Biopsia líquida y epigenética

Autores:

Víctor M.Lafuente, 747325

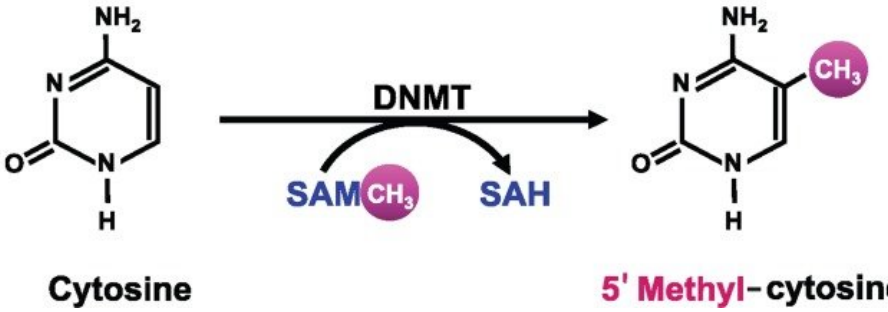
Alonso Muñoz García, 745016

Jose M.Romero, 740914

Universidad Zaragoza

Metilación del ADN

- La metilación es la adición de un grupo metilo (-CH3) a una molécula.
- Se observa tanto en el ADN como en proteínas.



Funciones de la Metilación

- La metilación es clave en el desarrollo embrionario y otras funciones vitales para el ser humano.
- Los grados de metilación influye en los niveles de expresión de un gen.
- La interacción entre la metilación de DNA y varias proteínas determinan el estado de la cromatina y están así implicados en la regulación génica (**cambios epigenéticos**).

Epigenética: ¿Qué es?

Mecanismo de regulación de la expresión de los genes sin modificar la secuencia de ADN.

- Establece relación entre influencias genéticas y ambientales que determinan un fenotipo.
- Las regulaciones genéticas son debidas a la metilación y modificación de histonas.
- Diversos estudios sobre la relación de la metilación y expresión genética y predicción de cáncer

Aberraciones de metilación del ADN

- Existen dos tipos de aberraciones
  - Hipometilación: Escasez de metilación
  - Hipermetilación: Exceso de metilación
- Se les relaciona con la desregularización de expresión gen.
  - activación o silenciamiento de genes
- Pueden generar:
  - Sobreexpresión o supresión de genes
  - Generar interacciones de proteínas con causas adversas

Tumores

- La desregularización epigenética puede provocar tumores
  - Por silenciamiento de genes supresores tumorales
  - Generar comportamientos cancerígenos en las células incluso antes de ser visibles

Métodos de extracción de Metilación del ADN

Requiere usar químicos para eliminar proteínas enlazadas y modificaciones en histona

- Chip on Chip
- Chip-Seq
- Infinium HumanMethylation450 BeadChip

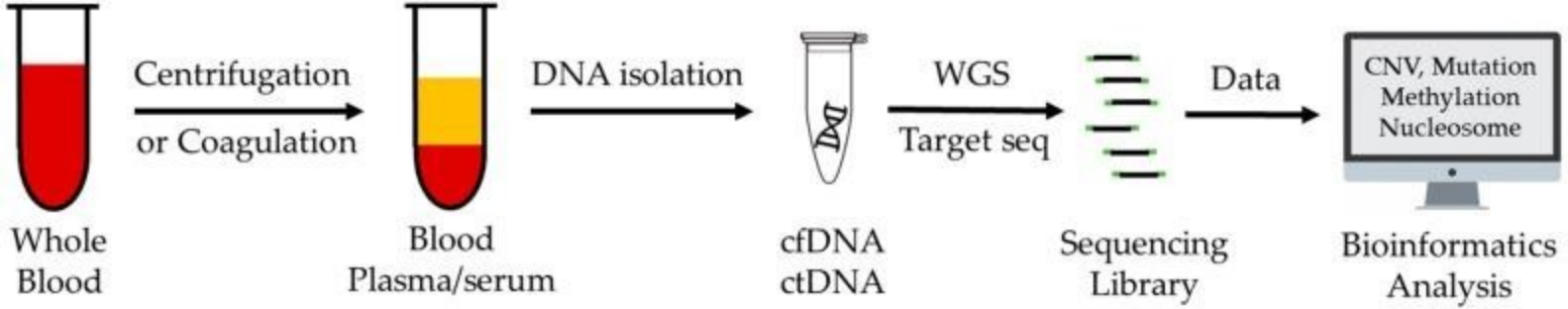
Detección de cáncer

- Antes se realizaba mediante análisis de tejido
- Análisis de tejidos presenta ciertas limitaciones
  - Disponibilidad
  - Heterogeneidad

- Solución: **BIOPSIA LÍQUIDA**

Biopsia líquida: ¿Qué es?

- Análisis molecular del ADN extracelular (cfDNA)
  - Se puede detectar en plasma (sangre), orina, heces...
  - Se obtiene coagulando la sangre o



Biopsia líquida:Ventajas

El análisis de cfDNA presenta ventajas respecto al análisis de tejido:

- Son no invasivas
- Capacidad de muestreo múltiple
- Cobertura integral para abordar problemas de heterogeneidad tumoral

Análisis del ctDNA

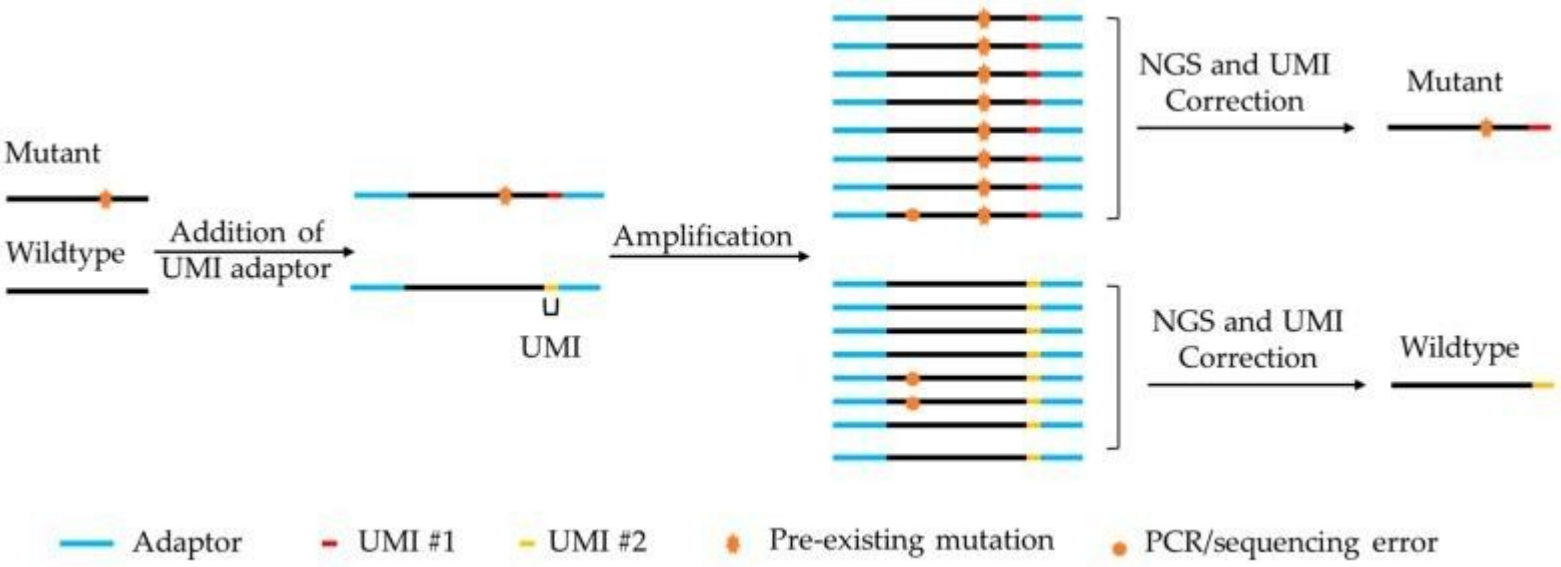
- Tipo concreto de cfDNA->Proviene de células tumorales
- Ayuda a detectar enfermedades tumorales
- Puede enriquecerse mediante eliminación de fragmentos más pequeños de cfDNA

Límites de análisis de ctDNA

- Fracción pequeña respecto del cfDNA ->Solución:Next Gen Sequencing
- Errores de amplificación por PCR

  -> Solución:Uso de UMIs
- No tiene aplicación clínica

UMIs

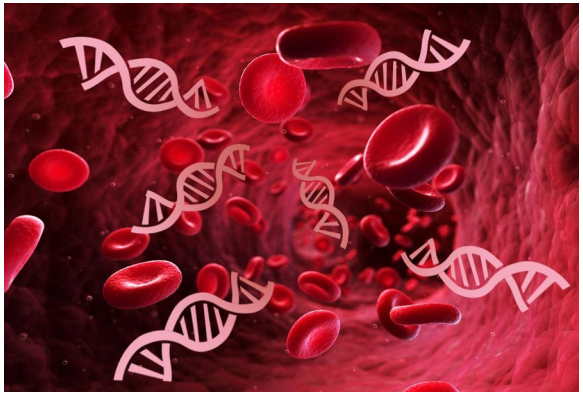

Mutant
Wildtype
Addition of
UMI adaptor
UMI
Amplification
NGS and UMI
Correction
Mutant
NGS and UMI
Correction
Wildtype
Adaptor
UMI #1
UMI #2
Pre-existing mutation
PCR/sequencing error

DDPCR(Digital Polymerse Chain Reaction)

- Diferencia entre PCR y DDPCR:
  - PCR ->Medida analógica DDPCR->Medida digital
  - Dividir muestra en muchas particiones->20.000 nanolitros
  - Mayor sensibilidad -> 0.0005%

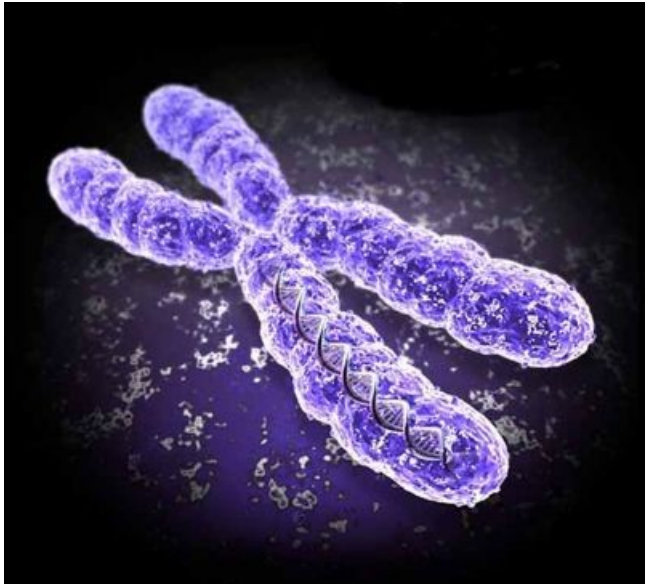
Alteraciones detectadas con ctDNA

- Mutaciones somáticas
- Variaciones en el número de copias (CNV)
- Cambios en la metilación del ADN



Mutaciones somáticas

- Afecta a las células somáticas
- Produce un genotipo distinto
- Puede ser detectado mediante ddPCR

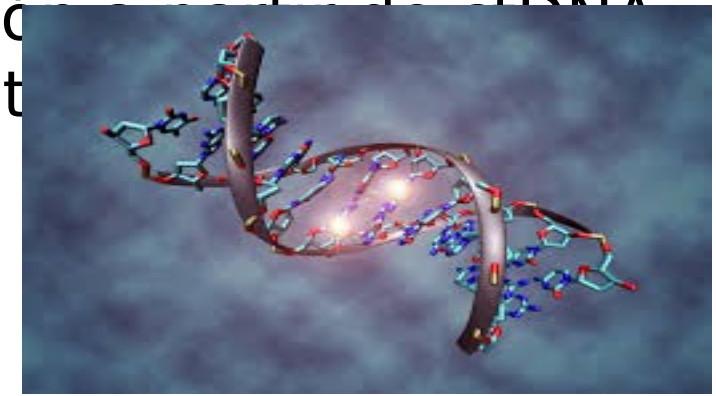


Variaciones en el número de copias (CNV)

- El número de copias de uno de los segmentos no coincide con el genoma de referencia
- Se consideran biomarcadores
- Se pueden detectar CNVs con NGS->DOC (Depth of Coverage)

Cambios en la metilación de ADN

- Proceso por el cual se añaden metilos al ADN
- Obtener metilació
  >Mapear origen t
- Cancer Locator



Cancer Locator

- Modelo probabilístico ->Determina proporción y localización tumor
- Estima nivel ctDNA en células normales y cancerosas
- Limitación:Clasificación depende de fracción de ctDNA de un tipo especifico
- Mejora ->Cancer Detector->Marcadores no son específicos del tejido

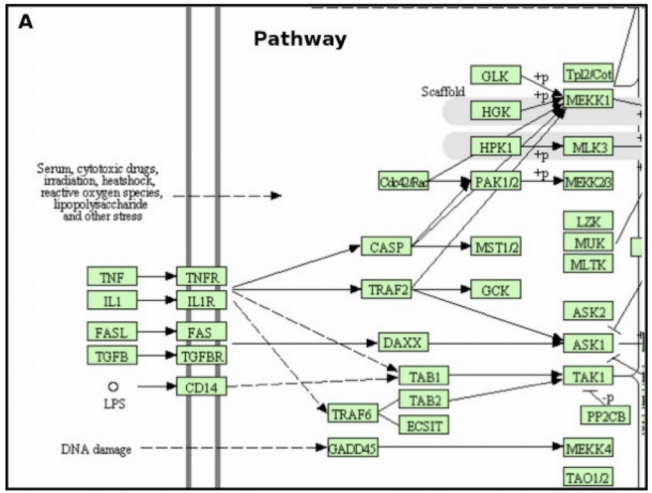
https://github.com/jasminezhoulab/CancerLocator

Futuro de la biopsia líquida

- ctDNA en canceres tempranos muy escaso -> Difícil de detectar
- Biomarcadores alternativos -> ARN
- Mejorar técnicas de aislamiento de ADN
- Personal capaz de interpretar datos
- Recopilación de datos

Pathways

Pathways es el término de la biología molecular que representa un modelo artificial simplificado de un proceso dentro de una célula o tejido



Al examinar los cambios en la expresión génica en un pathway, se pueden explorar sus causas biológicas

Detección cáncer de mama

- Análisis de miARN y ARNm y relaciones entre genes
- Se analizarán 4 tipos de cáncer:
  - Luminal A
  - Luminal B
  - Basal
  - HER2

Detección cáncer de mama: Pasos

1. Obtención de datos de la base TGCA
2. Obtener muestras combinadas de miARN y ARNm y trabajar por separado con cada subtipo de cáncer
3. Realizamos validación cruzada

Detección cáncer de mama

- Separar datos mediante Monte Carlo
- Identificar genes expresados diferencialmente
- Enriquecer pathways
- Matriz IS: **IS:|(Mx-My)|/(Sx+Sy)**
- Utilizamos matriz IS para clasificar

Detección Cáncer Pancreático [7]

- Búsqueda de biomarcadores
- Actualmente CA-19-9 (antígeno)
  - Lo provocan muchos tipos de cáncer
  - Falsos negativos

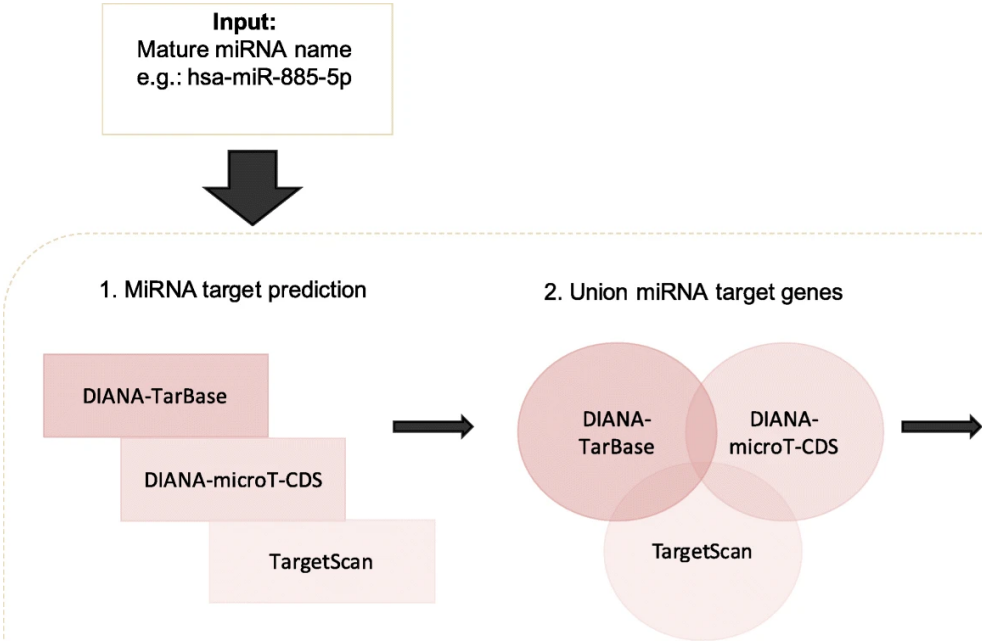
Cáncer Pancreático

Antes de comenzar el análisis:

Se sabe que hay 15 miARN circulantes que se expresan en pacientes con este tipo de cáncer en una fase madura de este [8].

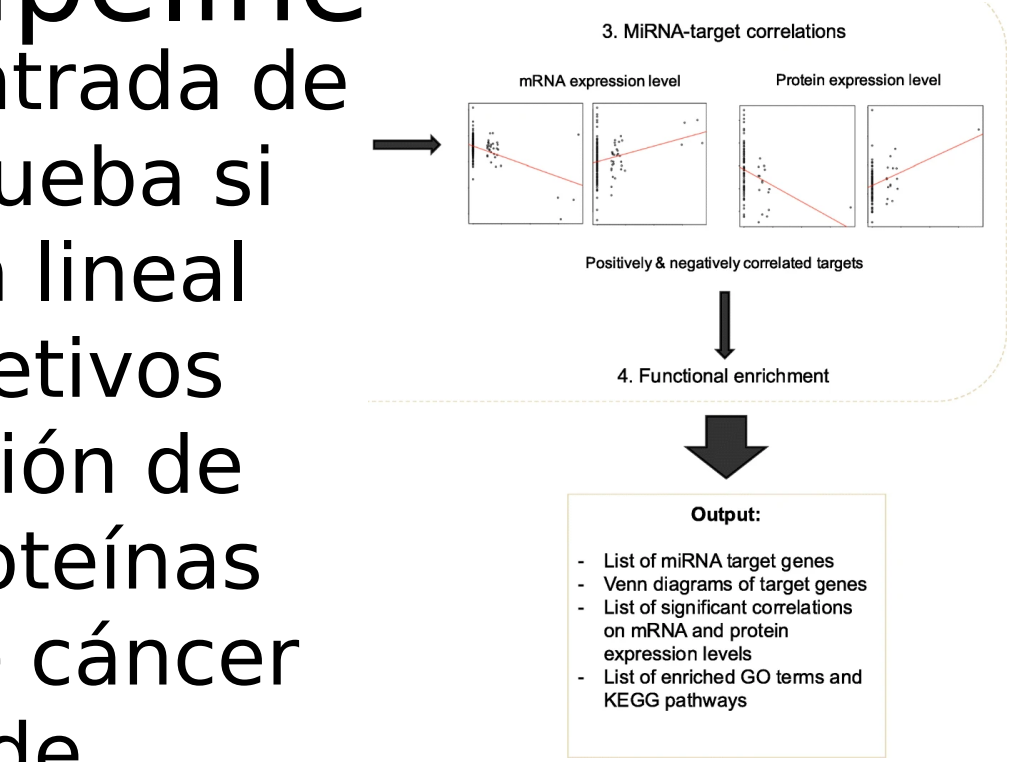
El pipeline

Se usan 3 bases de datos de predicción de objetivos de miRNA

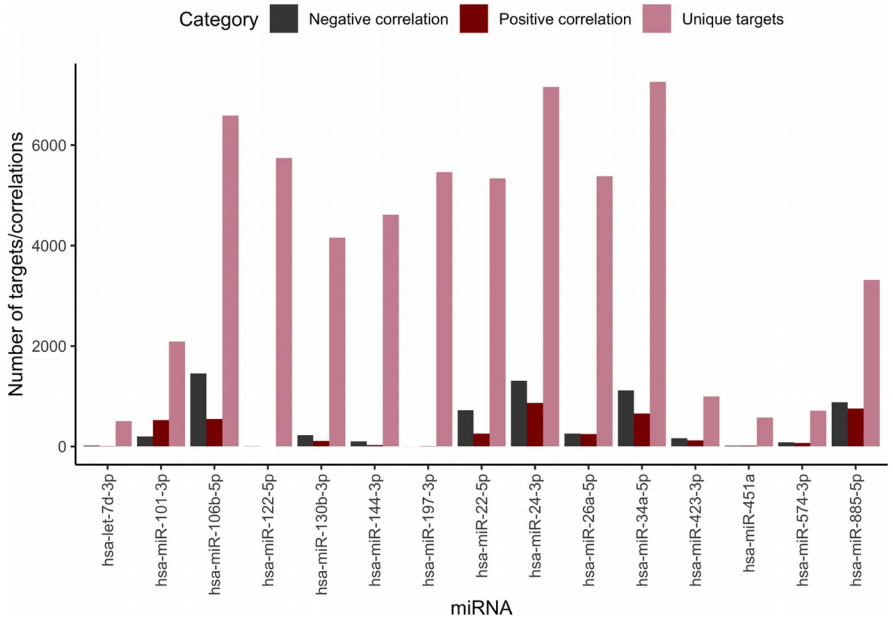


El pipeline

Con los datos de entrada de las bases se comprueba si hay una correlación lineal entre los genes objetivos sacados e información de miRNA, mRNA y proteínas sacada de tejido de cáncer pancreático se puede acceder a través de TCGA.



Gracias a esto se pueden ver qué genes pueden tener una correlación negativa o positiva con el cáncer.



Usando bases de datos de pathways como KEGG y GO se vio que:

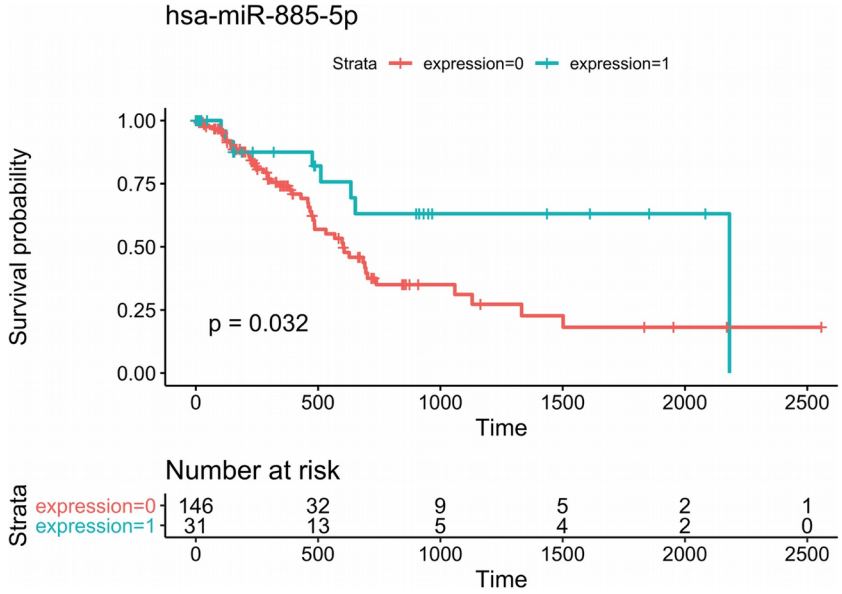
- No había genes comunes obtenidos de estos 15 que estuviesen relacionados con cáncer pancreático
- Pero sí que había comunes a 4 de ellos

Gene	hsa-miR-22-5p	hsa-miR-24-3p	hsa-miR-106b-5p	hsa-miR-885-5p
MAP 1B	- 0.24	- 0.20	- 0.40	0.33
NFIB	- 0.19	- 0.33	- 0.35	- 0.29
REV3L	- 0.22	- 0.28	- 0.30	0.18
LONRF2	- 0.24	- 0.26	- 0.34	0.34
TMTC1	- 0.24	- 0.22	- 0.21	0.21

Análisis supervivencia de Kaplan-Meier

Usando los datos de TCGA-PAAD.

Consiste en una estimación de cuál es la probabilidad de que un paciente siga vivo dada una condición (en este caso que hsa-miR-885-5p esté por debajo de la media (0) o por encima de la media(1))



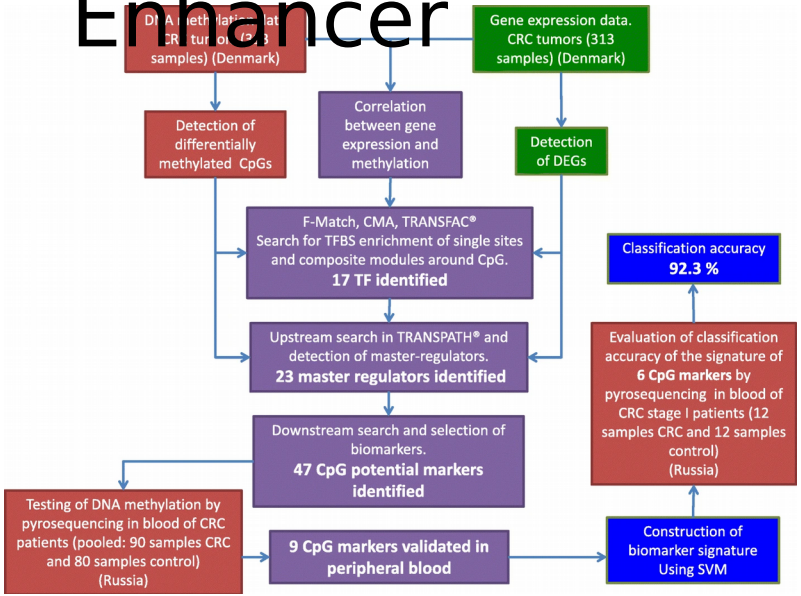
Detección cáncer colorrectal [10]

- Búsqueda de biomarcadores moleculares del cáncer colorrectal (CCR).
- Encontrar las potenciales **relaciones causales** entre **cambios epigenéticos** (DNA methylations) en zonas de regulación génica que afecta a **sitios de unión del factor de transcripción** (TFBS) y cambios de la

Universidad Zaragoza 1542

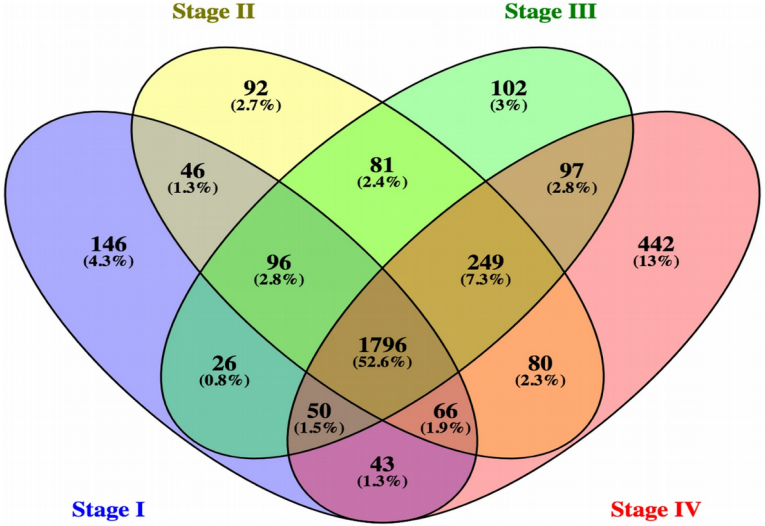
El pipeline utilizado MGE

- My Genome Enhancer



- Obtuvieron muestras de tejido tumoral y tejidos intestinales normales de 300 pacientes, proyecto SysCol
- Se utilizó Limma tool (R/Bioconductor para la construcción de modelos lineales) para la detección de DEGs (genes expresados diferencialmente). Comparando expresión génica entre

Se analizaron las muestras con las diferentes etapas del cáncer por separado para localizar entre qué etapas estaban localizados los DEG



Para encontrar las proteínas que eran capaces de modificar estas regiones génicas (y posiblemente sus alrededores).

CMA (una librería de aprendizaje supervisado para la clasificación de microarrays de grandes dimensiones). Sirve para construir un modelo de regiones reguladoras de genes.

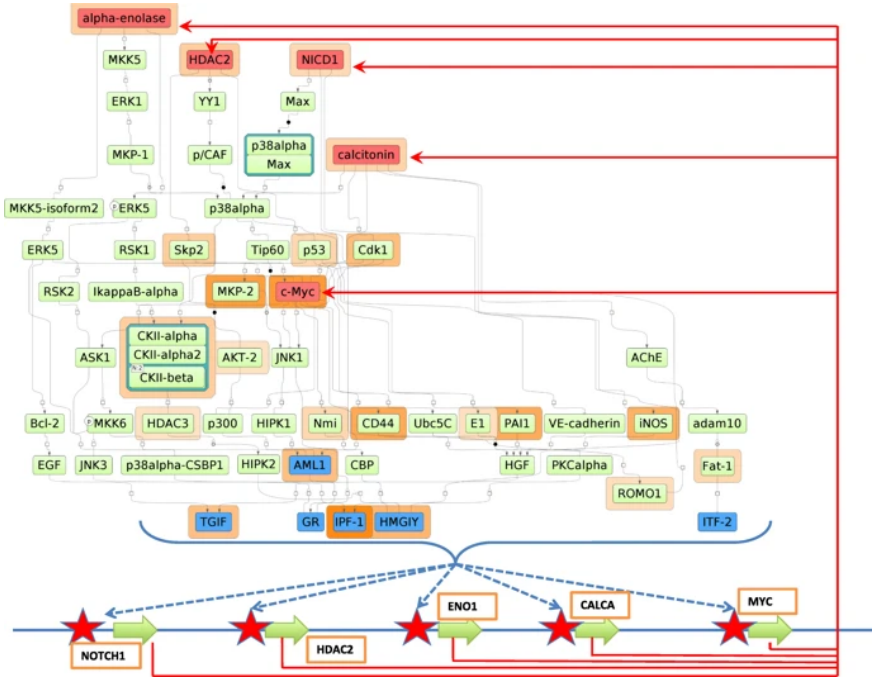
F-Match para identificar sitios de unión de factor de transcripción estadísticamente sobrerrepresentados TFBS.

Se modificó el algoritmo para detectar reguladores maestros (biomarcadores) para poder considerar el re-cableado producido por cambios epigenéticos como cambios en el estado de la metilación en regiones reguladoras de genes durante el desarrollo del cáncer.

Para ello se modificó los pathways para poder incluir bucles de retroalimentación.

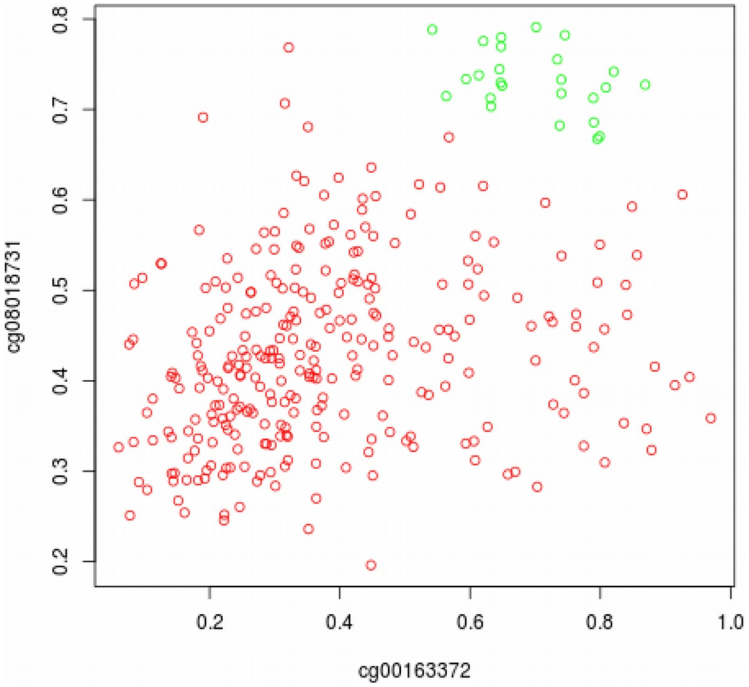
Walking pathways

- nodos rojos reguladores maestros
- nodos azules representan factores de transcripción
- estrellas rojas representan islas CpG metiladas
- flechas rojas muestran la traducción de los



Resultados

La combinación de la metilación de varios biomarcadores pueden ser necesarios para la detección del cáncer.



Se encontraron en total 47 biomarcadores que se redujeron a 9 comparando pacientes con cáncer y de control (la diferencia de metilación era mayor en estos).

Finalmente se redujo el número de biomarcadores a 6 utilizando SVM para construir una función de clasificación que mira el grado de metilación de todos los marcadores metidos.

Método de Inferencia (FEM)

- Estudio sobre la influencia de la metilación en la expresión de los genes
- Sirve también como clasificación tumoral
- Algoritmo Estadístico funcional supervisado.

Juntar datos

- Inferir DNAm, mRNA y PPIN en una única matriz
  - Los datos han de estar normalizados
- Con esta integración se pretende encontrar puntos máximos de solapamiento

Cálculos estadísticos

- Se calcula estadísticas de relación genéticas
  - Bayes, modelos de Regresión…
- Se realizan por separado
  - DNAm con POI (fenotipo de interés)
  - mARN con POI
- Al ser estadístico, ADN y ARN no tienen porqué estar relacionados

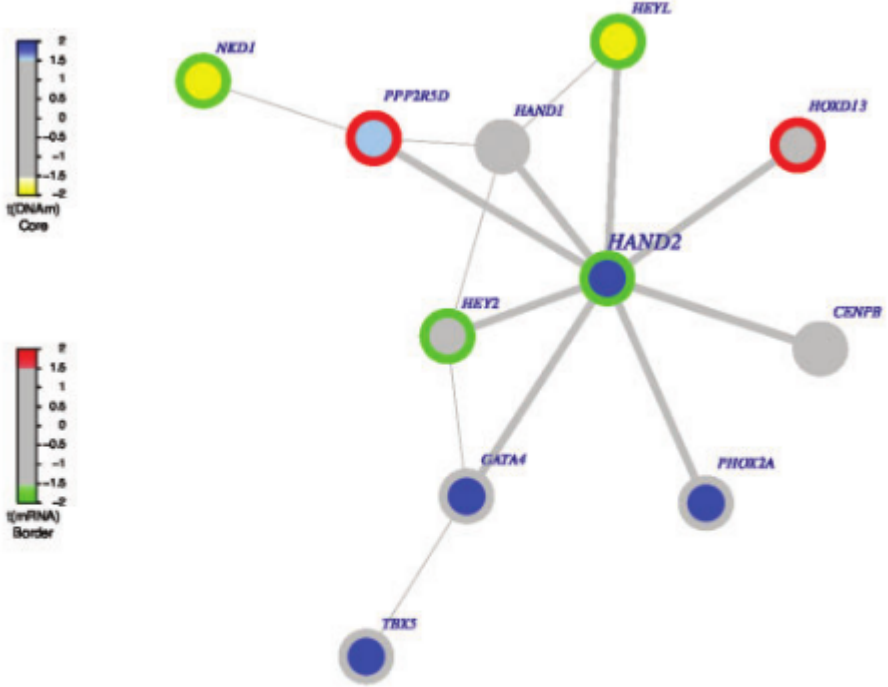
Generación de grafo pesado

- Relacionar relaciones entre genes y POI
  - Identificar puntos críticos entre Metiliación y expresión diferencial
- Cada nodo es un gene
- Cada arista es una relación entre genes
  - El peso de las aristas es la relación entre la metilación y la expresión diferencial de ambos genes

Generación de grafo pesado 2

- Un grafo pesado es un subgrafo cuya densidad de aristas (pesos) es superior a cualquier otro
- Extracción de subgrafo pesado más significativo
  - Algoritmo voraz Spin-Glass

Subgrafo pesado



Spin-glas para extraer GP

- Este algoritmo voraz emplea semillas para identificar los genes más influyentes
  - Top 100 de Bayes
- Posibles genes Aislados
  - Repetir proceso con aleatoriedad para evitar caer en genes aislados

Bibliografía

- Kang, S., Li, Q., Chen, Q. et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 18, 53 (2017). https://doi.org/10.1186/s13059-017-1191-5
- Web illumina: https://emea.illumina.com/science/technology/next-generation-sequencing.html
- HumanMethylation450: https://www.illumina.com/documents/products/datasheets/datasheet_humanmethylation450.pdf
- web EMBL-EBi https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/protein-protein-interaction-networks

Bibliografía

- Huang CC, Du M, Wang L. Bioinformatics Analysis for Circulating Cell-Free DNA in Cancer. *Cancers (Basel)*. 2019;11(6):805. Published 2019 Jun 11. doi:10.3390/cancers11060805 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6627444/
- Cava, C., Colaprico, A., Bertoli, G. et al. How interacting pathways are regulated by miRNAs in breast cancer subtypes. BMC Bioinformatics 17, 348 (2016). https://doi.org/10.1186/s12859-016-1196-1
- Franklin O, Jonsson P, Billing O, Lundberg E, Öhlund D, Nyström H, et al. Plasma micro-RNA alterations appear late in pancreatic Cancer. Ann Surg. 2018;267(4):775-81. https://www.ncbi.nlm.nih.gov/pubmed/28425921?dopt=Abstract

[8] Borgmästars, E., de Weerd, H.A., Lubovac-Pilav, Z. et al. miRFA: an automated pipeline for microRNA functional analysis with correlation support from TCGA and TCPA expression data in pancreatic cancer. BMC Bioinformatics 20, 393 (2019). https://doi.org/10.1186/s12859-019-2974-3

[9] Christoph Bock, Thomas Lengauer, Computational epigenetics, *Bioinformatics*, Volume 24, Issue 1, 1 January 2008, Pages 1–10, https://doi.org/10.1093/bioinformatics/btm546

[10] Kel, A., Boyarskikh, U., Stegmaier, P. et al. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. BMC Bioinformatics 20, 119 (2019). https://doi.org/10.1186/s12859-019-2687-7

[11] Yinming Jiao, Martin Widschwendter, Andrew E. Teschendorff, A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control, *Bioinformatics*, Volume 30, Issue 16, 15 August 2014, Pages 2360–2366, https://doi.org/10.1093/bioinformatics/btu316