

# Fondef

FONDO DE FOMENTO AL DESARROLLO  
CIENTÍFICO Y TECNOLÓGICO

## **TRANSFERENCIA DE RESULTADOS AL MINSITERIO DE SALUD**

**PROYECTO DE  
INVESTIGACIÓN Y DESARROLLO  
FONDEF**

**Título del Proyecto**

**GENÓMICA DE LA POBLACIÓN CHILENA: OBTENCIÓN DE  
PERFILES GENÉTICOS NECESARIOS EN INVESTIGACIÓN  
CLÍNICA, SALUD PÚBLICA Y MEDICINA FORENSE**

2 de Enero 2012 al 2 de Noviembre 2015

## RESUMEN

A comienzos del año 2012 cuando se iniciaba el presente proyecto, la información existente acerca de la composición genética de la población mixta chilena se había restringido a regiones muy particulares del genoma y a muestras de población originaria y de muy pocos centros urbanos: la estructura genética de la mayor parte de los chilenos se desconocía.

El presente proyecto estudió la variabilidad genética de la población mixta chilena, y la caracterizó mediante el estudio de marcadores genéticos informativos de ancestría repartidos por todo el genoma. Esta caracterización se hizo en base al estudio de genotipos de personas provenientes de 8 ciudades del país (n = 3216). La caracterización genotípica se hizo en 3 etapas: 1° secuenciación genómica a baja cobertura de 18 chilenos con alta ancestría amerindia, 2° genotipificación de alta densidad con plataforma AXIOM LAT1 (800-000 SNPs) de un subconjunto de 475 de los chilenos muestreados y 3° Genotipificación de los individuos restantes (2900) con un panel de 150 SNPs especialmente seleccionados, como los más informativos de ancestría para chilenos. Estos 150 SNPs se seleccionaron en base a la información recogida en las etapas de genotipificación 1° y 2°.

El resultado del proyecto es un perfil genético para cada una de las poblaciones muestreadas (frecuencias génicas y porcentaje de mezcla amerindia-caucásica-africana para los 150 marcadores estudiados) y la definición de un set acotado de marcadores genéticos capaces de distinguir entre subpoblaciones chilenas. Los componentes de ancestría para los chilenos resultaron ser en promedio 51 % Europeo, 5 % Africano y 44 % Amerindio, descomponiéndose este último componente en un 20 % Aymara y 24 % Mapuche. Estos porcentajes varían según la región de muestreo, de residencia, nivel socioeconómico y antecedentes de antepasados originarios, entre otras. Se montó un portal web con toda esta información con el fin de que sea utilizada por autoridades de salud e investigadores básicos y clínicos. Autoridades de salud podrán diseñar con estos resultados, políticas específicas de salud (prevención y tratamiento) de acuerdo al componente étnico propio de cada subpoblación (el que se correlaciona con su morbimortalidad) y los investigadores contarán con información necesaria e indispensable para los estudios de asociación entre genotipo y enfermedad.

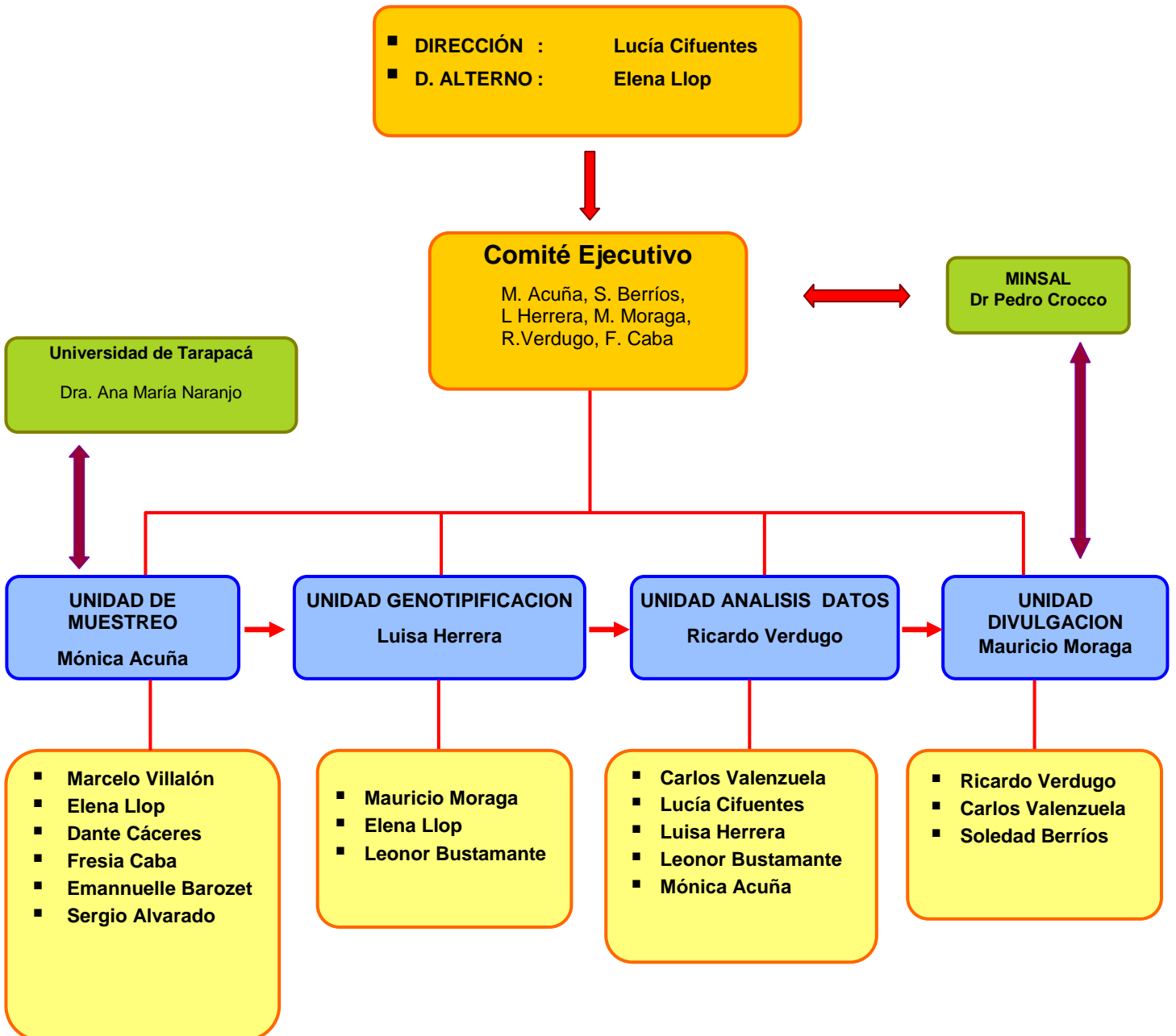
# INSTITUCIONES PARTICIPANTES

**Institución Beneficiaria:** Universidad de Chile, Facultad de Medicina

**Institución Mandante:** Ministerio de Salud

**Entidad Asociada:** Universidad de Tarapacá

Investigadores participantes: Los investigadores principales del Proyecto incluyen genetistas del Programa de Genética Humana de la Universidad de Chile (L. Cifuentes, E Llop, L Herrera, M Moraga, M Acuña, S Berríos, R Verdugo, C Valenzuela, L Bustamante), Salubristas ( S Alvarado, F Caba, D Cáceres, M Villalón), Socióloga (E Barozet) y participan también ingenieros bioinformáticos del Centro de Modelamiento matemático de la Facultad de Ingeniería de la U de Chile (A Maass, N Loira y A Digenova). El equipo estuvo organizado de acuerdo al siguiente organigrama.



## INTRODUCCION

Los avances científicos han permitido el estudio y conocimiento de la composición o estructura genética de diversas poblaciones humanas. A partir de esta información se ha logrado establecer las principales diferencias genéticas entre ellas. Los grupos genéticamente más diferenciados son: Caucásicos, Africanos, Asiáticos y Amerindios (Chung et al., 2009). El conocimiento genético poblacional humano tiene aplicaciones diversas, entre las cuales cabe destacar: a) individualización genética, b) cuantificación del riesgo poblacional a desarrollar un fenotipo patológico, c) identificación en el pool genético poblacional de los genotipos susceptibles a responder anómalamente a medicamentos (RAM), d) establecimiento de las variantes de los blancos farmacéuticos a nivel poblacional, e) especificación de la heterogeneidad genética de los fenotipos patológicos dentro y entre poblaciones.

Chile cuenta con información genética poblacional, que ha permitido dar respuestas a algunos de los puntos señalados en los ámbitos de la salud pública, la epidemiología y la medicina forense, información que se detallará más adelante. No obstante, del gran trabajo que han hecho los genetistas poblacionales chilenos, no ha sido trascendente para alcanzar el nivel en que se encuentra la discusión científica que se está llevando en los países desarrollados, orientada a resolver los puntos anteriormente señalados. Para ser parte activa y relevante de la discusión de las diversas hipótesis científicas que se están manejando, necesitamos dar un salto cualitativo y cuantitativo en términos de la descripción genético poblacional chilena, con marcadores informativos de ancestría o AIMS (*ancestry informative markers*) (Chung et al., 2009). Estos marcadores son los polimorfismos únicos de nucleótidos (SNPs) del DNA, distribuidos por todo el genoma.

Entre los años 1985 y 2010 se han publicado numerosos trabajos que describen con AIMS la estructura genética poblacional humana (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>). A la fecha, la base de datos de SNPs (dbSNP) del NIH (<http://www.ncbi.nlm.nih.gov/SNP/>) registra datos de millones de SNPs con sus frecuencias alélicas. Estudios comparativos muestran diferencias étnicas en las frecuencias dentro y entre poblaciones (Kosoy et al., 2009). Al inicio del presente proyecto no había ningún estudio de población chilena con este tipo de marcadores genéticos y los escasos estudios que han aparecido durante el curso de nuestro proyecto solo describen poblaciones muy particulares (chilenos elegidos por padecer una enfermedad en particular o bien reclutados en un punto acotado del país).

Las poblaciones latinoamericanas resultaron de una miscegenación de la población nativa, o amerindios, con los conquistadores españoles que ocurrió entre los siglos XVI y XVII y bajo el dominio ibérico posterior que continuó hasta el siglo XIX. En efecto, más de un millón de españoles y portugueses migraron al nuevo mundo, sumado al efecto de la llegada de más de 5 millones de esclavos africanos (Bedoya et al., 2006). Los efectos de estos cambios demográficos se reflejan en una serie de diferentes tipos y grados de miscegenación a través de toda Latinoamérica. En el caso de Sudamérica, a la llegada de los conquistadores europeos, los grupos amerindios mostraban una heterogénea distribución, ubicándose principalmente en torno al eje de los Andes y escasamente hacia las costas de Brasil y Argentina. Luego, los portugueses y españoles se distribuyeron por todo el continente exceptuando el extremo austral. La

llegada de los esclavos africanos se centró principalmente en las parte norte y en las costas del continente (Vieira et al., 2002).

En los párrafos subsiguientes, se puede apreciar que se han realizado una serie de estudios sobre la composición genética de poblaciones latinoamericanas en diversos países. Sin embargo, estos resultados no pueden extrapolarse directamente a la población chilena debido a que ésta presenta algunas significativas diferencias en su origen respecto a otras poblaciones de la región. Esta composición y su origen se describen a continuación.

### **Origen de la Población chilena**

La población humana chilena contemporánea ha emergido de una mezcla de dos componentes étnicos principales: caucásico y amerindio. Hace unos 45.000 años se separó de un tronco ancestral de la especie humana, desde una región de medio oriente, una rama que migró hacia el poniente (Europa, Nor-África, Asia-europea, Asia Menor) y constituyó el grupo denominado caucásico y otra rama que migró al oriente y al norte (Asia central, oriental y del norte) y que constituyó el grupo denominado mongoloide o mongoliano. Este último grupo, después de detenerse algunos milenios en la región siberiana cercana al Estrecho de Behring, invadió América del Norte y constituyó la mayor parte de los grupos étnicos conocidos como Amerindios. Existe un amplio consenso que el ingreso de los Amerindios a América ocurrió antes de los 20.000 años AP tal como lo soportan tanto la información molecular, morfológica y el registro arqueológico (Fagundes 2008, Gonzalez-Jose 2008, Goebel 2008).

Como es sabido el descubrimiento y colonización de América por parte de los europeos especialmente de la península ibérica y más tarde por franceses, ingleses y holandeses ocurren desde los viajes de Colón al fin del siglo XV. Los españoles llegan a Chile a mediados del siglo XVI (1535-1541) y desde entonces se ha producido una miscegenación étnica durante estas últimas 16 generaciones (considerando generaciones de 30 años promedio). Como Chile fue una colonia militar durante 300 años, la llegada de varones europeos fue mayor que la llegada de mujeres europeas, especialmente durante la Conquista y Colonia inicial. De allí que la miscegenación se dio natural y preferentemente entre varones europeos caucásicos y mujeres amerindias (Valenzuela, 1975; Valenzuela et al., 1987, Cifuentes et al. 2004).

De este conocimiento histórico, surge inmediatamente el interés por conocer el distinto destino evolutivo que los diferentes componentes genéticos tienen al producirse una mezcla racial tan asimétrica. La herencia nuclear autosómica en esta mezcla (hipotéticamente de sólo varones caucásicos y mujeres amerindias) es  $\frac{1}{2}$  caucásica y  $\frac{1}{2}$  amerindia. Por su parte, la herencia citoplasmática o mitocondrial es 100% amerindia, la del cromosoma Y es 100% caucásica y la del cromosoma X es  $\frac{2}{3}$  amerindia y  $\frac{1}{3}$  caucásica (Valenzuela 1975). Desde el siglo XVII, se agrega también un componente africano negroide, que se hizo presente a través de la llegada de esclavos venidos desde el Perú o desde otros países sudamericanos. Este componente llegó a ser cerca del 7% de la población (Thayer Ojeda, 1919). Nuestros estudios de las poblaciones urbanas han revelado que el componente africano es exiguo o inexistente y se hace mayor en zonas aisladas (Cifuentes et al., 1995).

El grupo amerindio que estaba presente en el territorio que hoy es reconocido como Chile no era homogéneo. Se sabe de una docena de grupos con culturas y lenguas o dialectos distintos. Sin embargo, la estructura genética diferencial de estos grupos no es bien conocida y sólo tenemos una apreciación de ella en base a marcadores genéticos clásicos (pre-genómicos). Esta apreciación no permite tener un panorama preciso sobre la constitución genética de los cromosomas. Por ejemplo, se ha descrito un gradiente del grupo Diego (Di), que va desde norte a sur, siendo más frecuente en atacameños, menos frecuente en mapuches y ausente en fueguinos (Etcheverry et al., 1967), resultados confirmados posteriormente por Matson et al., (1969). Otras diferencias en marcadores hemato-sero-antropológicos son descritas en un artículo de revisión de Etcheverry (1997). Los estudios de DNA mitocondrial muestran también un gradiente latitudinal norte sur tanto para poblaciones originarias como para la población urbana, con una alta frecuencia del haplogrupo B en el norte y una mayor frecuencia de los haplogrupos C y D en el sur (Moraga et al. 2000, Rocco et al. 2002, Garcia et al 2006). El DNA mitocondrial de herencia uniparental materna muestra además la asimetría en los apareamientos durante los primeros cien años de conquista, ya que casi la totalidad de los linajes maternos en poblaciones originarias son de origen amerindio, y sobre el 80% lo son en población urbana (Rocco et al, 2002).

El conocimiento histórico nos indica que la miscegenación mayor ha ocurrido con el grupo Picunche, hoy desaparecido, que dio origen a la población más numerosa de la zona central de Chile, incluyendo Santiago, Valparaíso y Concepción. La otra miscegenación importante fue en el Norte Chico con la población Diaguita (extinta), en el Norte grande con los Aymaras y en la Región de los Lagos y Chiloé con el grupo Huilliche y en mucho menor grado con los Chonos (extintos). La población Mapuche no se incorporó al acervo chileno, salvo en grupos aislados, durante la Conquista y la Colonia, pero empezó a asimilarse a la población chilena por migración a las grandes ciudades desde la Independencia y posteriormente después del proceso llamado Pacificación de la Araucanía, en el que se realizó también una incorporación mayor de la población Huilliche (Thayer Ojeda, 1919).

La población caucásica que llegó a Chile también es heterogénea. Es así como de la población hispana que arribó durante la Conquista, pueden diferenciarse castellanos, extremeños, andaluces, vascos, catalanes, gallegos y otros grupos (Thayer Ojeda, 1919). Además debemos agregar la llegada, desde la Conquista hasta nuestros días, de italianos, franceses, ingleses, alemanes y holandeses que arribaron, en especial, durante o después de los grandes conflictos bélicos, políticos o hambrunas en Europa (Valenzuela and Harb, 1977). Estudiar la traza genómica de estos grupos es un objetivo importante del Proyecto aunque de interés posterior en el tiempo.

Cabe destacar que, además de las diferencias antes mencionadas entre los grupos amerindios y entre los caucásicos llegados a Chile, se debe sumar un fenómeno de miscegenación no homogénea según los estratos socioeconómicos. Los españoles o europeos se reservaron los puestos y cargos de mayor importancia y dejaron los trabajos manuales y cargos menos relevantes para los mestizos o los indios. Pero aún los mestizos se estratificaron según sus proporciones de mezclas y así los que tenían más ancestros europeos podían asimilarse mejor a los estratos altos y los que tenían más ancestros amerindios se continuaban relacionando con su grupo o con grupos más amerindios. Así se constituyó una sociedad que en el siglo pasado fue tipificada en la población de Santiago como un gradiente sociogenético en el que el estrato

socioeconómico alto (5% de la población) no tenía componente genético nuclear amerindio, el estrato medio (20% de la población) con 20% de componente amerindio y un estrato bajo (75% de la población) con 30-40% de mezcla amerindia (Thayer Ojeda, 1919; Valenzuela, 1984; Valenzuela, 1998).

Una condición evolutiva muy importante a considerar desde la ciencia misma, que también tiene importancia de aplicación médica práctica es el componente de selección natural que ha operado desde los inicios de la miscegenación. Las poblaciones Caucásicas y Amerindias separadas por más de 40.000 años de historias evolutivas de adaptación a sus respectivos ambientes, especialmente a aquellos que pueden originar patologías, pero también a sus condiciones de producción endógena de patología (enfermedades hereditarias), se encuentran en este ambiente chileno con determinados agentes patógenos externos al que ofrecen tres grupos genómicos para adaptarse: el caucásico, el amerindio y el mixto.

Esta sucinta historia nos muestra las características de un laboratorio evolutivo natural de mezclas étnicas que habían estado separadas por más de 40.000 años y que juntándose evolucionan por unas 16 generaciones. Es evidente que los grupos caucásicos y amerindios diferían en historia evolutiva, acervo genético y la estructura genómica al momento de la miscegenación. Sin embargo, no tenemos un conocimiento fino de estas diferencias, en los grupos que constituyeron la población chilena, y tampoco conocemos la evolución genético-genómica que se ha dado para constituir la población actual. El objetivo principal de este proyecto es aproximarse a establecer esa estructura genómica y describir la subestructura poblacional.

### **Estudios actuales con paneles de marcadores genéticos informativos de ancestría en otras poblaciones humanas**

Desde el año 2002 a la fecha, se han publicado al menos 20 artículos que intentan describir la estructura genética de poblaciones tanto amerindias como de poblaciones mixtas de Latinoamérica. Estas aproximaciones utilizan diferentes paneles de marcadores genéticos informativos de ancestría (AIMs), en cuya selección se han utilizado muchas metodologías y criterios.

Existe una serie de iniciativas alrededor del mundo, similares a la que desarrollada en este proyecto. Si nos centramos en países en vías de desarrollo como el nuestro, podemos encontrar ejemplos como el de México, que desde el año 2004 dio inicio al análisis de entre 500 mil a 600 mil SNPs en una muestra de población mestiza de 1200 individuos, provenientes de diferentes lugares geográficos de México. Para ello el gobierno mexicano estableció el Instituto Nacional de Medicina Genómica (INMEGEN) (Séguin et al., 2008). Podemos citar también el caso de la India con su *Indian Genome Variation consortium* (IGV) que cuenta con una base de datos de más de 15000 individuos de varias subpoblaciones de la India. El IGV ha detectado, por ejemplo, un set de 405 SNPs de susceptibilidad para 75 genes asociados a enfermedades (Hardy et al., 2008a). En el caso de África, podemos citar el trabajo colaborativo entre instituciones públicas y privadas. Por ejemplo, se están analizando los genotipos de más de un millón de SNPs en cinco poblaciones nativas sudafricanas para establecer su relación con una serie de enfermedades complejas (Hardy et al., 2008b). Por su parte, en Asia podemos destacar una iniciativa que involucra a varios países del sudeste y del este asiático. En este estudio se analizan 54794 SNPs autosómicos en 1928 individuos que

representan a 73 poblaciones asiáticas. Los investigadores concluyen que existe una correlación entre la ancestría genética, las afiliaciones lingüísticas y geográficas. Además, fue posible demostrar que la principal fuente de las poblaciones del este asiático fueron las del sudeste asiático (HUGO Pan-Asian SNP Consortium, 2009). Otros ejemplos han sido mencionados en un trabajo de revisión de Hardy et al. (2008c), en el cual se destacan iniciativas como las de China, Kuwait, Iran y de países del África sub-Sahara (Nigeria, Kenya, Tanzania, Zinbawe y Sudáfrica).

Todos los proyectos similares al presente, tanto en países desarrollados como en países en vías de desarrollo tienen objetivos en común. Podemos tomar como ejemplo el caso de México, teniendo en cuenta ciertas similitudes en el origen de su población y de la población chilena. Con la creación del INMEGEN, el gobierno mexicano da una fuerte señal respecto a que la medicina genómica es realmente una prioridad. En este sentido, las variaciones genómicas propias de cada subpoblación deben tener implicancias tanto en el diagnóstico de enfermedades como en su tratamiento. Los resultados de los estudios de este instituto pueden ser utilizados para detectar subpoblaciones con riesgos genotípicos incrementados de presentar ciertas enfermedades, como las patologías cardiovasculares, diabetes y cáncer. En estas subpoblaciones de mayor riesgo se pueden realizar campañas de prevención y de promoción de salud. Se proyecta, por ejemplo, que en el caso de la diabetes, la contribución del INMEGEN podría reducir los costos relativos a su manejo en un 36% entre 2010 y 2025. Desde el punto de vista de la farmacogenómica y vacunogenómica, se pueden proyectar programas de vigilancia respecto a grupos de individuos que no respondan a una terapia o a una vacuna determinadas. En síntesis, las iniciativas como la emprendida por México pueden tener grandes proyecciones como: la generación de alianzas estratégicas a nivel nacional o regional para promover la investigación y desarrollo en medicina genómica, aplicando las tecnologías desarrolladas a los problemas comunes de salud, estableciendo programas de formación de recursos humanos en el área de la medicina genómica y transfiriendo el conocimiento genómico a productos y servicios (Hardy et al., 2008c; Séguin et al., 2008).

## **Situación en Chile**

En nuestro país a la fecha del inicio de nuestro proyecto no se había llevado a cabo ninguna iniciativa que intentara describir la estructura genética de la población chilena contemporánea en base a marcadores genéticos distribuidos a lo largo de todo el genoma. Sin embargo, en la literatura se podían encontrar una serie de estudios con este objetivo tanto en subpoblaciones de nuestra población mixta como en poblaciones amerindias chilenas en base a un número bajo de marcadores circunscritos a ciertos genes (o sus productos) o a sistemas génicos.

Entre los estudios chilenos, una tendencia ha sido el uso de marcadores del sistema HLA, grupos sanguíneos y/o proteínas séricas para describir la estructura genética tanto de poblaciones originarias como de poblaciones mestizas. Entre estos podemos citar estudios como los de Allison et al., (1978) en momias chilenas y peruanas, Haas et al., (1985) en mapuches argentinos y chilenos, Goede et al., (1985) en indios atacameños, Llop et al., (1993) en pehuenches, Cifuentes et al., (1995) en relación al componente negroide de la población de Santiago, Llop (1996) en cinco poblaciones amerindias, Harb et al., (1998) en cuatro localidades costeras, Llop et al (2006), en dos poblaciones originarias, Acuña et al., (2000), en tres localidades rurales



del norte chico y Droguett et al., (2008) en donantes y receptores de transplantes renales.

También es posible encontrar otra serie de estudios, que si bien no pretenden directamente describir la estructura genética de la población chilena, describen las frecuencias alélicas para una serie de marcadores, en especial STRs, en la población general en relación a estudios de paternidad y estudios forenses. Es así como en este grupo encontramos estudios como los de Aguirre et al. (1992), Armanet et al. (1995), Cifuentes et al. (1996), Jorquera et al. (1999) y Jorquera et al. (2008). Otro grupo de publicaciones también describen frecuencias de marcadores moleculares de herencia uniparental ya sea de la línea materna (DNA mitocondrial) o de la paterna (cromosoma Y) tanto en población urbana y rural como en poblaciones amerindias chilenas. En este grupo podemos mencionar los estudios de Merriwether et al. (1995) en población aymará, Cifuentes et al. (2004) en población mixta, Henríquez et al. (2004) y Rothhammer et al 2010 en changos, García et al, 2006 en población chilota rural, Moraga et al. (2005) en poblaciones amerindias prehispánicas del norte de Chile y Moraga et al (2010) en población Kawesqar actual y antigua.

En síntesis, en Chile, y en especial en nuestra Unidad Académica de la Universidad de Chile, se había ido acumulando un valioso conocimiento sobre la estructura genética de la población mixta chilena, obtenido a través de diferentes tipos de estudios y desde distintos enfoques metodológicos. Aún valorando el significativo aporte del conocimiento alcanzado, se hacía absolutamente necesaria una iniciativa como la de este proyecto, que involucre el estudio de múltiples marcadores genéticos distribuidos a lo largo de todo el genoma y que contemple el estudio de varias subpoblaciones para conocer la composición ancestral de nuestra población contemporánea.

En el último año, durante el curso del presente proyecto se publicó el primer trabajo que describe la ancestría en una población mixta chilena en base a un conjunto de AIMs; se trata de una publicación de Fuentes et al. 2014 en 923 chilenos residentes en la ciudad de Arica y genotificados con 40 marcadores que dan una ancestría promedio de 44.3, 51.9 y 3.8 % para los componentes Amerindio, Europeo y Africano respectivamente. Hace unos meses se agrega un trabajo en 313 chilenos que comprenden sujetos enrolados en un estudio de infección por virus Hanta y por Síndrome de microdelecion 22q11, los que se genotificaron con un array de 680.000 SNPs (Eyheramendy, 2015) encontrándose porcentajes de ancestría similares a los descritos por Fuentes. Estos han sido los primeros esfuerzos por describir la ancestría continental de algunos chilenos mestizos con herramientas modernas de la era genómica.

El presente proyecto buscó caracterizar genéticamente diversas subpoblaciones mixtas chilenas, en cuanto a su ancestría y en base a la información obtenida diseñar un panel acotado de AIMs informativos para población chilena. Adicionalmente, con el fin de divulgar los resultados del proyecto se implementaron sitios web de divulgación de la información recogida, para hacerla disponible tanto a los científicos usuarios de estos datos como la comunidad en general.

# METODOLOGÍAS DE INVESTIGACIÓN Y DESARROLLO

## 1- Muestreo

Se hizo un muestreo de 3216 chilenos no emparentados residentes en las ciudades de Arica, Iquique, La Serena, Coquimbo, Santiago, Chillán, Temuco y Puerto Montt. La mayor parte del muestreo se realizó en dadores de banco de sangre, con excepción de las ciudades de Arica e Iquique en que éste se realizó en Centros de Salud Familiar y para el caso de Puerto Montt en Consultorios y diversos sitios públicos. Se incluyeron además 39 tríos Padre Madre e Hijo-a en el estudio, contactados directamente por investigadores del proyecto.

En la ciudad de Santiago el muestreo de dadores de sangre se realizó en la casa del donante Santiago Norte del Hospital San José, en colectas a terreno realizadas por el Banco Metropolitano de dadores (en Universidades de Chile y privadas y Plaza de Puente Alto) y en Clínica Santa María.

Criterios de Inclusión: ser chileno mayor de edad y tener ambos padres nacidos en Chile.

Criterios de exclusión: ser extranjero o hijo de extranjeros

Se invitó a participar en el estudio a los individuos que cumplieran con los criterios señalados, después de explicarles los objetivos del proyecto y se les solicitó la firma de un documento de consentimiento informado. La participación consistió en la donación de 3 ml de sangre de aquella que habían donado al Centro de Dadores o 3 ml de saliva depositados en tubo Oragene (m.r.) para recolección de saliva (en el caso de los tríos y los habitantes de Puerto Montt) además de responder a un cuestionario acerca de antecedentes demográficos y sociales (ver “ficha de caracterización” en anexo 1), especialmente diseñado por Investigadora Socióloga de este Proyecto.

Los protocolos así como los documentos de consentimiento informados fueron revisados y aprobados por los comités de ética asistenciales correspondientes además del comité de ética de Fondecyt y de la Facultad de Medicina de la Universidad de Chile.

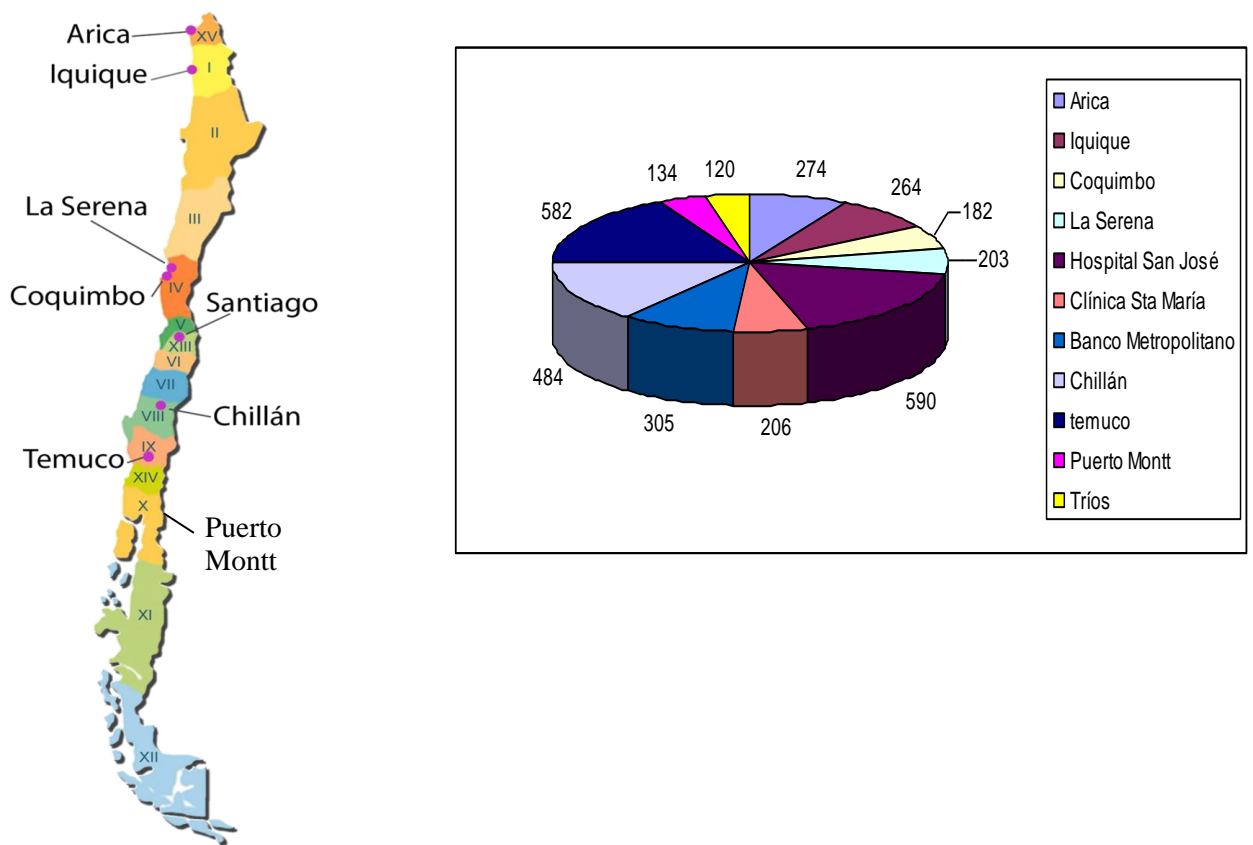
El número de individuos entrevistados en cada punto de muestreo de puede apreciar en la Figura 1.

## 2. Extracción de DNA a partir de muestras de sangre y saliva

El DNA fue extraído mediante el método de Lahiri y cols. Se evaluó la calidad del DNA extraído mediante evaluación visual de geles de agarosa teñidos con bromuro de etidio para determinar su nivel degradación y espectrofotometría UV. Todas las muestras seleccionadas para genotipificación tenían un score de “Good” o “Borderline” según este sistema de calificación. La cuantificación de DNA se hizo con el sistema PicoGreen.

## 3. Genotipificación de las muestras y análisis bioinformático

La genotipificación de las muestras de DNA se realizó en tres fases sucesivas de mayor a menor profundidad de análisis genómico y de menor a mayor número de individuos:



**Figura 1:** Número de chilenos reclutados en cada Punto de Muestreo

i. Primera fase de Genotipificación mediante secuenciación genómica:

Se seleccionaron 18 participantes del proyecto ChileGenómico por poseer ancestría amerindia Aymara o Mapuche. Se eligieron porque cumplían con al menos 2 de los siguientes criterios: (1) tener al menos 1 apellido originario, (2) poseer haplotipo de origen amerindio en el cromosoma Y (hombres) y (3) poseer haplotipo de origen amerindio en el DNA mitocondrial. Se seleccionaron 18 muestras que pasaron el control de calidad, 9 con ancestría Aymara y 9 con ancestría Mapuche para secuenciación de genoma completo a baja cobertura para la identificación de variantes propias de estas poblaciones. La secuenciación se realizó con la plataforma SOLID 5500 (Applied Biosystems) en OMICS Solutions en Santiago de Chile (<http://omics-solutions.cl/>). La secuenciación se realizó en 4 corridas entre enero y marzo de 2013. En la primera corrida, se utilizaron adaptadores con secuencias únicas (“código de barras” o más comúnmente llamado *barcoding*) para cada muestra, y así poder separar muestras secuenciadas en un mismo carril del *flow cell* de secuenciación. En las siguientes corridas, cada muestra fue secuenciada en un carril independiente y por lo tanto no se necesitó usar *barcoding*. Todas las secuencias fueron producidas con librerías pareadas con un tamaño de inserto de 250 pares de bases y largos de 75, 35 pares de bases respectivamente.

El control de calidad de las secuencias producidas, mapeo de lecturas de secuencia en el genoma de referencia y la determinación o “llamado” de variantes se realizó mediante un flujo de trabajo o *pipeline* implementado por el Laboratorio de Bioinformática y Matemática del Genoma (LBMG) del Centro de Modelamiento Matemático (CMM) de la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile. El mapeo de lecturas se realizó con el software de LIFESCOPE versión 2.5.1 (Applied Biosystems) utilizando la versión GRCh37 del genoma de referencia, la misma utilizada por el proyecto 1000 genomas ([www.1000genomes.org/](http://www.1000genomes.org/)). El llamado de variantes se realizó con SAMtools v0.1.19 ([samtools.sourceforge.net/](http://samtools.sourceforge.net/)) y GATK v2.7-4 ([www.broadinstitute.org/gatk/](http://www.broadinstitute.org/gatk/)). Solo lecturas con una calidad de mapeo mayor o igual a 20 (99% certeza) fueron utilizadas para el llamado de variantes. Dieciséis de las 18 muestras secuenciadas fueron también genotipadas con el microarray AXIOM LAT1 (800.000 SNPs) en la segunda fase de lo que permitió comprobar una replicabilidad de 95% de los genotipos obtenidos.

Los puntajes de calidad de variantes fueron re-calibrados mediante *machine learning* utilizando variantes de HapMap, 1000 Genomas y dbSNP v137 como “gold standard”. Las variantes no previamente conocidas en regiones con cobertura > 150x o presentes en repeticiones en serie fueron eliminadas para disminuir la tasa de falsos positivos generadas por lecturas de distintas regiones cromosómicas que fueras mapeadas equivocadamente a una misma región.

ii. Segunda fase de Genotipificación de alta densidad mediante Plataforma AXIOM LAT1 (817.000 SNPs):

A partir de la información genómica obtenida en la primera fase de secuenciación de 18 chilenos se eligió el panel AXIOM LAT1 de Affimetrix para esta segunda fase, ya que el conjunto de marcadores que este panel incluye aportaba la máxima la informatividad de ancestría a lo largo de todo el genoma en comparación con las otras plataformas comerciales disponibles, para población chilena. Para realizar esta elección se utilizó un algoritmo genético desarrollado en el LBMG que maximiza concomitantemente informatividad ( $F_{st}$ ) y cobertura del genoma.

Para esta fase de genotipificación se eligieron 358 chilenos no emparentados de las ciudades de Arica, Iquique, La Serena, Coquimbo, Santiago, Chillán y Temuco y 39 tríos Padre- Madre- Hijo-a.

Los 358 individuos no emparentados fueron aleatoriamente seleccionados del total de muestras de proyecto divididas en seis grupos de acuerdo a lugar de reclutamiento (Tabla 1):

Población	Reclutados	DNA <sup>1</sup>	Seleccionados	Tasa de selección <sup>2</sup> (%)
Chillan	492	450	60	13.3
Norte Chico	406	173	59	34.1
Norte Grande	520	102	59	57.8
SGOpriv <sup>3</sup>	527	281	60	21.4
SGOpub <sup>4</sup>	550	529	60	11.3
Temuco	583	457	60	13.1
Total	3078	1992	358	18.0

<sup>1</sup>Calidad óptima para microarrays a partir de geles de agarosa y espectrofotometría UV

<sup>2</sup>Seleccionados/DNA x 100

<sup>3</sup>Santiago, sistema privado de salud o educación

<sup>4</sup>Santiago, sistema público de salud o educación

**Tabla 1:** Individuos no emparentados, seleccionados para la etapa 2 de genotipificación.

Las muestras se tipificaron en el Institute for Human Genetics, University of California, San Francisco, mediante el uso de un microarray de alta densidad: AXIOM LAT1 (World Array 4, Affymetrix) y sus resultados se analizaron en el Laboratorio del Dr. Carlos Bustamante en la Escuela de Medicina de la Universidad de Stanford, en visita del Investigador Ricardo Verdugo.

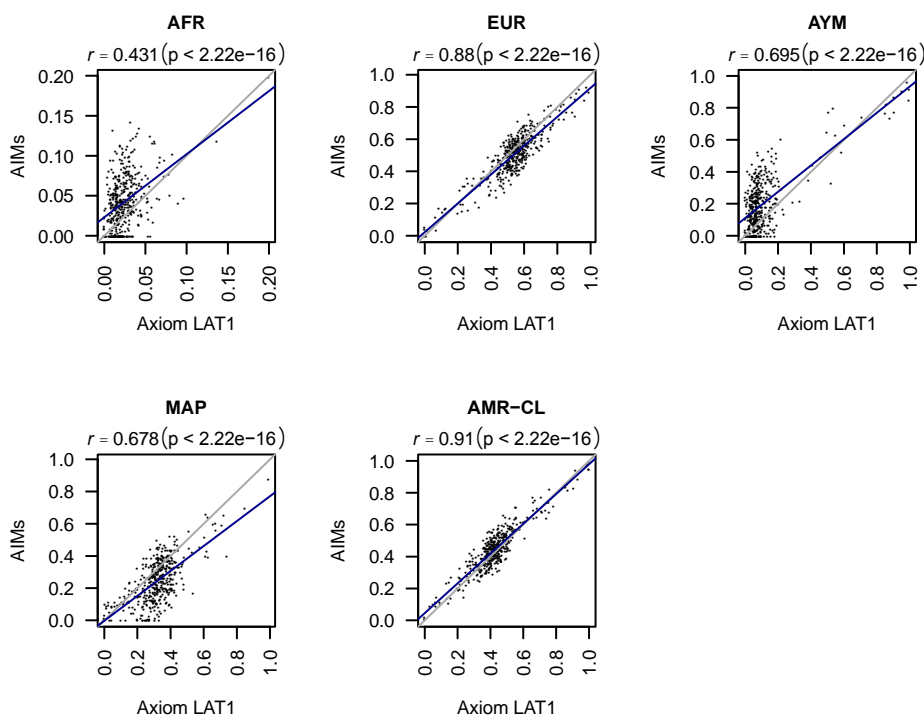
El envío de muestras fue realizado en hielo seco con placas de 96 pocillos Bio-Rad selladas con láminas aluminicas adhesivas de acuerdo al protocolo solicitado por UCSF.

- iii. Tercera fase de Genotipificación de 150 SNPs informativos de ancestría en los 2900 chilenos restantes:

Se diseñó un panel de acotado 150 SNPs compuesto por 143 SNPs conocidos (de la plataforma AXIOM LAT1) y 7 SNPs nuevos, descubiertos por secuenciación masiva en la 1<sup>o</sup> etapa del proyecto, cuyo objetivo es determinar con precisión los porcentajes del genoma de chilenos que tienen origen en cuatro posibles ancestrías: europea, africana, amerindia andina, amerindia del sur de Chile. Los cromosomas X, Y, y MT fueron excluidos del panel dado que esto se estudian con otras aproximaciones y buscamos maximizar la cobertura del los autosomas. Como referencia, se usaron datos de 30 individuos europeos, 30 individuos africanos del proyecto 1000 Genomas, 17 chilenos con ancestría Aymara y 31 chilenos con ancestría Mapuche generados por el proyecto ChileGenómico con el microarray AXIOM LAT1.

Para determinar informatividad sobre ancestría, se buscó un set de marcadores que maximizaran la diferenciación entre las muestras chilenas y los referentes europeos y africanos, entre europeos, amerindios y africanos y entre Aymara y Mapuche. La diferenciación genética se estimó mediante el estadístico  $I_n$  (Rosenberg et al. 2003). Se utilizó un algoritmo genético desarrollado por el equipo de ingenieros del proyecto (CMM, U. de Chile) que maximiza simultáneamente el  $I_n$  de todas las comparaciones así como la cobertura del

genoma para un tamaño de panel dado, en este caso 150. Se realizaron numerosos ensayos variando cada parámetro del algoritmo y se seleccionó el que maximizó la correlación entre las estimaciones de ancestría generadas y aquellas obtenidas mediante set completo de SNPs presentes en el microarray (Figura 2). Los SNPs nuevos fueron seleccionados con el objetivo de agregar informatividad al panel y maximizando la confianza en que representen verdaderas variantes y no errores de secuenciación. El set inicial de SNPs nuevos predichos por GATK fue de 260367 SNPs con genotipos imputados usando SNPtools v1.0 (Wang et al. 2013). Solo consideramos SNPs con un puntaje de calidad >20 (equivalente a una probabilidad de error < 1%) y cuyo alelo alternativo (distinto a la referencia) haya sido observado en al menos 20 de los 36 cromosomas secuenciados. Los genotipos que a la imputación recibieran una probabilidad menor al 50 por ciento no fueron considerados. Los genotipos en regiones genómicas que fueran etiquetados como de origen europeo o africano por PCAadmix (Brisbin et al. 2012) fueron reemplazados por datos perdidos antes del análisis. Esto generó una lista de 59 SNPs candidatos. Los datos de genotipificación para estos SNPs fueron inspeccionados visualmente en IGV v2.3.10 (Robinson et al. 2011) para eliminar aquellos flanqueados por otros SNP que pudieran interferir en el diseño del ensayo de genotipificación o que representaran genotipos idénticos con SNPs cercanos. Esto resultó en 7 SNPs confiables que fueron incluidos en el panel de AIMS. De estos, 3 fueron observados sólo en muestras con ancestría Aymara, uno tuvo baja frecuencia en Mapuches e intermedia en Aymara y tres tuvieron alta frecuencia en ambas poblaciones (>0.8).



**Figura 2.** Correlación en valores de ancestría estimados a partir del panel acotado de AIMS y el microarrays AXIOM LAT1 con 600.000 SNPs. AMR-CL denota ancestría amerindia chilena (AYM+MAP).

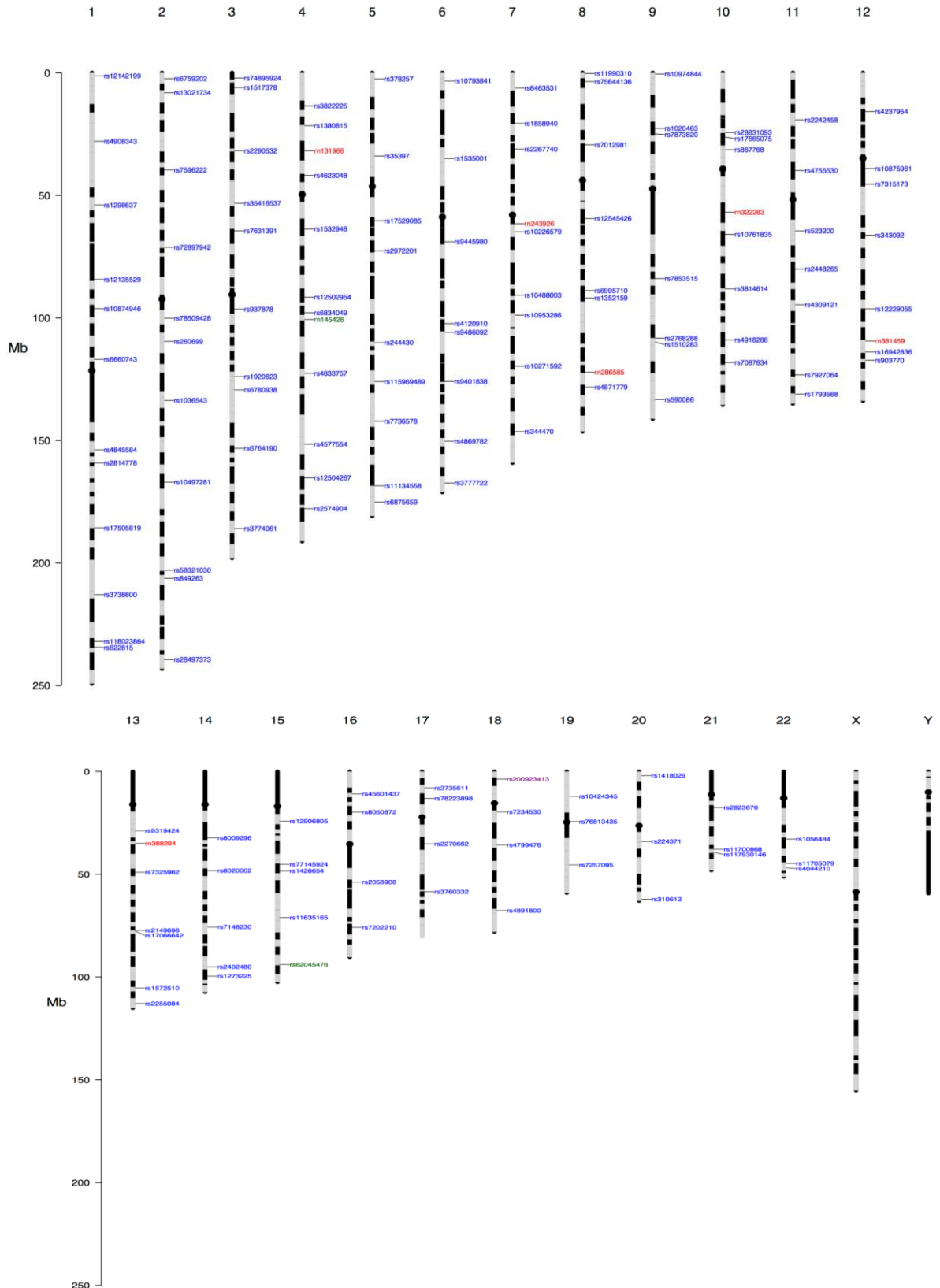
El panel final de marcadores se muestra en la Figura 3 y en el anexo 2

#### **4. Inferencias de Ancestría**

Las estimaciones de ancestría se hicieron mediante comparación de las frecuencias alélicas de los SNPs obtenidas en ChileGenómico con aquellas de poblaciones de referencia a través del programa Admixture (<https://www.genetics.ucla.edu/software/admixture/>).

#### **5. Comparaciones estadísticas**

Las comparaciones de los porcentajes de ancestría entre subpoblaciones chilenas se hicieron con Análisis de Varianza o prueba t de Student en el caso de solo 2 grupos y para el ajuste por variables confundentes se utilizó análisis de regresión múltiple. Las correlaciones entre estimaciones de ancestría se evaluaron con coeficiente de correlación de Pearson.



**Figura 3:** Ubicación cromosómica de los 150 marcadores informativos de ascendencia elegidos (azul). En rojo los marcadores nuevos de ChileGenomico, en verde y púrpura los que fallaron en la genotipificación. Las posiciones se muestran en megabases (Mb).



## **6. Divulgación de resultados**

Para la divulgación de resultados globales se construyó un sitio web de difusión para el público general que alberga las principales estimaciones de ancestría por región muestreada, además de entrevistas y noticias del proyecto.

Para divulgación de resultados de frecuencias alélicas y variantes estudiadas en chilenos, se construyó una base de datos que además alberga información de otras bases de datos públicas de referencia.

# **RESULTADOS**

## **1. Caracterización socioantropológica de subpoblaciones chilenas:**

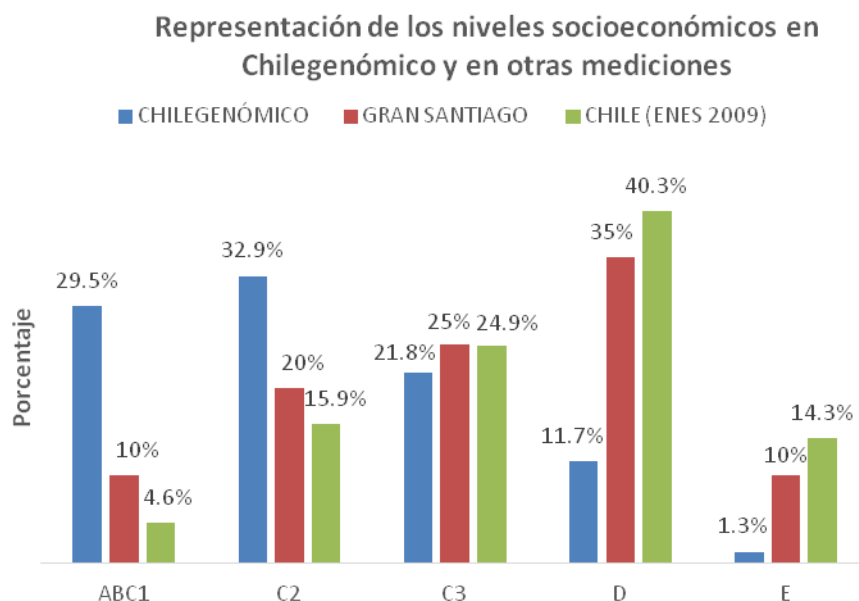
Las fichas de caracterización de los 3300 chilenos muestreados (46.8 % varones) fueron analizadas por Investigadora del Proyecto (Socióloga), previa digitación en duplicado. En la región de residencia y de nacimiento de los chilenos incluidos en el estudio están representadas todas las regiones del país, mayoritariamente las regiones Metropolitana, del Bío Bío y de la Araucanía (ver tabla 2).

El análisis de las fichas de los sujetos, reveló que el 17,5 % de los encuestados refirió tener algún antepasado originario y el 17.9 % se declaró perteneciente a alguno de nuestros pueblos originarios. Por otro lado el 24,0 % de los encuestados reportaba educación universitaria completa del principal sostenedor del hogar y un 0.2 % sin estudios.

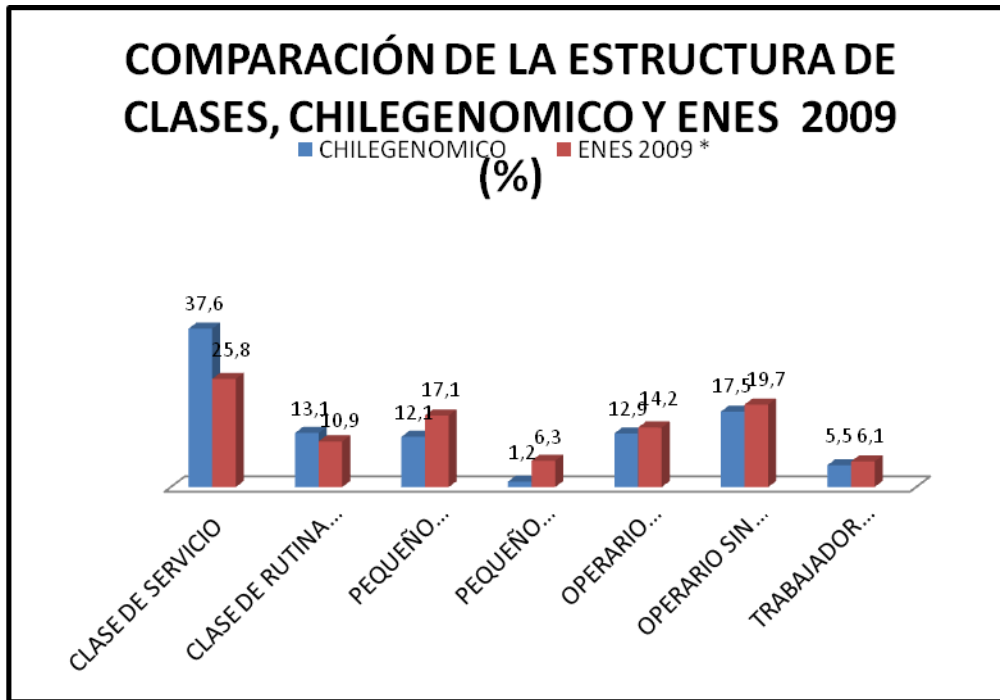
Para la clasificación social, se decidió usar el nivel socioeconómico (NSE) que mide la capacidad de consumo de los hogares, y también la clase social que determina la posición socio-ocupacional de las personas. El análisis de estas variables reveló que en la muestra de Chile Genómico están representados todos los niveles socioeconómicos y clases sociales, sin embargo están más representados los NSE ABC1 y C2 y menos los niveles D y E comparativamente con el Gran Santiago y con la encuesta ENES 2009 (que incluye 6000 chilenos representativos de todo el país) (ver figura 4). La representación de los distintos niveles socioeconómicos difiere entre los distintos puntos de muestreo, encontrándose una mayor proporción de individuos del grupo ABC1 en Clínica Santa María y Banco Metropolitano de tejidos y más individuos del estrato E en Hospital San José de Santiago. Respecto de la clase social, en ChileGenómico hay ligeras diferencias respecto de lo registrado en ENES 2009 (ver Figura 5) algo más de sujetos en la “clase de servicios” y un poco menos en “pequeño empresario” y “pequeño propietario agrícola”).

Lugar de Residencia - Región		
	Frecuencia	porcentaje
Región de Tarapacá	266	7.9
Región de Antofagasta	5	.1
Región de Atacama	10	.3
Región de Coquimbo	414	12.3
Región de Valparaíso	14	.4
Región del Libertador Bernardo O'Higgins	41	1.2
Región del Maule	14	.4
Región del Biobío	482	14.3
Región de la Araucanía	568	16.9
Región de los Lagos	139	4.1
Región de Aysén	4	.1
Región de Magallanes	3	.1
Región Metropolitana	1097	32.6
Región de los Ríos	11	.3
Región de Arica y Parinacota	289	8.6
no consignado	77	.4
Total	3370	100.0

**Tabla 2.** Distribución de sujetos estudiados según Región de Residencia



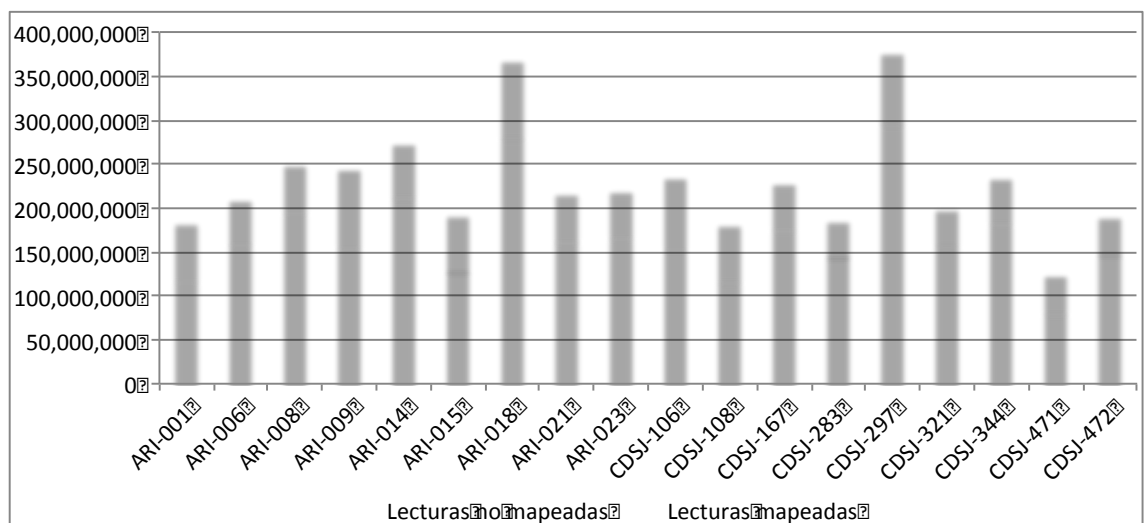
**Figura 4.** Estratos socioeconómicos representados en la muestra de 3100 chilenos estudiados en el marco del Proyecto Chilegenómico. El tamaño muestral de la Encuesta Nacional de Estratificación Social (ENES) 2009 corresponden a 3336 y 3295, respectivamente. Por su parte, la clasificación de los sujetos en el Gran Santiago en distintos grupos socioeconómicos - ABC1, C2, C3, D y E - está construida en base a cortes fijos que aseguran que los porcentajes de cada grupo correspondan a 10%, 20%, 25%, 35% y 10%, respectivamente.



**Figura 5.** Distribución de chilenos muestreados en Chilegenómico según clase social

## 2. Secuenciación de genomas chilenos:

El secuenciamiento de los 18 individuos, produjo un total de 2,000 millones de fragmentos representando un promedio de 114 millones de fragmentos o 228 millones de lecturas por individuo. Un 80% de las lecturas *forward* (75pb) y un 68% de las lecturas *reverse* (35pb) pudieron ser mapeadas en el genoma de referencia (Figura 6). La calidad promedio de alineamiento fue 34, que significa que el 99.9 % de las lecturas corresponden en la posición asignada del genoma de referencia. Los 18 genomas tuvieron una cobertura 3.14x, es decir que cada base fue secuenciada 3.14 veces en promedio (2x min, 5x max).



**Figura 6.** Número de lecturas de secuencia (*reads*) por individuo. Los códigos de muestra que comienzan con ARI identifican muestras con ancestría Aymara y aquellos con CDSJ identifican individuos con ancestría Mapuche.

### *i. Descubrimiento de nuevas variantes*

Se encontraron 4.064.554 millones de polimorfismos de 1 nucleótido (SNPs) en los 18 individuos secuenciados por GATK y 3,664,458 por SAMtools. Un total de 3.297.755 de variantes fueron detectadas por ambos programas. Se decidió utilizar las variantes determinadas por GATK dado que GATK está optimizado para buscar variantes en genomas Humanos y porque la razón entre transiciones y transversiones predichas por GATK (2.1) se acercó más a la observada en otras poblaciones, que las predichas por SAMtools, lo que indicaría que las predicciones realizadas con GATK tendrían menos falsos positivos que las de SAMtools.

Entre los 4 millones de SNPs detectados, 163.575 (4%) no han sido previamente descritos en ninguna población interrogada en el proyecto 1000 Genomas. A su vez, de las 307.885 inserciones y deleciones (INDELs) detectadas, 113,029 (37%) no han sido descritas y pueden representar variaciones particulares de la población Chilena o de poblaciones sudamericanas. Sin embargo, este hallazgo solo podrá ser confirmado con nuevos experimentos de secuenciación tanto en Chile como en otros países de la región.

### *ii. Análisis de Ancestría Global*

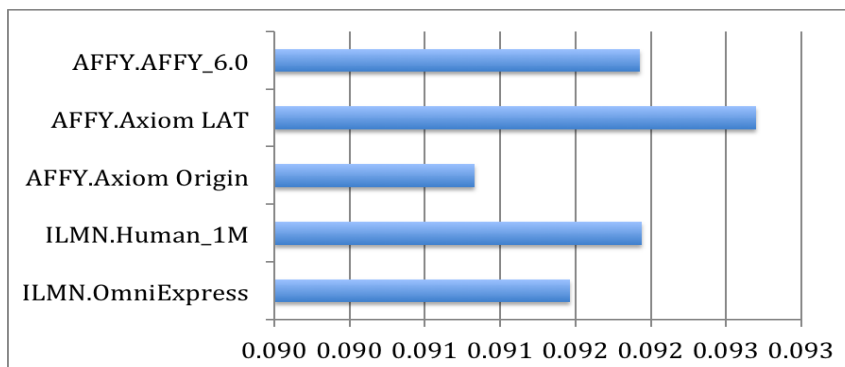
Mediante comparación de las frecuencias alélicas para los SNPs conocidos con poblaciones de referencia del proyecto ChileGenómico se pudo determinar los componentes de ancestría de los 18 individuos de genomas secuenciados. La inferencia de ancestría global fue realizada con el programa Admixture (<https://www.genetics.ucla.edu/software/admixture/>). Los individuos secuenciados mostraron un 73% de ancestría amerindia y 4% africana en promedio. Los resultados se resumen en la Tabla 3.

Punto de Muestreo	EU	AF	AM
ARI	19.6	4.2	76.2
CDSJ	25.9	3.5	70.6
Promedio	22.7	3.9	73.4

**Tabla 3.** Porcentaje de ancestría por continente

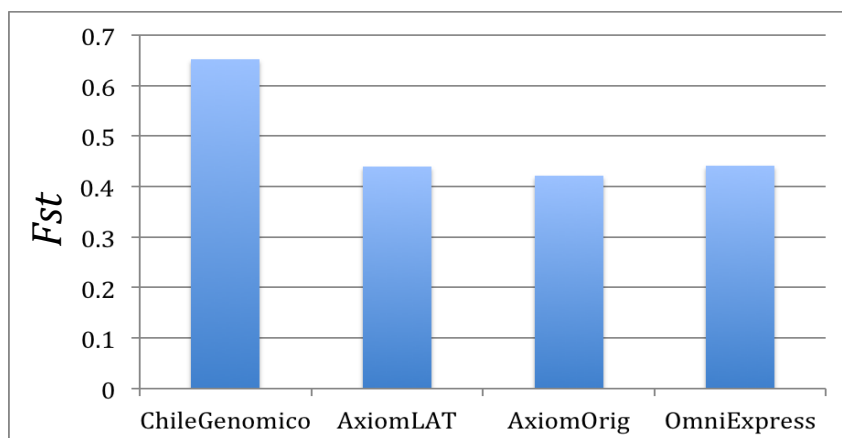
### *iii. Selección de un panel de AIMS para la población chilena*

El análisis de ancestría permite determinar qué SNPs contienen mayor información sobre el origen ancestral de cada región cromosómica. Estos marcadores se denominan *Ancestry Informative Markers* (AIMs). La informatividad de los AIMs fue cuantificada utilizando el Índice de Fijación ( $F_{st}$ ), que toma valores continuos entre 0 y 1 que indican el grado de informatividad. Con el objetivo de evaluar la mejor plataforma de genotipificación para individuos chilenos, calculamos el  $F_{st}$  promedio de los SNPs presentes en microarreglos comerciales de genotipificación disponibles en el mercado. Este análisis evidenció la plataforma AXIOM LAT1 (Affymetrix) como la más informativa Figura 7.



**Figura 7.** Informatividad de ancestría medida por  $F_{st}$  promedio de los SNPs incluidos en microarrays de genotipificación comerciales.

La informatividad de todos los polimorfismos identificados en los 18 individuos chilenos fue usada para elegir un panel de marcadores que maximicen la informatividad de ancestría a lo largo de todo el genoma. Se utilizó un algoritmo genético desarrollado en el LBMG que maximiza concomitantemente informatividad ( $F_{st}$ ) y cobertura del genoma. Para poder evaluar el valor de este panel, el mismo algoritmo fue también alimentado solo con marcadores presentes en microarreglos comerciales que resultaron interesantes a nivel global. Este análisis evidenció un aumento en informatividad de al menos 47% respecto de lo obtenible de paneles de igual tamaño en paneles comerciales (Figura 8).



**Figura 8.** Informatividad promedio de paneles de 2400 AIMs seleccionados para maximizar  $F_{st}$  y cobertura del genoma.

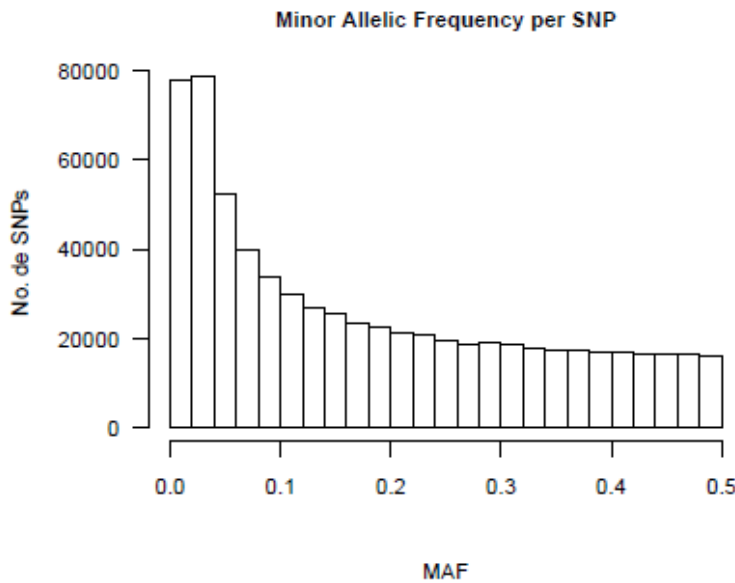
En virtud de estos análisis se decidió utilizar la plataforma AXIOM LAT1 de Affymetrix para la segunda fase de genotipificación.

### 3. Genotipificación de 475 chilenos con plataforma AXIOM LAT1 (World Array 4, Affymetrix):

Los 475 individuos seleccionados se genotipificaron con chip AXIOM LAT1 de Affymetrix que cuenta con un total de 817,810 marcadores SNP y ha sido utilizado en múltiples estudios poblacionales tanto para el estudio de la estructura poblacional, mezcla y ancestría. Se utilizaron los parámetros recomendados por Affymetrix para evaluar la calidad de los datos de microarray. Se requirió como mínimo una tasa de

genotificación (*call rate*)  $\geq 95\%$ , un valor de Fisher's Linear Discriminant (*FLD*)  $> 3.6$  y una diferencia de señal entre heterocigotos y el promedio de homocigotos menor a 0.5 (*HetSO*  $> -0.5$ ). De los las 358 muestras de individuos no-emparentados, 346 (96.7%) pasaron el control de calidad. De los 117 emparentados, 116 (99,1%) tuvieron buena calidad.

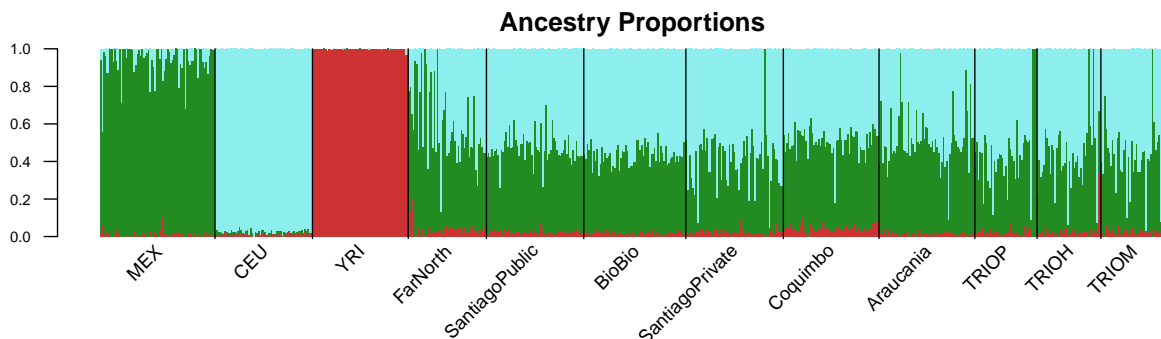
La Figura 9, resume las frecuencias alélicas estimadas en la muestra, para todos los SNPs.



**Figura 9:** Frecuencias alélicas de los SNPs genotificados con AXIOM LAT1

Los datos obtenidos fueron interceptados con datos de 60 individuos de ancestría europeo (CEU) y 59 individuos de ancestría africana (YRI) del proyecto Hapmap (<http://www.hapmap.org>). Como referencia amerindia, se usaron 71 muestras de un panel de indígenas mexicanos (MEX; Kumar et al. 2013)

Se utilizó el programa ADMIXTURE para estimar el porcentaje del genoma de cada individuo según origen continental, Europeo, Amerindio, y Africano. Se incluyeron las muestras CEU, MEX y YRI como referencia (Figura 10).



**Figura 10.** Componentes de ancestría para 475 individuos participantes de ChileGenómico y tres poblaciones de referencia (MEX nativo americano, CEU Europeo y YRI africanos).

Los resultados muestran un componente de ancestría amerindia significativo en la gran mayoría de las muestras chilenas. Algunos individuos muestran valores extremos, ya sea 100% ancestría amerindia o 100% de ancestría europea. Ningún participante mostró ancestría africana mayor al 34% (Tabla 4).

	<b>EUR</b>	<b>NAM</b>	<b>AFR</b>
Mín.	1.00E-05	1.00E-05	1.00E-05
Mediana	0.540	0.440	0.023
Promedio	0.530	0.450	0.027
SD	0.150	0.150	0.024
Máx.	0.970	1.000	0.340
<b>Población</b>	<b>EUR</b>	<b>NAM</b>	<b>AFR</b>
CEU	0.974	0.013	0.013
MEX	0.063	0.924	0.012
YRI	0.001	0.000	0.998
Araucania	0.483	0.500	0.017
BioBio	0.565	0.417	0.019
Coquimbo	0.492	0.465	0.043
Norte Grande	0.402	0.561	0.037
SantiagoPrivado	0.590	0.386	0.024
SantiagoPublico	0.527	0.448	0.025
TRIOH	0.562	0.408	0.030
TRIOM	0.556	0.418	0.025
TRIOP	0.566	0.409	0.025

**Tabla 4:** Estadísticos para tres componentes de ancestría  
TRIOH, TRIOM y TRIOP representan a hijo-as, madres y padres de los tríos.

#### **4. Genotipificación de 2963 chilenos con panel acotado de 150 SNPs diseñado durante el proyecto en virtud de los resultados de las dos etapas previas de genotipificación:**

Se envió un total de 2963 muestras (CLG) a genotipificar por LGC Limited (Middlesex, UK), para el panel de 150 SNPs elegidos. Los genotipos para dos marcadores, rs62045476 y rn145426 (novel, posición 4:100673238), fallaron la validación interna por LGC y sus datos no fueron recibidos. Tres muestras fueron enviadas en duplicado a modo de control. En este set de muestras se incluyeron 154 muestras que ya habían sido genotipificadas con AXIOM en etapa previa del proyecto, para probar mezcla de muestras y estimar tasa de error de genotipificación por el array. Estas incluyen 29 muestras de padres o madres del set de tríos.

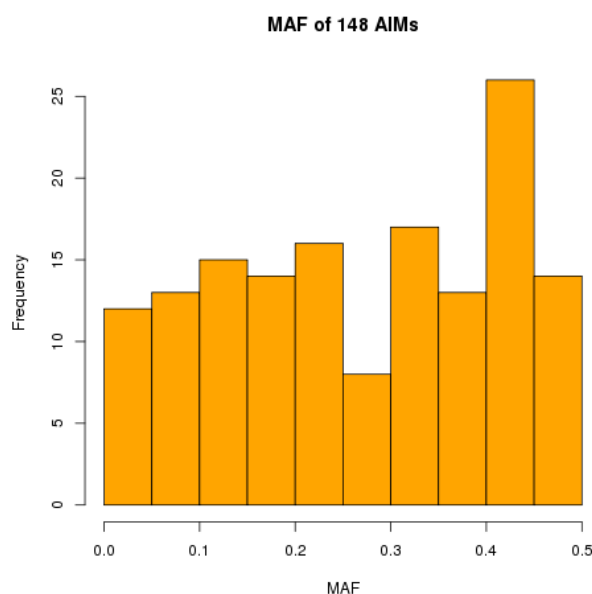
##### *i.Tasa de genotipificación:*

Utilizando plink (Purcell et al, 2014), se estimaron las tasas de datos perdidos por SNP y por Individuos/muestras. Un total de 1800 individuos (62,1%) no presentan datos perdidos. En 1047 individuos (35,9%) se observaron entre 1 y 5 genotipos perdidos y los 64 individuos restantes (2,3%) tienen entre 6 a 148 genotipos perdidos. Una muestra falló a nivel de todos los SNPs. A nivel de SNPs, la tasa de genotipos

perdidos es uniforme, con un promedio de 36 genotipos perdidos por SNP (1,2 %), en donde el mayor número de datos perdidos para un marcador es 63 y el menor 18.

ii. *Frecuencias alélicas:*

En cuanto a las *Minor Allele Frequency* o MAF se observan valores variables de frecuencia con un promedio de 0.27, destacándose un grupo de AIMs (aproximadamente 26) cuyos MAF están entre 0.4 y 0.45 (Figura 11). Al inspeccionar los MAF para cada AIM, se aprecia que el AIM *rs200923413* apareció como monomórfico (error de genotipificación que sólo fue capaz de detectar uno de los alelos), y luego le sigue un AIM novel obtenido de los datos de secuenciación, identificado como *rn381459*, que presentó un MAF menor a 0.01.



**Figura 11.** Histograma de *Minor Allele Frequency* para los 148 AIMs seleccionados.

iii. *Comparación con datos AXIOM:*

Algunas de las muestras obtenidas con el panel de AIMs, ya habían sido genotipificadas con AXIOM (154) y 2 muestras habían sido secuenciadas. Para estas dos últimas muestras, se utilizaron los datos de AXIOM.

En total, había 156 muestras y 142 SNPs en común entre AXIOM y CLG, ya que los 6 AIMs propuestos como nuevos no están presentes en los datos AXIOM. Se compararon los datos de los 142 SNPs conocidos en las 156 muestras con los datos de AXIOM utilizando R y se reportó el número de genotipos iguales entre AXIOM y CLG, evaluando a nivel de lugar de muestreo y a nivel de individuos. Los resultados se muestran en la Tabla 5 (no aparece muestras de Puerto Montt, pues éstas no se genotipificaron con AXIOM). La replicabilidad fue superior a 0.95 en todas las muestras y el promedio de replicabilidad fue superior a 0.97 en todas las poblaciones.



Población	N	SNPs	Total GT	Miss CLG	Miss Axiom	Match GT	Missmatch GT	Replicabilidad
ARI	9	142	1272	6	0	1244	28	0.9780
CDSJ	21	142	2977	4	1	2921	56	0.9812
CHI	20	142	2828	4	8	2771	57	0.9798
CLSM	9	142	1275	1	2	1252	23	0.9819
COQ	7	142	990	1	3	973	17	0.9828
IQQ	10	142	1403	3	14	1365	38	0.9728
MFMCH	3	142	407	17	2	398	9	0.9778
MPPA	6	142	845	3	4	830	15	0.9823
MPUC	3	142	425	0	1	418	7	0.9835
MUAB	3	142	425	0	0	418	7	0.9835
MUDP	1	142	141	0	1	138	3	0.9787
SER	11	142	1552	6	4	1525	27	0.9826
TEM	19	142	2667	12	19	2613	54	0.9797
TRIO	29	142	4051	52	15	3969	82	0.9797

**Tabla 5.** Comparación de genotipos idénticos entre CLG y AXIOM. GT denota Genotipo.

La replicabilidad fue estimada como  $Match_{GT}/Total_{GT}$ . Las columnas *MissCLG* y *MissAxiom* indican el número de genotipos perdidos en cada población.

ARI= Arica; CDSJ= Centro de Donantes de sangre Hospital San José Santiago; CHI= Chillán; CLSM= Clínica Santa María de Santiago, COQ=Coquimbo, IQQ= Iquique; MFMCH, MPPA, MPUC, MUAB, MUDP son colectas de sangre en terreno efectuadas por el Centro de Dadores Metropolitano Oriente en la Facultad de Ingeniería de la Universidad de Chile, Plaza de Puente Alto, Facultad de Medicina Universidad Católica, en Universidad Andrés Bello y en Universidad Diego Portales; SER= La Serena y TEM= Temuco

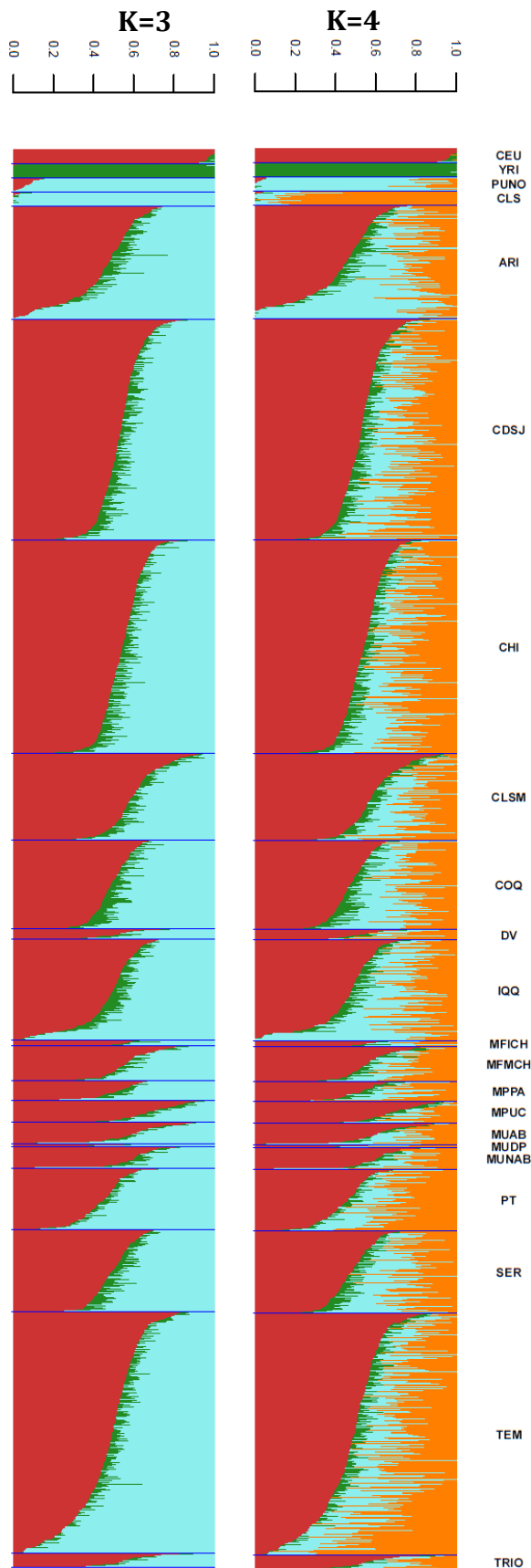
La genotipificación realizada en 2963 muestras chilenas presentó alta tasa de éxito y de reproducibilidad cuando fueron comparadas con datos independientes de genotipificación mediante la tecnología AXIOM. La reproducibilidad de los genotipos LGC, dentro de una misma corrida, fue prácticamente perfecta, con solo una diferencia en 3 muestras dada por un dato perdido.

#### *iv. Inferencia de ancestría:*

Utilizando el software ADMIXTURE inferimos componente de ancestría en los datos CLG (Alexander et al. 2009). Los datos de poblaciones de referencia de 30 individuos europeos y 30 individuos africanos fueron obtenidos del proyecto 1000 Genomas, poblaciones CEU y YRI respectivamente; como referente de ancestría Aymara se usaron 30 individuos de la localidad de Puno en Perú (aportados por Andrés Moreno de Stanford University) y de ancestría Mapuche, 30 individuos Pehuenches y Huilliches del sur de Chile (Llop 1996, Proyecto CONICYT DRI USA2013-0015). Cuatro SNPs se encontraban físicamente cerca presentando LD de  $r^2 \geq 0.2$ ; sin embargo, eliminarlos no afectó significativamente los resultados de ancestría y por lo tanto no fueron eliminados. Una prueba de validación cruzada sugirió usar un número de componentes  $K=3$  o  $K=4$  ancestrías. Los resultados de la inferencia de ancestría se resumen en la Tabla 6 y en la Figura 12. Los resultados muestran una ancestría amerindia promedio  $.44 \pm .14$  y discretas diferencias entre puntos de muestreo. Sin embargo, existe mayor heterogeneidad para la relación Andino/Mapuche a lo largo del país.

<b>Población</b>	<b>N</b>	<b>Europea</b>	<b>Africana</b>	<b>Amerindia</b>	<b>Aymara</b>	<b>Mapuche</b>
<i>Poblaciones de referencia</i>						
CEU	30	.97±.01	.02±.00	.00±.00	.01±.00	.00±.00
YRI	30	.00±.00	.99±.00	.00±.00	.00±.00	.00±.00
PUNO	30	.02±.00	.00±.00	.92±.01	.83±.02	.15±.02
CLS	30	.01±.01	.01±.00	.98±.01	.16±.01	.82±.02
<i>Muestras ChileGenómico</i>						
ARI	239	.41±.01	.05±.00	.50±.01	.36±.02	.18±.01
CDSJ	457	.53±.00	.04±.00	.42±.00	.18±.01	.25±.01
CHI	442	.54±.00	.04±.00	.42±.00	.17±.01	.25±.01
CLSM	181	.60±.01	.04±.00	.34±.01	.17±.01	.19±.01
COQ	182	.48±.01	.06±.00	.45±.01	.24±.01	.22±.01
DV	22	.54±.03	.04±.01	.41±.03	.16±.02	.26±.03
IQQ	211	.45±.01	.05±.00	.47±.01	.30±.01	.20±.01
MFICH	13	.60±.04	.05±.01	.35±.04	.14±.03	.21±.03
MFMCH	73	.57±.01	.05±.00	.37±.01	.16±.01	.22±.01
MPPA	41	.52±.02	.05±.01	.43±.02	.16±.02	.27±.02
MPUC	47	.70±.02	.05±.00	.25±.02	.10±.01	.15±.02
MUAB	45	.62±.03	.04±.00	.34±.03	.15±.02	.19±.02
MUDP	7	.57±.07	.05±.02	.38±.08	.16±.05	.23±.06
MUNAB	44	.60±.02	.05±.01	.35±.02	.15±.02	.21±.02
PT	130	.46±.01	.03±.00	.52±.01	.17±.01	.35±.01
SER	172	.48±.01	.05±.00	.45±.01	.24±.01	.23±.01
TEM	509	.47±.01	.04±.00	.49±.01	.18±.01	.31±.01
TRIO	28	.57±.02	.04±.01	.38±.02	.18±.02	.21±.02
<b>Total</b>						
<b>ChileGenómico</b>	<b>2963</b>	<b>.51±.14</b>	<b>.05±.04</b>	<b>.44±.14</b>	<b>.20±.15</b>	<b>.25±.13</b>

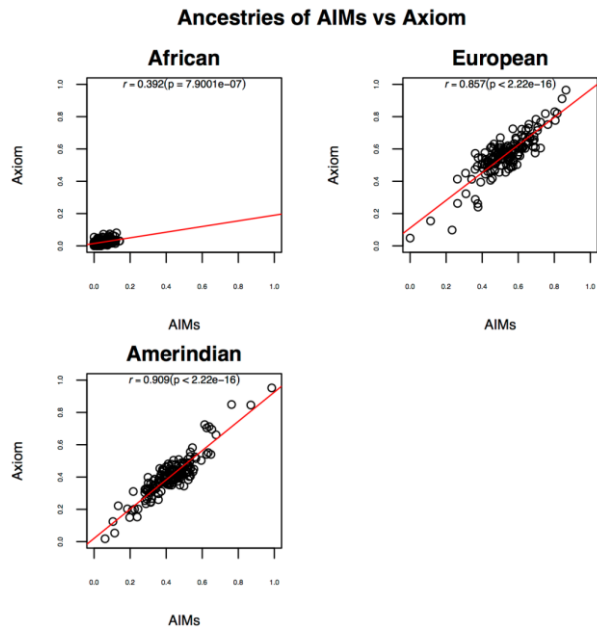
**Tabla 6.** Proporciones de ancestría según ADMIXTURE. Todos los valores fueron estimados con K=4, excepto la ancestría amerindia en la columna central que fue estimada con K=3. CEU, YRI, PUNO y CLS son poblaciones de referencia Europea, Africana, Aymara de Puno y Pehuenches y Huilliches del sur de Chile



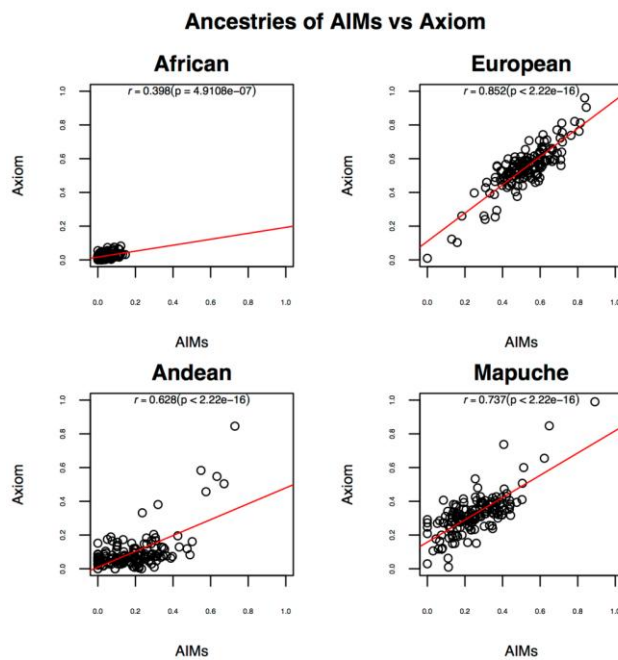
**Figura 12:** Inferencias de ancestría en 2991 participantes del proyecto ChileGenómico considerando K= 3 y K=4 poblaciones de referencia. Las poblaciones CEU (rojo) y YRI (verde) provienen del proyecto 1000G y representan ancestría Europea y Africana respectivamente. La población PUNO (celeste) proviene de UNA, Puno, en Perú y fueron obtenidos mediante colaboración con el grupo del Dr. Carlos Bustamante (Stanford University). Los individuos CLS (naranja) tienen ancestría Pehuenche y Huilliche y fueron seleccionados del proyecto Patagonia DNA por su alta ancestría amerindia.

v. *Replicabilidad de inferencias de ancestría*

Para 156 individuos que cuentan con genotipos de la plataforma AXIOM y del panel de AIMs, comparamos los resultados de la inferencia de ancestría utilizando ADMIXTURE. El análisis suponiendo un  $K=3$  muestra una correlación de 0.86 para inferencias de ancestría Europea y de 0.9 para ancestría Amerindia (Figura 13). Aumentando a  $K=4$ , detectamos una correlación de 0.87 para ancestría Andina del Norte (aymara) y 0,74 para la ancestría Mapuche (Figura 14). La ancestría Africana muestra una baja correlación (0.39), pero distinta de 0 ( $p=7 \times 10^{-7}$ ) dado que este es un componente muy menor en los participantes de ChileGenómico.



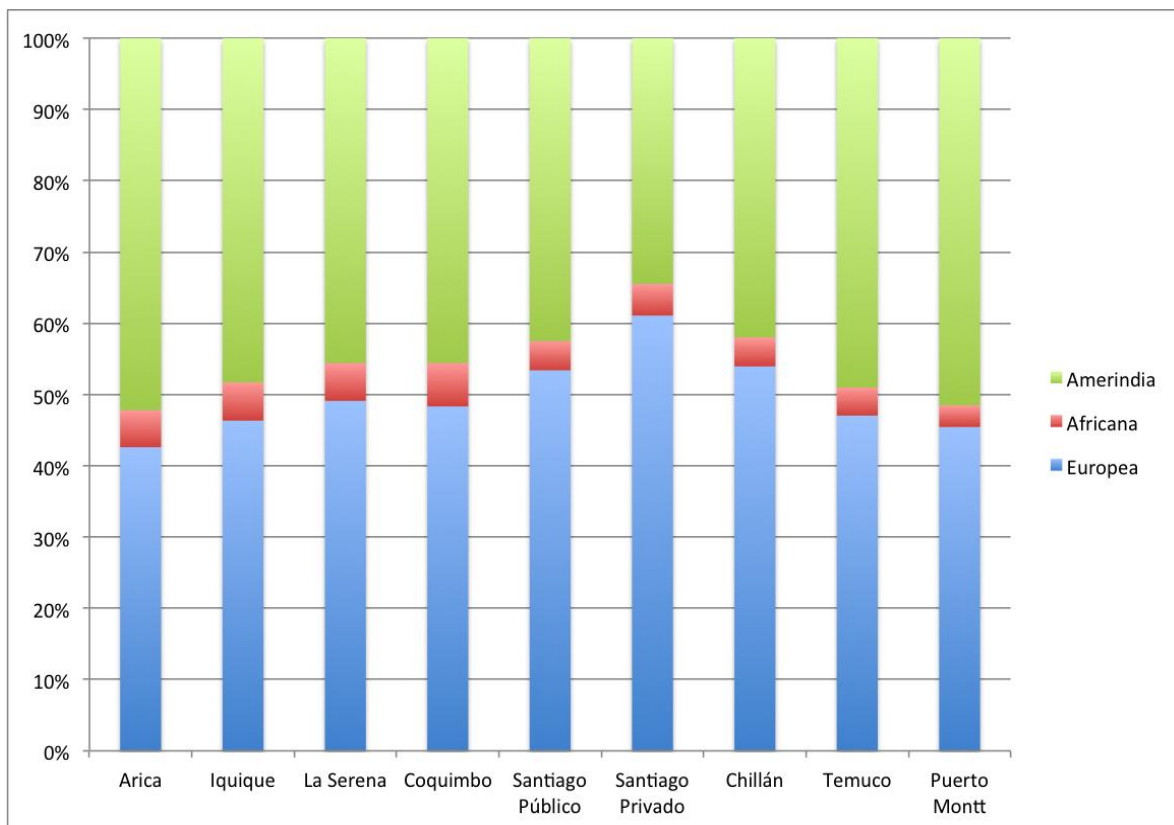
**Figura 13.** Comparación de valores de ancestría estimados a partir de datos de genoma completo (AXIOM) y del paneles de 147 marcadores conocidos con datos válidos e independientes con  $K=3$  (CLG).



**Figura 14.** Comparación de valores de ancestría estimados a partir de datos de genoma completo (AXIOM) y del paneles de 147 marcadores conocidos con datos válidos e independientes con  $K=4$  (CLG).

vi. Variación en la magnitud del componente Amerindio en distintos subgrupos de chilenos mixtos:

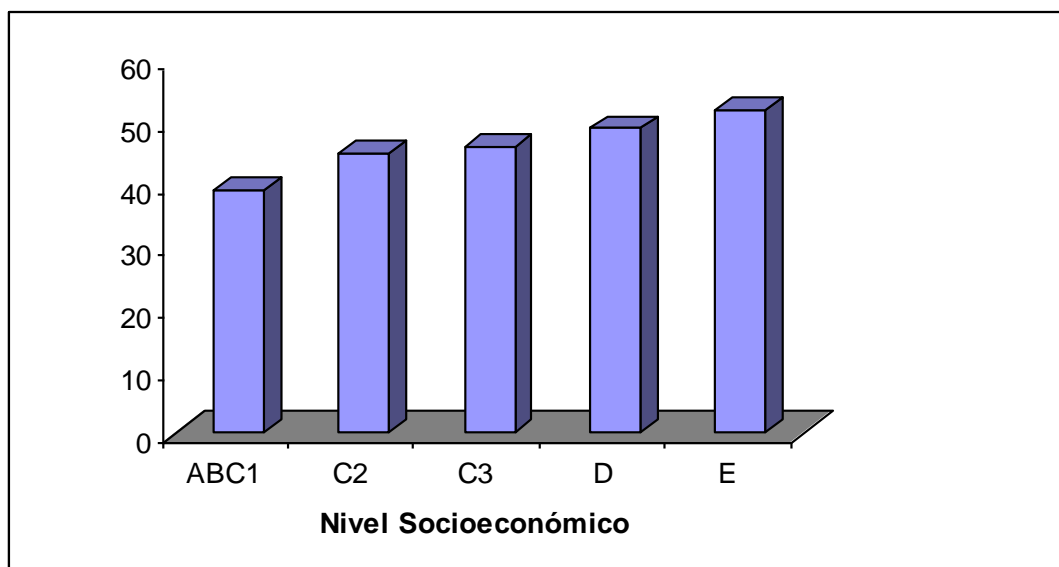
La magnitud de la ancestría Amerindia varía a lo largo de Chile según el punto de muestreo, como se observa en la Tabla 6 y Figura 15 ( $p < 0.001$ ), siendo mayor en los extremos sur y norte y más bajo en regiones centrales del país; la ancestría africana por su parte, está presente en niveles muy bajos en los chilenos de todas las regiones.



**Figura 15.** Porcentajes de ancestría Amerindia, Europea y africana promedios para cada región de muestreo.

La ancestría Amerindia promedio también depende del nivel socioeconómico de los sujetos, como se aprecia en la Figura 16, aunque se observa bastante varianza al interior de cada grupo como se puede apreciar en la Tabla 7. Al hacer comparaciones múltiples entre el nivel de ancestría amerindia entre los distintos niveles socioeconómicos se observa que el nivel ABC1 difiere de todo el resto, mientras que C2 y C3 no difieren significativamente entre sí, como tampoco lo hacen los grupos C3, D y E entre sí. En la Tabla 8 se observa el nivel de ancestría amerindia promedio en los grupos ABC1 y C2 C3 D y E de todas las ciudades muestreadas.

Al hacer un análisis de regresión lineal múltiple se observa que las diferencias en los niveles de ancestría Amerindia entre las distintas ciudades son estadísticamente significativas, una vez ajustadas por el nivel socioeconómico. También se comprueba el aumento del nivel de ancestría Amerindia a menor nivel socioeconómico, una vez hecho el ajuste por ciudad de muestreo. El nivel de ancestría Amerindia también es diferente entre las distintas clases sociales como se aprecia en la Figura 17.



**Figura 16:** Porcentajes de componente de ancestría Amerindia estimados con panel de 147 SNPs informativos, según nivel socioeconómico ( $p < 0.001$ )

PORCENTAJE DE ANCESTRÍA AMERINDIA

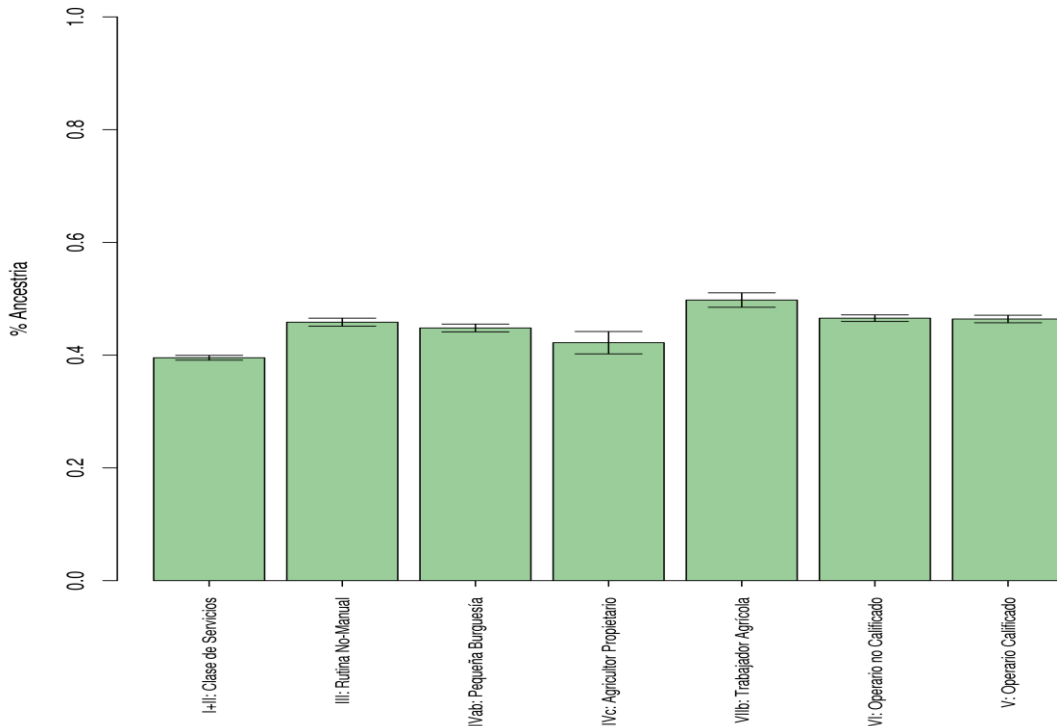
	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
ABC1	829	,38675738	,137091703	,004761392	,37741156	,39610320	,000010	,940623
C2	972	,44789414	,129776791	,004162593	,43972543	,45606286	,000010	,999980
C3	633	,46194871	,135144383	,005371507	,45140055	,47249687	,162239	,986291
D	338	,49262836	,154304344	,008393050	,47611899	,50913772	,099776	,997861
E	40	,51614343	,164319668	,025981221	,46359145	,56869540	,312082	,945207
S/D	32	,51078913	,164207153	,029027998	,45158613	,56999212	,148667	,984379
Total	2844	,44018561	,142109523	,002664762	,43496055	,44541067	,000010	,999980

**Tabla 7.** Niveles de ancestría Amerindia según estrato socioeconómico  $p < 0.001$

PORCENTAJE DE ANCESTRÍA AMERINDIA

Ciudad de Muestreo	ABC1		C2, C3, D y E		p
	n,	Promedio $\pm$ Error St.	n,	Promedio $\pm$ Error St.	
Arica	73,	44.3 $\pm$ 1.69	161,	53.0 $\pm$ 1.55	0.001
Iquique	92,	45.6 $\pm$ 1.23	112,	48.5 $\pm$ 1.60	0.159
La Serena	51,	43.1 $\pm$ 1.23	118,	45.5 $\pm$ 0.81	0.102
Coquimbo	32,	42.3 $\pm$ 8.24	146,	44.8 $\pm$ 0.65	0.126
Santiago Público	107,	39.6 $\pm$ 0.93	393,	43.5 $\pm$ 0.56	0.001
Santiago Privado	248,	30.7 $\pm$ 0.89	182,	38.9 $\pm$ 0.87	<0.001
Chillán	56,	37.2 $\pm$ 1.58	381,	42.3 $\pm$ 0.45	0.002
Temuco	125,	41.1 $\pm$ 1.21	381,	51.7 $\pm$ 0.84	<0.001
Puerto Montt	34,	47.6 $\pm$ 1.93	94,	53.4 $\pm$ 1.33	0.022

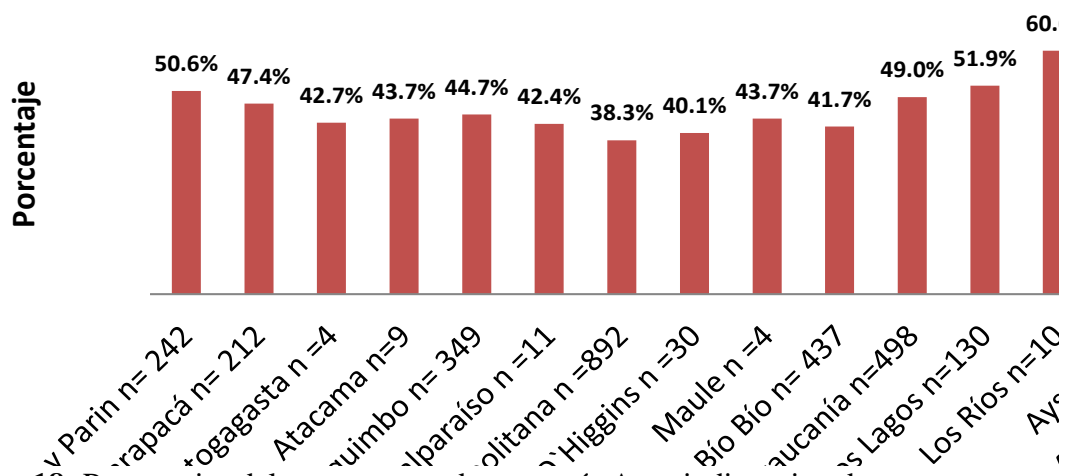
**Tabla 8:** Porcentajes promedio de ancestría amerindia en chilenos de NSE ABC1 y C2 C3 D y E de distintas ciudades



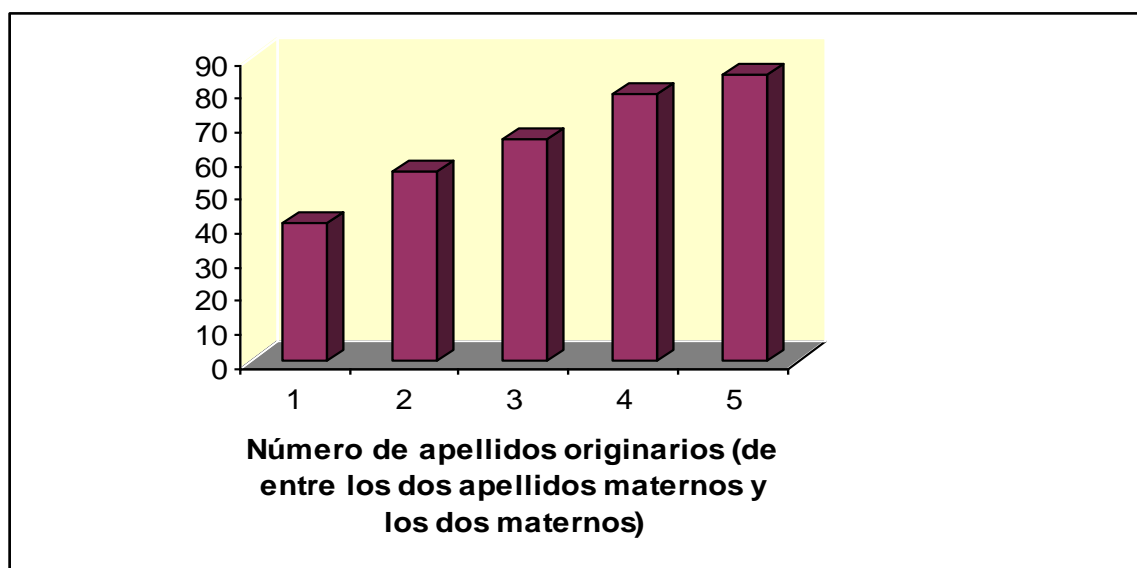
**Figura 17.** Porcentajes de componente de ancestría Amerindia estimados con panel de 147 SNPs informativos, según clase social ( $p < 0.001$ )

El nivel de ancestría Amerindia también es diferente según Región de residencia de los individuos, aunque para algunas regiones los individuos reclutados fueron muy pocos. (Figura 18). Finalmente el número de apellidos de origen Amerindio que tienen los sujetos se correlaciona con el nivel de ancestría Amerindia como se observa en la Figura 19, así como el antecedente de tener antepasados originarios (Figura 20)

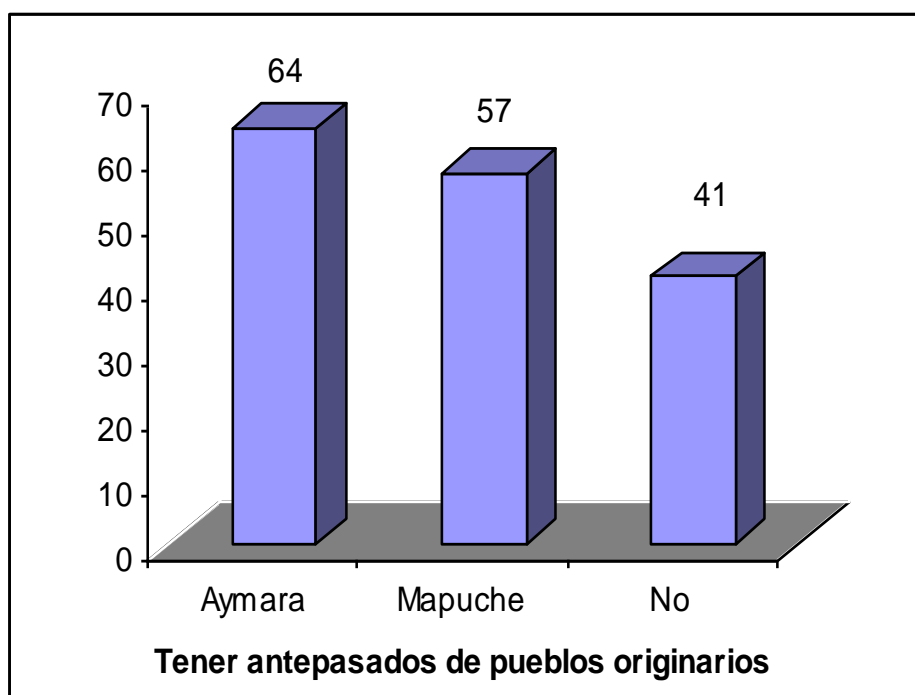
### Porcentaje de ancestría amerindia según región de residencia



**Figura 18:** Porcentajes del componente de ancestría Amerindia estimados con panel de 147 SNPs informativos, según Región de Residencia ( $p < 0.001$ )



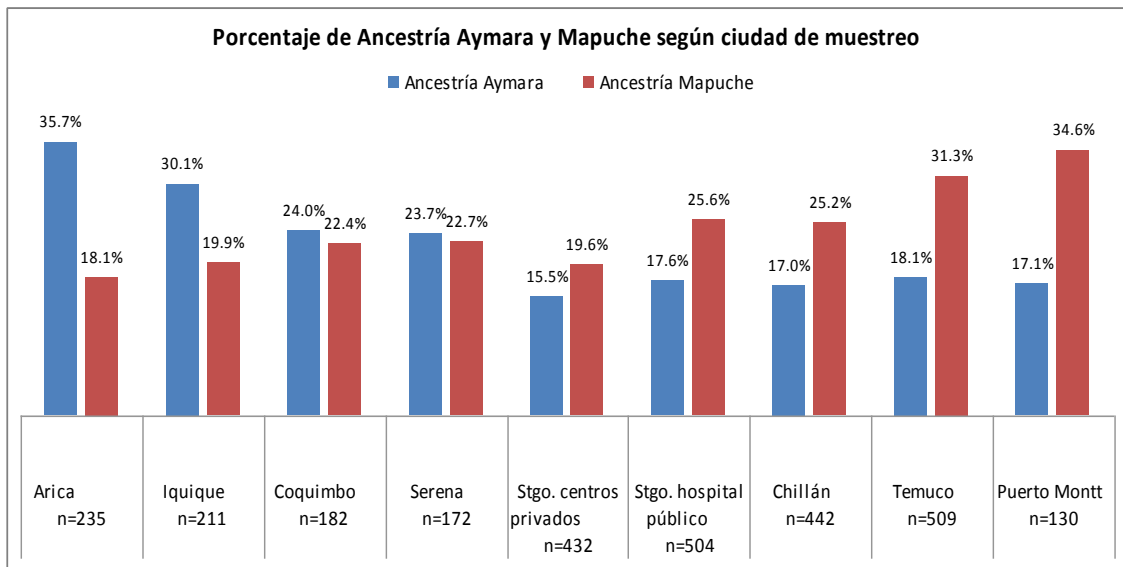
**Figura 19:** Porcentajes del componente de ancestría Amerindia estimados con panel de 147 SNPs informativos de ancestría, según el número de apellidos de origen Amerindio que tienen los sujetos ( $p < 0.001$ )



**Figura 20.** Porcentajes del componente de ancestría Amerindia estimados con panel de 147 SNPs informativos de ancestría, según si hay antecedentes de algún antepasado originario ( $p < 0.001$ )

Al descomponer el componente Amerindio según si se trata de Amerindio del Norte (Aymara) o Amerindio del sur (Mapuche) se observa que ambos componentes varían a lo largo del país (Figura 21). El componente aymara presenta un claro gradiente de norte sur y al revés ocurre con el componente mapuche que aumenta paulatinamente desde el norte hacia el sur del país.



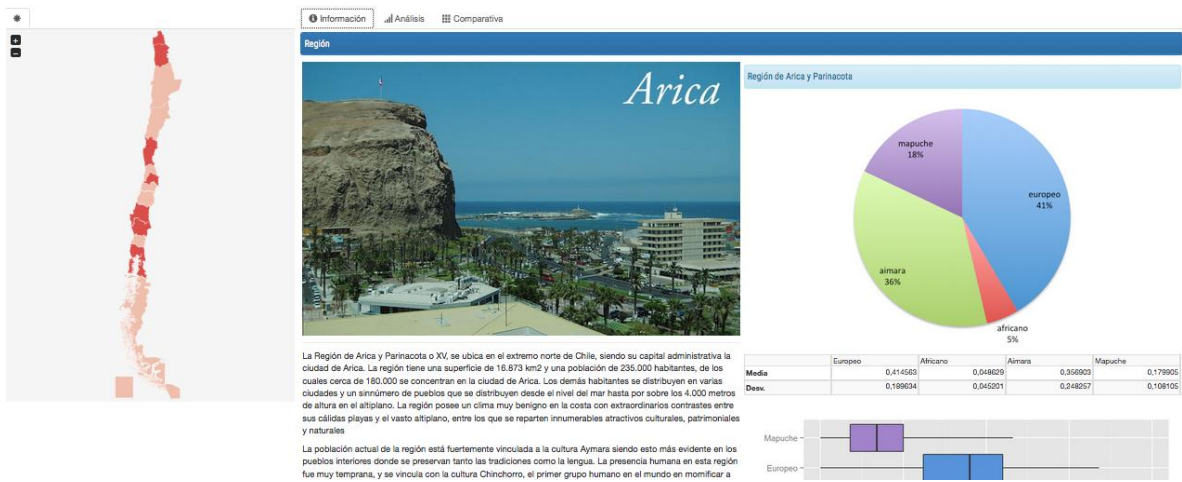


**Figura 21.** Porcentajes de ancestría aymara y mapuche en las distintas regiones del país.

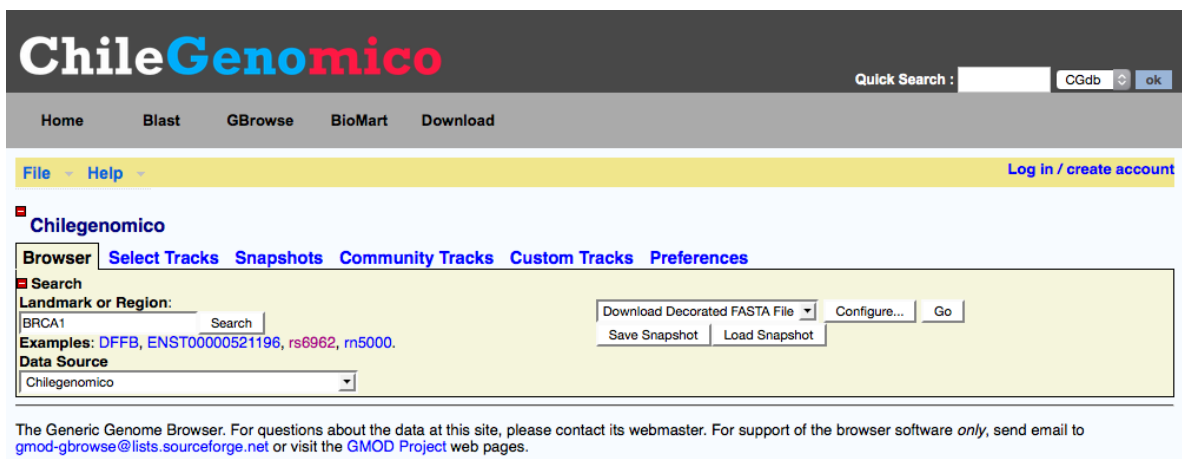
## DIVULGACIÓN DE RESULTADOS

Se ha implementado un sitio web (<http://www.chilegenomico.cl>) que contiene la mayor parte de los resultados, así como noticias y entrevistas vinculadas al proyecto. En este sitio se ha implementado una interfaz gráfica que permite desplegar algunos de los contenidos sobre posiciones georeferenciadas. Así, el usuario puede acceder a través del mouse a las distintas zonas del mapa de Chile y la información almacenada en ellas, permitiendo visualizar y analizar componentes de ancestría zona a zona del país, así como comparar diferentes zonas. Entre las visualizaciones se incorporan tablas resúmenes de estadística de ancestría, coloreado de zonas por ancestría (escogida por usuario) y diagramas tipo torta (ver Figura 21).

Adicionalmente, se ha construido una Base de datos CGdb ([chilegenomico.cmm.uchile.cl](http://chilegenomico.cmm.uchile.cl)), con acceso controlado que contiene la información de los 18 genomas chilenos secuenciados y las frecuencias alélicas para el panel acotado de SNPs en los 2900 chilenos; alberga también un total de 40 millones de SNPs almacenados, con 130 millones de frecuencias alélicas, un total de 11 poblaciones (1000G más 2 chilenas) y 17 plataformas de GWAS (Arrays), además de referencia cruzada con otras fuentes de información como Pubmed, dbSNP. Contiene un motor de búsqueda de variantes (GBrowse) (Figuras 22 y 23) y una herramienta Biomart que permite personalizar esa búsqueda según determinados filtros en población chilena y 1000G (ver Figura 24)



**Figura 21.** Inicio del visualizador interactivo. De izquierda a derecha, se muestra el mapa de Chile, información regional e información estadística de ancestría



**Figura 22.** Búsqueda de variantes por nombre de gen en GBrowse

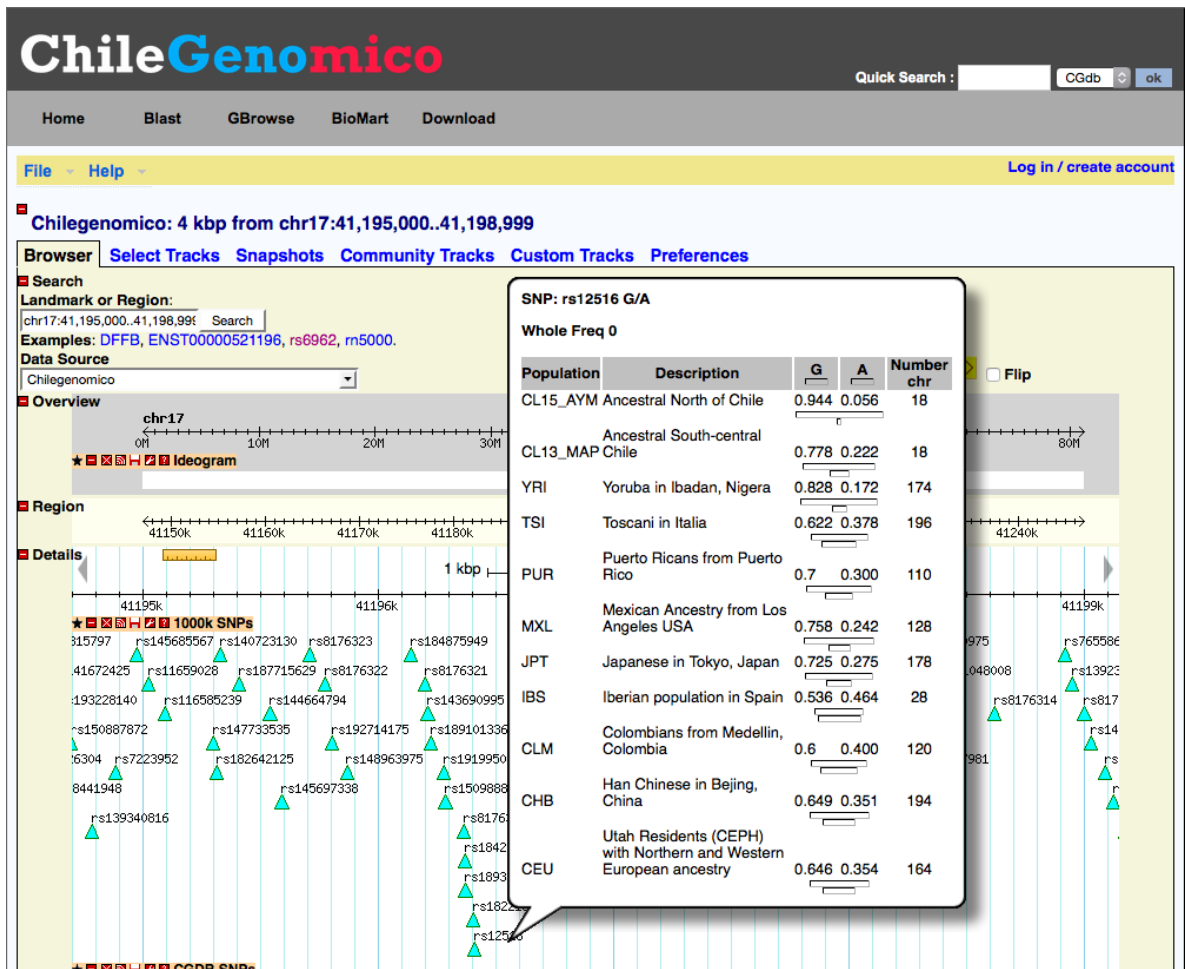


Figura 23. Frecuencias alélicas para un SNP en el gen BRCA1 en múltiples poblaciones visualizadas en la herramienta GBrowse

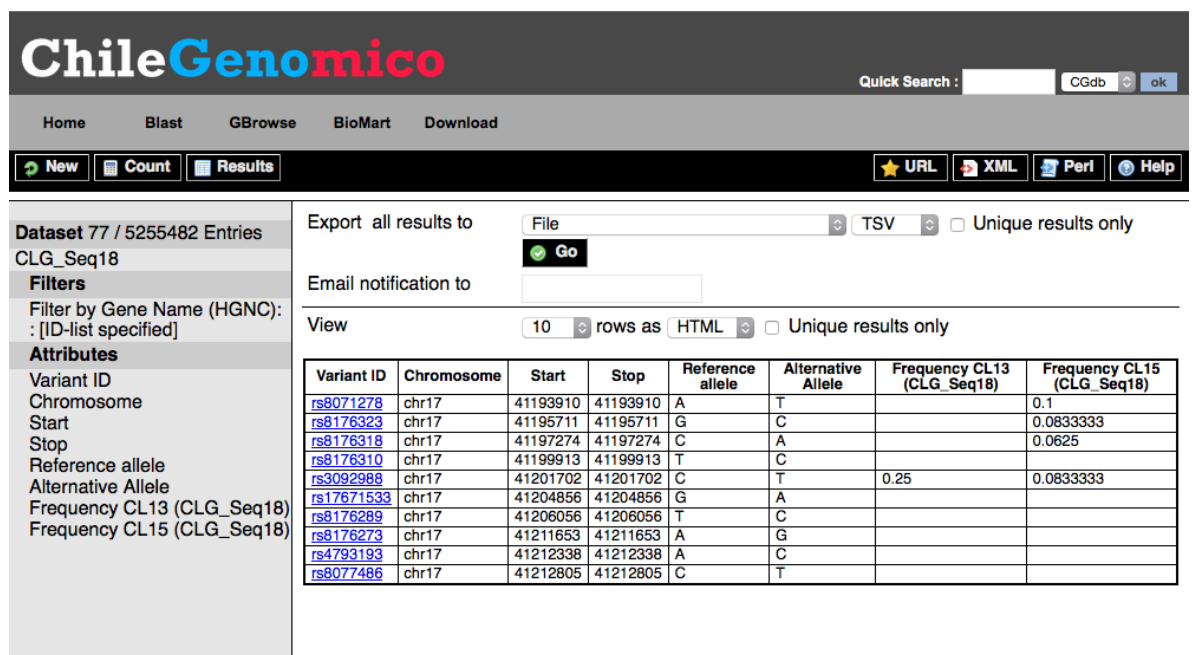


Figura 24. Resultado de búsqueda por la herramienta BioMart.

## CONCLUSIONES

Los resultados de nuestro proyecto de investigación corroboran la composición ancestral mixta de la población chilena mediante el uso de herramientas genómicas modernas que demuestran nuestra heterogeneidad. Esta ancestría principalmente nativo americana y europea, presenta variaciones porcentuales entre las distintas regiones del país y en distintos grupos sociales. El componente africano es pequeño, un poco más alto en el norte del país.

El desarrollo de esta investigación ha permitido el diseño y puesta a prueba de un panel acotado de marcadores genéticos informativos de ancestría adecuado para población chilena, el que podrá ser de utilidad para diversos grupos de investigadores en diversas áreas como determinación de factores genéticos de enfermedades, epidemiología, genética forenses, salud pública, etc.

Adicionalmente, este estudio ha permitido el desarrollo de capacidades en el país (análisis de ancestría a nivel genómico, desarrollo de capacidades bioinformáticas de análisis de datos genómicos, secuenciación de genomas humanos, etc) y ha generado información disponible para distintos públicos acerca de la ancestría de la población chilena, la cual se divulga en sitio web y en una base de datos para uso de científico.

## REFERENCIAS

Acuña M, Llop E, Rothhammer F. Genetic composition of Chilean population: rural communities of Elqui, Limari and Choapa valleys. *Rev Med Chil.* 2000 Jun;128(6):593-600.

Aguirre R, Blanco R, Cifuentes L, Chiffelle I, Armanet L, Vargas J, Jara L. Paternity study in Chilean families using DNA fingerprints and erythrocyte blood markers. *Rev Med Chil.* 1992 Oct;120(10):1095-101.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9): 1655–64. doi:10.1101/gr.094052.109.

Allison MJ, Hossaini AA, Munizaga J, Fung R. ABO blood groups in Chilean and Peruvian mummies. II. Results of agglutination-inhibition technique. *Am J Phys Anthropol.* 1978 Jul;49(1):139-42.

Armanet L, Aguirre R, Vargas J, Llop E, Castillo S, Cifuentes L. Paternity study applying DNA polymorphism: evaluation of methods traditionally used in Chile. *Rev Med Chil.* 1995 May;123(5):560-6.

Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A.* 2006 May 9;103(19):7234-9. Epub 2006 Apr 28.

Brisbin, Abra, Katarzyna Bryc, Jake Byrnes, Fouad Zakharia, Larsson Omberg, Jeremiah Degenhardt, Andrew Reynolds, Harry Ostrer, Jason G. Mezey, and Carlos D. Bustamante. 2012. "PCAdmix: Principal Components-Based Assignment of Ancestry along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations." *Human Biology* 84 (4): 343–64. doi:10.3378/027.084.0401.

Chung SA, Tian C, Taylor KE, Lee AT, Ortmann WA, Hom G, Graham RR, Nititham J, Kelly JA, Morrisey J, Wu H, Yin H, Alarcón-Riquelme ME, Tsao BP, Harley JB, Gaffney PM, Moser KL, Manzi S, Petri M, Gregersen PK, Langefeld CD, Behrens TW, European population substructure is associated with mucocutaneous manifestations and autoantibody production in systemic lupus erythematosus. *Arthritis Rheum.* 2009 Aug;60(8):2448-56

Cifuentes L, Armanet L, Nazer J. Genes of negroid origin in a urban population of Santiago. *Rev Med Chil.* 1995 May;123(5):649-50.

Cifuentes L, Jorquera H, Aguirre E, Moreno F. Paternity determination in cases of incest using multiallelic loci DNA polymorphisms. *Rev Med Chil.* 1996 Nov;124(11):1307-13.

Cifuentes L, Morales R, Sepúlveda D, Jorquera H, Acuña M. DYS19 and DYS199 loci in a Chilean population of mixed ancestry. *Am J Phys Anthropol.* 2004 Sep;125(1):85-9.

Droguett MA, Beltran R, Ardiles R, Raddatz N, Labraña C, Arenas A, Flores J, Alruiz P, Mezzano S, Ardiles L. Ethnic differences in HLA antigens in Chilean donors and recipients: data from the National Renal Transplantation Program. *Transplant Proc.* 2008 Nov;40(9):3247-50.

Etcheverry R, Guzmán C, Boris F. Investigación de grupos sanguíneos y de otros caracteres genéticos en indígenas de Chile y nativos pascuenses. *Rev Med Chile* 1967;95:599-613.

Etcheverry R. Razas humanas y hémat-sero-antropología. Origen de los indígenas de Chile y nativos pascuenses en el concierto de las razas humanas. *Rev Med Chile* 1997; 125:1073-1081.

Eyheramendy S, Martínez F, Manevy F et al. Genetic structure characterization of Chileans reflects historical immigration patterns. *Natur Commun* 2015: 17(6):6472

Fagundes N, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano F, Smith DG, Silva Jr WA, Zago M, Ribeiro-dos-Santos AK, Santos S, Petzl-Erler ML, Bonatto SL. Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. *Am J Hum Genet.* 2008: 82: 583-592.

Fuentes M, Pulgar I, Gallo C et al. Geografía génica de Chile. Distribución regional de los aportes genéticos americanos, europeos y africanos. *Rev Med Chile* 2014; 142:281-289

García F, Moraga M, Vera S, Henríquez H, Llop E, Aspillaga E, Rothhammer F. mtDNA microevolution in Southern Chile's archipelagos. *Am J Phys Anthropol.* 2006 Mar;129(3):473-81.

Goebel T, Waters MR, O'Rourke DH. The Late Pleistocene Dispersal of Modern Humans in the Americas. *Science* 2008: 319:1497-1502.

Goedde HW, Rothhammer F, Benkmann HG, Bogdanski P. Genetic studies in Atacameño Indians: serum protein and red cell-enzyme polymorphisms. *Ann Hum Biol.* 1985 May-Jun;12(3):251-9.

González-José R, Bortolini MC, Santos FR and Bonatto S. The Peopling of America: Craniofacial Shape Variation on a Continental Scale and its Interpretation From an Interdisciplinary View. *American Journal of Physical Anthropology* 2008;137:175–187.

Haas EJ, Salzano FM, Araujo HA, Grossman F, Barbetti A, Weimer TA, Franco MH, Verruno L, Nasif O, Morales VH, et al. HLA antigens and other genetic markers in the Mapuche Indians of Argentina. *Hum Hered.* 1985;35(5):306-13.

Harb Z, Llop E, Moreno R, Quiroz D. Coastal Chilean populations: genetic markers in four locations. *Rev Med Chil.* 1998; 126: 753-760.

Hardy BJ, Séguin B, Goodsaid F, Jimenez-Sanchez G, Singer PA, Daar AS. The next steps for genomic medicine: challenges and opportunities for the developing world. *Nat Rev Genet.* 2008 Oct;9 Suppl 1:S23-7.

Hardy BJ, Séguin B, Singer PA, Mukerji M, Brahmachari SK, Daar AS. From diversity to delivery: the case of the Indian Genome Variation initiative. *Nat Rev Genet.* 2008 Oct;9 Suppl 1:S9-14.

Hardy BJ, Séguin B, Ramesar R, Singer PA, Daar AS. South Africa: from species cradle to genomic applications. *Nat Rev Genet.* 2008 Oct;9 Suppl 1:S19-23.

Henríquez H, Moraga M, Llop E, Rothhammer F. Molecular and genetic characterization of Changos descendants living in Paposo Cove. *Rev Med Chil.* 2004 Jun;132(6):663-72.

HUGO Pan-Asian SNP Consortium, Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, Chaurasia A, Chen CH, Chen J, Chen YT, Chu J, Cutiongco-de la Paz EM, De Ungria MC, Delfin FC, Edo J, Fuchareon S, Ghang H, Gojobori T, Han J, Ho SF, Hoh BP, Huang W, Inoko H, Jha P, Jinam TA, Jin L, Jung J, Kangwanpong D, Kampuansai J, Kennedy GC, Khurana P, Kim HL, Kim K, Kim S, Kim WY, Kimm K, Kimura R, Koike T, Kulawonganunchai S, Kumar V, Lai PS, Lee JY, Lee S, Liu ET, Majumder PP, Mandapati KK, Marzuki S, Mitchell W, Mukerji M, Naritomi K, Ngamphiw C, Niikawa N, Nishida N, Oh B, Oh S, Ohashi J, Oka A, Ong R, Padilla CD, Palittapongarnpim P, Perdigon HB, Phipps ME, Png E, Sakaki Y, Salvador JM, Sandraling Y, Scaria V, Seielstad M, Sidek MR, Sinha A, Srikummool M, Sudoyo H, Sugano S, Suryadi H, Suzuki Y, Tabbada KA, Tan A, Tokunaga K, Tongsimma S, Villamor LP, Wang E, Wang Y, Wang H, Wu JY, Xiao H, Xu S, Yang JO, Shugart YY, Yoo HS, Yuan W, Zhao G, Zilfalil BA; Indian Genome Variation Consortium. Mapping human genetic diversity in Asia. *Science.* 2009; 326: 1541-1545.

Jorquera H, Acuña M, Cifuentes L, Aguirre E, Moreno F. Paternity analysis using four DNA markers amplified by polymerase chain reaction. *Rev Med Chil.* 1999 Sep;127(9):1057-62.

Jorquera GH, Acuña P M, Cifuentes L L. Kinship determination using DNA markers. *Rev Med Chil.* 2008 Feb;136(2):193-200. Epub 2008 May 7.

Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, Seldin MF. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat.* 2009 Jan;30(1):69-78.

Kumar R, Nguyens E, Roth L et al. Factors associated with degree of atopy in Latino children in a nationwide pediatric sample: The GALA II Study Allergy Clin Immunol. *J Allergy Clin Immunol.* 2013 Oct; 132(4): 896–905.

Lahiri DK, Bye S, Nurnberger JI Jr, Hodes ME, Crisp M. A non-organic and non-enzymatic extraction method gives higher yields of genomic DNA from whole-blood

samples than do nine other methods tested. *J Biochem Biophys Methods*. 1992 Dec;25(4):193-205.

Llop E, Harb Z, Acuña M, Moreno R, Barton S, Aspillaga E, Rothhammer F. Genetic composition of Chilean population: the Pehuenches from Trapa-Trapa. *Rev Med Chil*. 1993 May; 121(5):494-8.

Llop E, 1996. Genetic Composition of Chilean Aboriginal Populations: HLA and other genetic marker variation. *Am J Phys Anthropol* 101:325- 332

Llop E., Henríquez H., Moraga M., Castro M. y Rothhammer F. Brief Communication: Molecular Characterization of O Alleles at the ABO Locus in Chilean Aymara and Huilliche Indians. *Am. J. Phys. Anthrop.* 2006: 131:535-538,

Matson G, Sutton E, Etcheverry R, Swanson J, Robinson A. Distribution of hereditary blood groups among Indians in South America. *Am J Phys Anthropol* 1969; 27:157-194.

Merriwether DA, Ferrell RE, Rothhammer F. mtDNA D-loop 6-bp deletion found in the Chilean Aymara: not a unique marker for Chibcha-speaking Amerindians. *Am J Hum Genet*. 1995; 56: 812-813.

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty, R, Rothhammer F, Carvallo P. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol* 2000; 113:19–29.

Moraga M, Santoro CM, Standen VG, Carvallo P, Rothhammer F. Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile. *Am J Phys Anthropol*. 2005 Jun; 127(2):170-81.

Moraga M, de Saint Pierre M, Torres F, and Ríos J. Vínculos de parentesco por vía materna entre los últimos descendientes de la etnia Kawésqar y algunos entierros en los canales patagónicos: evidencia desde el estudio de linajes mitocondriales. *Magallania* 2010; 38(2):102-114.

Proyecto CONICYT DRI USA2013-0015 Genomic investigation of the human biodiversity in the pre-Columbian and contemporary Chilean Patagonia

Purcell, Shaun, and Christopher Chang. 2014. *PLINK v1.9*. Accessed December 2. <https://www.cog-genomics.org/plink2>.

Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov.. “Integrative Genomics Viewer.” *Nature Biotechnology* 2011; 29 (1): 24–26. doi:10.1038/nbt.1754.

Rocco P, Morales C, Moraga M, Miquel JF, Nervi F, Llop E, Carvallo P, Rothhammer F. Composición Genética de la Población Chilena. Distribución de Polimorfismos de DNA Mitocondrial en Grupos Originarios y en la Población Mixta de Santiago. *Rev Méd Chile* 2002; 130:125-131.



Rosenberg, Noah A, Lei M Li, Ryk Ward, and Jonathan K Pritchard. "Informativeness of Genetic Markers for Inference of Ancestry." *American Journal of Human Genetics* 2003; 73 (6): 1402–22. doi:10.1086/380416.

Rothhammer F, Moraga M, Santoro C, Arriaza B, Origen de los Changos. Análisis de ADNmt antiguo sugiere descendencia de pescadores de la cultura Chinchorro (7.900-4.000 A.P.) *Rev Med Chile* 2010;138: 251-256.

Séguin B, Hardy BJ, Singer PA, Daar AS. Genomics, public health and developing countries: the case of the Mexican National Institute of Genomic Medicine (INMEGEN). *Nat Rev Genet.* 2008 Oct;9 Suppl 1:S5-9.

Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, Silva G, Belmont JW, Hanson RL, Knowler WC, Gregersen PK, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet.* 2007 Jun;80(6):1014-23.

Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK, Seldin MF. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* 2008 Jan;4(1):e4.

Thayer Ojeda L. Elementos étnicos que han intervenido en la población de Chile. Imprenta, Litografía y Enc. "La Ilustración". Santiago, Chile, 1919.

Valenzuela CY. Dimorfismo sexual pondoestatural en una población chilena. Evidencia de genes para estatura en los cromosomas sexuales? *Rev Med Chile* 1975 103(5)322-26.

Valenzuela CY, Harb Z. Socioeconomic assortative mating in Santiago, Chile: a demonstration using stochastic matrices of mother-child relationships applied to ABO blood groups. *Soc Biol.* 1977; 24: 225-233.

Valenzuela CY, Acuña M, Harb Z. Gradiente sociogenético en la población chilena. *Rev Med Chile* 1987;115: 295-299.

Valenzuela CY. Sociogenetics Structure of the Chilean Population. En "Genetic Epidemiology of Diabetes in Chile. 1998. INTA, Universidad de Chile.

Vieira AR, Karras JC, Orioli IM, Castilla EE, Murray JC. Genetic origins in a South American clefting population. *Clin Genet.* 2002 Dec;62(6):458-63.

Wang, Yi, James Lu, Jin Yu, Richard A. Gibbs, and Fuli Yu. 2013. "An Integrative Variant Analysis Pipeline for Accurate Genotype/haplotype Inference in Population NGS Data." *Genome Research*, January, gr.146084.112. doi:10.1101/gr.146084.112

**ANEXO 1**  
**FICHA DE CARACTERIZACIÓN DEL PACIENTE**  
**PROYECTO: "GENÓMICA DE LA POBLACIÓN CHILENA: OBTENCIÓN DE PERFILES**  
**GENÉTICOS NECESARIOS EN INVESTIGACIÓN CLÍNICA, SALUD PÚBLICA Y MEDICINA**  
**FORENSE"**

<b>Correlativo individual</b>		<b>Fecha de aplicación</b>	/ /
-------------------------------	--	----------------------------	-----

DP	Datos personales [LLENE LA INFORMACIÓN O CIRCULE RESPUESTA CUANDO CORRESPONDA].		
DP1 Edad	[en años]		
DP2 Año de nacimiento			
<b>Si tiene menos de 18 años a más de 50 años, aquí termina aplicación de cuestionario.</b>			
DP3 Nombre del paciente			
DP4 Lugar de toma de la muestra			
DP5 Sexo	a. Masculino	b. Femenino	
DP6 Grupo sanguíneo ABO	a. AB	b. A	c. B d. O
DP7 Grupo sanguíneo Rh	a. Positivo	b. Negativo	
DP8 Lugar de nacimiento	a. ANOTE COMUNA Y REGIÓN	b. NS/NR [NO LEER]	
DP9 Lugar de residencia	a. ANOTE COMUNA Y REGIÓN	b. NS/NR [NO LEER]	
DP10 Primer apellido de su padre	a.	b. NS/NR [NO LEER]	
DP11 Segundo apellido de su padre	a.	b. NS/NR [NO LEER]	
DP12 Lugar de nacimiento de su padre	a. ANOTE COMUNA Y REGIÓN	b. NS/NR [NO LEER]	
DP13 Primer apellido de su madre	a.	b. NS/NR [NO LEER]	
DP14 Segundo apellido de su madre	a.	b. NS/NR [NO LEER]	
DP15. Lugar de nacimiento de su madre	a. ANOTE COMUNA Y REGION	b. NS/NR [NO LEER]	
DP16. Lugar de nacimiento de su ABUELO paterno	a. ANOTE REGION Y PAIS	b. NS/NR [NO LEER]	
DP17. Lugar de nacimiento de su ABUELA paterna	a. ANOTE REGION Y PAIS	b. NS/NR [NO LEER]	
DP18. Lugar de nacimiento de su ABUELO materno	a. ANOTE REGION Y PAIS	b. NS/NR [NO LEER]	
DP19. Lugar de nacimiento de su ABUELA materna	a. ANOTE REGION Y PAIS	b. NS/NR [NO LEER]	

P1	En Chile se reconoce la existencia de pueblos originarios indígenas ¿Es descendiente o tiene antepasados de alguno de ellos?  [LEER ALTERNATIVAS Y CIRCULE RESPUESTA].	a. Acalufe (kawésqar)	1
		b. Atacameño	2
		c. Aymara	3
		d. Colla	4
		e. Mapuche	5
		f. Quechua	6
		g. Rapa Nui	7
		h. Yámana (Yagán)	8
		i. Otros [cual] _____	9
		j. NS/NR [NO LEER]	99

P2	En Chile se reconoce la existencia de pueblos originarios indígenas  P1b. ¿Pertenece Ud. a alguno de ellos? [LEER ALTERNATIVAS Y CIRCULE RESPUESTA].	a. Acalufe (kawésqar)	1
		b. Atacameño	2
		c. Aymara	3
		d. Colla	4
		e. Mapuche	5
		f. Quechua	6
		g. Rapa Nui	7
		h. Yámana (Yagán)	8
		i. Otros [cual] _____	9
		j. NS/NR [NO LEER]	99

P3	¿A qué sistema de salud tiene acceso?  [LEER ALTERNATIVAS y CIRCULE RESPUESTA]	a. Indigente	1	Ir a P4
		b. FONASA	2	
		c. Privado	3	
		d. Otra	6	
		e. NS/NR [NO LEER]	9	

P4	Si tiene acceso a FONASA, indique el nivel.  [LEER ALTERNATIVAS y CIRCULE RESPUESTA]	a. Nivel A	1
		b. Nivel B	2
		c. Nivel C	3
		d. Nivel D	4
		e. NS/NR	9

P5	¿Cuáles de los siguientes bienes o servicios tiene actualmente en funcionamiento en su hogar? [CONSULTE UNO A UNO Y CIRCULE RESPUESTA].					
		Si	No		Si	No
	a. Refrigerador	1	2	h. TV Cable o Satelital (pagado)	1	2
	b. Lavadora automática	1	2	i. Conexión a Internet (pagada)	1	2
	c. Videograbador o DVD	1	2	j. Cámara de video	1	2
	d. Horno microondas	1	2	k. Calefont o similar/ Agua caliente	1	2
	e. Computador de escritorio (PC)	1	2	l. Teléfono celular con contrato	1	2
	f. Computador portátil (Notebook laptop)	1	2	m. Automóvil (uso particular)	1	2
	g. Teléfono fijo	1	2	n. Televisor a color	1	2

P6	¿Cuál es el nivel educacional formal del principal sostenedor del hogar (es decir la persona que aporta los mayores ingresos)- en el cual Ud. vive? – Ud. puede ser el principal sostenedor--[CIRCULE RESPUESTA].
----	---

Tipo de enseñanza	Nivel	Tipo de enseñanza	Nivel
a. Sin estudios	1	f. Técnica incompleta	6
b. Básica incompleta	2	g. Técnica completa / Universitaria incompleta	7
c. Básica completa	3	h. Universitaria completa	8
d. Media incompleta	4	i. NS/NR [NO LEER]	99
e. Media completa	5		

<b>P7</b>	¿En qué tipo de establecimiento el principal sostenedor del hogar en el cual Ud. vive, realizó su enseñanza escolar? <b>[CIRCULE RESPUESTA].</b>	
	a. Municipal, pública	1
	b. Subvencionada	2
	c. Privada	3
	d. NS/NR [NO LEER]	9

<b>P8</b>	En su actual ocupación, el principal sostenedor del hogar en el cual Ud. vive, ¿trabaja como? <b>[CIRCULE RESPUESTA].</b>	Empleado u obrero en empresa privada	1
		Empleado u obrero del sector público (incluso empresa pública o municipalidad)	2
		Miembro de FFAA y de Orden	3
		Patrón/a o empleador/a (contrata o paga a honorarios a uno/a o más trabajadores/as)	4
		Trabaja solo; no tiene empleados/as	5
		Familiar no remunerado	6
		Servicio doméstico	7
		No trabaja	8
		NS/NR [NO LEER]	9

<b>P9</b>	El principal sostenedor del hogar en el cual Ud. vive, en su trabajo, ¿tiene personas a cargo o es jefe de otras personas? <b>[CIRCULE RESPUESTA].</b>	Sí	1	<b>Ir a P10</b>
		No	2	
		NS/NR [NO LEER]	9	

<b>P10</b>	El principal sostenedor del hogar en el cual Ud. vive ¿A cuántas personas manda o tiene a su cargo en su trabajo? <b>[CIRCULE RESPUESTA].</b>	Menos de 10 personas	1
		Más de 10 y menos de 50 personas	2
		Más de 50 y menos de 100 personas	3
		Más de 100 personas	4
		NS/NR [NO LEER]	9

<b>P11</b>	¿Cuál es la ocupación u oficio actual del principal sostenedor del hogar en el cual Ud. vive? Describa sus principales tareas y funciones en el puesto de trabajo actual. <b>Ej. Mecánico automotriz en taller de reparación; empleado de contabilidad; vendedor en mostrador de comercio; cajera en supermercado, etc.</b> <b>[ANOTAR LO QUE CONTESTE EL/LA ENTREVISTADO/A Y PEDIR QUE PROFUNDICE SI LA RESPUESTA ES MUY GENERAL].</b>	
------------	---	--

**Con esta pregunta terminamos la entrevista, muchas gracias por la información que compartió con nosotros.  
Su ayuda es muy importante para nuestro trabajo.**

**AGRADEZCA EN NOMBRE PROPIO Y DEL PROYECTO**

1. OBSERVACIONES DEL ENCUESTADOR :			
1	[SI ENCUESTADO/A REHUSÓ DECIR SU EDAD]: ¿Qué edad le estima Ud.?	EDAD	
2	¿Qué nivel de confiabilidad y sinceridad le atribuye al encuestado/a? [CIRCULE RESPUESTA]	Totalmente confiable Relativamente confiable Poco confiable Nada confiable	1 2 3 4
3	<p>a. PARA EL(LA) ENCUESTADOR(A): <b>¿Tiene algún comentario o información adicional que no fue registrada en la encuesta?</b> [Anote algún aspecto relevante que no ha sido abordado en la encuesta].</p> <p>Comentario</p>		

## ANEXO 2

### LISTADO DE SNPs INCLUIDOS EN EL PANEL ACOTADO

cromosoma	posición	rsID	ref	alt
chr10	108987492	rs4918288	A	G
chr10	118171016	rs7087634	C	T
chr10	24347747	rs28831093	A	C
chr10	26101206	rs17665075	C	T
chr10	31414304	rs867768	C	T
chr10	56881122	rs322283	G	C
chr10	65892071	rs10761835	C	T
chr10	88128281	rs3814614	A	G
chr11	123261385	rs7927064	C	T
chr11	131106301	rs1793568	G	T
chr1	116949306	rs6660743	A	G
chr11	19188122	rs2242458	A	G
chr1	1249187	rs12142199	A	G
chr11	39889167	rs4755530	A	G
chr1	153885261	rs4845584	A	C
chr1	159174683	rs2814778	C	T
chr11	64532579	rs523200	A	C
chr11	80086923	rs2448265	A	G
chr1	185740656	rs17505819	C	T
chr11	94591289	rs4309121	C	T
chr12	109427241	rs381459	A	G
chr12	113916920	rs16942836	A	G
chr12	117291957	rs903770	A	G
chr1	212960866	rs3738800	C	T
chr12	15801799	rs4237954	A	C
chr1	231987419	rs118023864	A	G
chr1	234358216	rs622815	C	T
chr12	39068198	rs10875961	G	T
chr12	45394863	rs7315173	A	G
chr12	66250940	rs343092	G	T
chr1	27931698	rs4908343	A	G
chr12	96328422	rs12229055	A	G
chr13	105381134	rs1572510	C	T
chr13	112889456	rs2255084	G	T
chr13	28847135	rs9319424	C	T
chr13	34991773	rs388294	C	G
chr13	49067103	rs7325962	C	T
chr13	77289977	rs2149698	A	G
chr13	77336326	rs17066642	C	T
chr14	32328112	rs8009296	C	T
chr14	48252483	rs8020002	C	T
chr14	75677699	rs7148230	G	T
chr14	95062298	rs2402480	C	T
chr14	99609217	rs1273225	C	T
chr15	24196452	rs12906805	A	G

chr1	53958603	rs1298637	C	T
chr15	45167563	rs77145924	C	T
chr15	48426484	rs1426654	A	G
chr15	71139706	rs11635165	C	T
chr16	10989754	rs45601437	A	G
chr16	19803846	rs8050872	C	T
chr16	53806145	rs2058908	C	T
chr16	75820543	rs7202210	C	T
chr17	13124790	rs78223898	C	T
chr17	35219205	rs2270662	C	T
chr17	58508618	rs3760332	C	T
chr17	8048283	rs2735611	A	G
chr18	19683998	rs7234530	A	G
chr18	35762424	rs4799476	C	T
chr1	84293113	rs12135529	C	T
chr18	67683698	rs4891800	C	T
chr19	12188317	rs10424345	A	G
chr19	24376384	rs76813435	A	G
chr19	45451647	rs7257095	C	G
chr1	96236784	rs10874946	G	T
chr20	2060151	rs1418029	A	G
chr20	34074831	rs224371	A	G
chr20	62181508	rs310612	A	G
chr2	100134353	rs78509428	C	T
chr2	109567445	rs260699	A	G
chr21	17611773	rs2823676	C	T
chr2	133676214	rs1036543	A	G
chr21	37873237	rs11700868	C	T
chr21	39202275	rs117930146	G	T
chr2	167016102	rs10497281	A	G
chr2	202975501	rs58321030	A	G
chr2	206278900	rs849263	A	G
chr22	32908738	rs1056484	C	T
chr2	2372168	rs6759202	C	T
chr2	239417840	rs28497373	A	G
chr22	44754963	rs11705079	A	G
chr22	46786315	rs4044210	C	T
chr2	39611382	rs7596222	C	T
chr2	71194556	rs72897942	A	C
chr2	8118123	rs13021734	A	G
chr3	123941100	rs1920623	A	C
chr3	129401505	rs6780938	A	G
chr3	153258272	rs6764190	A	G
chr3	185964708	rs3774061	C	T
chr3	2134399	rs74895924	A	G
chr3	31789582	rs2290532	C	T
chr3	53243913	rs35416537	A	G
chr3	6042658	rs1517378	C	T
chr3	64514393	rs7631391	A	G
chr3	96472739	rs937878	C	T

chr4	122633535	rs4833757	A	G
chr4	13483049	rs3822225	A	G
chr4	151531352	rs4577554	C	T
chr4	165251843	rs12504267	G	T
chr4	177844568	rs2574904	C	T
chr4	21552230	rs1380815	A	G
chr4	31841506	rn131966	C	T
chr4	41833487	rs4623048	C	T
chr4	63728612	rs1532948	C	T
chr4	91527074	rs12502954	A	G
chr4	97906295	rs6834049	A	G
chr5	110106634	rs244430	C	T
chr5	126009267	rs115969489	C	T
chr5	142142348	rs7736578	A	G
chr5	168522312	rs11134558	A	G
chr5	175158653	rs6875659	A	G
chr5	2518738	rs378257	C	T
chr5	33951116	rs35397	G	T
chr5	60368781	rs17529085	G	T
chr5	72636566	rs2972201	A	G
chr6	102323696	rs4120910	C	T
chr6	105883147	rs9486092	A	G
chr6	125961396	rs9401838	C	T
chr6	150303241	rs4869782	A	G
chr6	167352104	rs3777722	A	C
chr6	3285286	rs10793841	A	G
chr6	34927280	rs1535001	A	G
chr6	68930738	rs9445980	A	G
chr7	119694410	rs10271592	C	T
chr7	146413497	rs344470	C	T
chr7	20636239	rs1858940	C	T
chr7	31137003	rs2267740	C	T
chr7	61540368	rn243926	T	C
chr7	6255327	rs6463531	C	T
chr7	64865118	rs10226579	A	G
chr7	90747682	rs10488003	C	T
chr7	98887802	rs10953286	C	T
chr8	122220756	rn286585	G	T
chr8	128279007	rs4871779	A	G
chr8	259934	rs11990310	C	T
chr8	29383034	rs7012981	C	T
chr8	3545212	rs75644136	C	T
chr8	59453461	rs12545426	G	T
chr8	88852543	rs6995710	C	T
chr8	91897549	rs1352159	C	T
chr9	108316653	rs2768288	A	G
chr9	109695950	rs1510283	G	T
chr9	133343659	rs590086	C	T
chr9	22667225	rs1020463	C	T
chr9	24883179	rs7873820	A	C



chr9	499516	rs10974844	A	G
chr9	83932430	rs7853515	A	G