap:	Mapa de haplotipos humanos.

HPP-BI:	Iniciativa Cerebro del Proyecto Proteoma Humano.
HPRC:	El Consortium Reference de Pangenoma Humano.
IP:	Inmunoprecipitación.
ISH:	Hibridación <i>in situ</i> .
kb	kilobases.
KEGG:	Enciclopedia de genes y genomas de Kioto.
LBHN:	Proyecto de red biológica pulmonar del Atlas Celular Humano.
MERFIS:	Técnica altamente multiplexada y avanzada derivada del smFISH,
MPSS:	Secuenciación masiva de firmas en paralelo.
MS:	Espectrometría de masas.
MVP:	Posiciones variables de metilación.
NGS:	Secuenciación de Nueva Generación.
NHGRI:	Instituto Nacional de Investigación del Genoma Humano.
NIAID:	Instituto Nacional de Alergias y Enfermedades Infecciosas.
NIH: Institutes	Instituto Nacional de Salud (por sus siglas en inglés National of Health).
PCR:	Reacción en cadena de la polimerasa.
PGH:	Proyecto Genoma Humano.
pH:	Potencial de hidrógeno.
PTM:	Modificación postraducciona de proteínas.
RT-PCR:	PCR en tiempo real.
RT-qPCR:	Retrotranscripción y PCR cuantitativa.
SAGE:	Análisis en serie de la expresión génica.
SBS:	Secuenciación por síntesis.
scARN-seq:	Secuenciación de ARN unicelular.
SCREEN:	Buscador de elementos reguladores candidato ENCODE.

SEGWAY:	Paquete de <i>software</i> para analizar múltiples pistas de datos de genómica funcional.
SLAM-seq:	Etiquetado y multiplexación de una sola célula mediante secuenciación.
smFISH	FISH de molécula única.
SMRT:	Single Molecule Real-Time Sequencing.
SNP	Polimorfismos de un solo nucleótido.
SOLiD:	Secuenciación por ligación y detección de oligonucleótidos.
Tag SNP:	SNP etiqueta.
tARN:	Ácido desoxirribonucleico de transferencia.
TLC:	Cromatografía en capa fina.
TOF-MS:	Espectrometría de masas basada en tiempo de vuelo.
UCSC:	Universidad de California, Santa Cruz.
WE:	Exoma complete.
WG:	Genoma completo.
YAC:	Cromosomas artificiales de levaduras.
ZMW:	Secuenciación de guía de onda de modo cero.

Esta obra
se terminó de imprimir
el 29 de septiembre de 2025,
en los talleres gráficos de
GRUPO EDITORIAL IBÁÑEZ,
Cra. 69 Bis No. 36-20 Sur
Tels.: 2300731 - 2386035
Bogotá D.C. - Colombia

Resulta fascinante comprender los conceptos fundamentales de las ómicas y las diferentes tecnologías empleadas en sus análisis. Estos conocimientos permitirán a los lectores adquirir una comprensión más amplia de las aplicaciones de las ciencias ómicas en áreas como la biotecnología y disciplinas afines.

En el primer capítulo, se abordan una amplia variedad de temas fundamentales en estas disciplinas emergentes, desde su introducción hasta la exploración de conceptos clave como la genómica, la proteómica, la transcriptómica y la metabolómica. Se ofrece un detallado análisis de la evolución histórica de las ciencias ómicas, cubriendo desde sus primeros avances hasta las eras preómica y ómica, además de examinar proyectos innovadores en desarrollo.

El segundo capítulo aborda una amplia gama de herramientas y métodos utilizados en el estudio de disciplinas relacionadas con las ciencias ómicas. Comienza con una introducción que establece el contexto y la importancia de estas técnicas. Luego, se detallan las técnicas en genómica, que incluyen diferentes generaciones de secuenciación y microarreglos. A continuación, se exploran las técnicas en transcriptómica, abarcando desde enfoques basados en hibridación hasta métodos avanzados como ARN-seq y transcriptómica espacial.

