



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

ZIENTZIA
ETA TEKNOLOGIA
FAKULTATEA
FACULTAD
DE CIENCIA
Y TECNOLOGÍA



Gradu Amaierako Lana
Biologiako Gradua

Aldakortasun genetikoa giza populazioetan: hautespenaren eragina 22. kromosomako geneetan

Egilea:
Aitor Zabala Garcia
Zuzendaria:
Mikel Iriondo Orensanz

1.1 LABURPENA	2
1.2 ABSTRACT	2
2. SARRERA	3
3. HELBURUA	4
4. METODOLOGIA	5
4.1- Datu-basearen eraketa	5
4.2- Hautespenaren eragina izan duten SNPen identifikazioa	7
4.3- Hautespenaren eragina izan duten geneen identifikazioa	7
4.4- APOBEC3G eta APOBEC3H geneen hautespenaren azterketa	7
5. EMAITZAK	8
6. EZTABAIDA	14
6.1- APOBEC3G eta APOBEC3H geneak:	14
6.2- Erretrobirusekiko harremana: GIB birusa	15
6.3- APOBEC3G eta APOBEC3H geneetako hautespena:	16
6.4- APOBEC3G eta APOBEC3H geneen hautespenaren zergatia	17
7. ONDORIOAK	18
8. BIBLIOGRAFIA	19
9. ERANSKINAK	24
1. ERANSKINA	25
2. ERANSKINA	27
3. ERANSKINA	34

1.1 LABURPENA

Giza aldakortasuna populazio baten eta populazioen arteko genomak bereizten dituzten desberdintasun genetikoaren multzoa da. Giza aldakortasuna populazio arbasoen historiaren erakusle da, eta populazioek jasan duten hautespen naturala aztertzea ahalbidetzen du. Lan honen helburua izan da giza populazio desberdinen 22. kromosomaren aldakortasuna aztertzea, eta hautespen naturalaren eragina jasan duten 22. kromosomaren geneak identifikatzea eta aztertzea. Horretarako, 1000 Genomen datu-basean *Minor Allele Frequency* (MAF) 0,4 baino altuagoa zuten 22. kromosomako *Single Nucleotide Polymorphism* (SNP) hautatu ziren, eta SNP horien *fixation index* (F_{ST}) kalkulatu ziren. Emaitzek agerian utzi zituzten hautespen naturalaren eraginpean zeuden geneak, eta populazio bakoitzak oztopoei aurre egiteko garatutako moldapenak. Aurrera begira, populazio bakoitzaren historia zehatzago ezagut dezagun, etorkizuneko ikerketek hautespenaren eraginpean dauden geneak aztertu eta hautespenaren kausak azter ditzakete. Esate baterako, APOBEC3 geneek jasan duten hautespena aztertu zen, eta ondorioztatu zen APOBEC geneek zuten aldakortasun genetikoaren eragile zela gizakia Afrikatik irten zenean topatutako erretrobirus berriak.

1.2 ABSTRACT

Human diversity differs within and between population genomes. It provides information about the ancestry and the natural selection's effect among populations. The objective of this work is to analyze the diversity of the 22nd chromosome in different populations, and to identify and analyze regions under natural selection for further research. For doing so, *Single Nucleotide Polymorphisms* (SNPs) with a *Minor Allele Frequency* (MAF) higher than 0.4 were chosen from the 1000 Genome Project and their *fixation indexes* (F_{ST}) were calculated. The results uncovered genes under natural selection and the adaptations each population developed to overcome changes. Looking ahead, future studies can analyze genes under natural selection in order to clarify reasons for that selection and more information about the populations themselves. For instance, selection in the APOBEC3 genes was analyzed in this work and retroviruses were identified (found by humans in the out of Africa event) as the main responsible for the high diversity found in those genes.

2. SARRERA

Gizakiaren lehen aztarnak duela 200.000 urtekoak dira, eta Afrikan aurkitu dira (Day, 1969; McDougall, Brown, & Fleagle, 2005). Aztarnak horiek iradokitzen dute gizakiaren (*Homo sapiens*) sorrera Afrikako hainbat lekutan gertatu zela, eta gerora Afrikatik irten eta beste kontinenteak kolonizatu zituela. Gizaki modernoa (*Homo sapiens sapiens*) duela 100.000 urte sortu zen Etiopian (Haile-Selassie, Asfaw, & White, 2004), eta duela 65.000 urte atera zen Afrikatik (Forster, 2004; Mondal et al., 2016; Thangaraj et al., 2005). Bai Afrikan bai Afrikatik kanpo, gizakia populazioetan banatzen hasi zen.

Gizakia populazioetan banatzeko prozesu horretan mekanismo mikroebolutiboek lan egin dute. Mekanismo mikroebolutiboek aleloen maiztasun aldatzen dituzte, eta lau dira: mutazioa, jito genetikoa, gene-fluxua eta hautespen naturala. Mutazioak material genetikoan gertatzen diren aldaketak dira, eta giza populazioetan eragiten duten desberdintzapena txikia da (Hartl, Clark, & Clark, 1997), espezie nahiko berria garelako. Jito genetikoak populazioak desberdintzatzen ditu, eta gene-fluxuak, aldiz, populazioak berdintzen ditu. Oro har, prozesu horiek oreka bilatzen dute, eta genoma osoan dute eragina (Kimura, 1991).

Hautespen naturalak, ordea, gene konkretuetan du eragina, eta ingurunera hoberen moldatzen den fenotipoaren genotipoa faboratzen du. Afrikatik irtetean, egoera berriekin topatu zen gizakia, eta egoera berri horietara moldatu behar izan zuen. Alde batetik, klima-egoera eta patogeno berriekin topatu zen eta, bestetik, gizakiak garatu zituen abeltzaintzak eta nekazaritzak gizakiaren egoera aldatu zuten. Aldaketa horiek guztiek giza genomaren eboluzioa bultzatu dute (Diamond, 2002).

Hautespen naturala populazio batean aztertzean, hiru hautespen natural mota daude: positiboa edo direkzionala, negatiboa edo purifikatzailea, eta orekatzailea. Hautespen natural positiboak alelo berria faboratzen du, negatiboak alelo berria kaltetzen du, eta orekatzaileak aleloen arteko oreka lortzen du (Barreiro eta al., 2008)

Populazio asko direnean, ordea, hautespen naturala beste modu batean aztertu daiteke, bi hautespen natural mota daudelarik: dibertsifikatzailea eta orekatzailea. Hautespen natural dibertsifikatzaileak populazio batean alelo bat eta beste populazioan beste aleloa faboratzen du, eta populazioak desberdintzatzen ditu (Barreiro et al., 2008).

Hautespén natural orekatzaileak, aldiz, populazioetan bi aleloak faboratuz bi aleloen arteko oreka lortzen du, eta giza populazioak berdintzen ditu. Bi prozesu horietan giza aldakortasuna sortzen da. Giza aldakortasunaren % 85 populazio barnekoa da, % 5 populazioen artekoa eta % 10 kontinenteen artekoa (Jorde & Wooding, 2004). Hortaz, kontinenteen eta populazioen artean portzentaje txikian bada ere aldakortasuna egon badago.

Giza aldakortasuna aztertzeke *fixation index* (F_{ST}) balioa erabil daiteke, zeinak populazio barneko aldakortasuna eta populazioen arteko aldakortasuna aztertzea ahalbidetzen duen. F_{ST} balio altuak agerian uzten du aldakortasuna populazioen artekoa dela, F_{ST} balio baxuak, aldiz, aldakortasuna populazio barnekoa dela (Beaumont, 2005).

Giza aldakortasuna aztertzeke markatzaile genetikoak erabiltzen dira, horien artean, *Single Nucleotide Polymorphisms* (SNPak). Giza aldakortasunaren azterketan pausu garrantzitsua izan da giza genomaren sekuentziazio osoa lortzea. 2000. urtean, giza genoma sekuentziatzea lortu zen (“A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,” 2001; Venter et al., 2001), eta horrek hainbat proiekturi bidea ireki zion. Adibide garbia da, 1000 Genoma Proiektua, zeinak munduko eskualde desberdinetako indibiduen genomen sekuentziak biltzen dituen. Genoma horiek aztertuta, giza aldakortasuna azter daiteke.

3. HELBURUA:

Lan honen helburua izan da giza populazio desberdinen 22. kromosomaren aldakortasuna aztertzea, eta hautespén naturalaren eragina jasan duten 22. kromosomaren geneak identifikatzea eta aztertzea.

1. taula: eratutako datu-basearen 26 populazioen kodea, izena, jatorria eta lagindutako indibiduo kopurua

Kodea	Populazioaren izena	Jatorria	N
ACB	Afrikar karibearrak, Barbados	Afrika	96
ASW	Afrika zaharreko amerikarrak, Estatu Batu mendebalde	Afrika	61
BEB	Bangaliak, Bangladesch	India	86
CDX	Dai txinatarrak, Xishuangbanna	Asia Ekialdea	93
CEU	Ipar eta mendebaldeko Europar zaharreko Utharrak	Europa	99
CHB	Han txinatarrak, Beijing	Asia Ekialdea	103
CHS	Han txinatarrak, Txina hego-ekialdea	Asia Ekialdea	105
CLM	Konlonbiarrak, Medellin	Amerika	94
ESN	Esanak, Nigeria	Afrika	99
FIN	Finlandiarrak, Finlandia	Europa	99
GBR	Britainiarrak, Ingalaterra eta Eskozia	Europa	91
GIH	Gujarati indiarak, Houston (Texas)	India	103
GWD	Ganbiarrak, mendebaldeko Ganbiako eskualdea	Afrika	113
IBS	Iberiar populazioa, Espainia	Europa	107
ITU	Telugu indiarak, Britainia Handia	India	102
JPT	Japoniarrak, Tokio	Asia Ekialdea	104
KHV	Kinhak, Ho Chi Minh City (Vietnam)	Asia Ekialdea	98
LWK	Luhyarak, Webuye (Kenia)	Afrika	99
MSL	Mendeak, Sierra Leona	Afrika	85
MXL	Mexikar zaharrak, Los Angeles	Amerika	64
PEL	Peruarrak, Lima	Amerika	86
PJL	Punjabiak, Lahore (Pakistan)	India	96
PUR	Puerto Rikokoak, Puerto Rico	Amerika	104
STU	Sri Lankako Tamilak, Britainia Handia	India	102
TSI	Toskanakoak, Italia	Europa	107
YRI	Yorubak, Ibadan (Nigeria)	Afrika	108

4.2- Hautespenaren eraginpean zeuden SNPen identifikazioa:

F_{ST} balioa eta hari zegokion q balioa izan ziren hautespenaren eragina aztertzeko kalkulaturako parametroak, eta parametro horiek *BayeScan v2.1* (Foll & Gaggiotti, 2008) programaren bidez kalkulatu ziren. *BayeScan* programa erabiltzeko, jatorrizko fitxategia, *Structure* formatuan zegoena, *GESTE/BayeScan* formatura itzuli zen *PGDSpider v2.1.1.5* (Lischer & Excoffier, 2012) programa erabiliz. Fitxategia behin formatu egokira itzulita, *BayeScan* programak F_{ST} balioak eta q balioak kalkulatu zituen. Azkenik, gene-hautespena harremana *R studio* (R Core Team, 2014) programaren bidez grafiko batean irudikatu zen, non x ardatzean q balioaren logaritmoak eta y ardatzean F_{ST} balioak irudikatu ziren.

4.3- Hautespenaren eraginpean zeuden geneen identifikazioa:

Hautespenaren eragina izan zuten SNPak identifikatuta, FDR zuzenketa egin zen hautespen garbia jasan zuten SNPen geneak identifikatzeko. SNP horiei zegokien geneak identifikatu ziren NCBI dbSNP datu-basearen bidez (<https://www.ncbi.nlm.nih.gov/snp/>). Genearen identifikazioa egin ostean, gene bakoitzari zegokion eta hautespenaren eragina jasan zuten SNPak zenbatu ziren, eta lortutako datuak taula batean bildu ziren. Alde batetik, F_{ST} balioa altua ($F_{ST} > 0,09$) eta q esangarria ($q < 0,05$) zuten SNPak zituzten geneen taula egin zen eta, bestetik, F_{ST} balio baxua ($F_{ST} < 0,09$) eta q esangarria ($q < 0,05$) zuten SNPak zituzten geneen taula egin zen.

4.4- Hautespenaren eraginpean zeuden geneen azterketa:

Hautespena jasan zuten SNPak hautatu ziren, eta SNP horien arteko lotura-desoreka aztertu zen *HaploView v4.2* (Barrett, Fry, Maller, & Daly, 2005) programaren bidez. *HaploView* programa erabiltzeko, jatorrizko fitxategia *PED* formatura itzuli zen eta, itzulketa hori egiteko, *PGDSpider* programa erabili zen.

HaploView programak lotura-desorekaren neurria (D') eta LOD balioa erlazionatzen ditu. LOD balioa L1 eta L0 balioen arteko zatiduraren logaritmoa da, non L1 lotura-desorekan zeuden datuen probabilitatea eta L0 lotura-orekan zeuden datuen probabilitatea diren. D' eta LOD balioen arteko erlazioa koloreen arabera irudikatzen du *HaploView* programak (2. taula).

2. taula: LOD eta D balioen erlazioaren arabera HaploView programak erabilitako koloreak

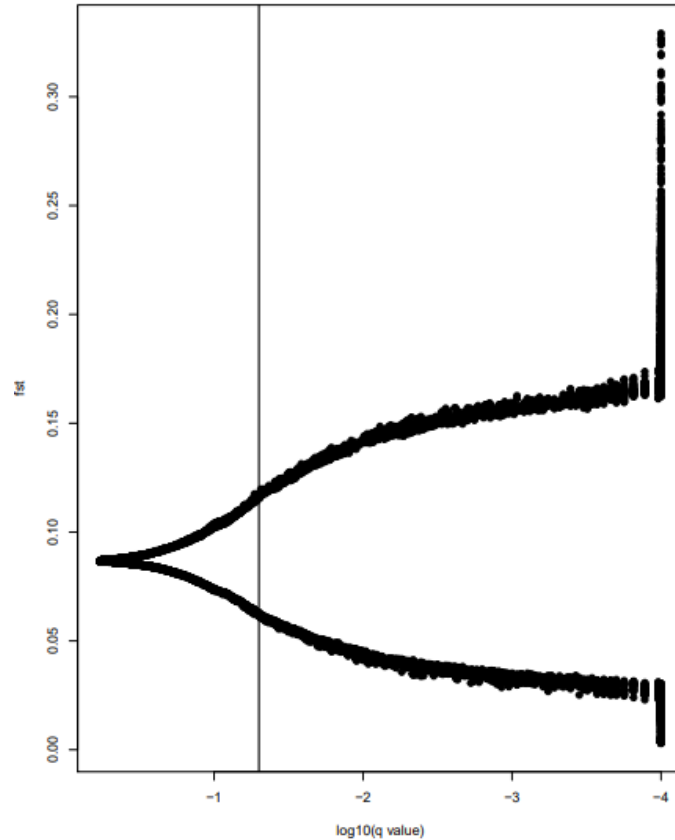
	$D' < 1$	$D' = 1$
LOD < 2	Zuria	Urdina
LOD > 2	Arrosa	Gorria

Hautespenaren eragina izan zuten SNPen populazioko maiztasunak kalkulatzeko *Genepop v4.2* (Rousset, 2008) programa erabili zen eta, horretarako, jatorrizko fitxategia *Genepop* formatura itzuli zen *PGDSpider* programaren bidez. Lortutako maiztasunen taula eranskin moduan bildu zen. Bestalde, SNPen arteko lotura-desorekatik ondorioztatu ziren SNP-taldeen aleloen populazioko maiztasunak mapa batean biltzeko, *Geography of Genetic Variants Browser* (Marcus & Novembre, 2017) programa erabili zen.

Azkenik, geneen hautespena aztertzeko, NCBI datu-basea (<https://www.ncbi.nlm.nih.gov/>) eta lortutako artikulu zientifikoek erabilitako bibliografia erabili zen.

5. EMAITZAK

22 kromosoman MAF balioa 0,4 baino handiagoa zuten 14341 SNP identifikatu ziren eta, hortaz, 35.909.864 genotipo zituen datu-basea eratu zen. 14341 SNP horietatik 5456 ziren hautespena jasan zutenak ($q < 0,05$), zeintzuk 2766-k F_{ST} altua eta 2690-k F_{ST} baxua zuten. Hautespena jasan ez zuten SNPen F_{ST} balioa 0,09 zen, eta SNP horien aleloen maiztasunen desberdintasuna jito genetikoaren eta gene-fluxuaren arteko orekak eragin zuen. Aztertutako SNPen q balioaren batezbestekoa grafikoko lerro bertikalari zegokion eta 0,050 zen (2. irudia)



2. irudia: SNP bakoitzari zegokion F_{ST} eta q balioa erlazionatzen dituen grafikoa, non x ardatzean F_{ST} balioa eta y ardatzean q balioaren logaritmoa irudikatuta dauden.

SNP bakoitzaren hautespenaren erakusle ziren parametroen FDR zuzenketa egin zen, eta zuzenketa egin ostean hautespenaren eragin jasan zuten ($q < 0,05$) 4819 SNP identifikatu ziren. Hasierako datuetatik % 88,32 izan ziren hautespen argia jasan zuten SNPak, eta 637 SNP (% 11,68) izan ziren hautespen faltsua izan zezaketen SNPak. Hautespenaren eraginpean zeuden 4819 SNPetatik F_{ST} balio altua ($F_{ST} > 0,09$) zuten 2453 SNP (% 50,90) identifikatu ziren, eta hautespen dibertsifikatzailea jasan zuten 249 gene identifikatu ziren (2. eranskina). F_{ST} baxua ($F_{ST} < 0,09$) zuten 2366 SNP (% 49,10) identifikatu ziren, eta hautespen orekatzailea jasan zuten 235 gene identifikatu ziren (3. eranskina).

Hautespeneren eraginpean ($q < 0,05$) zeuden geneetatik APOBEC3G eta APOBEC3H geneak hautatu eta sakon aztertu ziren. APOBEC3G eta APOBEC3H geneetan hautespenaren eragina zuten 21 SNP identifikatu ziren, eta 6 SNP APOBEC3G geneari eta 15 SNP APOBEC3H geneari zegokion. 21 SNP horietatik 14-k F_{ST} balio altua

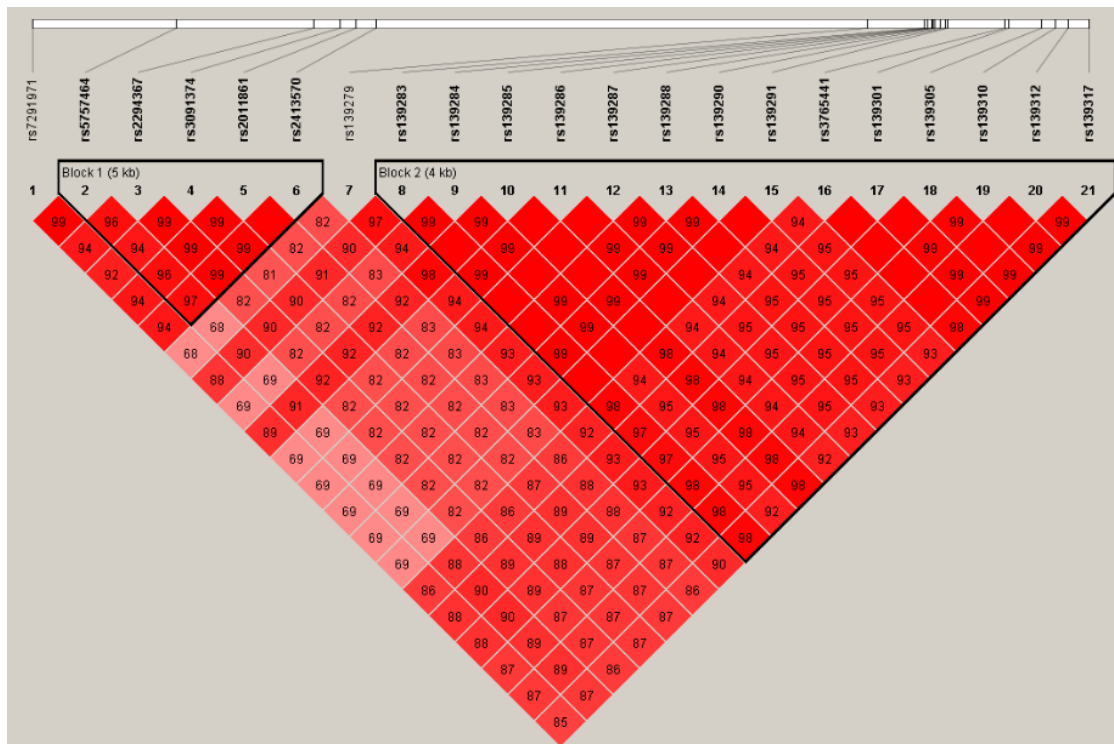
($F_{ST} > 0,09$) zuten eta 7-k F_{ST} balio baxua ($F_{ST} < 0,09$). F_{ST} balio baxua zuten SNPak APOBEC3H genean zeuden (3. taula).

3. taula: APOBEC3 geneetan hautespenaren eragina zuten SNPei buruzko informazioa, non SNPari dagokion genea, kokapena eta F_{ST} balioa biltzen diren.

SNP	Genea	Kokapena	FST
rs7291971	APOBEC3G	39472174	0,1769
rs5757464	APOBEC3G	39475964	0,1598
rs2294367	APOBEC3G	39479574	0,1559
rs3091374	APOBEC3G	39480245	0,1425
rs2011861	APOBEC3G	39480663	0,1558
rs2413570	APOBEC3G	39481187	0,1559
rs139279	APOBEC3H	39494084	0,0392
rs139283	APOBEC3H	39495584	0,1516
rs139284	APOBEC3H	39495670	0,0486
rs139285	APOBEC3H	39495780	0,1642
rs139286	APOBEC3H	39495800	0,0479
rs139287	APOBEC3H	39495870	0,0487
rs139288	APOBEC3H	39496016	0,0460
rs139290	APOBEC3H	39496134	0,0462
rs139291	APOBEC3H	39496199	0,0459
rs3765441	APOBEC3H	39497701	0,1486
rs139301	APOBEC3H	39497811	0,1652
rs139305	APOBEC3H	39498677	0,1657
rs139310	APOBEC3H	39499015	0,1550
rs139312	APOBEC3H	39499369	0,1556
rs139317	APOBEC3H	39499902	0,1323

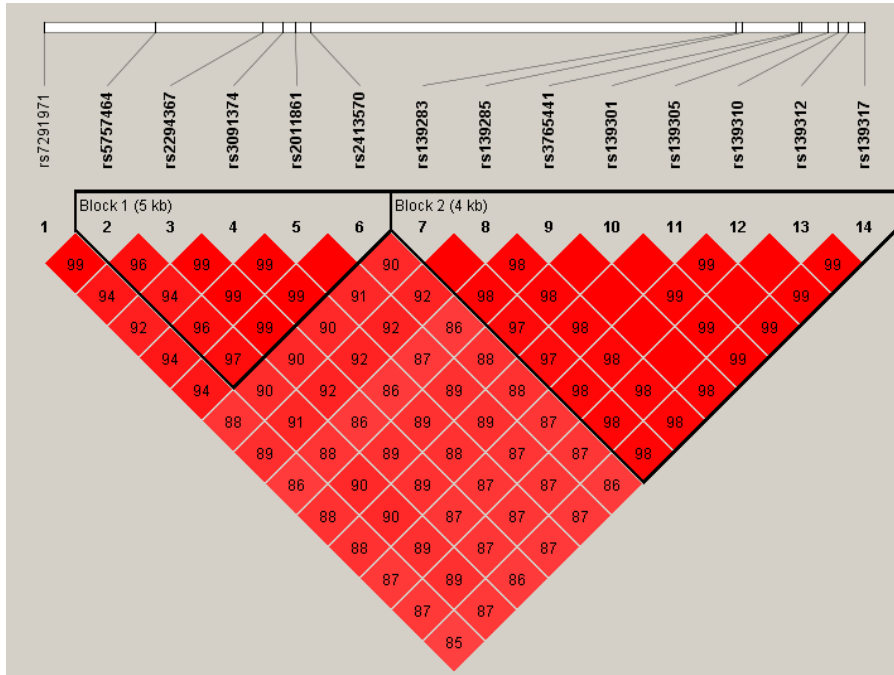
21 SNP horien arteko lotura-desoreka aztertu zen, eta ondorioztatu zen bi bloke nagusi zeudela, eta bloke horiek gene bakoitzaren SNPak biltzen zituela. Egoera horretan, lehenengo blokean APOBEC3G genearen azkeneko 6 SNPak ageri ziren, eta bigarren blokean APOBEC3H genearen azkeneko 14 SNPak. APOBEC3G eta APOBEC3H geneen lehenengo SNPak (rs7291971 eta rs139279, hurrenez hurren) bi blokeetatik kanpo geratzen ziren (3. Irudia).

APOBEC3G genean zeuden SNPek APOBEC3H genean zeuden SNP-talde batekin lotura-desoreka harreman txikia aurkezten zuen, eta beste SNP-talde batekin, aldiz, lotura-desoreka harreman handia erakusten zuen. Hain zuzen ere, F_{ST} balioa altua ($F_{ST} > 0,09$) zuen SNP-taldearekin lotura-desoreka harreman handia erakusten zuen, eta F_{ST} balio baxua ($F_{ST} < 0,09$) zuen SNP-taldearekin lotura-desoreka harreman txikia (3. irudia).



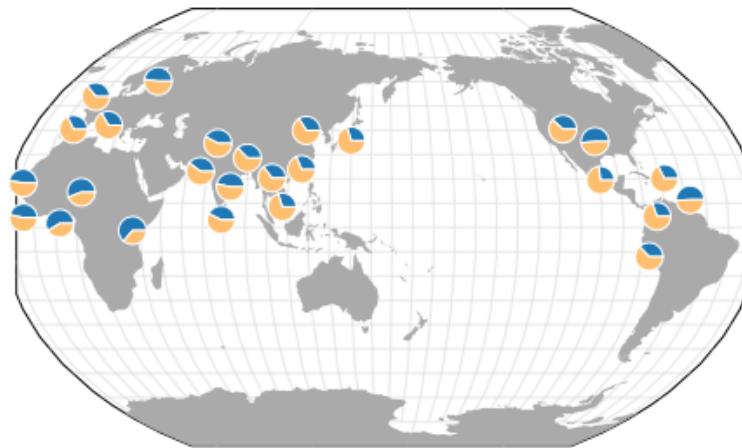
3. irudia: APOBEC3G eta APOBEC3H geneetan hautespenaren eragina izan zuten SNPen arteko lotura-desoreka, non bi bloke nagusi bereizten ziren. Bloke bakoitzean identifikatu ziren SNPak gene bakoitzari zegokion, lehenengo blokea APOBEC3G geneari eta bigarren blokea APOBEC3H geneari, hain zuzen ere.

Bestalde, F_{ST} balio altuak ($F_{ST} > 0,09$) erakusten zuten SNPen arteko lotura-desoreka harremana argiago irudikatu zen (4. Irudia), eta 3. irudian ondorioztatutako bi blokeak azaldu ziren. Lehenengo blokean, APOBEC3G genearen azkeneko 7 SNPak zeuden eta, bigarren blokean, APOBEC3H genearen 8 SNPak. Azkenik, APOBEC3G genean zegoen rs7291971 SNPa aipatutako blokeetatik kanpo ageri zen, bere geneko beste SNPekin lotura-desoreka harreman handia zuelarik.



4. irudia: APOBEC3G eta APOBEC3H geneetan hautespenaren eragina zuten eta F_{ST} balio altua ($F_{ST} > 0,09$) zuten SNPen arteko lotura-desoreka, non bi bloke nagusi bereizten ziren. Hain zuzen ere, bloke bakoitzean zeuden SNPak gene bakoitzari zegokion.

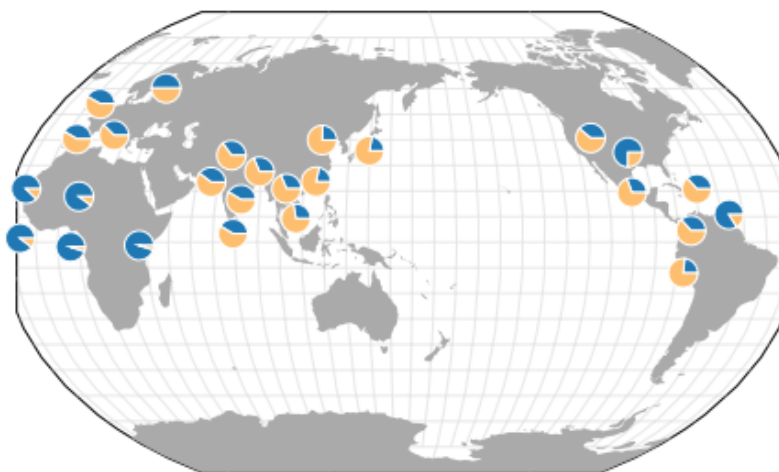
Aztertutako SNPek bi alelo zituzten, eta aleloen munduko banaketa aztertu zen. F_{ST} balio baxua ($F_{ST} < 0,05$) zuten SNPei dagokienez, ez zen populazioetan alelo nagusirik ageri (5. irudia). F_{ST} balio baxua izanda, hautespen orekatzaileak giza populazioetan bi aleloak izatea faboratu zuen eta, hala, populazioak oro har ez ziren argi desberdintzen, aleloen populazioen maiztasunak 0,6 eta 0,3 artean zeudelarik (1. eranskina).



5. irudia: rs139284 aleloen banaketa, non kolore urdina T aleloari eta laranja C aleloari dagokion

F_{ST} balio altua zuten ($F_{ST} > 0,09$) SNPei dagokienez, SNP guztien aleloen banaketa oro har bera zen: Afrikan alelo bat zen nagusi eta beste aleloak Asia Ekialdean zuen maiztasun handien. Europan, bi aleloak agertzen ziren, baina bi alelo horietatik Asia Ekialdeko aleloa da maizen agertzen zena. Bestalde, Amerikan ez zen aleloen banaketa argirik ikusi (6. irudia eta 7. irudia). Amerikan aleloen banaketa argia ez egotea espero zen emaitza zen. Izan ere, aztertu diren Amerikako populazioen jatorria ez da bera, hau da, badira Afrikarekin harreman handiagoa zuten populazioak eta Eurasiarekin harreman handiagoa zuten populazioak (1. taula). Harreman desberdintasun horrek eragin zuen aleloen banaketa Amerikako populazio guztietan bera ez izatea. Aleloen maiztasunak zehatzago ikus daitezke 1. eranskinean.

F_{ST} balio altuko ($F_{ST} > 0,09$) SNPetan ondorioztatutako bi blokeetako SNPen aleloen maiztasunak populazioetan irudikatzean, agerian utzi zen desberdintasunak desberdintasun bi SNPek populazioetan zuten maiztasunak antzekoak zirela eta, era berean, Afrikatik urrundu ahala Afrikako aleloa desagertzen zihoala argi ikusi zen (6. irudia eta 7. irudia).



6. irudia: rs5757464 aleloen banaketa, non kolore urdina G aleloari eta laranja A aleloari dagokion



7. irudia: rs139305 aleloen banaketa, non kolore urdina A aleloari eta laranja C aleloari dagokion

6. EZTABAIDA

Lan honetan hautespen naturalaren eragina giza populazioetan aztertu nahi izan da eta, horretarako, 1000 genoma datu-basetik 22. kromosomako MAF balioa 0,4 baino altuagoa zuten SNPak hautatu ziren. Hautespen naturala hainbat populazioetan aztertzean, bi hautespen natural mota daude: dibertsifikatzailea eta orekatzailea. Hautespen natural dibertsifikatzaileak gene bateko alelo bat populazio batzuetan faboratzen du, eta beste populazioetan beste aleloa; hautespen orekatzaileak, aldiz, populazioetan bi aleloak faboratzen ditu. Hautespen naturalak eragindako desberdintasunak neurtzeko F_{ST} balioak kalkulatu dira, non F_{ST} altuek ($F_{ST} > 0,09$) hautespen natural dibertsifikatzailearen eragina eta F_{ST} baxuek ($F_{ST} < 0,09$) hautespen natural orekatzailearen eragina islatzen duten. Hautespen naturalaren eragina jasan zuten geneetarik, APOBEC3G eta APOBEC3H geneak aztertu ziren.

6.1- APOBEC3G eta APOBEC3H geneak:

APOBEC3 gene-familian zazpi gene bereizten dira: APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G eta APOBEC3H. Gene horiek APOBEC familiako proteinak sintetizatzen dituzte. APOBEC proteinak zitosina desaturasa entzimak dira, eta zitosina urazilora desaminatzen dute (Jarmuz et al., 2002). Mekanismo horrek funtzio antibiralak ematen dizkio geneari (Jarmuz et al., 2002). Izan ere, erretrovirusen genomako uraziloak desaminatzean, genoma birikoa hipermutatzea eta

degradatzea eragiten du eta, ondorioz, erretrovirusak, erretrovirus endogenoak eta erretrotransposonak inhibitzen ditu (H. P. Bogerd, 2006; Hal P. Bogerd et al., 2006; Esnault et al., 2005; Esnault, Millet, Schwartz, & Heidmann, 2006)

Gizakian ez ezik, ardi latxan eta asaaf arditean ere APOBEC3 geneek erretrovirusekin duten harremana aztertu da. APOBEC3 genearen A3Z1 genean polimorfismoak aurkitu dira, eta bi SNP polimorfikoen aleloen banaketa aztertuz, bi SNP Visna/Maedi gaixotasun erretrovirikoarekin erlazionatu dira. (Esparza-Baquer et al., 2015). Harreman horrek, APOBEC3 geneek erretrovirusekin duten harremana sendotzen du.

Bestalde, erretrovirusekin ez ezik APOBEC3G eta APOBEC3H geneek birusekin duten erlazioa aztertu da, Hepatitis B birusarekin duen erlazioa, esate baterako. Ikerketek agerian utzi dute APOBEC3 geneak Hepatitis B birusaren hazkundera inhibitzen duela, eta giza populazioko APOBEC3 geneen aldakortasun genetikoaren eragileetako bat dela (Kidd, Newman, Tuzun, Kaul, & Eichler, 2007). Gauzak horrela, baliteke erretrovirusak ez ezik birusen hazkundera inhibi dezaketela eta, hartara, APOBEC3 geneek birusekin duten harremana uste baino handiagoa izatea. Hala ere, birusen eta APOBEC3 geneen arteko harremana ez da gene horiek erretrovirusekin duten harremana adinako sendoa eta, hortaz, nahiz eta birusak kontuan hartuta izan behar ditugun, ezin dira gure lanaren ardatza izan.

6.2- Erretrovirusekiko harremana: GIB birusa

Erretrovirusei buruz hitz egitean *Giza Immunoeskasiaren Birusa* (GIB) aipatu behar da derrigorrez, eta APOBEC3 geneak GIB birusarekin maiz erlazionatu dira, gainera. GIB birusak *Hartutako Immuno Eskasairen Sindromea* (HIES) eragiten du, eta gaur egun hainbati eragiten dion gaixotasuna da. HIESak immunitate-sistema ahultzen du, eta horren eraginez gaixotasun oportunistak pairatzeko probabilitatea handitzen da, tuberkulosia kasu. Bi HIES mota bereizten dira, 1. motakoa eta 2. motakoa.

GIB birusaren lehen aztarnak 1980ko hamarkadakoak dira. Ikerketen arabera, 1. motako HIE-a txinpantzeetatik gizakira 1920ko hamarkadan pasa zen gaur egun den Kongoko Errepublika Demokratikoko eskualdeetan (Gao et al., 1999; Rambaut, Robertson, Pybus, Peeters, & Holmes, 2001). Transmisio horren eragile nagusia txinpantzeen ehiza izan zen (Sharp & Hahn, 2011). Izan ere, ehizatzen ziren txinpantze horiek jan egiten ziren maiz eta, gainera, ehiza horretan sortutako zauriek txinpantzearen arteko odolarekiko kontaktua ahalbidetzen zuten. 2. motako HIESa txinpantzeetatik transmititu beharrean,

mangabey tximino grisetatik izan zen eta 1. motako HIESaren transmisio bera izan omen zuen (Chen et al., 1997).

Gizakiak birusa eskuratu zuenean, beste eskualdeetara zabaldu zen eta, gaur egun, 1. motako HIESa kontinente guztietan zabaldua dago. 2. motako HIESak, ordea, ez du infekzio-gaitasuna hain handia eta, horregatik, ez die pertsona askori eragiten eta Afrika Mendebaldeko herri batzuetan soilik aurkitu daiteke, besteak beste, Malin, Mauritania, Nigerian eta Sierra Leonan (Sharp & Hahn, 2011).

APOBEC3G eta APOBEC3H geneen funtzio antibirala ikusita, hainbat ikerlarik gene horien eta GIB birusaren arteko erlazioa aztertu dute, eta ondorioztatu dute hautespen naturalak eragina izan duela gene horiengan (Cagliani et al., 2011; Esnault et al., 2005; Zhang & Webb, 2004). Gainera, GIB birusaren eta primateen arteko erlazioa aztertuta, ondorioztatu da APOBEC3G genea primateetan duela 30 milioi urtetatik hautespen natural gogorraren eraginpean dagoela (Sawyer, Emerman, & Malik, 2004).

GIB birusaren Vif proteina ez da infekziosoa APOBEC3G genearen presentzian (Mangeat et al., 2003), baina arazoa da GIB birusak APOBEC3G geneari aurre egiteko mekanismoak garatzen dituela entzimen proteosomak degradatuz (Marin, Rose, Kozak, & Kabat, 2003; Stopak, de Noronha, Yonemoto, & Greene, 2003). Hortaz, APOBEC3 geneen eta birusaren arteko lehia gertatzen ari da gaur egun.

GIB birusarekin laburbilduz, APOBEC3G eta APOBEC3H, eta GIB birusaren artean harreman zuzena dago, zeinak APOBEC3G eta APOBEC3H geneak hautespen naturalaren eraginpean jartzen dituen. HIESa duela 100 urte agertu zen gizakian eta, ondorioz, ezin da gene horien aleloen banaketaren eragile nagusia izan. Hala ere, primateen eta GIB birusaren arteko erlazioa APOBEC3G genearekin aztertzean, agerian utzi da duela 30 milioi urtetatik hautespen natural gogorraren eraginpean egon dela eta, hortaz, gaur eguneko primateen APOBEC3G genearen aleloen banaketan eragina izan duela ondorioztatu da (Sawyer et al., 2004).

6.3- APOBEC3G eta APOBEC3H geneetako hautespena:

APOBEC3G eta APOBEC3H geneetan hautespena jasan duten SNPen F_{ST} balioa ez da bera eta, egoera horretan, hautespenak desberdin eragin du SNPetan. F_{ST} balioa altua ($F_{ST} > 0,05$) duten SNPek populazioen arteko desberdintasunak sendotu dituzte, hau da,

hautespen dibertsifikatzaileak populazioak desberdindu ditu. F_{ST} balio baxua ($F_{ST} < 0,05$) duten SNPek, aldiz, populazioen barruko desberdintasunak eragin dituzte, hau da, hautespen orekatzaileak populazioen barruko indibiduoak desberdindu ditu.

6.4- APOBEC3G eta APOBEC3H geneen hautespenaren zergatia:

Gizakia duela 60.000 urte Afrikatik irten zen (Forster, 2004; Mondal et al., 2016; Thangaraj et al., 2005), eta jakina da gizakia Afrikatik irten zenean patogeno berriekin topo egin zuela, eta patogeno berrien erlazioak giza genoma aldaraztea eragin zuela. Alelo berriak sortu ziren, eta hautespen naturalak patogeno horietara hobeto moldatu zen aleloa faboratu zuen. Gainera, azken ikerketek agerian utzi dute birusek uste baino eragin handiagoa izan zutela giza genomaren aldaketan, hain handia non proteinen adaptazioen eragilerik handienak birusak direla proposatu den, proteinen adaptazioen % 30 eragiten dutelarik. (Enard, Cai, Gwennap, & Petrov, 2016)

Egoera horren aurrean, eta APOBEC3G eta APOBEC3H geneen funtzio antibirala aztertuta, ondorioztatu da APOBEC3G eta APOBEC3H geneen hautespenaren dibertsifikatzailearen eragilea erretrovirusak izan zirela. Gizakia Afrikatik irten zenean, erretrovirus berriekin topatu zen, eta haiei aurre egiteko erretrovirusen aurka egiten zuten proteinek adaptazio berri bat garatu zuten. Proteinen adaptazio horrek ahalbidetu zuen alelo berriaren nagusitasuna, Asian nagusi den aleloaren nagusitasuna, alegia.

Alelo berria Afrikatik kanpo sortu zenean, alelo hori zuten indibiduoak faboratu zituen, eta arrakasta handiagoa lortu zuten. Arrakasta horrek indibiduo horien eta haien ondorengoen iraunkortasuna handiagotu zuen eta, horregatik, alelo berriak belaunaldiz belaunaldi transmititu eta nagusitu zen. Gaur egun, alelo berriak dirau.

Hala ere, F_{ST} balio baxua ($F_{ST} < 0,05$) duten SNPek hautespen orekatzailea gertatu dela agerian uzten dute, eta hautespen orekatzaileak populazio guztietan bi aleloak agertzea eragin du. Populazioetan bi aleloak agertzea oso garrantzitsua dela detektatu dugu, eta seguruenik proteinen adierazpenerako edo erregulaziorako izan da garrantzitsua bi aleloak giza populazioetan orekan mantentzea.

SNP horiek Afrikan duten maiztasuna aztertuta, bi aleloak beste populazioetan bezala nagusi direla ikusirik, ondorioztatu da gizakia Afrikatik irten baino lehen hautespen orekatzaileak lan egin zuela. Hau da, gizakia Afrikan sortu zenean, APOBEC3 proteinen

adierazpen edo erregulazio egokirako F_{ST} baxua ($F_{ST} < 0,05$) duten SNP horien bi aleloak Afrikako populazioetan egonkortu ziren. Bi aleloen populazioetako presentzia hain garrantzitsua izanik, gizakia Afrikatik irten zenean sortutako populazioetan jatorrizko maiztasunak oro har mantendu dira.

7. ONDORIOAK:

APOBEC3G eta APOBEC3H geneek hautespena natural dibertsifikatzailea eta orekatzailea jasan dutela detektatu da. APOBEC3G eta APOBEC3H geneek funtzio antibiralak dituzte, eta erretrovirusak, erretrovirus endogenoak eta erretrotransposonak inhibitzen dituzte. APOBEC3 proteinen adierazpen edo erregulazio egokirako F_{ST} balio txikiak ($F_{ST} < 0,09$) duten SNPak detektatu dira, zeintzuk hautespen natural orekatzaileak populazioetan SNPen bi aleloen maiztasunen oreka faboratu duen. Hautespen natural orekatzailea Afrikako populazioetan egin zuen lan, eta gizakia Afrikatik irteteen sortutako giza populazioek maiztasun antzekoak izaten jarraitu dute, bi aleloen oreka proteinaren adierazpenerako edo erregulaziorako funtsezkoa baita. Gizakia Afrikatik irten zenean, gizakia erretrovirus berriekin topatu zen, eta patogeno horietara hobeto moldatzen ziren APOBEC3G eta APOBEC3H geneen genotipoa faboratu zuen hautespen natural dibertsifikatzaileak. Egoera horren ondorioz, Afrikako eta Afrika kanpoko populazioen maiztasunak desberdina dira F_{ST} altua ($F_{ST} > 0,05$) duten SNPei dagokienez.

Etorkizunera begira, oso interesgarria izango litzateke APOBEC3 geneen aldakortasun genetiko osoaren azterketa egitea. Egoera horretan, SNP bakoitzaren funtzio zehatza zein den jakin genezake, eta hautespen naturalaren eragina zehatzago azter genezake. Bestalde, GIB birusak gaur egun duen eragina kontuan hartuta, gene horiek gehiago aztertuko balira, erretrovirusaren aurka egiteko tratamendu genetikoak gara litezke.

8. BIBLIOGRAFIA

A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. (2001). *Nature*, 409(6822), 928–933. <https://doi.org/10.1038/35057149>

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*, 40(3), 340–345. <https://doi.org/10.1038/ng.78>

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–265. <https://doi.org/10.1093/bioinformatics/bth457>

Beaumont, M. A. (2005). Adaptation and speciation: what can F_{st} tell us? *Trends in Ecology & Evolution*, 20(8), 435–440. <https://doi.org/10.1016/J.TREE.2005.05.017>

Bogerd, H. P. (2006). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Research*, 34(1), 89–95. <https://doi.org/10.1093/nar/gkj416>

Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O’Shea, K. S., Moran, J. V., & Cullen, B. R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8780. <https://doi.org/10.1073/PNAS.0603313103>

Cagliani, R., Riva, S., Fumagalli, M., Biasin, M., Caputo, S. Lo, Mazzotta, F., ... Sironi, M. (2011). A positively selected apobec3h haplotype is associated with natural resistance to hiv-1 infection. *Evolution*, 65(11), 3311–3322. <https://doi.org/10.1111/j.1558-5646.2011.01368.x>

Chen, Z., Luckay, A., Sodora, D. L., Telfer, P., Reed, P., Gettie, A., ... Marx, P. A. (1997). Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *Journal of Virology*, 71(5), 3953–3960. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9094672>

- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Day, M. H. (1969). Omo human skeletal remains. *Nature*, 222(5199), 1135–1138. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5788978>
- Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*, 418(6898), 700–707. <https://doi.org/10.1038/nature01019>
- Enard, D., Cai, L., Gwennap, C., & Petrov, D. A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *ELife*, 5. <https://doi.org/10.7554/eLife.12469>
- Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., ... Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, 433(7024), 430–433. <https://doi.org/10.1038/nature03238>
- Esnault, C., Millet, J., Schwartz, O., & Heidmann, T. (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Research*, 34(5), 1522. <https://doi.org/10.1093/NAR/GKL054>
- Esparza-Baquer, A., Larruskain, A., Mateo-Abad, M., Minguijón, E., Juste, R. A., Benavides, J., ... Jugo, B. M. (2015). SNPs in APOBEC3 cytosine deaminases and their association with Visna/Maedi disease progression. *Veterinary Immunology and Immunopathology*, 163(3–4), 125–133. <https://doi.org/10.1016/j.vetimm.2014.11.007>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forster, P. (2004). Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1442), 255–264. <https://doi.org/10.1098/rstb.2003.1394>

- Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., ... Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, *397*(6718), 436–441. <https://doi.org/10.1038/17130>
- Haile-Selassie, Y., Asfaw, B., & White, T. D. (2004). Hominid cranial remains from upper pleistocene deposits at Aduma, Middle Awash, Ethiopia. *American Journal of Physical Anthropology*, *123*(1), 1–10. <https://doi.org/10.1002/ajpa.10330>
- Hartl, D., Clark, A., & Clark, A. (1997). Principles of population genetics. Retrieved from <http://www.academia.edu/download/28139957/183030028.pdf>
- Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., & Navaratnam, N. (2002). An Anthropoid-Specific Locus of Orphan C to U RNA-Editing Enzymes on Chromosome 22. *Genomics*, *79*(3), 285–296. <https://doi.org/10.1006/geno.2002.6718>
- Jorde, L. B., & Wooding, S. P. (2004). Genetic variation, classification and “race.” *Nature Genetics*, *36*(S11), S28–S33. <https://doi.org/10.1038/ng1435>
- Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R., & Eichler, E. E. (2007). Population Stratification of a Common APOBEC Gene Deletion Polymorphism. *PLoS Genetics*, *3*(4), e63. <https://doi.org/10.1371/journal.pgen.0030063>
- Kimura, M. (1991). The neutral theory of molecular evolution: a review of recent evidence. *Idengaku Zasshi*, *66*(4), 367–386. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1954033>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, *424*(6944), 99–103. <https://doi.org/10.1038/nature01709>
- Marcus, J. H., & Novembre, J. (2017). Visualizing the geography of genetic variants. *Bioinformatics*, *33*(4), 594. <https://doi.org/10.1093/BIOINFORMATICS/BTW643>

- Marin, M., Rose, K. M., Kozak, S. L., & Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Medicine*, 9(11), 1398–1403. <https://doi.org/10.1038/nm946>
- McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, 433(7027), 733–736. <https://doi.org/10.1038/nature03258>
- Mondal, M., Casals, F., Xu, T., Dall’Olio, G. M., Pybus, M., Netea, M. G., ... Bertranpetit, J. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nature Genetics*, 48(9), 1066–1070. <https://doi.org/10.1038/ng.3621>
- Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M., & Holmes, E. C. (2001). Phylogeny and the origin of HIV-1. *Nature*, 410(6832), 1047–1048. <https://doi.org/10.1038/35074179>
- R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>.
- Rousset, F. (2008). genepop’007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Sawyer, S. L., Emerman, M., & Malik, H. S. (2004). Ancient Adaptive Evolution of the Primate Antiviral DNA-Editing Enzyme APOBEC3G. *PLoS Biology*, 2(9), e275. <https://doi.org/10.1371/journal.pbio.0020275>
- Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1(1), a006841. <https://doi.org/10.1101/cshperspect.a006841>
- Stopak, K., de Noronha, C., Yonemoto, W., & Greene, W. C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Molecular Cell*, 12(3), 591–601. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14527406>

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A., & Singh, L. (2005). Reconstructing the Origin of Andaman Islanders. *Science*, 308(5724), 996–996. <https://doi.org/10.1126/science.1109987>

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), 1304–1351. <https://doi.org/10.1126/science.1058040>

Zhang, J., & Webb, D. M. (2004). Rapid evolution of primate antiviral enzyme APOBEC3G. *Human Molecular Genetics*, 13(16), 1785–1791. <https://doi.org/10.1093/hmg/ddh183>

9. ERANSKINAK

1. ERANSKINA

1. Eranskina: APOBEC3 geneetan hautespenaren eraginpean zeuden SNPak biltzen dira, eta SNP horietako aleloetako batek aztertutako 26 giza populazioetan zuen maiztasuna ageri da.

SNP	ACB	ASW	BEB	CDX	CEU	CHB	CHS	CLM	ESN	FIN	GBR	GIH	GWD	IBS	ITU	JPT	KHV	LWK	MSL	MXL	PEL	PIL	PUR	STU	TSI	YRI
rs7291971	0.156	0.230	0.686	0.688	0.611	0.752	0.800	0.628	0.051	0.500	0.577	0.612	0.119	0.584	0.593	0.774	0.722	0.025	0.094	0.688	0.771	0.646	0.630	0.583	0.626	0.037
rs5757464	0.161	0.254	0.686	0.688	0.611	0.752	0.790	0.633	0.101	0.500	0.577	0.612	0.124	0.584	0.593	0.779	0.717	0.051	0.106	0.688	0.771	0.651	0.630	0.583	0.626	0.051
rs2294367	0.130	0.246	0.657	0.624	0.571	0.680	0.700	0.617	0.066	0.439	0.533	0.617	0.124	0.542	0.588	0.707	0.677	0.056	0.082	0.664	0.647	0.641	0.577	0.574	0.593	0.042
rs3091374	0.146	0.262	0.657	0.624	0.576	0.680	0.700	0.617	0.066	0.439	0.533	0.617	0.133	0.542	0.588	0.712	0.677	0.061	0.147	0.664	0.647	0.630	0.577	0.574	0.593	0.046
rs2011861	0.130	0.246	0.657	0.624	0.571	0.680	0.700	0.617	0.066	0.439	0.533	0.617	0.124	0.542	0.588	0.712	0.677	0.056	0.076	0.664	0.647	0.635	0.572	0.574	0.593	0.042
rs2413570	0.130	0.246	0.651	0.624	0.571	0.680	0.700	0.617	0.066	0.434	0.533	0.612	0.124	0.542	0.588	0.712	0.677	0.056	0.076	0.664	0.647	0.625	0.572	0.574	0.589	0.042
rs139283	0.130	0.238	0.610	0.618	0.545	0.646	0.686	0.638	0.066	0.460	0.549	0.597	0.084	0.561	0.515	0.688	0.672	0.051	0.076	0.672	0.606	0.557	0.577	0.544	0.607	0.056
rs139285	0.120	0.213	0.610	0.618	0.545	0.646	0.681	0.638	0.061	0.460	0.549	0.597	0.084	0.556	0.515	0.688	0.672	0.030	0.076	0.672	0.606	0.557	0.577	0.544	0.607	0.028
rs3765441	0.141	0.238	0.628	0.618	0.545	0.655	0.690	0.649	0.091	0.439	0.533	0.597	0.088	0.556	0.515	0.707	0.672	0.051	0.082	0.688	0.724	0.578	0.582	0.549	0.603	0.083
rs139301	0.130	0.213	0.628	0.618	0.540	0.650	0.690	0.649	0.071	0.439	0.533	0.597	0.080	0.556	0.515	0.707	0.672	0.030	0.076	0.688	0.724	0.578	0.582	0.549	0.603	0.037
rs139305	0.130	0.213	0.628	0.618	0.540	0.650	0.690	0.649	0.071	0.439	0.533	0.597	0.080	0.556	0.515	0.707	0.672	0.030	0.076	0.688	0.724	0.573	0.582	0.544	0.603	0.037
rs139310	0.141	0.238	0.628	0.618	0.535	0.650	0.686	0.649	0.076	0.439	0.522	0.592	0.080	0.556	0.515	0.707	0.667	0.051	0.076	0.688	0.724	0.573	0.582	0.544	0.603	0.065
rs139312	0.141	0.238	0.628	0.618	0.540	0.650	0.690	0.649	0.076	0.439	0.533	0.597	0.080	0.556	0.515	0.707	0.672	0.051	0.076	0.688	0.724	0.573	0.582	0.544	0.603	0.065
rs139317	0.156	0.246	0.628	0.618	0.540	0.650	0.690	0.649	0.081	0.439	0.533	0.597	0.088	0.556	0.515	0.707	0.667	0.106	0.082	0.688	0.724	0.573	0.582	0.544	0.603	0.079
rs139279	0.510	0.533	0.372	0.349	0.384	0.340	0.310	0.303	0.561	0.480	0.363	0.393	0.473	0.322	0.466	0.298	0.303	0.631	0.465	0.273	0.376	0.422	0.327	0.431	0.336	0.583
rs139284	0.510	0.541	0.384	0.349	0.389	0.345	0.310	0.298	0.616	0.480	0.363	0.379	0.487	0.327	0.475	0.298	0.303	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.616
rs139286	0.510	0.541	0.378	0.349	0.389	0.345	0.314	0.298	0.616	0.480	0.363	0.379	0.487	0.332	0.475	0.298	0.303	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.616
rs139287	0.510	0.541	0.378	0.349	0.389	0.345	0.310	0.298	0.616	0.480	0.363	0.379	0.487	0.327	0.475	0.298	0.303	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.616
rs139288	0.505	0.541	0.378	0.349	0.389	0.345	0.310	0.298	0.606	0.480	0.363	0.379	0.487	0.332	0.475	0.298	0.303	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.606
rs139290	0.505	0.541	0.378	0.349	0.389	0.345	0.310	0.298	0.606	0.480	0.363	0.379	0.487	0.327	0.475	0.298	0.308	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.606
rs139291	0.505	0.541	0.378	0.349	0.389	0.345	0.310	0.298	0.606	0.480	0.363	0.379	0.487	0.327	0.471	0.298	0.303	0.636	0.471	0.281	0.376	0.427	0.322	0.426	0.332	0.606

2. ERANSKINA

2. eranskina: Hautespén dibertsifikatzailea jasan duten 22. kromosomako geneak, gene bakoitzean hautespén orekatzailea jasan duten SNPen kopurua eta SNPen batzbesteko F_{ST} a

Genea	SNP	F_{ST}
POTEH	1	0,1387
CCT8L2	1	0,1909
XKR3	1	0,2518
CECR2	38	0,1787
SLC25A18	8	0,1616
ATP6V1E1	7	0,1511
BCL2L13	2	0,1645
BID	9	0,1688
MICAL3	11	0,1828
LINC01634	1	0,1318
LOC105372853	1	0,2041
TUBA8	8	0,2131
USP18	2	0,1530
LINC01663	8	0,1594
DGCR6	11	0,2057
PRODH	5	0,1658
DGCR5	12	0,1624
LINC01311	1	0,1702
LOC105372859	1	0,1716
TXNRD2	1	0,1780
LOC105372863	2	0,2052
RTN4R	1	0,1567
KLHL22	1	0,1525
MED15	1	0,1748
LRRC74B	3	0,2406
BCRP2	5	0,1488
FAM230B	1	0,1409
GGT2	2	0,1873
LOC100996335	2	0,1481
PI4KAP2	9	0,2334
MIR130B	2	0,1217
MIR301B	2	0,1217
LOC107985532	3	0,1383
PRAMENP	1	0,1333
LOC102724653	1	0,1598
LOC105379602	37	0,1682
ZNF280A	31	0,1692
LOC105372948	4	0,1564
BCR	11	0,1904
CES5AP1	8	0,1849
ZDHHC8P1	2	0,1815
LINC01658	2	0,1903

Genea	SNP	F_{ST}
LOC107985580	2	0,1768
LOC105372957	2	0,2808
IGLL1	3	0,2004
DRICH1	3	0,1685
GUSBP11	8	0,1414
SMARCB1	1	0,1492
GSTT4	17	0,1730
POM121L10P	1	0,1451
BCRP3	1	0,1725
SGSM1	7	0,2102
KIAA1671	18	0,1771
CRYBB2	3	0,1498
IGLL3P	1	0,1558
LRP5L	6	0,1401
CRYBB2P1	3	0,1409
MYO18B	33	0,1705
SEZ6L	10	0,2106
TPST2	4	0,1419
CRYBB1	1	0,1681
MIATNB	3	0,1623
LINC01422	1	0,1618
LOC107985534	5	0,1816
LOC105372981	5	0,1658
MN1	1	0,1665
TTC28	6	0,1438
CHEK2	3	0,1432
ZNRF3	7	0,1360
KREMEN1	11	0,1671
LOC101929638	2	0,2339
EMID1	5	0,1905
RFPL1	8	0,1659
CABP7	5	0,1588
ZMAT5	9	0,1673
UQCR10	3	0,1479
ASCC2	7	0,1580
MTMR3	1	0,1396
CCDC157	1	0,1291
SF3A1	1	0,1291
SEC14L3	1	0,1192
PES1	2	0,1222
TCN2	2	0,1222
DUSP18	9	0,1705

Genea	SNP	F_{ST}
SLC35E4	10	0,1716
SMTN	2	0,1613
PLA2G3	1	0,1451
RNF185	5	0,1814
LIMK2	1	0,1837
PIK3IP1	2	0,1825
DRG1	1	0,1798
EIF4ENIF1	4	0,1743
SFI1	1	0,1524
LINC02558	1	0,1603
SLC5A1	6	0,1684
AP1B1P1	5	0,1720
C22orf42	5	0,1826
RFPL2	27	0,1520
SLC5A4-AS1	6	0,1719
SLC5A4	6	0,1630
RFPL3	1	0,1220
LOC339666	1	0,1690
BPIFC	1	0,1540
LL22NC03-104C7.1	2	0,1369
SYN3	39	0,1775
LARGE1	46	0,1919
LINC01643	12	0,1384
LOC105373012	2	0,1522
ISX-AS1	57	0,1982
LINC01399	60	0,1808
LOC105373017	11	0,1820
HMGXB4	21	0,1785
TOM1	19	0,1724
MCM5	5	0,2033
RASD2	6	0,1681
LOC107985590	1	0,2093
RBFOX2	1	0,2400
LOC102723507	5	0,1631
APOL3	20	0,1970
APOL4	15	0,1618
APOL2	5	0,1962
APOL1	2	0,1523
LOC112268296	1	0,1483
MYH9	31	0,1933
LOC105377199	8	0,2043
TXN2	2	0,1998

Genea	SNP	F _{ST}
FOXRED2	8	0,2009
EIF3D	4	0,1836
CACNG2	4	0,1394
LOC105373021	16	0,1700
IFT27	4	0,1642
PVALB	4	0,1784
NCF4-AS1	1	0,1522
CSF2RB	12	0,1578
LOC112268295	3	0,1514
TMPRSS6	5	0,1605
CYTH4	14	0,1587
ELFN2	2	0,1297
CDC42EP1	1	0,2030
TRIOBP	1	0,1503
KCNJ4	1	0,1564
CBX6	1	0,1773
LOC105373032	1	0,1660
APOBEC3G	6	0,1578
LOC101927202	1	0,1608
APOBEC3H	8	0,1548
PDGFB	2	0,1656
RPL3	3	0,1507
SNORD139	2	0,1632
SNORD83A	2	0,1632
SNORD43	1	0,1259
SYNGR1	6	0,1852
MGAT3	6	0,1880
CACNA1I	40	0,2052
FAM83F	4	0,1579
TNRC6B	8	0,2000
MKL1	17	0,1627
MCHR1	2	0,2282
SLC25A17	7	0,1765
ST13	2	0,1890
XPNPEP3	6	0,1743
RBX1	1	0,1480
EP300	10	0,1448
EP300-AS1	2	0,1500
ZC3H7B	6	0,1413
TEF	2	0,1472
TOB2	2	0,1472
PHF5A	1	0,1329

Genea	SNP	F_{ST}
ACO2	5	0,1502
POLR3H	1	0,1608
CCDC134	1	0,1255
SREBF2	1	0,1285
TNFRSF13C	2	0,2197
CENPM	4	0,2242
LINC00634	8	0,2207
SEPT3	11	0,1837
WBP2NL	35	0,1845
NAGA	2	0,2176
NFAM1	4	0,1672
SERHL	8	0,2078
RRP7A	3	0,1792
SERHL2	10	0,1603
RRP7BP	5	0,1679
POLDIP3	3	0,1630
CYB5R3	10	0,1960
A4GALT	2	0,1499
PACSIN2	4	0,2247
TTLL1	11	0,1645
BIK	2	0,2238
MCAT	4	0,1884
TSPO	9	0,2692
FLJ30901	1	0,1637
SCUBE1	23	0,1855
LOC105373053	1	0,1786
MPPED1	2	0,1725
EFCAB6-AS1	2	0,2270
EFCAB6	7	0,2045
PNPLA5	19	0,1766
PARVB	2	0,1513
SHISAL1	1	0,2291
LINC00207	3	0,1957
LOC105373060	2	0,2299
PRR5	7	0,1611
PRR5-ARHGAP8	16	0,1521
ARHGAP8	5	0,1498
PHF21B	3	0,1835
NUP50-AS1	1	0,1536
UPK3A	5	0,1853
FAM118A	25	0,1550
SMC1B	17	0,1408

Genea	SNP	F_{ST}
RIBC2	4	0,1253
FBLN1	1	0,1305
LOC107985535	2	0,2493
WNT7B	3	0,1491
LOC730668	3	0,1809
LINC00899	4	0,2048
PRR34	1	0,1953
PRR34-AS1	3	0,1699
LOC642648	1	0,1887
MIRLET7BHG	15	0,2728
MIRLET7B	2	0,2852
PPARA	2	0,2413
CELSR1	11	0,1434
LOC105373075	7	0,1351
GRAMD4	23	0,1777
CERK	6	0,1584
TBC1D22A	41	0,1527
LINC00898	1	0,1565
LOC284930	20	0,1566
FAM19A5	46	0,1669
MIR4535	1	0,1592
LINC01310	3	0,1910
LOC105373086	2	0,1438
C22orf34	29	0,1511
BRD1	8	0,1872
LOC105377205	1	0,3255
ZBED4	32	0,1935
ALG12	54	0,1901
LOC107985553	1	0,1370
PIM3	2	0,1413
IL17REL	1	0,1682
MIR6821	1	0,1682
TTLL8	1	0,1314
PLXNB2	1	0,1497
PPP6R2	4	0,1567
LOC105373098	1	0,1843
CHKB-CPT1B	5	0,1650
CPT1B	5	0,1650
SHANK3	2	0,1848

3. ERANSKINA

3. eranskina: Hautespen orekatzailea jasan duten 22. kromosomako geneak, gene bakoitzean hautespen orekatzailea jasan duten SNPen kopurua eta SNP horien batzbesteko F_{ST} a

Genea	SNP	F_{ST}
BMS1P22,	1	0,0063
DUXAP8	1	0,0063
LOC107987324	1	0,0266
LINC01297	2	0,0199
POTEH	1	0,0063
TPTEP1	1	0,0330
XKR3	16	0,0171
HSFY1P1	2	0,0333
GAB4	3	0,0313
CECR7	4	0,0322
LINC01664	1	0,0472
TMEM121B	1	0,0472
ADA2	4	0,0411
CECR2	12	0,0351
SLC25A18	7	0,0282
ATP6V1E1	1	0,0183
BCL2L13	1	0,0113
MICAL3	4	0,0194
LOC105372853	1	0,0501
USP18	1	0,0221
LINC01662	3	0,0183
LOC105377182	10	0,0323
GGT3P	3	0,0210
LOC102725072	4	0,0076
DGCR2	71	0,0529
CLTCL1	5	0,0168
LOC105372859	2	0,0318
HIRA	2	0,0262
LOC105372859	1	0,0397
MRPL40	1	0,0400
LOC105372860	1	0,0316
UFD1	13	0,0318
UFD1	2	0,0320
LOC105372861	1	0,0275
TBX1	2	0,0449
TXNRD2	8	0,0363
COMT	3	0,0460
TANGO2	2	0,0437
MIR185	2	0,0437
LOC105372863	1	0,0039
RTN4R	3	0,0388
LOC105372864	2	0,0399

Genea	SNP	F_{ST}
TMEM191B	4	0,0327
PI4KAP1	10	0,0327
LINC01660	3	0,0074
ZNF74	1	0,0456
LOC107985584	1	0,0376
PI4KA	9	0,0454
AIFM3	5	0,0347
LZTR1	2	0,0343
P2RX6	4	0,0429
TUBA3FP	4	0,0429
FAM230B	2	0,0082
GGT2	6	0,0164
POM121L8P	1	0,0299
RIMBP3B	1	0,0105
YPEL1	1	0,0253
MAPK1	11	0,0423
PPM1F	4	0,0373
TOP3B	1	0,0509
VPREB1	1	0,0376
BMS1P20	7	0,0406
LOC101929255	8	0,0512
LOC105379601	1	0,0365
PRAME	1	0,0148
LL22NC03-102D1.18	15	0,0261
LOC107985542	12	0,0364
LOC107985543	1	0,0569
GNAZ	4	0,0350
RSPH14	44	0,0331
RAB36	29	0,0289
BCR	9	0,0395
LINC01658	7	0,0318
LINC02557	2	0,0101
PCAT14	4	0,0449
DRICH1	4	0,0329
GUSBP11	4	0,0437
RGL4	1	0,0451
VPREB3	1	0,0508
SLC2A11	1	0,0352
LOC100652871	2	0,0282
GGT5	6	0,0286
POM121L9P	4	0,0258
SPECC1L-ADORA2A	14	0,0285
SPECC1L	11	0,0279

Genea	SNP	F_{ST}
ADORA2A	3	0,0310
ADORA2A-AS1	21	0,0327
UPB1	4	0,0490
SNRPD3	1	0,0059
GGT1	2	0,0054
LOC105372962	1	0,0358
PIWIL3	26	0,0335
KIAA1671	1	0,0162
LRP5L	1	0,0033
MYO18B	13	0,0241
SEZ6L	8	0,0408
ASPHD2	2	0,0278
TFIP11	2	0,0470
TPST2	21	0,0373
MIAT	4	0,0234
MIATNB	8	0,0321
LOC110091768	1	0,0583
LOC107985582	1	0,0149
LOC284898	1	0,0063
LOC107985534	3	0,0397
LOC102724900	2	0,0318
MN1	1	0,0345
PITPNB	1	0,0398
TTC28-AS1	1	0,0184
TTC28	7	0,0324
ZNRF3	44	0,0291
C22orf31	2	0,0460
KREMEN1	3	0,0144
LIF-AS1	2	0,0430
LIF	2	0,0430
SF3A1	1	0,0235
SEC14L6	1	0,0320
GAL3ST1	2	0,0317
PES1	20	0,0034
OSBP2	22	0,0353
PLA2G3	1	0,0294
SLC5A4	101	0,0337
LOC339666	1	0,0260
BPIFC	2	0,0033
FBXO7	2	0,0468
SYN3	32	0,0328
LARGE1	91	0,0392
LINC01643	6	0,0382

Genea	SNP	F_{ST}
ISX-AS1	21	0,0323
HMGXB4	2	0,0480
TOM1	2	0,0480
LOC107985590	1	0,0598
LOC102723507	1	0,0424
CACNG2	8	0,0448
LOC105373021	2	0,0204
PVALB	1	0,0446
TEX33	5	0,0287
TST	14	0,0253
TMPRSS6	18	0,0297
IL2RB	5	0,0263
SSTR3	1	0,0329
RAC2	7	0,0326
CARD10	1	0,0347
SH3BP1	2	0,0142
LOC101927051	10	0,0260
PDXP	10	0,0273
LGALS1	2	0,0392
H1FO	1	0,0506
GCAT	1	0,0495
ANKRD54	1	0,0078
POLR2F	2	0,0338
SLC16A8	2	0,0287
BAIAP2L2	3	0,0290
PLA2G6	12	0,0337
TMEM184B	8	0,0320
LOC400927-CSNK1E	1	0,0495
LOC105373030	2	0,0479
FAM227A	1	0,0406
LOC105373032	2	0,0500
APOBEC3H	7	0,0461
GRAP2	1	0,0038
TNRC6B	1	0,0049
C22orf46	1	0,0273
SNU13	1	0,0273
LOC112267923	1	0,0401
LOC112268342	1	0,0401
LOC102723722	1	0,0401
LOC105377203	1	0,0211
LINC01315	2	0,0529
NFAM1	4	0,0565

Genea	SNP	F_{ST}
POLDIP3	1	0,0127
LOC100506679	1	0,0322
TLL1	1	0,0580
TLL12	2	0,0342
SCUBE1	3	0,0409
MPPED1	9	0,0181
EFCAB6	2	0,0433
SULT4A1	1	0,0259
SAMM50	7	0,0494
PARVB	20	0,0327
PARVG	9	0,0174
LOC107985571	1	0,0385
SHISAL1	7	0,0283
LINC01656	1	0,0529
LOC105373059	6	0,0239
PRR5	1	0,0306
ARHGAP8, PRR5-ARHGAP8	32	0,0331
ARHGAP8, PRR5-ARHGAP8	32	0,0331
PHF21B	3	0,0330
NUP50-AS1	4	0,0326
NUP50	4	0,0433
LOC105373064	4	0,0433
KIAA0930	4	0,0354
LOC105373066	1	0,0394
FBLN1	1	0,0394
LOC105373069	3	0,0391
LOC105373067	2	0,0482
LOC105373068	1	0,0207
ATXN10	2	0,0215
WNT7B	1	0,0369
CELSR1	2	0,0395
GRAMD4	8	0,0265
CERK	2	0,0302
LOC107985566	1	0,0473
TBC1D22A	58	0,0343
LOC339685	2	0,0422
LINC01644	3	0,0162
LINC00898	1	0,0514
LOC284930	5	0,0139
FAM19A5	42	0,0328
LINC01310	2	0,0127
C22orf34	6	0,0425

Genea	SNP	F_{ST}
IL17REL	1	0,0241
TTLL8	19	0,0507
MLC1	3	0,0236
MOV10L1	1	0,0036
PANX2	2	0,0330
TRABD	2	0,0230
SELENOO	6	0,0270
TUBGCP6	2	0,0145
MAPK12	2	0,0349
PLXNB2	8	0,0440
PPP6R2	8	0,0906
SBF1	2	0,0173
ADM2	1	0,0494
MIOX	1	0,0494
LOC105373098	1	0,1843
KLHDC7B	2	0,0178
LOC107985568	1	0,1843
CHKB-CPT1B, CPT1B	3	0,1761
CHKB-AS1	4	0,0910
ARSA	1	0,0548
SHANK3	6	0,0817
RABL2B	2	0,0216
RPL23AP82	3	0,0204