Ako rozumieť genetickej rôznorodosti?

Mišo Hledík

IST Austria, Barton & Tkačik groups

GJH

19. Marec, 2018


I S T AUSTRIA
Institute of Science and Technology

Mat, fyz, inf v biológii?

- $80\ 000\ 000\ 000$ neurónov, ako sú poprepájané?
  - Elektrické napätie v čase zo stoviek z nich naraz
- Genóm $ATGC \ldots CTGA$, dlhý $3\ 000\ 000\ 000$ báz
  - Skladanie sekvencie z miliónov chybových utržkov
  - Porovnávanie DNA $100\ 000$ jedincov
- $20\ 000$ rôznych génov, $\sim 10\ 000\ 000$ *bežných* mutácií
  - Meranie koncentrácie $20\ 000$ druhov RNA naraz v $10\ 000$ bunkách naraz

Veľké otázky genetiky

Genotyp → fenotyp

Do akej miery sú vlastnosti dedičné? (Nature vs. Nurture)

Kde v genóme sú zakódované?

Akým spôsobom sa efekty mutácií skladajú?

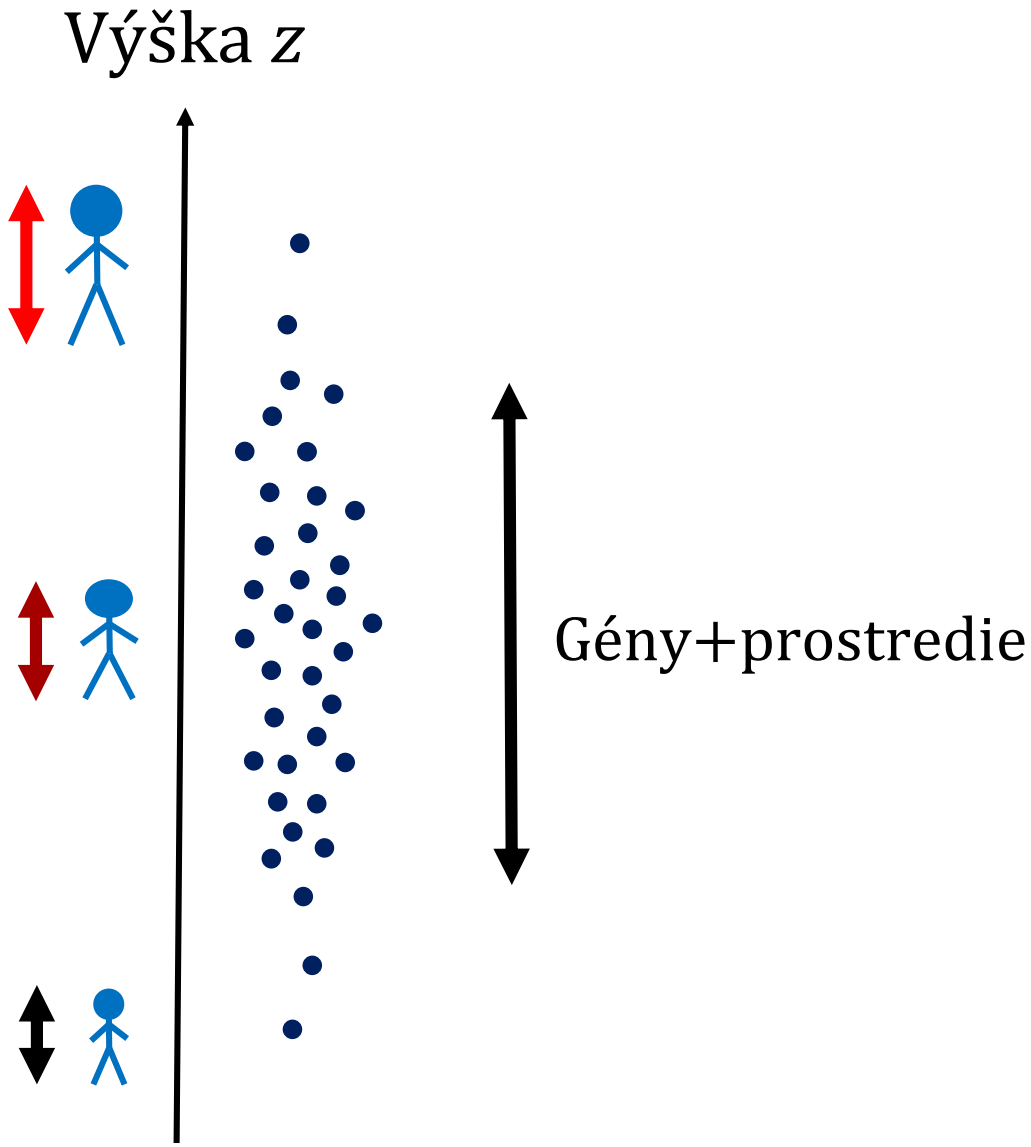
Zopár silných vs. veľa slabých efektov?

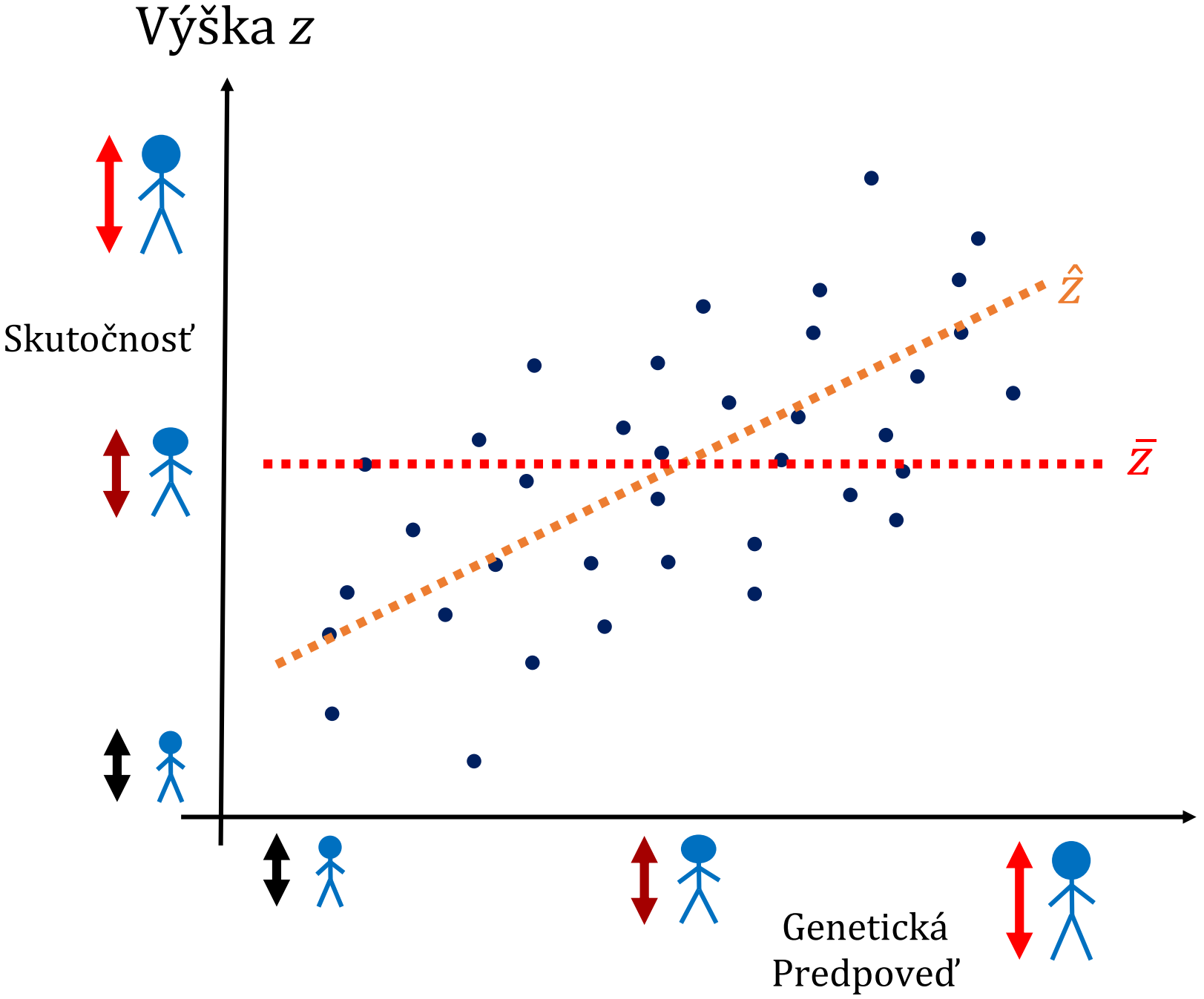
Bežné vs. vzácne mutácie?

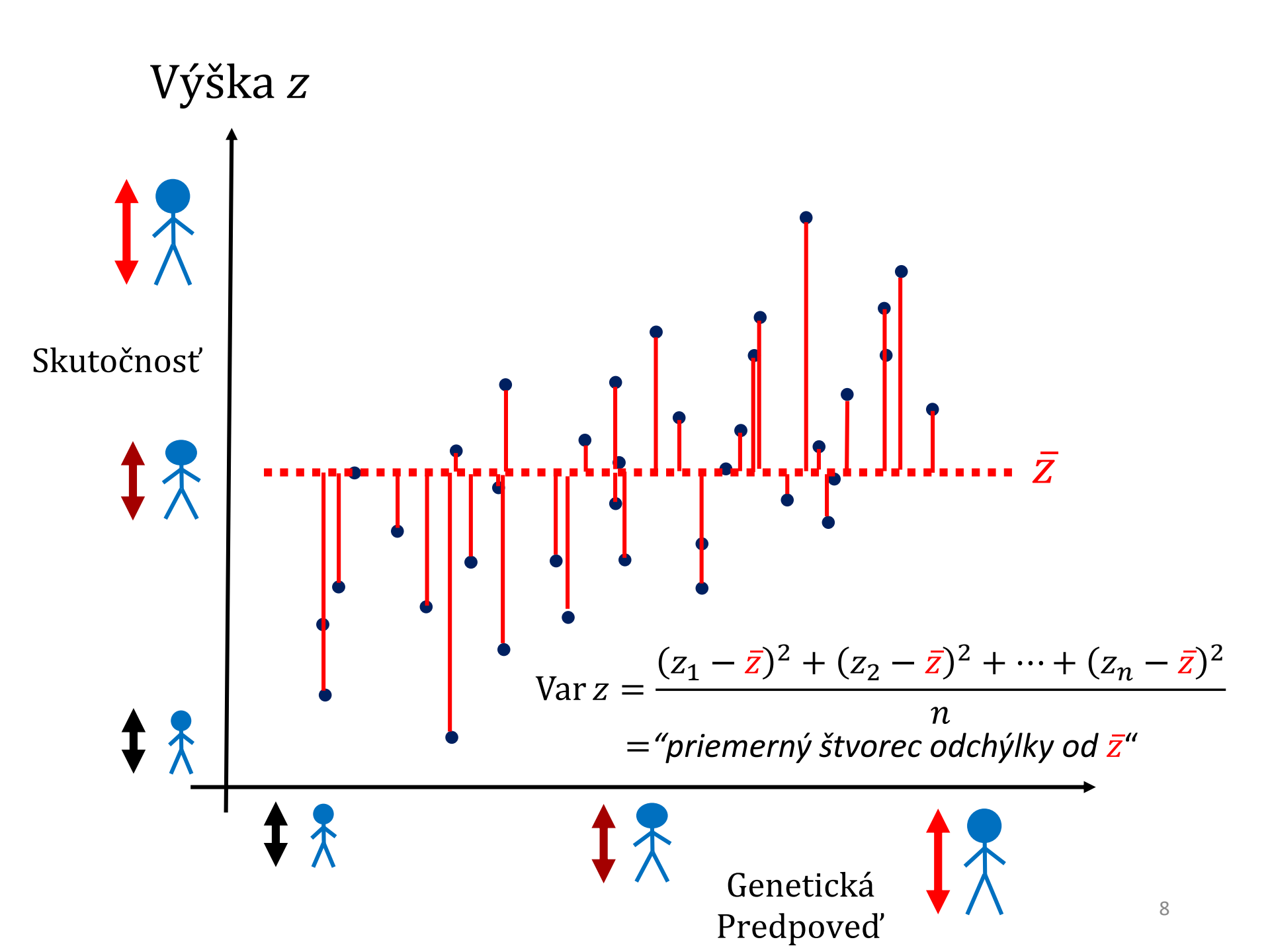
Porozumenie evolúcii, individualizovaná diagnostika a liečba

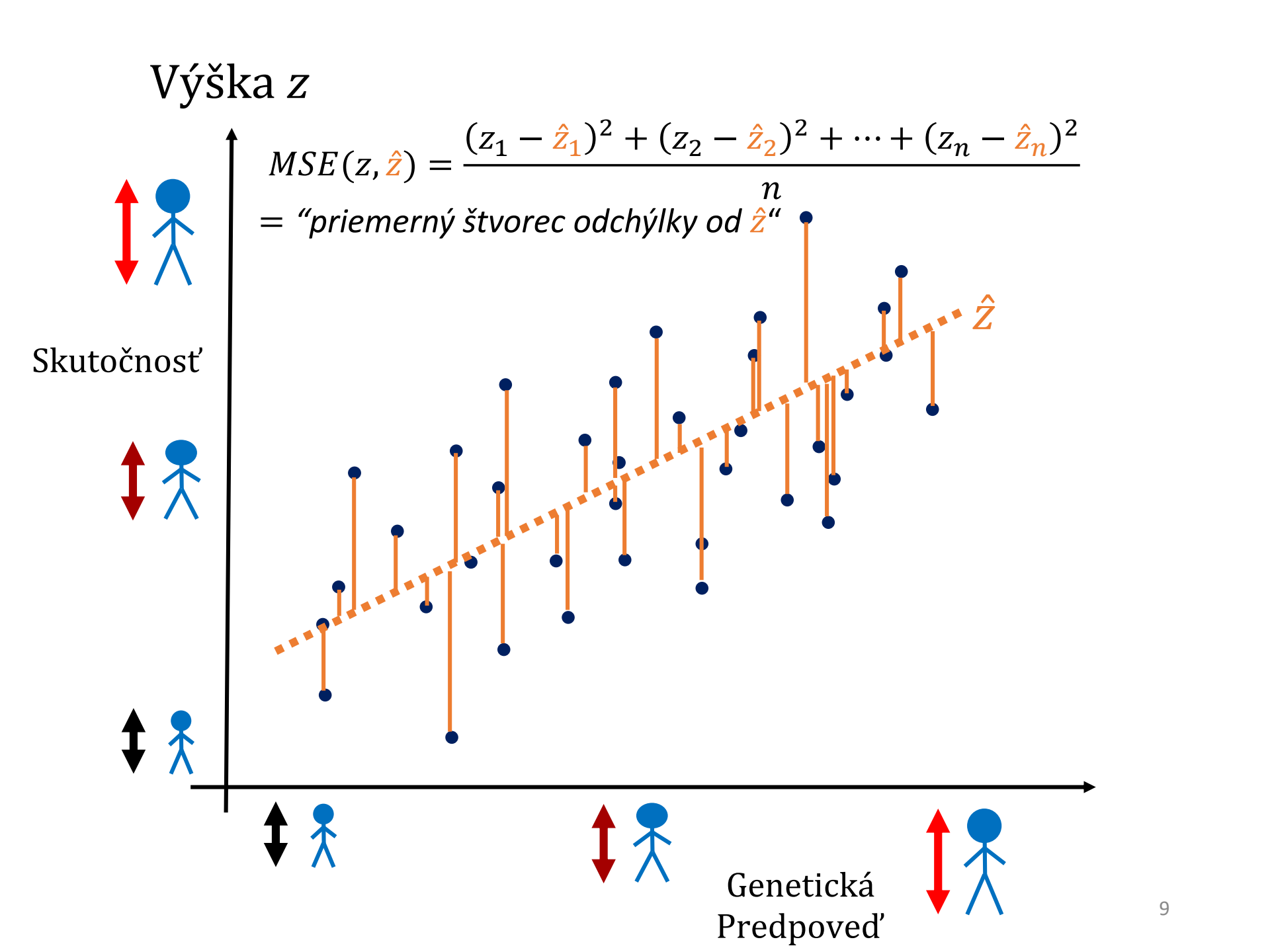
https://www.genomicsengland.co.uk/the-100000-genomes-project/understanding-genomics/

Klasická genetika


Výška z
Gény+prostredie


Výška z
Skutočnosť
$\hat{z}$
$\bar{z}$
Genetická
Predpoveď


Výška z
Skutočnosť
$\bar{z}$
$\text{Var } z = \frac{(z_1 - \bar{z})^2 + (z_2 - \bar{z})^2 + \cdots + (z_n - \bar{z})^2}{n}$
$=$“priemerný štvorec odchýlky od $\bar{z}$“
Genetická
Predpoveď

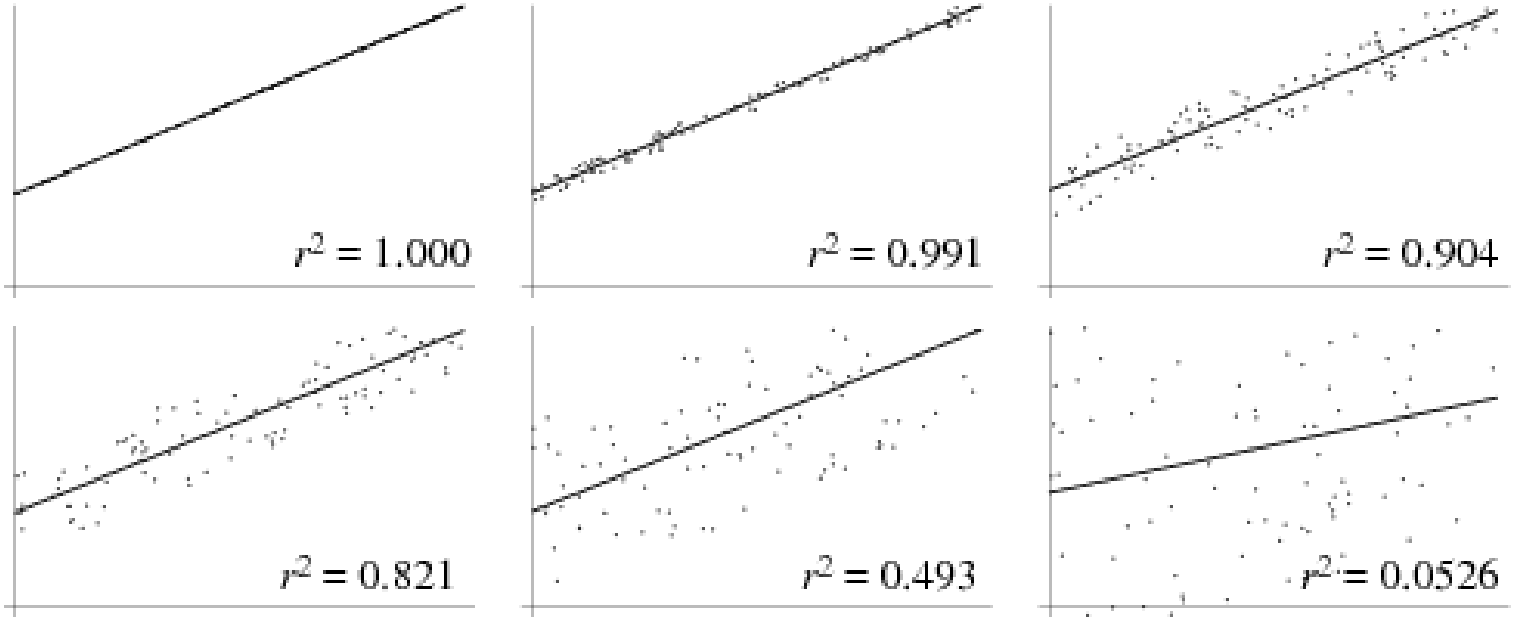

Výška z
$MSE(z,\hat{z}) = \frac{(z_1-\hat{z}_1)^2+(z_2-\hat{z}_2)^2+\cdots+(z_n-\hat{z}_n)^2}{n}$
$=$ "priemerný štvorec odchýlky od $\hat{z}$"
$\hat{z}$
Skutočnosť
Genetická
Predpoveď

Dedičnosť: $H^2 = 1 - \frac{MSE(z,\hat{z})}{\mathrm{Var}\, z}$

=„podiel variancie vysvetlený genetikou“

Aditívny model: $z = z_{mut1} + z_{mut2} + \cdots + z_{Prostredie}$

Dedičnosť $h^2 \approx 2(r^2_{jednovaječné} - r^2_{dvojvaječné\ dvojičky})$



http://mathworld.wolfram.com/CorrelationCoefficient.html

Klasická genetika

Základné pravidlo biológie:

Vždy je to zložitejšie, než si myslíš.

- Neaditívne efekty génov?
  - Dáta: aspoň 50% dedičnosti je aditívna*
- Efekty génov závisia na prostredí?
- Dedičnosť sa v čase mení

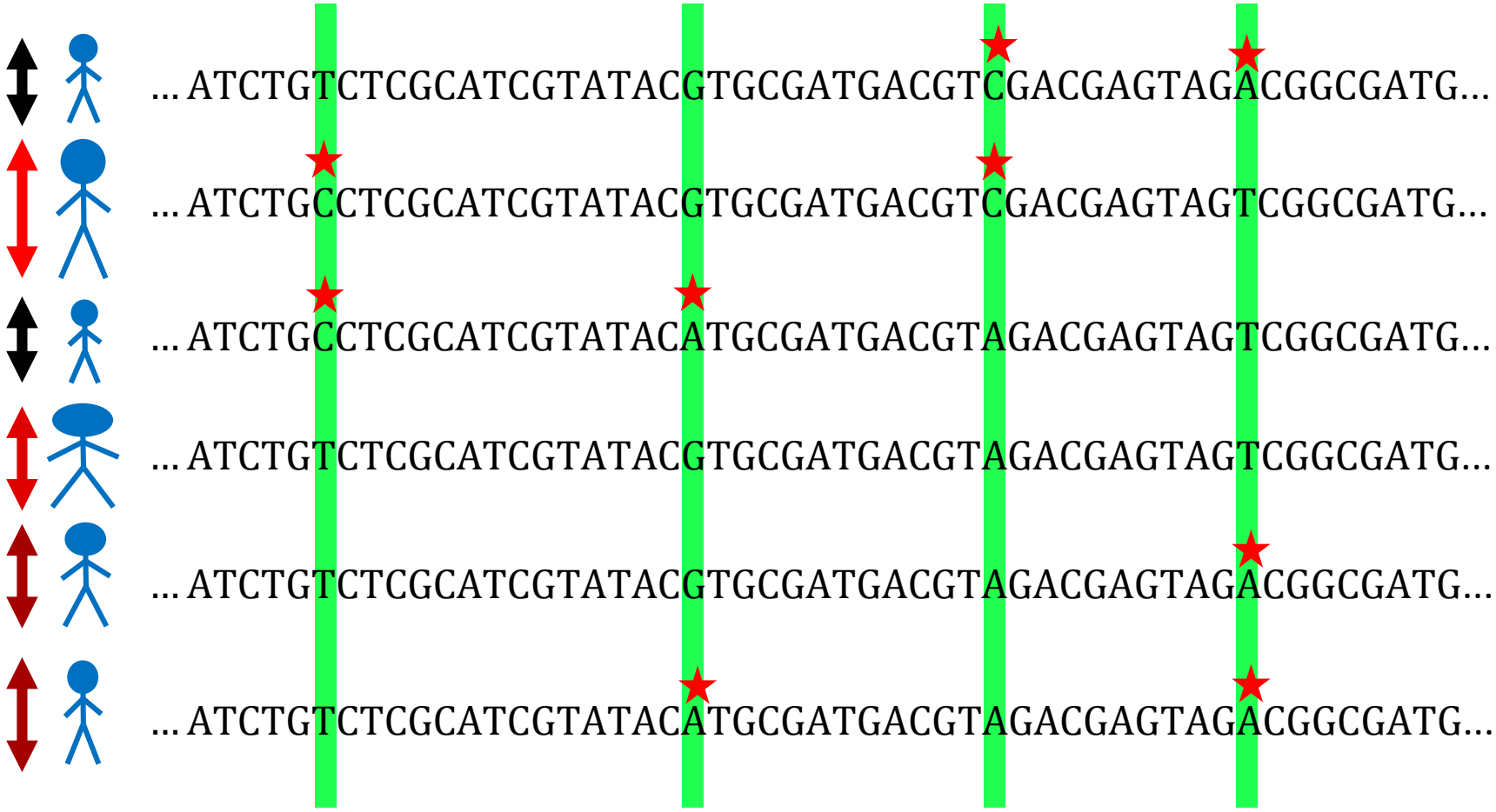
Trait or Disease	h^2 Pedigree Studies
Type 1 diabetes	0.9[98]
Type 2 diabetes	0.3–0.6[100]
Obesity (BMI)	0.4–0.6[101,102]
Crohn's disease	0.6–0.8[103]
Ulcerative colitis	0.5[103]
Multiple sclerosis	0.3–0.8[104]
Ankylosing spondylitis	>0.90[105]
Rheumatoid arthritis	0.6[107]
Schizophrenia	0.7–0.8[108]
Bipolar disorder	0.6–0.7[108]
Breast cancer	0.3[110]
Von Willebrand factor	0.66–0.75[112,113]
Height	0.8[115,116]
Bone mineral density	0.6-0.8[117]
QT interval	0.37–0.60[119,120]
HDL cholesterol	0.5[122]
Platelet count	0.8[123]

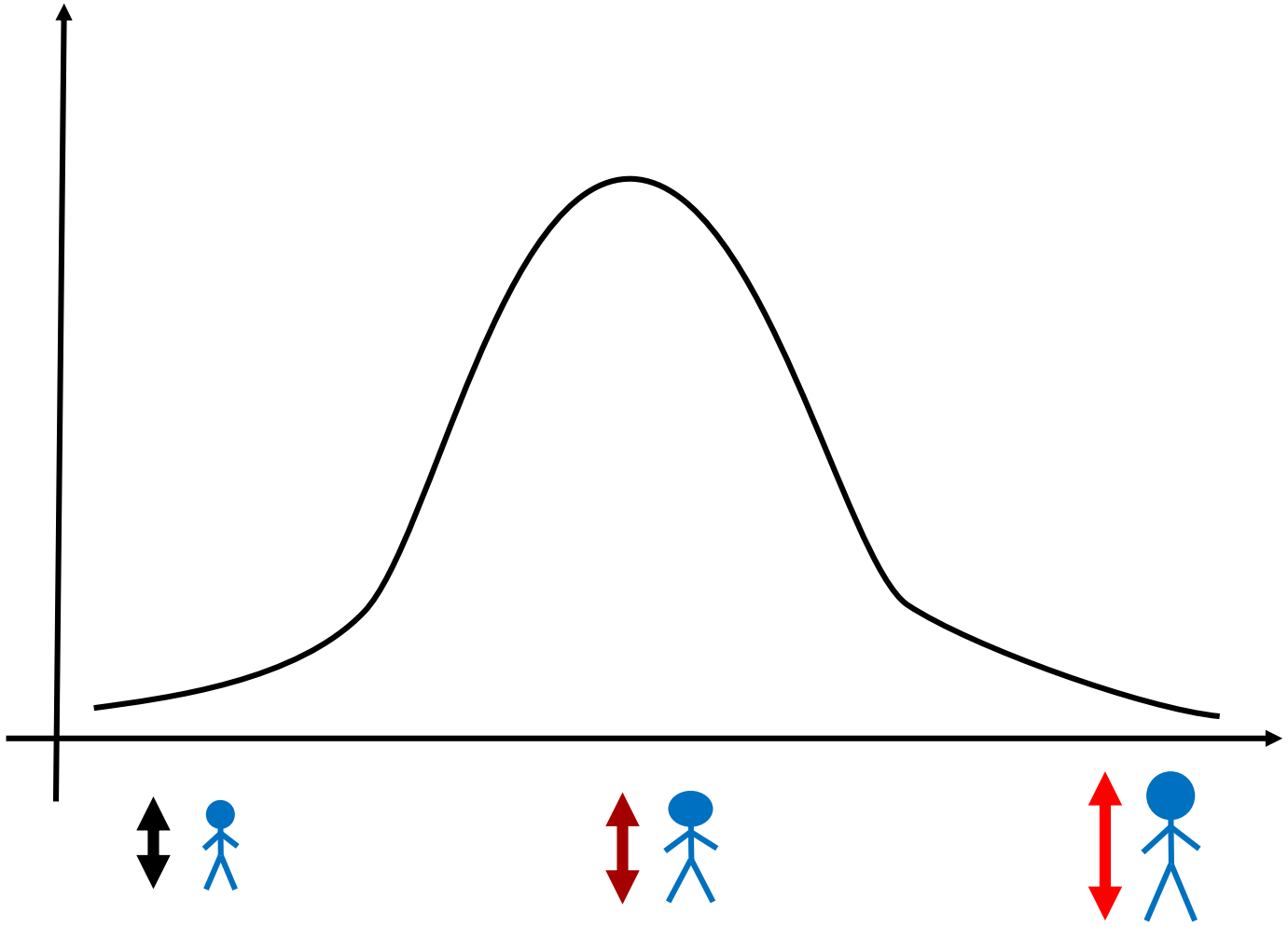
Peter M. Visscher, Matthew A. Brown, Mark I. McCarthy, and Jian Yang (2012). *Five Years of GWAS Discovery.* The American Journal of Human Genetics 90, 7–24

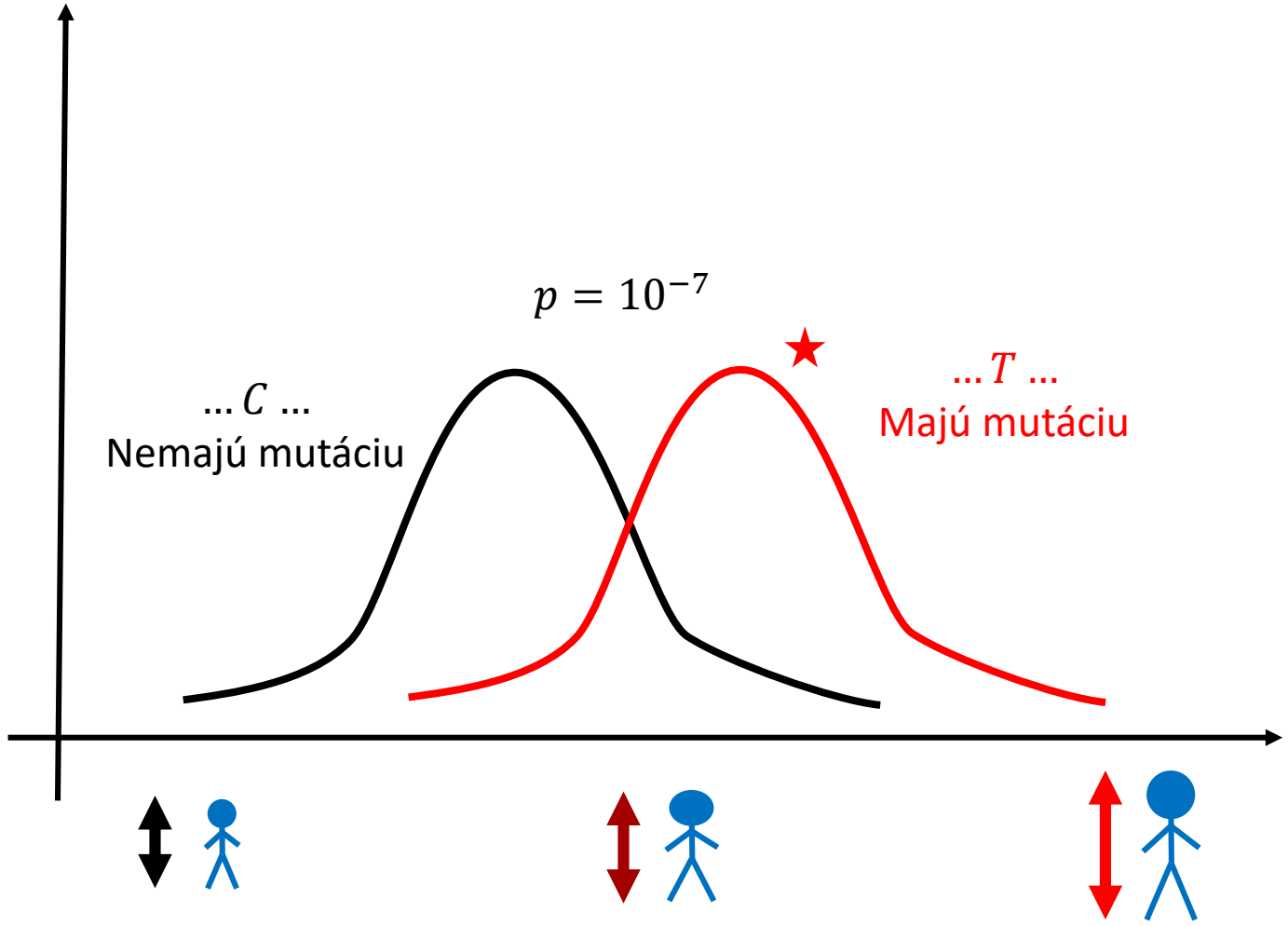
*William G. Hill, Michael E. Goddard, Peter M. Visscher, *Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits*. PLoS Genet 4(2): e1000008 (2008). doi:10.1371/journal.pgen.1000008

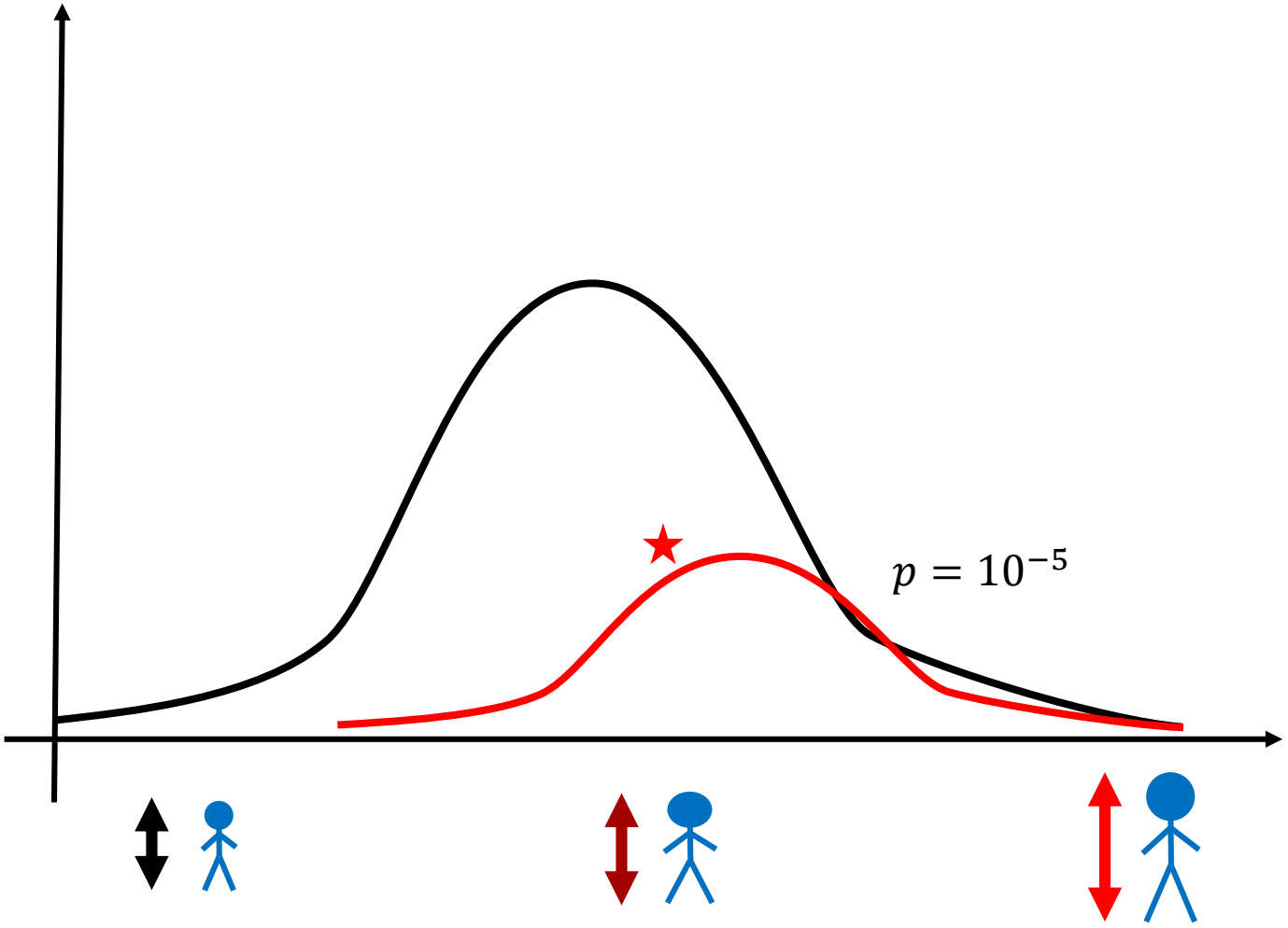
GWAS

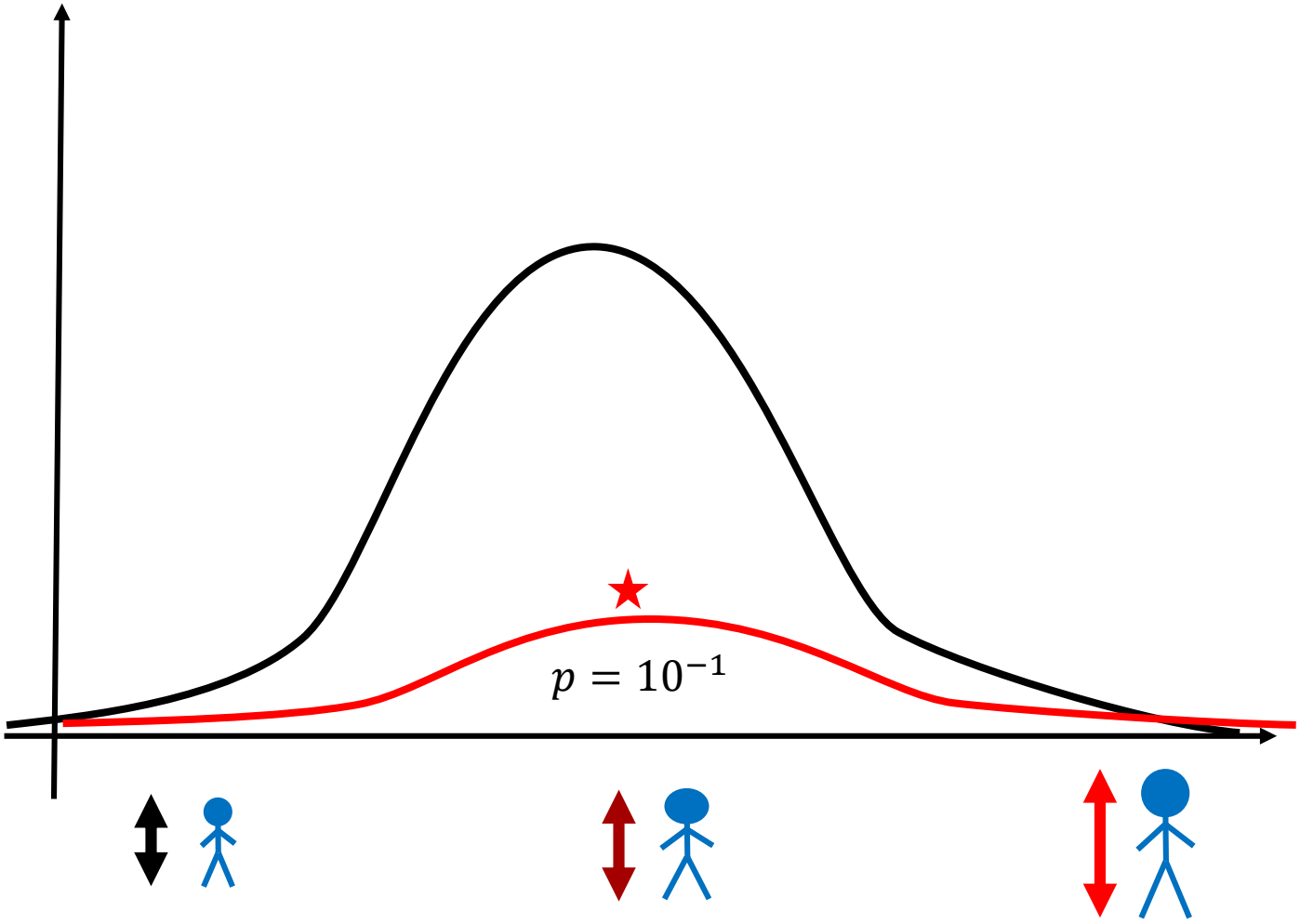
Genome-wide association studies

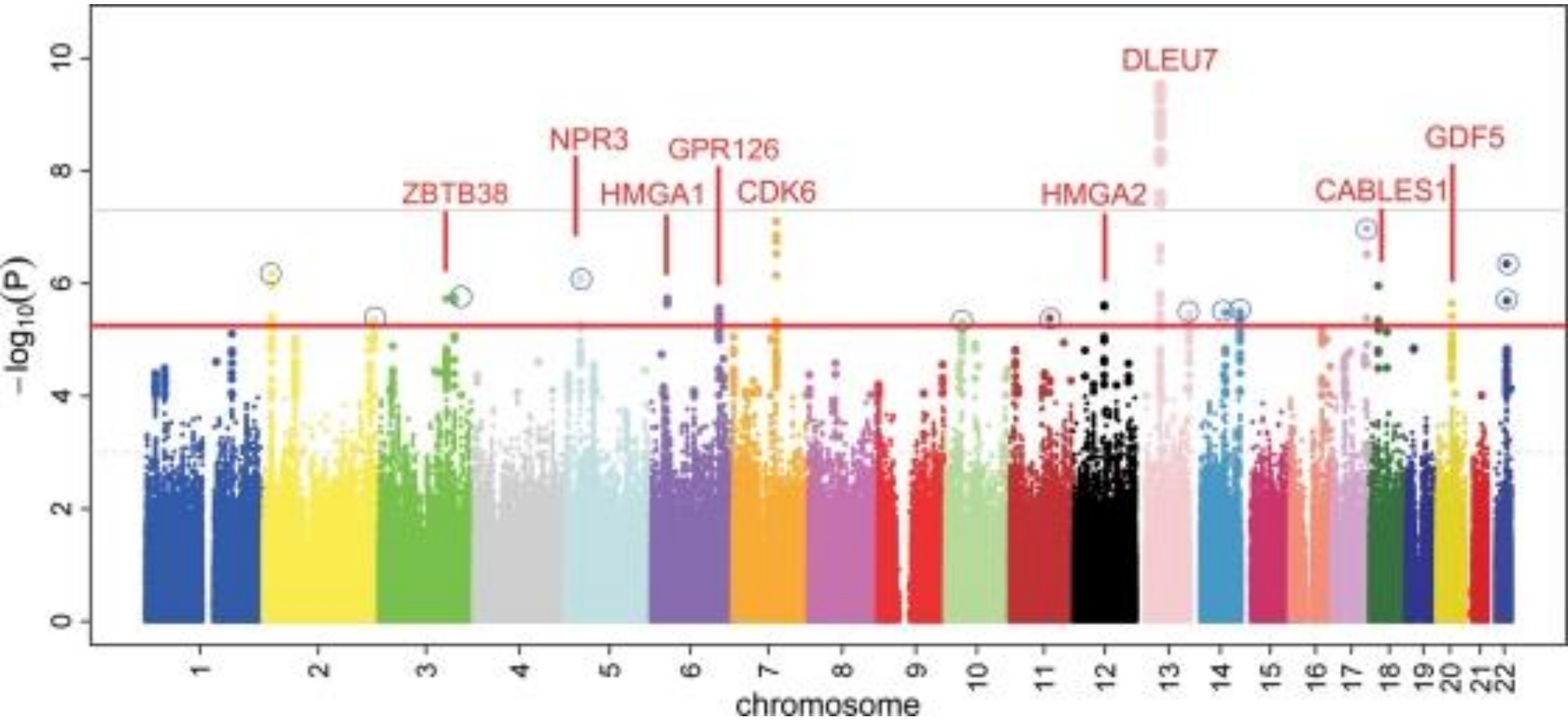

…ATCTGTCTCGCATCGTATACGTGCGATGACGTCGACGAGTAGACGGCGATG…
…ATCTGCCTCGCATCGTATACGTGCGATGACGTCGACGAGTAGTCGGCGATG…
…ATCTGCCTCGCATCGTATACATGCGATGACGTAGACGAGTAGTCGGCGATG…
…ATCTGTCTCGCATCGTATACGTGCGATGACGTAGACGAGTAGTCGGCGATG…
…ATCTGTCTCGCATCGTATACGTGCGATGACGTAGACGAGTAGACGGCGATG…
…ATCTGTCTCGCATCGTATACATGCGATGACGTAGACGAGTAGACGGCGATG…



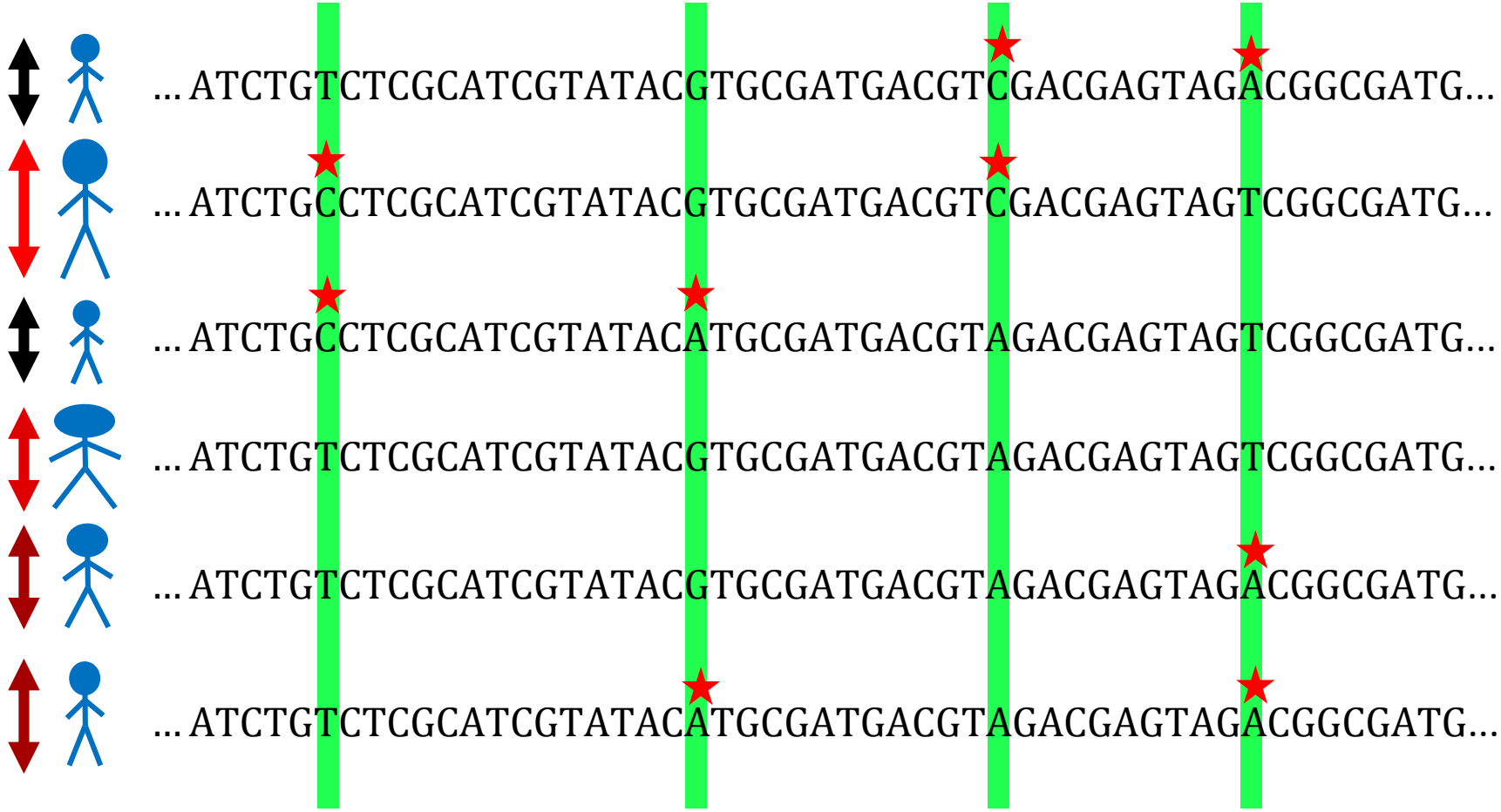

$p = 10^{-7}$
… C …
Nemajú mutáciu
… T …
Majú mutáciu


$p = 10^{-5}$


$p = 10^{-1}$



Estrada, Karol et al. (2009). *A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation*. Human molecular genetics. 18. 3516-24. 10.1093/hmg/ddp296.



$$\hat{z}_1 = 0\alpha_1 + 0\alpha_2 + 1\alpha_3 + 1\alpha_4 + \cdots$$
$$\hat{z}_2 = 1\alpha_1 + 0\alpha_2 + 1\alpha_3 + 0\alpha_4 + \cdots$$
$$\vdots$$
$$\hat{z}_i = x_{i1}\alpha_1 + x_{i2}\alpha_2 + x_{i3}\alpha_3 + x_{i4}\alpha_4 + \cdots$$

Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h^2 Pedigree Studies	h^2 GWAS Hits[a]	h^2 All GWAS SNPs[b]
Type 1 diabetes	0.9[98]	0.6[99,c]	0.3[12]
Type 2 diabetes	0.3–0.6[100]	0.05-0.10[34]	
Obesity (BMI)	0.4–0.6[101,102]	0.01-0.02[36]	0.2[14]
Crohn's disease	0.6–0.8[103]	0.1[11]	0.4[12]
Ulcerative colitis	0.5[103]	0.05[12]	
Multiple sclerosis	0.3–0.8[104]	0.1[45]	
Ankylosing spondylitis	>0.90[105]	0.2[106]	
Rheumatoid arthritis	0.6[107]		
Schizophrenia	0.7–0.8[108]	0.01[79]	0.3[109]
Bipolar disorder	0.6–0.7[108]	0.02[79]	0.4[12]
Breast cancer	0.3[110]	0.08[111]	
Von Willebrand factor	0.66–0.75[112,113]	0.13[114]	0.25[14]
Height	0.8[115,116]	0.1[13]	0.5[13,14]
Bone mineral density	0.6-0.8[117]	0.05[118]	
QT interval	0.37–0.60[119,120]	0.07[121]	0.2[14]
HDL cholesterol	0.5[122]	0.1[57]	
Platelet count	0.8[123]	0.05–0.1[58]	

Peter M. Visscher, Matthew A. Brown, Mark I. McCarthy, and Jian Yang (**2012**). *Five Years of GWAS Discovery.* The American Journal of Human Genetics 90, 7–24

Defining the role of common variation in the genomic and biological architecture of adult human height

A full list of authors and affiliations appears at the end of the article.

[#] These authors contributed equally to this work.

Abstract

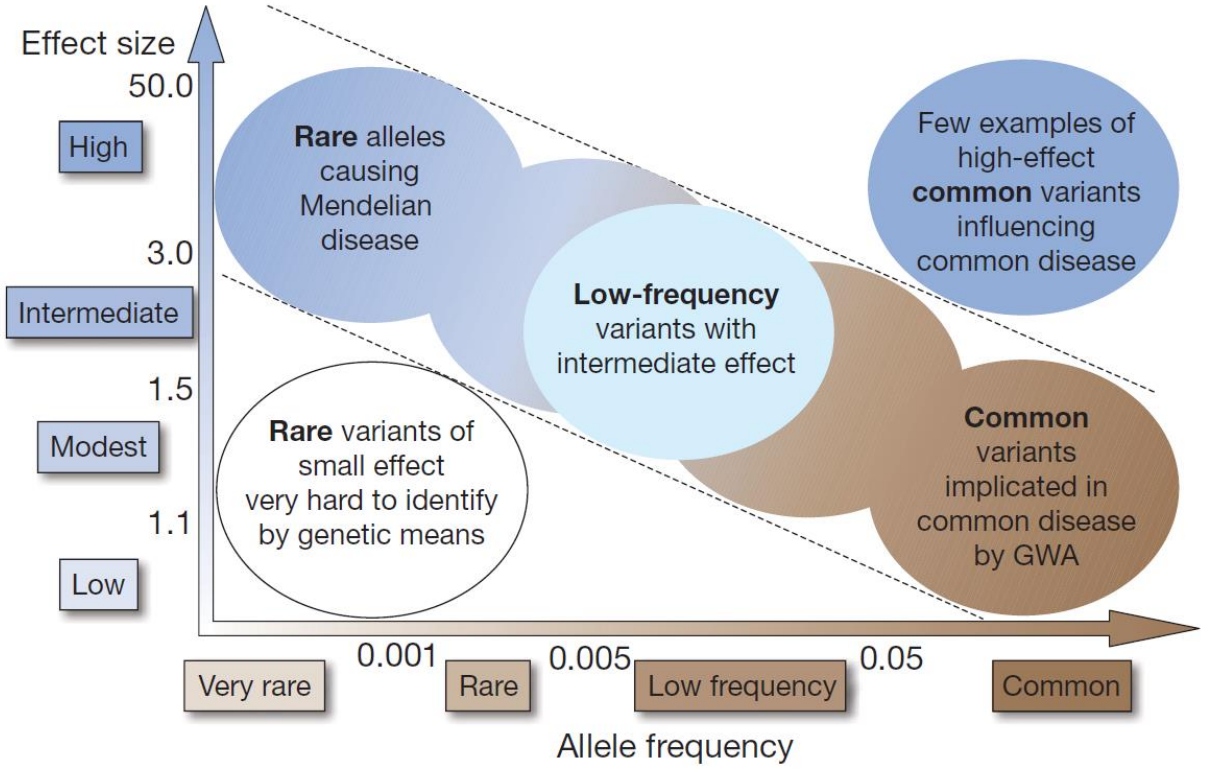
Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explain one-fifth of heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured the majority (60%) of heritability. The 697 variants clustered in 423 loci enriched for genes, pathways, and tissue-types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/beta-catenin, and chondroitin sulfate-related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.

Andrew R Wood et al. (**2014**). Nat Genet. 46(11): 1173–1186. doi:10.1038/ng.3097

GWAS

Základné pravidlo biológie...

- **Veľmi veľa veľmi malých efektov**
- Neaditívne efekty génov? → Deep learning?
- Prostredie, korelácie medzi mutáciami, ...



Teri A. Manolio, Francis S. Collins, Peter M. Visscher (2009). *Finding the missing heritability of complex diseases*. Nature vol. 461, pp. 747–753

Populačná genetika

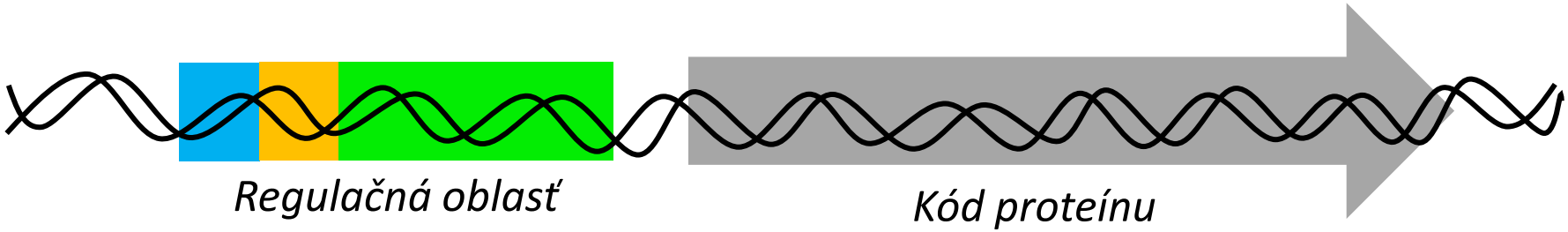
(simulácie v Mathematica)

Čo robia mutácie?

Expresia a regulácia génov

Centrálna dogma: DNA → RNA → Protein

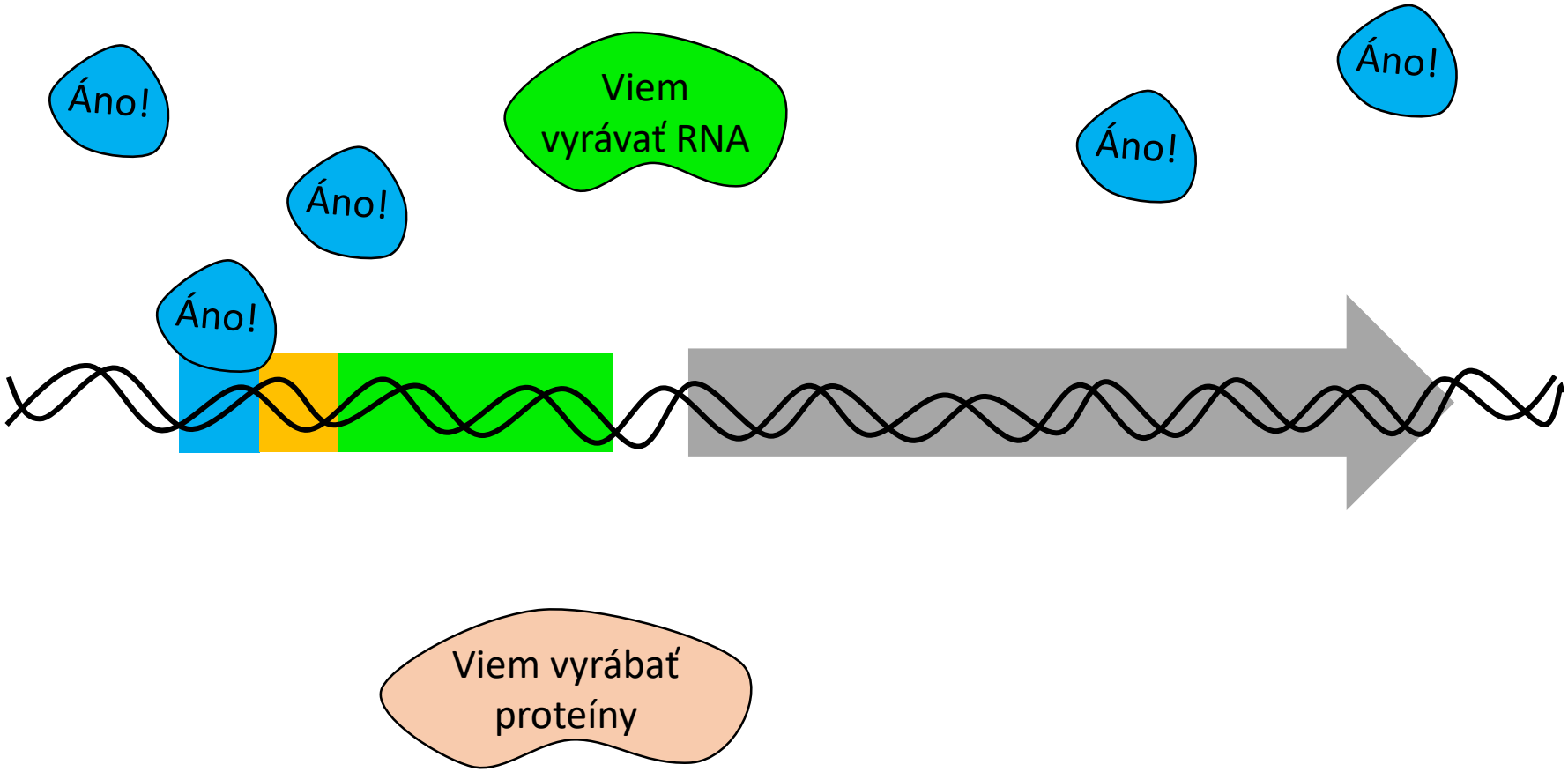
Ja viem vyrábať RNA



Ja viem vyrábať proteíny

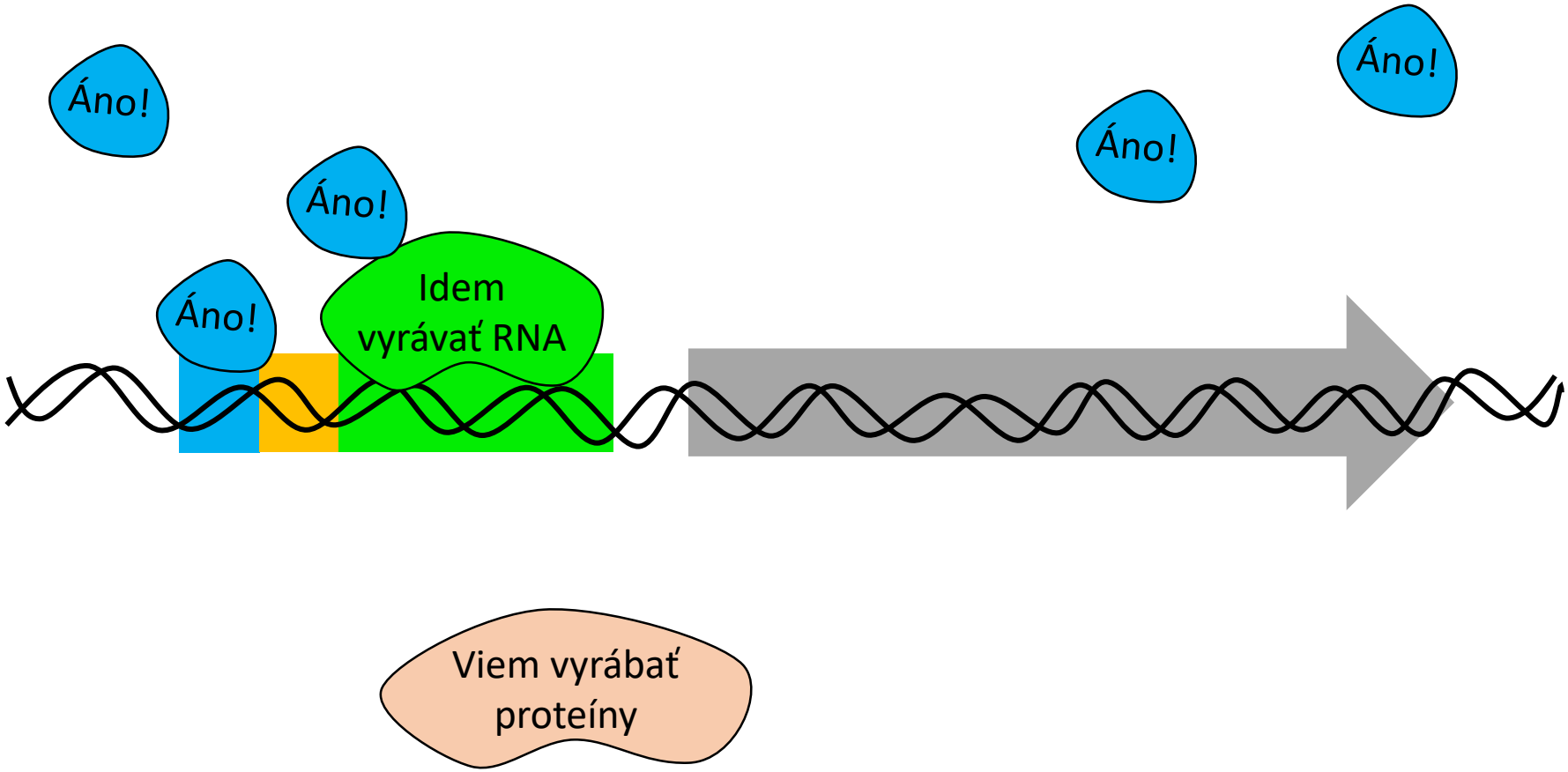
Expresia a regulácia génov

Centrálna dogma: DNA → RNA → Protein



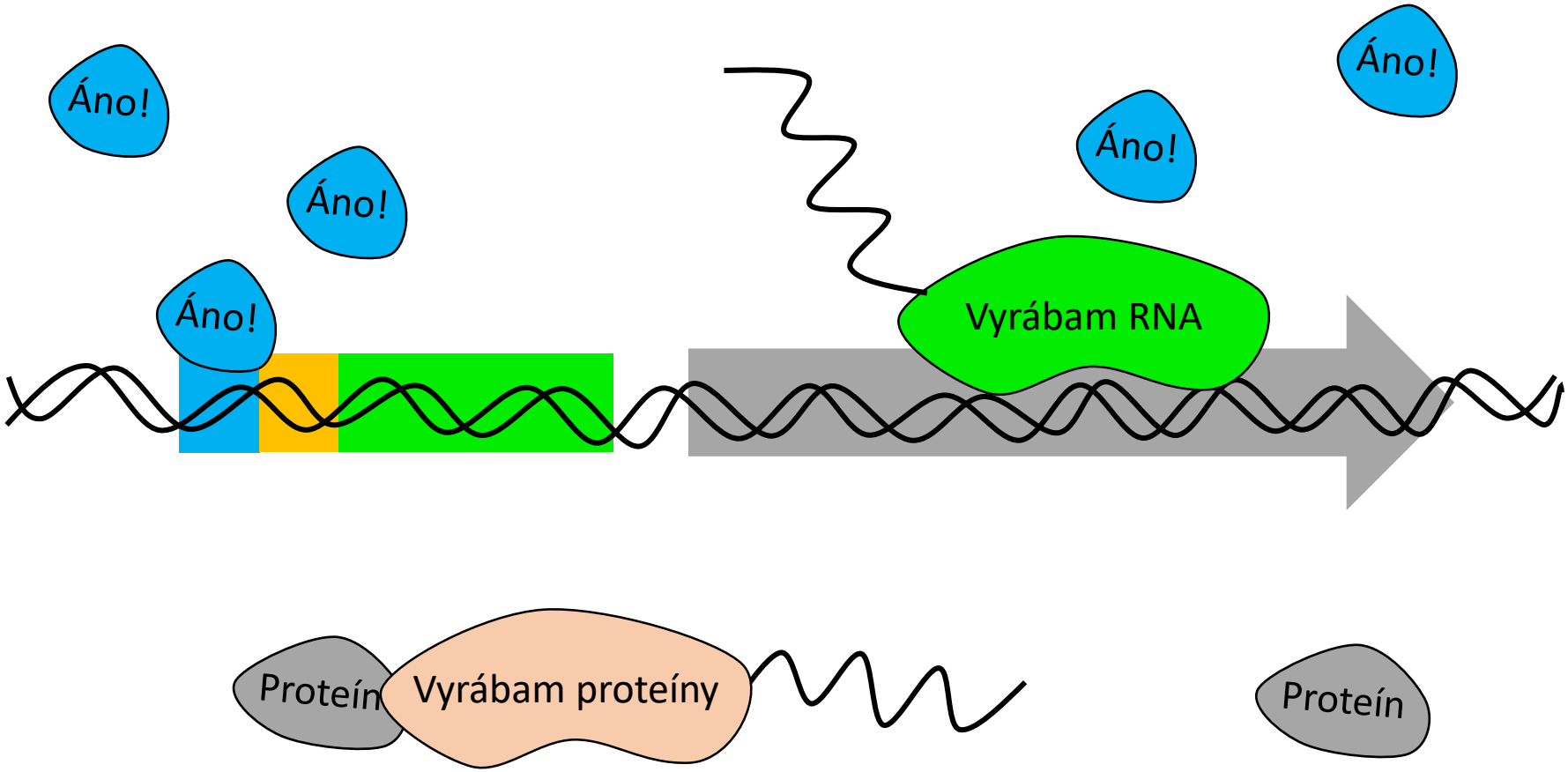
Expresia a regulácia génov

Centrálna dogma: DNA → RNA → Protein



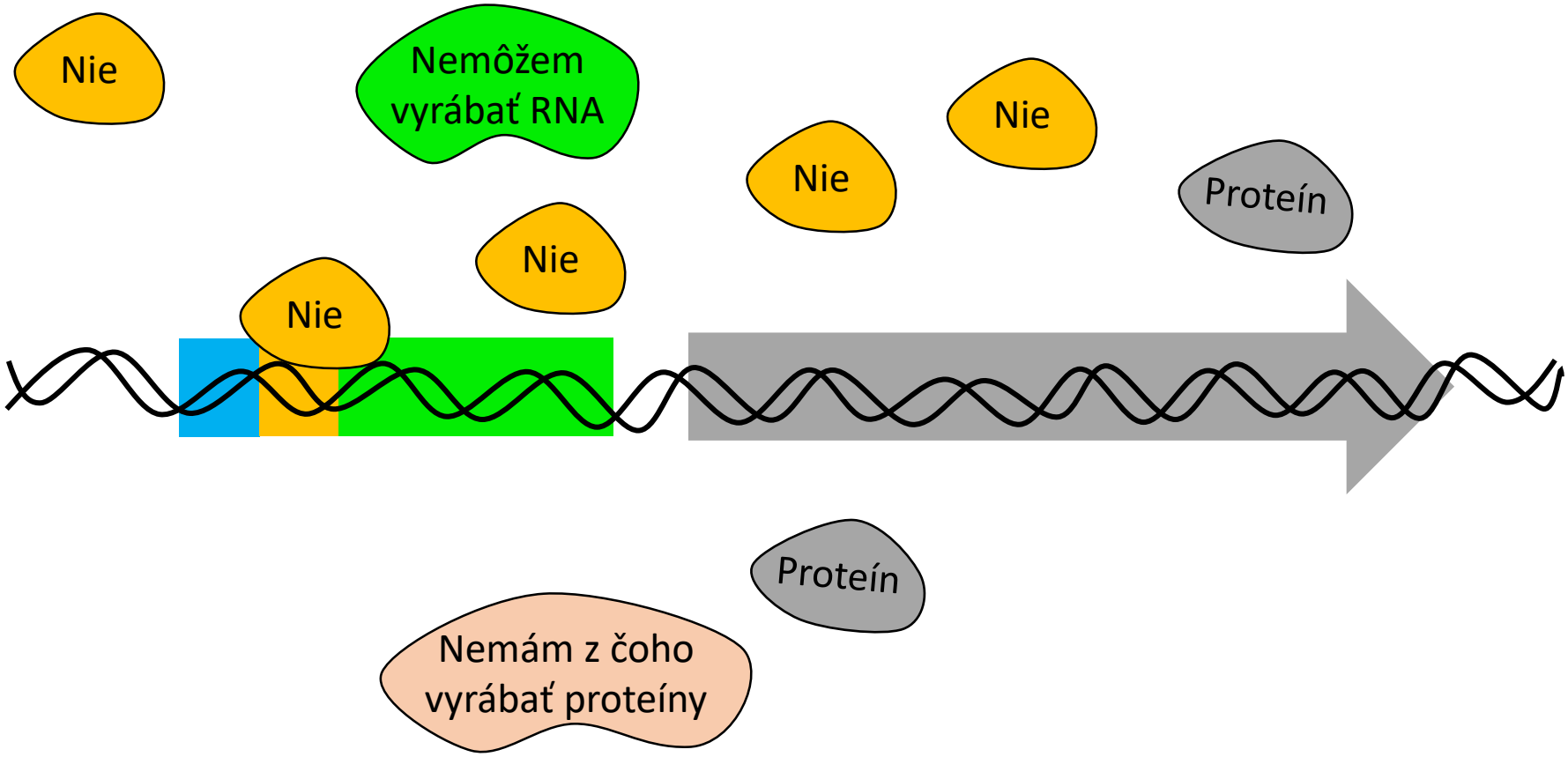
Expresia a regulácia génov

Centrálna dogma: DNA → RNA → Protein

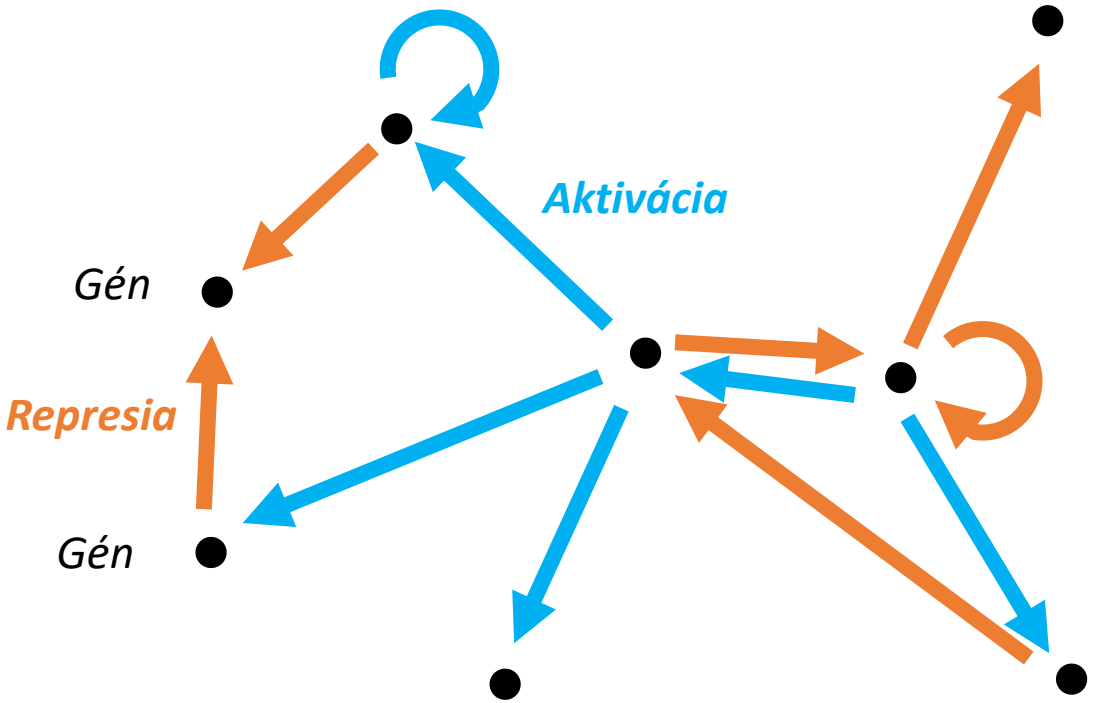


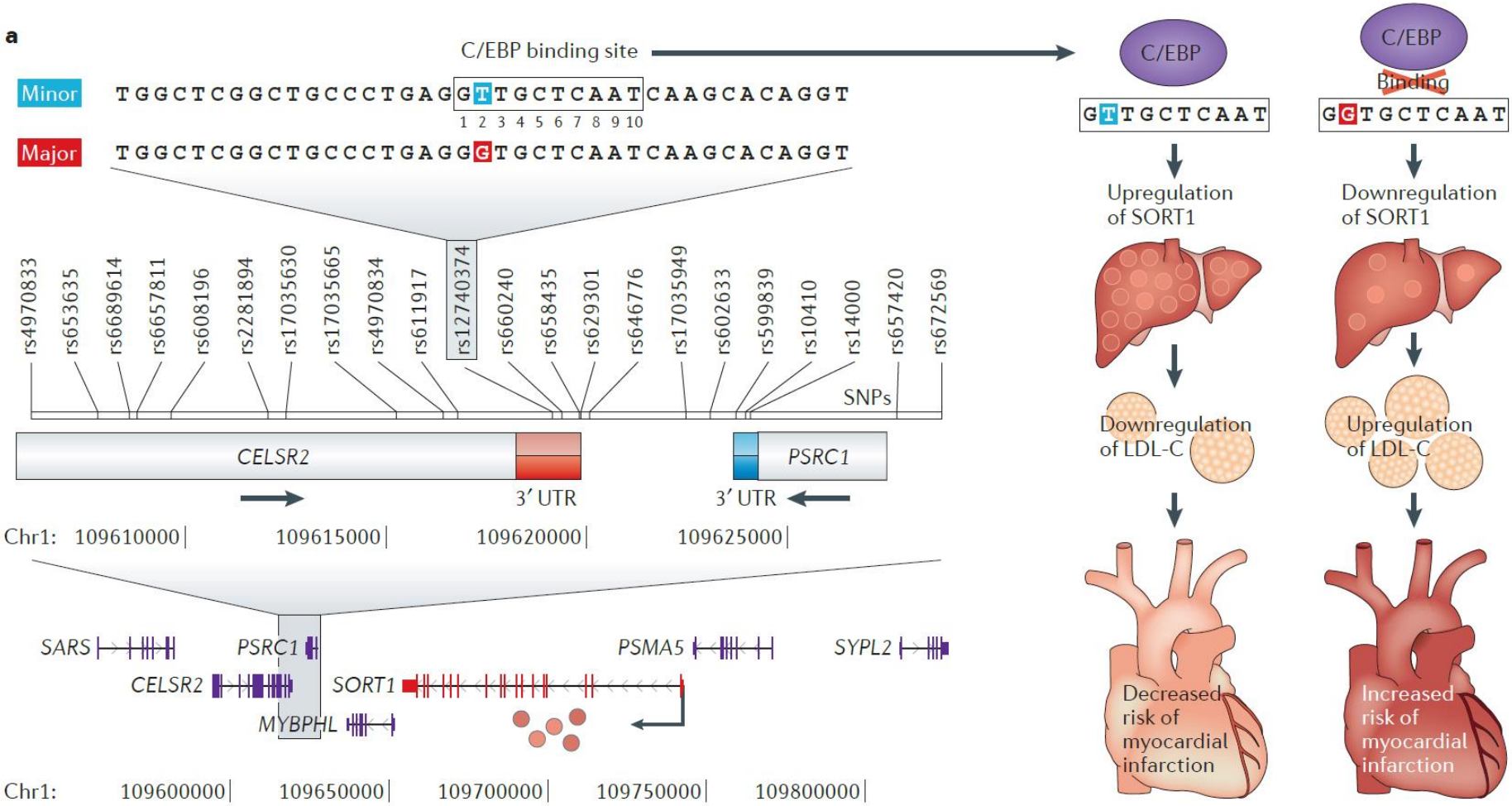
Expresia a regulácia génov

Centrálna dogma: DNA → RNA → Protein



Genetická regulačná sieť





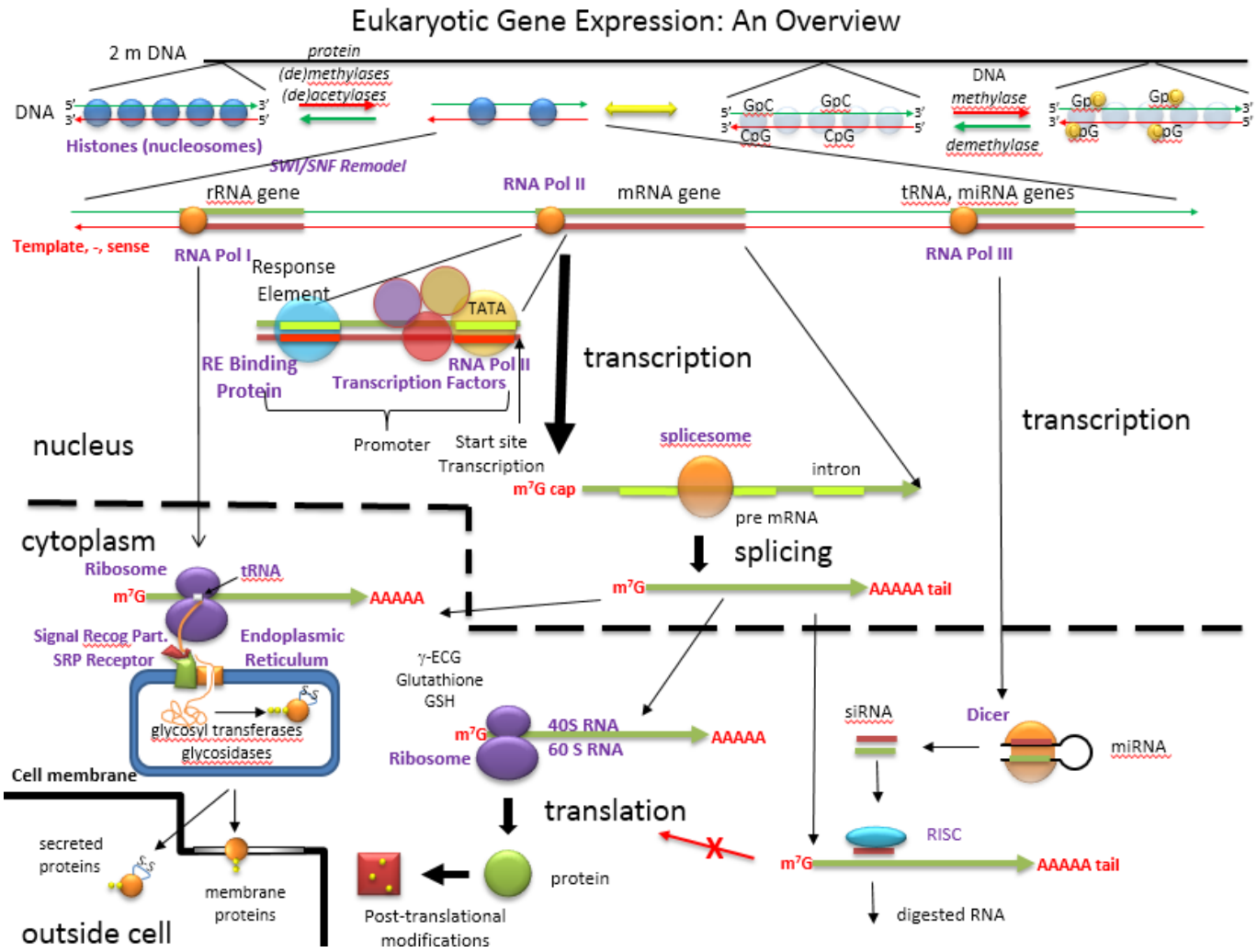
Frank W. Albert, Leonid Kruglyak, *The role of regulatory variation in complex traits and disease*. Nature Reviews Genetics 16, 197{212 (2015). doi:10.1038/nrg3891

Table 1. Examples Linking Variable TF-DNA Binding to Phenotypic Variation Arranged by Date of Characterization

Phenotype	Affected Gene	Causal Variant Position Relative to TSS	Affected Binding Site	TFBS Outcome	Reference(s)
Hereditary persistance of fetal haemoglobin	*HBG*	−175 bp	GATA1; TAL1	Gain	(Martin et al., 1989; Wienert et al., 2015)
Haemophilia B Leyden	*F9*	−20 bp; 10 bp; −6 bp	HNF4a; C/EBPa; OC1/OC2	Loss	(Reijnen et al., 1992; Crossley and Brownlee, 1990; Funnell et al., 2013)
Haemophilia B Brandenburg	*F9*	−26 bp	AR	Loss	(Crossley et al., 1992)
Delta-thalassemia	*HBD*	−77 bp	GATA1	Loss	(Matsuda et al., 1992)
Duffy blood antigen/chemokine receptor expression	*DARC*	−46 bp	GATA1	Loss	(Tournamille et al., 1995)
Familial combined hyperlipidemia	*LPL*	−39 bp	OCT1	Loss	(Yang et al., 1995)
Bernard-Soulier syndrome	*GP1BB*	−133 bp	GATA1	Loss	(Ludlow et al., 1996)
Osteoporosis	*COLIAI*	+2 kb	Sp1	Gain	(Grant et al., 1996)
Maturity-onset diabetes of the young	*HNF1A*	−58 bp	HNF4A	Loss	(Gragnoli et al., 1997)
Asthma	*IL10*	−509 bp	YY1	Gain	(Hobbs et al., 1998)
Pyruvate kinase deficiency	*PKLR*	−72 bp	GATA1	Loss	(Manco et al., 2000)
Congenital erythropoietic porphyria	*UROS*	−70 bp; −90 bp	GATA1; CP2	Loss	(Solis et al., 2001)
Psoriasis	*SLC9A3R1*	−237 bp	RUNX1	Loss	(Helms et al., 2003)
Systemic lupus erythematosus	*FASLG*	−844 bp	CEBPB	Loss	(Wu et al., 2003)
Esophageal cancer	*COX-2*	−1195 bp	c-MYB	Gain	(Zhang et al., 2005)
Treacher Collins syndrome	*TCOF1*	−346 bp	YY1	Loss	(Masotti et al., 2005)
Alpha-thalassemia	*HBA*	−13 bp	GATA1	Gain	(De Gobbi et al., 2006)
Holoprosencephaly	*SHH*	−460 kb	SIX3	Loss	(Jeong et al., 2008)
Various cancers	*TERT*	−187 bp	ETS2	Loss	(Xu et al., 2008)
Nonsyndromic cleft lip	*IRF6*	−14 kb	AP2	Loss	(Rahimov et al., 2008)
Pierre Robin syndrome	*SOX9*	−1.44 Mb	MSX1	Loss	(Benko et al., 2009)

Bart Deplancke, Daniel Alpern, Vincent Gardeux (2016). *The Genetics of Transcription Factor DNA Binding Variation*. Cell 166/3, 538 – 554.

Základné pravidlo biológie...



Henry Jakubowski (2016). https://employees.csbsju.edu/hjakubowski/classes/ch331/bind/olbindtransciption.html

Physical model of gene regulation

Reaction rate dynamics:

$$\frac{d}{dt}x_i = \dot{x}_i = -\lambda_i x_i + m_i f_i(x)$$

Product i concentration — x_i

Degradation — $-\lambda_i x_i$

Synthesis rate — m_i

Occupancy of active state — $f_i(x)$

Monod-Wyman-Changeaux – binding in thermodynamic equilibrium:

$$f_i(x) = \frac{1}{1 + e^{-\Delta G_i(x)}} = \sigma\big(\Delta G_i(x)\big)$$

Free energy difference: active vs. inactive state — $\Delta G_i(x)$

$$\Delta G_i(x) = \sum_{jk} \log \frac{1 + x_j / K^A_{ijk}}{1 + x_j / K^I_{ijk}} + \Delta G^0_i$$

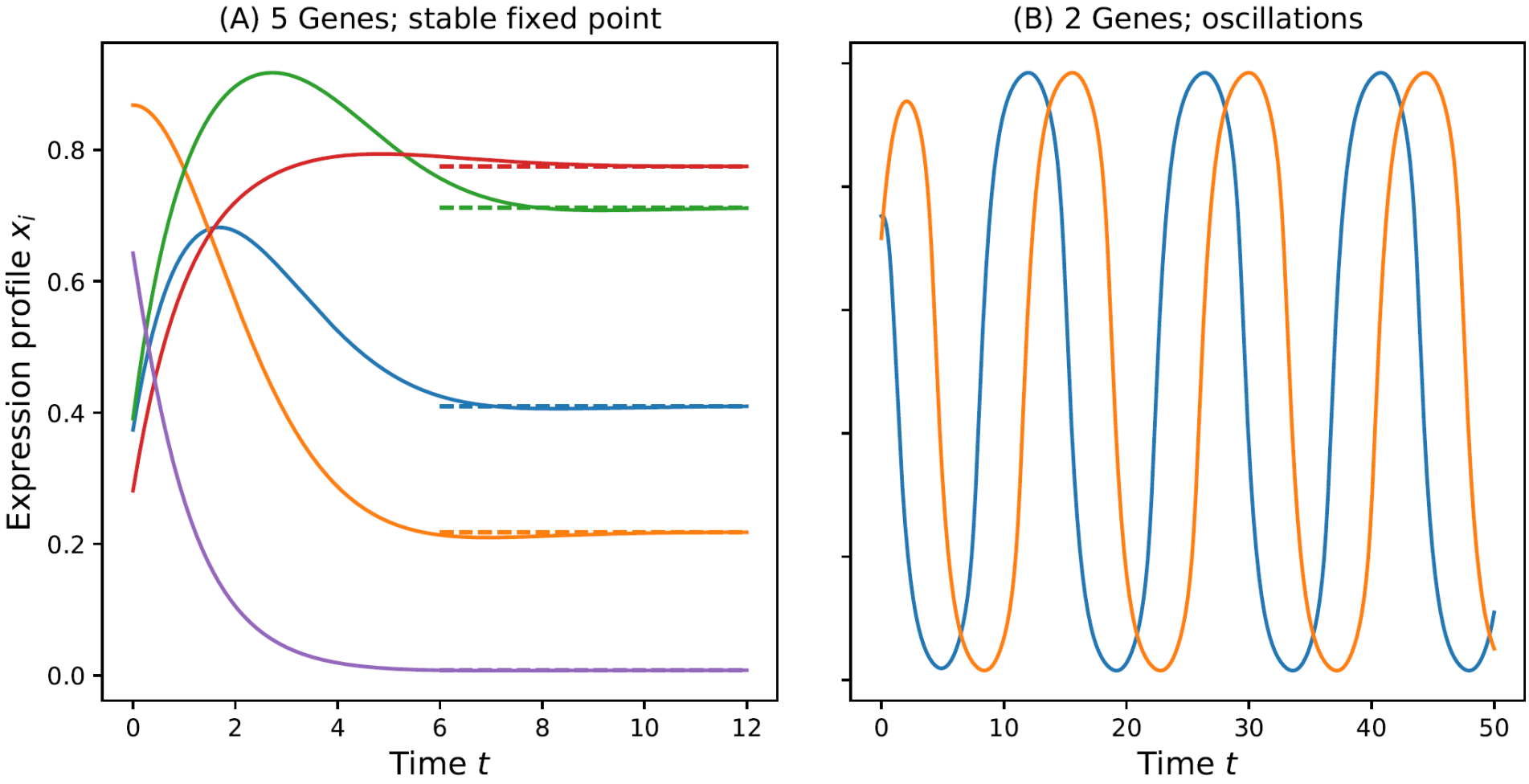
Other products — x_j

TF dissociation rate gene i, TF j, BS k — K^A_{ijk}, K^I_{ijk}

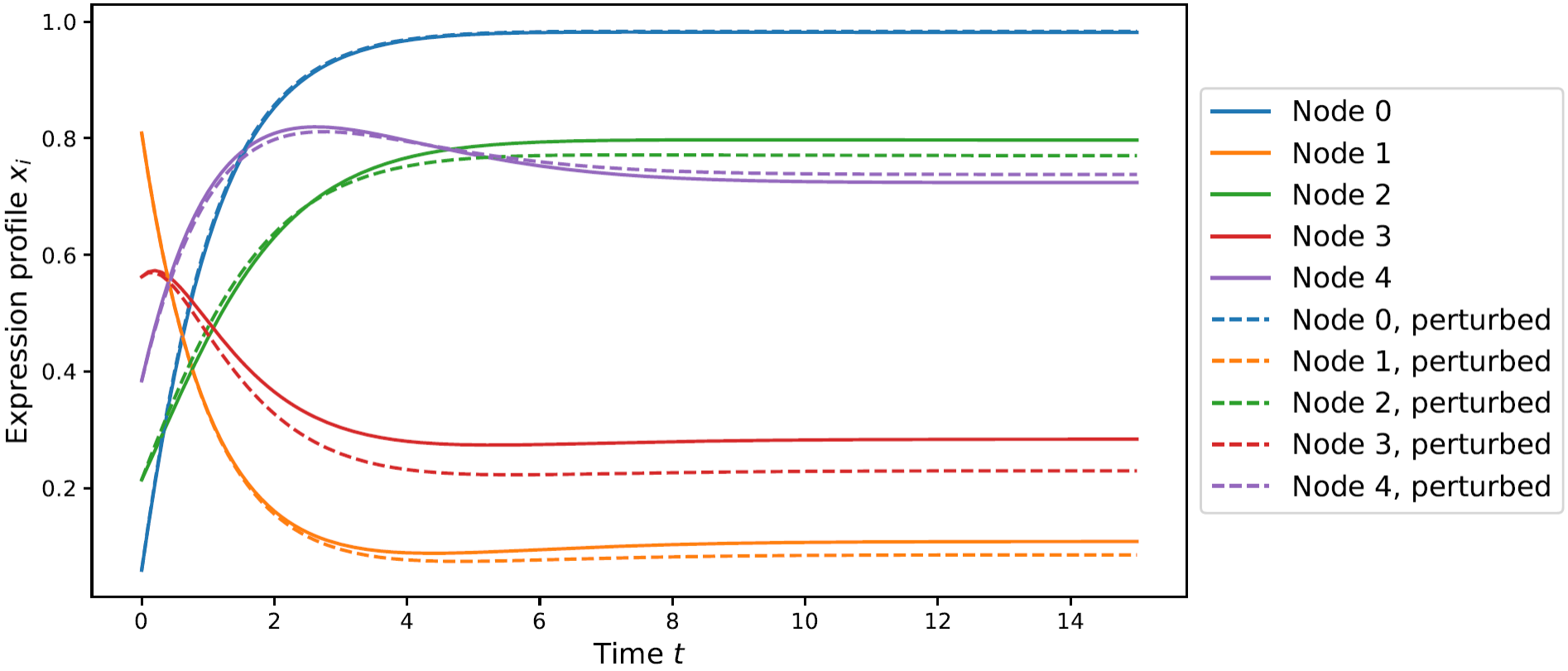
ΔG_i at zero TFs bound — ΔG^0_i

J. Monod, J. Wyman, J.-P. Changeux, J. Mol. Biol 12, 88–118 (1965)
L. Bintu et al., Curr Opin Genet Dev. 15(2): 116–124 (2005)

Správanie genetických sietí

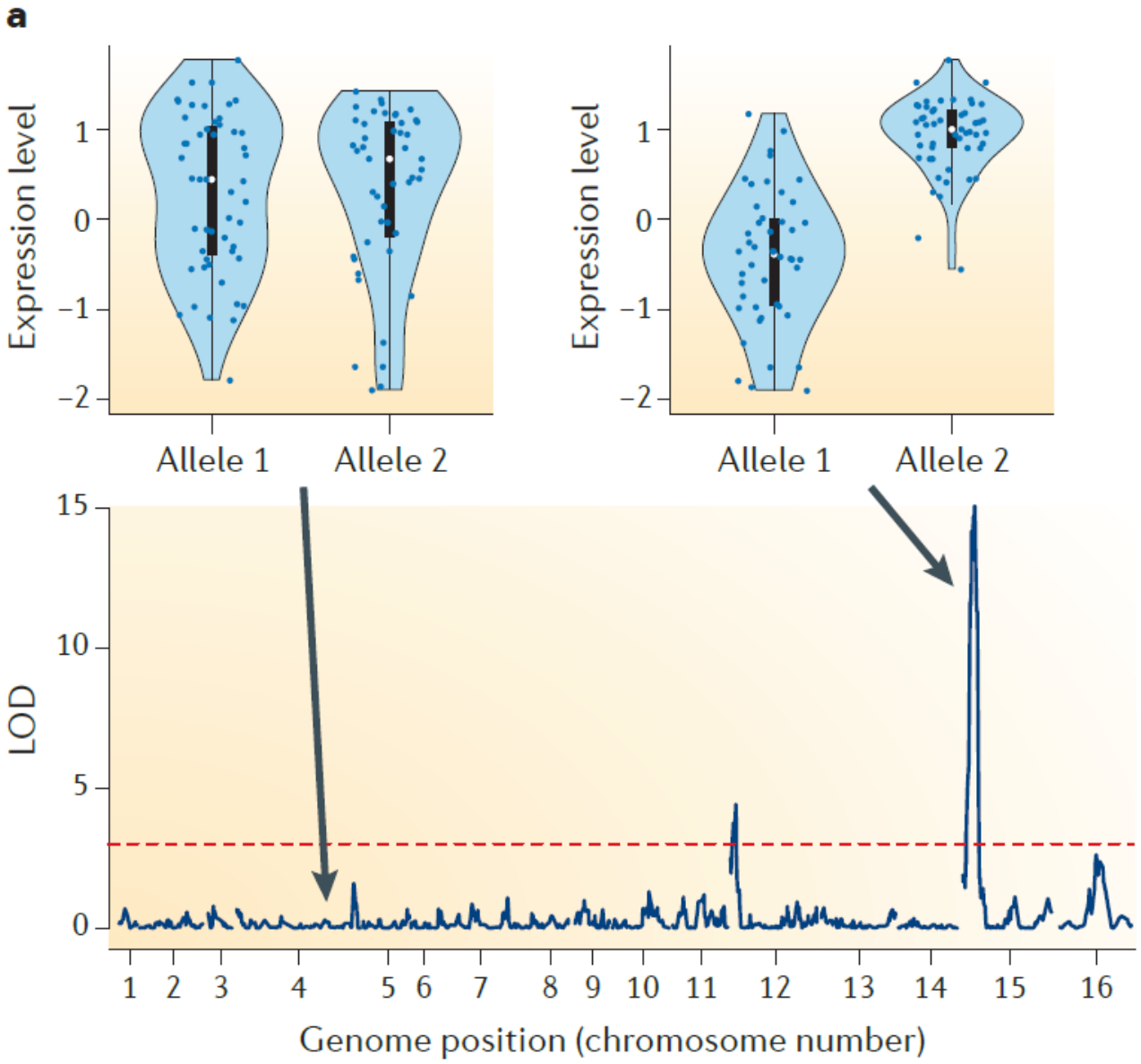


Efekt mutácie

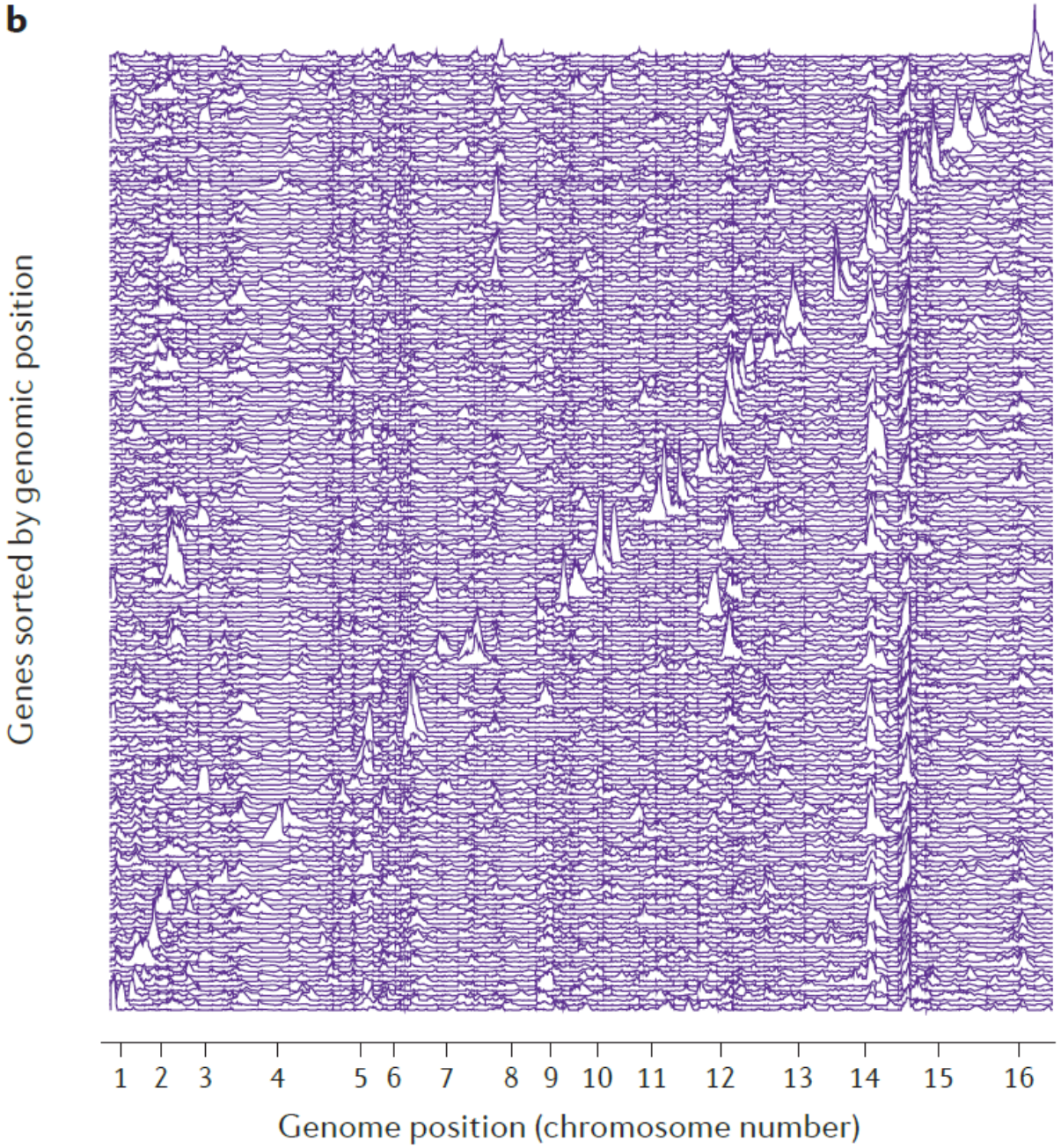


Mutácia: $R \to R + \delta R$

Expresia: $\tilde{x} \to \tilde{x} + \delta\tilde{x}(\delta R)$



Frank W. Albert, Leonid Kruglyak, *The role of regulatory variation in complex traits and disease*. Nature Reviews Genetics 16, 197{212 (2015). doi:10.1038/nrg3891



Frank W. Albert, Leonid Kruglyak, *The role of regulatory variation in complex traits and disease*. Nature Reviews Genetics 16, 197{212 (2015). doi:10.1038/nrg3891