



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

I. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2015eko maiatzaren 13, 14 eta 15
Durango, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

OSASUNA

**Gaixotasun zeliakoarekin lotutako
eskualde genomiko berrien
bilaketa jatorri genetikoan
oinarritua**

*K. Garcia-Etxebarria,
L. Plaza-Izurieta eta J. R. Bilbao*

762-767 or.
<https://dx.doi.org/10.26876/ikergazte.i.106>

ANTOLATZAILEA:



BABESLEAK:



LAGUNTZAILEAK:



Gaixotasun zeliakoarekin lotutako eskualde genomiko berrien bilaketa jatorri genetikoan oinarritua

Garcia-Etxebarria K, Plaza-Izurietta L eta Bilbao JR

*Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila, BioCruces Osasun Ikerketako
Institutua, Euskal Herriko Unibertsitatea-UPV/EHU, Leioa, Bizkaia
koldo.garcia@ehu.eus*

Laburpena

Gaixotasun zeliakoak oinarri genetiko konplexua du. Gaixotasunean eragin dezaketen hainbat eskualde genomiko identifikatu badira ere, hauek populazioz populazio alda daitezke. Hortaz azpipopulazioen banaketan oinarritutako hurbilketa proposatzen dugu eta hau Estatu Espainoleko hainbat indibiduo aztertze erabili dugu pilotu modura. Indibiduoak beren jatorri genetikoaren arabera sailkatuta asoziazio-analisiak egiten badira, azpipopulazio bakoitzean agertzen diren gaixotasunarekin lotutako eskualde genomikoak ezberdinak dira eta orain arte gaixotasun zeliakorekin ez lotuak. Azkenik, etorkizunean, hurbilketa hau Europa osoko indibiduoetan erabiltzeko asmoa daukagu eskualde genomiko berriak aurkitzeko.

Hitz gakoak: gaixotasun zeliakoa, asoziazio-analisisa, estratifikazioa

Abstract

Celiac disease has a complex genetic basis. Although various genomic regions that could be related to the disease have been identified, these could vary among populations. Thus, we propose an approach based on the separation of subpopulations and we test it as pilot to analyze individuals from Spain. When association analyses are performed after individuals are classified by their genetic origin, the genomic regions that are related to the disease are different in each population and not previously related to the celiac disease. Finally, in the future, we will use this approach in a large European sample to find new genomic regions.

Keywords: celiac disease, association analysis, stratification

1. Sarrera eta motibazioa

Gaixotasun zeliakoa gaixotasun autoimmune bat da zeinean hainbat zerealek duten gliadinaren aurrean hesteko zelulen aurka erantzun immunea gertatzen den. Gaixotasun kroniko hau Europar jatorriko populazioan nahiko komuna da eta oinarri genetiko konplexua du., 6. kromosoman kokatzen den HLA eskualdeak gaixotasunaren osagai genetikoaren %40a azaltzen duenez gaixotasun zeliakoarekin lotuta egon daitezkeen beste eskualde genomiko aurkitzeko interes handia dago. Beste eskualde horiek detektatzeko *ImmunoChip* deitutako lanabesa erabili izan da. Lanabes honen bidez gaixotasun autoimmuneekin lotuta dauden eskualdeen azterketa sakona egiten da eskualde horretako SNP kopuru handia aztertzen baitira (Trynka et al., 2011; Fernandez-Jimenez et al., 2014). SNPak (nukleotido bakarreko polimorfismo) gaixotasun genetikoak aztertze erabili ohi diren markatzaile genetikoak dira, aztertze errezak eta informatiboak baitira. Behin indibiduo bakoitzaren SNPak zehaztu ostean, SNP hauen maiztasuna gaixoetan eta ez-gaixoetan konparatuta, gaixotasuna edukitzeko arriskua kalkulatu da asoziazio-analisisa deritzon lanabes estatistikoaren bidez. *ImmunoChip* asoziazio-analisiak interesgarriak izan daitezkeen eskualde genomikoetan egitea ahalbidetzen ditu.

Esan bezala, *ImmunoChip*a gaixotasun zeliakoa aztertze erabili da eta gaixotasunarekin 40 eskualde genomiko erlazionatuta egon daitezkeela ondorioztatu da (Trynka et al., 2011). Analisi horretan, gaitasun estatistikoa handitzeko, jatorri ezberdinetako indibiduoak nahastu ziren, baina populazio bakoitzeko emaitzek ez dute zertan etorri behar emaitza orokorrarekin. Hortaz, aztertutako indibiduoek azpipopulazioak eratu ditzazketenez, azpipulazioa hauek aztertu nahi ditugu gaixotasun zeliakoarekin lotuta egon daitezkeen eskualde berriak bilatzeko. Orain arte, azpoipoluzioetako banaketa indibiduoaren jatorri etno-geografikoaren arabera egin izan da; gure

proposamena ordea, ikerketa honetan bertan aztertutako SNPein eraikitako profil genetikoaren arabera sailkatu ditugu, aurreiritzirik ezarri gabe.

2. Ikerketaren helburuak

Gaixotasun zeliakoa aztertzeko egindako asoziazio analisiak indibiduo guztiak batera aztertuta egin direnez interesgarriak izan daitezkeen gaixotasun zeliakoarekin lotutako eskualde genomikoak ezkutuan egon daitezke eta oraindik ezezagunak izan. Hortaz, ihes egiten zaizkigun balizko eskualde genomiko horiek aztertzeko, lehen pausu eta azterketa-pilotu bezala, Estatu Espainoleko laginak aztertu ditugu helburu hauekin:

- Aztertutako indibiduen distantzia genetikoetan oinarritutako azpipopulazioak eratzeko SNP egokienak aukeratu, hau da, informazio gehien ematen dutenak eta zarata gutxien sartzen dutenak.
- SNP horietan oinarrituta azpipopulazioak eratzeko metodo egokiena aukeratu. Horretarako bi metodo erkatuko dira: bata distantzia matematikoetan oinarritua, bestea indibiduen jatorri genetikoan oinarritua.
- Behin metodorik onena aukeratuta, azpipopulazioetan oinarritutako asoziazio analisiak burutu eta horrela gaixotasun zeliakoarekin lotuta egon daitezkeen eskualde genomiko berriak aurkitu.

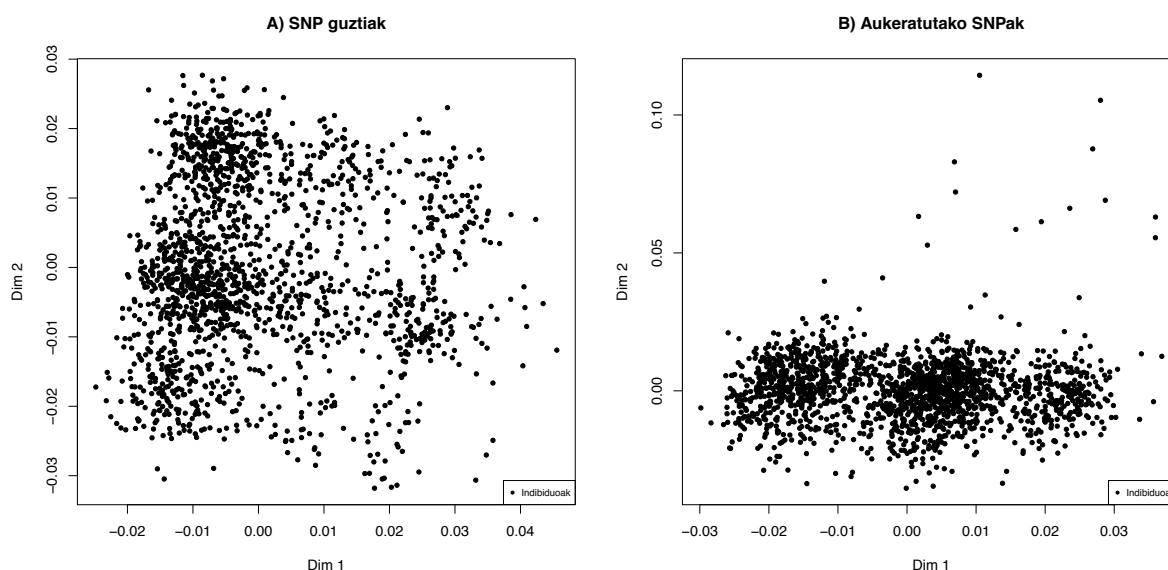
3. Ikerketaren muina

Lan honetarako Estatu Espainoleko 1863 indibiduoren (1052 gaixo, 711 ez-gaixo) 152468 SNP erabili dira aurretik publikatutako datuetan oinarrituta (Trynka et al., 2011).

3.1 Markatzaileen aukeraketa

Lan honetan erabiltzen diren datuak *ImmunoChipean* oinarrituta daudenez, lehen pausua *ImmunoChipean* presente dauden SNPetatik zeintzuk erabiliko diren hautatzea da. Pausu hau beharrezkoa da *ImmunoChipa* hainbat eskualdetan SNP asko baititu eta hauek beraien artean beti batera joan daitezkeelako bloke bat eratuz. Batera doazen SNPen bloke hauen eragina indibiduoen arteko antzekotasunaren kalkuluan saihesteko bloke hauek zeintzuk diren ebatzi ostean bloke bakoitzetik zoriz SNP bat aukeratu da.

1. irudia: Aztertutako indibiduoen banaketa erabilitako SNPen arabera. Puntu bakoitzak indibiduo bat adierazten du eta Puntuen arteko distantziak indibiduoen arteko distantzia genetikoak adierazten dute

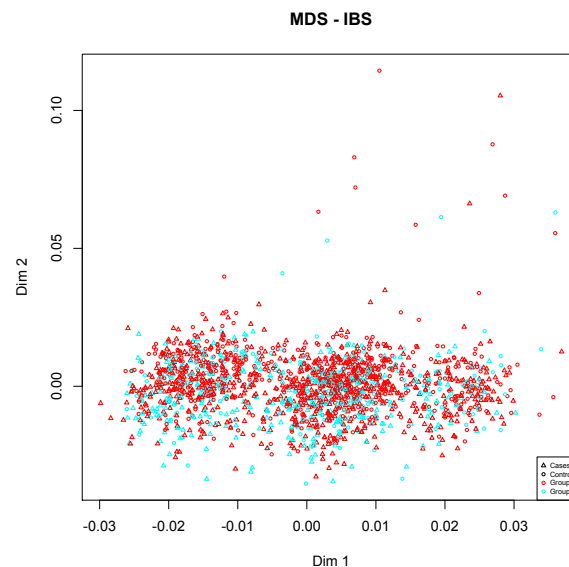


Azterketa-pilotu bezala erabili ziren indibiduoak eta erabilgarriak ziren SNPak aztertuta, lehenengo blokeak definitu ziren eta gero gaixotasunarekin erlazioa izan zezaketen SNPak kendu ziren, guztira 5761 SNPein geldituta. Analisi hau Plink (Purcell et al., 2007) programaren bidez burutu zen. Programa honekin ere aztertutako indibiduen arteko distantzia genetikoak kalkulatu zen IBS (Identity-by-state) prozedura erabilita eta distantzia hauek MDS (multidimensional scaling) bidez irudikatu ziren. SNP guztiak eta bloke bakoitzetik SNP bakarra hautatuta egindako MDSak konparatuta (1. irudia) aukeratutako SNPak erabilita talde trinkoagoak osotzen zirela ikusi zen. Hortaz, bloke bakoitzetik SNP bakarra erabiltzea lehenestea erabaki genuen.

3.2 Taldekatze metodoa

Behin bloke bakoitzetik SNP bat erabiliko dela zehaztuta indibiduoak taldekatzeko metodologia aukeratu zen. Horretarako bi hurbilketa proposatu genituen.

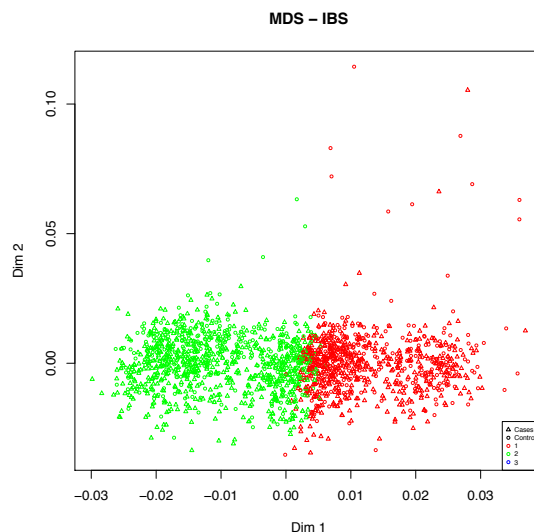
- 2. irudia: Indibiduen sailkapena PAM metodoa erabilita. Puntu bakoitzak indibiduo bat adierazten du, puntu gorriak 1 azpipopulazioan sailkatu direnak eta zelesteak 2 azpipulazioan. Puntuen arteko distantziek indibiduen arteko distantzia genetikoak adierazten dute.**



Lehenengo hurbilketa distantzietan oinarritu zen. Horretarako alde aurretik zenbat azpitalde nahi ziren definitu behar zenez, ahalik eta modu objektiboenean egiteko, R lengoaiari (R Development Core Team, 2008) eskuragarri dagoen “Cluster” paketea (Maechler et al., 2015) erabili zen. PAM metodoa erabilita azpitalde kopuru optimoa 2 zela kalkulatu zen. Hala ere, aztertutako indibiduoak 2 azpitalde horietan sailkatzerakoan, grafikoki behintzat, ez ziren bi azpitaldeak argi banatzen (2. irudia).

Bigarren hurbilketa indibiduen jatorri genetikoak erabiltzea izan zen. Zenbat jatorri erabiliko ziren zehaztu behar zenez, Admixture (Alexander et al., 2009) programa erabilita hainbat jatorri kalkulatu ziren (1 eta 15 artean) eta balidazio gurutzatuaren emaitzetan oinarrituta jatorri kopurua definitu zen. Kopuru optimoa 2 jatorri finkatzea izan zen eta indibiduo bakoitza jatorri genetiko horien arabera taldekatu zen (3. irudia). Jatorri genetikoan oinarritutako irizpide hau erabilita bi taldeak nahiko argi banatzen ziren eta, hortaz, lan honetan proposatzen diren helburuak betetzeko behintzat, hurbilketa hau aproposena zirudien.

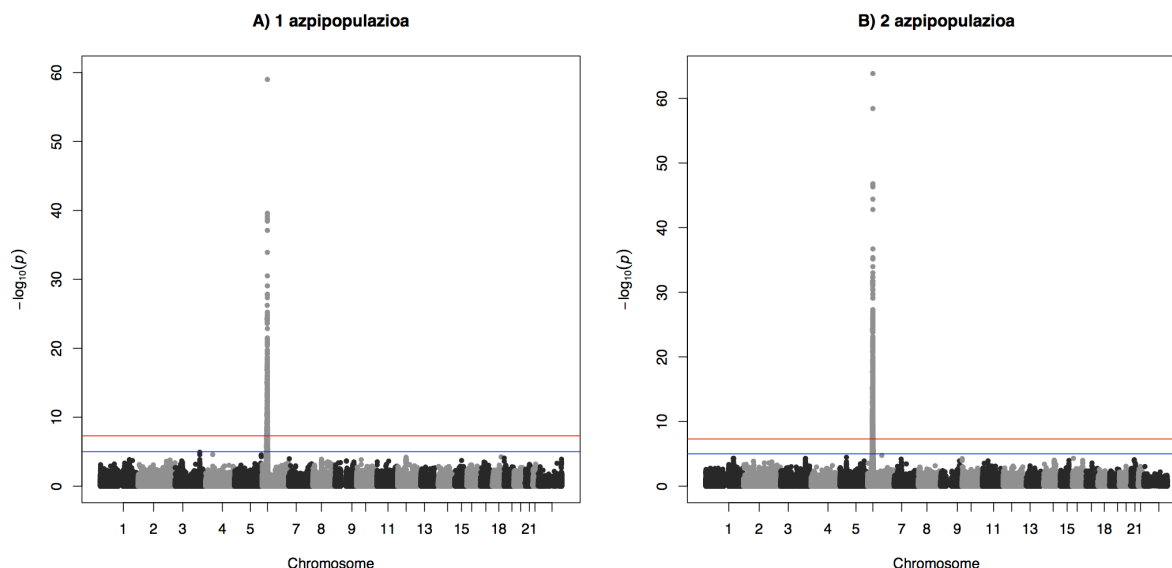
3. irudia: Indibiduen sailkapena bere jatorria erabilita. Puntu bakoitzak indibiduo bat adierazten du, puntu gorriak 1 azpipopulazioan sailkatu direnak eta berdeak 2 azpipopulazioan. Puntuen arteko distantziek indibiduen arteko distantzia genetiko adierazten dute.



3.3 Azpipopulazioen asoziazio-analisiak

Indibiduoak jatorrian oinarritutako bi azpipopulaziotan banatu ostean gaixotasunarekiko asoziazio analisia burutu zen bakoitzean (4. irudia). Itxaron zitekeen moduan, 6. kromosoman kokatu ziren seinalerik esangarrienak baina eremu horretatik kanpo, azpipopulazio bakoitzean gaixotasun zeliakoarekin lotuta egon zitezkeen eremu genomiko desberdinak identifikatu ziren.

4. irudia: Asoziazio analisiak azpipopulazio bakoitzean. Puntua gero eta gorago SNP horren lotura gaixotasunaekin esangarriagoa da. Marra gorriak Bonferroni zuzenketa estatistikoaren maila adieraztu du; urdinak konbentzioz erabiltzen den $p < 10^{-5}$ muga.



1 azpipopulazioan (1. taula) 6 eremu ezberdinetako 9 SNP esangarri aurkitu ziren eta 2 azpipopulazioan (2. taula), ordea, 9 eremu ezberdinetako 17 SNP. Gainera eskualde hauek ez zien gainjarri, hots, azpipopulazio bakoitzean gaixotasun zeliakoan ere eragin dezaketen eskualde genomikoak ezberdinak izan ziren. Salbuespen bakarra 3. kromosomako eskualde bat izan zen, 1 azpipopulazioan mutur bat esangarria bait zen eta 2 azpipopulazioan beste muturra.

Bestalde, aipatutako 3. kromosomako eskualdea izan ezik, ez ziren mota honetako indibiduo gehiagorekin egindako lanean aurkitu eta lan horretan gaixotasun zeliakoarekin aurkitu ziren eskualde genomikoak ez ziren hurbilketaekin detektatu (Trynka et al., 2011).

1. taula: 1 azpipopulazioko 6. kromosomatik kanpo dauden SNP esangarriak.

SNP	Chr	Pos	Genea
imm_3_161095192	chr3	159612498	SCHIP1
imm_3_161103127	chr3	159620433	
rs12331086	chr4	47281414	GABRB1
lkg_5_173269741	chr5	173337135	CPEB4
lkg_5_173330278	chr5	173397672	
lkg_5_173348013	chr5	173415407	C5orf47
imm_12_54675560	chr12	56389293	ANKRD33
rs2156336	chr18	58117967	
rs6510868	chr19	655608	RNF126

Esangarriak ziren SNP hauen kokapenari dagokiola, 1 azpipopulazioko SNPak 6 gene ezberdinetan kokatu ziren eta 3 SNP gene arteko eskualdeetan (1. taula); 2 azpipopulazioaren kasuan, 6 gene ezberdinetan eta 6 SNP gene arteko eskualdeetan (2. taula). Azpipopulazio bakoitzean aurkitu ziren geneak ez dira orain arte gaixotasun zeliakoarekin lotu eta gene arteko eskualdeetan kokatzen ziren SNPen funtzioa zehazteko dago.

2. taula : 2 azpipopulazioko 6. kromosomatik kanpo dauden SNP esangarriak.

SNP	Chr	Pos	Genea
imm_1_181641341	chr1	183374718	SRGAP2 / NMNAT2
imm_1_181643509	chr1	183376886	SRGAP2 / NMNAT2
imm_1_181645752	chr1	183379129	SRGAP2 / NMNAT2
imm_1_181646901	chr1	183380278	SRGAP2 / NMNAT2
imm_1_181652935	chr1	183386312	SRGAP2 / NMNAT2
imm_1_181653795	chr1	183387172	SRGAP2 / NMNAT2
rs35100811	chr3	161084046	C3orf57
rs2404341	chr3	161092405	
rs2404342	chr3	161092562	
rs11710489	chr3	161093332	
ccc-5-40593683-C-T	chr5	40557926	
rs1525235	chr7	46875785	POLR2J2 / DQ599799
vh_10_1058639	chr10	1068639	IDI2 / C10orf110
lkg_14_87673873	chr14	88604120	
rs11077049	chr16	6604090	A2BP1
rs2070370	chr21	36157698	
rs2268278	chr21	36184939	AML1/AMP19 fusion RUNX1

4. Ondorioak

Lan-pilotu honen emaitzen harira:

- Azpipopulazioak eratzeko markatzaile egokienak independenteki banatzen diren SNP bana erabiltzea da, horrek analisiari bereizmena eta argitasuna eman dio.
- Azpipopulazioak eratzeko metodo egokiena jatorriaren arabera sailkatzea da. Horrela eginez gero eratzen diren taldeak argiagoak dira eta biologikoki zentzua daukate.
- Azpipopulazio bakoitzean gaixotasun zeliakoarekin lotuta egon daitezkeen eskualde genomikoak ezberdinak dira eta hautagai berriak bilatzeko erabilgarria. Honekin gaixotasuna aztertze bide berriak irekitzea espero dugu.

5. Etorkizunerako planteatzeko den norabidea

Gaixotasun zeliakoarekin lotutako eskualde genomiko berriak bilatzeko indibiduoak jatorriaren arabera sailkatzea hurbilketa aproposa izan daitekeela ikusita, metodologia hau lagin kopuru handiago batekin erabiliko dugu. Horretarako lehenago aipatutako *ImmunoChip*arekin egindako gaixotasun zeliakoaren azterketan (Trynka et al., 2011) erabilitako datu guztiekin hurbilketa hau erabiliko dugu. Behin emaitzak lortuta gaixotasun zeliakoarekin lotutako eskualde genomiko berriak proposatzea espero dugu eta hauek gene arteko eskualdeak badira, karakterizazio sakonago bat egin.

6. Erreferentziak

- Alexander, D.H., Novembre, J. eta Lange K. (2009): Fast model-based estimation of ancestry in unrelated individuals, *Genome Research*, 19, 1655–1664.
- Fernandez-Jimenez, N., Plaza-Izurieta L., eta Bilbao J. R. (2014): Genetic Markers in Celiac Disease, in L. Rodrigo eta A. S. Peña (arg.), *Celiac Disease and Non-Celiac Gluten Sensitivity*, OmniaScience, Barcelona, 103-121.
- Maechler, M et al. (2015): cluster: Cluster Analysis Basics and Extensions. R package version 2.0.1.
- Purcell, S. et al. (2007): PLINK: a toolset for whole-genome association and population-based linkage analysis, *American Journal of Human Genetics*, 81, 559-575.
- R Development Core Team (2008): R: A language and environment for statistical computing. Vienna, Austria.
- Trynka, G. et al. (2011): Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease, *Nature genetics*, 43, 1193-1201.

7. Eskerrak eta oharrak

Trynka, G et al (2011) artikuluan parte hartu zuten ikertzaile eta kolaboratzaile guztieei eskerrak eman nahi dizkiegu.

Lan honek *Wellcome Trust Case- Control Consortium*ek sortutako datuak erabiltzen ditu. Datuen sortzean lagundu duten ikertzaileen zerrenda osoa www.wtccc.org.uk helbidean eskuragarri dago. Egitasmoarentzako diru-funtsa *Wellcome Trust*ek emana da 076113, 085475 eta 090355 sarietan.

This study makes use of data generated by the Wellcome Trust Case- Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113, 085475 and 090355.