MUNI
MED

Analyze of the genome – GWAS and GRS

Structure

- Terminology
- The human genome project
- HapMap
- GWAS
- GWAS and oral cavity disease
- GRS

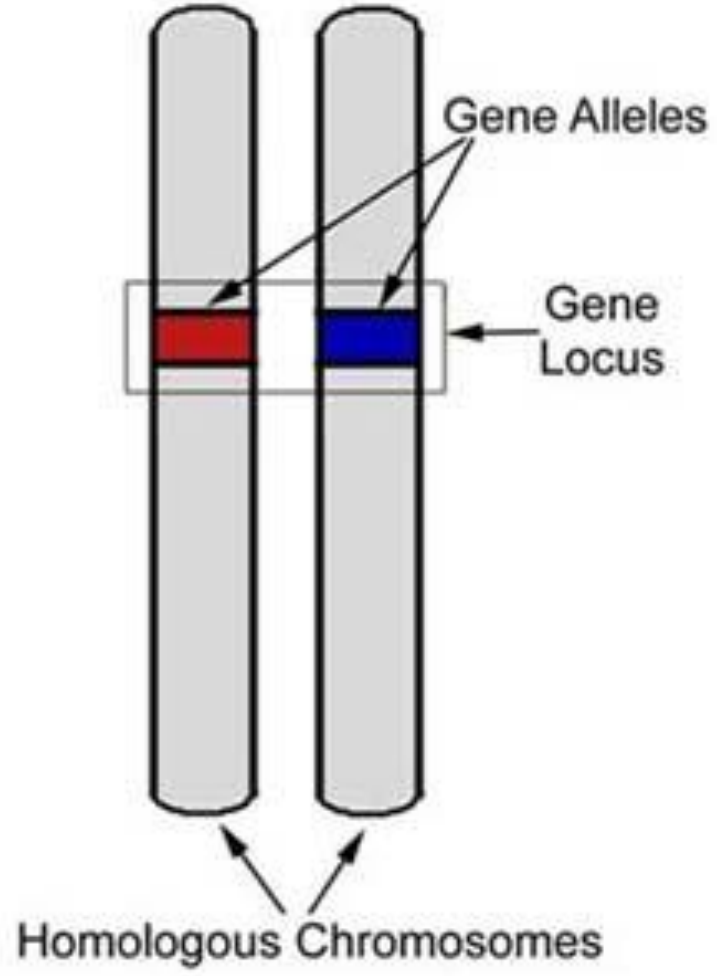
Terminology

- Allele
- Locus
- Single nucleotide polymorphism (SNP)
- Haplotype
- Linkage disequilibrium (LD)
- Imputation
- Genome wide association studies (GWAS)
- Genetic risk score (GRS)

MUNI MED

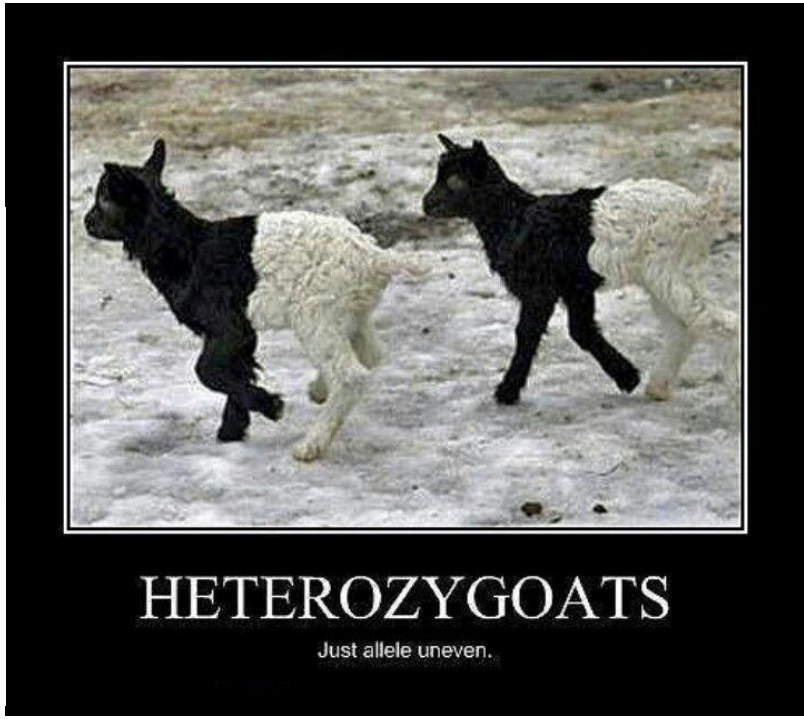
Allele a locus

- Allele is specific variant of the gene
- Locus determine specific position on the chromosome



MUNI MED

Single nucleotide polymorphism (SNP)

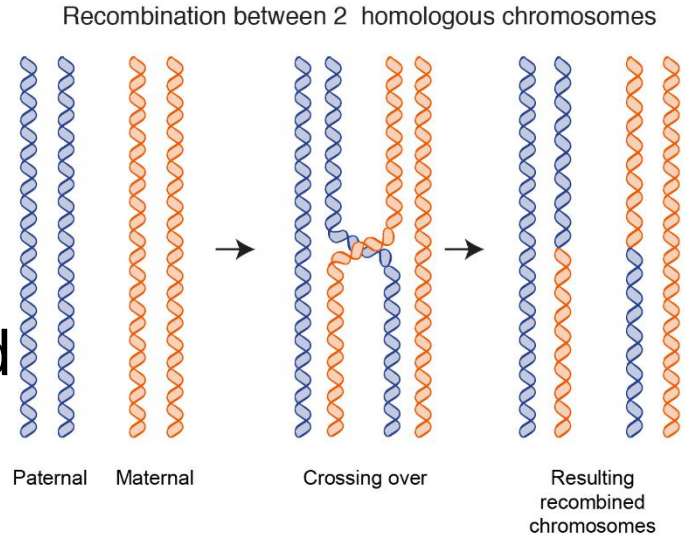


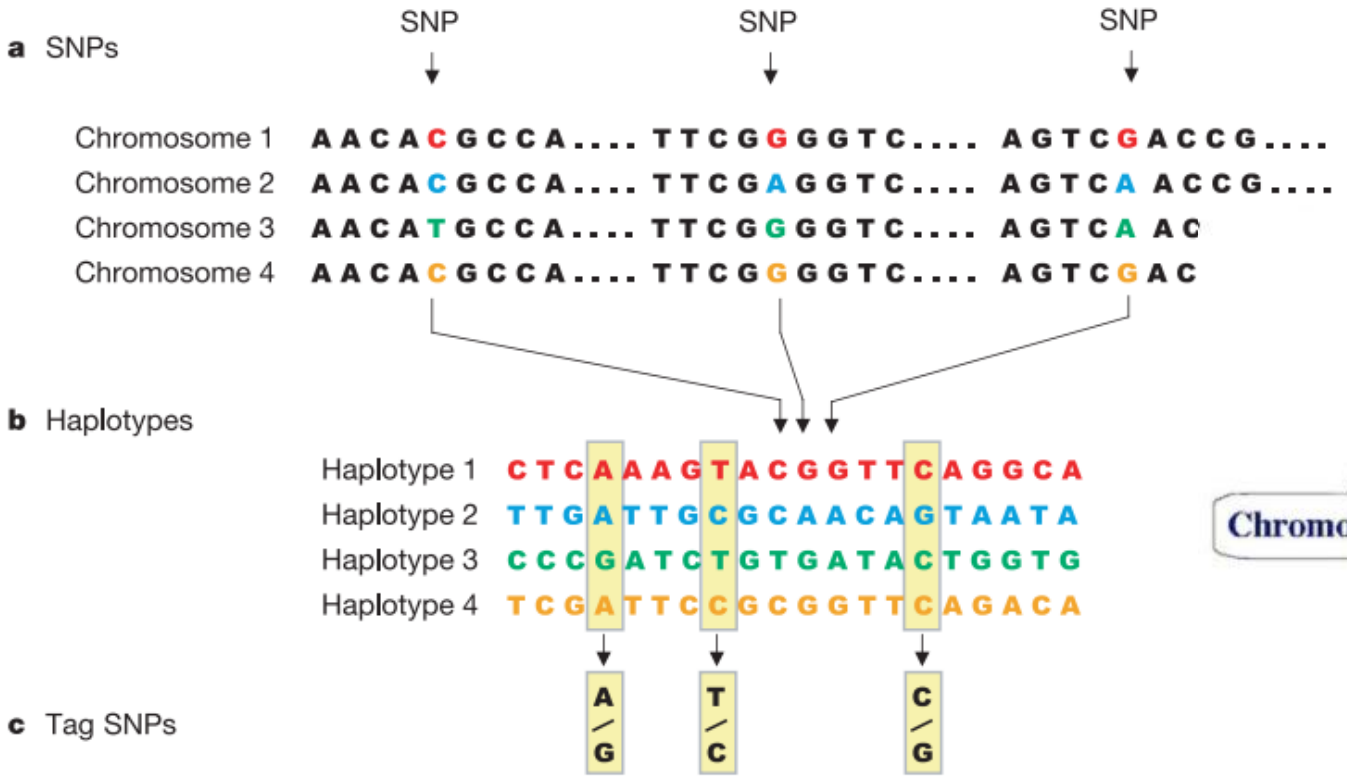
One nucleotide change with frequency higher than 1% in given population. This change does not have to impact function of the gene or protein.

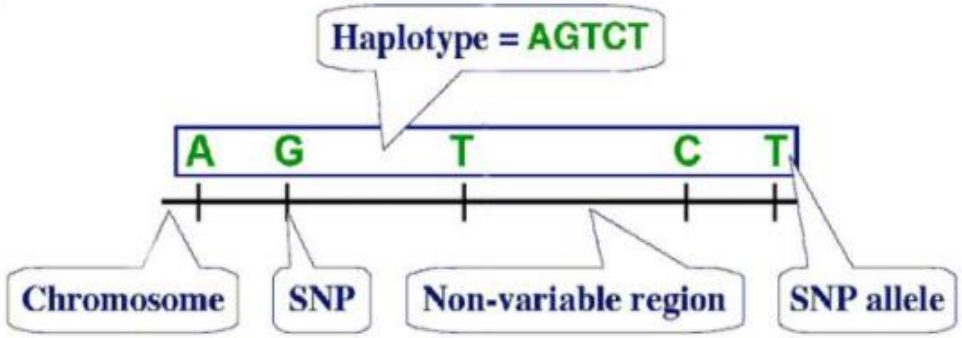
MUNI MED

Haplotype

– It is combination of alleles on different parts of the DNA (usually one chromosome or its part) which are inherited together

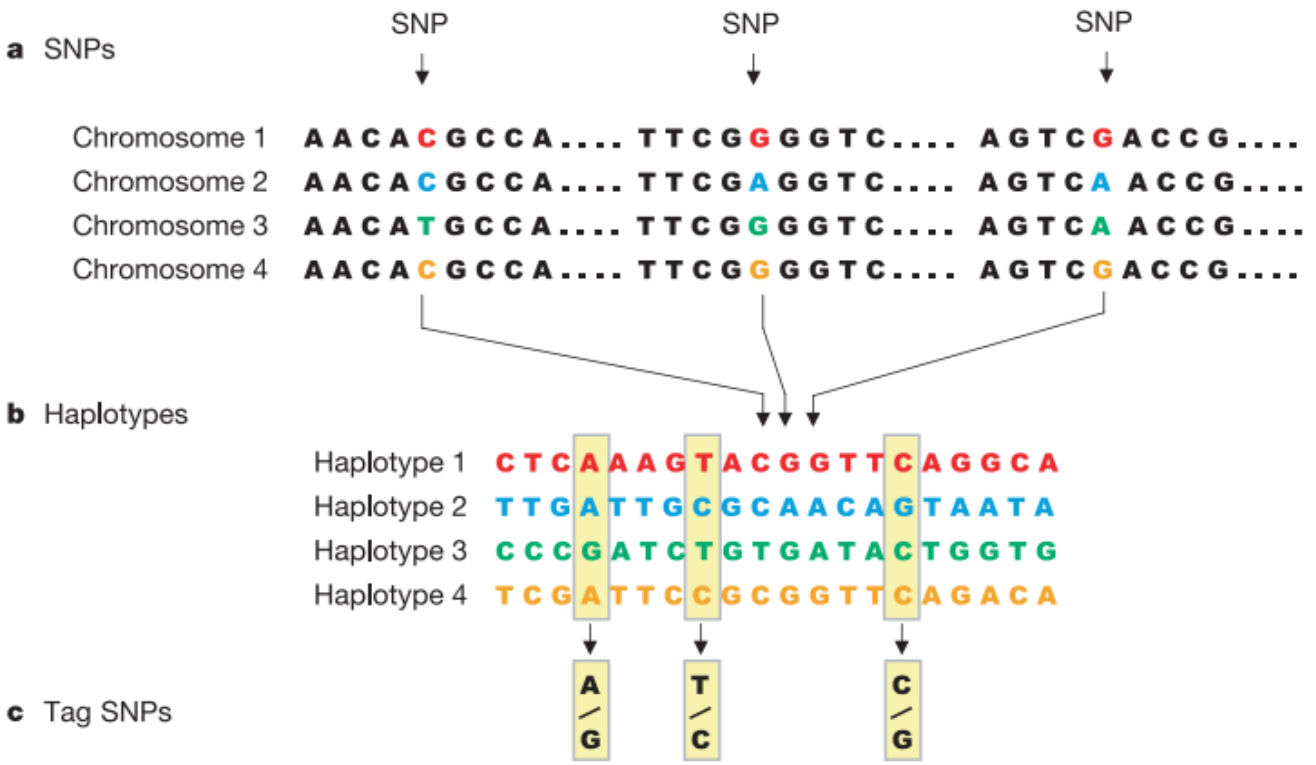






MUNI MED

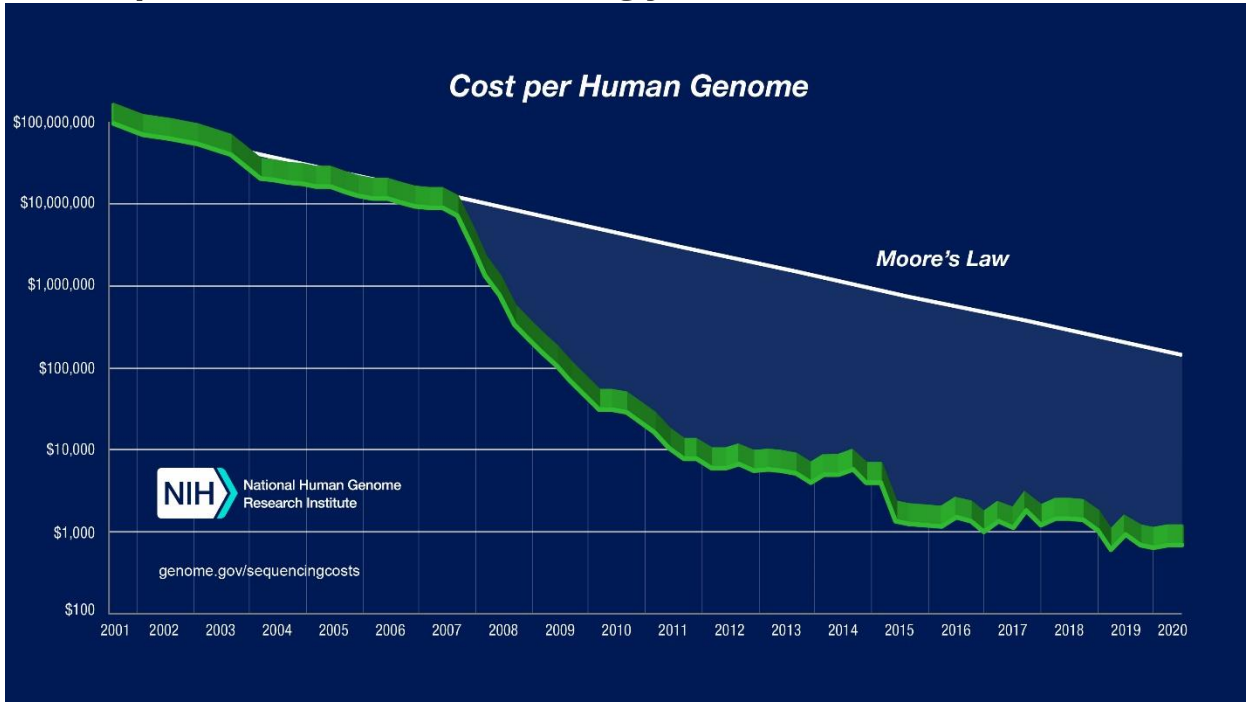
Linkage disequilibrium (LD) and imputation



MUNI MED

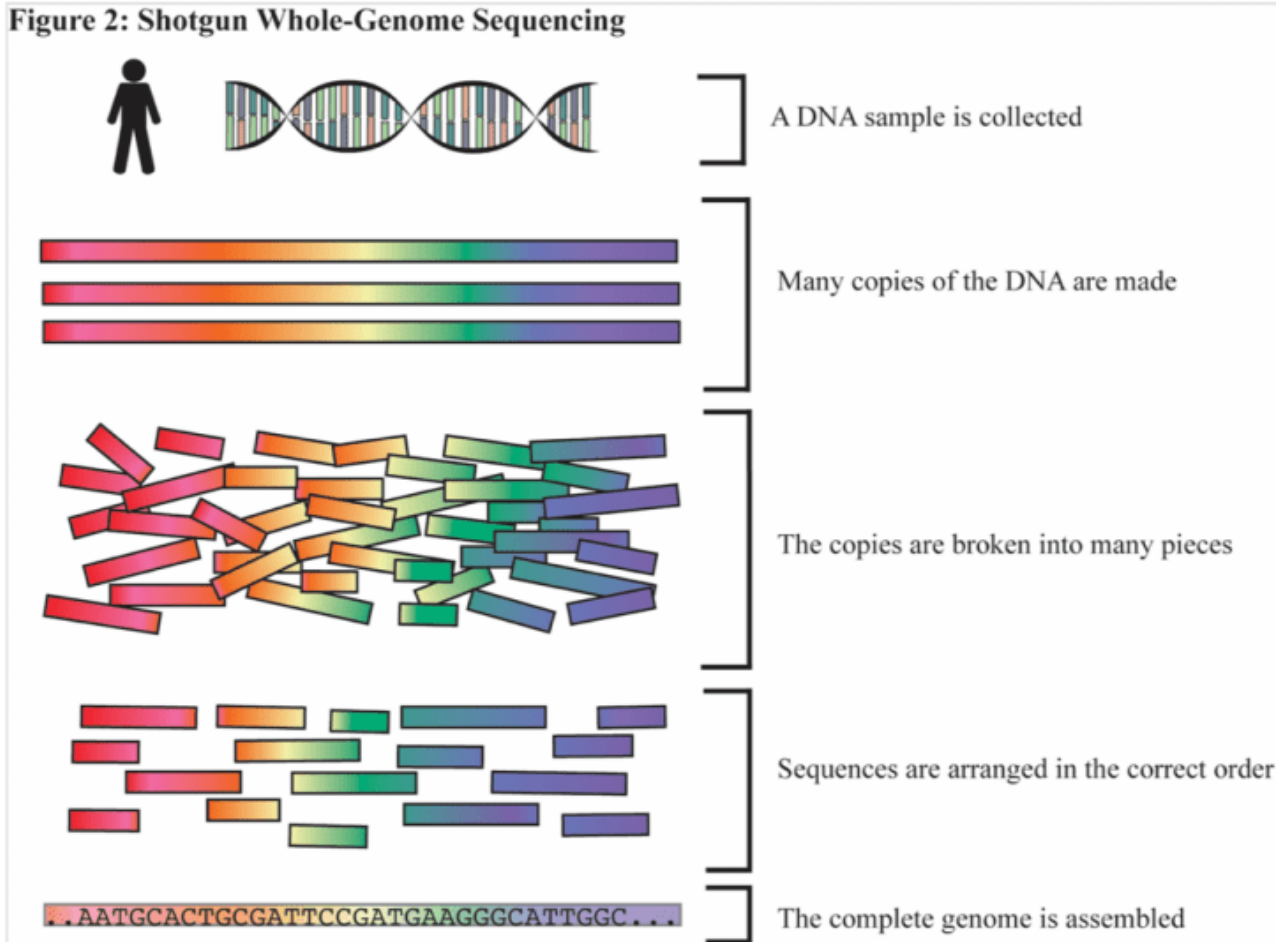
Human genome project

- Started in 1980´s, results published in 2001
- Estimated cost approx. $3 billions and 50 thousand „man-years“
  - Approx. 1/3 of cost for moon landing
- At the beginning under the jurisdiction of Department of Energy of the USA (labs and scientist all over the world were enrolled), later private company started to compete (Celera Genomics)
- Race in the sequencing has begun



Human genome project

- Celera wanted to keep its result private and sell them for profit – in the contrast to the government project
- Results were published at 15.2.2001 in the Nature (gov.) and 16.2.2001 in the Science (Celera project)
- Map of human genome was established, but without variability between individuals



MUNI MED

The HapMap Project

- DNA between each other is different in only about 0.1% of nucleotides - most commonly SNPs, which is known about 10 millions. These SNPs represents about 90% of total genome variability (rest are mutation, deletion and insertion)
- Based on math and statistic approx. 45 unrelated samples should be able to find 99% of all haplotypes with frequency higher than 5%

The HapMap Project

- Started in 2002 – two phases – at first production of „blank map“ and then fill up the blank spaces
- In the first phase was found about 1 mil of SNPs – results in 2005
- Second phase found another 2 mil of SNPs – results in 2007
- **Discovery of approx. 1 mil of LD blocks**
- Scientists from all over the world were enrolled
- Samples from USA, China, Japan, Kenya, UK, Canada

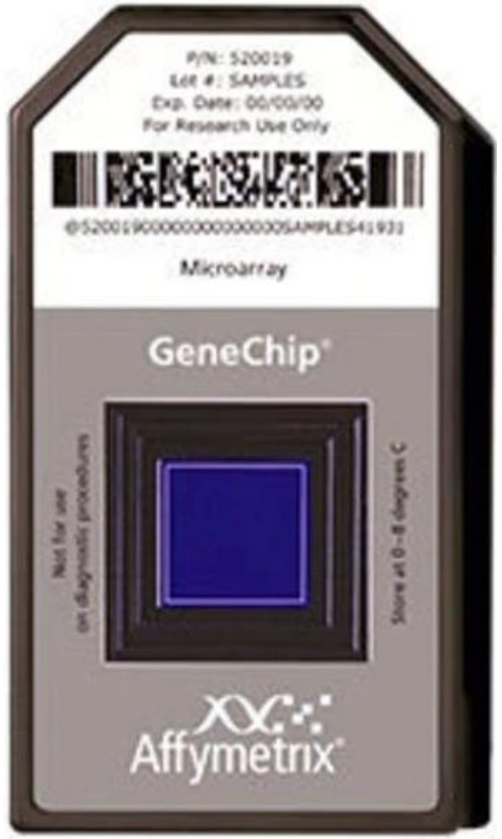
Genome-wide association studies (GWAS)

Combination of epidemiologic studies and new possibilities of genotyping

Ten thousands up to hundred of thousands of SNPs are determined (+ imputation and LD)

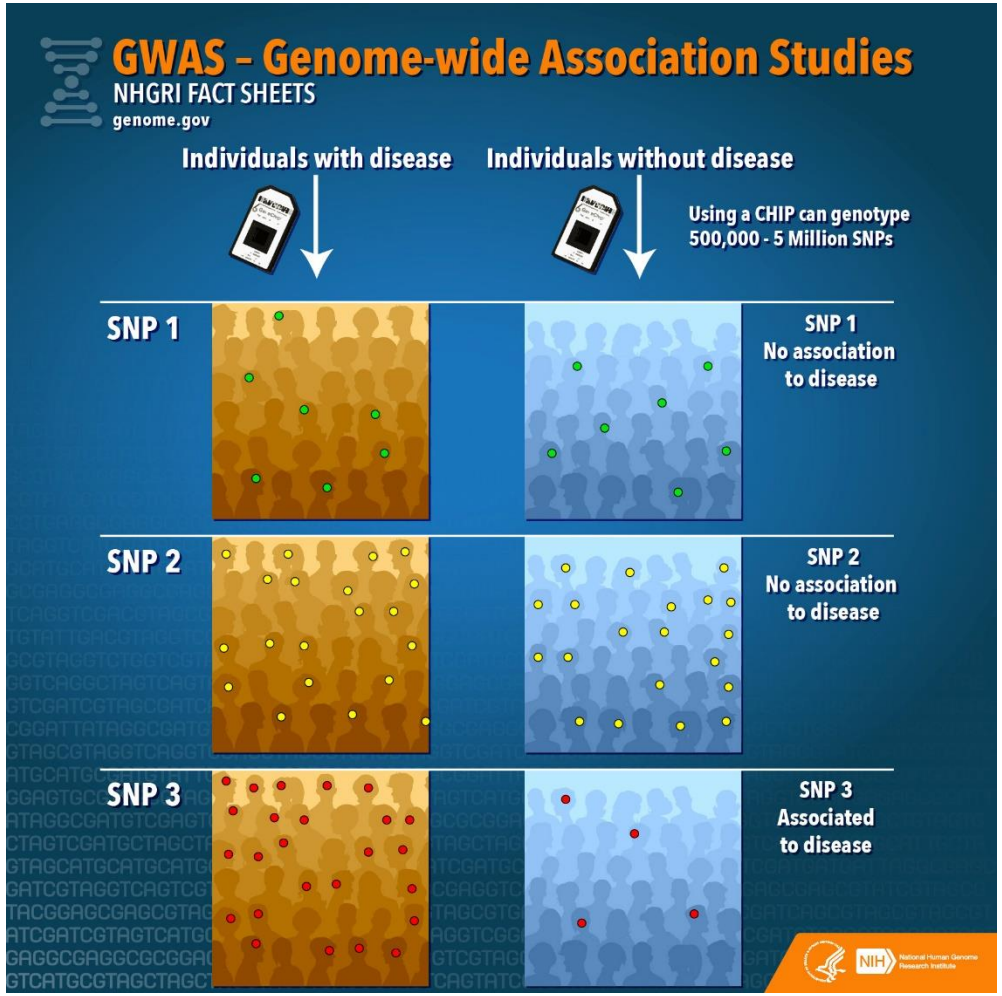
Need for huge set of patients, thousands more likely tens of thousands (control group + group with studied phenotype)

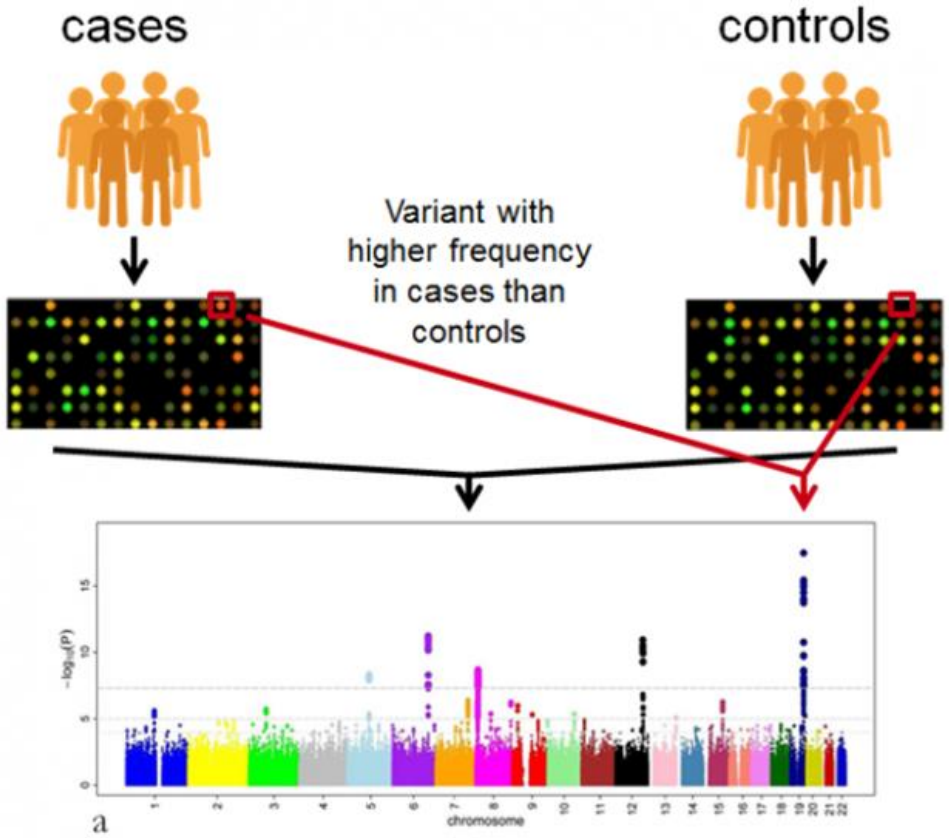
Necessary to proper describe phenotype of both, patients and control group



MUNI MED

GWAS

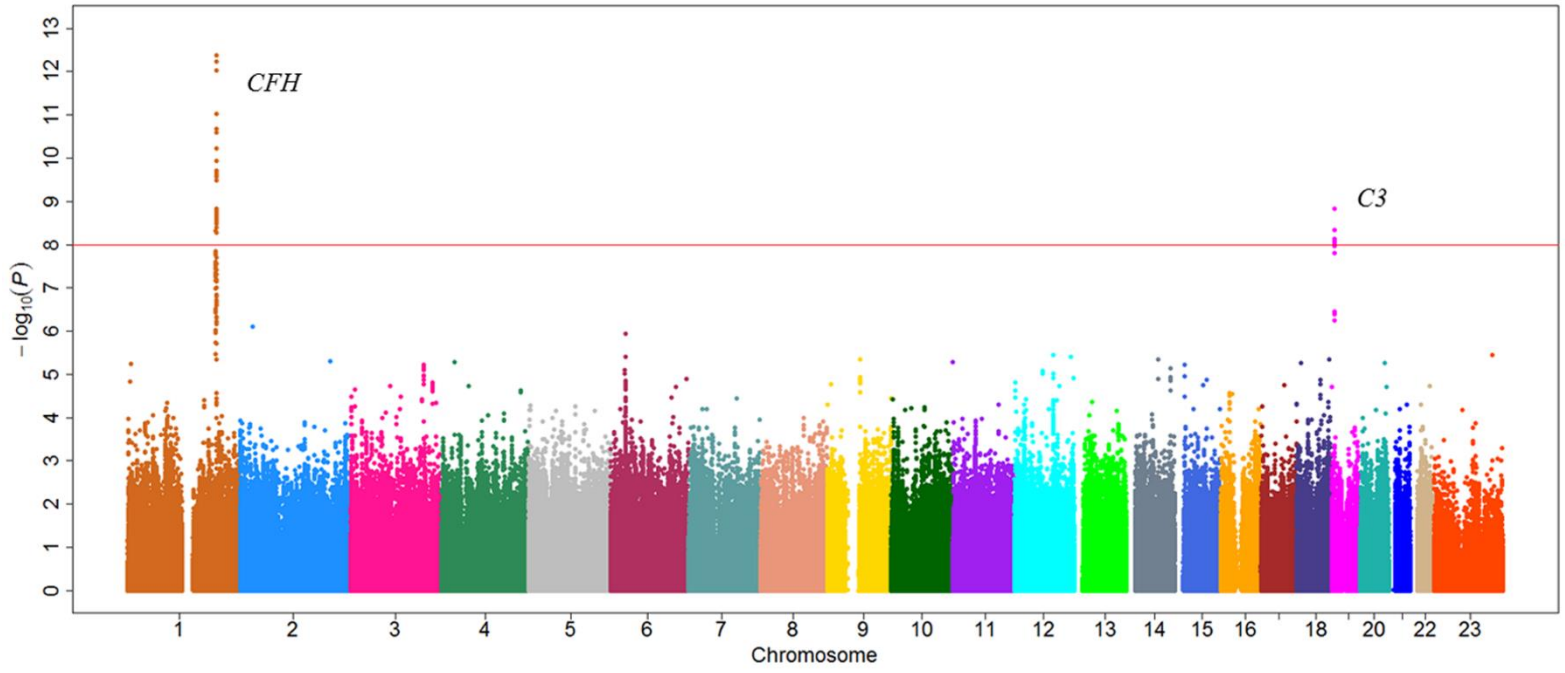




MUNI
MED

GWAS

- Great computing power is necessary for evaluation (approx few GBs for one patient and approx. 15 TB for 10 000 patients)
- As statistically significant is considered $P < 5*10^{-8}$
- P values between $1*10^{-6}$ to $5*10^{-8}$ are further replicated for possible association



MUNI MED

GWAS – pros and cons

+

- Successful method for findings of new variants associated with given phenotype
  - Approx. 40 000 SNPs associated with different traits (cancers, T2DM, anorexia, depression, schizophrenia, BMI, insomnia,…)
- Could lead to discovery of new biological mechanism
  - Study of associated SNPs and their function
- Wide clinical application
  - Identification of risk groups of patients
  - Genetic risk score
- GWAS are able to explain differences between various ethnics in the complex trait
  - E.g. T2DM

-

- Each variant, by itself, have very limited indicative power
- Huge amount of patient is needed
  - Due to high demand for statistical power
- SNPs associated in GWAs represent only portion of inheritability of complex diseases
  - It is estimated that 1/3 to 2/3 of total heritability of complex diseases
- GWAS are able to find only locus associated with trait, not specific SNP
  - Another steps for determination of specific SNP are needed
- Can not find all variants associated with defined trait
  - Hard to find common variant with low effect or very rare variants with big impact

MUNI MED

GWAS – pros and cons

+

- Can find genetic variants with low frequency in population
  - Bigger set of patients, rarer SNPs can be associated
- Data can be used in another use
  - Determination of ancestry, estimate place of birth, forensic analysis, paternity,…
- Data can be loaded and shared to public databases
- Data presented so far represent only tip of the iceberg
  - Bigger set of patients the better information we can get
- Reliable genotyping technology
- Cheap method (price/performance ratio)

-

- Population stratification
  - Differences in allele frequency between patient and controls can be caused by different ancestry rather than association for the gene with specific trait
- Limited clinical predictive ability
  - Rare to predict disease based on specific variant
  - GRS
- Need to know genetic background of investigated population
  - LD can differ between ethnics
  - Could be problem in native Americans, island nation in Pacific, Pygmy
- Does not count with gene-environment interaction
- Big team with various expert is needed for this kind of study

What does the studies say?

- First GWAS studying childhood caries
- 1305 children at age 3-12 years
- Genotyped 580 000 SNPs, with imputation 1,4 M SNPs
- No significant SNPs found

J Dent Res. 2011 Dec;90(12):1457-62. doi: 10.1177/0022034511422910. Epub 2011 Sep 21.

Genome-wide association scan for childhood caries implicates novel genes

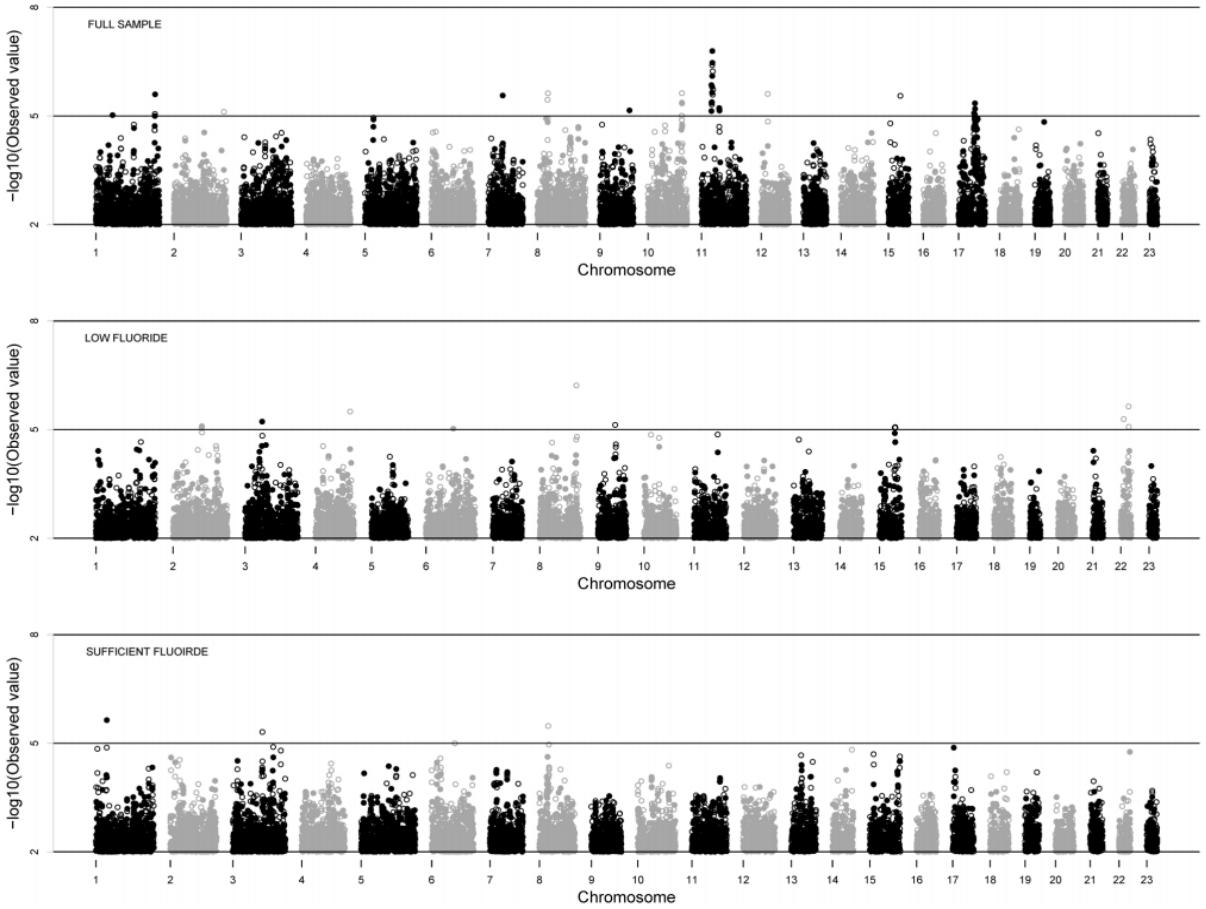
J R Shaffer [1], X Wang, E Feingold, M Lee, F Begum, D E Weeks, K T Cuenco, M M Barmada, S K Wendell, D R Crosslin, C C Laurie, K F Doheny, E W Pugh, Q Zhang, B Feenstra, F Geller, H A Boyd, H Zhang, M Melbye, J C Murray, R J Weyant, R Crout, D W McNeil, S M Levy, R L Slayton, M C Willing, B Broffitt, A R Vieira, M L Marazita

Affiliations + expand

PMID: 21940522 PMCID: PMC3215757 DOI: 10.1177/0022034511422910

MUNI MED

Shaffer et al.



- No significant SNPs were found

MUNI MED

Shaffer et al.

- 920 participants at age18-75 years
- 520 000 SNPs
- Patients were divided into groups based on DMFS (decay-missing-filled surface index)

J Dent Res. 2013 Jan;92(1):38-44. doi: 10.1177/0022034512463579. Epub 2012 Oct 11.

GWAS of dental caries patterns in the permanent dentition

J R Shaffer [1], E Feingold, X Wang, M Lee, K Tcuenco, D E Weeks, R J Weyant, R Crout, D W McNeil, M L Marazita

Affiliations + expand

PMID: 23064961 PMCID: PMC3521449 DOI: 10.1177/0022034512463579

MUNI MED

- Two significant locus were found
  - AJAP1 – involved in development of the tooth together with MMP
  - LYZL2 – lysozyme-like gene, bacteriolytic factor
- Another 31 „suspicious“ loci

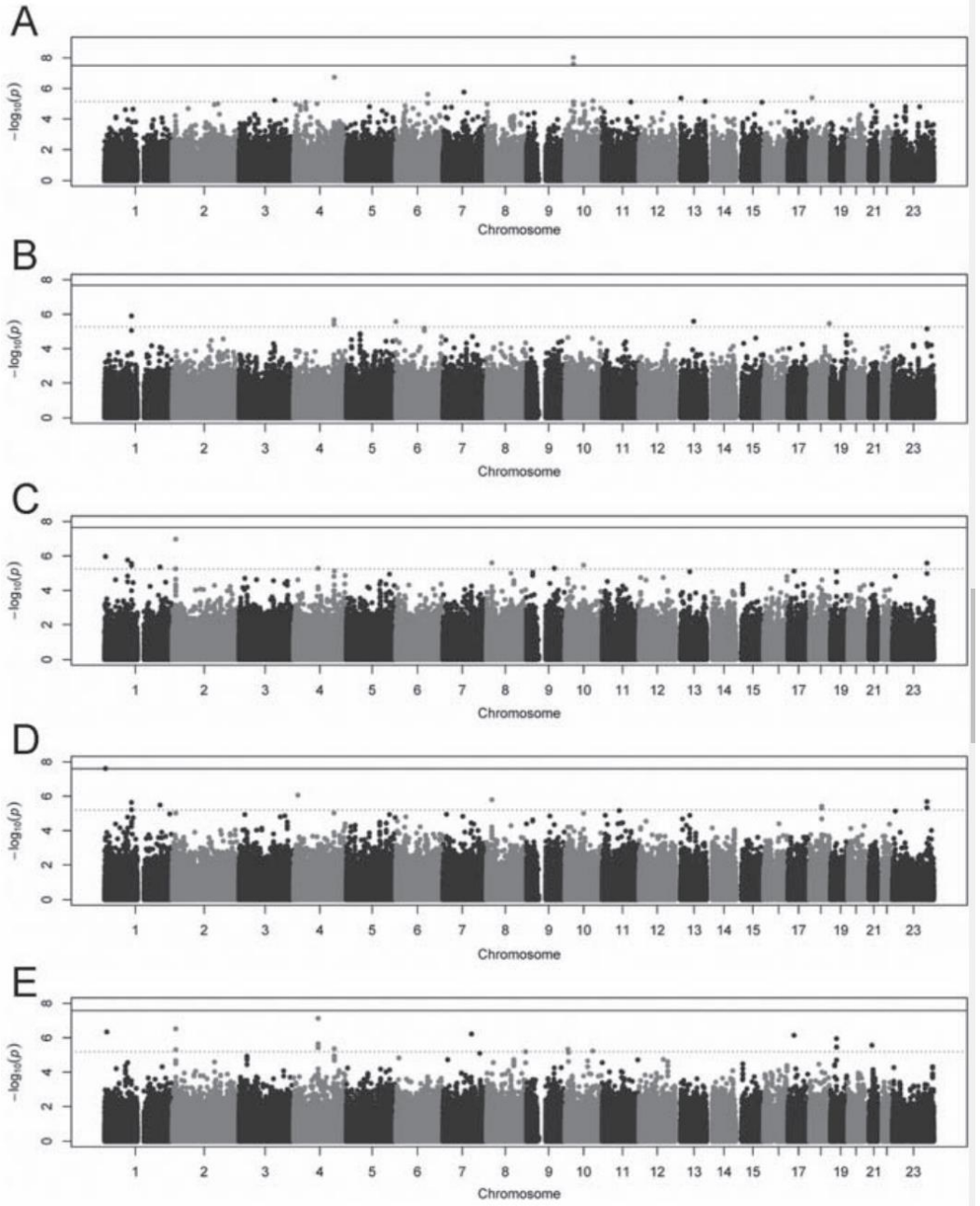


Figure. Manhattan plots showing GWAS results for **(A)** DMFS2, **(B)** DMFS3, **(C)** DMFS5, **(D)** $DMFS5_{max}$, and **(E)** $DMFS5_{mand}$. Solid lines represent thresholds for genome-wide significance (p value < $10^{-7.3}$). Dotted lines represent thresholds for suggestive significance (p value < 10^{-5})

MUNI MED

Zeng et al.

- Two sets of patients – 1006 children at age 3-12 (SM) and 979 children at age 4-14 (PF)
  - DMFS divided into two phenotypes – smooth teeth surface and teeth with fissure
- Genotyped 530 000 SNPs, with imputation 1 200 000 SNPs

› Caries Res. 2014;48(4):330-8. doi: 10.1159/000356299.

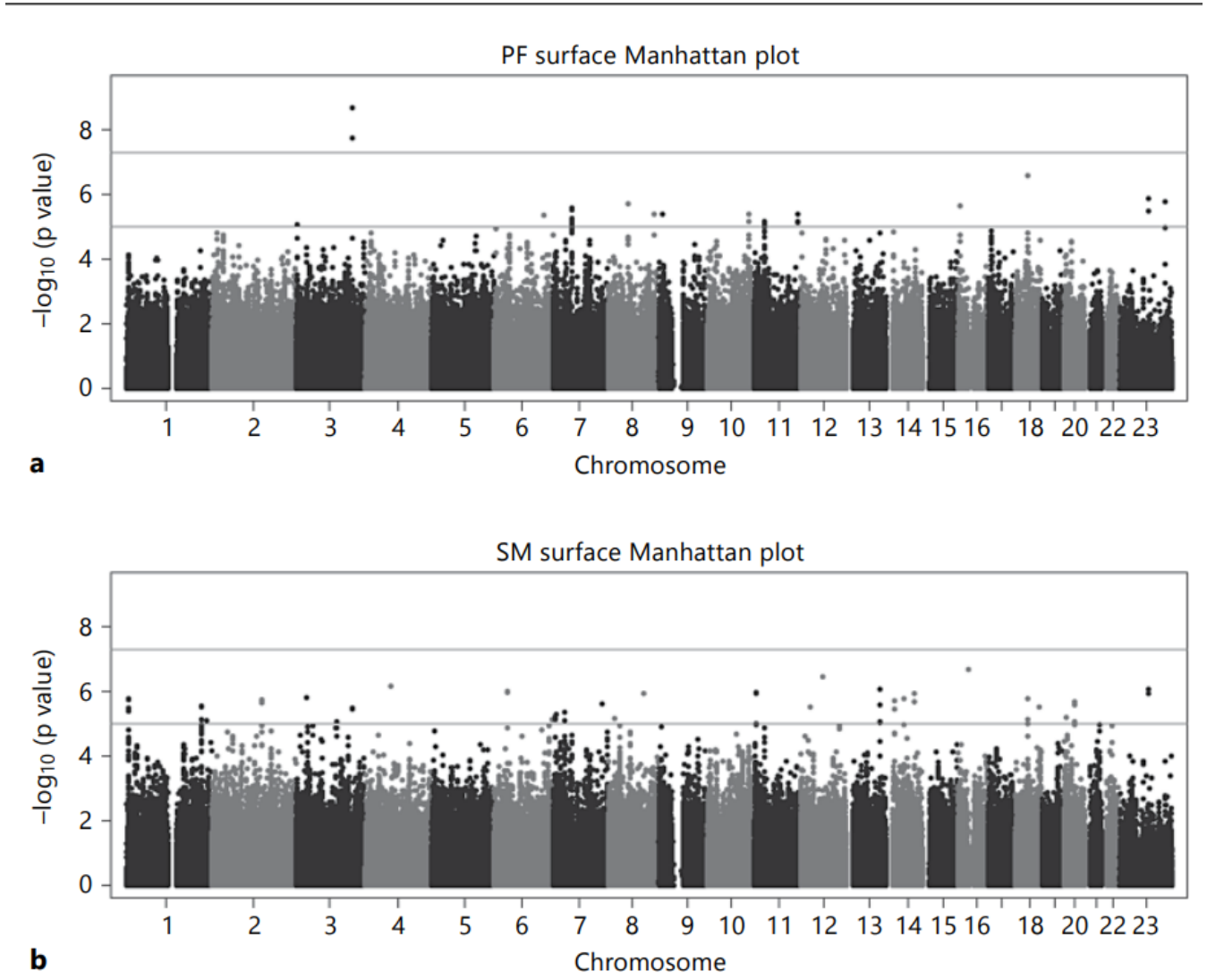
Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries

Z Zeng, E Feingold, X Wang, D E Weeks, M Lee, D T Cuenco, B Broffitt, R J Weyant, R Crout, D W McNeil, S M Levy, M L Marazita, J R Shaffer

PMID: 24556642 PMCID: PMC4043868 DOI: 10.1159/000356299

MUNI MED

- In PF group KPNA4 gene was significantly associated
- No statistically significant association in SM group
- Another 5 suspicious loci



MUNI MED

Shungin et al.

- Two biobanks were used – UKB and GLIDE (Gene-lifestyle interactions in dental endpoints)
  - Over 500 000 patients
- Genotyped approx. 500 000 SNPs + imputation (together 8.9M SNPs)
- 47 new variants were associated with dental caries

Article | Open Access | Published: 24 June 2019

Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data

Dmitry Shungin, Simon Haworth ✉, [...] Ingegerd Johansson

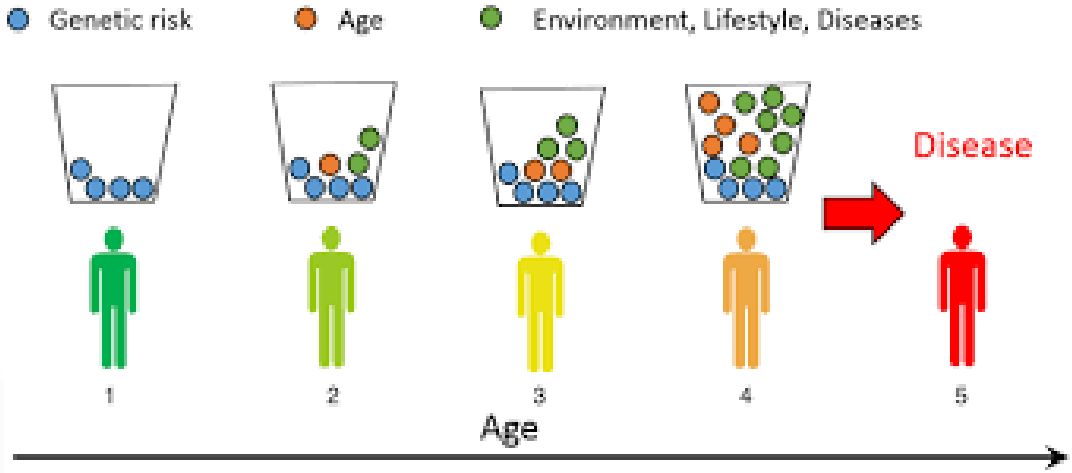
Nature Communications **10**, Article number: 2773 (2019) | Cite this article

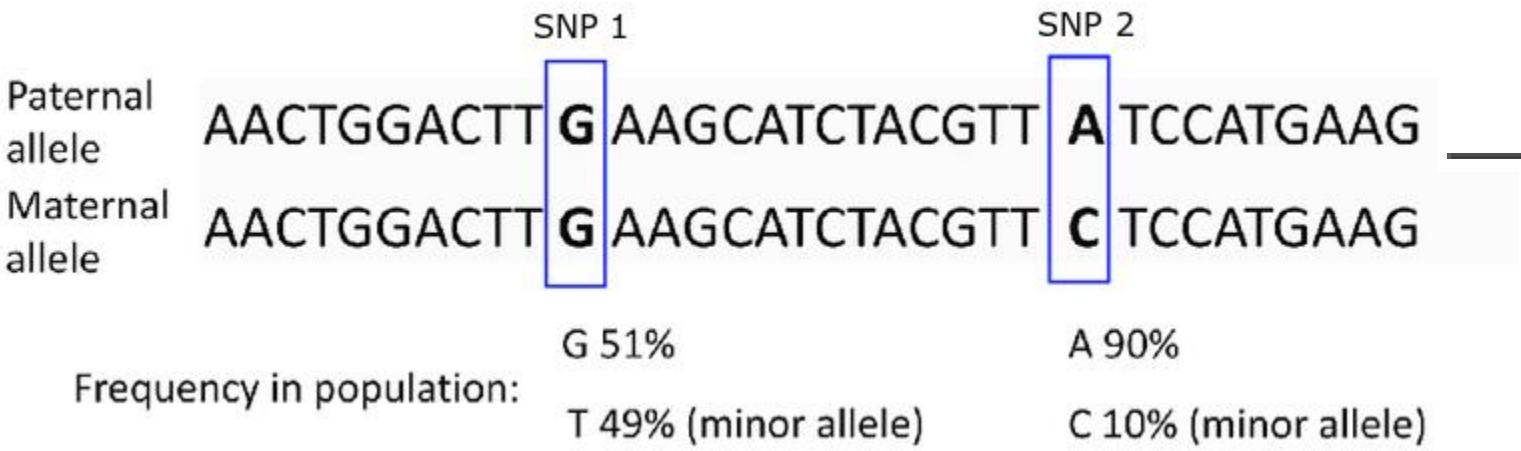
7904 Accesses | **30** Citations | **129** Altmetric | Metrics

MUNI MED

Genetic/polygenic risk score (GRS/PRS)

- Number, determining risk of development of observed phenotype
- Weighted and unweighted

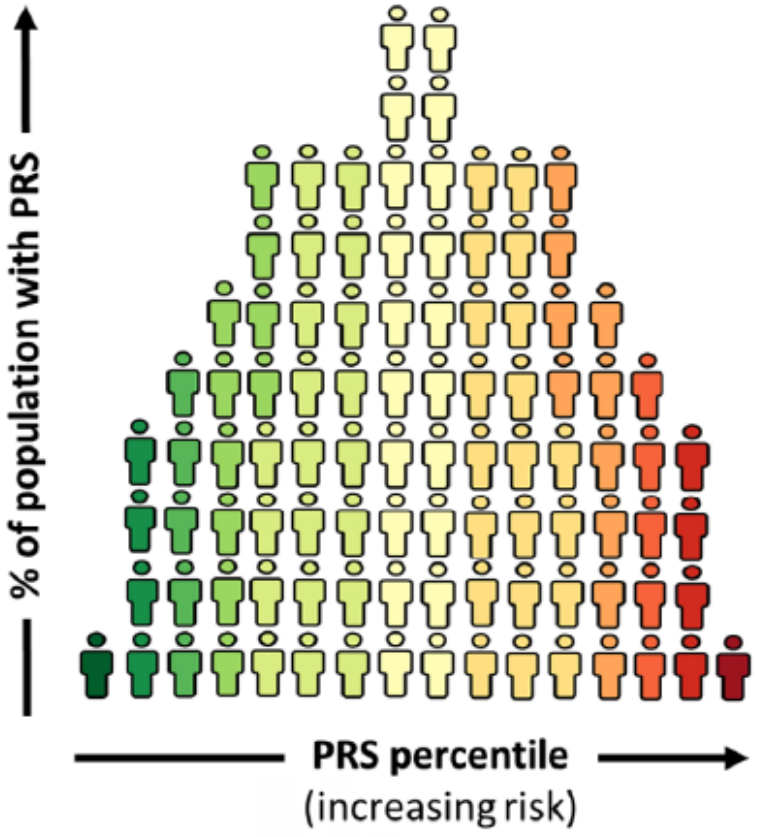




MUNI MED

Genetic/polygenic risk score (GRS/PRS)

PRS percentile	Risk of disease vs. reference group
0-1	Lowest
1-5	
5-10	
10-20	
20-40	
40-60 (reference)	1
60-80	
80-90	
90-95	
95-99	
99-100	Highest



Source: RGA

MUNI MED

Morelli et al.

- 40 most strongly associated SNPs from the GWAS and they constructed unweighted GRS
  - Theoretical values 0-80, mean 37,1 ± 3,9; range of values 24 – 52
  - European-American population

Review > Periodontol 2000. 2020 Feb;82(1):143-156. doi: 10.1111/prd.12320.

Genomics of periodontal disease and tooth morbidity

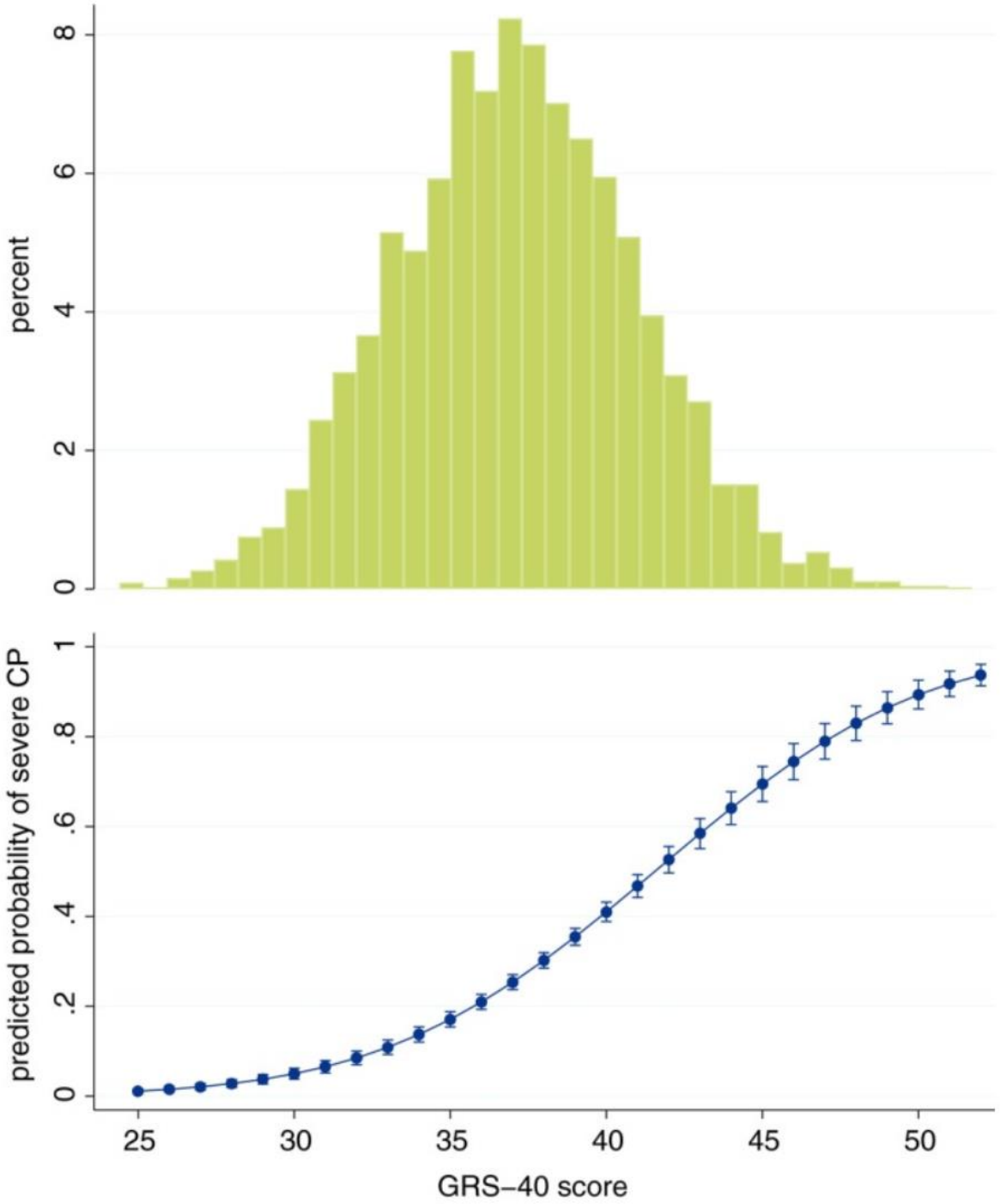
Thiago Morelli [1], Cary S Agler [2], Kimon Divaris [3 4]

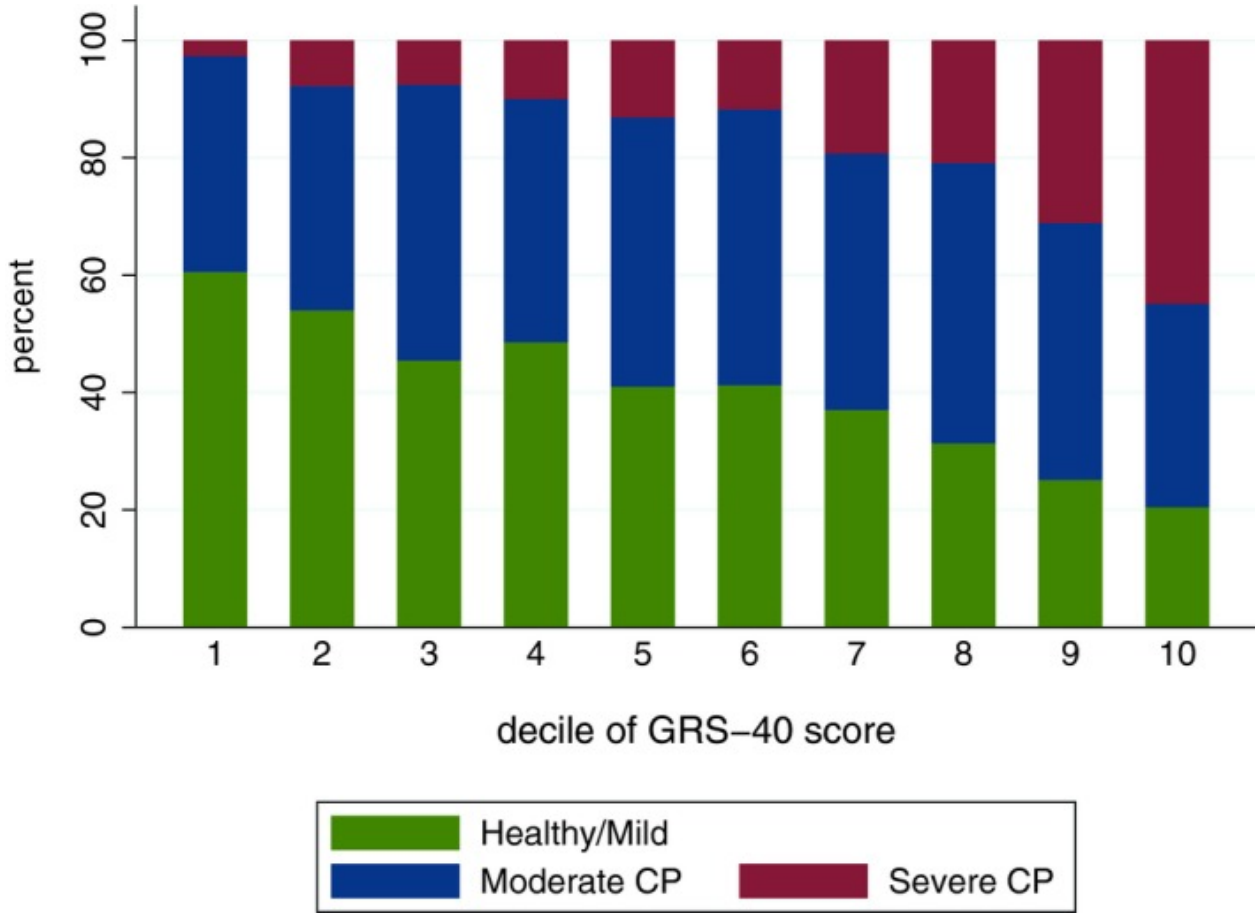
Affiliations + expand

PMID: 31850632 PMCID: PMC6972532 DOI: 10.1111/prd.12320

Free PMC article

MUNI MED


percent
0
2
4
6
8
predicted probability of severe CP
0
.2
.4
.6
.8
1
25
30
35
40
45
50
GRS–40 score


percent
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
10
decile of GRS–40 score
Healthy/Mild
Moderate CP
Severe CP

MUNI MED

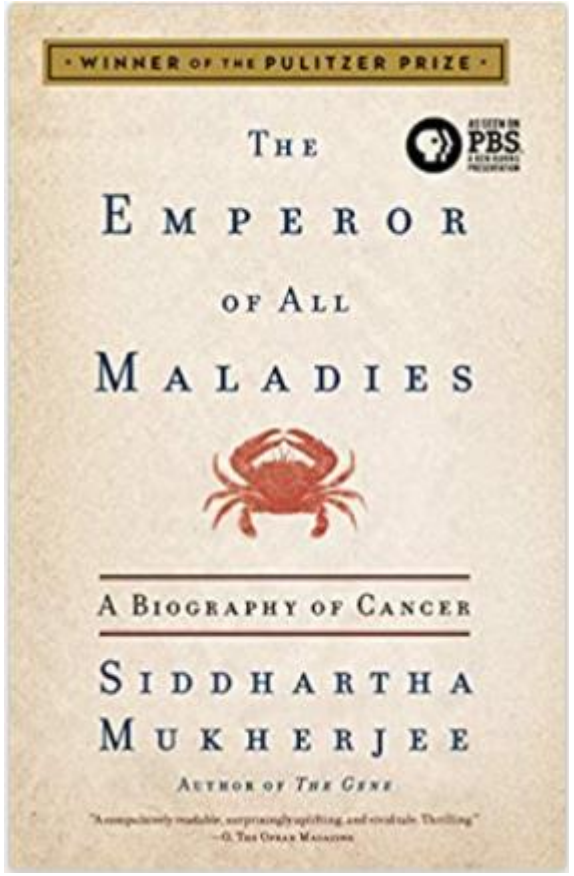
Morelli et al.

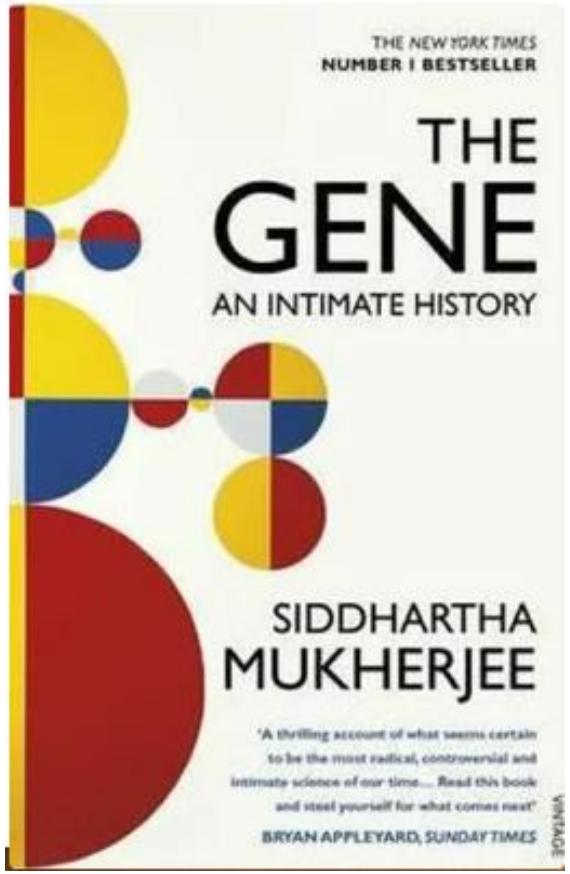
- Authors posted three reasons why these score need further adjustment
  - SNPs used in this study were associated only on one set of patients – does not have to be true for other ethic groups. At first validation ad replication of the results are needed
  - Participants were at middle age and only European-American ancestry
  - Other factors than genetic may play a role on progression of diseases in the oral cavity (habits, socio-economical status, dental care access)
- Tendency to create universal GRS for all people capable of determining individual risk for particular disease. These individuals could be under more frequent screening, they could alter their habits,…

Conclusion

- Era before GWAS
- What are the GWAS – pros and cons
- Summarizing of recent GWAS studies
- Construction of GRS

Recommended literature





Interview with Eric Lander: https://www.ceskatelevize.cz/porady/10441294653-hyde-park-civilizace/220411058090919/

MUNI MED