

Practical: Effect of the IL6R gene on IL-6R concentration

Family based association & follow up analyses

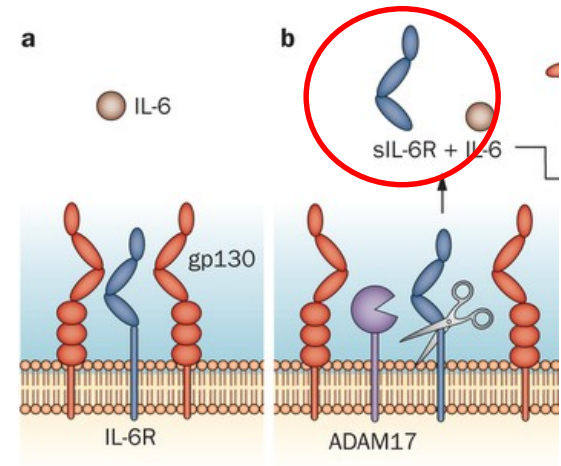
Behav Genet (2014) 44:368–382
DOI 10.1007/s10519-014-9656-8

ORIGINAL RESEARCH

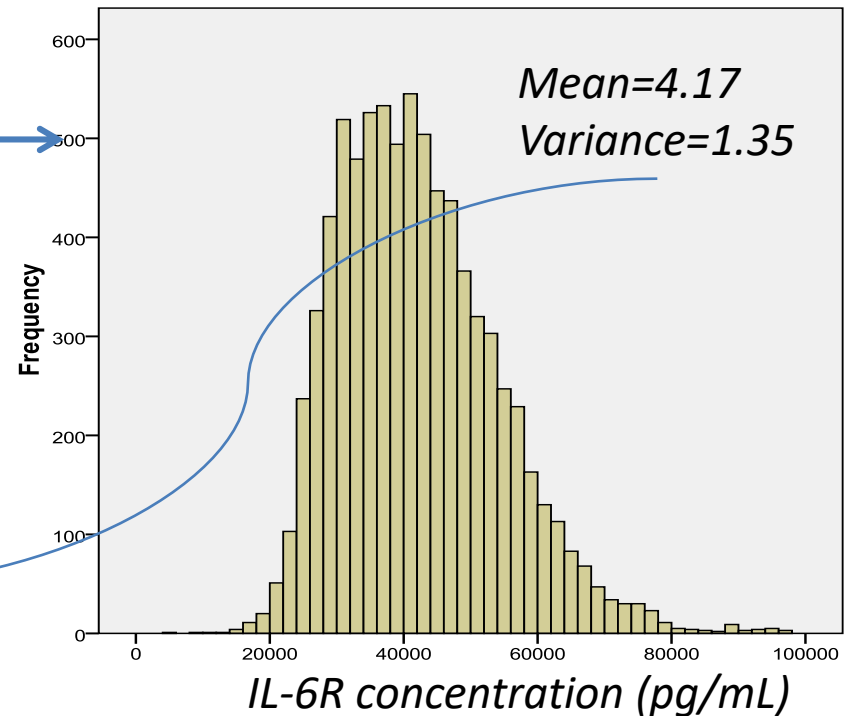
The Contribution of the Functional *IL6R* Polymorphism rs2228145, eQTLs and Other Genome-Wide SNPs to the Heritability of Plasma sIL-6R Levels

Jenny van Dongen · Rick Jansen · Dirk Smit · Jouke-Jan Hottenga · Hamdi Mbarek ·
Gonneke Willemsen · Cornelis Klufft · AAGC Collaborators · Brenda W. J. Penninx ·
Manuel A. Ferreira · Dorret I. Boomsma · Eco J. C. de Geus

- We measured soluble IL-6R concentration in blood in ~5000 individuals (from the Netherlands Twin Register)



- sIL-6R concentration in blood is a **quantitative trait**



Estimated in Mx

Genetics → IL-6R concentration → common disease

- IL-6R protein is encoded by the ***IL6R*** gene (chromosome 1)
- *IL6R* gene important for **several common diseases**
 - Asthma¹
 - Coronary heart disease²
 - Type 1 diabetes³

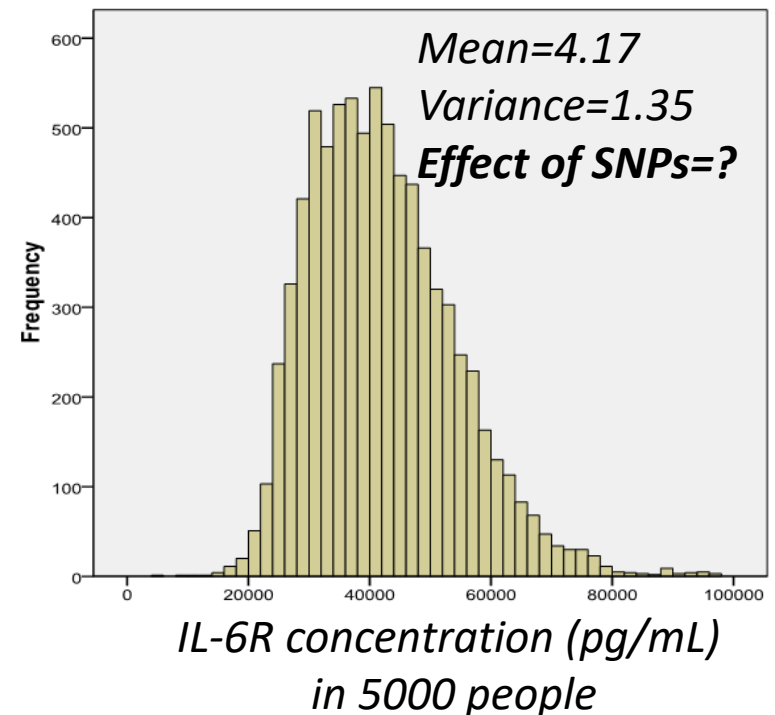
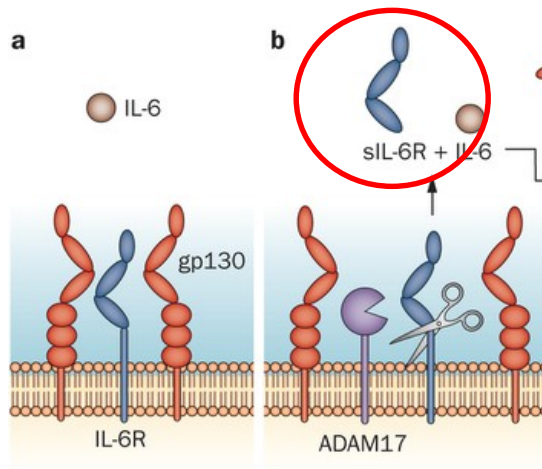
¹Ferreira M.A. *et al* Lancet 2011

²*IL6R* consortium Lancet 2012

³Ferreira R.C. *et al* PLoS Genetics 2013

Genetics → IL-6R concentration

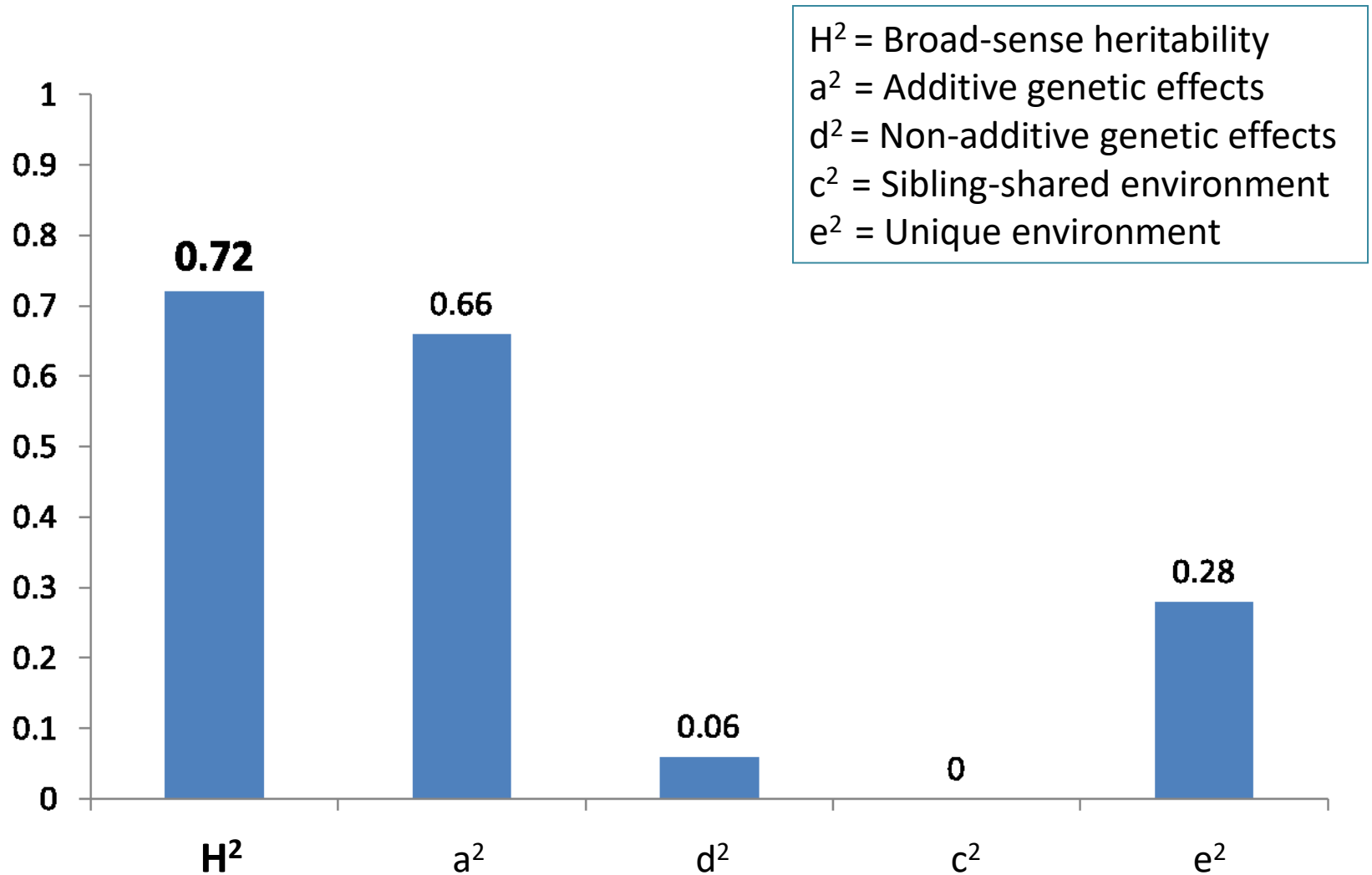
- Functional studies: SNP affects sIL-6R production (non-synonymous missense)
 - Rs2228145: **Large effect** on sIL-6R level (allele C increases sIL-6R concentration)
 - **How much of all variance does this SNP explain?**
 - **How much of the variance is explained by other variants?**



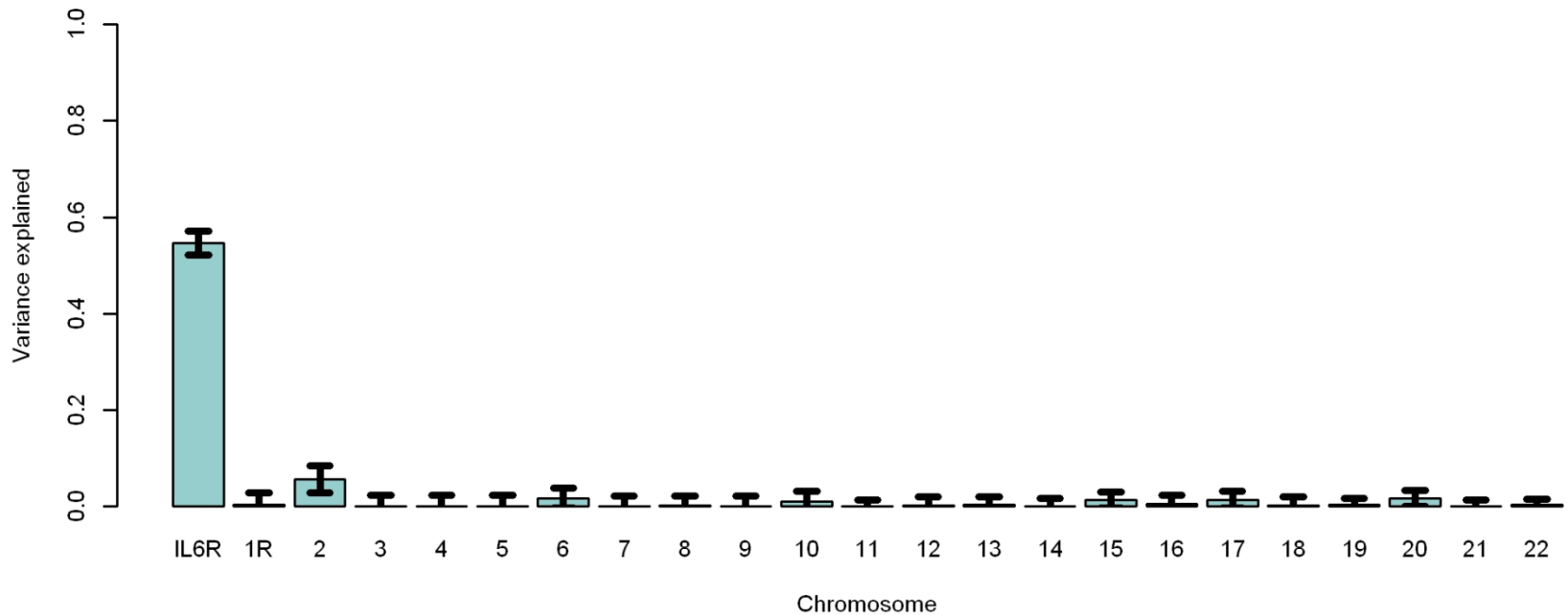
Methods

- We measured IL-6R concentration in ~5000 **twins & parents & siblings**
- We estimated **Heritability**: Variance of sIL-6R level explained by total genetic effects (Mx)
- We measured genome-wide SNP genotypes of the same subjects:
 - How much variance is explained by **all SNPs in the genome** (Genomewide-complex trait analysis, GCTA)?
 - How much variance is explained by **all genetic variation in the *IL6R* gene** (linkage analysis)?
 - How much variance is explained by **the SNP rs2228145?**
 - *GWAS & eQTL analysis to identify novel variants associated with *IL6R* protein and gene expression levels*

Heritability of sIL-6R level (twin-family data)



Variance explained by chromosome-wide SNPs (GCTA)

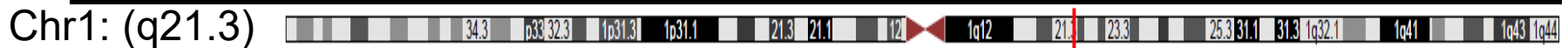
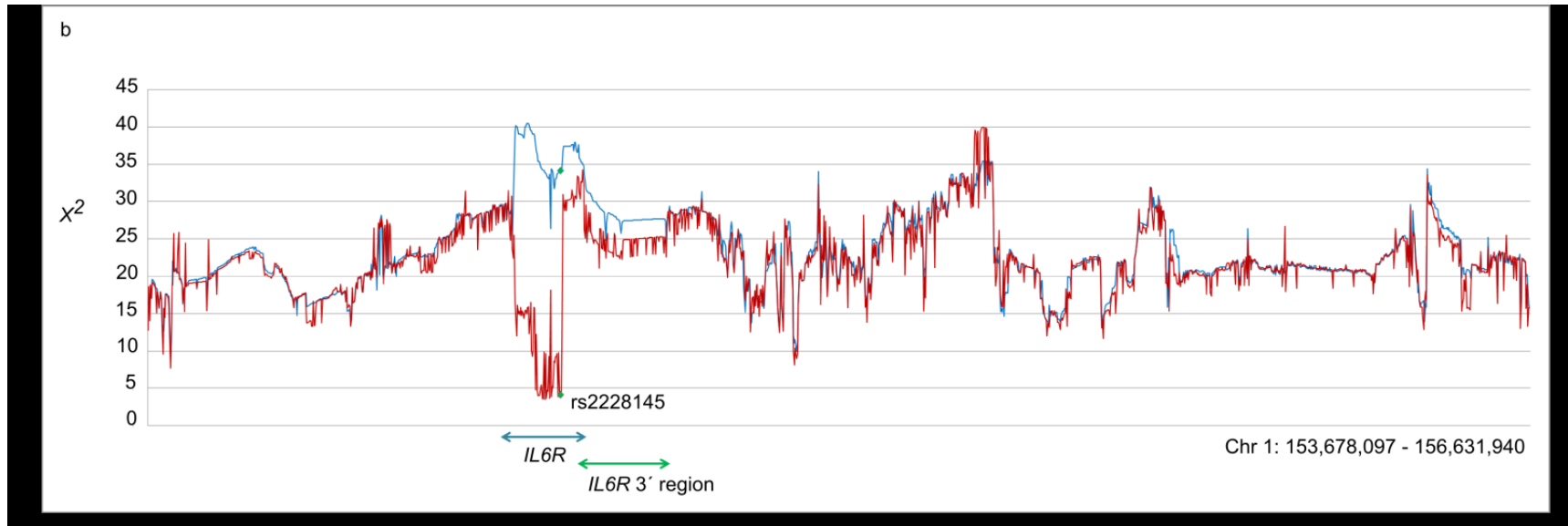


SNPs in the *IL6R* gene on Chromosome 1 (+/- 10MB): **54.7 %** (SE=2.5%)

Combined linkage and association analysis (qtdt)

— Chi-squared from linkage test

— Chi-squared from linkage test – while modeling association for individual SNPs



IL6R region:

1. Variance explained by linkage (V_A/V_{total}): **69 %**
2. Variance explained by linkage after correction for rs2228145: **..%**

Practical

Original analyses paper

GWAS on 4846 Ss
(twin families)
in plink with --family

biometrical model

Note: for --fam,
use plink 1.07!
This option is no
longer supported
in newer plink
versions

1 SNP (rs2228145)

This practical

3000 related Ss
Association analysis in R gee

2096 unrelated Ss
• Association analysis
• Allele frequencies
• Genotype frequencies and means
In plink 1.9

Imputed genotype
data (best guess)

biometrical model

/faculty/jenny/2019/tuesday

```
mkdir practical_family
```

```
cp -r /faculty/jenny/2019/tuesday/* practical_family
```

```
cd practical_family
```

Plink –association analysis in unrelated individuals

- Data

plink_covar_unrel.txt

rs2228145_plink_unrel.map

rs2228145_plink_unrel.ped

- Covariates (plink_covar.txt)

zage	= z-score of age
PC1_NL PC2_NL PC3_NL	= Dutch ancestry PCs
PC3_chip_effect PC5_chip_effect PC1_buccal	= PCs to correct for chip and DNA source (buccal/blood)

- Run association test (1 SNP) - sIL6R, correcting for 7 covariates

- We use plink version 1.9

```
plink --file rs2228145_plink_unrel --covar plink_covar_unrel.txt --linear --assoc --qt-means --freq
```

Output plink

- `plink.qassoc.means (--qt-means)`

```
CHR SNP VALUE Q1 Q2 Q3  
1 rs2220145 FE20 C C A A  
1 rs2220145 COM3 200 86 60  
1 rs2220145 FE21 0.165 0.478 0.579  
1 rs2220145 FE20 5.628 4.527 3.224  
1 rs2220145 SD 1.068 0.821 0.624
```

- `plink.freq (--freq)`

```
CHR SNP AL AL2 HF NCHROMS  
1 rs2220145 C A 0.5702 2539
```

- `plink.assoc.linear (--linear --assoc)`

```
CHR SNP BP AL TEST MISS BETA STAT P  
1 rs2220145 154426970 C A0 1010 1.216 43.36 7.653e-202  
1 rs2220145 154426970 C C01 1010 0.1599 0.873 1.244e-15  
1 rs2220145 154426970 C C02 1010 -2.259 -1.001 0.3169  
1 rs2220145 154426970 C C03 1010 -2.66 -0.9670 0.3333  
1 rs2220145 154426970 C C04 1010 -0.4344 -0.1404 0.8804  
1 rs2220145 154426970 C C05 1010 -4.206 -2.591 0.009635  
1 rs2220145 154426970 C C06 1010 -0.02435 -0.01024 0.9054  
1 rs2220145 154426970 C C07 1010 10.06 0.3405 0.3963
```

Gee – association analysis in family data (related individuals)

- We will use the R-package gee to test the association between our SNP and sIL-6R, in data from family members
- Open the R-script [association_rs2228145_gee.r](#) (click on it, it will open in R-studio)
- Run the script line by line

Gee – association analysis in family data (related individuals)

- Data

plink_covar.txt

rs2228145_plink.map

rs2228145_plink.ped

- Covariates (plink_covar.txt)

zage

= z-score of age

PC1_NL PC2_NL PC3_NL

= Dutch ancestry PCs

PC3_chip_effect PC5_chip_effect PC1_buccal
source (buccal/blood)

= PCs to correct for chip and DNA

- R gee command

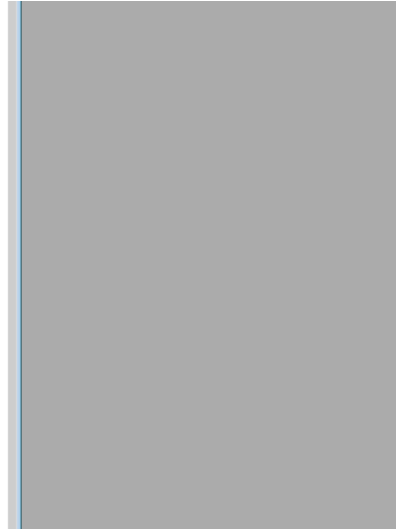
```
gee(sIL6R~genonum + zage + PC1_NL + PC2_NL + PC3_NL + PC3_chip_effect +  
PC5_chip_effect + PC1_buccal,data=data, id=FAMID, family=gaussian,  
corstr="exchangeable", maxiter=100, na.action=na.omit)
```

R output

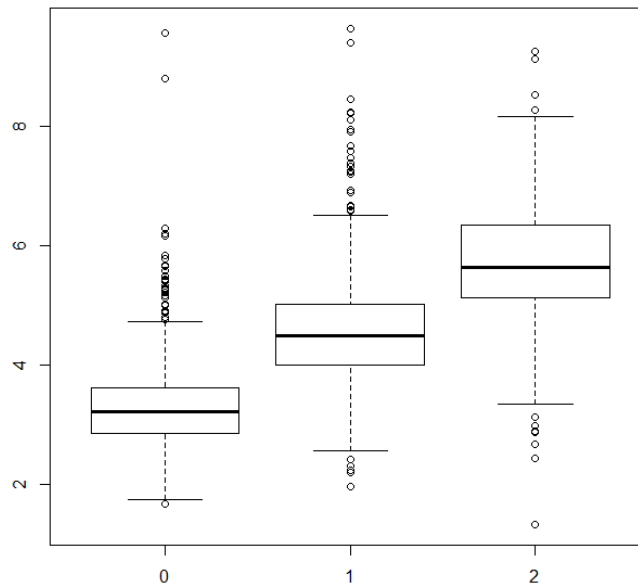
```
> results

```

	Estimate	RobustSE	Robusta	Pval
(Intercept)	3.2009476	0.03101160	103.1898373	0.00000e+00
genom	1.2270942	0.02568134	47.78158138	0.00000e+00
age	0.17856874	0.01688357	10.61080906	2.65899e-26
PC1_WU	-2.22687489	1.91889437	-1.16880746	2.457202e-01
PC2_WU	-3.45382650	1.57648882	-1.94932270	1.801404e-01
PC3_WU	0.03283707	2.7404517	0.01189414	9.804383e-01
PC3_chip_effect	-3.62049889	1.49874450	-2.41537134	1.571817e-02
PC5_chip_effect	-1.04564505	1.62689607	-0.61621406	5.377533e-01
PC1_housal	11.61605341	9.75824743	1.19026119	2.338438e-01



IL6R concentration



Number of C alleles (0=AA, 1=AC, 2=CC)

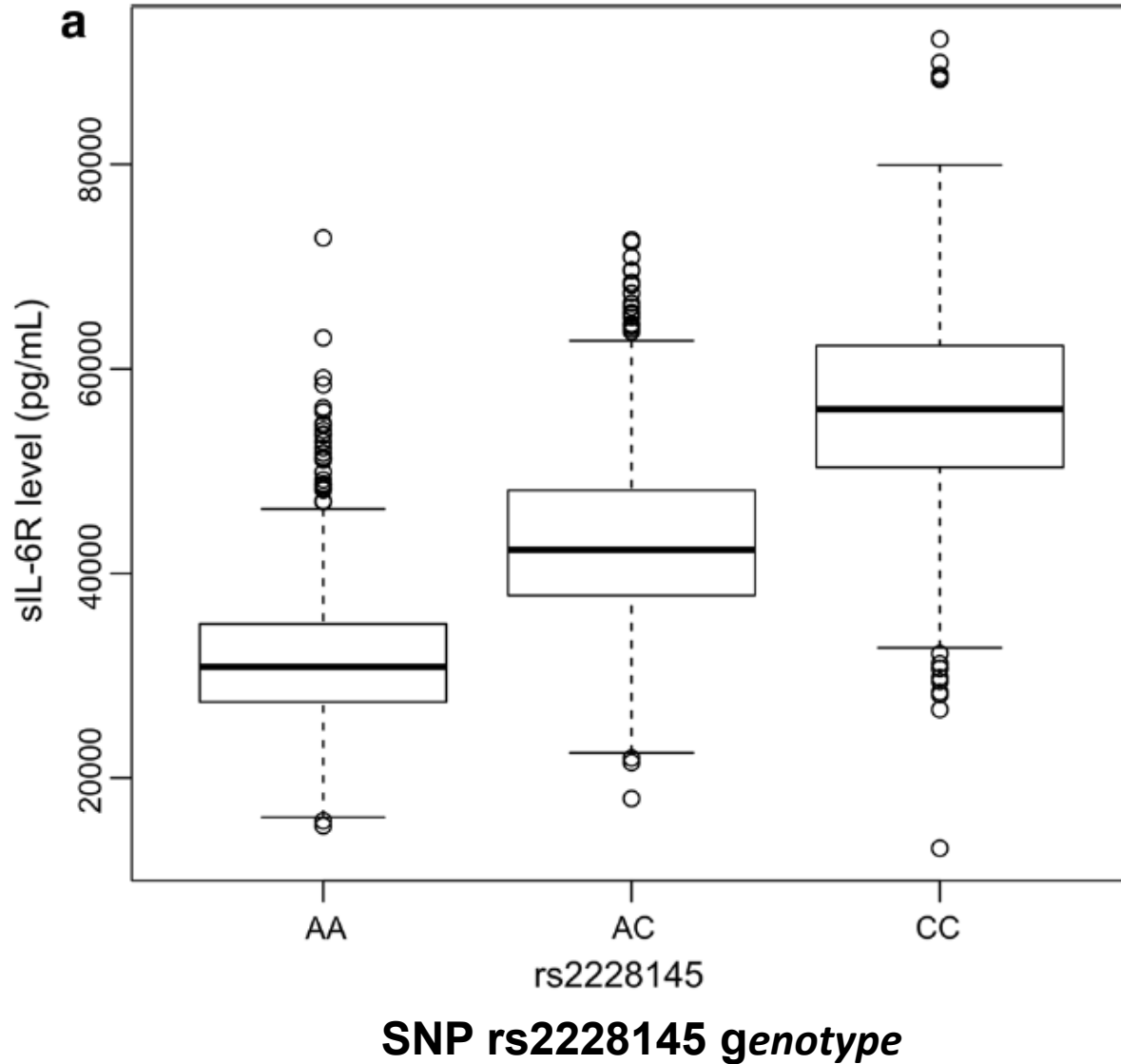
Note

- Allele-frequency estimation based on data from related individuals
- Sib-pair (David Duffy)
- <https://genepi.qimr.edu.au/staff/davidD/#sib-pair>

Biometrical model

Rs2228145: **Large effect** on sIL-6R level (allele C increases sIL-6R concentration)

sIL-6R concentration



Exercise: *Effect of the IL6R gene on IL-6R concentration*

INFORMATION

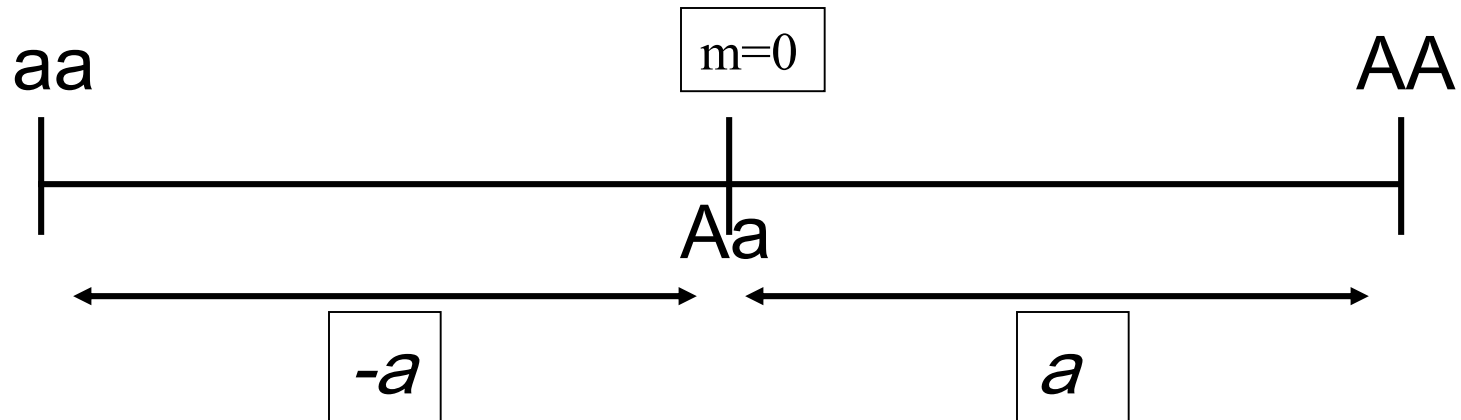
- The SNP (single nucleotide polymorphism) has 2 alleles:
 - Minor allele: C, frequency: $p=0.39$
 - Major Allele: A, frequency: $q =0.61$
- Mean IL-6R concentration of each genotype:
 - CC: 5.698 (10^{-8} g/mL)
 - CA: 4.418 (10^{-8} g/mL)
 - AA: 3.238 (10^{-8} g/mL)
- Total Variance of IL-6R concentration=1.35

QUESTIONS (Falconer & MacKay; 1996: Introduction to quantitative genetics)

1. Calculate genotypic values (a and d) (page 109)
2. Calculate the genotype frequencies (page 7)
3. Calculate the mean IL6-R concentration in the population (page 110)
4. Calculate how much of the variance is explained by this SNP
(*Variance= Sum of squared deviations from the mean*)

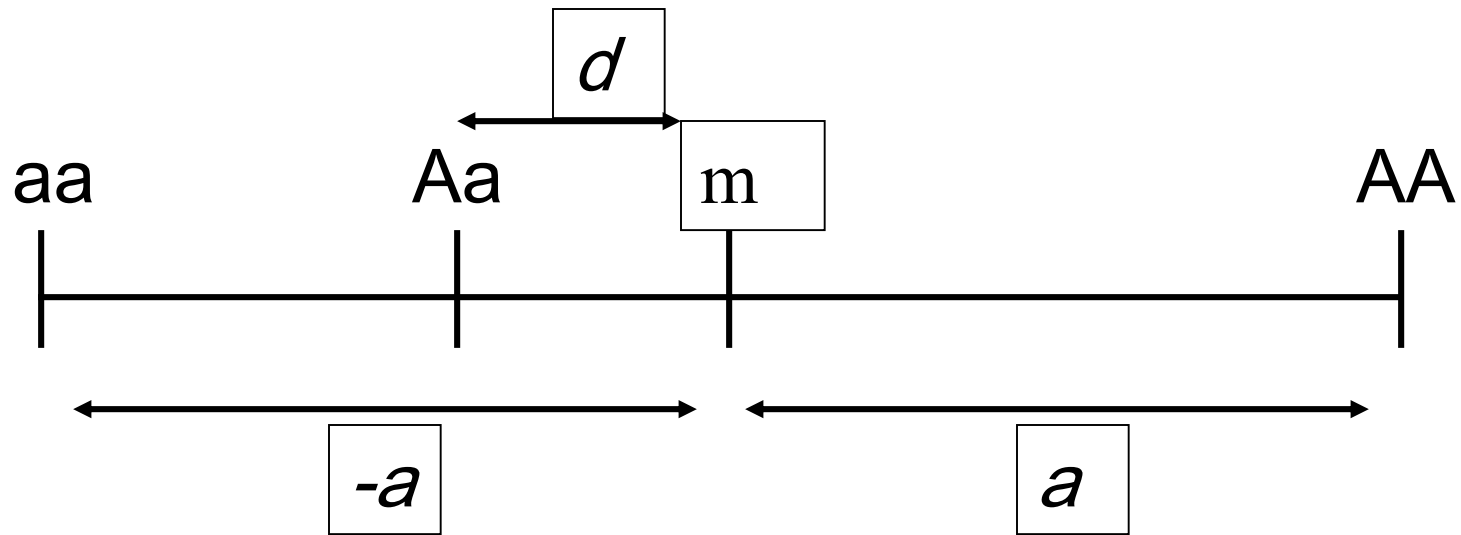
Extra: Calculate the average effect of the alleles (page 113)

Model: gene with 2 alleles A and a
and 3 genotypes AA, Aa and aa



The difference on a quantitative scale between AA and aa is $2a$.
The middle (m) is zero and the value of Aa is 0 (no dominance).

Model: gene with 2 alleles A and a
and 3 genotypes AA, Aa and aa



The deviation from m (middle) of the heterozygote Aa is d :
partial dominance.

Genotype (i)	AA	Aa	aa
Frequency (f)	p^2	$2pq$	q^2
Genotypic effect (x)	a	d	-a

Mean?

Genotype (i)	AA	Aa	aa
Frequency (f)	p^2	$2pq$	q^2
Genotypic effect (x)	a	d	-a
f * x	$p^2 a$	$2pqd$	$-q^2 a$

mean: $p^2 a + 2pqd - q^2 a =$ (recall $p+q = 1$)

$$a(p^2 - q^2) + 2pqd =$$

$$a(p-q)(p+q) + 2pqd =$$

$$\mathbf{Mean = a(p-q) + 2pqd}$$

$a(p-q)$: attributable to homozygotes

$2pqd$: attributable to heterozygotes

Genotype (i)	AA	Aa	aa
Frequency (f)	p^2	$2pq$	q^2
Genotypic effect (x)	a	d	-a
$f * x$	$p^2 a$	$2pqd$	$- q^2 a$

mean: $p^2 a + 2pqd - q^2 a = a(p-q) + 2pqd$

Variation: $2pq[a+d(q-p)]^2 + (2pqd)^2$

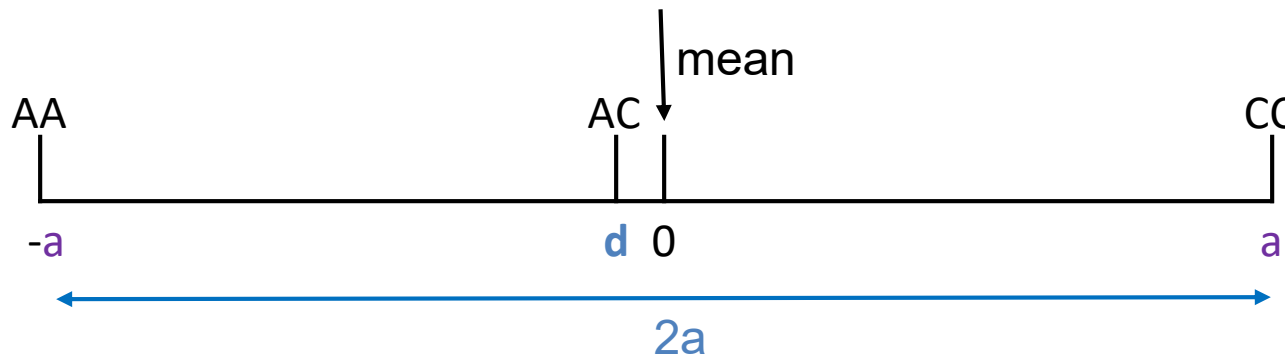
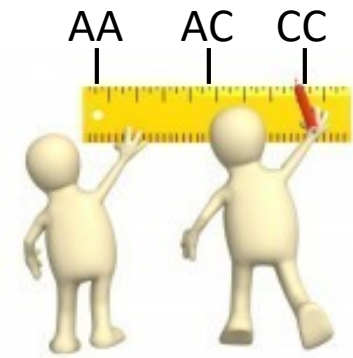
Population variation depends on 'a' (difference between homozygote individuals), 'd' (deviation of heterozygote persons from zero) and on allele frequency (p & q).

Mean IL-6R concentration of each genotype:

CC: 5.698 / CA: 4.418 / AA: 3.238 (10^{-8} g/mL)

Total Variance of IL-6R concentration=1.35

Frequencies: C, frequency: $p=0.39$ / A, frequency: $q=0.61$



QUESTIONS (Falconer & MacKay; 1996: Introduction to quantitative genetics)

1. Calculate genotypic values (a and d) (page 109)
2. Calculate the genotype frequencies (page 7)
3. Calculate the mean IL6-R concentration in the population (page 110)
4. Calculate how much of the variance is explained by this SNP

(*Variance= Sum of squared deviations from the mean*)

extra: Calculate the average effect of the alleles (page 113)