

# Análisis multi-ómico con RAMEN: identificando interacciones gen-ambiente en la metilación del ADN

**Autores:** Erick I. Navarro-Delgado<sup>1,2</sup>, Michael S. Kobor<sup>1,2</sup>, Keegan Korthauer<sup>1,2</sup>.

(1) Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, Vancouver, BC, Canadá. (2) The University of British Columbia, Vancouver, BC, Canadá.

**Palabras clave:** Metilación del ADN, interacción gen-ambiente, multi-ómica

## Resumen

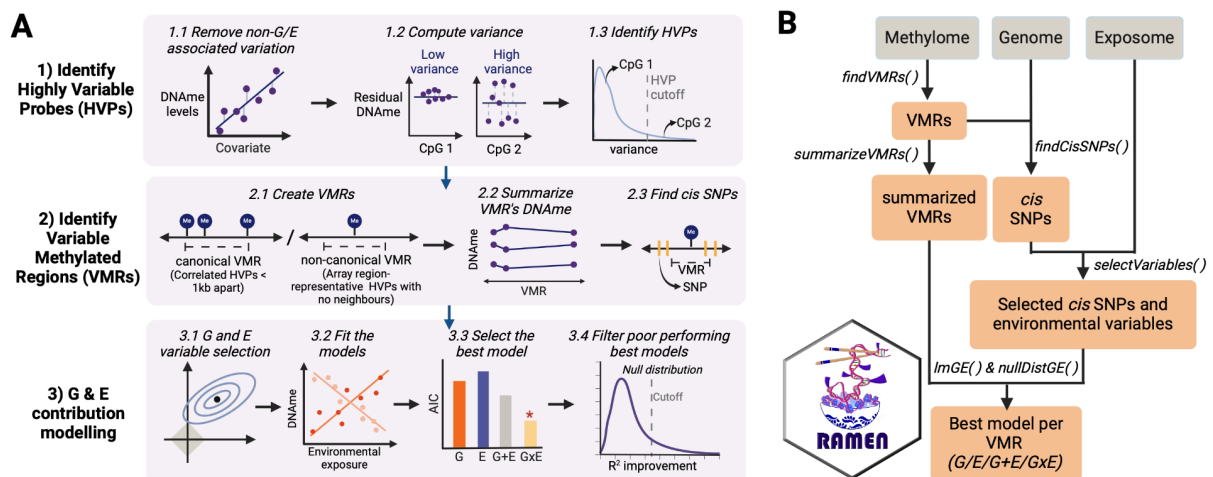
**INTRODUCCIÓN:** El periodo prenatal es una etapa altamente sensible en el desarrollo humano. En este periodo puede ocurrir la “programación fetal”, un fenómeno en el que exposiciones ambientales pueden desencadenar mecanismos biológicos duraderos que alteran la salud en etapas postnatales. Uno de los procesos moleculares que podrían mediar este fenómeno es la metilación del ADN (meADN). Esta marca molecular, definida como el depósito de un grupo metilo a la molécula de ADN, es capaz de alterar mecanismos celulares como lo es la expresión de genes. Acorde a su relevancia en la programación fetal, estudios recientes han asociado principalmente la variación de la meADN con diferencias genéticas (G) y ambientales (environment; E). Sin embargo, aún no queda claro 1) en qué partes del mapa genómico existe una mayor variabilidad de meADN en etapas de vida temprana, 2) con qué frecuencia los factores ambientales y genéticos contribuyen a esta variabilidad, y 3) si las diferencias genéticas y ambientales se asocian con esta variabilidad de manera individual, en conjunto, o mediante interacciones. Para resolver esta pregunta, desarrollamos RAMEN (Regional Association of Methylome variability with the Exposome and geNome; <https://ericknavarro.github.io/RAMEN>), un paquete en R que identifica regiones de ADN con alta variabilidad de meADN (Variable Methylated Region; VMR) y estima la contribución de las diferencias ambientales y genéticas a la meADN en cada región.

**MÉTODO/RESULTADOS:** En este proyecto, utilizamos RAMEN (**Figura 1**) en muestras de sangre del cordón umbilical de CHILD y PREDO, dos estudios nacionales de Canadá y Finlandia (n total = 1,662) para integrar datos multi-ómicos del genoma (conjunto de ADN; 1,260,703 variantes genéticas), exposoma (conjunto de exposición ambiental; 94 variables) y metiloma (conjunto de sitios con meADN; 785,689 sitios). Identificamos 28,480 VMRs en CHILD y 17,578 VMRs en PREDO, utilizando criterios de proximidad genómica (1 kilobase), varianza (90° percentil) y correlación en meADN ( $r > 0.15$ ). Para reducir el universo de posibles modelos estadísticos a comparar, RAMEN implementa una selección de variables para cada VMR basado en LASSO, un algoritmo de aprendizaje de máquina. Posteriormente, nuestro paquete ajusta modelos lineares

# ANÁLISIS MULTI-ÓMICO CON RAMEN: IDENTIFICANDO INTERACCIONES GEN-AMBIENTE EN LA METILACIÓN DEL ADN

univariables G y E, además de bivariables aditivos (G+E) y de interacción (GxE). El mejor modelo por cada región genómica se escoge posteriormente utilizando su Akaike Information Content (AIC). Finalmente, para remover asociaciones originadas por azar, implementamos un método computacionalmente rápido basado en permutaciones.

**CONCLUSIÓN:** Con un porcentaje alto de modelos seleccionados (59%), una mayor proporción de varianza atribuida (promedio = 0.22), y una consistencia total en ambos estudios (100%), identificamos a la genética como el mayor contribuidor a la variabilidad de la meADN, usualmente en combinación con el ambiente prenatal (44.4%). Nuestro estudio demuestra la relevancia de la genética en los patrones divergentes de meADN en la vida temprana, y ayuda a entender los factores que pueden estar involucrados en la programación fetal. Adicionalmente, demostramos la utilidad de RAMEN para extraer patrones de variabilidad regional en microarreglos de meADN y modelar sus fuentes de variabilidad removiendo asociaciones azarasas a través de técnicas estadísticas de vanguardia. Este paquete puede utilizarse gratuitamente para hacer análisis de contribución gen-ambiente en datos de meADN de forma robusta.



**Figura 1. Metodología implementada en RAMEN.** A) Flujo de análisis estadístico implementado en el paquete. B) Mapa conceptual de los input y outputs de las funciones de RAMEN.