

Family Finder -testi ja genomilaajuisten tulosten hyödyntäminen

Jaakko Häkkinen, 29.6.2013

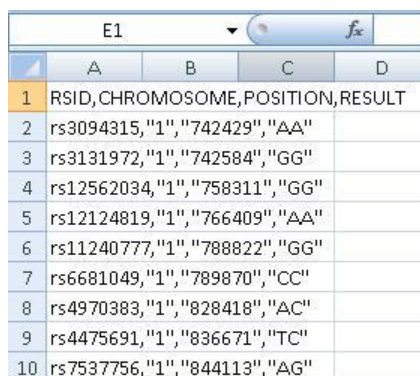
Testin tilaaminen

FamilyTreeDNA tarjoaa isälinja- (Y-kromosomin DNA) ja äitilinjatestin (mitokondrion DNA) lisäksi ns. genomilaajuisen testin, jossa analysoidaan yhteensä lähes miljoona kohtaa kaikista 22:sta autosomista (ne kromosomit, joiden toinen puolisko peritään isältä ja toinen äidiltä). Maksaminen hoituu kätevimmin luottokortilla, ja sen jälkeen saa testipaketin eli kolme kaavinta, joilla rapsutellaan tovi posken sisäpintaa, minkä jälkeen ne pakataan ohjeiden mukaan ja lähetetään Yhdysvaltoihin. Koska tulokset koskevat suunnilleen myös sisarusia ja kertovat jotain vanhemmistakin, voi perheenjäsenille ehdottaa kulujen jakamista. (<http://www.familyreedna.com/>)

Testeistä saa yleensä alennuksen projektien kautta, joten kannattaa liittyä Finland DNA -projektiin (<http://www.familyreedna.com/group-join-request.aspx?group=Finland&vGroup=finland>) samalla kun tilaa testipaketin ("Option B: Join").

Testitulokset

FamilyTreeDNA:n Family Finder -testin tulos lähetetään suurena tiedostona, jossa on runsaat 700 000 riviä, jokaisella kaksi alleelia (A, G, C ja T eri yhdistelminä). Toinen alleeli on peritty isältä ja toinen äidiltä, mutta valitettavasti järjestys on satunnainen eikä mene niin, että esimerkiksi vasemmanpuoleinen olisi peritty aina samalta vanhemmalta.



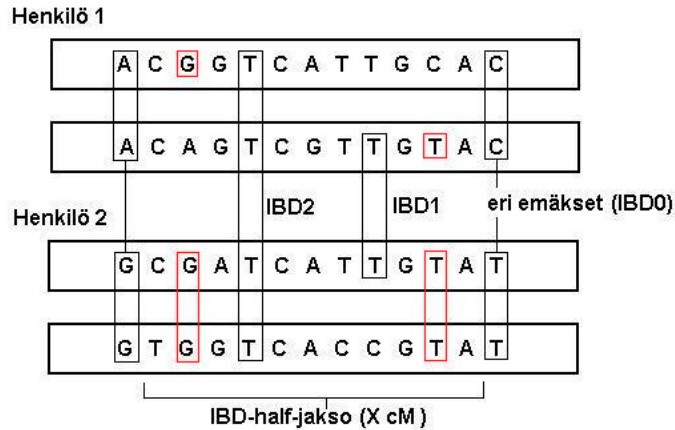
	A	B	C	D
1	RSID, CHROMOSOME, POSITION, RESULT			
2	rs3094315	"1", "742429"	"AA"	
3	rs3131972	"1", "742584"	"GG"	
4	rs12562034	"1", "758311"	"GG"	
5	rs12124819	"1", "766409"	"AA"	
6	rs11240777	"1", "788822"	"GG"	
7	rs6681049	"1", "789870"	"CC"	
8	rs4970383	"1", "828418"	"AC"	
9	rs4475691	"1", "836671"	"TC"	
10	rs7537756	"1", "844113"	"AG"	

Itse tiedostoa ei ole välttämätöntä edes avata, vaan tulosten analysoinnin hoitavat erilaiset ohjelmat. Excel ei oikein jaksa edes avata tiedostoa, joten jos haluaa katsella tuloksiaan, suosittelen ohjelmaa Notepad++, jonka voi ladata koneelleen ilmaiseksi (<http://notepad-plus-plus.org/>).

Pelkästä tiedostosta ei vielä näe yhtään mitään, eikä geenitestien hintaan sisälly minkäänlaista tulosten analysointia. Tämä on varmasti monelle testauttajalle suuri pettymys. Onneksi asianharrastajat, jotka hallitsevat geneetikkojen käyttämät ohjelmat, tarjoavat Internetissä ilmaiseksi erilaisia analysointityökaluja ja -projekteja. Genomilaajuista dataa voidaan analysoida useammalla mahdollisella tavalla – ne ovat kaikki oikeita tapoja, mutta ne saattavat kertoa hieman eri asioita.

IBD-taso (yksilötason sukulaisuus)

IBD-jakso (*identity by descent* eli peritysti samanlainen) tarkoittaa sellaista yksittäisistä emäksistä koostuvaa jaksoa, joka on kahdella yksilöllä samanlainen. Mitä pidempiä IBD-jaksoja, sitä läheisempää sukua henkilöt ovat, koska rekombinaatio pilkkoo IBD-jaksoja joka sukupolvi entistä lyhyemmiksi. Family Finder -testi tuottaa IBD-half-jaksoja, jolloin ei vaadita että molemmat emäkset olisivat samat, vaan riittää että edes jompikumpi emäs on vertailuhenkilöillä sama. Tällainen IBD-half-jakso päättyy vasta kun tulee kohta, jossa kumpikaan emäs ei ole vertailuhenkilöillä sama.



IBD2 merkitsee, että molemmat emäkset (alleelit) ovat samoja, ja IBD1 merkitsee, että vain toinen emäs on henkilöille yhteinen. IBD0 merkitsee että molemmat emäkset ovat erejä, jolloin IBD-half-jakso katkeaa.

IBD-half-menetelmän etuna on, että se toimii suuremmalla aikasyvyydellä. Jokaisessa sukupolvessa voi tapahtua mutaatioita, ja yhdenkin emäksen muutos johtaisi IBD-jakson katkeamiseen, jos vaadittaisiin että molempien emästen pitäisi olla samat. Toisaalta IBD-half-menetelmä ei pysty täysin luotettavasti erottamaan yksilötason sukulaisuutta väestötason sukulaisuudesta: pienten väestöjen yleisesti kokema geneettinen ajautuminen johtaa harvinaisten alleelien katoamiseen, jolloin yhä useammassa kohdassa on samaan väestöön kuuluvilla aivan samat alleelit (emäkset).

Esimerkiksi suomalaisilla on yleensä pitkiä IBD-half-jaksoja sellaistenkin suomalaisten kanssa, joiden juuret ovat aivan eri alueella – alleelien samanlaisuus (eli että ainakin toinen alleeli on yhteinen) johtuukin tällöin yleensä siitä, että suomalaisten geenipoolissa on vain tiettyjä alleleja, eikä taustalta välttämättä löydy todellista sukututkimuksen keinoin löydettävää sukulaisuutta. Mutta mitä pidempi IBD-half-jakso on kyseessä, sitä todennäköisemmin se kuvastaa todellista sukulaisuutta, koska samasta väestöpohjasta tulevillakin sattuu silloin tällöin kohtia, joissa mutaatioiden vuoksi molemmat alleelit ovat erejä. Tällöin IBD-half-jaksot jäävät lyhyemmiksi.

Eurooppalaisilla, joiden isovanhemmat edustavat samaa väestöä, homotsygoottisuus on yleensä noin 70 % tai hieman enemmän. Tämä merkitsee, että useammassa kuin kahdessa kohdassa kolmesta molemmilta vanhemmilta on peritty sama alleeli, ja vain joka kolmannessa tai neljännessä kohdassa on eri alleelit. Vertailtaessa kahta samaan väestöön kuuluvaa yksilöä on todennäköistä, että erialleelisuutta esiintyy aivan samoissa kohdissa: juuri niissä kohdissa (25–30 %), joissa väestön geenipoolissa yleensä on vielä kaksi alleelia jaossa.

Koska toinen alleeli on yleensä selvästi yleisempi, voidaan sanoa että niissä kohdissa (70–75 %), joissa väestön geenipoolissa on vain yksi alleeli, yksilöillä on pääasiassa kaksi samaa alleelia (IBD2), mutta niissä kohdissa (25–30 %), joissa väestön geenipoolissa on kaksi alleelia tarjolla, yksilöillä on silti usein kaksi samaa alleelia (IBD2), mutta toisinaan vain yksi sama alleeli (IBD1). Jos yleisemmän alleelin **A** esiintymistiheys on 80 % ja harvinaisemman alleelin **a** esiintymistiheys 20 %, on fenotyypin **Aa** todennäköisyys $0,8 \times 0,2 \times 2$ (koska myös **aA** lasketaan) = 32 %, fenotyypin **AA** todennäköisyys $0,8 \times 0,8 = 64$ % ja fenotyypin **aa** todennäköisyys $0,2 \times 0,2 = 4$ %. Tällöin aina kaksi tapausta kolmesta sellaisesta, joissa väestössä on kaksi alleelia jaossa, tuottaa silti yksilössä saman tuloksen kuin jos jaossa olisi vain yksi alleeli.

IBD-half-jaksojen pituus

IBD-half-jaksojen pituus ilmoitetaan centiMorganeina (cM). Family Finderin osuissa ilmoitetaan IBD-half-jaksot yhteensä (142,62 cM) sekä pisin yksittäinen jakso (28,86 cM); lisäksi annetaan arvio siitä, kuinka läheiseen sukulaisuuteen nämä luvut keskimäärin yleensä perustuvat.

Name	GENEALOGICAL MATCHES BASED ON DNA					Known Relationship
	Match Date	Relationship Range	Suggested Relationship	Shared cM	Longest Block	
[Redacted] Y-DNA67 FMS	6.7.2012	2nd Cousin - 3rd Cousin	2nd Cousin	142,62	28,86	ASSIGN
[Redacted] Y-DNA12	6.7.2012	2nd Cousin - 4th Cousin	3rd Cousin	124,21	24,53	ASSIGN
[Redacted] Y-DNA67	6.7.2012	2nd Cousin - 4th Cousin	3rd Cousin	123,33	20,80	ASSIGN
[Redacted] Y-DNA111 FMS	6.7.2012	2nd Cousin - 4th Cousin	3rd Cousin	90,48	19,05	ASSIGN

Osumien yhteystiedot pääsee näkemään, ja sitä kautta voi alkaa jäljittää yhteisiä esivanhempia. Hyödyllistä on myös ladata FamilyTreeDNA:n sivulle sukututkimusohjelmalla tehty GEDCOM-muotoon talletettu sukupuu (merkitty punaisella ympyrällä), jota osumat pääsevät tarkastelemaan. Tuttuja sukunimiä ja seutuja saattaa löytyä jo ensivilkaisulla. Ilmainen ja selkeäkäyttöinen sukututkimusohjelma on esimerkiksi Ahnenblatt; myös maksullisesta Sukujutut-ohjelmasta on saatavilla rajoitettu ilmaisversio.

Kromosomikartta

Hyödyllinen toiminto on myös *Chromosome browser*, jonka avulla voi verrata missä kohdissa yhteiset IBD-half-jaksot ovat. Seuraavassa kuvassa näkyvät kolmen osumieni joukosta valitun henkilön yli 1 cM:n mittaiset IBD-half-jaksot, jotka ovat yhteisiä kanssani. Lisäksi nähdään, mitkä niistä kohdista, jotka he jakavat minun kanssani, he jakavat myös keskenään. Esimerkiksi henkilöillä **1** ja **2** on yhteisiä kohtia kromosomissa 6, ja henkilöillä **1** ja **3** on yhteisiä kohtia kromosomeissa 4 ja 8. On kuitenkin huomattava, että henkilöt saattavat jakaa enemmänkin kohtia keskenään – ne eivät vain näy minun tulosteni kautta, koska minä en jaa niitä kohtia heidän kanssaan.

Joka tapauksessa yhdistämällä kromosomikartan tiedot henkilöiden esivanhempien asuinpaikkoihin on mahdollista saada käsitys siitä, mikä jakso missäkin kromosomissa todennäköisesti periytyy miltäkin alueelta. Tämä on työlästä, mutta muiden geenisukututkijoiden kanssa voi vaihtaa tietoa alan foorumeilla.

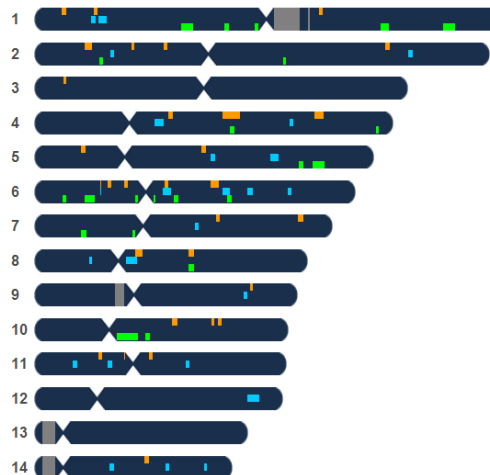
Compare List 1+ cM Remove

- Shared Segments: 38 Remove
- Shared Segments: 32 Remove
- Shared Segments: 25 Remove

Reset Defaults Clear Compare List

Filter Matches by... 41 - 50 of 231

- [Redacted]
- [Redacted]



IBS-taso (yksilö- ja väestötason sukulaisuus)

IBS-tasolla (*identity by state* eli muodollinen samankaltaisuus) verrataan kahden henkilön jokaisen SNP:n emäksiä yksitellen toisiinsa: jos molemmilla on GG, tulos on IBS = 2; jos toisella on GG ja toisella AG, tulos on IBS = 1; jos toisella on GG ja toisella AA, tulos on IBS = 0. Lopuksi kaikki nämä tulokset lasketaan yhteen, jaetaan ensin kahdella ja sitten vielä SNP-kohtien lukumäärällä. Lopputuloksena on lukuarvo, joka kertoo kuinka suuri osa kahden henkilön emäksistä on yhteisiä. Kahdella samaan väestöön kuuluvalla tämä arvo voi olla esimerkiksi 0,7634, mikä tarkoittaa että yli 76 % kaikkien SNP-kohtien emäksistä on yhteisiä.

Suuruusluokka on lähes sama kuin homotsygoottisuusaste, ja näillä tasoilla onkin yhteys: kun noin 3/4:ssa kohdista on molemmilta vanhemmilta peritty sama alleeli (homotsygoottisuus), on erittäin todennäköistä, että myös muilla samaan väestöön kuuluvilla on aivan sama alleeli perittynä molemmilta vanhemmiltaan. Kun siis väestössä on jonkin SNP:n kohdalla saavutettu tilanne, että harvinaisempi alleeli A on kadonnut ja kaikki ovat homotsygootteja GG (sekä isältä että äidiltä G-alleeli), on kaikilla tässä kohdassa IBS = 2.

IBS-tasollakin samankaltaisuus kasvaa sukulaisuuden myötä: samoja esivanhempia jakavien välillä IBS-arvo on suunnilleen 0,77, isovanhemman ja lapsenlapsen välillä 0,80 ja vanhemman ja lapsen välillä peräti 0,85. Kuvaa mutkistaa kuitenkin se, että jotkut henkilöt jakavat kaikkien kanssa korkeampia arvoja kuin jotkut toiset. Tästä syystä IBS-arvo täytyy suhteuttaa aina suurempaan otokseen, ja mieluummin lisäksi ulkomaalaisiin. Tällöin saadaan selville eri alueiden suomalaisten IBS-profiili, johon kyseinen yksilöllinen häiriötekijä ei vaikuta. Esimerkiksi länsirannikon suomalaiset ovat lähempänä baltteja kuin venäläisiä, riippumatta siitä, onko heillä tavallista korkeampi vai matalampi IBS-arvo muiden kanssa.

Admix-laskuritaso (väestötason sukulaisuus)

Kolmas taso, jonka avulla voidaan analysoida omaa tulosta, on sekoituslaskurien taso (*admixture* = sekoitus). Tällaisia laskureita on erilaisia, riippuen siitä, mitä väestöjä on saatu vertailuun. Ensinnäkin laitetaan ohjelma etsimään käsillä olevista geeninäytteistä geneettisiä "napoja", joiden sekoitukseksi muut väestöt voidaan sujuvimmin selittää, ja valmis laskuri sitten katsoo kuinka henkilön geenit jakautuvat näiden napojen kesken. Suomalaisten kannalta ei ole juurikaan hyötyä sellaisista laskureista, joissa on mukana suomalainen napa, koska kaikki suomalaiset saavat hyvin korkean arvon siihen, ja vain pari–kolmekymmentä prosenttia geeneistä jää muiden napojen jaettavaksi (summa on aina 100 %).

Admix-laskurit kertovat totuuden, mutteivät koskaan koko totuutta. Jos esimerkiksi laskurissa on mukana venäläinen napa ja jakuuttinapa, suomalaiset saattavat saada tulokseksi 20 % venäläinen ja 5 % jakuutti, loppujen 75 % geeneistä edustaessa läntisempiä eurooppalaisia napoja. Jos kuitenkin laskuria luotaessa mukana oli lisäksi mordvalaisia ja manseja, nämä navat vievät suurimman osan vastaavassa järjestyksessä venäläiseltä ja jakuuttinavalta. Eli laskurituloksen tarkkuus riippuu siitä, kuinka paljon erilaisia väestöjä napoja laskettaessa on ollut käytössä. Eri laskurien tulokset ovat silti samansuuntaisia: länsisuomalaisilla on enemmän yhteistä skandinaavien ja keskieurooppalaisten kanssa kuin itäsuomalaisilla.

Oman datatiedoston voi ladata GEDmatchiin (<http://gedmatch.com/>), joka vaatii nykyään rekisteröitymisen mutta on ilmainen, ja sen jälkeen sen voi ajaa eri admix-laskureissa. Kit-numeron eteen laitetaan "F":

Select the project:

Gedmatch.com > Admixture
Esimerkki: MDLP K=12

Select how you want to process it:

- Admixture Proportions (with link to Oracle)
- Admixture Proportions by Chromosome
- Chromosome Painting
- Chromosome Painting - Reduced Size
- Paint differences between 2 kits, 1 chromosome
- Paint differences between 2 kits, 22 chromosomes, reduced size

Enter your kit number:

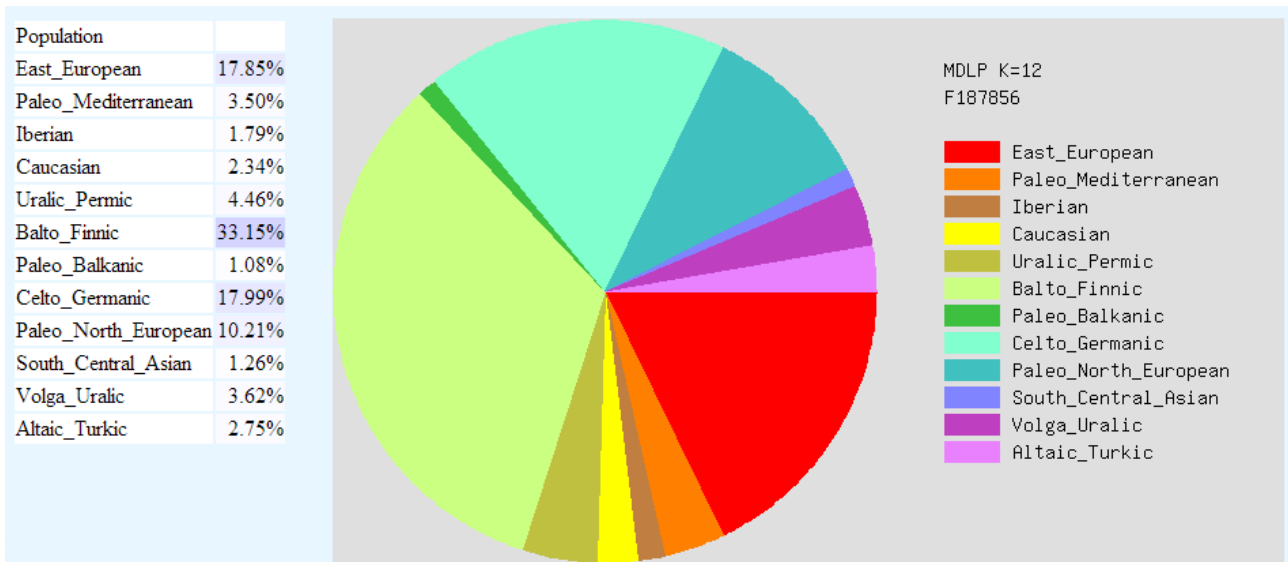
Select the 'calculator' model to use:

Optional - If all 4 of your grandparents are from the same ethnic group (example: Polish, Swedish, or Nigerian), you can help the creator of this project in his research by providing that information to him.

If you provide your ethnic group, that group name and your AP-mix results will be forwarded to him by GEDmatch. **No raw DNA data will be sent.**
 If you do not provide ethnicity information, no information will be forwarded to him.

Enter your Ethnicity:

Kun ajo on pysähtynyt, tulos näyttää tältä:



Sivun alareunasta voi monen laskurin kohdalla valita "Oracle", joka antaa tämännäköisiä tuloksia:

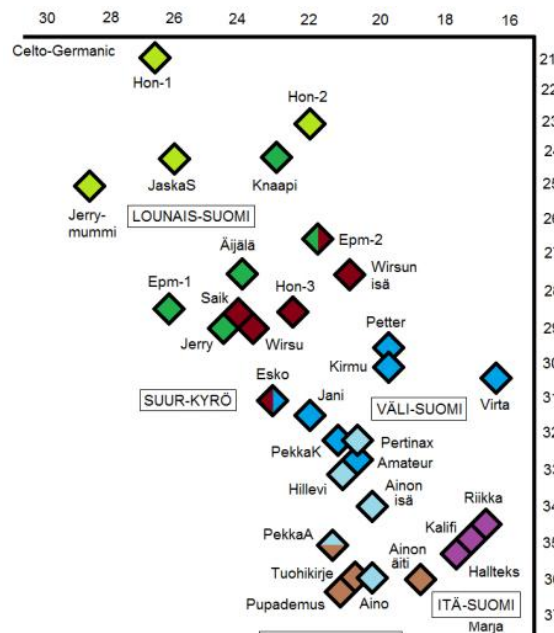
Single Population Sharing:		
#	Population (source)	Distance
1	FIN (Finn)	13.06
2	NRUS (North-Russian)	21.42
3	ETN (Estonian)	29.07
4	KM (Komi)	29.99
5	GER (German)	31.23
6	TTR (Tatar)	33.71
7	HNG (Hungarian)	34.28
8	MLD (Moldovian)	36.38
9	UKR (Ukrainian)	36.71
10	MRD (Erzya)	36.74
11	SLV (Slovenian)	37.09
12	RUS (Russian)	37.35
13	SLK (Slovak)	37.79
14	BLG (Bulgarian)	38.73
15	SWD (Swedish)	39.14

Mixed Mode Population Sharing:			
#	Primary Population (source)	Secondary Population (source)	Distance
1	75.9% FIN (Finn)	24.1% TTR (Tatar)	7.93
2	79.6% FIN (Finn)	20.4% MRD (Erzya)	9.39
3	79.1% FIN (Finn)	20.9% HNG (Hungarian)	9.76
4	86.1% FIN (Finn)	13.9% BSHK (Bashkir)	9.82
5	76.7% FIN (Finn)	23.3% KM (Komi)	9.83
6	80.9% FIN (Finn)	19.1% SLK (Slovak)	9.84
7	77.7% FIN (Finn)	22.3% GER (German)	9.94
8	81% FIN (Finn)	19% SLV (Slovenian)	10.03
9	82.7% FIN (Finn)	17.3% WUKR (West-Ukrainian)	10.07
10	82.3% FIN (Finn)	17.7% BSN (Bosnian)	10.08
11	90.1% FIN (Finn)	9.9% SAM (Sami)	10.12
12	81.4% FIN (Finn)	18.6% RUS (Russian)	10.14

Kaikki nämä tulossivut kannattaa tallettaa omalle koneelle: hiiren oikea painike > Tallenna nimellä. Eri alu-
eiden suomalaisia voidaan sitten vertailla, kuten kyseisen laskurin (jossa on suomalaisittain järkevät ja ha-
vainnolliset tekijät) kohdalla täällä: <http://www.elisanet.fi/alkupera/K12.pdf>

Ryhmäkohtaiset keskiarvot sekä vertailunäytteitä:

	Baltic-Finnic	East European	Celto-Germanic	North European	Iberian	"Uralic"
Itäsuomalaiset, itä?	38,3	20,0	16,3	7,4	3,8	12,0
Itäsuomalaiset, länsi?	35,0	20,8	17,0	6,8	2,9	11,3
Karjalais-kainuulaiset?	40,2	18,2	18,6	7,5	2,8	8,4
Suurkyröläis-itäiset	35,8	19,4	20,5	7,5	4,2	10,0
Hämäläis-itäiset	33,8	21,6	20,6	7,4	3,2	7,8
Välisuomalaiset	30,9	22,9	20,0	6,0	5,0	8,8
Suurkyröläiset	28,9	22,3	22,8	6,4	4,7	8,9
Lounaissuomalaiset, sisämaa?	27,1	22,7	24,0	5,7	6,4	7,5
Lounaissuomalaiset, rannikko?	23,3	24,5	25,8	4,9	5,8	7,1



Myös naapurikansoihin vertailu paljastaa havainnollisesti, ketkä muistuttavat geneettisesti ketäkin.

Terveysgeenit: Promethease ja SNPedia

Family Finder -datatiedostonsa voi ajaa myös Promethease-ohjelmassa, joka käy läpi kaikki ne genom-
kohdat, jotka on geenitutkimuksissa yhdistetty johonkin terveydentilaan. Esimerkiksi kohdassa rs3775291
minulla on toiselta vanhemmalta peritty A-alleeli, joka vähentää riskiä saada verkkokalvon rappeuma van-
hemmalla iällä. (<http://snpedia.com/index.php/Promethease>)

<p>Good news (hide)</p> <p>rs3775291(A,G) Magnitude: 2.1 Frequency: 47.8% Repute: Good References: 12</p>	<p>0.71x decreased risk for dry age related macular degeneration (hide)</p> <p>rs3775291 is a SNP in the TLR3 gene associated with an amino acid change in the corresponding protein. In the orientation as shown in dbSNP, the more common rs3775291(G) allele encodes a leucine while the rarer rs3775291(A) allele encodes a phenylalanine. rs3775291 has been associated with one form of age related macular degeneration (ARMD), specifically, the form known as 'dry' ARMD, also known as geographic atrophy. A study of 3 case-control groups of Americans of European descent, comprising about 900 ARMD patients in total, indicated that having one rs3775291(A) allele reduces the odds of having dry ARMD about 30%, and being a rs3775291(A,A) homozygote cuts your odds by more than half. Specifically, the odds ratio of having dry ARMD with one minor allele is 0.712 (CI: 0.50-1.0), and fo...</p>
<p>rs3746444(T,T) Magnitude: 2.1 Repute: Good References: 18</p>	<p>decreased cancer risk source (hide)</p> <p>microRNA 1,009 breast cancer cases and 1,093 controls in a population of Chinese women *rs3746444 in hsa-mir-499 A>G OR 1.25; 95% CI, 1.02-1.51 Genetic variants of miRNA sequences and non-small cell lung cancer survival. MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine. Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. SNPs in human miRNA genes affect biogenesis and function. Common genetic variants in pre-microRNAs are associated with risk of coal workers' pneumoconiosis. MicroRNA polymorphisms: a giant leap towards personalized medicine. Common genetic variants in pre-microRNAs and risk of gallbladder cancer in North Indian population. Association of pre-microRNAs genetic varia...</p>

Viiden dollarin lisämaksulla saa ajettua nopeasti laajan analyysin. Tulokset esitetään kategorioittain: Hyvät uutiset (Good news), Huonot uutiset (Bad news) jne.

Homotsygoottisuus

Homotsygoottinen merkitsee, että tietyssä kohdassa kummaltakin vanhemmalta peritty alleeli on sama. Eurooppalaisilla yli 70 % kaikista kohdista on homotsygoottisia. Suhteellisesti korkea homotsygoottisuus ei suoraan kerro sisäsiittoisuudesta, koska pienillä väestöillä (joita kaikki pohjoisen väestöt historiallisesti ovat) vain on geneettinen vaihtelu hieman vähentynyt. Homotsygoottisuuden voi selvittää täällä: <http://www.math.mun.ca/~dapike/FF23utils/roh.php>.

Yleisen ja kromosomikohtaisen homotsygoottisuuden lisäksi laskuri kertoo myös ROHien (*region of homozygosity*) määrän ja pituuden. ROHien pituus (pitkä on yli 10 Mb) kertovat luotettavammin sisäsiittoisuudesta, mutta ei ole syytä pelkoon, elleivät vanhempasi ole lähisukulaisia keskenään. Serkusavioliittoja harrastaviin kansoihin verrattuna suomalaiset ja eurooppalaiset yleensäkin eivät ole sisäsiittoisia. (Ks. <http://www.elisanet.fi/alkupera/Asutushistoria.pdf>)

Muinainen Suomi -foorumilla voi kysellä lisää ja keskustella testistä ja testituloksista!

<http://p1.foorumi.info/muinainensuomi/index.php>