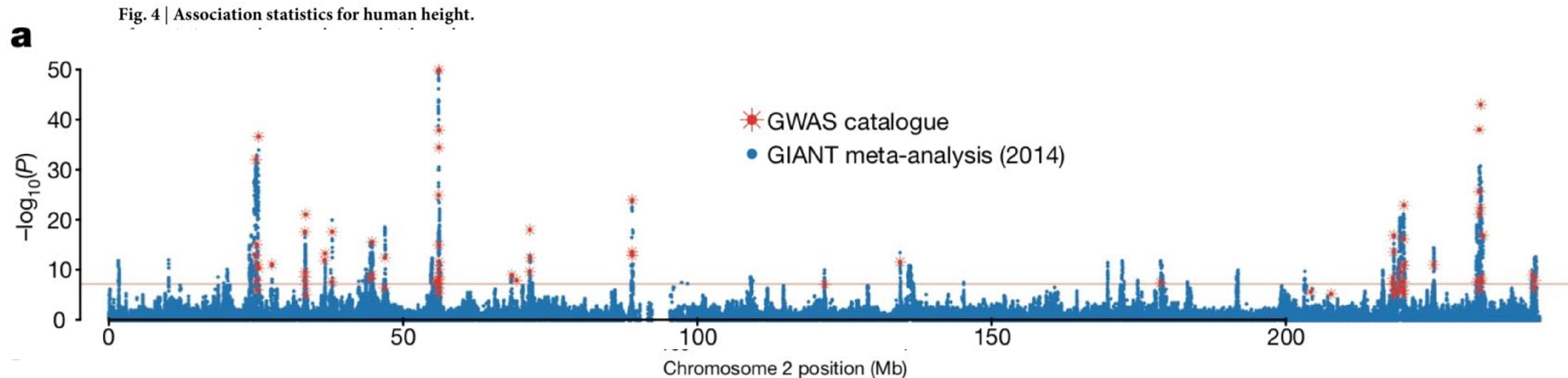


Genetyczne czynniki ryzyka

- Badania GWAS (*Genome-Wide Association Studies*) – asocjacje na skalę genomu
- Korelacja wariantów genetycznych z ryzykiem choroby
- Znajduje się wiele korelacji, ale żadna nie może być decydującym czynnikiem
- Tajemnica brakującej odziedziczalności

Wizualizacja GWAS - Manhattan plot



The UK Biobank resource with deep phenotyping and genomic data

Clare Bycroft^{1,13}, Colin Freeman^{1,13}, Desislava Petkova^{1,12,13}, Gavin Band¹, Lloyd T. Elliott², Kevin Sharp², Allan Motyer³, Damjan Vukcevic^{3,4}, Olivier Delaneau^{5,6,7}, Jared O'Connell⁸, Adrian Cortes^{1,9}, Samantha Welsh¹⁰, Alan Young¹¹, Mark Effingham¹⁰, Gil McVean^{1,11}, Stephen Leslie^{3,4}, Naomi Allen¹¹, Peter Donnelly^{1,2,14} & Jonathan Marchini^{1,2,14*}

Istotność wartości p w GWAS

- $p < 5 \cdot 10^{-8}$ próg zwyczajowy (dla populacji europejskiej)
- można też wyliczyć istotność korzystając z różnych poprawek statystycznych
 - Bonferroniego (prosta, ale zbyt restrykcyjna)
 - Benjaminiiego–Hochberga (FDR - False Discovery Ratio)

Dane

- Informacja o osobach (w tym ew. rodowody) + fenotypy
- Genotypy
- Pozycje miejsc zmiennych w genomie (mapa)

Format PLINK

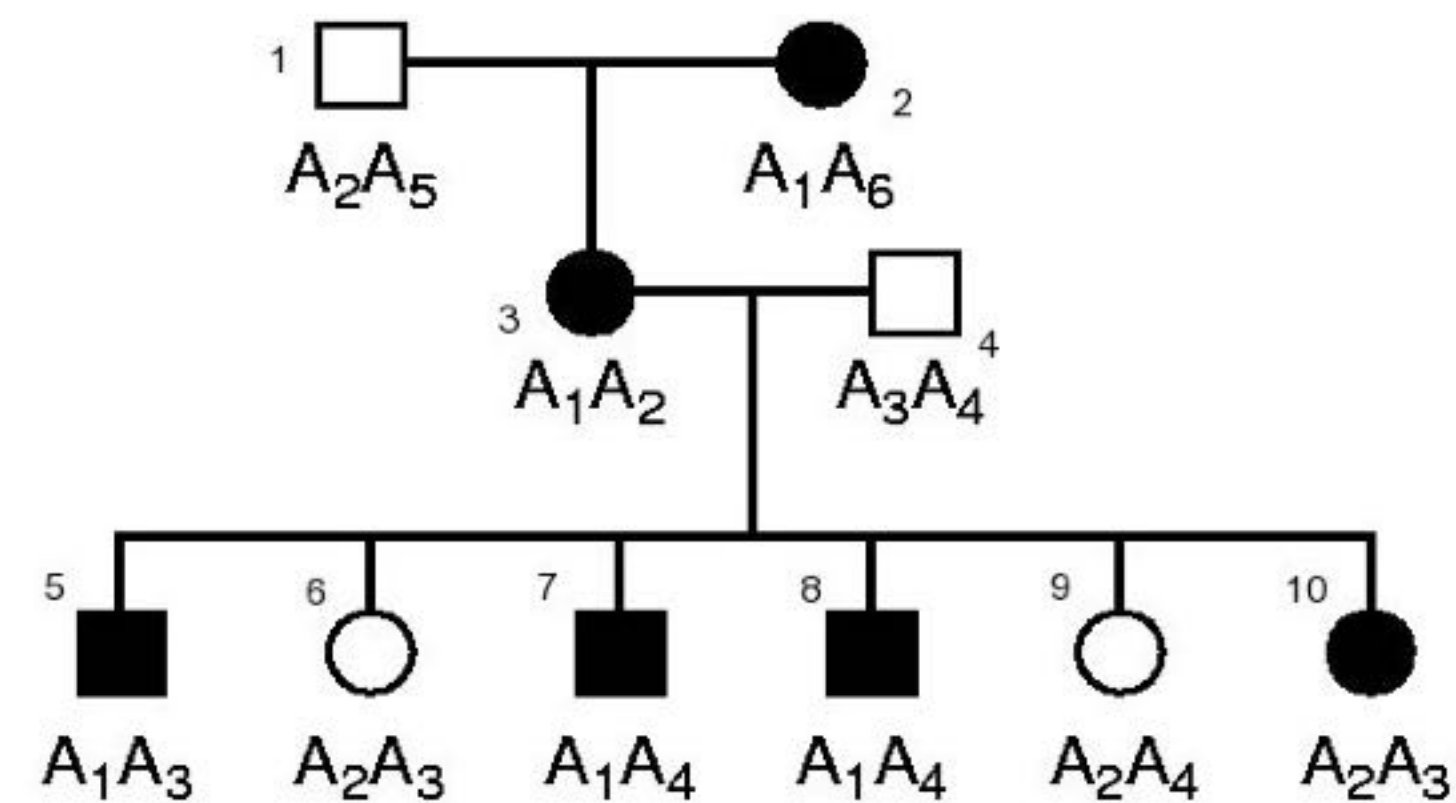
Dane:

- rodowód i genotypy: plik .ped
- mapa: plik .map

Dane w formacie binarnym

- rodowód: plik .fam (pierwsze 6 kolumn pliku .ped)
- mapa: plik .bim (dodatkowe dane: allele i MAF)
- genotypy markera: plik .bed (skompresowana wersja pliku tekstowego)

Kodowanie rodowodu



Płeć: 1 mężczyzna, 2 kobieta, 0 nieznana
Choroba: 1 zdrowa(y), 2 chora(y), 0 nieznana
0 zawsze oznacza nieznane/brak danych!!!

Rodzina	osoba	ojciec	matka	płeć	choroba	marker1a1	marker1a2
001	1	0	0	1	1	A	A
001	2	0	0	2	2	A	C
001	3	1	2	2	2	C	C
001	4	0	0	1	1	3	4
001	5	4	3	1	2	A	C

Liczba spacji nieistotna (minimum 1)

Pliki mapy

CHROMOSOME	MARKER	POSITION
1	Marker1	0

format .map

1	rs12073590	0.029735	1205155	C	A
1	rs6685064	0.029785	1211292	T	C
1	rs61559999	0.030045	1235792	T	C
1	rs62623580	0.030111	1254255	A	G

format .bim

Inne częste formaty

VCF

```
##fileformat=VCFv4.2
##contig=<ID=2,length=51304566>
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE1	SAMPLE2	SAMPLE3	SAMPLE4	SAMPLE5	SAMPLE6	SAMPLE7
2	81170	.	C	T	.	.	AC=9;AN=7424	GT:DP:GQ	0/0:4:12	0/0:3:9	0/1:1:3	0/1:9:24	1/0:4:12	0/0:5:15	0/0:4:12
2	81171	.	G	A	.	.	AC=6;AN=7446	GT:DP:GQ	0/1:4:12	0/0:3:9	0/0:1:3	0/0:9:24	0/1:4:12	0/1:5:15	0/0:4:12
2	81182	.	A	G	.	.	AC=5;AN=7506	GT:DP:GQ	0/0:5:15	0/0:4:12	0/0:5:15	0/0:9:24	0/0:4:12	0/0:4:12	0/0:4:12
2	81204	.	T	G	.	.	AC=2;AN=7542	GT:DP:GQ	1/0:5:15	0/0:9:27	0/0:10:30	0/0:15:39	0/0:9:27	1/0:13:39	0/1:14:42

BCF

2	81170	.	C	T	.	.	AC=9;AN=7424	GT:0/0:0/0:0/1:0/1:1/0:0/0:0/0	DP:4:3:1:9:4:5:4	GQ:12: 9: 3:24:12:15:12
2	81171	.	G	A	.	.	AC=6;AN=7446	GT:0/1:0/0:0/0:0/0:0/1:0/1:0/0	DP:4:3:1:9:4:5:4	GQ:12: 9: 3:24:12:15:12
2	81182	.	A	G	.	.	AC=5;AN=7506	GT:0/0:0/0:0/0:0/0:0/0:0/0:0/0	DP:5:4:5:9:4:4:4	GQ:15:12:15:24:12:12:12
2	81204	.	T	G	.	.	AC=2;AN=7542	GT:1/0:0/0:0/0:0/0:0/0:1/0:0/1	DP:5:9:10:15:9:13:14	GQ:15:27:30:39:27:39:42

plink (w wersji 1.9b lub wyższej) odczytuje pliki vcf i bcf i przekształca na .bed/.bim/.fam

Program plink

`plink --file nazwa`

wczytuje pliki nazwa.ped nazwa.map

`plink --bfile nazwa`

wczytuje pliki nazwa.fam nazwa.bim nazwa.bed

Konwersja formatów

```
plink --file przyklad --make-bed --out nowanazwa
```

```
plink --bfile przyklad --recode --tab --out nowanazwa
```

```
plink --vcf przyklad.vcf --make-bed --out nowanazwa
```

W pliku VCF nie ma osobnych ID osoby i rodziny! Domyślnie każda osoba jest traktowana jako będąca z odrębnej rodziny, chyba że w jej ID jest znak '_', wtedy służy jako delimiter.

Przykładowe możliwości plink

`--assoc`

analiza asocjacji

`--assoc --adjust`

j.w. z poprawkami statystycznymi

`--out nowanazwa`

Przykładowe możliwości plink

`--keep / --remove lista`

zachowuje / usuwa osoby o ID z pliku lista

`--keep-fam --remove-fam`

j.w., tylko dla ID rodziny a nie osoby

`--extract / --exclude lista`

zachowuje / usuwa markery o nazwach z pliku lista

+ wiele innych (filtrowanie wg. chromosomów, MAF, SNP vs. indele, i wiele innych)

`--make-bed --out nowanazwa`

zapisuje efekt filtrowania w nowym pliku

Interpretacja wg. bazy genomu człowieka

- Na stronie ncbi.nlm.nih.gov, w sekcji *Genetics & Medicine*, baza ClinVar, narzędzie *Variation Viewer*
- Formaty zapytania (przykłady)
 - rs28693734 - identyfikator SNP (z bazy dbSNP), o ile dostępny
 - chr1:215210111 - chromosom:pozycja (uwaga na genom referencyjny)