

## Estudio de la relación de la ancestría genética con el riesgo al cáncer colorrectal en Colombia.

### Study of the relationship between genetic ancestry and colorectal cancer risk in Colombia

Angel Criollo-Rayó<sup>1</sup>, Mabel Elena Bohórquez<sup>1</sup>, Paul Lott<sup>2</sup>, Angel Carracedo<sup>3</sup>, Ian Tomlinson<sup>4</sup>, Consorcio CHIBCHA<sup>5</sup>, Luis Carvajal<sup>1,2</sup>, Magdalena Echeverry<sup>1</sup>.

1. Grupo de Citogenética, Filogenia y Evolución de Poblaciones, Facultades de Ciencias y Facultad de Ciencias de la Salud, Universidad del Tolima. Ibagué, Colombia.
2. Genome Center, Department of Biochemistry and Molecular Medicine, School of Medicine- University of California, Davis. GBSF, 451 Health Science Drive Davis, California. 95616-8816. USA.
3. Fundación Pública Galega de Medicina Xenómica (SERGAS)-CIBERER, Universidade de Santiago de Compostela, Santiago de Compostela, España.
4. The Wellcome Trust Centre for Human Genetics, University of Oxford, Reino Unido.
5. Lista completa en el anexo.

Recibido: Junio 25 del 2019

Aceptado: Septiembre 15 del 2019

\*Correspondencia del autor: Angel Criollo-Rayó

E-mail: alcriollora@ut.edu.co

#### Resumen

**Introducción.** El cáncer colorrectal es una carga para la salud pública en Colombia y el mundo. Estudios de asociación genética han identificado regiones cromosómicas asociadas a esta enfermedad, mostrando riesgo variable entre poblaciones, debido a la historia demográfica y la ancestría genética. **Objetivo.** Estudiar el riesgo que aportan 20 marcadores al cáncer colorrectal en Colombia, empleando 955 casos y 972 controles del consorcio CHIBCHA, analizando conjuntamente el efecto de la ancestría genética global y local. **Metodología.** Las muestras se genotipificaron usando microarreglos Axyom Affymetrix LAT y CUSTOME, para obtener los genotipos genómicos globales, incluyendo 20 SNPs de riesgo. Los análisis estadísticos se realizaron en PLINK (asociaciones), ADMIXTURE (ancestría global), Elai (ancestría local) y R (modelos logísticos). **Resultados.** Once regiones cromosómicas resultaron asociadas presentando ORs entre 1.14 y 1.41 ( $p < 0.05$ ): 18q21.1, 19q13.11, 10p14, 14q22.2, 20p12.3, 8q23.3, 6p21.2, 15q13.3 y 8q24.21. Una mayor ancestría europea se asoció con el riesgo a nivel global (OR=3.016, IC 95%:1.162-7.894,  $p=0.00325$ ), y a nivel cromosómico local se detectaron las regiones 6q23.2 (OR<sub>ajustado</sub>=1.378, IC95%: 1.202-1.580,  $P_{ajustado}=4.2e-6$ ) y 4p13 (OR<sub>ajustado</sub>=1.301, IC95%:1.137-1.489;  $P_{ajustado}=0.00013$ ). **Conclusiones.** La ancestría podría considerarse un factor en la explicación de la susceptibilidad en Colombia, indicando que la mezcla genética de origen amerindio y europeo, influye en la estructura poblacional y explicaría las diferencias en la incidencia del CCR entre poblaciones latinas y europeas.

**Palabras clave:** Asociación genética, población, haplotipo, ancestría genética, susceptibilidad genética.

## Abstract

**Introduction:** Colorectal cancer is a public health burden in the world and Colombia. Recent genome wide association studies have identified chromosomal regions associated with the disease, depicting variable risk between populations, owing to the demographic history and genetic ancestry. **Objective:** We aimed to study the colorectal cancer risk in Colombia provided for 20 genetic markers, by using 955 cases and 972 controls from the CHIBCHA consortium, in the context of global and local genetic ancestry. **Methodology:** The samples were genotyped using Axyom Affymetrix LAT and CUSTOME array in order to obtain the global genome genotypes including 20 risk SNPs. Statistical analysis was performed in PLINK (associations), ADMIXTURE (global ancestry), Elai (local ancestry) and R language (logistic models). **Results:** Eleven chromosomal regions were associated with ORs ranging between 1.14-1.41 ( $p < 0.05$ ): 18q21.1, 19q13.11, 10p14, 14q2.2, 20p12.3, 8q23.3, 6p21.2, 15q13.3 y 8q24.21. On average, a higher global European ancestry was associated with colorectal cancer risk (OR=3.016, IC 95%:1.162-7.894,  $p=0.00325$ ). At the local chromosomal level two regions presented a significant increment of European ancestry 6q23.2 (OR adjusted=1.378, CI95%: 1.202-1.580,  $p$  adjusted =4.2e-6) and 4p13 (OR adjusted =1.301, CI95%:1.137-1.489;  $p$  adjusted =0.00013). **Conclusions:** Genetic ancestry can be considered as a relevant factor for the colorectal cancer susceptibility in Colombia. Both Native American and European ancestry are accounting for the most part of population structure in the sample we studied, which could explain the differences for the colorectal cancer incidence between Latin American and European populations.

**Keywords:** genetic association study, SNPs, population, haplotype, genetic ancestry, genetic susceptibility.

## Introducción:

El cáncer colorrectal (CCR) es una de las malignidades que se presentan con mayor frecuencia en el mundo; en términos del número de casos nuevos diagnosticados al año (1.6 millones para el 2013), y, en el efecto que tiene la carga de dicho padecimiento en la sociedad, estimado como el número de años de vida ajustados por la enfermedad (1-3). No obstante el avance en metodologías para diagnóstico, tamizaje (4, 5), manejo operatorio y oncológico (6), el CCR esta generando alrededor de 771.000 muertes al año (en 2013), representando cerca de la mitad de todas las muertes causadas por cáncer gastrointestinal, siendo la probabilidad de desarrollarlo más alta en hombres que en mujeres (1:21 para hombres, 1:43 para mujeres) (7). A nivel mundial, el cáncer de colon y recto ocupa el tercer lugar en incidencia y el cuarto en mortalidad; particularmente, en el caso de los países desarrollados, es el segundo en ambos parámetros, mientras que en las naciones en vía de desarrollo ocupa el cuarto (8). La situación en Colombia refleja en parte el panorama global de las transiciones socio-demográficas en América Latina (9, 10), siguiendo un incremento gradual entre 1985 y 2010 (11, 12), representando la quinta causa de muerte por cáncer en ambos sexos. A la fecha, no se implementan planes de tamizaje, y la mortalidad es alta en áreas urbanas colombianas debido al aumento de la incidencia y a los cambios en el estilo de vida en la población colombiana (12, 13). Las diferencias observadas entre distintas regiones con-

tinental pueden deberse al efecto conjunto de factores ambientales -determinados en parte por el estilo de vida y la historia demográfica- y genéticos, que generan patrones de gradientes geográficos en la incidencia (8). Se estima que la contribución genética, correspondiente a la heredabilidad de la susceptibilidad al CCR, en los casos con historia familiar, puede llegar a un 35% (Paul Lichtenstein et al., 2000); que incluye aquellos *loci* con un bajo impacto individual en dicha heredabilidad (14). La mayoría de estas variantes han sido validadas y/o replicadas en poblaciones caucásicas; pero con estudios limitados en poblaciones mestizas como la colombiana. El impacto que las variantes conocidas pueda tener en un trasfondo de mezcla genética, varía con la extensión del desequilibrio de ligamiento -DL-, el mestizaje, y las recombinaciones meióticas que han generado mosaicos cromosómicos y nuevos haplotipos (15). La ancestría genética, entendida como la proporción del genoma en un individuo, que proviene de una o varias poblaciones ancestrales, genera un DL que es importante para mapear variantes asociadas a enfermedades que muestran un riesgo diferencial debido a la ancestría genética, mediante el mapeo de la mezcla (16), mostrando aplicabilidad en poblaciones Latinas (17). Empleando esta metodología se han detectado *loci* de susceptibilidad para diferentes rasgos en poblaciones hispanas/latinas, tales como la presión arterial (18), el asma (19) y cáncer de seno (20).

Entre distintas poblaciones continentales existen varia-

ciones en las frecuencias alélicas y en el desequilibrio de ligamiento -DL-, lo cual puede incrementarse por la mezcla genética, como la ocurrida en la Región Andina colombiana, objeto de estudio en este trabajo, en las que se produjo un mestizaje entre nativos americanos y europeos principalmente (21), producto del oleaje migratorio durante el periodo de la colonia. La implementación de las anteriores estrategias, ayudan a clarificar la etiología molecular del CCR y permiten generar herramientas para impactar en la eficiencia del manejo y prevención del CCR en la población colombiana y así mismo aportar al desarrollo de posibles paneles de marcadores para el tamizaje. Por lo anterior, con el fin de analizar la ancestría genética y su relación con el riesgo a desarrollar CCR en Colombia, se estudiaron 20 regiones cromosómicas conocidas, sus variaciones genómicas, la estructura genética de la población (incluyendo las proporciones ancestrales), algunos factores socioeconómicos y el efecto de la ancestría local, en un total de 955 y 968 controles, captados en el marco del proyecto «*Genetic Study of Common Bowel Cancer in Hispania and the Americas –CHIBCHA-*», en Colombia, para lo cual se tipificaron 1.169.944 SNPs, en la búsqueda de ampliar el conocimiento de la genética del síndrome. CHIBCHA es una iniciativa internacional, apoyada en el consorcio que lleva el mismo nombre y del cual hacen parte distintos centros de investigación del Reino Unido, España, Portugal, Brasil, México, Uruguay y Colombia, financiado por la Comisión Europea en el 7 marco (<https://cordis.europa.eu/project/rcn/91245/reporting/en>).

## Materiales y Métodos

**Poblaciones y muestras.** Las muestras de sangre se recolectaron en el marco del proyecto: «*Genetic Study of Common Bowel Cancer in Hispania and the Americas –CHIBCHA-*» (<http://www.well.ox.ac.uk/CHIBCHA/>) en los centros: Hospital Federico Lleras Acosta (Ibagué), Instituto Nacional de Cancerología (Bogotá D.C), Hospitales Pablo Tobón Uribe y San Vicente de Paul (Medellín), Hospital Universitario Fernando Moncaleano Perdomo (Neiva) y Registro Poblacional de Cáncer de Pasto, entre otros. Se empleó un total de 1923 participantes (955 casos de CCR y 968 controles), incluyendo casos menores de 75 años con diagnóstico de adenocarcinoma/adenoma de colon y recto; captados entre 2005 y 2013. Los controles cumplieron con los siguientes criterios de inclusión: Personas sin diagnóstico de cáncer, en particular neoplasia colorrectal; sin historia familiar de cáncer hasta el segundo grado de con-

sanguinidad y mayores de 55 años. La documentación del proyecto fue aprobada por el comité de Bioética de la Universidad del Tolima y de cada una de las instituciones colaboradoras. El protocolo de investigación se ciñó a los principios de la declaración de Helsinki.

**Genotipificación y control de calidad.** Se extrajo el ADN a partir de las muestras de sangre de casos y controles empleando el equipo automatizado MAXWELL-16 (*Promega® Madison, Wisconsin, USA*) y la cuantificación se desarrolló empleando un equipo Nanodrop ND-2000 (*Thermo Fischer Scientific, Waltham, MA USA*). Se genotipificaron 1.169.207 de SNPs distribuidos en dos microarreglos diferentes (Custom-550434 y LAT), usando la plataforma *GenTitan Axyom Affymetrix* (*Thermo Fischer Scientific, Waltham, MA USA*) en la Universidad Santiago de Compostela. Los datos genotípicos se sometieron a un control de calidad siguiendo protocolos reportados (22). Después de aplicar los filtros, un total de 1.169.944 marcadores quedaron para los análisis genómicos.

**Análisis estadístico y bioinformático.** Se desarrollaron los análisis de asociación en 20 SNPs de riesgo reportados previamente en otras poblaciones, y los estudios genómicos globales (ancestría global y cromosómica local).

**Ancestría global.** La ancestría genética se estimó mediante ADMIXTURE V.2, usando 87.359 SNPs. Se ejecutó el programa empezando con un  $k=2$  hasta  $k=5$ , con 100 réplicas, incorporando los casos y controles e individuos de las poblaciones parentales: África (YRI=108), Europa (CEU=99, GBR=104, IBS=107) y Nativos Americanos (Tlapanec=5, Quechua-Peruanos=24, Mexicanos-Nahua/Mixtec/Maya=34, Aymara-Bolivia=25). Los resultados se analizaron en R, bajo estadísticos no paramétricos (*U-Mann Whitney*).

**Análisis de asociación en 20 SNPs de susceptibilidad.** Las asociaciones alélicas se realizaron en PLINK (23), y empleando lenguaje de programación en R (Versión 3.2.1; <http://www.r-project.org/>) (24). Los modelos de regresión logística para la susceptibilidad al CCR tuvieron en cuenta las proporciones ancestrales globales, el sexo, el estrato socioeconómico, el nivel de educación y los componentes principales (PCs). Los PCs se calcularon en mediante el programa SmartPCA (25) y fueron usados para corregir el efecto de la estructuración poblacional.

**Ancestría local.** Se calculó la ancestría local en los 22 autosomas usando ELAI (26). Las simulaciones se desarrollaron empleando poblaciones de referencia de África, Europa y nativos americanos, usando como parámetros el número de generaciones desde que ocurrió la mezcla genética ( $y=20$ ) y, el número de poblaciones parentales a considerar ( $S=3$ ). El análisis de los resultados se realizó en lenguaje de R, mediante modelos de regresión logística para asociar la proporción ancestral por locus con el riesgo al CCR y mapear las regiones cromosómicas significativas ( $p < 10E-5$ ), usando las bases de datos (Genome Browser y ENSAMBL).

## Resultados.

Estudios previos del grupo han definido las características demográficas, clínicas y patológicas de la muestras empleadas en el presente estudio (27).

### Ancestría global en casos de CCR y en los controles.

En la tabla 1 se presentan las proporciones ancestrales en casos y controles de cada ciudad muestreada (tabla 1). Los casos y los controles presentaron principalmente ancestrías amerindia y europea, con variación amplia en cada ciudad. El componente africano fue alto ( $>10\%$ ) en algunas poblaciones: Medellín y Cali. En general,

existe una mayor proporción nativa americana en las poblaciones de Pasto, Bogotá y Neiva, mientras que es baja en ciudades del eje cafetero y Medellín.

A nivel global, la ancestría amerindia fue significativamente mayor en controles ( $x = 0.39 \pm 0.13$ ) comparada con los casos ( $x = 0.37 \pm 0.126$ ) (*U de Mann-Whitney*,  $p=0.0001621$ ). La proporción ancestral europea también fue significativamente diferente (*U Mann-Whitney*,  $p=0.0003423$ ), siendo mayor en los casos ( $x = 0.550 \pm 0.13$ ) que en los controles ( $x = 0.527 \pm 0.13$ ) (**Fig. 1**). La ancestría africana fue del  $\sim 8\%$  ( $d.s= 0.09$ ) en casos y controles, con diferencias significativas entre los dos grupos comparados ( $p=0.00016$ ). El componente nativo y el europeo fueron los de mayor proporción, mientras que la africana fue baja, pues el 75% de los individuos presentaron menos de un 10% de esta ancestría en promedio.

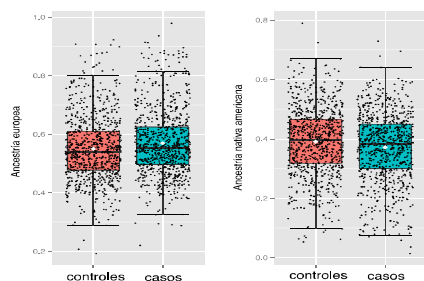
A nivel de ciudad, las diferencias entre casos y controles no son homogéneas, existen algunas donde las variaciones son más pronunciadas en unas que en otras; pero en general, aquellos donde las ancestrías nativas o europeas fueron las más altas, es donde existe significancia en las diferencias de las medias.

**Tabla 1.** Proporciones de ancestría nativas americanas europeas y africanas en los casos y controles (*media $\pm$ d.s.*), de acuerdo con la ciudad donde se captó la muestra.

Grupo	Ciudad	NAM	EUR	AFR
Controles	Bogotá	0,431 $\pm$ 0,116	0,515 $\pm$ 0,104	0,054 $\pm$ 0,073
	Bucaramanga	0,405 $\pm$ 0,071	0,561 $\pm$ 0,074	0,034 $\pm$ 0,028
	Cali	0,391 $\pm$ 0,162	0,421 $\pm$ 0,166	0,188 $\pm$ 0,233
	Ibagué	0,406 $\pm$ 0,102	0,535 $\pm$ 0,104	0,059 $\pm$ 0,065
	Medellín	0,296 $\pm$ 0,108	0,601 $\pm$ 0,137	0,103 $\pm$ 0,087
	Manizales	0,360 $\pm$ 0,196	0,563 $\pm$ 0,191	0,077 $\pm$ 0,071
	Neiva	0,435 $\pm$ 0,111	0,518 $\pm$ 0,107	0,046 $\pm$ 0,023
	Pasto	0,575 $\pm$ 0,152	0,392 $\pm$ 0,140	0,033 $\pm$ 0,085
	Bogotá	0,423 $\pm$ 0,094	0,525 $\pm$ 0,093	0,052 $\pm$ 0,075
	Bucaramanga	0,371 $\pm$ 0,110	0,583 $\pm$ 0,117	0,046 $\pm$ 0,072
Casos	Cali	0,383 $\pm$ 0,126	0,436 $\pm$ 0,144	0,181 $\pm$ 0,208
	Ibagué	0,402 $\pm$ 0,106	0,537 $\pm$ 0,108	0,061 $\pm$ 0,069
	Medellín	0,263 $\pm$ 0,097	0,635 $\pm$ 0,141	0,102 $\pm$ 0,089
	Manizales	0,288 $\pm$ 0,159	0,667 $\pm$ 0,190	0,045 $\pm$ 0,055
	Neiva	0,411 $\pm$ 0,114	0,533 $\pm$ 0,118	0,055 $\pm$ 0,036
	Pasto	0,481 $\pm$ 0,130	0,481 $\pm$ 0,121	0,038 $\pm$ 0,043

Ancestrías: NAM, nativas americanas; EUR europeas y (AFR africanas).

**Relación entre la ancestría genética global y el riesgo al CCR.** Se aplicó un modelo de regresión logística para la asociación entre las ancestrías europea/nativa y el CCR.



**Figura 1:** Proporciones de ancestría europea (a), y nativa americana (b) en casos de CCR y en controles. Se observan los cuartiles, el rango intercuartílico, la media (rombo blanco) y los puntos negros correspondientes a los individuos. Fuente: El autor.

Se definió como variable dependiente el fenotipo -con o sin la enfermedad (CCR)- y como variable independiente una de las ancestrías a la vez, dada la correlación entre NAM y EUR ( $r^2 = -0.764$ ,  $IC95\% = -0.782-0.744$ ,  $p < 2.2e^{-16}$ ). Adicionalmente se verificó la presencia de autocorrelación a través del factor de inflación de la varianza (*variance inflation factor* -*VIF*-).

En los modelos de regresión logística para cada ancestría:

$P(\text{fenotipo}|\text{ancestría}) = 1 / (1 + e^{-(\beta_0 + \beta_1 * \text{ancestría} + \dots)})$ , resultaron significativos los coeficientes para

EUR ( $\beta_1 = 1.45 \pm 0.3985$ ,  $p = 0.000267$ ) y NAM ( $\beta_1 = -1.5053 \pm 0.39$ ,  $p = 0.000114$ ). Adicionalmente, por cada unidad de incremento en EUR, el riesgo (OR) de padecer CCR se incrementó en 4.274 ( $IC95\% = 1.966-9.388$ ), mientras que para NAM se observó un OR de 0.222 ( $IC95\% = 0.103-0.475$ ). El modelo con las ancestrías fue significativamente diferente del nulo ( $H_0: \beta_1 = 0$ , solo el intersepto) (test de Wald,  $PEUR = 0.00027$ ,  $PNAM = 0.00011$ ). Al adicionar gradualmente las covariables: sexo, origen geográfico, estrato socioeconómico y nivel educativo (Tabla 2), se observó significancia en los cinco modelos ( $m1 = 0.00027$ ,  $m2 = 2e-04$ ,  $m3 = 2.5e-05$ ,  $m4 = 0.00016$ ,  $m5 = 0.002$ ).

La adición de todas las covariables en el modelo 5 presentó la mejor bondad de ajuste ( $p = 6.48E-04$ ), mostrado en el valor de verosimilitud más bajo entre los modelos ( $LogLik = -1003.1$ ), y el OR fue significativo ( $OR = 4.244$ ,  $IC95\% = 1.701-10.68$ ). El análisis de la ancestría NAM, fue similar, observándose un incremento en la bondad de ajuste a medida que se adicionaron covariables ( $LogLik_{\text{basal}} = -1030.1$  v.s  $LogLik_{\text{completo}} = -999.15$ ) (datos no mostrados). En el modelo completo, la ancestría NAM se encontró asociada con el CCR ( $OR = 0.113$ ,  $IC95\%: 0.039-0.316$ ).

**Tabla 2:** Regresiones logísticas en cinco modelos anidados para el fenotipo (casos/control).

Modelo	Variable	Coefficiente	Error estándar	Z	Pr(> z )
<b>m1.fenotipo~EUR</b> LogLik:-1030.9	(Intercepto)	-0,6893018	0,2235106	-3,083978	0,002042525
	<b>EUR</b>	1,4526075	0,3984763	3,645405	0,000266971
<b>m2.fenotipo ~EUR+GE</b> LogLik:-1028.4	(Intercepto)	-0,5739953	0,2293603	-2,502592	0,01232875
	<b>EUR</b>	1,4873788	0,39934	3,724592	0,000195631
	GE(Mujeres)	-0,2356554	0,1052472	-2,239067	0,025151589
<b>m3.fenotipo~EUR+GE+DEP</b> LogLik:-1021.4	(Intercepto)	-0,916488	0,2874985	-3,18780081	0,001433593
	<b>EUR</b>	1,85985268	0,4415393	4,21220223	2,52893E-05
	GE(Mujeres)	-0,22379373	0,106329	-2,10472992	0,035314824
<b>m4.fenotipo~EUR+GE+DEP+EST</b> LogLik:-1015.6	(Intercepto)	-0,73642801	0,293229	-2,51144366	0,012023848
	<b>EUR</b>	1,73939282	0,4600407	3,78095451	0,000156228
	GE(Mujeres)	-0,22379373	0,106329	-2,10472992	0,035314824
	Estrato_medio	-0,26040313	0,1141608	-2,28102127	0,022547188
<b>m5.fenotipo~EUR+GE+DEP+EST+EDU</b> LogLik:-1003.1	(Intercepto)	-0,64006873	0,2959652	-2,16264865	0,03056821
	<b>EUR</b>	1,44539835	0,4682979	3,08649304	0,002025327
	GE(Mujeres)	-0,15855907	0,1084598	-1,4619161	0,1437642
	Estrato_medio	-0,46057403	0,1223729	-3,76369291	0,0001674225
	Estrato_alto	0,01756536	0,2766457	0,06349408	0,9493731
	Educacion_medio	0,36341426	0,1255301	2,89503635	0,003791146
	Educación_alto	0,90141173	0,1905402	4,73082175	0,000002236128



El modelo basal (m1) fue ajustado por las covariables: sexo, origen geográfico, estrato socioeconómico y nivel educativo. El origen geográfico, correspondió al departamento/región de nacimiento. La tabla reporta el logaritmo de la verosimilitud de cada modelo (*LogLik*) y el valor estimado del coeficiente (beta).

Se analizaron rangos de la ancestría europea mediante una regresión logística, ajustando para las cuatro covariables ya mencionadas, y tomando como nivel basal para comparar, el rango de ancestría europea más bajo (0-25%). Los rangos más significativos fueron entre 50-75% ( $P=0.049$ ) y entre 75-100% ( $P=0.0407$ ), al comparar con el nivel basal; presentando *ORs* ajustados de 2.30 ( $IC95\%=1.019-5.442$ ) y de 2.64 ( $IC95\%=1.057-6.863$ ) respectivamente. Para la ancestría NAM, el intervalo significativo estuvo entre 50-75% (*OR* ajustado=0.591,  $IC95\%=0.376-0.926$ ,  $P=0.022$ ), reflejando el efecto protector frente al incremento del riesgo. Tabla 3. Algunas de las covariables resultaron significativas cómo, por ejemplo, el estrato socioeconómico y el nivel educativo, tanto como cuando se analizó la ancestría como variable continua o categórica.

**Tabla 3.** Frecuencias alélicas y *ORs* en casos de CCR y controles para los 20 SNPs analizados.

Región	SNP	Alelos	AR	FAR Casos	FAR Controles	OR alélico	IC95%	P	P*
18q21.1	rs4939827	T/C	T	0.401	0.331	1.351	(1.184-1.541)	7.35E-06	0.000147
19q13.11	rs10411210	C/T	C	0.864	0.826	1.34	(1.124-1.598)	0.001079	0.02158
10p14	rs10795668	G/A	G	0.701	0.655	1.233	(1.077-1.412)	0.0024	0.04799
14q22.2	rs4444235	C/T	C	0.452	0.408	1.201	(1.056-1.364)	0.005061	0.1012
20p12.3	rs961253	A/C	A	0.372	0.329	1.204	(1.054-1.375)	0.006159	0.1232
8q23.3	rs16892766	C/A	C	0.069	0.049	1.419	(1.082-1.861)	0.01109	0.2218
8q24.21	rs10505477	A/G	A	0.567	0.528	1.167	(1.028-1.325)	0.01709	0.3419
8q24.21	rs7014346	A/G	A	0.338	0.304	1.169	(1.021-1.339)	0.0236	0.472
15q13.3	rs4779584	T/C	T	0.287	0.255	1.175	(1.019-1.355)	0.02636	0.5271
6p21.2	rs1321311	A/C	A	0.235	0.207	1.183	(1.015-1.377)	0.03102	0.6203
8q24.21	rs6983267	G/T	G	0.575	0.542	1.144	(1.007-1.299)	0.03849	0.7698

AR: alelo de riesgo, FAR: frecuencia del alelo de riesgo, IC: intervalo de confianza, P: valor p crudo, P\*: valor p ajustado.

**Interacciones entre los SNPs conocidos y la ancestría.** Once de los 20 marcadores se asociaron significativamente con el riesgo al CCR en la población colombiana ( $P<0.05$ ), pero solo tres después del ajuste de Bonferroni (*rs4939827*,  $P=0.000147$ ; *rs10411210*,  $P=0.02158$  y *rs10795668*,  $P=0.04799$ ). Se analizó la relación de los 20 SNPs en conjunto con la ancestría, mediante la implementación del siguiente modelo de regresión logística:  $P(\text{fenotipo}/\text{ancestría}, \text{SNP}) = 1 / (1 + e^{-(\beta_0 + \beta_1 * \text{ancestría} + \beta_2 * \text{SNP} + i)})$ , adicionando las covariables sociodemográficas y los PCs calculados a partir de 87.359 SNPs. En la tabla 4 se muestran los resultados de las asociaciones con un SNP de riesgo a la vez, se aprecian solo los resultados para las variables de interés.

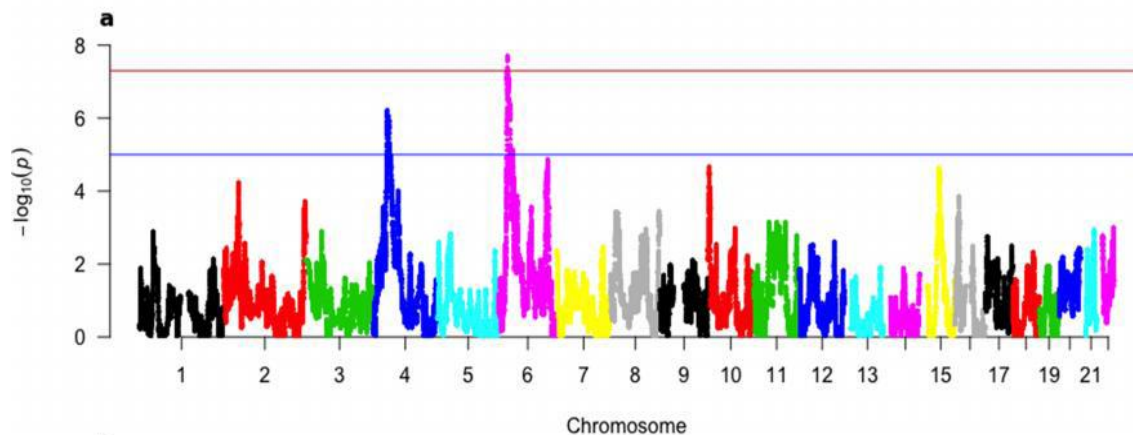
**Tabla 4.** Variantes SNPs significativamente asociadas con el riesgo al CCR, mediante modelos de regresión logística. Los modelos incluyeron la adición de la ancestría europea en cada caso.

CRO	SNP	AR	TEST	SE	COEF	OR	IC95%	P
8	rs16892766	C	ADD	0,1747	2.366	1.512	(1.074-2.129)	0,01798
			EUR	0,5806	3.013	5.752	(1.843-17,95)	0,002585
10	rs10795668	G	ADD	0,08085	2.051	1,18	(1.007-1.383)	0,04026
			EUR	0,5822	3.084	6.023	(1.924-18,85)	0,002042
14	rs4444235	C	ADD	0,07721	2,47	1,21	(1,04-1.408)	0,01349
			EUR	0,587	2.738	4.988	(1.579-15,76)	0,006186
18	rs4939827	T	ADD	0,07923	2.817	1,25	(1,07-1,46)	0,00485
			EUR	0,5846	2.786	5.098	(1.621-16,03)	0,005332
19	rs10411210	C	ADD	0,1049	2.078	1.244	(1.012-1.527)	0,03768
			EUR	0,5809	3.033	5.823	(1.865-18,18)	0,002424

CRO: cromosoma, ADD: modelo aditivo para cada SNP, EUR: ancestría europea, SE: error estándar, COEF: coeficiente en la regresión, IC95%: intervalo de confianza al 95%, P: valor p.

Dado que la ancestría europea se encontró correlacionada con el PC2 ( $corr=-0.8261193$ ,  $p < 2.2e-16$ ) y el PC4 ( $corr=-0.74$ ,  $p < 2.2e-16$ ), fue necesario excluirlos de los modelos; por otro lado, ambos PCs estuvieron correlacionados entre sí ( $corr=0.62$ ,  $p\text{-val} < 2.2e-16$ ). Finalmente se obtuvieron cinco marcadores significativamente asociados, y también la ancestría europea. Al incluir en un solo modelo a los 5 SNPs, las variables sociodemográficas y los componentes principales 1 y 3, solo tres marcadores permanecieron significativos (rs16892766  $p=0.03236$ , rs444423  $p=0.01570$ , rs4939827  $p=0.03752$ ), además de la ancestría europea (OR=4.374, IC95%=1.37-14.21,  $p=0.01346$ ).

**Mapeo de la mezcla.** Se calculó la ancestría por cada locus, evaluado a lo largo de todos los autosomas, y después de realizar la asociación entre casos y controles mediante modelos de regresión logística (mapeo de la mezcla), se encontraron dos regiones cromosómicas con una ancestría europea significativamente más alta en casos que en controles así: **4p13** entre 41.8 y 45 Mb ( $OR_{crudo}=1.380$ ,  $IC95\%=1.216-1.566$ ;  $p_{crudo}=5.92e-07$ ) y **6q22.3** entre 24.1 y 30 Mb ( $OR_{crudo}=1.45$ ,  $IC95\%=1.275-1.653$ ;  $p_{crudo}=2.02e-8$ ), lo cual se mantuvo después del ajuste por ancestría global tanto en 6q22.3 ( $OR_{ajustado}=1.378$ ,  $IC95\%=1.202-1.580$ ,  $P_{ajustado}=4.2e-6$ ), como para 4p13 ( $OR_{ajustado}=1.301$ ,  $IC95\%=1.137-1.489$ ;  $P_{ajustado}=0.00013$ ) (Fig. 2). En la región del cromosoma 6 se encontraron varios genes en una ventana de 500 Kb alrededor del SNP de mayor asociación, tales como el LRRC16A (chr6:25279428 – 25620530), entre otros. En cuanto la región 4p13, es posible visualizar varias características en una ventana de 500 Kb, alrededor del SNP con mayor asociación, tales como el gen SHISA3 (Homólogo de la proteína *shisa 3*); también se encuentra ATP8A1 (Transportadora de fosfolípidos 8A1 con actividad ATPasa).



**Figura 2.** Gráfico de Manhattan para el logaritmo negativo de la probabilidad cruda de asociación entre la dosis de ancestría europea en cada locus autosómico evaluado, cuando se compararon casos de CCR y controles colombianos. Se observan los picos principales detectados mediante el mapeo de la mezcla, localizados en el cromosoma cuatro y el seis.

## Discusión.

Las características genéticas de la población colombiana actual, fueron definidas por la diversidad y densidad de las poblaciones indígenas en periodos precolombinos y los patrones de mestizaje ocurridos entre dichas poblaciones originales con europeos y africanos. La variación en la ancestría genética observada, originada por la historia demográfica de la región andina colombiana, puede emplearse para explicar la susceptibilidad a enfermedades como el CCR. Por ejemplo, la alta incidencia del CCR en europeos (68-44/100.000) comparada con poblaciones nativas (34/100.000), sugiere que en las poblaciones mestizas como latinos, dicha enfermedad al ser de carácter complejo, se presente en tasas intermedias (28), lo cual así sucede (58-50/100.000) (29); por lo tanto la variación genética en relación a la contribución del riesgo, pudo haber cambiado de frecuencias por procesos estocásticos de deriva durante colonia, especialmente debido al mestizaje.

En el presente estudio se observó mayor proporción del componente ancestral europeo en los casos que en controles, y viceversa para la proporción NAM. Ambas ancestrías fueron altamente significativas en los modelos logísticos analizados, indicando que son variables críticas en la predicción del CCR en poblaciones mestizas colombianas. Adicionalmente, se encontró que presentar una ancestría mayor al 50%, está asociada con el riesgo para el CCR. Lo anterior indica que uno o varios alelos con altas frecuencias en la población europea, y presentes en la población andina colombiana, pueden incrementar el riesgo al CCR en la misma. Aunque dicha asociación también podría deberse a factores medioambientales comunes en la región, no evaluados. En este sentido, es probable que estén presentes interacciones gen-medio ambiente, con la capacidad de explicar las diferencias observadas.

Aunque en este trabajo el riesgo a desarrollar CCR también estuvo asociado a variables como el nivel educativo y el socioeconómico, dado que los casos fueron recolectados en instituciones públicas de salud, existe un sesgo ya que las personas con capacidad económica mayor frecuentemente recurren a centros privados, en lugar de usar el régimen subsidiado (30). En consecuencia, es necesario controlar la presencia de estratificación poblacional como posible variable de confusión en los estudios de asociación en Colombia.

El papel de la ancestría nativa americana en el riesgo a CCR, sugiere que los Nativos americanos podrían tener una menor probabilidad de padecer la enfermedad. No obstante, los estudios o reportes sobre la incidencia en poblaciones indígenas son limitados. En algunas poblaciones indígenas latinoamericanas el cáncer de cérvix, estómago, próstata y vesícula biliar (31) son los más frecuentes. Con relación al CCR, se ha reportado una menor incidencia en nativos de Alaska entre los años 1999-2013 (32), y en general Norteamérica existen bajas tasas de incidencia y mortalidad por CCR (33). Para Colombia, el Ministerio de Seguridad Social en 2016 (34), encontró que el cáncer gástrico es el de mayor representación en indígenas. Estudios indican que en las mujeres de la Amazonía existe un alto riesgo para el cáncer de cérvix, mientras que en los hombres del cauca lo es para el cáncer gástrico, en ambos casos, la densidad de población indígena está dentro de las más altas del país (31).

Con respecto al riesgo que aportaron los 20 SNPs y su relación con la ancestría, estos datos proporcionan un

soporte para las causas genéticas comunes en la susceptibilidad del CCR, cuando se comparan distintas poblaciones como la colombiana y la europea. La asociación de la ancestría europea global continuó aportando al riesgo en conjunto con los SNPs. Dentro de las regiones asociadas se han identificado las variantes funcionales como en 18q21.1 caracterizada por la presencia del gen *SMAD7* (35), en 19q13.11 está la proteína 2 de unión *Rho GTPasa (RHPN2)* (36), vinculada con la regulación del citoesqueleto de actina y la motilidad celular, y por tanto con el crecimiento celular maligno (37). En cuanto a 8q23.3, hay secuencias reguladoras de la transcripción, y transcritos predichos: *EIF3H* y *C8orf53*, que influyen en la regulación del crecimiento celular y la viabilidad en distintos tipos de carcinoma tales como el de seno, próstata y hepatocelular (38). En el caso de 10p14, no hay genes predichos a una distancia de 0.4 Mb alrededor del SNP para esta región; es una variante intergénica con los pseudogenes *KRT8P37* (pseudogene 37 keratin 8) y *KRT8P16* (pseudogene 16 keratin 8). Existe también un cDNA correspondiente a un transcrito de secuencia *ALU* de 234 pb (39) asociado con cáncer gástrico, glándula mamaria (*MUTYH*, subfamilia *AluYb8*) (40) y el síndrome de *Peutz Jaeger* entre otras (*STK11*, subfamilia *AluY*) (41).

El empleo del mapeo de la mezcla permitió asociar la alta ancestría europea en los *loci* 6q22.3 y 4p13, que explican significativamente la susceptibilidad al CCR en la muestra. Este resultado es relevante para la comprensión del impacto que tiene el mestizaje en la población colombiana y lo importante que puede llegar a ser esta metodología en el hallazgo de genes de susceptibilidad de enfermedades complejas como el CCR o el carcinoma de glándula mamaria, y por tanto para detectar regiones que expliquen parte de la heredabilidad perdida de las mismas (20).

Algunas secuencias que hacen parte de 6q22.3 y 4p13 corresponden a pseudogenes, vinculados con distintos tipos de cáncer (42). Alrededor de 6q22.3 se observaron varios genes, como el *LRR16A (Leucine-rich repeat-containing protein 16A)*, localizado en 25.27 Mb, promueve la protrusión de los lamelipodios y la migración celular (43), asociado con el conteo de plaquetas (44). Otro gen es el *SCGN (secretogogina)* a 25.65 Mb, que produce una proteína de unión a calcio, asociada con la calbindina *D-28K* y la calretinina, vinculado con proliferación celular y su sobreexpresión está asociada con la inhibición de la apoptosis, una alta metástasis y quimioresistencia (45) y expresión considerable en el tejido



colónico (46). El gen *GMNN* (*Geminin*) localizado a 24.77 Mb, es un inhibidor de la replicación del DNA que previene la incorporación del complejo MCM en el complejo prereplicación, por lo tanto es degradado durante las fases mitóticas del ciclo celular (47), y asociado con el CCR (48, 49). La región 4p13 (41.8-45 Mb) se localizan genes que han sido estudiados en el CCR, por ejemplo, el *SHISA3* en la posición 42.39 Mb, un gen supresor de tumores, importante en la atenuación de las vías de señalización *FGF* y *WNT* (50) y podría actuar como un gen supresor de tumores acelerando la degradación de la *-catenina* (51), un elemento crucial en la vía canónica *WNT* en el epitelio colónico, que regula la proliferación y diferenciación de las células madres y progenitoras (52). otro gen es el *ATP8A1*, localizado a 42.41 Mb, vinculado con el crecimiento y comportamiento invasivo en cáncer de pulmón (53), y está implicada en mecanismos activos durante la migración celular (54). Los genes anteriores, entre otros localizados en 6q22.3 y 4p13, pueden constituir las variantes funcionales implicadas en el riesgo, pero se requieren estudios fisiológicos celulares para aclarar el rol biológico en el CCR y su aporte a la heredabilidad.

### Conclusiones.

La comprensión de la historia natural del CCR en Colombia puede incrementar si se cuenta con datos complementarios de poblaciones indígenas, para tener un conocimiento amplio de su genética. Aunque algunas regiones identificadas en poblaciones europeas fueron replicadas en el presente trabajo, y otras nuevas (6q22.3 y 4p13), el papel de la ancestría nativa no es concluyente, a pesar de su relación opuesta con el incremento del riesgo al CCR. El significado biológico de las regiones que resultaron asociadas bajo las diferentes metodologías (replicación de las regiones conocidas y mapeo de la mezcla) revela un trasfondo genético-molecular común, en lo que hace referencia al desarrollo y patogénesis del carcinoma colorrectal en las poblaciones humanas. Este trasfondo se concreta en vías de señalización celular, previamente identificadas en estudios en europeos y asiáticos como: *Wnt*, *TGF*, *EGFR*, *FGF*, las cuales regulan procesos homeostáticos del crecimiento, diferenciación y la migración celular y hacen parte de una red, no muy conocida, en la que falta tener en cuenta

a las secuencias intergénicas reguladoras, que pueden estar siendo no sólo afectadas por mutaciones heredables, sino también reguladas por procesos epigenéticos o a través de la interacción con factores ambientales.

**Agradecimientos:** Agradecemos a los donantes que participaron en el estudio. También a las instituciones de salud, y a aquellas que financiaron el proyecto: Colciencias, la Unión Europea, la Universidad de California en Davis y la Universidad del Tolima. AAC-R recibió financiamiento de Colciencias: “Becas Doctorales Nacionales, Convocatoria 628 2011”.

**Conflicto de intereses y financiación.** Los autores declaran que no hay conflictos de interés.

**Anexo:** Lista completa Consorcio CHIBCHA

Ma. Magdalena Echeverry de Polanco, Mabel Elena Bohórquez, Rodrigo Prieto, Angel Criollo, Carolina Ramírez, Ana Patricia Estrada, Jhon Jairo Suárez (*Grupo de Citogenética Filogenia y Evolución de Poblaciones, Universidad del Tolima, Colombia*); Augusto Rojas Martinez (*Center for Research and Development in Health Sciences, Universidad Autónoma de Nuevo León, Monterrey, Mexico*); Silvia Rogatto, Samuel Aguiar Jnr, Ericka Maria Monteiro Santos (*Department of Urology, School of Medicine, UNESP - São Paulo State University, Botucatu, Brazil*); Monica Sans, Valentina Colistro, Pedro C. Hidalgo, Patricia Mut (*Department of Biological Anthropology, College of Humanities and Educational Sciences, University of the Republic, Magallanes, Montevideo, Uruguay*); Angel Carracedo, Clara Ruiz Ponte, Ines Quntela Garcia (*Fundacion Publica Galega de Medicina Xenomica, CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Hospital Clinico, Santiago de Compostela, Galicia, Spain*); Sergi Castellvi-Bel (*Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, IDIBAPS, University of Barcelona, Barcelona, Catalonia, Spain*); Manuel Teixeira (*Department of Genetics, Portuguese Oncology Institute, Rua Dr. António Bernardino de Almeida, Porto, Portugal*).

## Referencias

1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. (2014) Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin.* 64(4):252-71.
2. Kriza C, Emmert M, Wahlster P, Niederländer C, Kolominsky-Rabas P. (2013) Cost of illness in colorectal cancer: An international review. *Pharmacoeconomics.* 31(7):577-88.
3. Soerjomataram I, Lortet-Tieulent J, Parkin DM, Ferlay J, Mathers C, Forman D, et al. (2012) Global burden of cancer in 2008: a systematic analysis of disability-adjusted life-years in 12 world regions. *Lancet.* 380(9856):1840-50.
4. Burt RW, Barthel JS, Dunn KB, David DS, Drelichman E, Ford JM, et al. (2010) Colorectal cancer screening. *J Natl Compr Canc Netw.*8(1):8-61.
5. Davies RJ, Miller R, Coleman N. (2005) Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer.* 5(3):199-209.
6. Costi R, Leonardi F, Zanoni D, Violi V, Roncoroni L. (2014) Palliative care and end-stage colorectal cancer management: The surgeon meets the oncologist. *World J Gastroent.* 20(24):7602-21.
7. Global Burden of Disease Cancer C. (2015) The Global Burden of Cancer 2013. *JAMA oncology.* 1(4):505-27.
8. Center MM, Jemal A, Smith RA, Ward E. (2009) Worldwide variations in colorectal cancer. *CA Cancer J Clin.* 59(6):366-78.
9. Bray F, Piñeros M. (2016) Cancer patterns, trends and projections in Latin America and the Caribbean: a global context. *Salud Pub Mex.* 58(2):104-17.
10. Curado MP, de Souza DLB. (2014) Cancer Burden in Latin America and the Caribbean. *Ann Glob Health.* 80(5):370-7.
11. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. (2016) Global patterns and trends in colorectal cancer incidence and mortality. *Gut.* 66(4): 683-91.
12. Piñeros M, Gamboa O, Hernández-Suárez G, Pardo C, Bray F. (2013) Patterns and trends in cancer mortality in Colombia 1984–2008. *Cancer Epidemiol.* 37(3):233-9.
13. Ramírez R, Agredo RA. (2012) El sedentarismo es un factor predictor de hipertrigliceridemia, obesidad central y sobrepeso. *Rev Colomb Cardiol.* 19:75-9.
14. Peters U, Bien S, Zubair N. (2015) Genetic Architecture of Colorectal Cancer. *Gut.* 64(10):1623-36.
15. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. (2014) *Human evolutionary genetics.* 2 ed. New York-USA: Garland Science, Taylor & Francis Group, LLC. 670 p.
16. Shriner D. (2013) Overview of Admixture Mapping. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al].* CHAPTER: Unit1.23-Unit1.
17. Browning SR, Grinde K, Plantinga A, Gogarten SM, Stilp AM, Kaplan RC, et al. (2016) Local Ancestry Inference in a Large US-Based Hispanic/Latino Study: Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *G3 (Bethesda).* 6(6):1525-34.
18. Sofer T, Baier LJ, Browning SR, Thornton TA, Talavera GA, Wassertheil-Smoller S, et al. (2017) Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits. *PLoS One.* 12(11): e0188400.
19. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, et al. (2014) GWAS and admixture mapping identify different asthma-associated loci in Latinos: The GALA II Study. *J allergy clin immunol.* 134(2):295-305.
20. Fejerman L, Chen GK, Eng C, Huntsman S, Hu D, Williams A, et al. Admixture Mapping Identifies a Locus on 6q25 Associated with breast cancer risk in US Latinas. *Hum Mol Genet.* 2012.
21. Criollo-Rayo AA, Bohorquez M, Prieto R, Howarth K, Culma C, Carracedo A, et al. (2018) Native American gene continuity to the modern admixed population from the Colombian Andes: Implication for biomedical, population and forensic studies. *Forensic Sci Int Genet.* 36: e1-e7.
22. Anderson C, Pettersson F, Clarke G, Cardon L, Morris A, Zondervan K. (2010) Data quality control in genetic case-control association studies. *Nat Protocols.* 5(9):1564-73.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.*

- 81(3):559-75.
24. R Core Team. (2014) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
  25. Price AL, Zaitlen NA, Reich D, Patterson N. (2010) New approaches to population stratification in genome-wide association studies. *Nat rev Genet.* 11(7):459-63.
  26. Guan Y. (2014) Detecting Structure of Haplotypes and Local Ancestry. *Genetics.* 196(3): 625-642.
  27. Bohórquez M, Sahasrabudhe R, Criollo A, Sanabria-Salas MC, Vélez A, Castro JM, et al. (2016) Clinical manifestations of colorectal cancer patients from a large multicenter study in Colombia. *Medicine.* 95(40): e4883.
  28. Risch N, Merikangas k. (1999) The Future of Genetic Studies of Complex Human Diseases. *Science.* 273:1516-7.
  29. Perdue DG, Haverkamp D, Perkins C, Daley CM, Provost E. (2014) Geographic variation in colorectal cancer incidence and mortality, age of onset, and stage at diagnosis among american indian and Alaska native people, 1990–2009. *Am J Public Health.* 104(Suppl 3): S404-S14.
  30. Webster PC. (2012) Health in Colombia: a system in crisis. *Can Med Assoc J.* 184(6): E289-E90.
  31. Moore SP, Forman D, Piñeros M, Fernández SM, Oliveira Santos M, Bray F. (2014) Cancer in indigenous people in Latin America and the Caribbean: a review. *Cancer Med.* 3(1):70-80.
  32. Center for Disease Control and Prevention. (2016) Colorectal cancer rates by race and ethnicity: division of cancer prevention and control, Centers for Disease Control and Prevention. [updated 16-junio-2016]. Available from: <https://http://www.cdc.gov/cancer/colorectal/statistics/race.htm>.
  33. Cobb N, Paisano RE. (1998) Patterns of cancer mortality among Native Americans. *Cancer.* 83(11):2377-83.
  34. Sandoval-Castaño EA. (2016) Perfil de Salud de la Población Indígena, y Perfil de salud de la población indígena, y medición de desigualdades en salud. Colombia 2016. Bogotá: Ministerio de Salud y Protección Social. Producto desarrollado en el marco del contrato No. 074 de 2016.
  35. Fortini BK, Tring S, Plummer SJ, Edlund CK, Moreno V, Bresalier RS, et al. (2014) Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One.* 9(11): e111914.
  36. Study, C. O. G. E. N. T. (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet.* 40 (12):1426-35.
  37. Chang Y-WE, Marlin JW, Chance TW, Jakobi R. (2006) RhoA mediates cyclooxygenase-2 signaling to disrupt the formation of adherens junctions and increase cell motility. *Cancer Res.* 66(24):11700-8.
  38. Okamoto H, Yasui K, Zhao C, Arii S, Inazawa J. (2003) PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated with large hepatocellular carcinomas. *Hepatology.* 38(5):1242-9.
  39. Liu WM, Maraia RJ, Rubin CM, Schmid CW. (1994) Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.* 22(6):1087-95.
  40. Zhu M, Chen X, Zhang H, Xiao N, Zhu C, He Q, et al. (2011) AluYb8 Insertion in the MUTYH Gene and risk of early-onset breast and gastric cancers in the Chinese population. *Asian Pac J Cancer Prev.* 12(6):1451-5.
  41. Borun P, De Rosa M, Nedoszytko B, Walkowiak J, Plawski A. (2015) Specific Alu elements involved in a significant percentage of copy number variations of the STK11 gene in patients with Peutz-Jeghers syndrome. *Fam Cancer.* 14(3):455-61.
  42. Xiao-Jie L, Ai-Mei G, Li-Juan J, Jiang X. (2015) Pseudogene in cancer: real functions and promising signature. *J Med Genet.* 52(1):17-24.
  43. Lanier MH, McConnell P, Cooper JA. (2016) Cell migration and invadopodia formation require a membrane-binding domain of CARMIL2. *J Biol Chem.* 291(3):1076-91.
  44. Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, et al. (2011) New gene functions in megakaryopoiesis and platelet formation. *Nature.* 480(7376):201-208.
  45. Bai Y, Sun Y, Peng J, Liao H, Gao H, Guo Y, et al. (2014) Overexpression of secretagogen inhibits cell apoptosis and induces chemoresistance in small cell lung cancer under the regulation of miR-494. *Oncotarget.* 5(17):7760-75.

46. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody based proteomics. *Proteomics. Mol Cel Prot*: 13(2):397-406.
47. Wu M, Lu W, Santos RE, Frattini MG, Kelly TJ. (2014) Geminin inhibits a late step in the formation of human pre-replicative complexes. *J Biol Chem*. 289(44):30810-21.
48. Kushwaha PP, Rapalli KC, Kumar S. (2016) Geminin a multi task protein involved in cancer pathophysiology and developmental process: A review. *Biochimie*. 131:115-27.
49. Nishihara K, Shomori K, Tamura T, Fujioka S, Ogawa T, Ito H. (2009) Immunohistochemical expression of geminin in colorectal cancer: Implication of prognostic significance. *Oncol Rep*. 21:1189-95.
50. Furushima K, Yamamoto A, Nagano T, Shibata M, Miyachi H, Abe T, et al. (2007) Mouse homologues of Shisa antagonistic to Wnt and Fgf signalings. *Dev Biol*. 306(2):480-92.
51. Chun-Chieh C, Hsuan-Yu C, Kang-Yi S, Qi-Sheng H, Bo-Shiun Y, Ching-Hsien C, et al. (2014) Shisa3 is associated with prolonged survival through promoting  $\beta$ -catenin degradation in lung cancer. *Am J Respir Crit Care Med*. 190(4):433-44.
52. Kim BM, Mao J, Taketo MM, Shivdasani RA. (2007) Phases of canonical Wnt signaling during the development of mouse intestinal epithelium. *Gastroenterology*. 33(2):529-38.
53. Dong W, Yao C, Teng X, Chai J, Yang X, Li B. (2016) MiR-140-3p suppressed cell growth and invasion by downregulating the expression of ATP8A1 in non-small cell lung cancer. *Tumor Biol*. 37(3):2973-85.
54. Kato U, Inadome H, Yamamoto M, Emoto K, Kobayashi T, Umeda M. (2013) Role for Phospholipid Flippase Complex of ATP8A1 and CDC50A proteins in cell migration. *J Biol Chem*. 288(7):4922-34.