

Simposio:

Epidemiología genética en poblaciones latinoamericanas

Asociación de genes con enfermedades complejas en poblaciones latinoamericanas: dificultades y ejemplificación en cáncer colorrectal.

Mónica Sans¹, Valentina Colistro¹, Augusto Rojas²

CHIBCHA Consortium

1. Departamento de Antropología Biológica, Facultad de Humanidades y Ciencias de la Educación, Universidad de la República, Uruguay- 2. Centro de Investigación y Desarrollo en Ciencias de la Salud, Universidad Autónoma de Nuevo León, Monterrey, México

Antecedentes y Sumario

- América Latina (América): poblaciones tri (o cuatri)-híbridas, **”experimentos naturales para estudios genético-epidemiológicos y antropológicos”** (Chakraborty y Weiss 1988)
- La ancestría es fundamental para la presencia/ausencia/frecuencia de determinadas enfermedades
-
- Diferentes SNPs/genes **asociados** a las enfermedades complejas, según ancestría
 - Ancestría como elemento confusor en los estudios



Cáncer colorrectal y ancestría

- Proyecto “CHIBCHA”: 7mo. Programa marco, CE.
- Responsable: Ian Tomlinson;
- Objetivo general: validar SNPs previamente detectados y detectar nuevos SNPs en A.L. asociados a cáncer colorrectal
 - **Particular: Contribuir al conocimiento de las causas genéticas que se relacionan con el cáncer colorrectal (CCR) en México, con énfasis en los elementos confusores relacionados a ancestría y subestructuración poblacional**

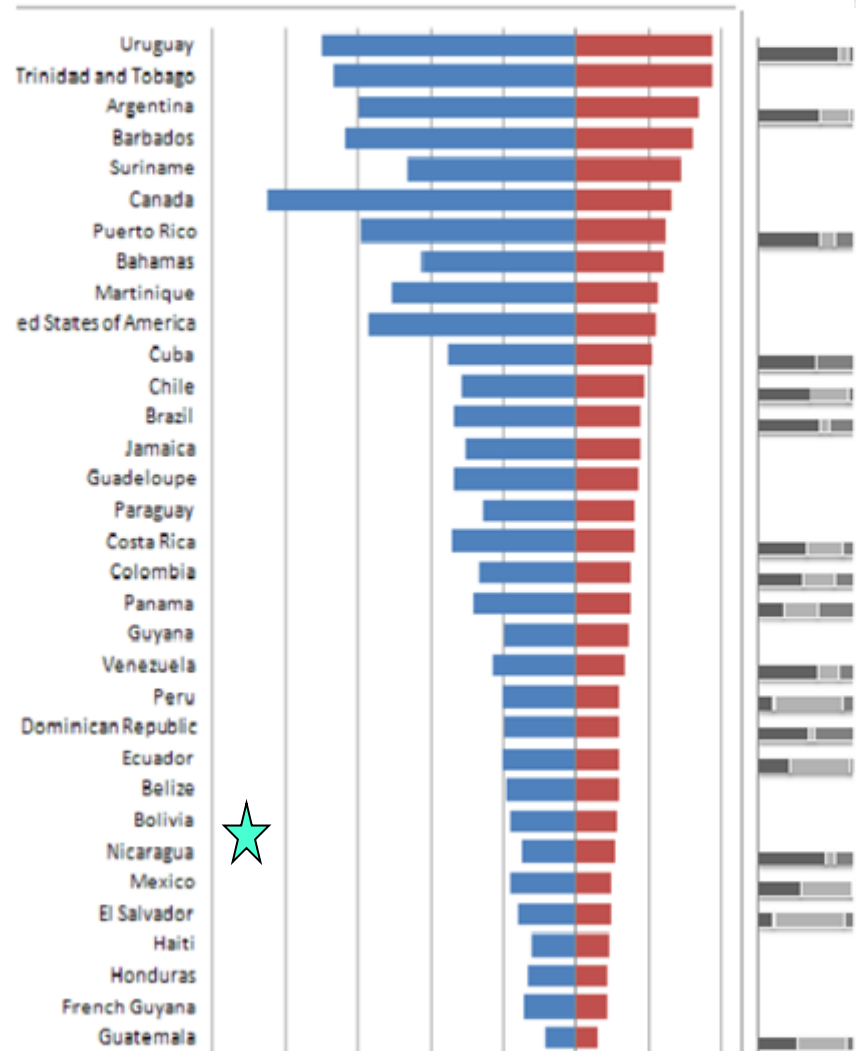
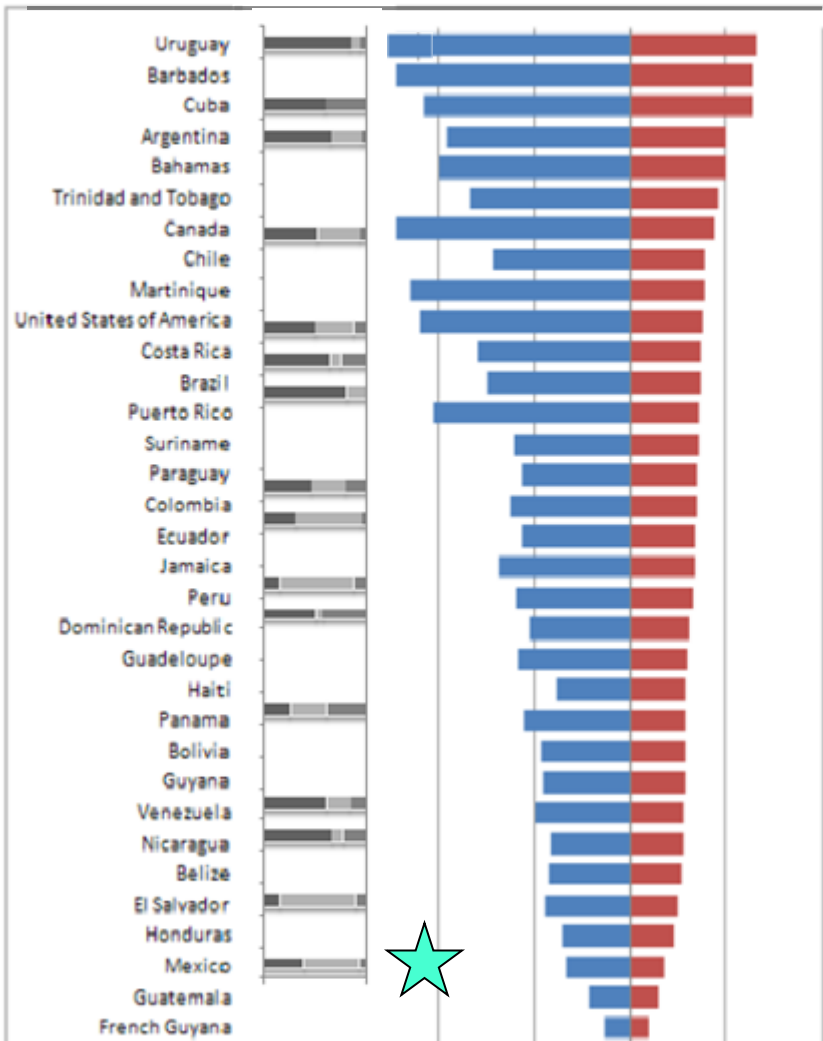
Cáncer colorrectal (OPS): azul: incidencia; marrón: mortalidad

negro: ancestría europea, gris claro: indígena; gris oscuro: africana

Colorectal cancer incidence and mortality age standardized rates, by country and sex, 2012

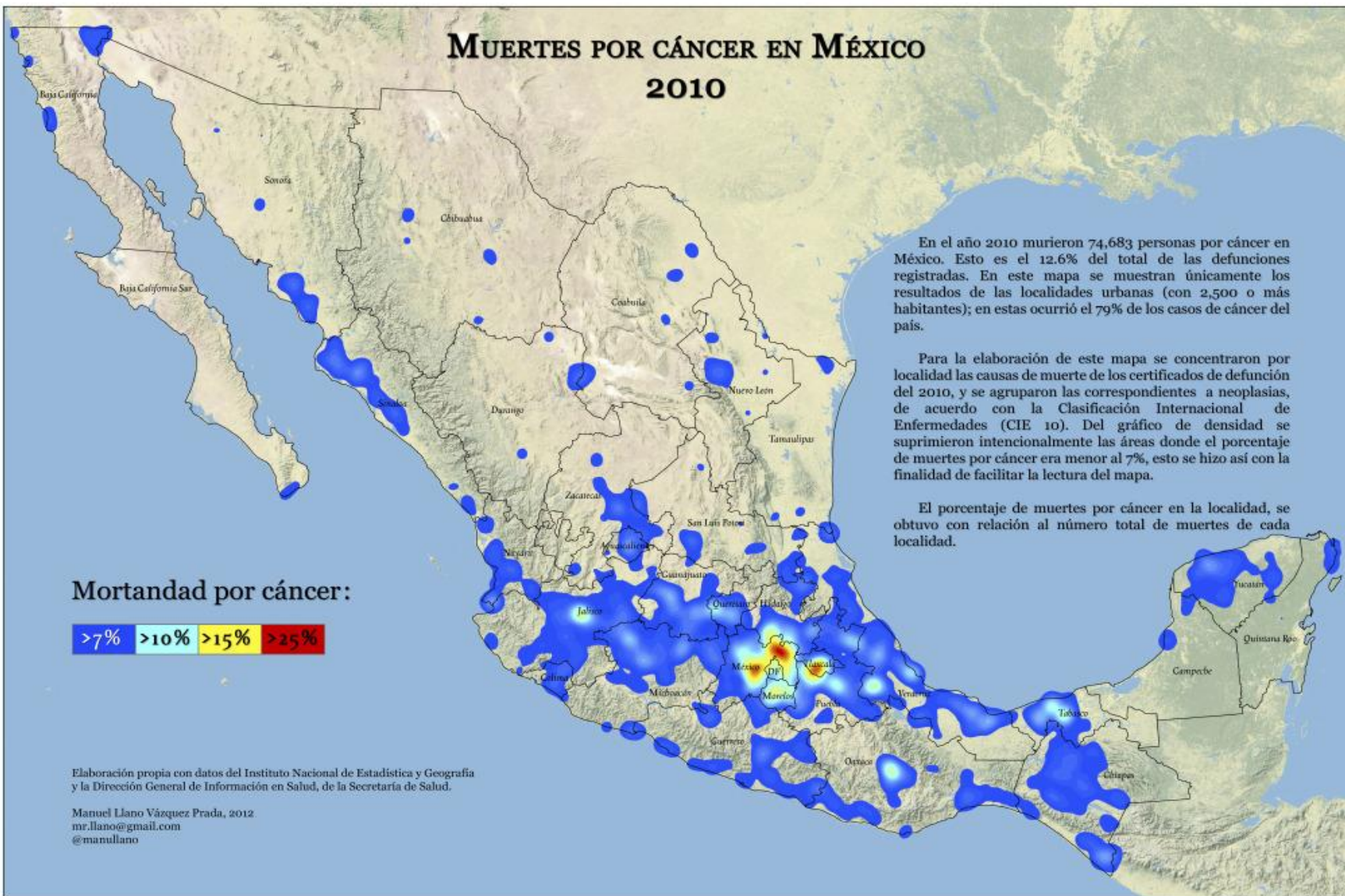
Women

Men



Cáncer en México



MUERTES POR CÁNCER EN MÉXICO 2010



Cáncer colorrectal en México

- 5to -4to. lugar entre los distintos cánceres, en aumento
- 6500 casos nuevos por año
- 4000 muertes/año, detección tardía
- 22% de los casos en < 40 años

Porcentaje medio de cambio anual (PMCA) de las tasas estandarizadas de mortalidad, según tipo de cáncer (2000-2010)⁵

Tipo de cáncer	Sexo	Número de muertes (2010)	TEM (2010)	PMCA
Todos los cánceres	Hombres	33.263	78,30	-1,0*
	Mujeres	34.745	69,39	-1,0*
Cáncer de mama		4.966	9,92	0,51*
Cáncer cervicouterino		3.864	7,69	-4,61*
Cáncer colorrectal 	Hombres	2.090	4,90	1,98* 
	Mujeres	1.892	3,79	0,59
Cáncer de hígado	Hombres	2.539	6,12	-0,93*
	Mujeres	2.746	5,60	-1,16*
Cáncer de pulmón	Hombres	4.300	10,47	-3,7*
	Mujeres	2.331	4,70	-1,96*
Cáncer de próstata		5.404	13,62	-0,73*
Cáncer de estómago	Hombres	2.944	6,98	-2,56*
	Mujeres	2.556	5,08	-2,3*

Notas: TEM: Tasa estandarizada de mortalidad; PMCA: Porcentaje medio de cambio anual; *: El PMCA es estadísticamente significativo distinto de cero para p=0.05.

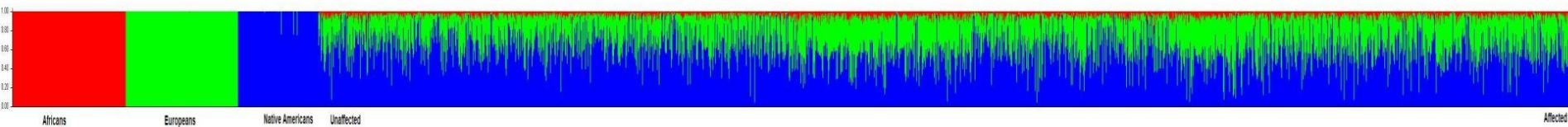
Estudios de asociación caso-control (GWAS y otros):

ancestría como elemento confusor

PRIMER PASO: “Data Quality”: eliminación de falsos positivos y falsos negativos

Estrategias: enfatizar individuo o marcadores.

- Anderson et al (2010) sugiere **enfatizar marcadores (SNPs)** para minimizar su pérdida.

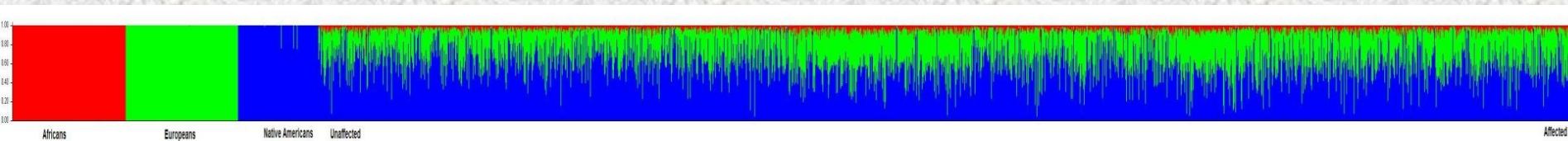


Control de calidad GWA (Anderson et al 2010)

- Por Marcadores:
 - 1) identificación de SNPs con exceso de genotipos faltantes,
 - **2) identificación de SNPs con desvío significativo del equilibrio Hardy-Weinberg,**
 - **3) identificación de SNPs con diferencias significativas de tasas de genotipado entre casos y controles**
 - **4) remoción de marcadores con frecuencia muy baja.**
- Por Individuos:
 - 1) identificación de individuos con información discordante de sexo,
 - 2) identificación de individuos con genotipos faltantes o tasa de heterocigocidad fuera de rango
 - **3) identificación de individuos duplicados o relacionados**
 - **4) identificación de individuos con ancestría divergente**

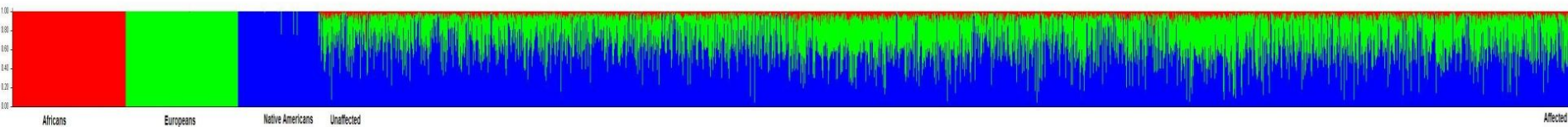
CHIBCHA-Cáncer colorrectal en México:SNPs

- 1er. Control de Calidad (inicial: 387.948 “Custom Array”, 781.996 “Latino Array”) :
1.169.944 (387.948 “Custom Array”, 781.996 “Latino Array”)



CHIBCHA-Cáncer colorrectal en México:SNPs

- SNPs: N=1.169.944 (inicial: 387.948 “Custom Array”, 781.996 “Latino Array”)
- Se excluyeron:
 - 2.625 SNPS en Cr. Y y mitocondriales
 - 53 SNPs con diferente imputación de genotipos (call rates) en casos y controles
 - 90 SNPS con falta de datos genotípicos
 - 3.761 SNPs por estar en desequilibrio HE.
 - 141.381 por estar bajo el umbral de MAF (frecuencia <0.01?)
- **ANALIZADOS (FINAL): 1.006.658 SNPs**



CHIBCHA-Cáncer colorrectal en México: muestra

- Casos y controles, inicial: 2.041 individuos
- Después primer control calidad (falta de datos): 1.881

	Casos	Controles	Totales
Mujeres	449	356	817
Hombres	449	614	1073
Totales	899	972	1881

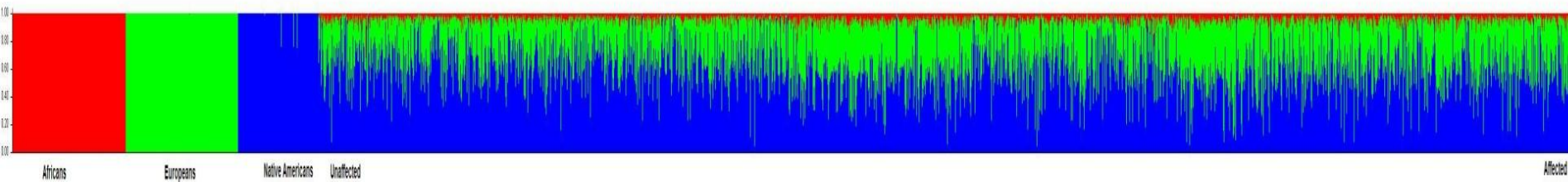


Mezcla (aportes):

Mezcla poblacional (2.010 AIMs)

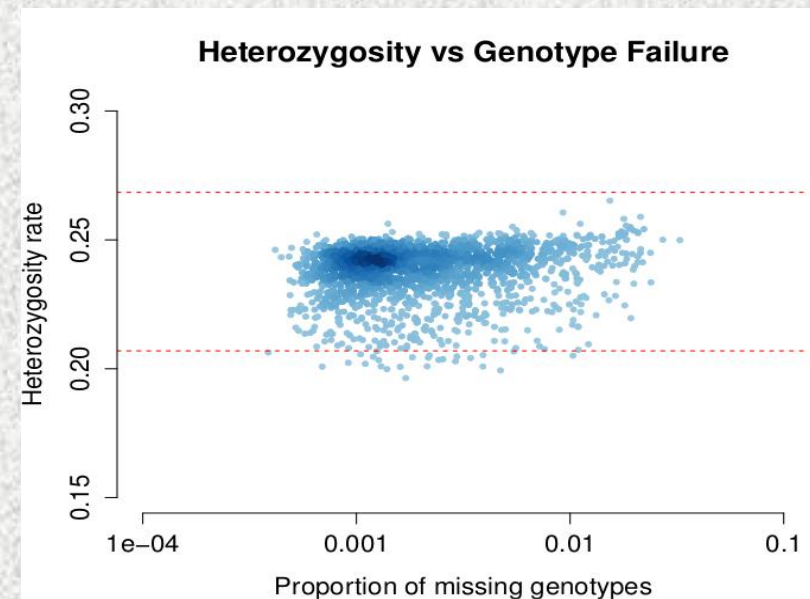
- Poblaciones parentales:
 - Europeos (CEU + Ibéricos) (1000 Genomas);
 - Africanos: (Yoruba de Nigeria + Gambia + Esan de Nigeria) (1000 Genomas);
 - Nativo-americanos (Maya, Nahua, Aymara, Quechua) (Mao et al 2007)

	Africano	Europeo	Nativo
Casos	4.6	37.2	58.2
Controles	4.2	33.7	62.0



Control de calidad – segunda etapa

- Individuos: N=1881
 - Discordancia en el sexo (47 individuos removidos)
 - Fallas en tasas de heterocigocidad y genotipado (“outliers”) (removidos: 26 por heterocigocidad, 0 por genotipado) FIGURA
 - Ancestría divergente: 20 removidos por tener aporte africano >20%
 - **Duplicados o relacionados (parientes): 402!! (PLINK)**

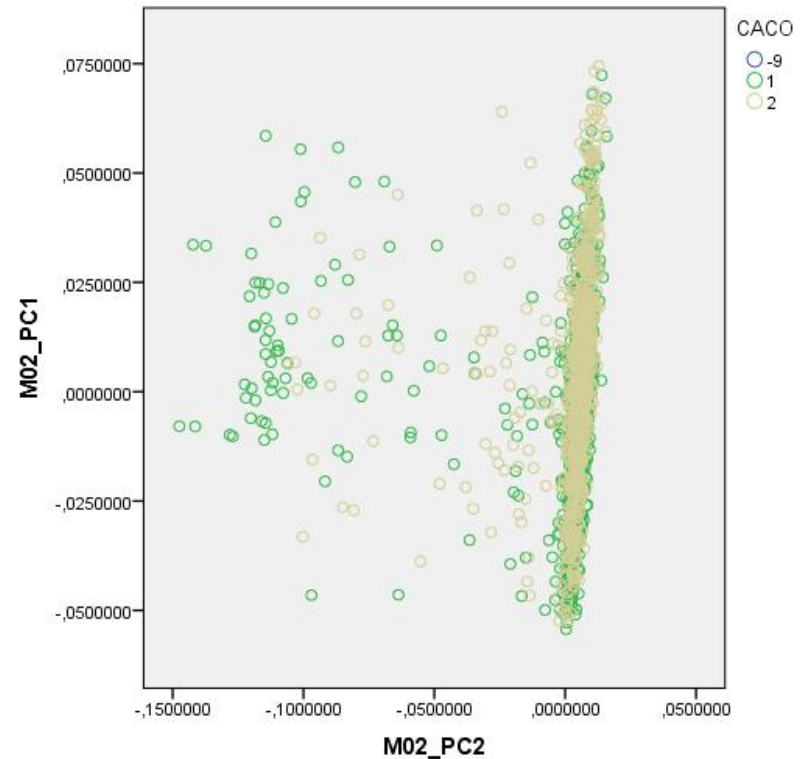
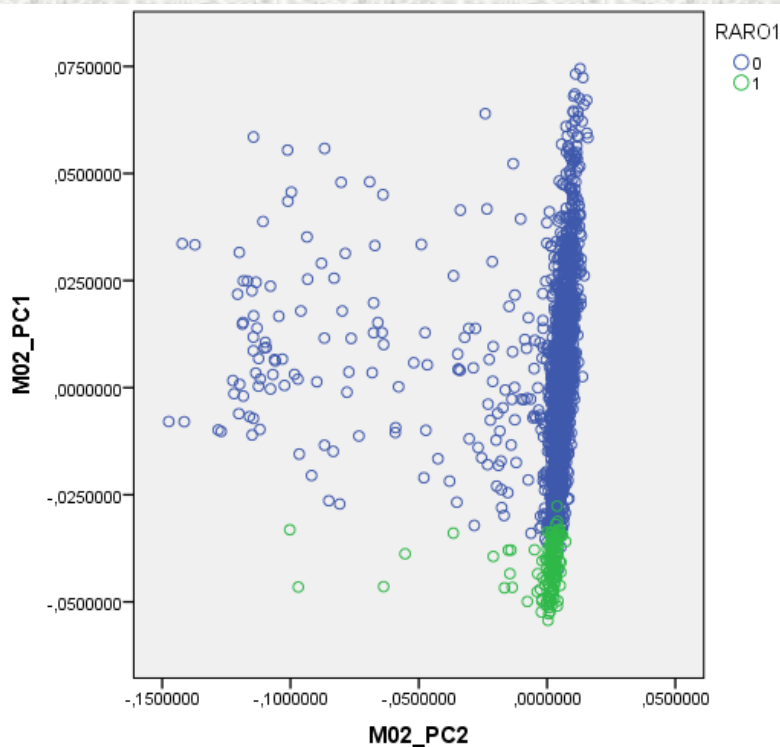


Control de calidad – segunda etapa

- 402 individuos duplicados o parientes (idénticos por descendencia- IBD) (verde en la figura):
 - “Custom array”: 184 (92 pares)
 - **“Latino array”: 373 interrelacionados**

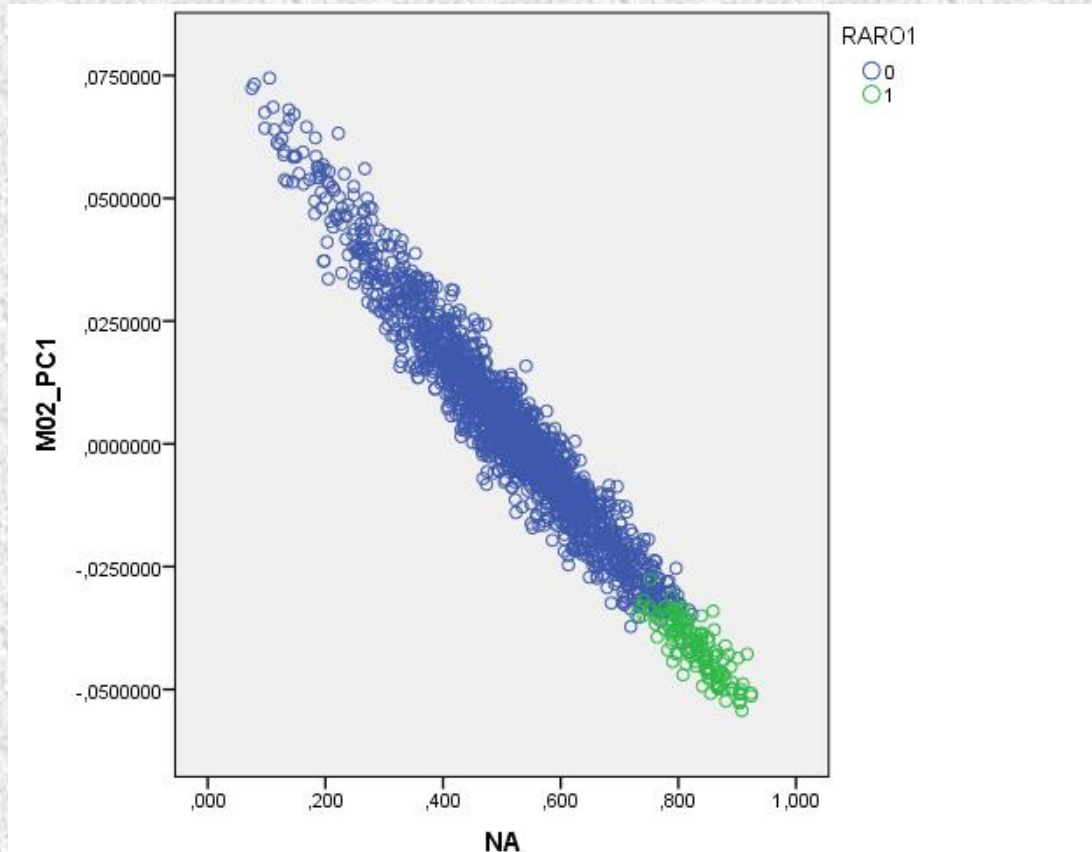
Azul: bajo IBD; Verde: alto IBD

Casos (verde) o controles (beige)



Control de calidad – segunda etapa

- Parentesco (alto IBD) relacionado a mestizaje/origen étnico
 - Artificio especialmente generado por el “LATINO array” (enriquecido por AIMs - ancestry informative markers)



Eje X (NA): aporte nativo

Eigenstrat (proyección
datos de mexicanos sobre
datos 1000 genomas: 4 PC

**Se removieron
70 individuos por alto IBD**

FINAL: 1712 individuos:

- 881 casos, 831 controles

- 1012 hombres, 700 mujeres

REAP (Thornton et al 2012)

- REAP (*relatedness estimation in admixed populations*), para estimación de “identity by descent” (IBD) en poblaciones estructuradas y con ancestría mixta.
 - Frecuencias alélicas de SNPs individuo-específicas, calculadas sobre ancestrías de GWAs
 - Individuos a eliminar: $N = 59$ (vs. 70 Eigenstrat)

Resultados: Genes y SNPs

- Bonferroni: 12 SNPs ($p \leq 0.05$), ninguno coincidente con Europa, 3 coincidentes con Colombia (CHIBCHA)
- FDR (*false discovery rate*): 101 SNPs ($p \leq 0.05$), 36 SNPs ($p \leq 0.01$)
- p: varios cientos.

Introducción y objetivos

El Cáncer Colorrectal (CRC) es común en ambos sexos, ocupando el 4º lugar en prevalencia y mortalidad por cáncer. Estudios de asociación de todo el genoma (GWAS), basados en poblaciones europeas, han identificado genes asociados a CRC así como las regiones: 8q23.3, 8q24.21, 10p14, 11q23.1, 15q14 y 15q21. Estas variantes explican <5% del riesgo genético a CRC, pero su efecto combinado, junto con variantes no detectadas, puede ser considerablemente mayor. La evidencia sugiere que estos estudios pueden no haber sido lo suficientemente grandes para detectar SNPs asociados a un riesgo relativo bajo o moderado. Las poblaciones mestizas latinoamericanas son una buena oportunidad para validar los genes previamente identificados y para detectar nuevos genes asociados.

Nuestro objetivo es identificar nuevos SNPs asociados a CRC y validar los SNPs que han sido determinados previamente.

Materiales y Métodos

1.006.658 SNPs analizados | 881 Casos | 831 Corroles

Se tomaron muestras en tres localidades de México.

1000 Genomes

Europeos (CEU + Españoles) | Africanos (Nigeria + Gambia)

Nativo Americanos | Mayá, Náhuatl, Aymara y Quechua

Software:

- Control de Calidad del Genotipado: Axon Analysis Suite / SNPtoolbox
- Control de Calidad poblacional: PLINK
- Estructura poblacional: Eigenstrat
- Análisis de asociación genómico: PLINK

Resultados

Análisis de estructura poblacional respecto a las poblaciones parentales. Se evidencia el aporte minoritario africano y se demuestra la ascendencia mestizada principalmente entre europeos y nativo americanos.

Análisis de asociación genómico ajustado por ascendencia. El gráfico muestra los p-valor de los SNPs por cromosoma. Se observan 17 SNPs asociados (p-valor < 0.000001). En la tabla se muestran sus p-valor no ajustados y luego del test de corrección por múltiples (FDR). En caso de considerar un p-valor < 0.01 se asocian 134 SNPs.

SNP	Cromosoma	Coordenada	Ref.	P-valor	FDR
rs711399	12	92.2	---	8.54 x 10 ⁻⁶	0.0006224
rs12184208	22	424.32	---	5.69 x 10 ⁻⁶	0.0002234
rs74362465	1	451.1	TTLL7	6.86 x 10 ⁻⁶	0.001205
rs121600661	23	423.2	TBC1D16	3.25 x 10 ⁻⁶	0.0003768
rs7971969	18	412.2	---	2.573 x 10 ⁻⁶	0.0005708
rs798112	7	145.1	NHIF3	1.84 x 10 ⁻⁶	0.0005488
rs6849527	1	106.22	---	1.84 x 10 ⁻⁶	0.0005488
rs13360905	4	106.13	NEL1	8.66 x 10 ⁻⁷	0.00024
rs13349841	4	103.2	U072823	9.55 x 10 ⁻⁷	0.00024
rs4451224	12	104.22	FRS39	6.86 x 10 ⁻⁷	0.00027
rs14602327	17	211.1	CHC3	6.65 x 10 ⁻⁷	0.00027
rs15490452	6	327.3	CDAL1	4.64 x 10 ⁻⁷	0.0004709
Abv_1713686	2	111.2	ILR12	2.66 x 10 ⁻⁷	0.00026
rs7430381	1	442.13	TAF1L	2.36 x 10 ⁻⁷	0.000449
rs117602366	10	111.23	KIAA042	1.78 x 10 ⁻⁷	0.000252
Abv_1048474	23	427.2	LINC00323	1.07 x 10 ⁻⁷	0.001343
rs1344000	17	423.33	ANDRO1	1.05 x 10 ⁻⁷	0.001343

* SNPs detectados como asociados en población mestizaje.
 ** SNP rs7430381 corresponde a gen asociado a CRC.

Conclusiones

- De los SNPs detectados en asociación a CRC, tres fueron validados en otro estudio con poblaciones mestizadas (Colombia) donde se trabajó con el mismo panel de SNPs.
- Un SNP se asocia significativamente al gen CHD3 y otros dos SNPs (no mostrados en la tabla) se asocia significativamente (p-valor < 0.01) al gen BMP4. Ambos genes fueron descritos previamente en asociación con CRC.
- Los SNPs que muestran asociación en este estudio podrían explicar parte del riesgo genético al CRC. Resta dilucidar su efecto en combinación con otros marcadores propuestos.
- Se evidencia la complejidad de trabajar con poblaciones mestizadas dada su estructura genética debido a la ascendencia genética, lo cual dificulta la reproducibilidad de resultados previos.

Agradecimientos:

Agustín Rojas-Martínez, Raquel Cruz-Guerrero, Angel Carracedo, Ian Tomlinson, Paul Loh, Luis Canvajal-Carmona, Pedro Hidalgo y Patricia Mut.

Referencias

- Tomlinson et al. 2008 Nat Genet
- Mao et al. 2007 Am J Hum Gen
- Anderson et al. 2010 Nat Protoc
- Purcell et al. 2007 Am J Hum Gen
- Price et al. 2006 Nature Genetics

poster #430

- FDR: 17 genes ($p \leq 0.01$), 76 genes ($p \leq 0.05$)
 - De 41 genes reportados como asociados a CCR en Europa solo 2 coinciden: BMP4 y CHD3.
 - De 22 SNPs analizados con relación a CCR en Europa, ninguno coincide.

Conclusiones

- Dificultades de análisis de las poblaciones mestizadas
 - Dificultades en determinar la ancestría /p.parentales
 - Desconocimiento de (sub)estructuraciones
 - **Dificultad de repetir asociaciones encontradas en otras regiones geográficas**
 - **¿Inaplicabilidad de otros resultados/estrategias?**

En breve...

- Ancestría individual por SNP
- Reanalizar con REAP
- Relacionar con datos de Colombia **y de Brasil**

Equipo de investigadores

- Rocío Ortiz y colaboradores (CHIBCHA México)
- Raquel Cruz, Ian Tomlinson, Angel Carracedo, Paul Lott, Luis Carvajal (CHIBCHA –Europa)
- Pedro C. Hidalgo, Patricia Mut (CHIBCHA – Uruguay)

Agradecimientos

- ALAG, Bernardo Bertoni

Gracias!!!