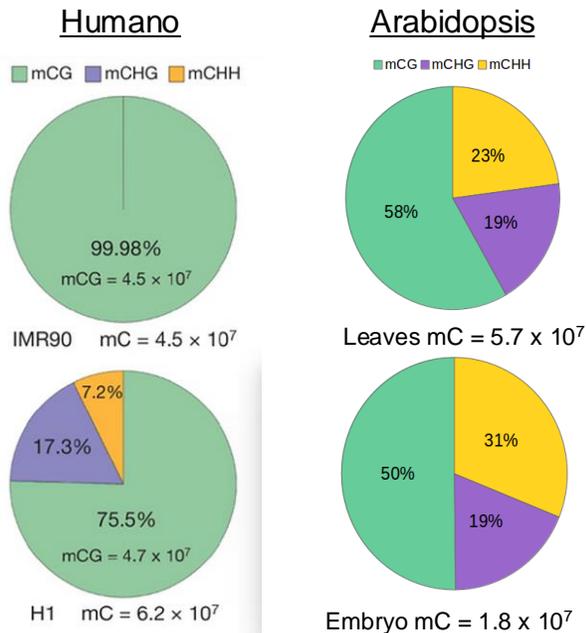
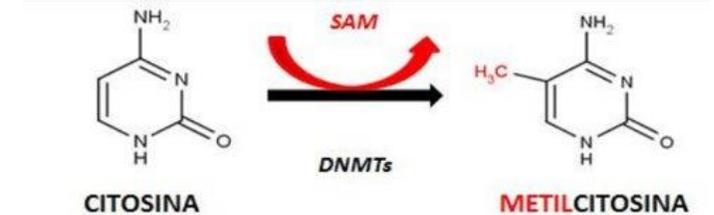


# Análisis metilación del ADN

Guillermo Barturen Briñas  
([gbarturen@ugr.es](mailto:gbarturen@ugr.es))

# ¿Qué es un metiloma?

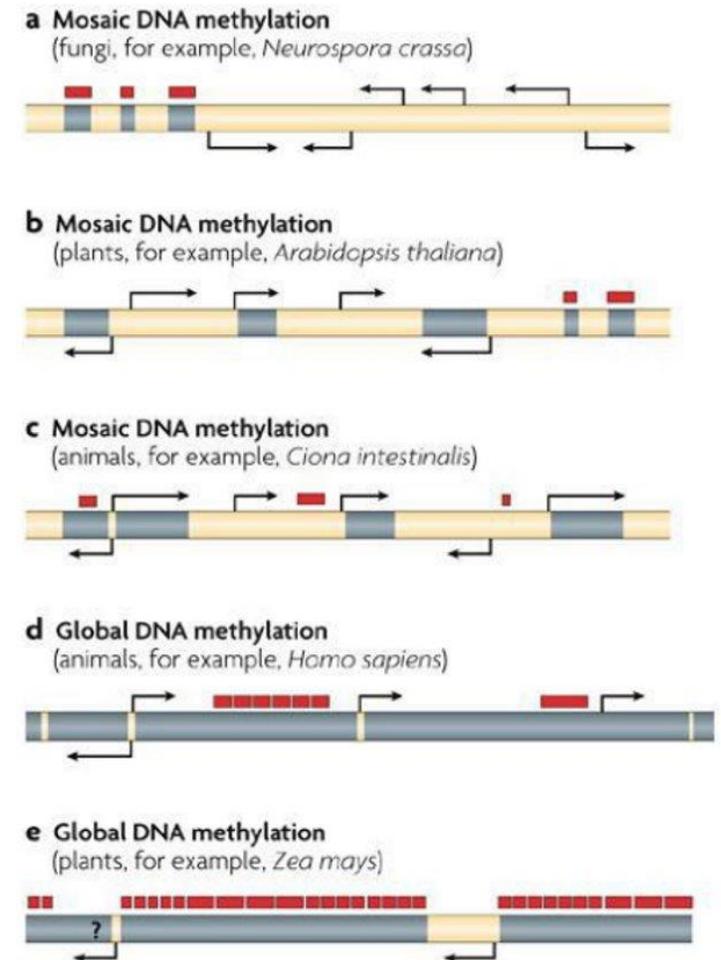
- El metiloma de una célula se puede definir como la localización de metil-citosinas a lo largo de la cadena de ADN (patrón de metilación).
- Este patrón de metilación es característico de cada tipo celular y/o condición fisiopatológica.
- En eucariotas sólo ciertas regiones del genoma son susceptibles de metilarse. El estudio de estos patrones puede revelar información acerca del **potencial transcripcional** de ciertas regiones o de la **interacción de otras moléculas con el ADN**.



- Aunque los contextos de metilación en eucariotas son similares debido a la alta conservación de los dominios catalíticos de las metiltransferasas, la distribución de los grupos metilo en las secuencias diana es diferente entre especies y tejidos dentro de una misma especie. La secuencia con mayor porcentaje de grupos metilo invariablemente de la especie es **CpG**, mientras que otros contextos como CpHpG o CpHpH se han observado aumentados en plantas y tejidos/células no diferenciadas. En plantas por ejemplo, la metilación en contextos no CpG se ha visto asociada a transposones silenciados.

# ¿Qué es un metiloma?

- En diferentes especies podemos encontrar patrones de metilación muy diversos, distinguiendo dos principales en organismos eucariotas:
  - Patrones mosaicos:** donde sólo ciertas regiones del ADN presentan grupos metilo, como cuerpos génicos o elementos transponibles.
  - Patrones globales:** la mayoría de las regiones susceptibles de ser metiladas se encuentran metiladas, mientras que sólo algunas carecen de grupos metilo. Entre estos podemos diferenciar el patrón global en animales y en plantas, el primero se ha hipotetizado que puede haber surgido como mecanismo de defensa ante procesos autoinmunes (TLR9, reconoce sitios CpG no metilados como defensa a infecciones), mientras que el segundo suele darse en plantas con genomas grandes con un alto porcentaje de elementos transponibles.



# Análisis masivo de metilomas

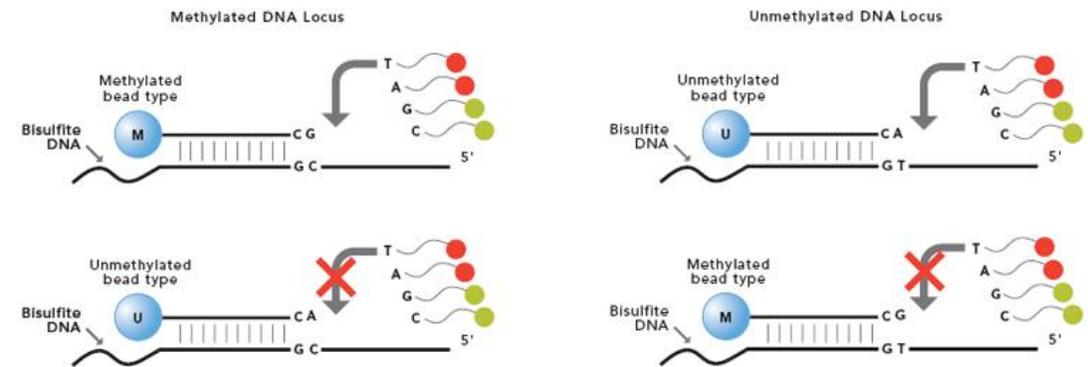
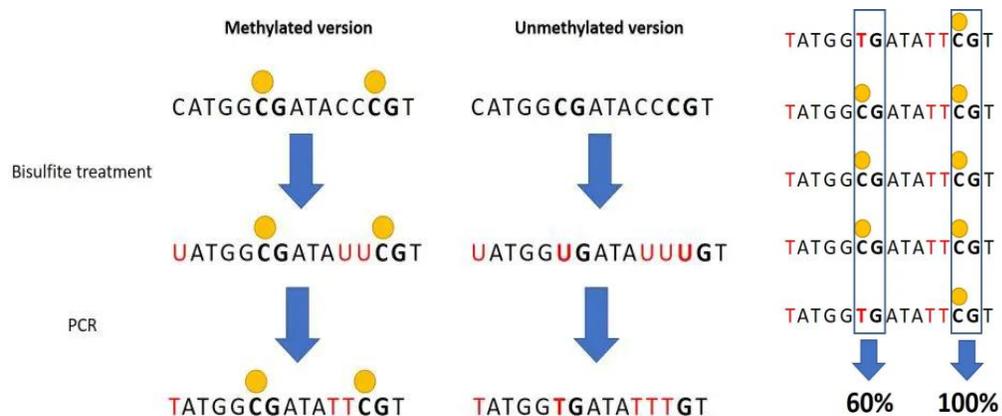
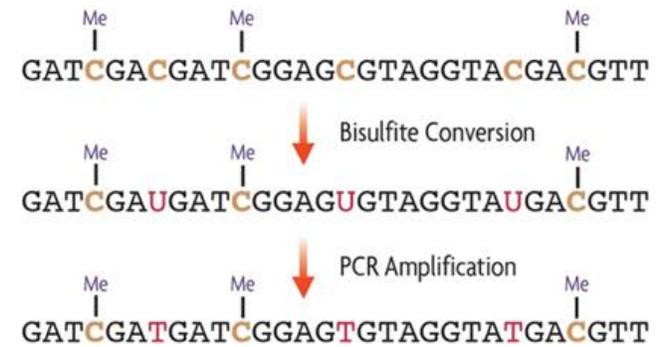
- La metilación del ADN es eliminada durante los procesos estándar de biología molecular, como el clonaje en bacterias o la amplificación mediante PCR, y no puede ser identificada mediante hibridación ya que los grupos metilo se localizan en el surco mayor del ADN.
- Por lo tanto, a lo largo de los años se han ido desarrollando numerosos **pretratamientos del ADN** que posteriormente permitan revelar su presencia o ausencia en las citosinas, como la digestión enzimática, el enriquecimiento por afinidad o la conversión por bisulfito, seguidos de procedimientos de cuantificación como la hibridación en microarrays o secuenciación entre otros.

Technology	Features							Potential sources of bias							
	Unambiguous identification of CpG measured	In cis co-methylation information	Non-CpG methylation information	Allele-specific measurement capability	Good coverage of regions with low CpG density	Compatible with low amounts of input DNA	Full repeat-masked genome coverage	Copy-number variation bias	Fragment size bias	Incomplete bisulphite conversion bias	Bisulphite PCR bias	Cross-hybridization bias	DNA methylation status bias	GC content bias	CpG density bias
Infinium	(•)					•				•	•	•			
Enzyme-chip	(•)	(•)			(•)			•				•		•	
MeDIP-chip							•					•		•	•
BSPP	•	•	•	•						•	•		•		
BC-seq	•	•	•	•						•	•		•		
RRBS	•	•	•	•		•				•	•				
Enzyme-seq	•	•		•	(•)	•		•							
MeDIP-seq				•			•	•						•	•
WGSBS	•	•	•	•	•	•	•			•	•				

\*• indicates that the method has this feature or potentially has this bias; \*(•) indicates that the method has this feature to a limited extent or in some circumstances. BC-seq, bisulphite conversion followed by capture and sequencing; BSPP, bisulphite padlock probes; -chip, followed by microarray; MeDIP, methylated DNA immunoprecipitation; RRBS, reduced representation bisulphite sequencing; -seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

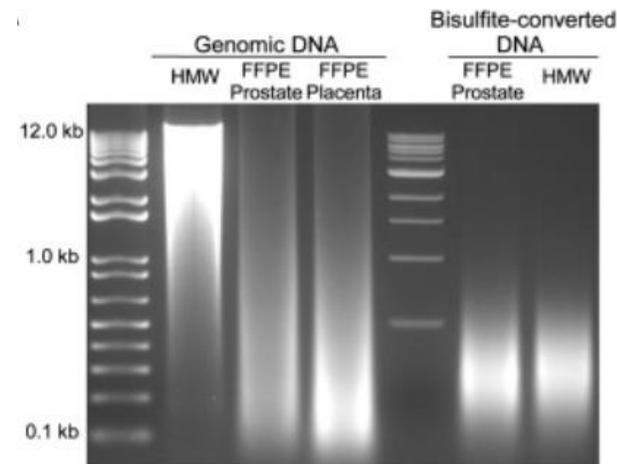
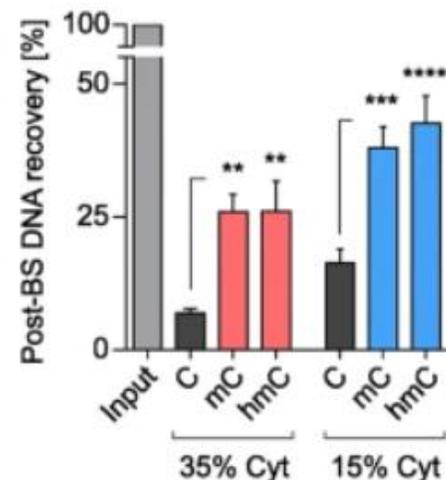
# Análisis masivo de metilomas

- Actualmente el pretratamiento más utilizado para interrogar el estado de metilación de las citosinas de un genoma es la **conversión por bisulfito**. Este pretratamiento se basa en la incubación de una muestra de ADN con bisulfito de sodio, el bisulfito induce la **desaminación espontánea de las citosinas no metiladas**, convirtiéndolas en uracilos que posteriormente durante la amplificación del ADN mediante PCR serán identificados como timinas.
- Una vez generados los cambios artificiales en las secuencias de ADN, el estado de metilación puede ser interrogado mediante múltiples métodos. Actualmente, los más utilizados son la hibridación del ADN en microarrays (similar a los microarrays de variantes de secuencia) y la secuenciación masiva.



# Sesgos en el pretratamiento con bisulfito

- Los sesgos que se introducen en la secuencia durante el tratamiento con bisulfito pueden provenir de:
  - i. La **degradación que produce el bisulfito durante la conversión del ADN** resulta en la pérdida de un porcentaje importante del material de partida (optimización importante). Además, es difícil de detectar lecturas largas una vez convertidas, ya que tras la conversión una gran mayoría de los fragmentos se encuentran entorno a las 500 Mbs.
  - ii. El bisulfito presenta a su vez una **tasa de conversión variable**, que suele ser controlada mediante la inclusión de muestras no-metiladas control en el proceso. Además, en ocasiones este sesgo es más acentuado en los extremos de los fragmentos de ADN.



# Sesgos en el proceso de cuantificación

- Dependiendo del método de cuantificación que utilicemos también debemos tener en cuenta una serie de factores que pueden influir en los niveles de metilación.

## Microarrays

- Conocimiento previo
- Limitados a metilación en CpGs
- Interroga ~ 1M CpGs (EPIC Illumina)
- Sesgos en la amplificación por PCR (bisulfito)
- Hibridación cruzada
  
- Permite cuantificación de bajas cantidades de ADN (min. ~250ng)
- Información por nucleótido
- Coste asequible para grandes cohortes

## Secuenciación

- Conocimiento previo
- Sesgos en la amplificación por PCR (bisulfito)
- Elevado coste por muestra
- Análisis computacional demandante
- Alineamiento con alfabeto reducido
  
- Permite cuantificación de bajas cantidades de ADN (min. ~50-200ng)
- Información por nucleótido
- Metiloma completo, todos los contextos
- Información alelo específica

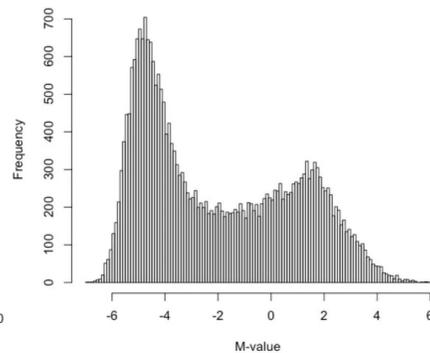
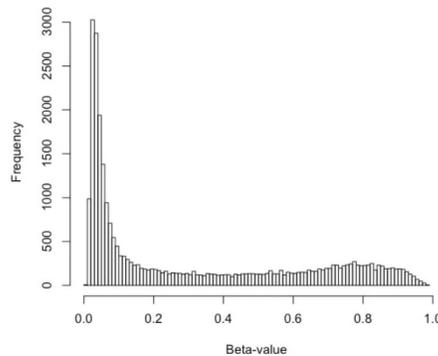
# Cuantificación de la metilación

- La cuantificación en microarrays se basa en la relación entre las intensidades de las sondas con C (Metiladas) y T (No-Metiladas). En secuenciación se basa en la cuantificación de las lecturas frente a la referencia, donde C/C (Metilada) y T/C (No-Metilada).



$$M_i = \log_2 \left( \frac{\max(y_{i,methy}, 0) + \alpha}{\max(y_{i,unmethy}, 0) + \alpha} \right)$$

$$Beta_i = \frac{\max(y_{i,methy}, 0)}{\max(y_{i,unmethy}, 0) + \max(y_{i,methy}, 0) + \alpha}$$



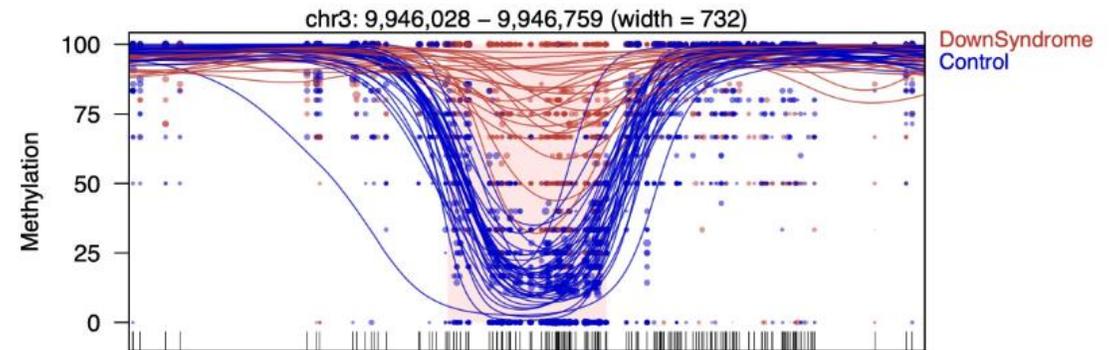
<https://doi.org/10.1186/1471-2105-11-587>

$$Beta_i = \frac{2^{M_i}}{2^{M_i} + 1}; M_i = \log_2 \left( \frac{Beta_i}{1 - Beta_i} \right)$$

- Suelen utilizarse dos medidas para cuantificar la metilación:
  - M-Values:** Siguen una distribución pseudo-normal con una escala infinita.
  - B-Values:** Presenta una escala limitada de 0-1 y es fácilmente interpretable.
- Ambas medidas son interconvertibles, usándose los M-values para comprobaciones estadísticas ya que su distribución se asemeja a una distribución normal y los B-Values para representar e interpretar gráficamente los resultados.

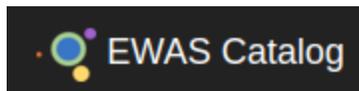
# DMRs (Regiones diferenciales de metilación)

- Al abordar análisis diferenciales de metilación en microarrays suele considerarse que las posiciones interrogadas son independientes. Sin embargo, al igual que ocurre con las variantes de secuencia esta asunción no siempre es cierta. En el caso de la secuenciación masiva, donde se interrogan virtualmente todas las posiciones del genoma esta asunción es incluso más arriesgada.
- El mantenimiento de la metilación ocurre de manera estocástica en zonas accesibles por las metiltransferasas, por lo que en regiones del ADN donde se están produciendo interacciones con el ADN la **metilación se modifica en numerosos sitios agrupados y co-regulados**, por lo tanto no son independientes.
- Para evitar esta asunción, los análisis diferenciales en datos de secuenciación masiva suelen abordarse desde la perspectiva de **DMRs (regiones diferencialmente metiladas)**.
- Este no es un problema trivial y existen numerosos métodos para identificarlas, que tienen en cuenta tanto su longitud, como el número de CpGs que contienen, la distancia entre regiones y por supuesto los niveles de metilación. Además, los límites de estas regiones son difusas presentando diferencias máximas en el centro de la región y un gradiente de diferencias en sus bordes lo que dificulta su identificación.



# Funcionalidad de la metilación diferencial

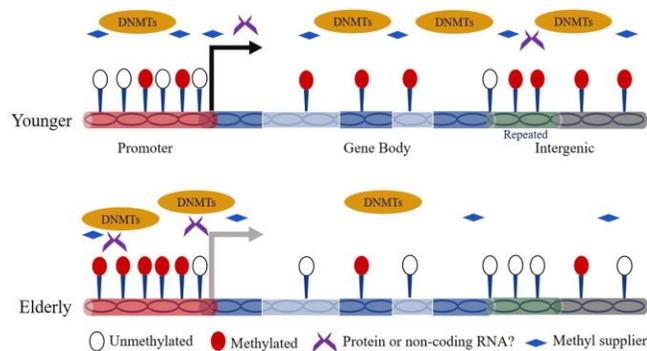
- Al igual que ocurre con otros datos moleculares diferentes a la abundancia diferencial de genes y/o proteínas, la interpretación de los cambios de metilación no siempre tienen una interpretación directa, por lo que se requiere de análisis adicionales para interpretar los resultados.
  - La **asociación por proximidad** (co-localización) entre la metilación y los genes que pueda estar regulando es una aproximación ingenua, aunque válida ya que muchos de los cambios de metilación afectan de manera local.
  - **Identificación de motivos de unión** en las regiones donde se producen los cambios para identificar los factores de transcripción que se encuentran interaccionando de manera diferencial.
  - **Asociación con otros datos moleculares** como expresión de genes o genotipos.
  - **Información previa** de los cambios de metilación (EWAS Catalog).



<http://www.ewascatalog.org/>

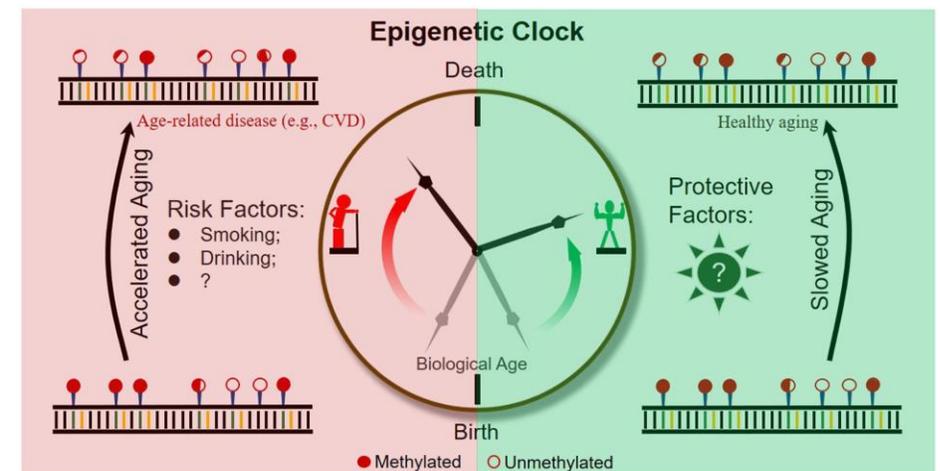
# La metilación como reloj epigenético

- La metilación es una marca que debe ser mantenida en las sucesivas divisiones celulares. La edad y ciertos factores ambientales pueden provocar cambios en su mantenimiento lo que conlleva cambios específicos que son fácilmente detectables y **correlacionan relativamente bien con la edad cronológica de los individuos**.



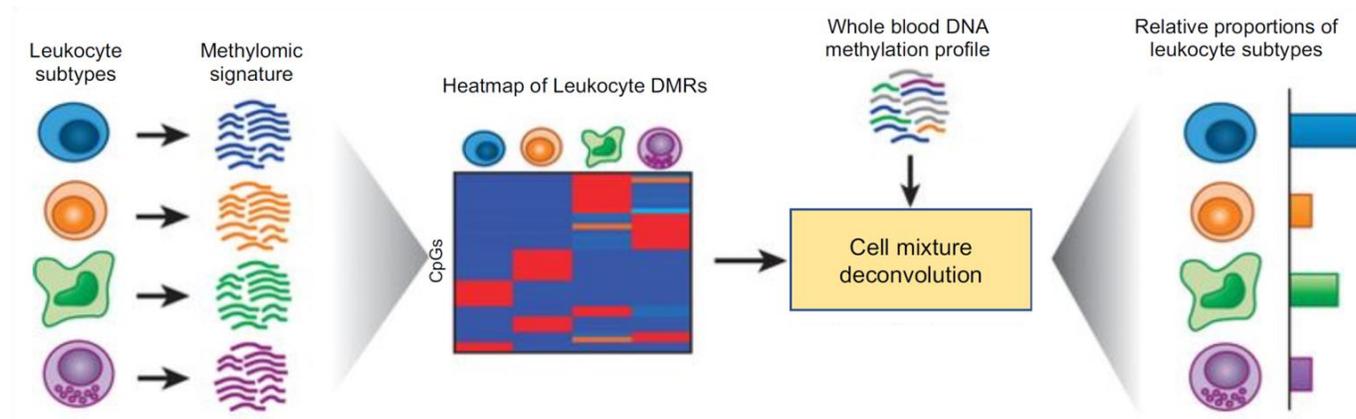
- En general, con la edad se ha identificado una tendencia a cambios de metilación que inducen una menor expresión de genes con hipermetilaciones en promotores e hipometilaciones en sus cuerpos génicos.

- Durante los últimos años, en base a estos patrones de metilación se han entrenado modelos estadísticos para predecir lo que se denomina **edad biológica de los individuos**. Por lo que comparando la edad biológica y la edad cronológica de los individuos podemos encontrar individuos con un **envejecimiento biológico acelerado o decelerado**. Debemos tener en cuenta que estos modelos son tejido/muestra específica, pudiendo encontrar individuos con edades biológicas diferentes en función del tejido analizado.



# La metilación como marcador celular

- La metilación es una marca relativamente estable y con **patrones específicos dependientes del tipo celular** y otras condiciones fisiopatológicas. La complejidad celular de muestras heterogéneas como muestras de sangre o tumorales limita la interpretación funcional de los resultados, ya que las señales obtenidas difícilmente pueden asociarse a tipos celulares concretos.
- A lo largo de los últimos años, se ha aprovechado que la **metilación refleja el proceso de diferenciación de los diferentes subtipos celulares inmunes** desde células madre de la médula ósea para de-convolucionar estas señales heterogéneas de metilación en los tipos celulares que la componen. En base a esta premisa se han entrenado numerosos modelos estadísticos que actualmente nos permiten determinar la composición celular de muestras complejas e inferir los tipos celulares donde se producen los cambios de metilación.



# La metilación como marcador de daño tisular

- Se conoce como ADN circulante a las moléculas/fragmentos de ADN liberadas al torrente sanguíneo mediante procesos de apoptosis, necrosis o secreción activa. Este **ADN circulante mantiene las marcas de metilación específicas del tipo celular o de la condición del tejido de donde se liberan**, por lo que se están empezando a utilizar como biomarcadores para el seguimiento del daño tisular provocado por diversas enfermedades, principalmente tumores.

- La identificación de patrones de metilación específicos en los tejidos afectados y su identificación en muestras de ADN circulante puede ser en un futuro próximo una **herramienta no invasiva** que nos permita: hacer seguimientos de pacientes para medir su respuesta a los tratamientos, identificar daño tisular de manera temprana, identificar de manera específica el origen primario de tumores o evaluar el pronóstico y la evolución de una enfermedad.

