Л е к ц и я № 7

Полногеномный анализ ассоциаций

Разработал профессор П. М. Васильев
Кафедра фармакологии и биоинформатики

Для студентов, обучающихся по направлению 06.03.01 «Биология»
профили Биохимия, Генетика
при изучении дисциплины «Биоинформатика»

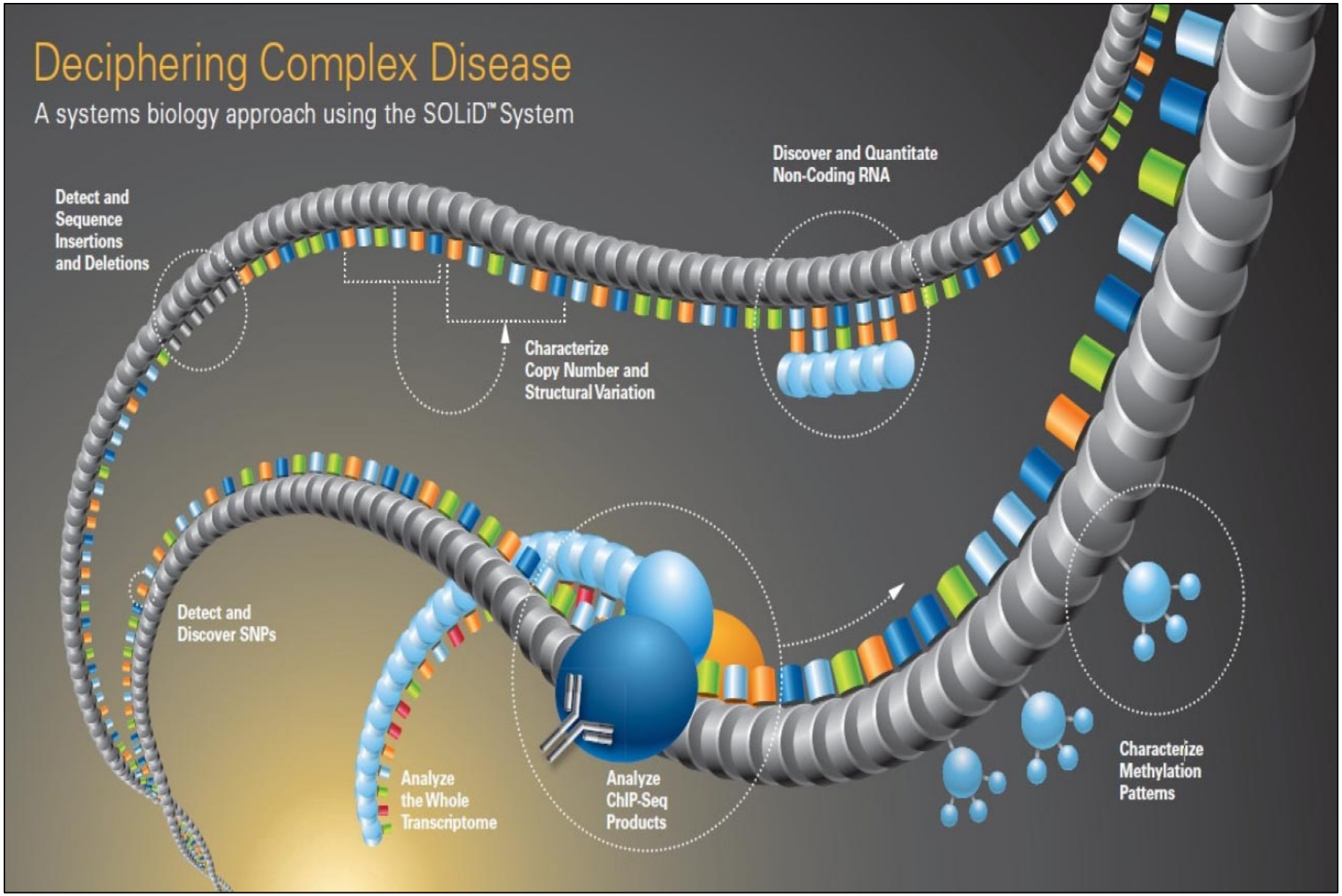
План лекции

- **Полногеномный анализ ассоциаций GWAS**
- **Моногенные заболевания**
- **Однонуклеотидный полиморфизм SNP**
- **Мультифакторные заболевания**

Характеристики генома

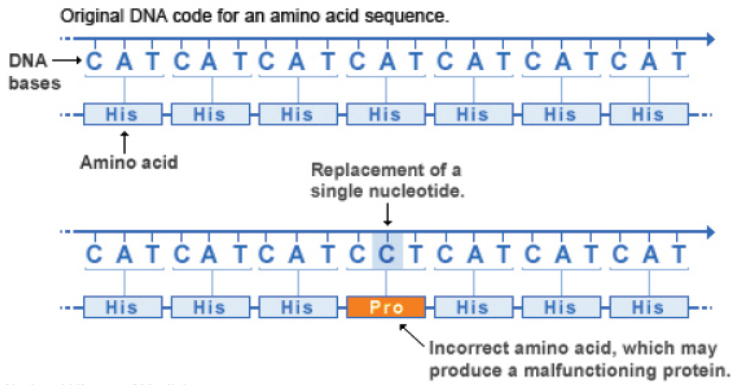
- 24 хромосомы
- ~3.2 млрд. пар нуклеотидов
- ~25 тыс. генов, ~1.3% генома
- Один ген — 1000-3000 оснований
- Два генома идентичны на ~99.9 %
- Фенотипические признаки и все заболевания — ~0.1 % генома, т.е. его 1/1000 мутированная часть

Факторы заболеваний



Мутации

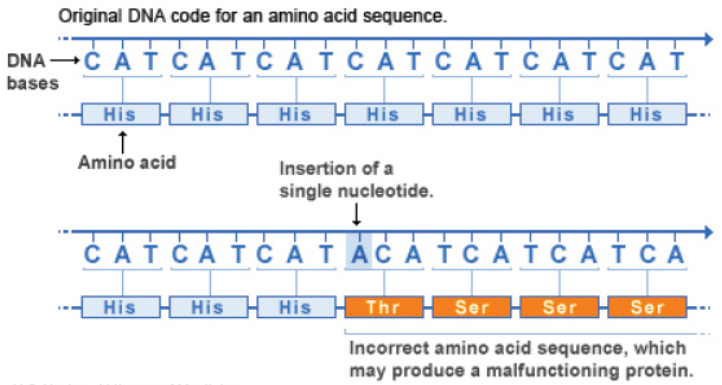
MISSENSE MUTATION — Замены



U.S. National Library of Medicine

This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.

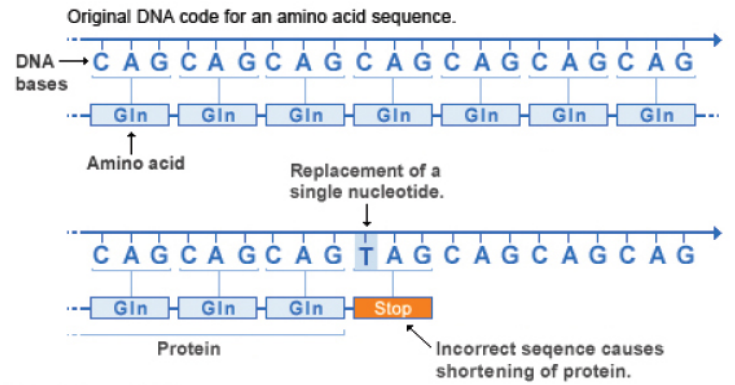
INSERTION — Вставки



U.S. National Library of Medicine

An insertion changes the number of DNA bases in a gene by adding a piece of DNA. As a result, the protein made by the gene may not function properly.

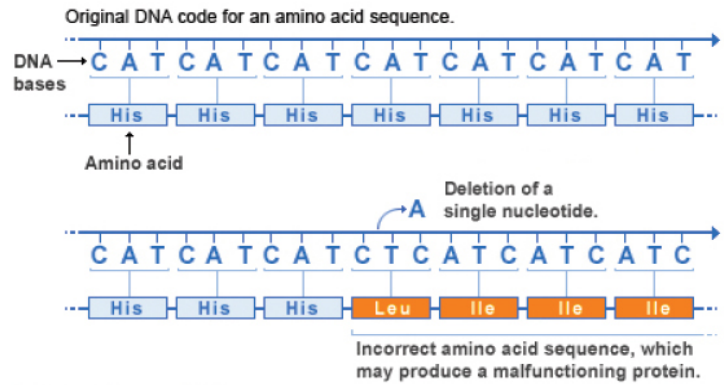
NONSENSE MUTATION — Нонсенсы



U.S. National Library of Medicine

A nonsense mutation is also a change in one DNA base pair. Instead of sub stituting one amino acid for another, however, the altered DNA sequence prematurely signals the cell to stop building a protein. This type of mutation results in a shortened protein that may function improperly or not at all.

DELETION — Делеции

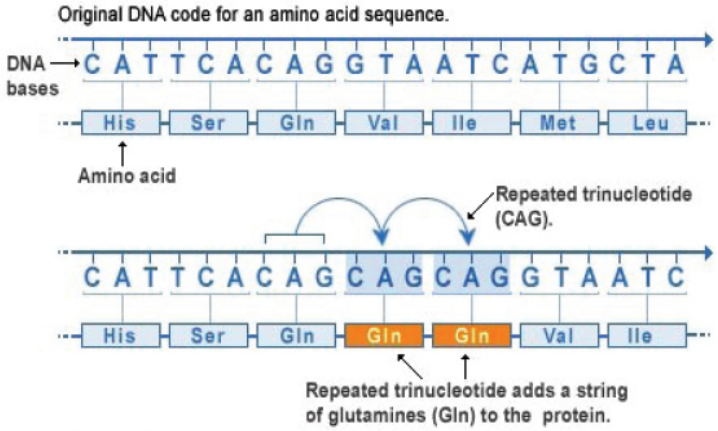


U.S. National Library of Medicine

A deletion changes the number of DNA bases by removing a piece of DNA. Small deletions may remove one or a few base pairs within a gene, while larger deletions can remove an entire gene or several neighboring genes. The deleted DNA may alter the function of the resulting protein(s).

Мутации

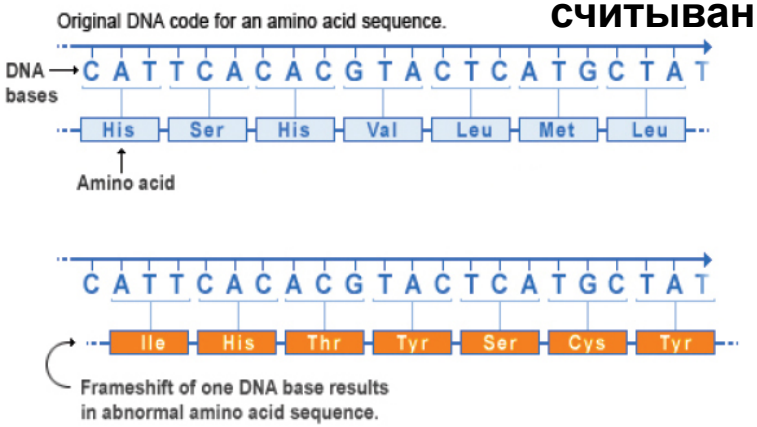
REPEAT EXPANSION — Повторы



U.S. National Library of Medicine

Nucleotide repeats are short DNA sequences that are repeated a number of times in a row. For example, a trinucleotide repeat is made up of 3-base-pair sequences, and a tetranucleotide repeat is made up of 4-base-pair sequences. A repeat expansion is a mutation that increases the number of times that the short DNA sequence is repeated. This type of mutation can cause the resulting protein to function improperly.

FRAMESHIFT MUTATION — Изменение рамки считывания

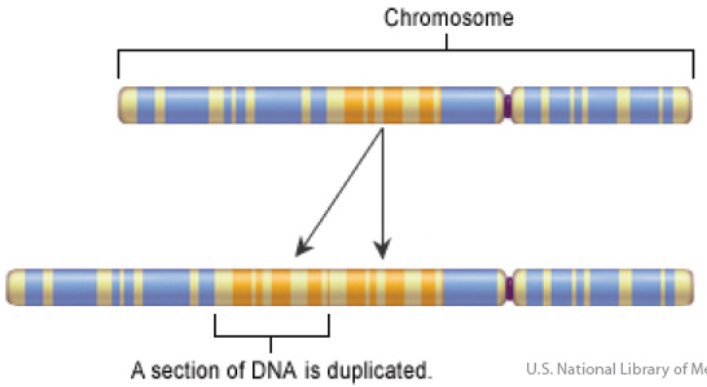


U.S. National Library of Medicine

This type of mutation occurs when the addition or loss of DNA bases changes a gene's reading frame. A reading frame consists of groups of 3 bases that each code for one amino acid. A frameshift mutation shifts the grouping of these bases and changes the code for amino acids. The resulting protein is usually nonfunctional. Insertions, deletions, and duplications can all be frameshift mutations.

DUPLICATION — Дублирование

A duplication consists of a piece of DNA that is abnormally copied one or more times. This type of mutation may alter the function of the resulting protein.



U.S. National Library of Medicine

Source: http://ghr.nlm.nih.gov/handbook/mutationsanddisorders/possiblemutations

GWAS

Genome-Wide Association Studies

Полногеномный анализ ассоциаций

GDAs

Gene-Disease Associations

Ассоциации ген – болезнь

VDAs

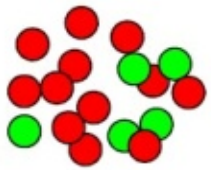
Variant-Disease Associations

Ассоциации варианты (SNP) – болезнь

GWAS

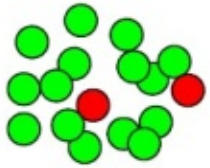
Выявление «генов предрасположенности» к заболеванию проводится путем сопоставлений частот генотипов у больных и здоровых

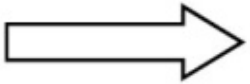
Группа больных



$P_{больные}$ >> $P_{контроль}$

Контроль (здоровые)



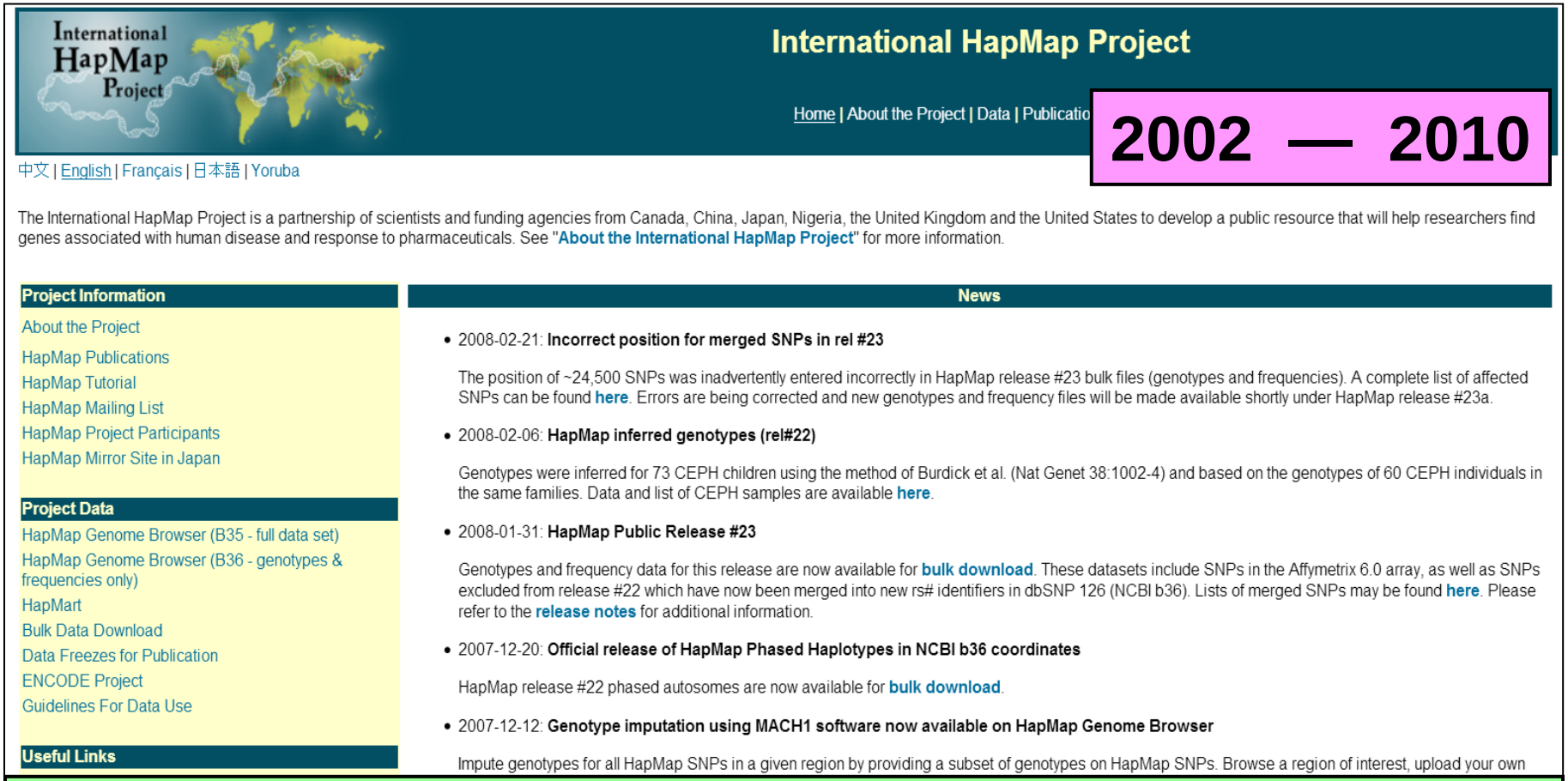


● - генотип, указывающий на предрасположенность к заболеванию.

OR – количественная мера предрасположенности (**O**dd **R**atio), показывает во сколько раз повышена вероятность заболеть для носителя «плохого» генотипа

$$\mathbf{OR} = \frac{P_{больные}(1 - P_{контроль})}{P_{контроль}(1 - P_{больные})}$$

Международный проект HapMap



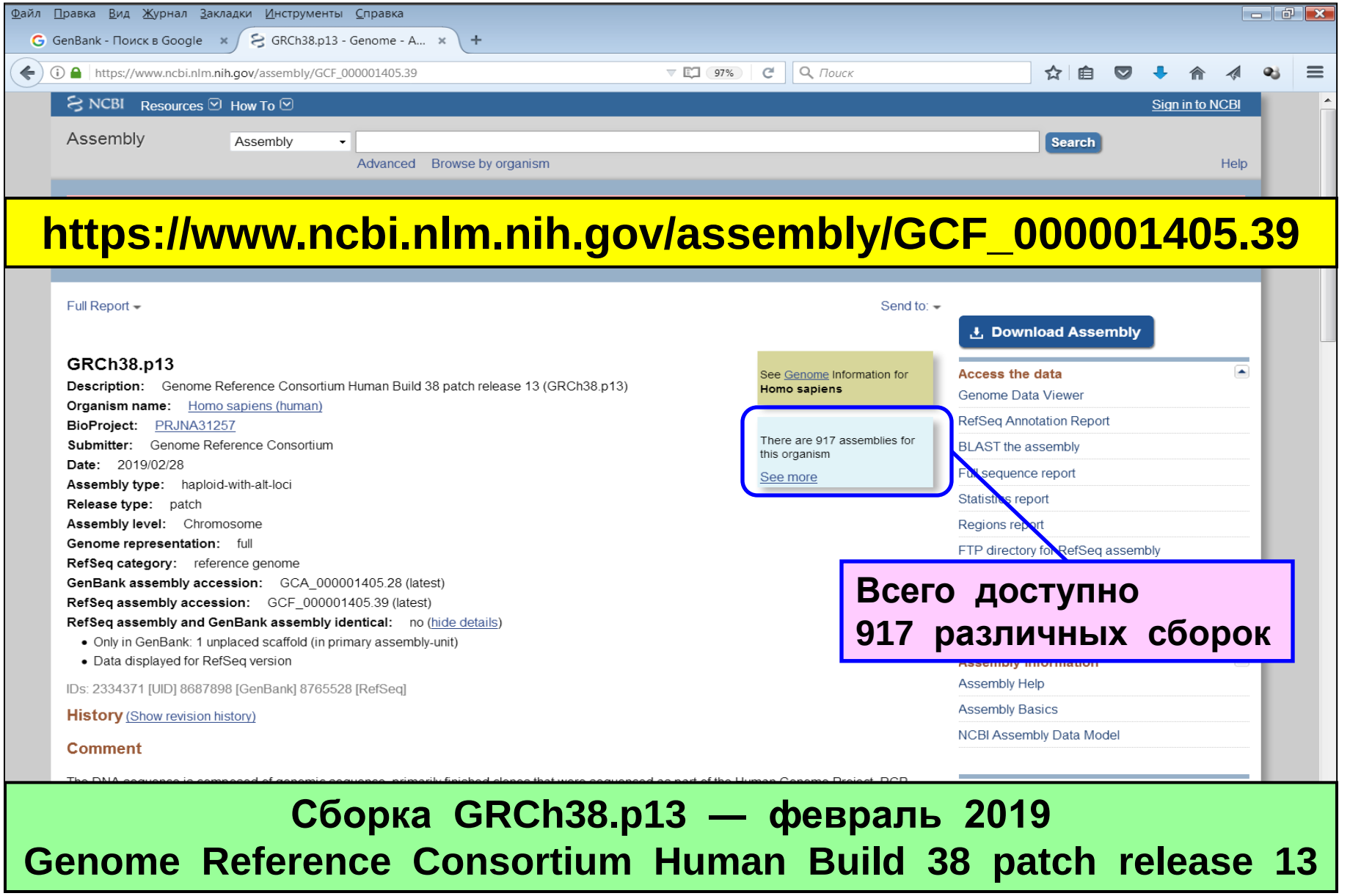
Карта гаплотипа генома человека

Описание общих закономерностей наследственной генетической изменчивости людей

Генетические варианты, влияющие на здоровье, болезни и реакции на лекарственные препараты и факторы окружающей среды

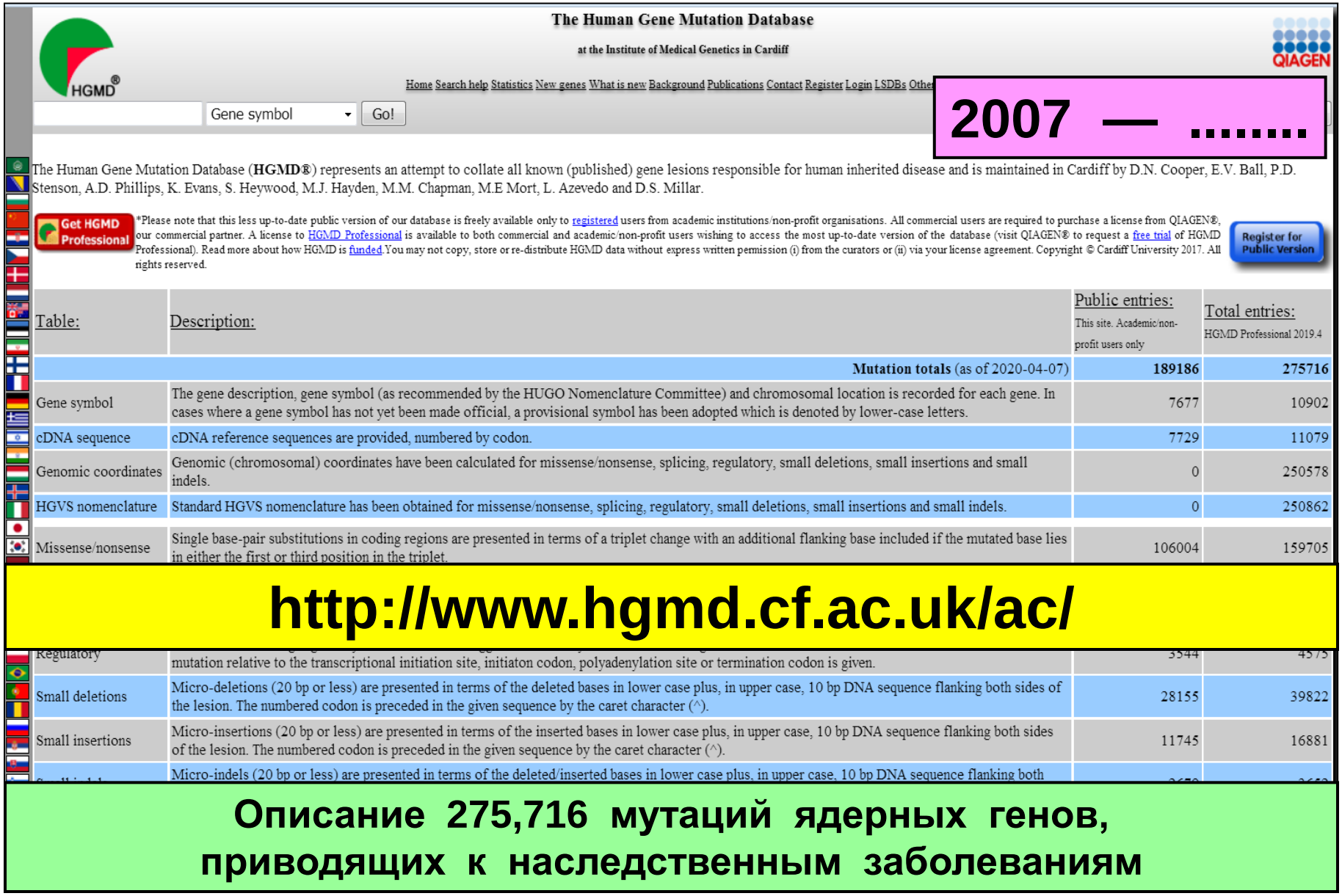
1000 Genomes

Референсный геном человека



Сборка GRCh38.p13 — февраль 2019
Genome Reference Consortium Human Build 38 patch release 13

HGMD — Human Gene Mutation Database



The Human Gene Mutation Database

at the Institute of Medical Genetics in Cardiff

Home Search help Statistics New genes What is new Background Publications Contact Register Login LSDBs Other

Gene symbol Go!

2007 — .....

The Human Gene Mutation Database (**HGMD®**) represents an attempt to collate all known (published) gene lesions responsible for human inherited disease and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E Mort, L. Azevedo and D.S. Millar.

Get HGMD Professional

*Please note that this less up-to-date public version of our database is freely available only to registered users from academic institutions/non-profit organisations. All commercial users are required to purchase a license from QIAGEN®, our commercial partner. A license to HGMD Professional is available to both commercial and academic/non-profit users wishing to access the most up-to-date version of the database (visit QIAGEN® to request a free trial of HGMD Professional). Read more about how HGMD is funded.You may not copy, store or re-distribute HGMD data without express written permission (i) from the curators or (ii) via your license agreement. Copyright © Cardiff University 2017. All rights reserved.

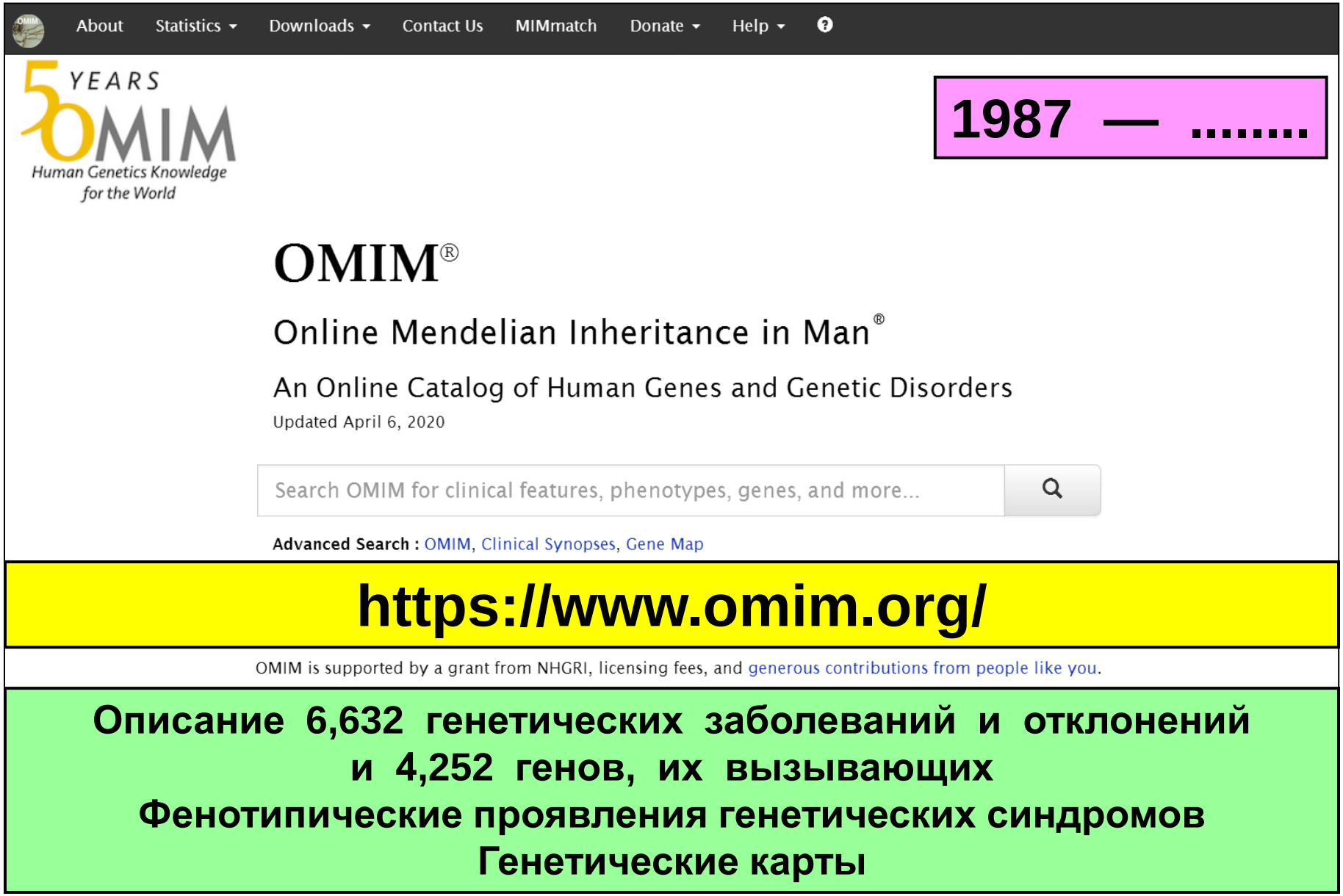
Register for Public Version

Table:	Description:	Public entries: This site. Academic/non-profit users only	Total entries: HGMD Professional 2019.4
	Mutation totals (as of 2020-04-07)	**189186**	**275716**
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	7677	10902
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	7729	11079
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	250578
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	250862
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	106004	159705
Regulatory	mutation relative to the transcriptional initiation site, initiaton codon, polyadenylation site or termination codon is given.	3544	4575
Small deletions	Micro-deletions (20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	28155	39822
Small insertions	Micro-insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	11745	16881
	Micro-indels (20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both		

http://www.hgmd.cf.ac.uk/ac/

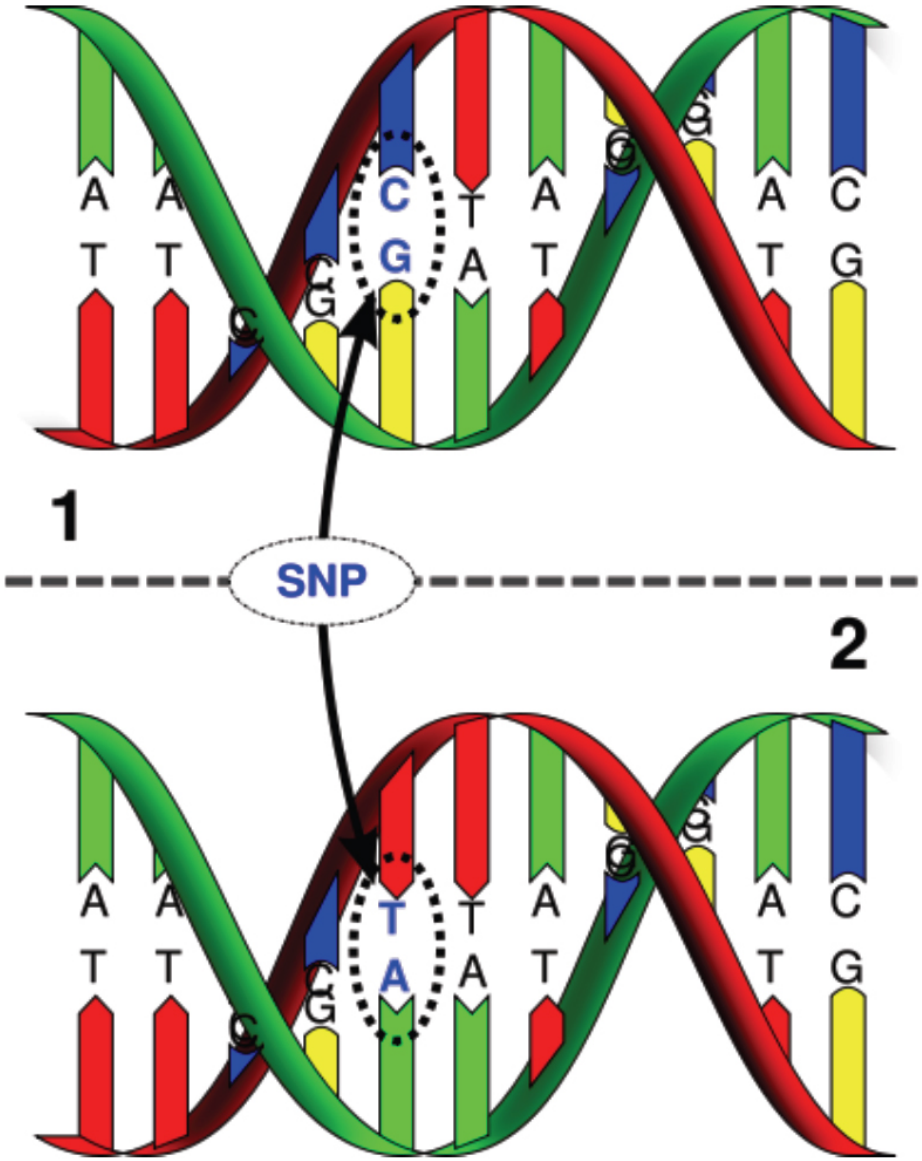
Описание 275,716 мутаций ядерных генов, приводящих к наследственным заболеваниям

OMIM — Online Mendelian Inheritance in Man



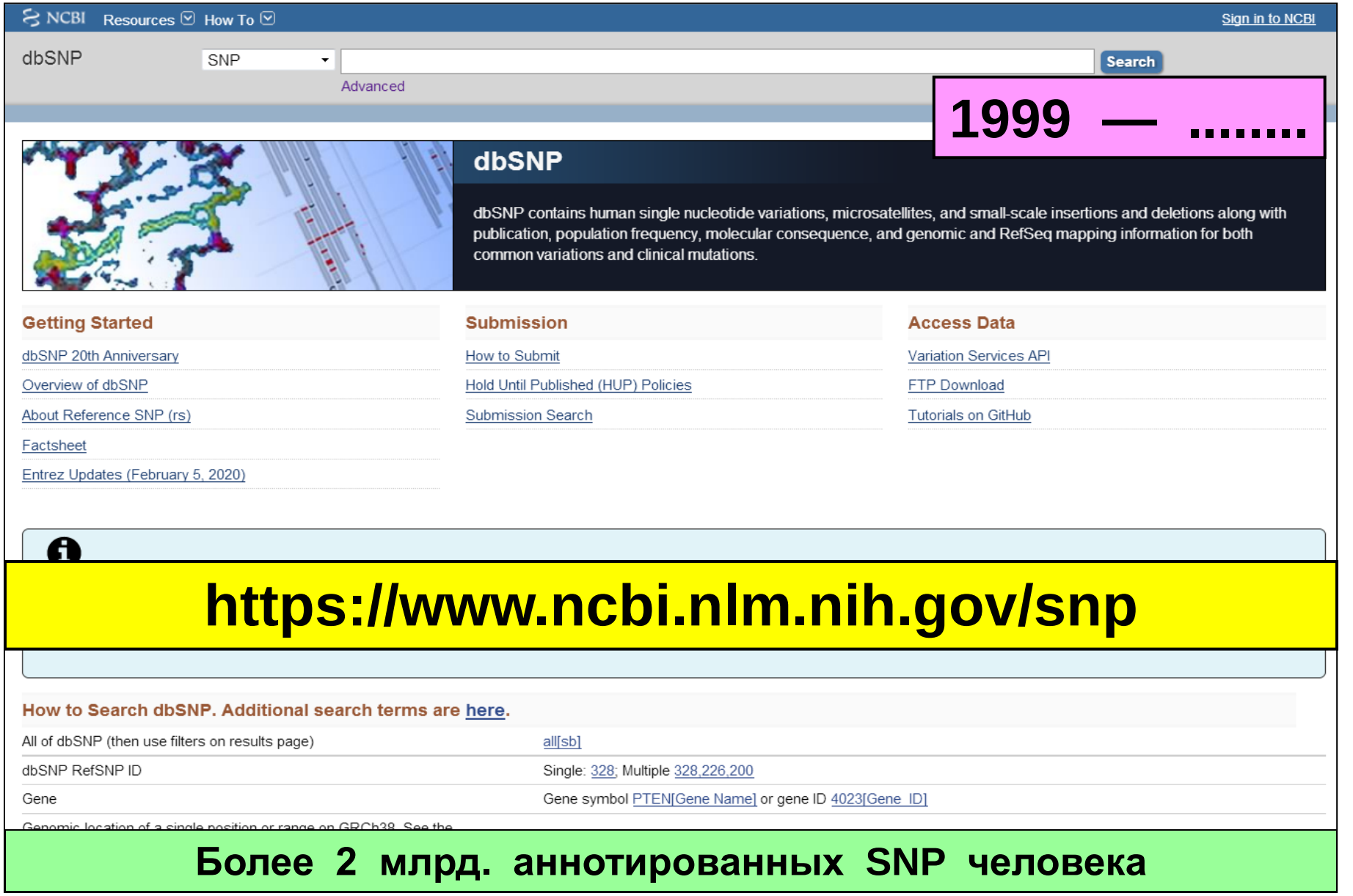
Описание 6,632 генетических заболеваний и отклонений и 4,252 генов, их вызывающих
Фенотипические проявления генетических синдромов
Генетические карты

Однонуклеотидный полиморфизм



~10 млрд. SNP

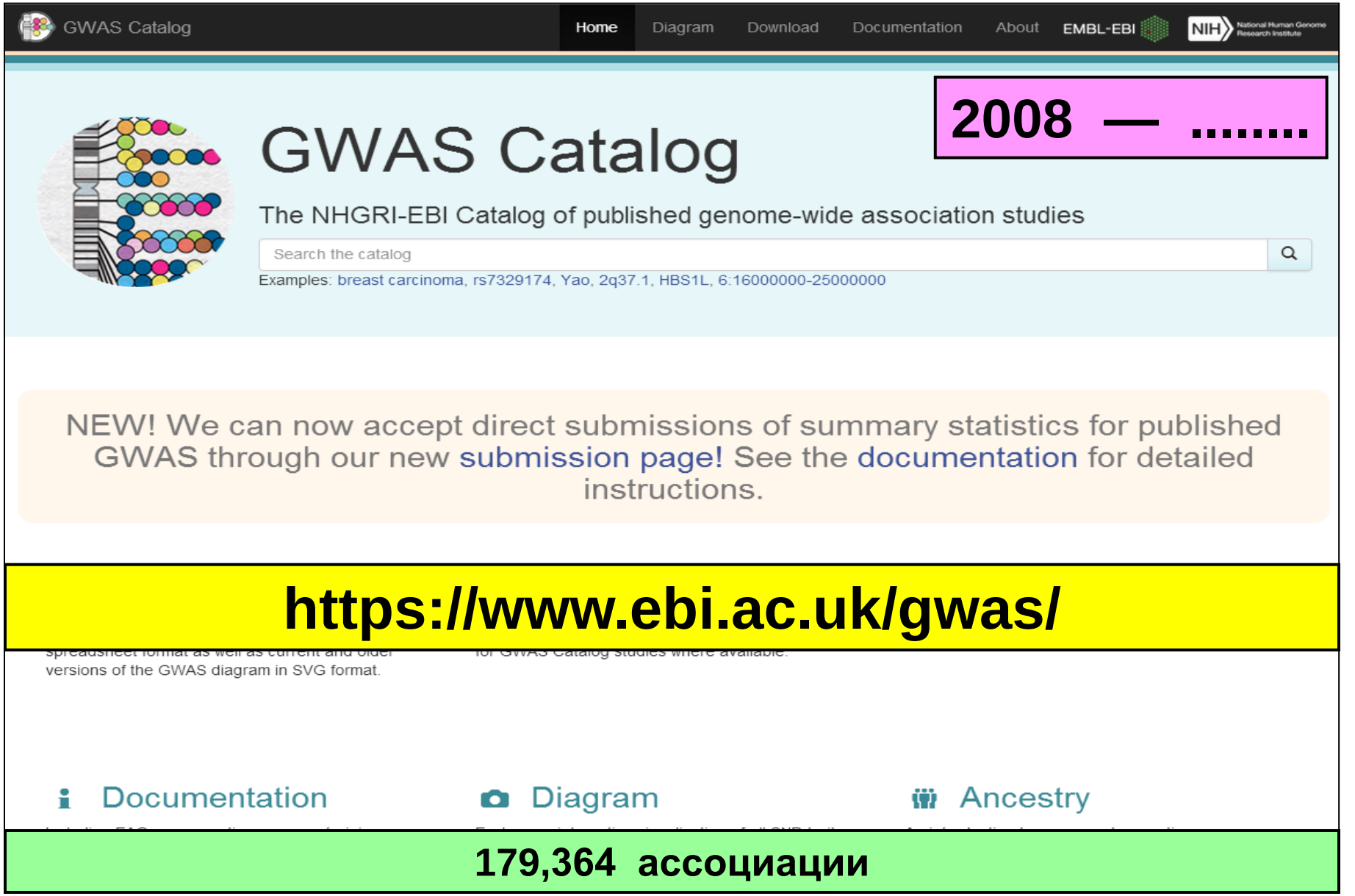
Базы данных GWAS — dbSNP



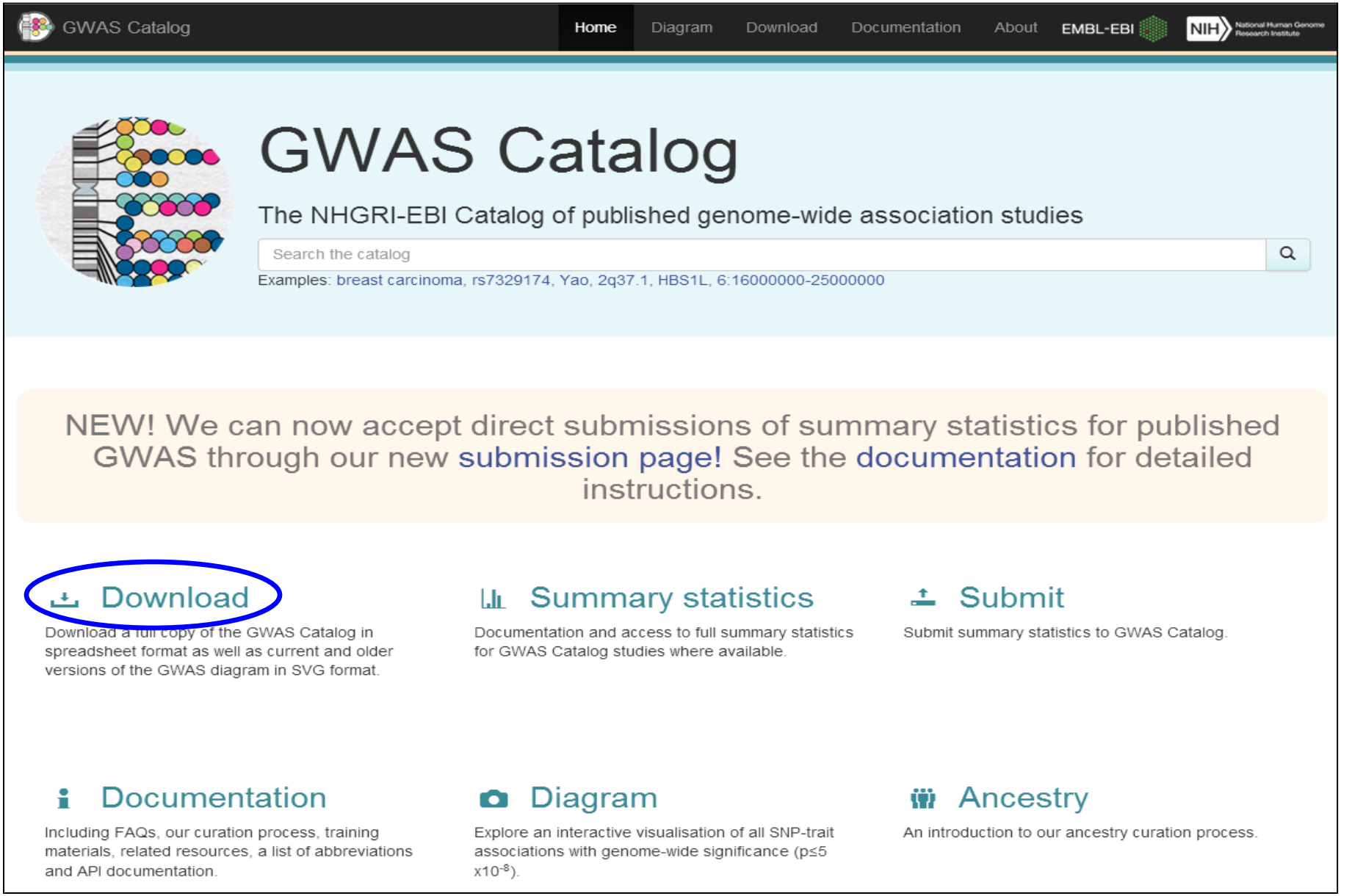
Базы данных GWAS — DisGeNET

- 628,685 GDAs между 17,549 генами и 24,166 заболеваниями и отклонениями
- 210,498 VDAs между 117,337 SNPs и 10,358 заболеваниями и фенотипами

GWAS Catalog



GWAS Catalog



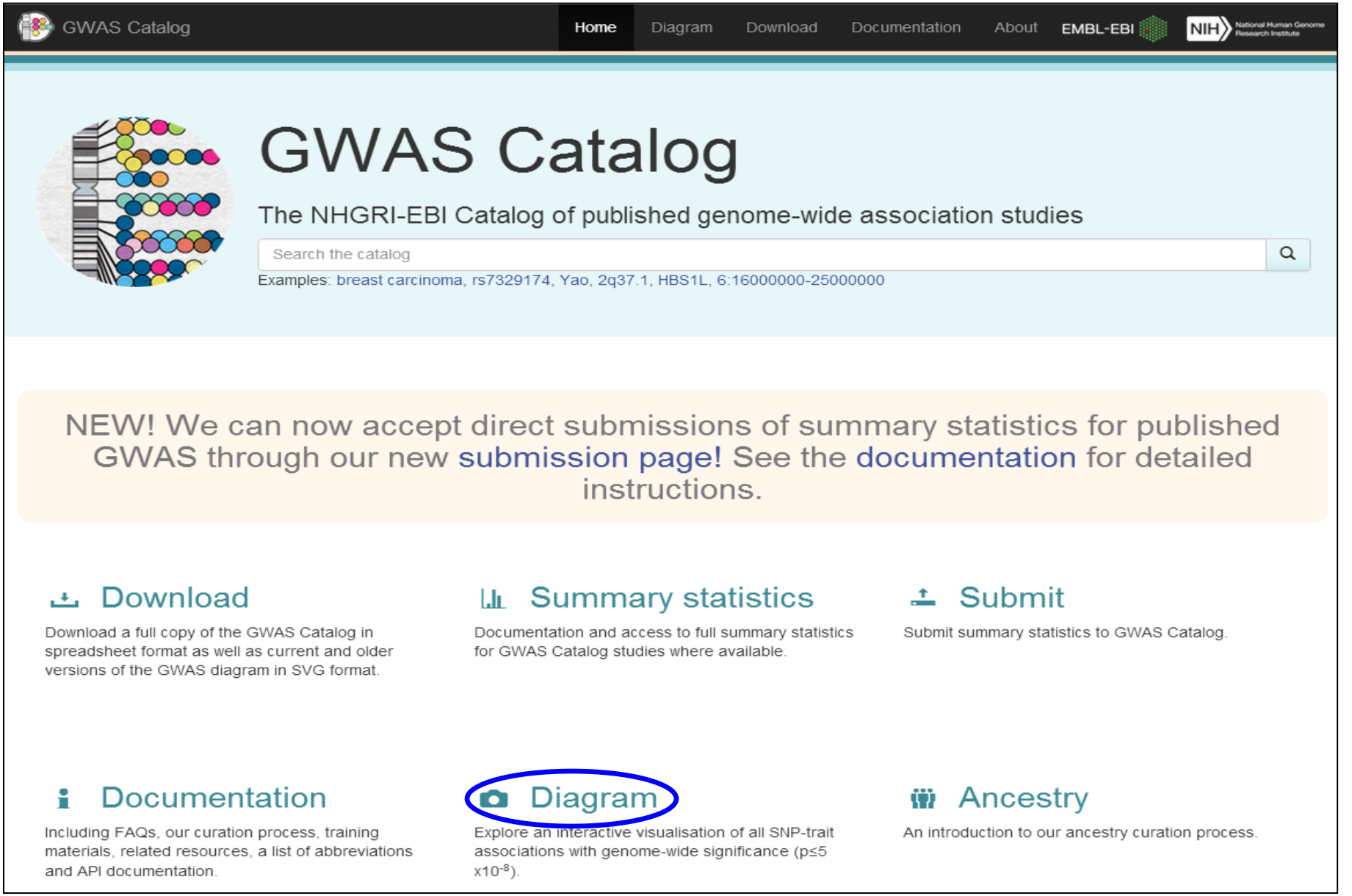
GWAS Catalog

	A	H	K	M	O	U	Y	AA	AB	AE
1	DATE ADDED	DISEASE/TRA	REGION	CHR_POS	MAPPED_GENE	STRONGEST SNP-RISK ALLELE	CONTEXT	RISK ALLELE FREQUENCY	P-VALUE	OR or BETA
494	2008-06-16	Type 2 diabetes	16q12.2	53782363	FTO	rs8050136-A	intron_variant	0.4	7.00E-14	1.23
495	2008-06-16	Type 2 diabetes	3q27.2	185793899	IGF2BP2	rs4402960-T	intron_variant	0.32	9.00E-16	1.14
496	2008-06-16	Type 2 diabetes	8q24.11	117172544	SLC30A8	rs13266634-C	missense_variant	0.69	5.00E-08	1.12
497	2008-06-16	Type 2 diabetes	6p22.3	20660803	CDKAL1	rs10946398-C	intron_variant	0.32	7.00E-08	1.16
498	2008-06-16	Type 2 diabetes	10q23.3	92702302	HHEX - IDE	rs5015480-C	intergenic_variant	0.57	5.00E-06	1.13
499	2008-06-16	Type 2 diabetes	10q25.2	112994339	TCF7L2	rs7901695-C	intron_variant	NR	1.00E-48	1.37
500	2008-06-16	Type 2 diabetes	9p21.3	22029548	AL359922.1, CDKN2	rs564398-T	3_prime_UTR_variant	0.56	1.00E-06	1.13
501	2008-06-16	Type 2 diabetes	11p15.1	17387083	KCNJ11	rs5215-C	missense_variant	NR	5.00E-11	1.14
502	2008-06-16	Type 2 diabetes	3p25.2	12351626	PPARG	rs1801282-C	missense_variant	NR	2.00E-06	1.14
503	2008-06-16	Type 2 diabetes	9p21.3	22134095	CDKN2B-AS1 - ALI	rs1081[illegible]	intergenic_variant	0.83	5.00E-06	1.19
525										
529										
530										
531										
532										
533										
534										
535										
661										
662										
663										
664										
665										
666										
667										
668										
669										
670	2011-05-24	Type 2 diabetes	1q32.2	207478831	CR2	rs17045328-G	intron_variant	0.3	7.00E-06	1.38
671	2011-05-24	Type 2 diabetes	6q13	70579486	SDHAF4	rs1048886-G	missense_variant	0.18	3.00E-08	1.54
672	2011-05-24	Type 2 diabetes	6q24.1	139952510	AL050338.2	rs642858-A	intron_variant	0.4	2.00E-06	1.35
673	2011-05-24	Type 2 diabetes	7q22.1	100892456	ACHE	rs7636-A				
674	2011-05-24	Type 2 diabetes	10q26.3	131149699	TCERG1L	rs10741243-G				
675	2011-05-24	Type 2 diabetes	3q12.3	102484201	ZPLD1 - AC063938.	rs2063640-A				
676	2011-05-24	Type 2 diabetes	19q13.2	39090097	ACP7	rs472265-G				
677	2011-05-24	Type 2 diabetes	3q23	142712158	AC072028.1, PLS1	rs3773506-C				
779	2009-04-10	Type 2 diabetes	11p15.4	2837316	KCNQ1	rs2237897-C				
780	2009-04-10	Type 2 diabetes	6p22.3	20657634	CDKAL1	rs4712524-G	intron_variant	0.42	3.00E-10	1.22
781	2009-04-10	Type 2 diabetes	3q27.2	185812502	IGF2BP2	rs6769511-C	intron_variant	0.32	1.00E-09	1.23
1416	2008-09-09	Type 2 diabetes	2q12.1	105221141	LINC01918 - GPR45	rs6712932-?	intergenic_variant	NR	6.00E-06	1.52

1. Хромосома 8, длинное плечо, область 2, бэнд 4, суб-бэнд 11
2. Позиция — 117172544
3. Ген SLC30A8 — цинковый транспортер 8, ZNT8
4. Рисковая аллель C — цитозин
5. Тип мутации — миссенс
6. Общая частота встречаемости рисковой аллели 0.68 %
7. Частота встречаемости рисковой аллели у больных — $5 \cdot 10^{-8}$
8. Мера предрасположенности (Odd Ratio) — 1.12 раза

Сахарный диабет типа 2
1,705 SNP

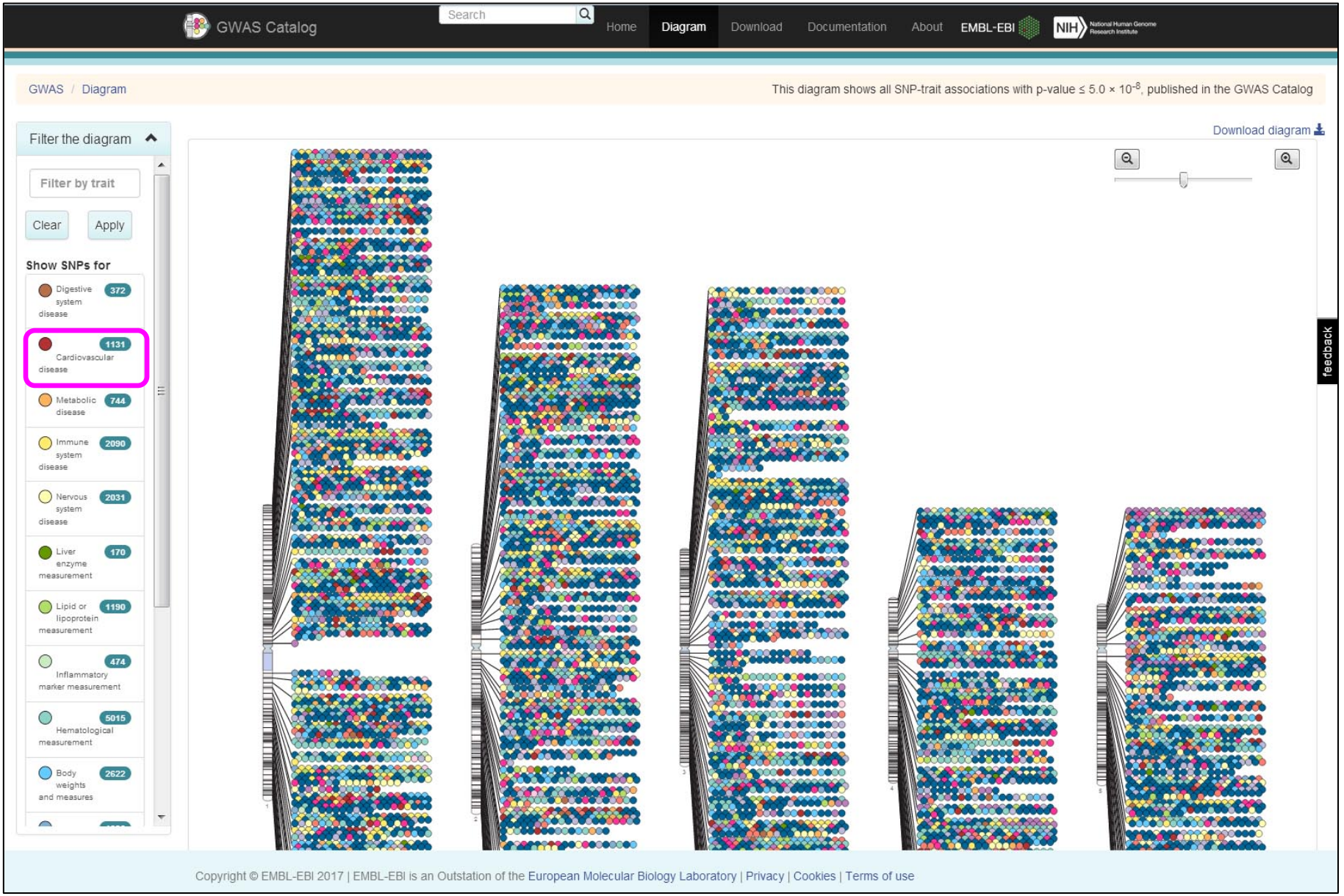
GWAS Catalog



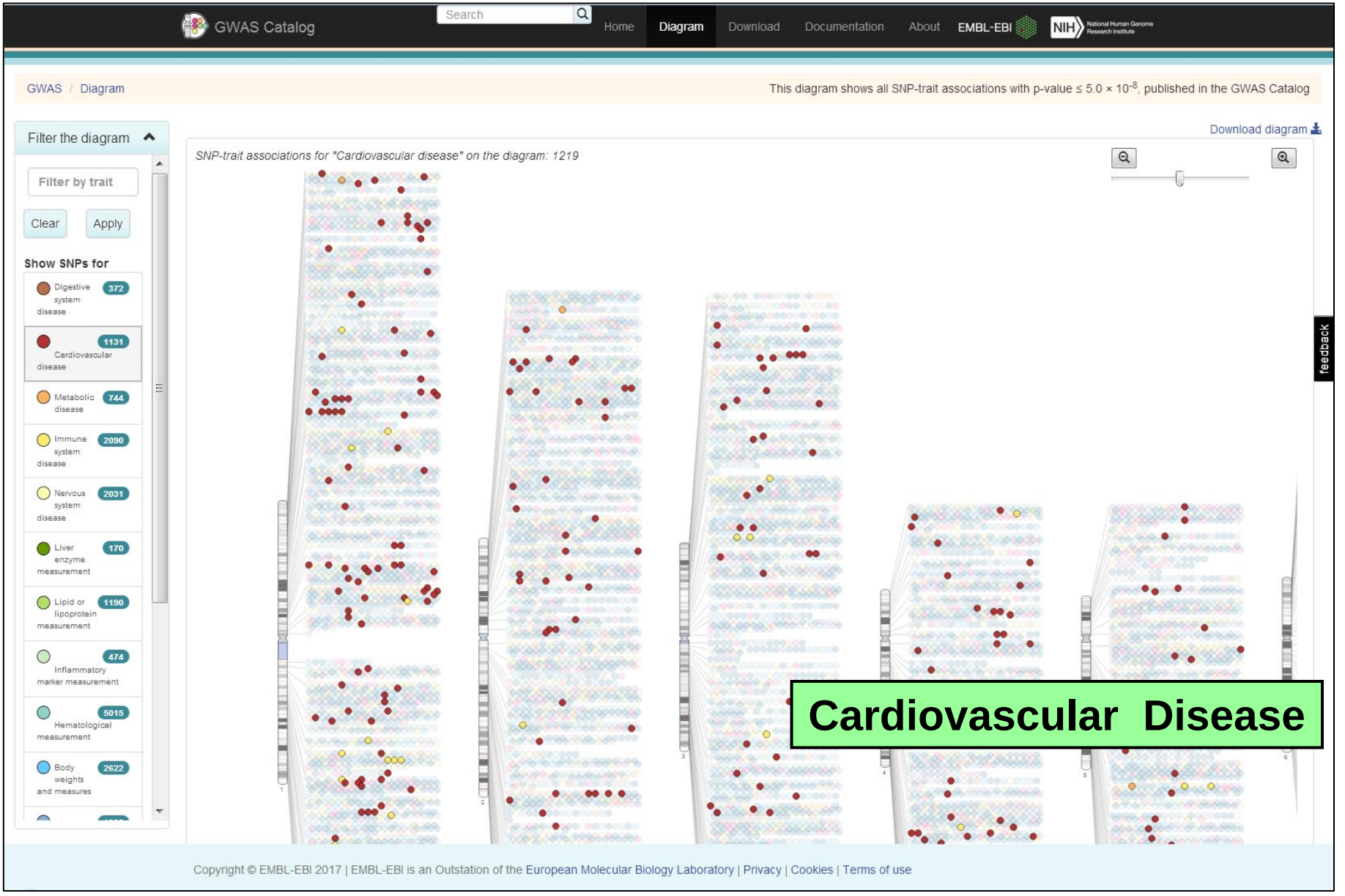
GWAS Diagram

https://www.ebi.ac.uk/gwas/diagram

GWAS Diagram

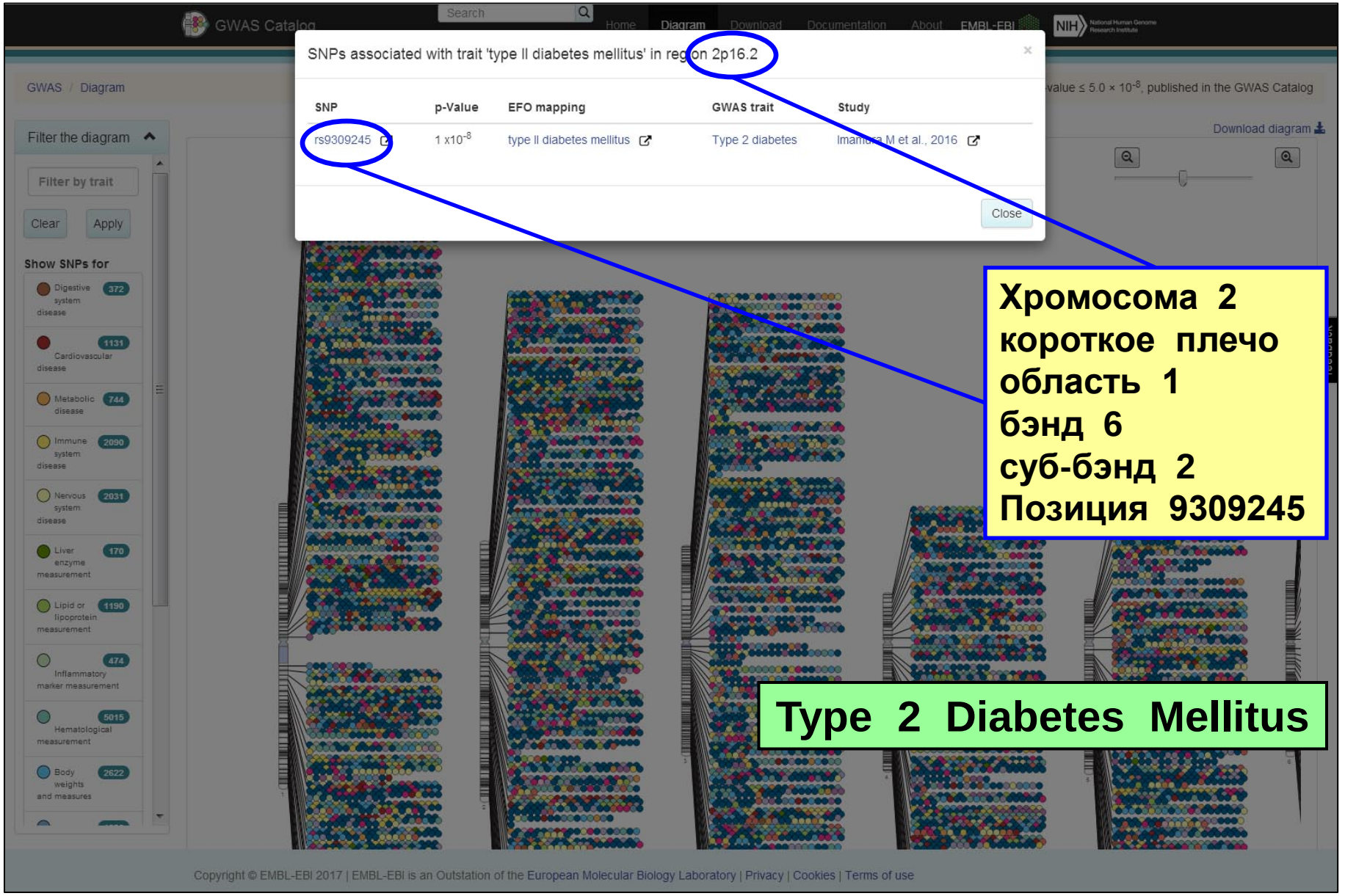


GWAS Diagram



GWAS Diagram

GWAS Diagram



GWAS

Table 1 GWA studies showing SNPs association with cancer.

Disease	Chromosomal location	Gene	Strongest SNPs-risk allele	*P*-value	Reference
Breast cancer	10q26.13	*FGFR2*	rs2981582-G	2×10^{-76}	[11,36-38]
	16q12.1	*TNCR9/LOC643714*	rs3803662-C	1×10^{-36}	
	5q11.2	*MAP3K1*	rs889312-A	7×10^{-20}	
	11p15.5	*LSP1*	rs3817198-T	3×10^{-9}	
	8q24.21	Intergenic	rs13281615-T	5×10^{-12}	
	2q34	*ERBB4*	rs13393577-C/T	8.8×10^{-14}	
	10q26.13	*FGFR2*	rs1219648-G	1×10^{-10}	
	2q35	Intergenic	rs13387042-A	1×10^{-13}	
Basal cell carcinoma	1q42.13	*RHOU*	rs801114-G	6×10^{-12}	[39-44]
	1p36.13	*PADI4, PADI6, RCC2, ARHGEF10L*	rs7538876-A	4×10^{-12}	
	12q12.13	*KRT5*	rs11170164-A	2×10^{-9}	
	9p21	*CDKN2A/B*	rs2151280-C	7×10^{-9}	
	8q24.21	Intergenic	rs16901979-A	1×10^{-12}	
	8q24.21	Intergenic	rs6983267-G	9×10^{-13}	
	17q12	*TCF2*	rs4430796-A	1×10^{-11}	
	11p15	*IGF2, IGF2A, INS, TH*	rs7127900-A	3×10^{-33}	
Colorectal cancer	8q24.21	*ORF, DQ515897*	rs10505477-A	3×10^{-11}	[45,46]
	18q21.1	*SMAD7*	rs4939827-T	1×10^{-12}	
Lung cancer	15q25.1	*CHRNA3, CHRNA5, CHRNB4, PSMA4, LOC123688*	rs8034191-C	5×10^{-20}	[47]
	6p22.1	*TRNAA-UGC*	rs4324798-A	2×10^{-8}	[48]
Melanoma	20q11.22	*CDC91L1*	rs910873-T	1×10^{-15}	[49,50]
	22q13.1	Intergenic	rs2284063-G	2×10^{-9}	
Neuroblastoma	6p22.3	*FLJ22536, FLJ44180*	rs6939340-G	9×10^{-15}	[51,52]
	2q35	*BARD1*	rs6435862-G	9×10^{-18}	
Ovarian cancer	9p22.2	Intergenic	rs3814113-T	5×10^{-19}	[53]
Thyroid cancer	9q22.23	*F*	513-A	2×10^{-27}	[15]
	14q13.3	*N*	289-T	2×10^{-9}	

Опухолевые заболевания

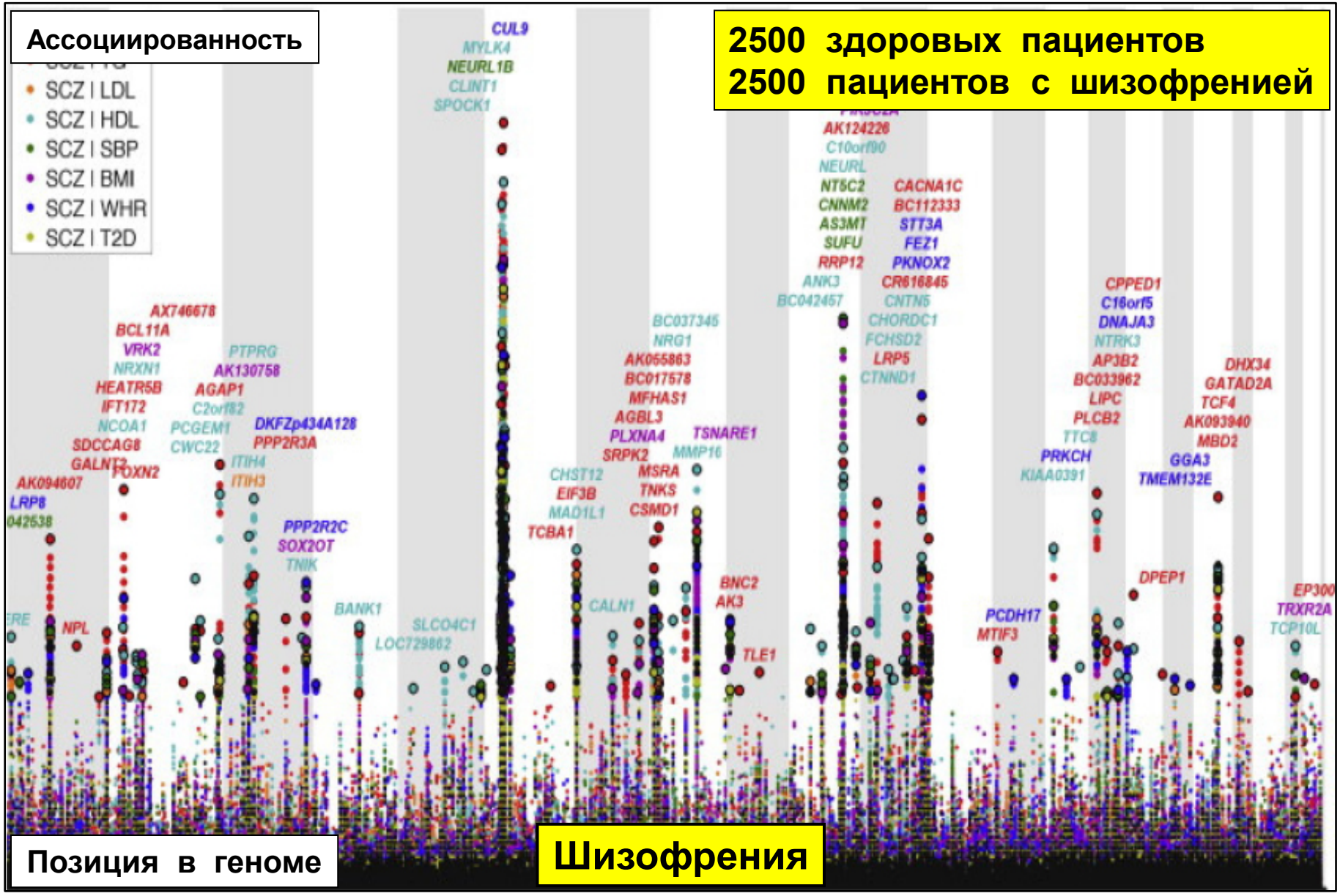
GWAS

Table 2 SNPs significantly associated with infectious disease phenotypes in genome-wide studies.

Disease	SNP location	Strongest SNPs-risk allele	*P*-value	Reference
Creutzfeldt–Jakob disease	PRNP	rs1799990	2.0×10^{-27}	[54]
Dengue shock Syndrome	MICB	rs3132468	4.4×10^{-11}	[55]
Hepatitis B	HLA-DPA1	rs3077	2.3×10^{-38}	[56]
	HLA-DPB1	rs9277535	6.3×10^{-39}	
Hepatitis C	IL28B	rs8099917	6.1×10^{-9}	[57]
HIV-1 and AIDS	HLA-C	rs9264942	5.9×10^{-32}	[58–64]
	HLA-B, HCP5	rs2395029	4.5×10^{-35}	
	HLA-B	rs2523608	5.6×10^{-10}	
	HLA-C	rs9264942	2.8×10^{-35}	
	MICA	rs4418214	1.4×10^{-34}	
	HLA-B, HCP5	rs2395029	9.7×10^{-26}	
	PSORS1C3	rs3131018	4.2×10^{-16}	
	HLA-B	rs2523608	8.9×10^{-20}	
	Intergenic	rs2255221	3.5×10^{-14}	
	HLA-B	rs2523590	1.7×10^{-13}	
	Intergenic	rs9262632	1.0×10^{-8}	
	ZNRD1, RNF39	rs9261174	1.8×10^{-8}	
	PARD3B	rs11884476	3.4×10^{-9}	
	HLA-B, HCP5	rs2395029	6.8×10^{-10}	
	CXCR6	rs2234358	9.7×10^{-10}	
Leprosy	LACC1	rs3764147	3.7×10^{-54}	[65]
	NOD2	rs9302752	3.8×10^{-40}	
	RIPK2	rs42490	1.4×10^{-16}	
	CCDC122	rs3088362	1.4×10^{-31}	
	TNFSF15	rs6478108	3.4×10^{-21}	
Meningococcal disease	CFH	rs1065489	2.2×10^{-11}	[66]
Severe malaria			3.7×10^{-11}	[67]
Tuberculosis			6.8×10^{-9}	[68]

Предрасположенность к инфекциям

GWAS


Ассоциированность
SCZ | LDL
SCZ | HDL
SCZ | SBP
SCZ | BMI
SCZ | WHR
SCZ | T2D
2500 здоровых пациентов
2500 пациентов с шизофренией
Позиция в геноме
Шизофрения
CUL9
MYLK4
NEURL1B
CLINT1
SPOCK1
AX746678
BCL11A
VRK2
NRXN1
HEATR5B
IFT172
NCOA1
SDCCAG8
GALNT2
AK094607
LRP8
042638
FOXN2
PTPRG
AK130758
AGAP1
C2orf82
PCGEM1
CWC22
DKFZp434A128
PPP2R3A
ITIH4
ITIH3
PPP2R2C
SOX2OT
TNIK
BANK1
SLCO4C1
LOC729862
CHST12
EIF3B
MAD1L1
TCBA1
CALN1
BC037345
NRG1
AK055863
BC017578
MFHAS1
AGBL3
PLXNA4
SRPK2
MSRA
TNKS
CSMD1
TSNARE1
MMP16
BNC2
AK3
TLE1
AK124226
C10orf90
NEURL
NT5C2
CNNM2
AS3MT
SUFU
RRP12
ANK3
BC042457
CACNA1C
BC112333
STT3A
FEZ1
PKNOX2
CR616845
CNTN5
CHORDC1
FCHSD2
LRP5
CTNND1
PCDH17
MTIF3
CPPED1
C16orf5
DNAJA3
NTRK3
AP3B2
BC033962
LIPC
PLCB2
TTC8
PRKCH
KIAA0391
DPEP1
DHX34
GATAD2A
TCF4
AK093940
MBD2
GGA3
TMEM132E
EP300
TRXR2A
TCP10L


To be continued ...