MUNI
MED

Novinky v analýze genomu – GWAS, funkční genomika

Osnova

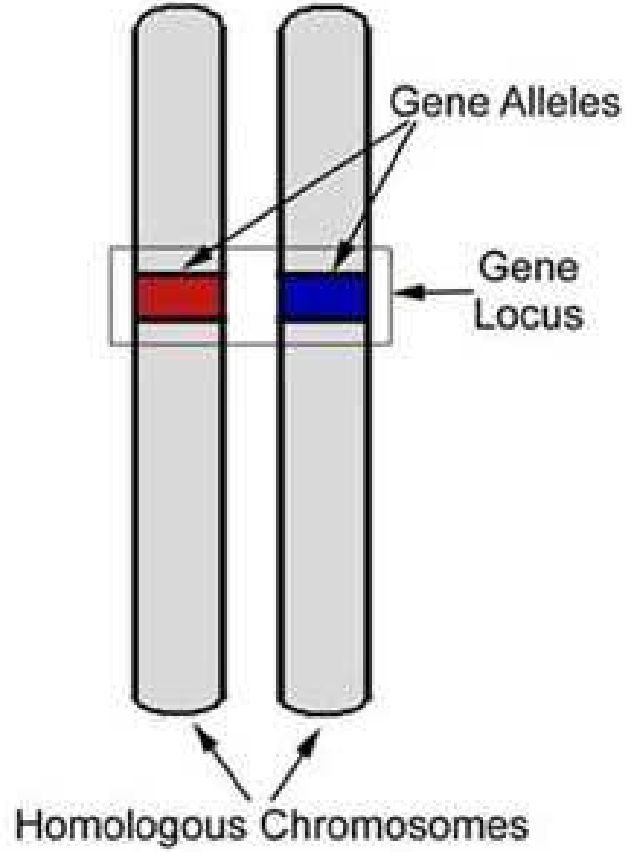
- Vymezení pojmů
- Projekt lidského genomu
- HapMap
- GWAS
- GWAS a nemoci dutiny ústní
- GRS

Vymezení pojmů

- Alela
- Lokus
- Single nucleotide polymorphism (SNP)
- Haplotyp
- Linkage disequilibrium (LD)
- Imputace
- Genome wide association studies (GWAS)
- Genetic risk score (GRS)

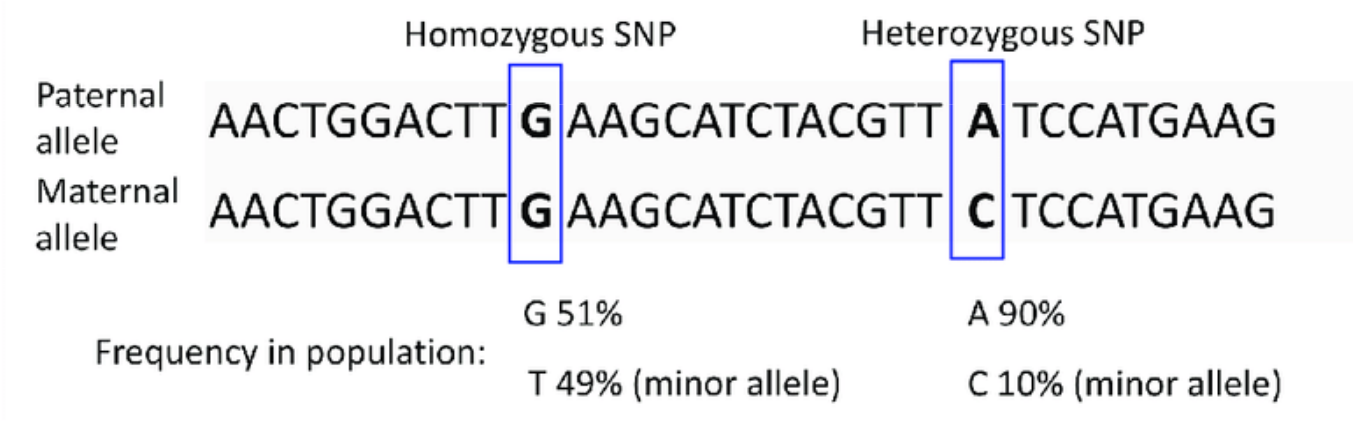
Alela a lokus

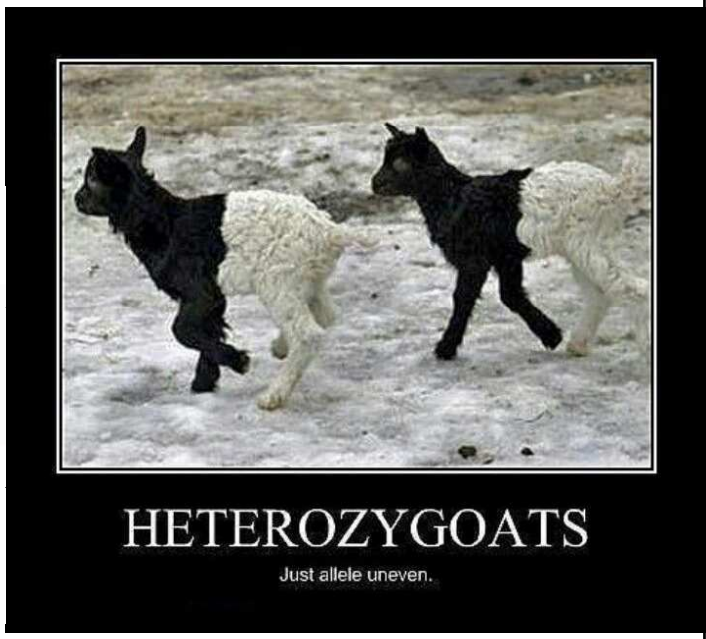
- Alela je konkrétní varianta genu
- Lokus je konkrétní pozice na chromozomu



MUNI MED

Single nucleotide polymorphism (SNP)



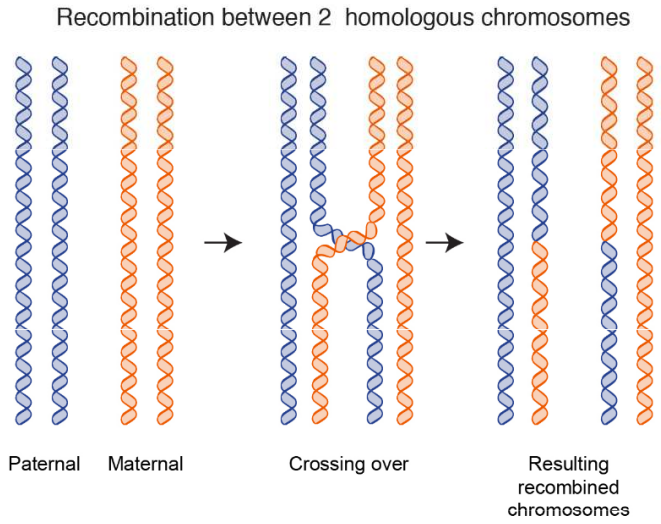


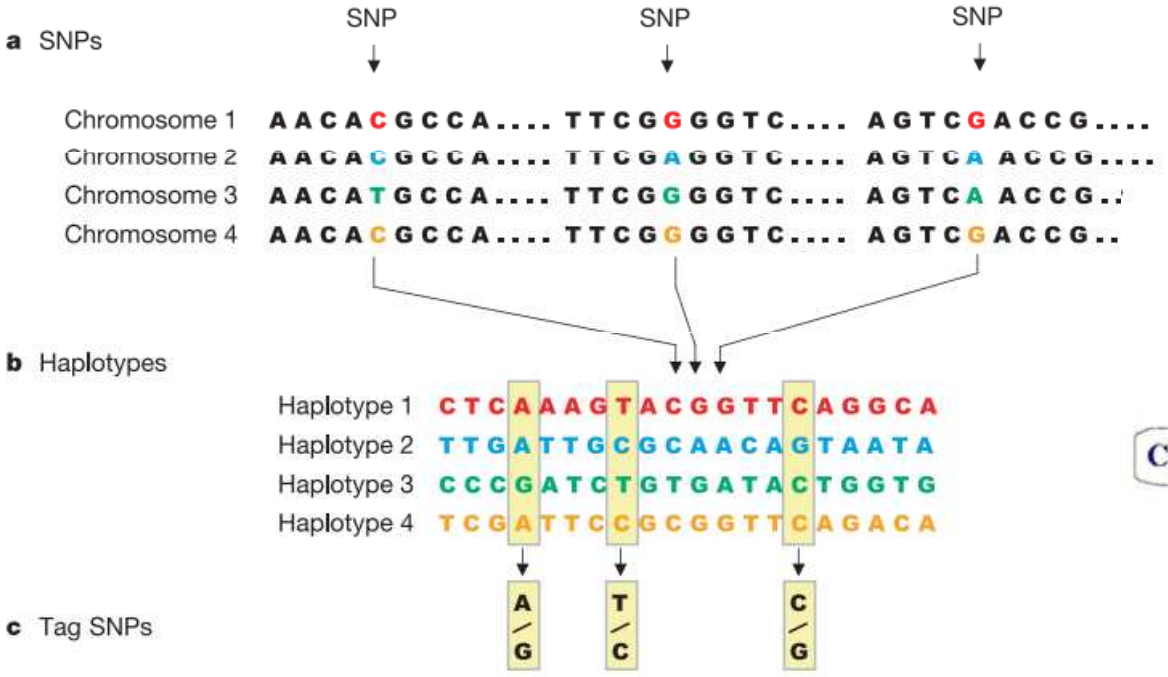
Jedná se o jednobodovou záměnu nukleotidu s frekvencí v populaci větší než 1%. Tato záměna může a nemusí mít vliv na funkci genu/proteinu.

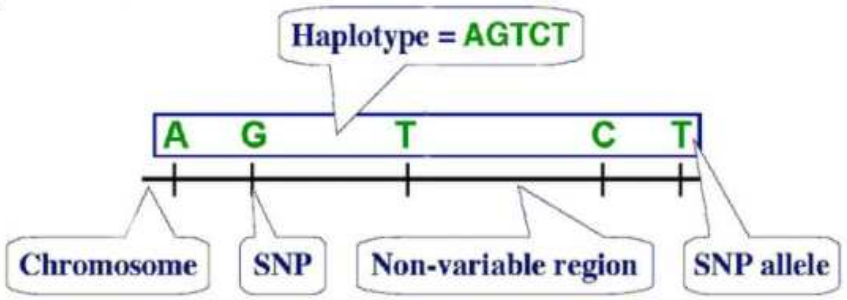
MUNI MED

Haplotyp

– Jedná se o kombinaci alel na různých místech DNA (jednom chromozomu nebo jeho oblasti), které se dědí společně



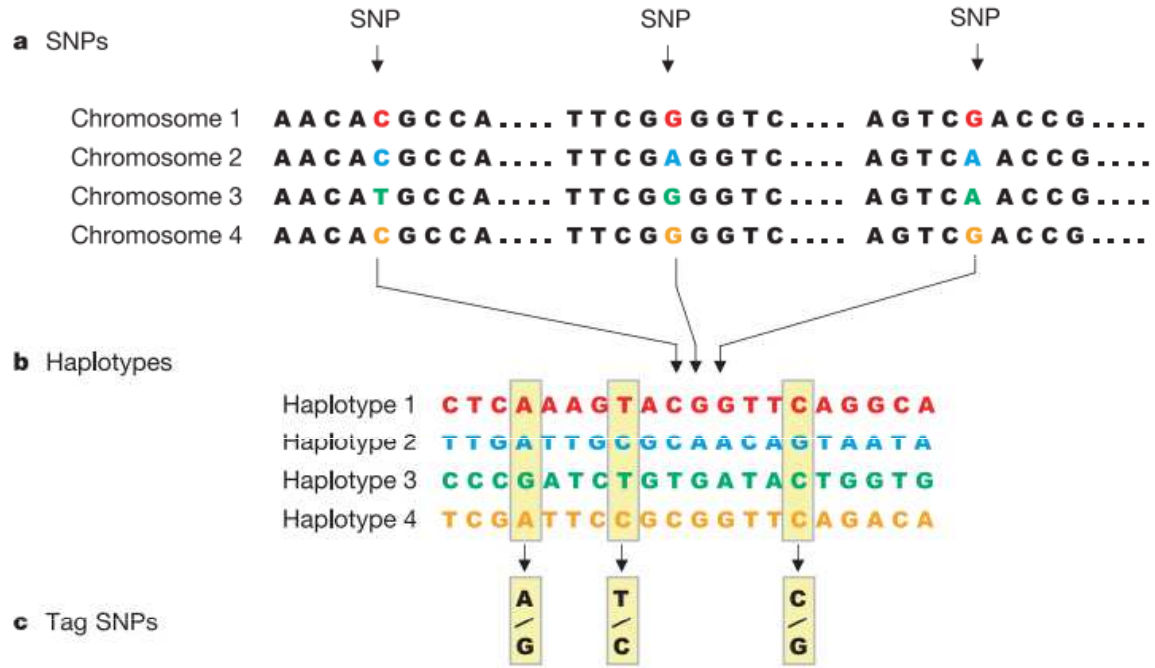




MUNI MED

Linkage disequilibrium (LD) a imputace

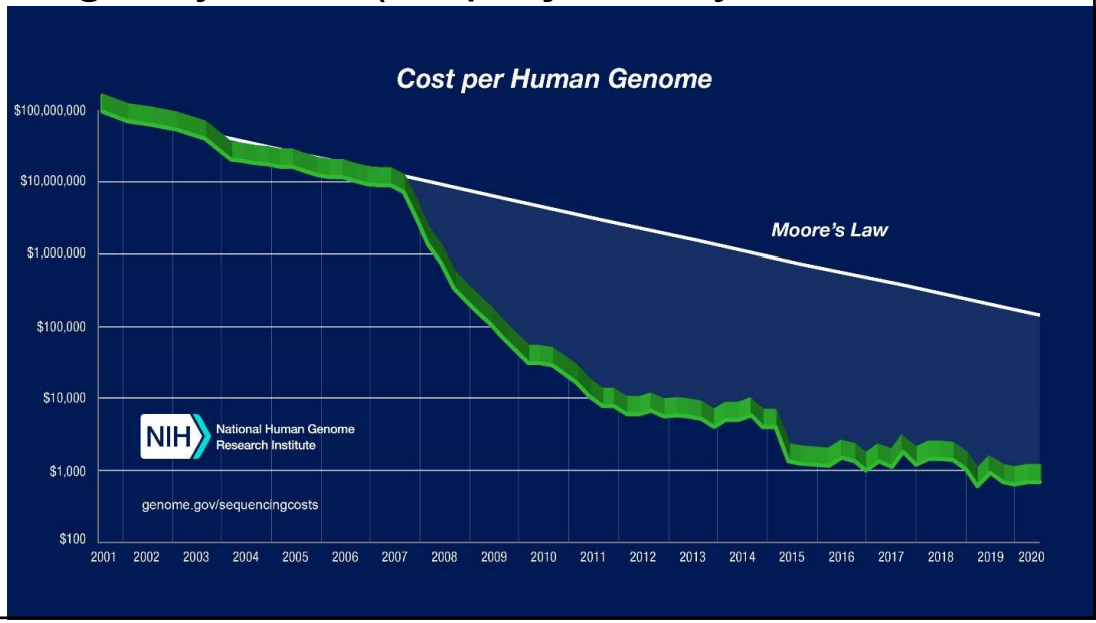
– Česky - vazbová nerovnováha



MUNI MED

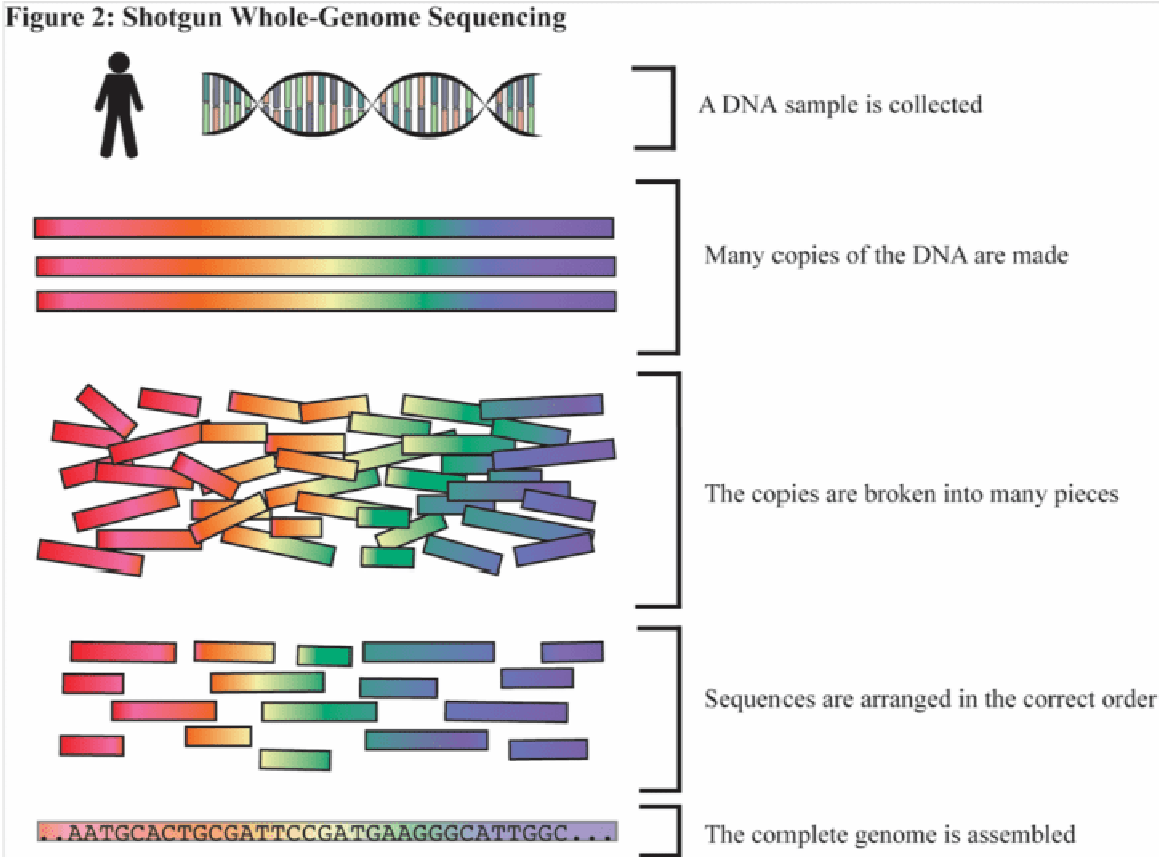
Projekt lidského genomu

- Projekt zahájen v 80. letech, publikování výsledků 2001
- Odhadované náklady cca $3 mld a 50 tisíc „člověkolet“
  - Cca 1/3 úsilí a nákladů pro přistání na měsíci
- Nejprve pod patronátem ministerstva energetiky USA (do projektu byli zapojené laboratoře a vědci z celého světa), později vznikla konkurence v podobě soukromé společnosti (Celera Genomics).
- Probíhal tak závod v sekvenování



Projekt lidského genomu

- Celera si chtěla nechat výsledky pro sebe a zároveň je prodávat „předplatitelům“ – velká rozepře s vládním projektem.
- Výsledky nakonec publikovány 15.2.2001 v Nature (vládní projekt) a 16.2.2001 v časopise Science (Celera)
- Podařilo se získat jakousi mapu lidského genomu, ovšem bez variability mezi jednotlivci



MUNI MED

Projekt HapMap

- DNA mezi jednotlivci se liší jen asi v 0,1% nukleotidů – nejčastěji SNP, kterých je známo cca 10 miliónů. To představuje 90% veškeré variability v genomu (zbytek jsou mutace, delece, inzerce)
- Cca 45 nepříbuzných vzorků by mělo být schopno najít 99% všech haplotypů s frekvencí větší než 5%

Projekt HapMap

- Zahájen 2002 – dvě fáze – nejprve tvorba jakési hrubé mapy a poté zaplnění prázdných míst
- V první fázi cca 1 milión SNP – výsledky 2005
- Druhá fáze cca další 2 milióny SNP - výsledky 2007
- **Objev asi 1 miliónu LD bloků**
- Účast vědců z celého světa
- Vzorky z USA, Číny, Japonska, Keni, UK, Kanady

Genome-wide association studies (GWAS)

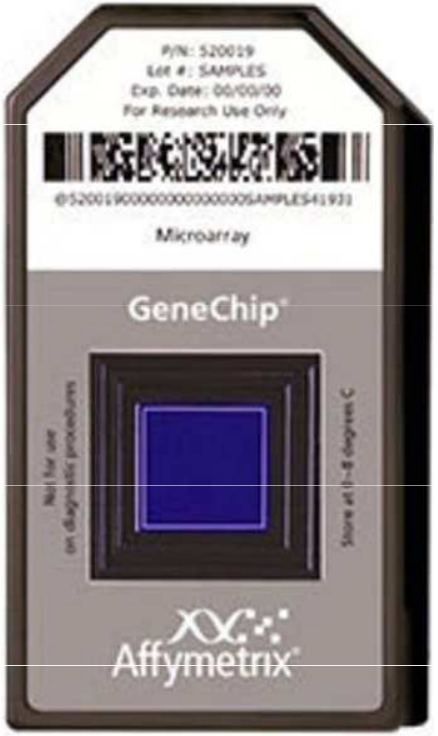
Celogenomové asociační studie

Kombinace epidemiologické studie a nových možností genotypizace

Stanovují se desetitisíce až stovky tisíc SNP (imputace a LD)

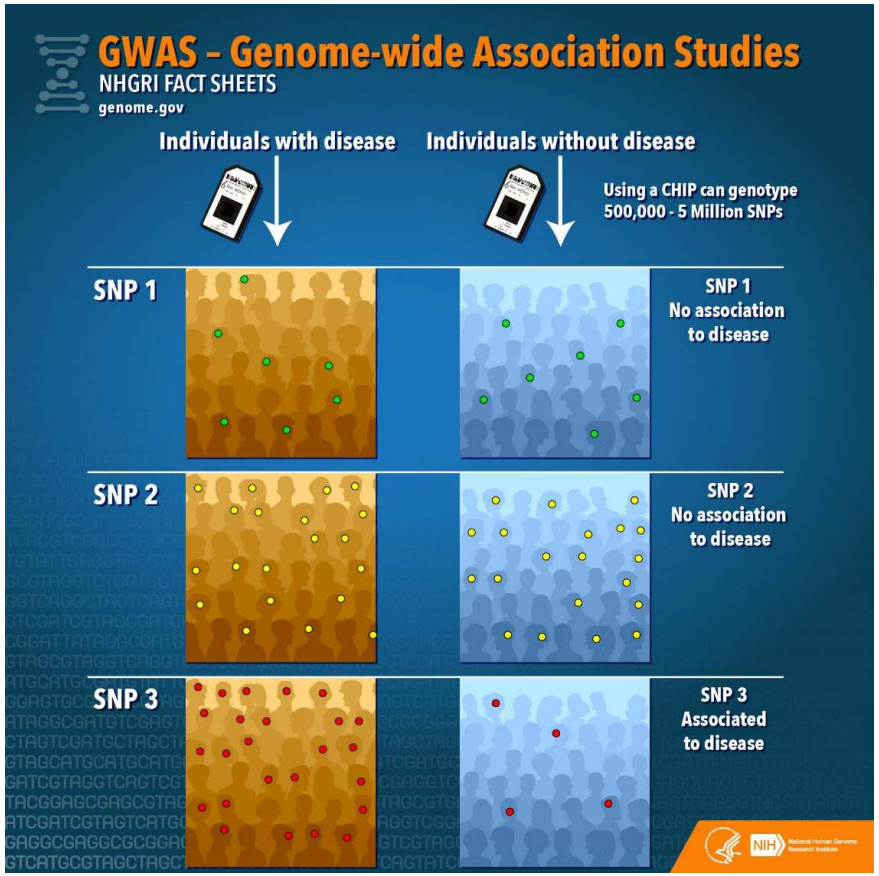
Potřeba velikého souboru pacientů, tisíce, spíše desetitisíce pacientů (kontrolní skupina a pacienty s daným fenotypem)

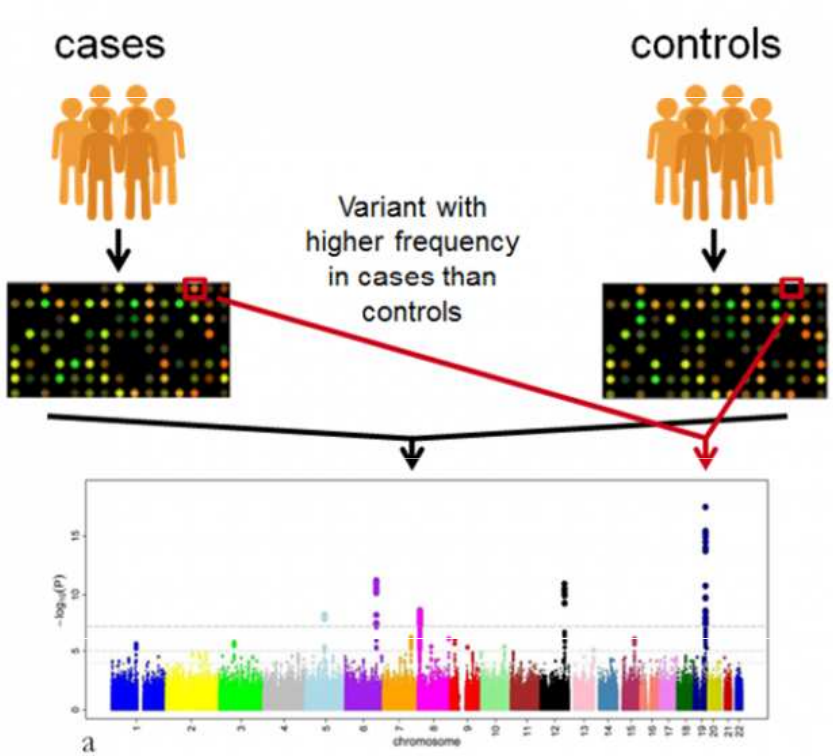
Potřeba mít dobře popsaný fenotyp pacientů i kontrol



MUNI MED

GWAS

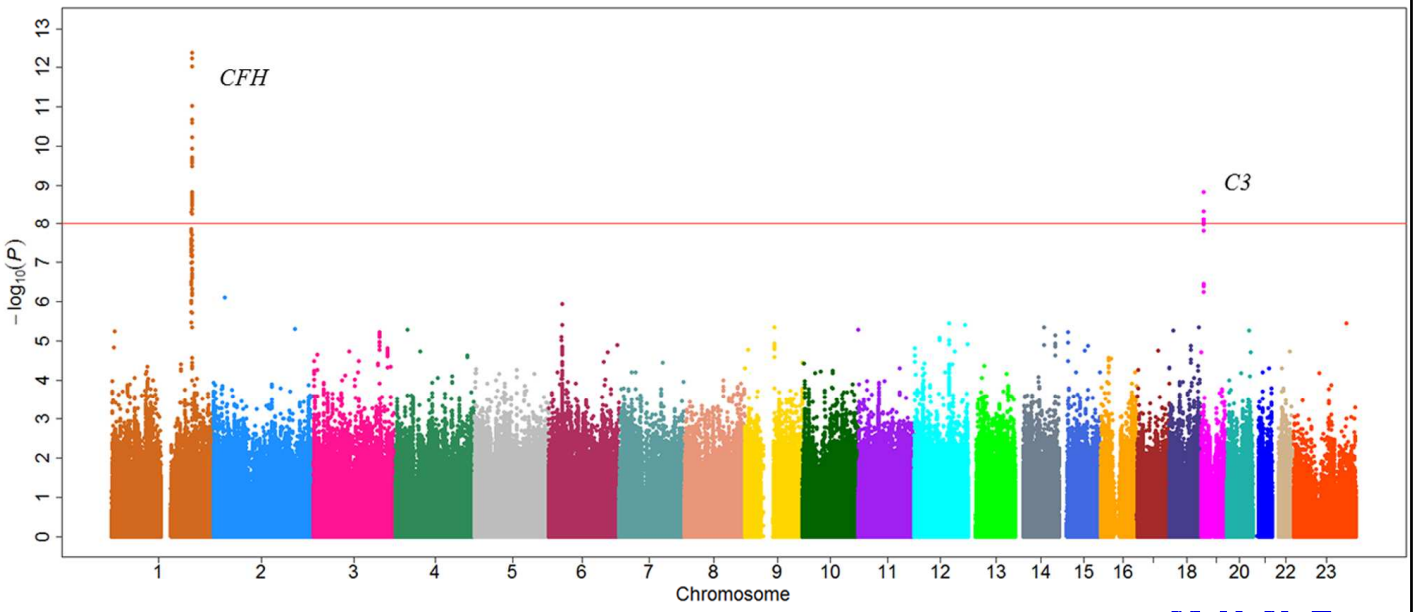




MUNI MED

GWAS

- Pro hodnocení je potřeba velká výpočetní síla a skladovací kapacita (několik set GB na jednoho pacienta, cca 15 TB na 10 000 pacientů)
- Jako statisticky významné se považuje $P < 5*10^{-8}$
- P hodnoty pro SNP mezi $1*10^{-6}$ a $5*10^{-8}$ se replikují pro možnou asociaci



MUNI MED

GWAS – výhody a nevýhody

+

- Velmi úspěšná metoda pro nalezení nových variant asociovaných s daným znakem
  - Cca 40 000 SNP asociovaných s různými rysy (rakoviny, T2DM, anorexie, deprese, schizofrenie, BMI, nespavost,…)
- Mohou vést k objevu nových biologických mechanismů
  - Rsp. studium nalezených SNP a jejich funkcí
- Široké klinické aplikace
  - Identifikace rizikových skupin/pacientů
  - Genetické rizikové skóre
- GWAS mohou poskytnou vysvětlení pro rozdílnost mezi různými etniky u komplexních znaků
  - Např. T2DM

-

- Každá varianta má sama o sobě malou vypovídající hodnotu
- Je třeba velikého souboru pacientů
  - Kvůli vysokému nároku na statistický rozdíl
- Varianty asociované v GWAS představují jen zlomek celkové „dědičnosti“ komplexních chorob
  - Odhady, že SNP odhalí cca 1/3 – 2/3 celkové dědičnosti kompl. chorob
- GWAS označují pouze lokus asociovaný se znakem
  - Pro identifikaci kauzálních variant je třeba provést další kroky
- Nemohou nalézt všechny varianty asociované s daným znakem
  - Problém s nalezením běžných variant s malým účinkem nebo velmi vzácných variant s velkým účinkem

GWAS – výhody a nevýhody

+

- Jsou schopny nalézt varianty s nízkou frekvencí výskytu
  - Čím větší soubor, tím vzácnější varianty lze nalézt a asociovat
- Data se dají využít i jinak než identifikaci genů
  - Určení předků, odhadnutí místa narození, forenzní analýza, určení otcovství,…
- Data se snadno sdílí přes veřejné databáze
- Data doposud prezentovaná představují jen špičku ledovce
  - Čím více dat o pacientech a větší soubory, tím přesnější informace budeme schopni zjistit
- Jedná se o spolehlivou genotypizační technologii
- Relativně levná metoda (poměr cena/výkon)

-

- Population stratification
  - Rozdíl ve frekvenci jednotlivých alel mezi pacienty a kontrolami může být způsoben rozdílnými předky spíše než asociací genu s chorobou
- Omezená klinicky prediktivní schopnost
  - Málo kdy lze díky konkrétní variantě spolehlivě předpovídat nemoc
  - Viz GRS
- Potřeba znát genetický základ zkoumané populace
  - LD se může lišit mezi jednotlivými etniky – potřebuji znát genom dané populace
  - Problém např. u Indiánů, ostrovních národů v Pacifiku, Pygmejů
- Neberou v potaz interakci gen-prostředí
- Potřeba pracovat s velkým týmem, s různou specializací

Co říkají konkrétní studie?

- První GWAS zkoumající kazivost zubů u dětí
- 1305 dětí ve věku 3-12 let
- Genotypizováno 580 000 SNPs, pomocí imputace dohromady 1,4 M SNPs
- Žádné signifikantní SNP nenalezeny

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

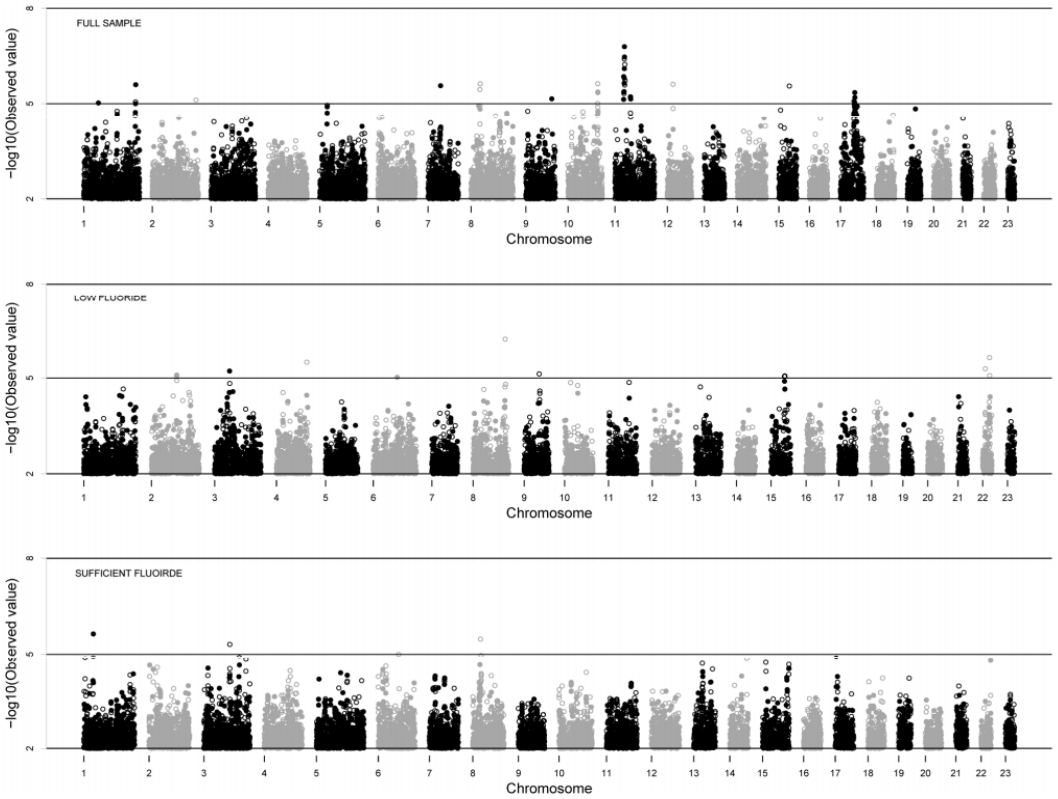
J R Shaffer [1], X Wang, E Feingold, M Lee, F Begum, D E Weeks, K T Cuenco, M M Barmada, S K Wendell, D R Crosslin, C C Laurie, K F Doheny, E W Pugh, Q Zhang, B Feenstra, F Geller, H A Boyd, H Zhang, M Melbye, J C Murray, R J Weyant, R Crout, D W McNeil, S M Levy, R L Slayton, M C Willing, B Broffitt,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

Affiliations + expand

PMID: 21940522 PMCID: PMC3215757 DOI: 10.1177/0022034511422910

MUNI MED

Shaffer et al.



– Žádné signifikantní SNP nenalezeny

MUNI MED

Shaffer et al.

- 920 participantů ve věku 18-75 let
- 520 000 SNPs
- Pacienti rozděleni podle 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

- Celkem 2 signifikantní lokusy
  - AJAP1 – podílí se na vývoji zubů společně s MMP
  - LYZL2 – lysozyme-like gene, bakteriolytický faktor
- Dalších 31 „podezřelých“ lokusů

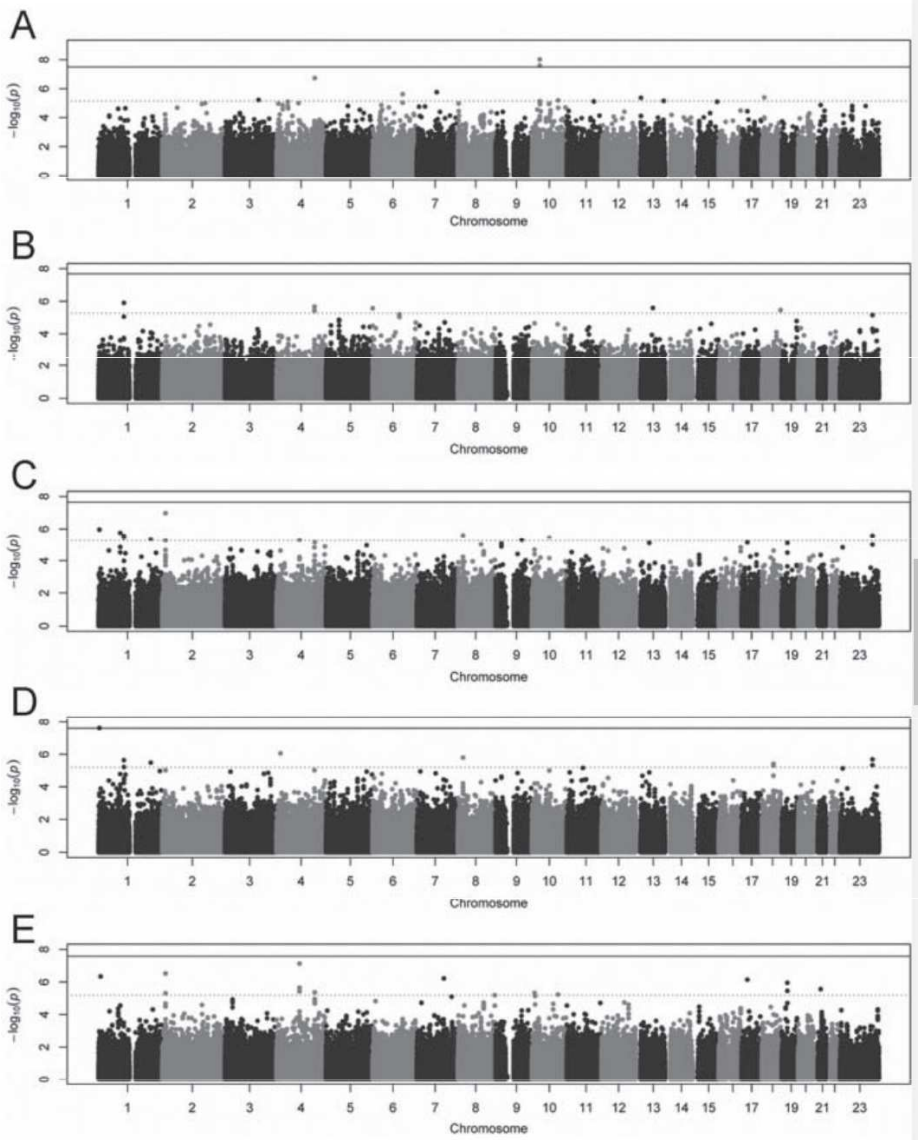


Figure. Manhattan plots showing GWAS results for **(A)** DMFS2, **(B)** DMFS3, **(C)** DMFS5, **(D** $DMFS5_{max}$, and **(E)** $DMFS5_{mand}$. Solid lines represent thresholds for genome-wide significance (p value < $10^{-7.3}$). Dotted lines represent thresholds for suggestive significance (p value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

MUNI MED

Zeng et al.

- Dva soubory 1006 dětí ve věku 3-12 (SM) a 979 dětí ve věku 4-14 (PF)
  - Autoři rozdělili DMFS na dva fenotypy – s hladkým povrchem zubů a zuby s fisuramy
- Genotypizováno 530 000 SNPs, s imputací 1 200 000 SNPs

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

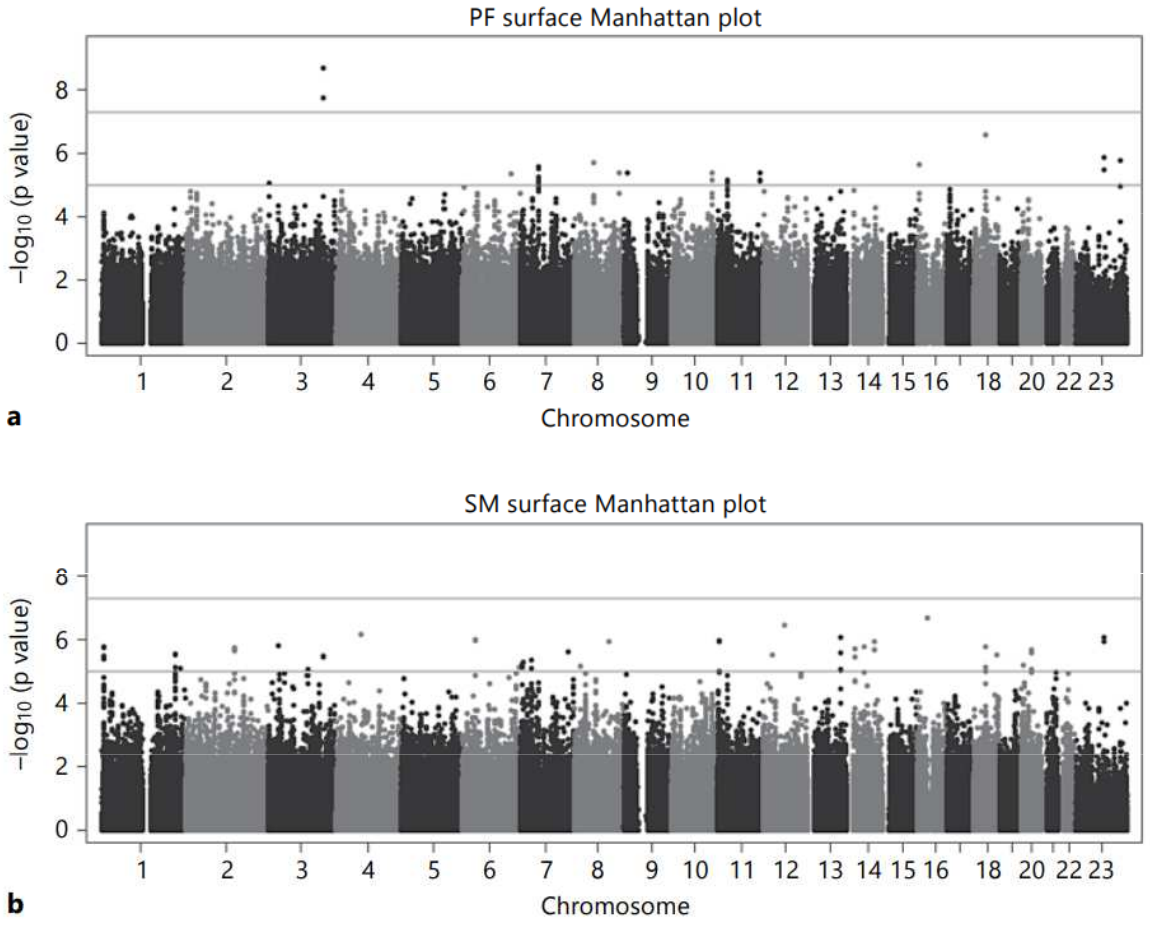
Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries

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

- U PF skupiny byl asociován gen KPNA4
- U SM skupiny nebyla nalezena žádná asociace
- Dalších 5 podezřelých lokusů



MUNI MED

Shungin et al.

- Využití dvou biobank – UKB a GLIDE (Gene-lifestyle interactions in dental endpoints)
  - Dohromady přes 500 000 pacientů
- Genotypyzováno cca 500 000 SNP + imputace (dohromady cca 8,9M SNPs)
- Celkem asociováno 47 nových variant

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

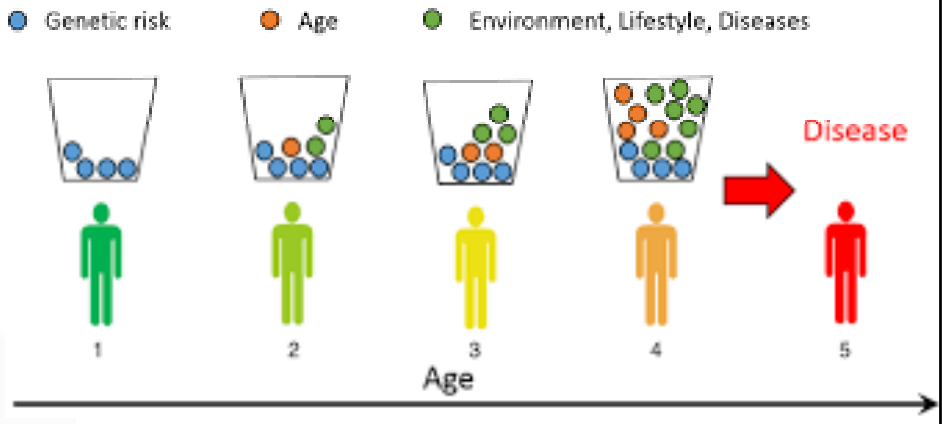
Nature Communications **10**, Article number: 2773 (2019) | Cite this article

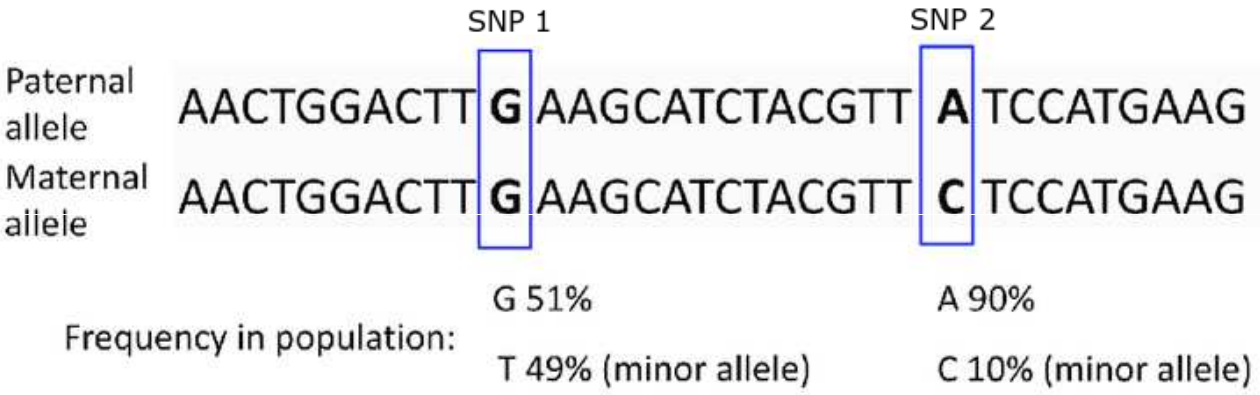
7904 Accesses | **30** Citations | **129** Altmetric | Metrics

MUNI MED

Genetické/polygenní rizikové skóre (GRS/PRS)

- Číslo, které vyjadřuje riziko sledovaného fenotypu
- Vážené a nevážené GRS



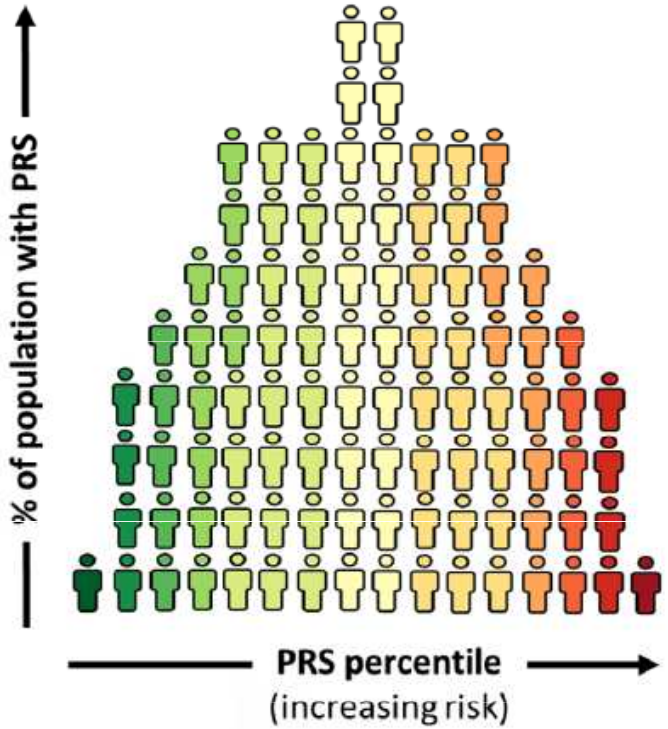


MUNI MED

Genetické/polygenní rizikové skóre (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.

- Vzali 40 nejsilněji asociovaných SNP z GWASu a sestavili nevážené GRS
  - Teoretické hodnoty 0-80, průměr byl 37,1 ± 3,9; rozmezí 24 – 52
  - Evropsko-americká populace

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

Genomics of periodontal disease and tooth morbidity

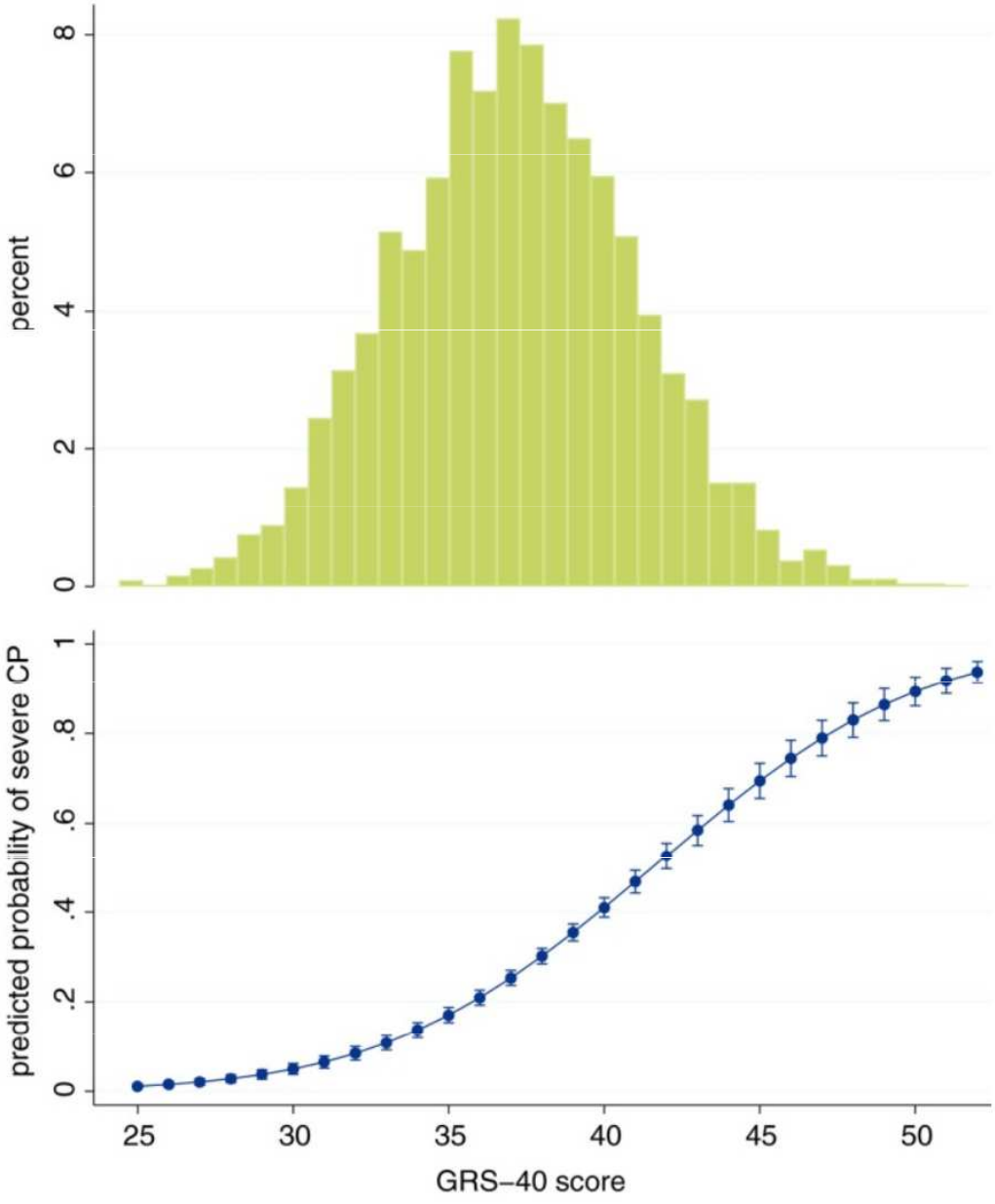
Thiago Morelli [1], Cary S Agler [2], Kimon Divaris [3 4]

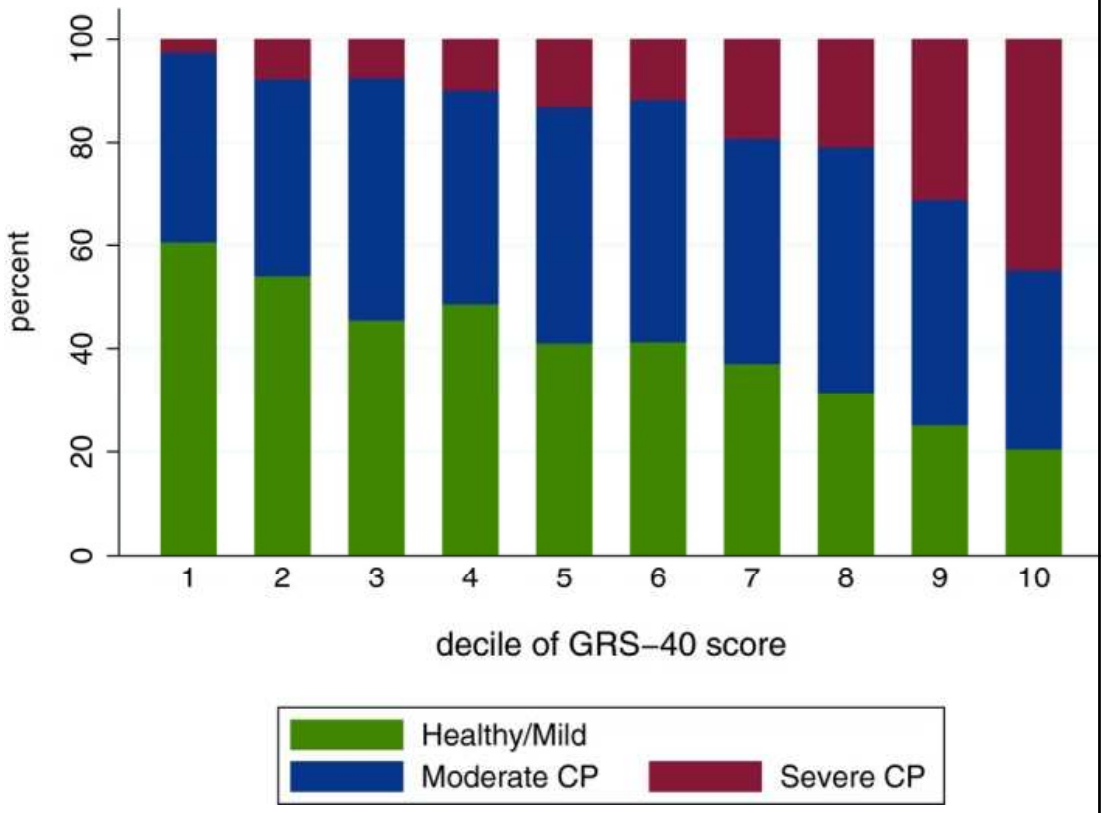
Affiliations + expand

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

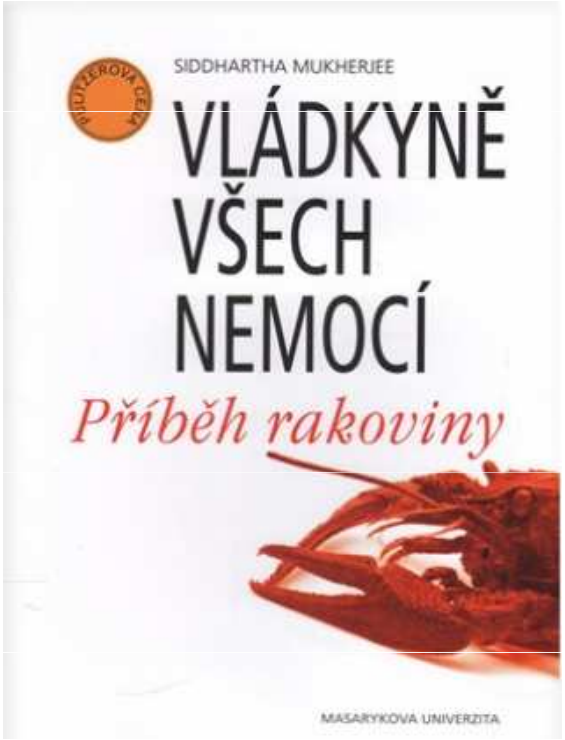
Morelli et al.

- Sami autoři uvádějí tři důvody, proč toto skóre je třeba i dále upravit
  - SNP použité v této studii byli asociované pouze na jedné kohortě pacientů – nemusí tedy platit pro všechna etnika, na jiné kohortě. Pro zavedení do praxe je potřeba nejprve zvalidovat a zreplikovat výsledky.
  - Participanti jen v středním věku, s Evropsko-Americkými předky
  - Další faktory, jiné než genetické, se účastní na rozvoji nemoci v dutině ústní (návyky, socioekonomický status, přístup k zubní péči,…)
- Snaha vytvořit univerzální GRS, které by bylo schopno určit jedince s větší predispozicí pro danou nemoc. Tito jedinci by byli pod častější kontrolou svého lékaře, mohou upravit návyky,…

Závěr

- Shrnutí před GWASové éry
- Co to jsou GWAS, jaké jsou jejich výhody a nevýhody
- Shrnutí recentních GWAS studií
- Využití informací z GWAS pro tvorbu GRS

Doporučená literatura



Příběhy vědy: rakovina Bi0001
Příběhy vědy: gen Bi0002
https://www.ceskatelevize.cz/porady/10441294653-hyde-park-civilizace/220411058090919/

MUNI MED