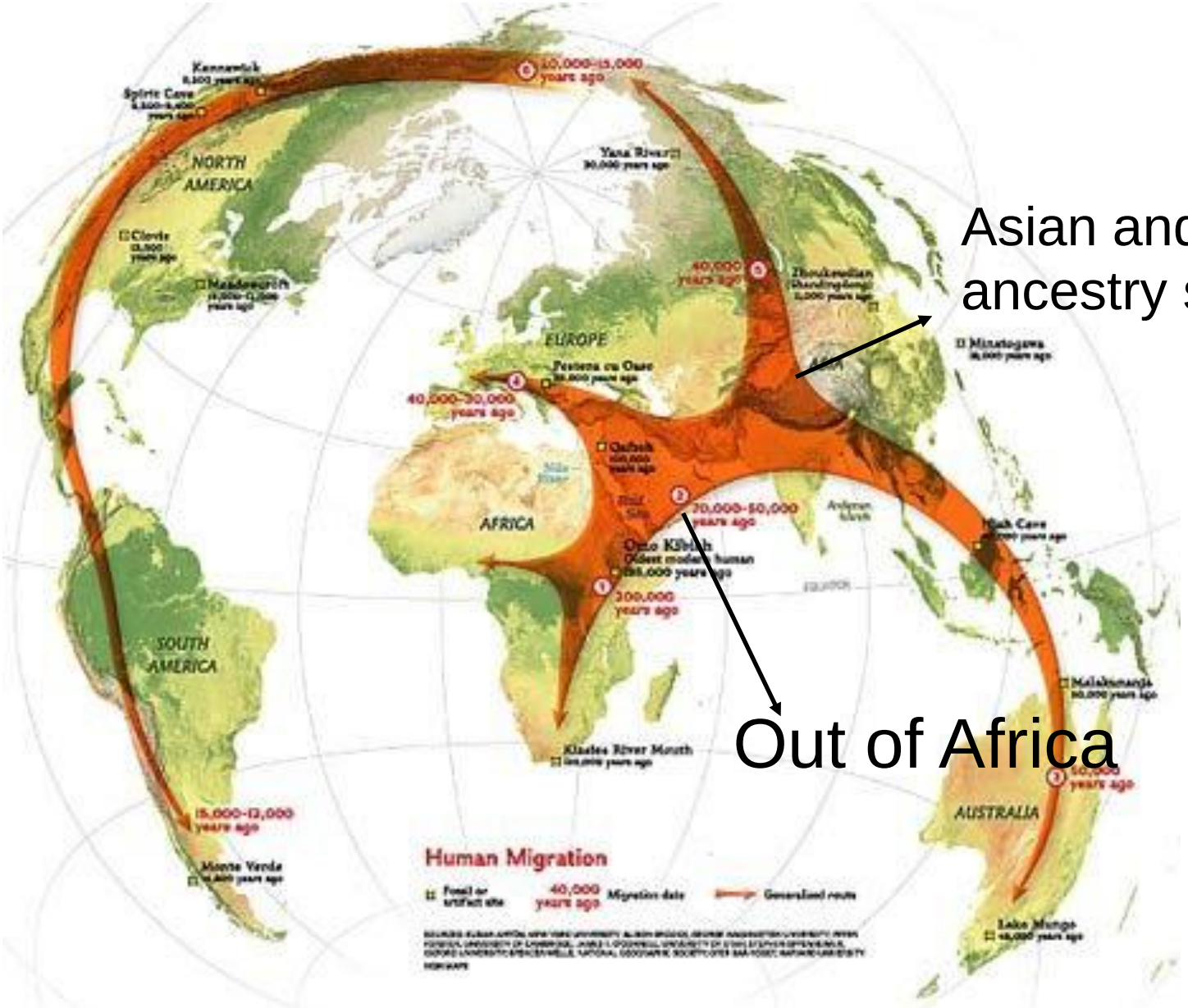
Origin and the spectrum of rare variants

——lesson learned from Autism and Psoriasis study

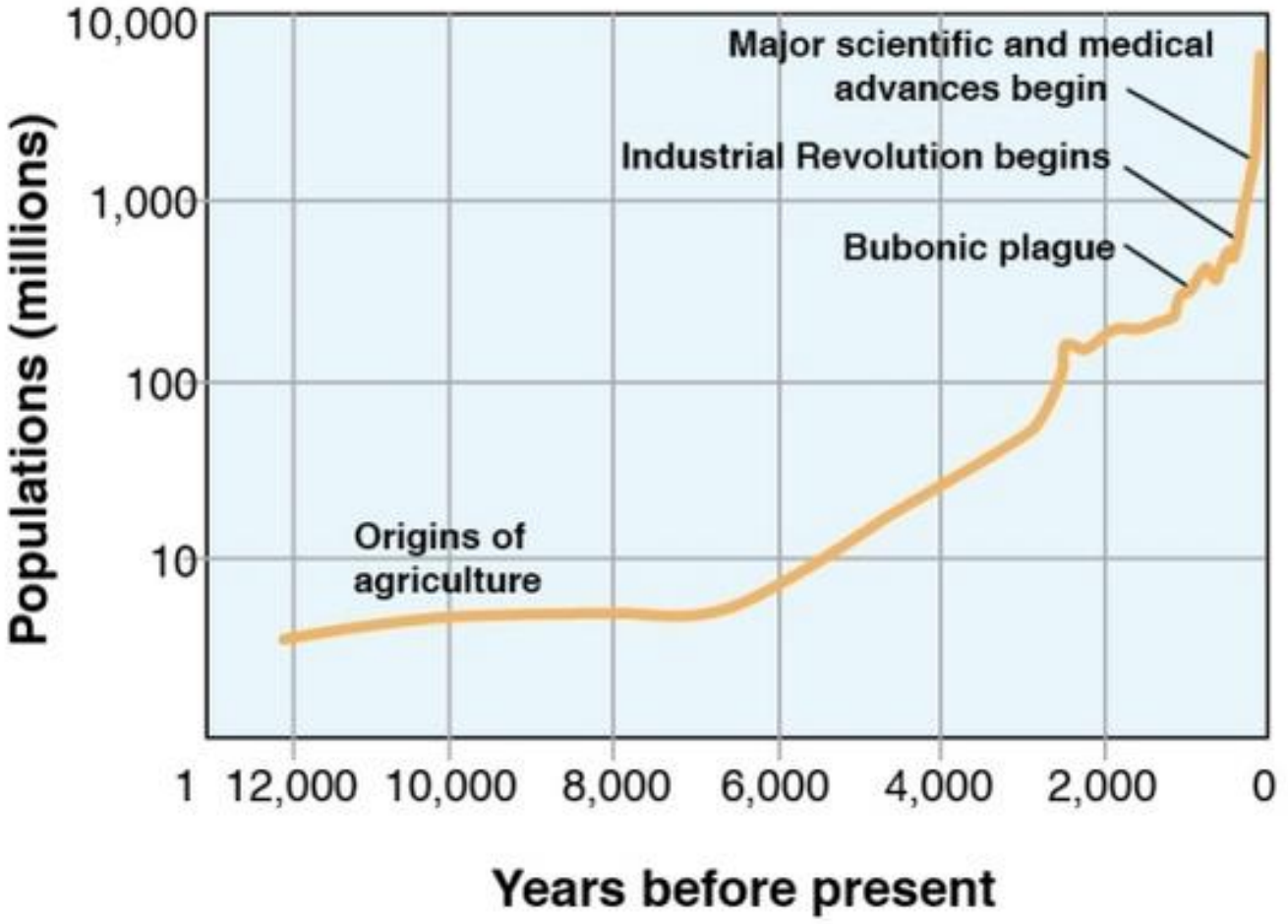
Xin Jin, BGI

IRDiRC, Shenzhen

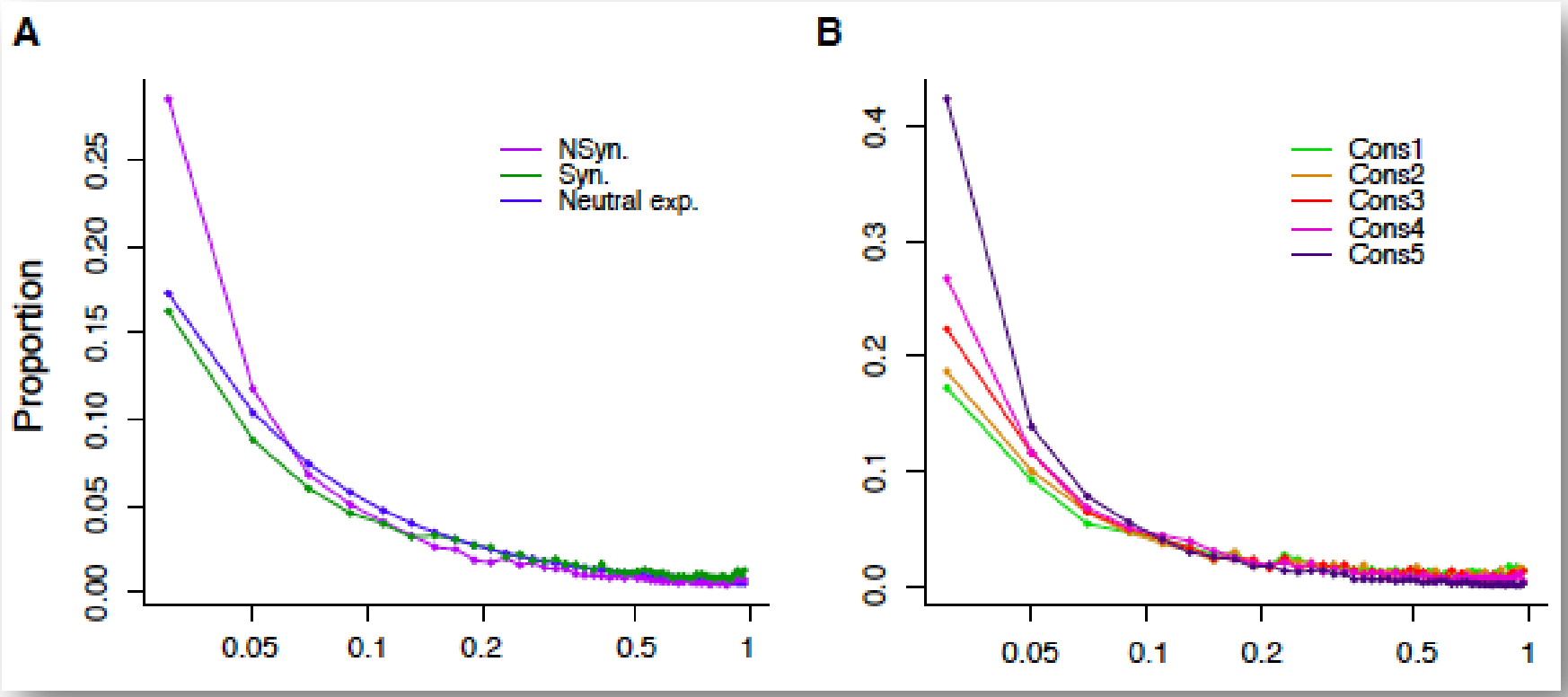
Origin of modern Human



Recent rapid growth

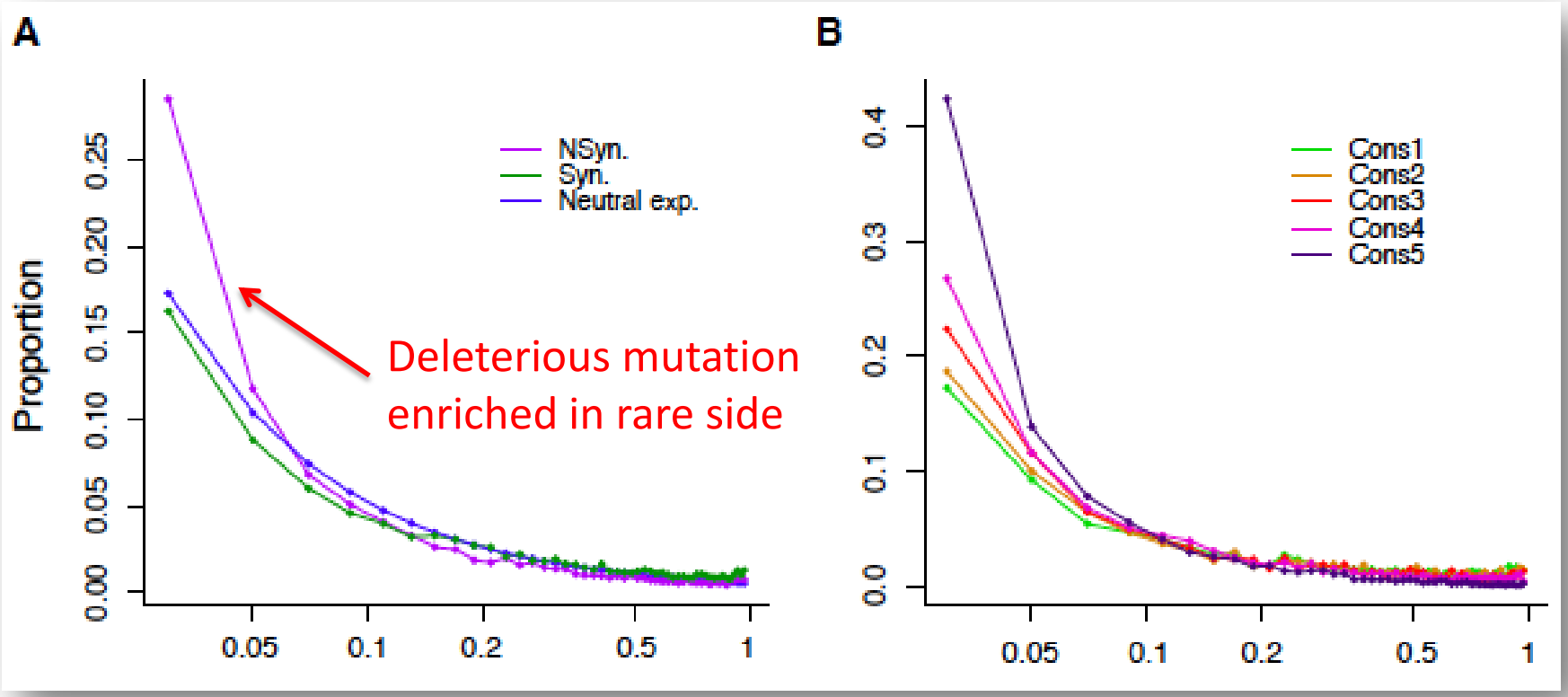


Power of nature selection



Y. Li, et.al. 2010. Nature Genetics

Power of nature selection



Y. Li, et.al. 2010. Nature Genetics

NGS accelerate MD research by identifying rare mutation

Brain Advance Access published November 23, 2010
doi:10.1093/brain/awq323 Brain 2010: Page 1 of 9 | 1
BRAIN
A JOURNAL OF NEUROLOGY

TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing

Jun Ling Wang,[1,2,*] Xu Yang,[3,*] Kun Xia,[2,*] Zheng Mao Hu,[2] Ling Weng,[1] Xin Jin,[3,4] Hong Jiang,[1,5] Peng Zhang,[3] Lu Shen,[1,5] Ji Feng Guo,[1,5] Nan li,[1] Ying Rui Li,[3] Li Fang Lei,[1] Jie Zhou,[1] Juan Du,[1] Ya Fang Zhou,[1] Qian Pan,[2] Jian Wang,[3] Jun Wang,[3,6] Rui Qiang Li[3,6] and Bei Sha Tang[1,2,5]

1 Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan province 410008, China
2 National Key Lab of Medical Genetics of China, Changsha, Hunan province 410008, China
3 BGI-Shenzhen, Shenzhen, Guangdong province 518083, China
4 Innovative Program for Undergraduate Students, School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China
5 Neurodegenerative Disorders Research Center, Central South University, Changsha, Hunan province 410008, China
6 Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark

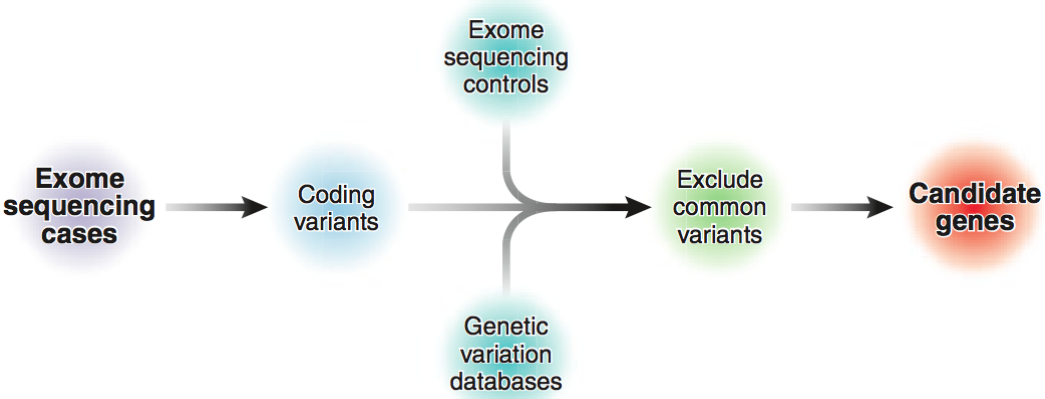
*These authors contributed equally to this work.

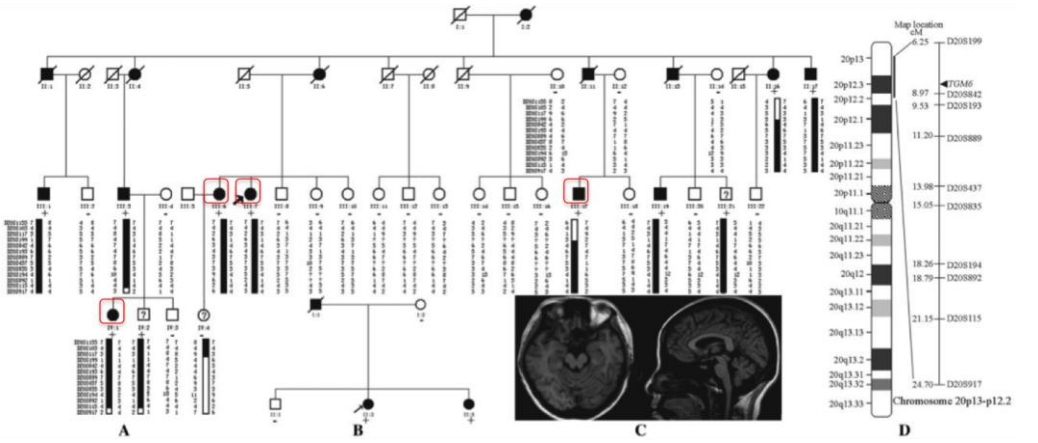
Correspondence to: Dr Bei Sha Tang,
Department of Neurology,
Xiangya Hospital,
Central South University, Changsha,
Hunan Province 410008, China
E-mail: bstang7398@yahoo.com.cn

Correspondence may also be addressed to: Rui Qiang Li,
BGI-Shenzhen, Beishan Industrial Zone,
Yantian Distince,
Shenzhen, 518083,
China.
E-mail: lirq@genomics.org.cn

Autosomal-dominant spinocerebellar ataxias constitute a large, heterogeneous group of progressive neurodegenerative diseases with multiple types. To date, classical genetic studies have revealed 31 distinct genetic forms of spinocerebellar ataxias and identified 19 causative genes. Traditional positional cloning strategies, however, have limitations for finding causative genes of rare Mendelian disorders. Here, we used a combined strategy of exome sequencing and linkage analysis to identify a novel spinocerebellar ataxia causative gene, *TGM6*. We sequenced the whole exome of four patients in a Chinese four-generation spinocerebellar ataxia family and identified a missense mutation, c.1550T–G transition (L517W), in exon 10 of *TGM6*. This change is at a highly conserved position, is predicted to have a functional impact, and completely cosegregated with the phenotype. The exome results were validated using linkage analysis. The mutation we identified using exome sequencing was located in the same region (20p13–12.2) as that identified by linkage analysis, which cross-validated *TGM6* as the causative spinocerebellar ataxia gene in this family. We also showed that the causative gene could be mapped by a combined method of linkage analysis and sequencing of one sample from the family. We further confirmed our finding by identifying another missense mutation c.980A–G transition (D327G) in exon seven of *TGM6* in an additional spinocerebellar ataxia family, which

Received August 7, 2010. Revised September 20, 2010. Accepted September 25, 2010.





NGS accelerate MD research by identifying rare mutation

Brain Advance Access published November 23, 2010
doi:10.1093/brain/awq323 Brain 2010: Page 1 of 9 | 1
BRAIN
A JOURNAL OF NEUROLOGY

TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing

Jun Ling Wang,[1,2,*] Xu Yang,[3,*] Kun Xia,[2,*] Zheng Mao Hu,[2] Ling Weng,[1] Xin Jin,[3,4] Hong Jiang,[1,5] Peng Zhang,[3] Lu Shen,[1,5] Ji Feng Guo,[1,5] Nan li,[1] Ying Rui Li,[3] Li Fang Lei,[1] Jie Zhou,[1] Juan Du,[1] Ya Fang Zhou,[1] Qian Pan,[2] Jian Wang,[3] Jun Wang,[3,6] Rui Qiang Li[3,6] and Bei Sha Tang[1,2,5]

1 Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan province 410008, China
2 National Key Lab of Medical Genetics of China, Changsha, Hunan province 410008, China
3 BGI-Shenzhen, Shenzhen, Guangdong province 518083, China
4 Innovative Program for Undergraduate Students, School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China
5 Neurodegenerative Disorders Research Center, Central South University, Changsha, Hunan province 410008, China
6 Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark

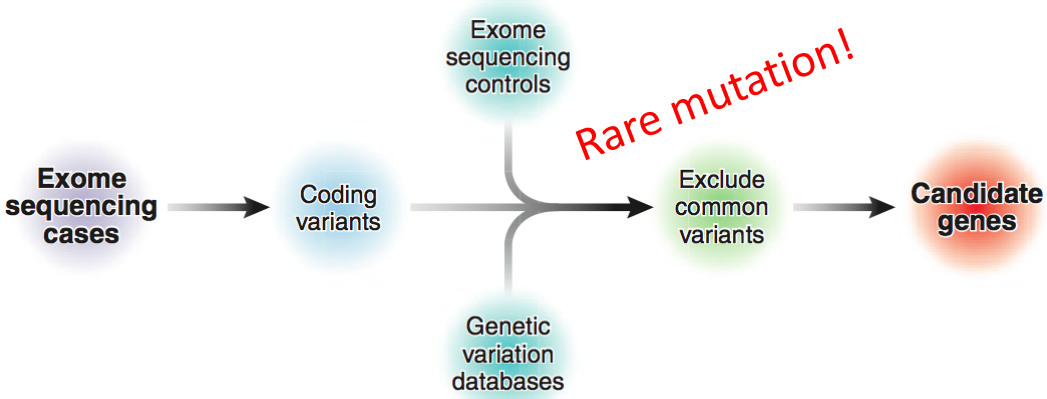
*These authors contributed equally to this work.

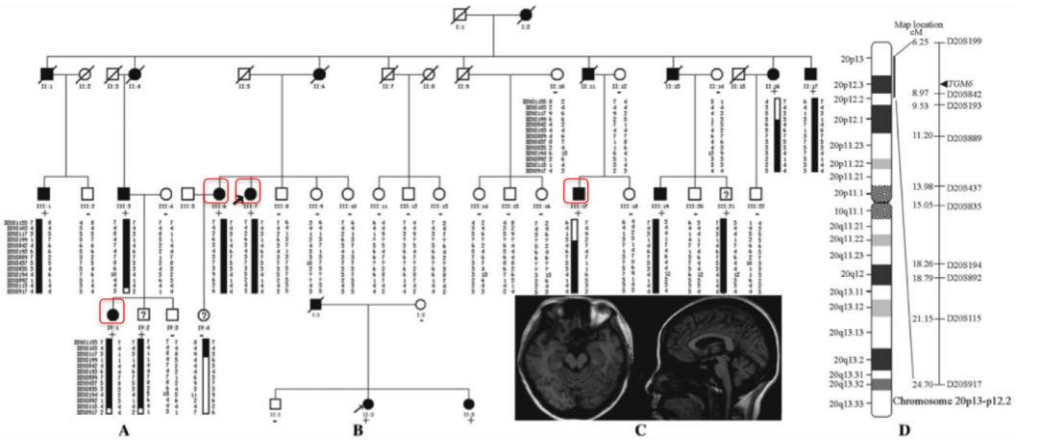
Correspondence to: Dr Bei Sha Tang,
Department of Neurology,
Xiangya Hospital,
Central South University, Changsha,
Hunan Province 410008, China
E-mail: bstang7398@yahoo.com.cn

Correspondence may also be addressed to: Rui Qiang Li,
BGI-Shenzhen, Beishan Industrial Zone,
Yantian Distince,
Shenzhen, 518083,
China.
E-mail: lirq@genomics.org.cn

Autosomal-dominant spinocerebellar ataxias constitute a large, heterogeneous group of progressive neurodegenerative diseases with multiple types. To date, classical genetic studies have revealed 31 distinct genetic forms of spinocerebellar ataxias and identified 19 causative genes. Traditional positional cloning strategies, however, have limitations for finding causative genes of rare Mendelian disorders. Here, we used a combined strategy of exome sequencing and linkage analysis to identify a novel spinocerebellar ataxia causative gene, *TGM6*. We sequenced the whole exome of four patients in a Chinese four-generation spinocerebellar ataxia family and identified a missense mutation, c.1550T–G transition (L517W), in exon 10 of *TGM6*. This change is at a highly conserved position, is predicted to have a functional impact, and completely cosegregated with the phenotype. The exome results were validated using linkage analysis. The mutation we identified using exome sequencing was located in the same region (20p13–12.2) as that identified by linkage analysis, which cross-validated *TGM6* as the causative spinocerebellar ataxia gene in this family. We also showed that the causative gene could be mapped by a combined method of linkage analysis and sequencing of one sample from the family. We further confirmed our finding by identifying another missense mutation c.980A–G transition (D327G) in exon seven of *TGM6* in an additional spinocerebellar ataxia family, which

Received August 7, 2010. Revised September 20, 2010. Accepted September 25, 2010.

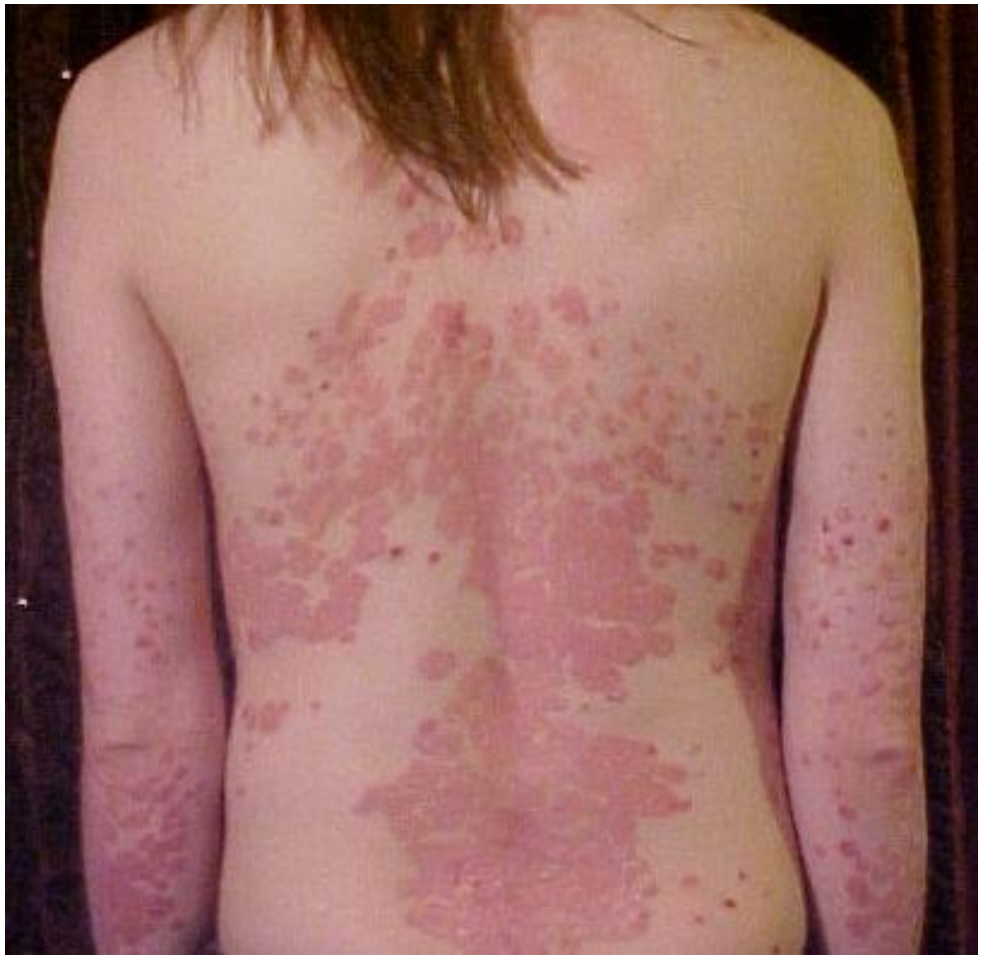




Picture spectrum of rare variants by psoriasis study

Psoriasis is a chronic, potentially disfiguring, immune-mediated inflammatory disease of the skin.

Affecting up to 3% of individuals of Western European descent and 0.1–0.3% of individuals of East Asian ancestry.

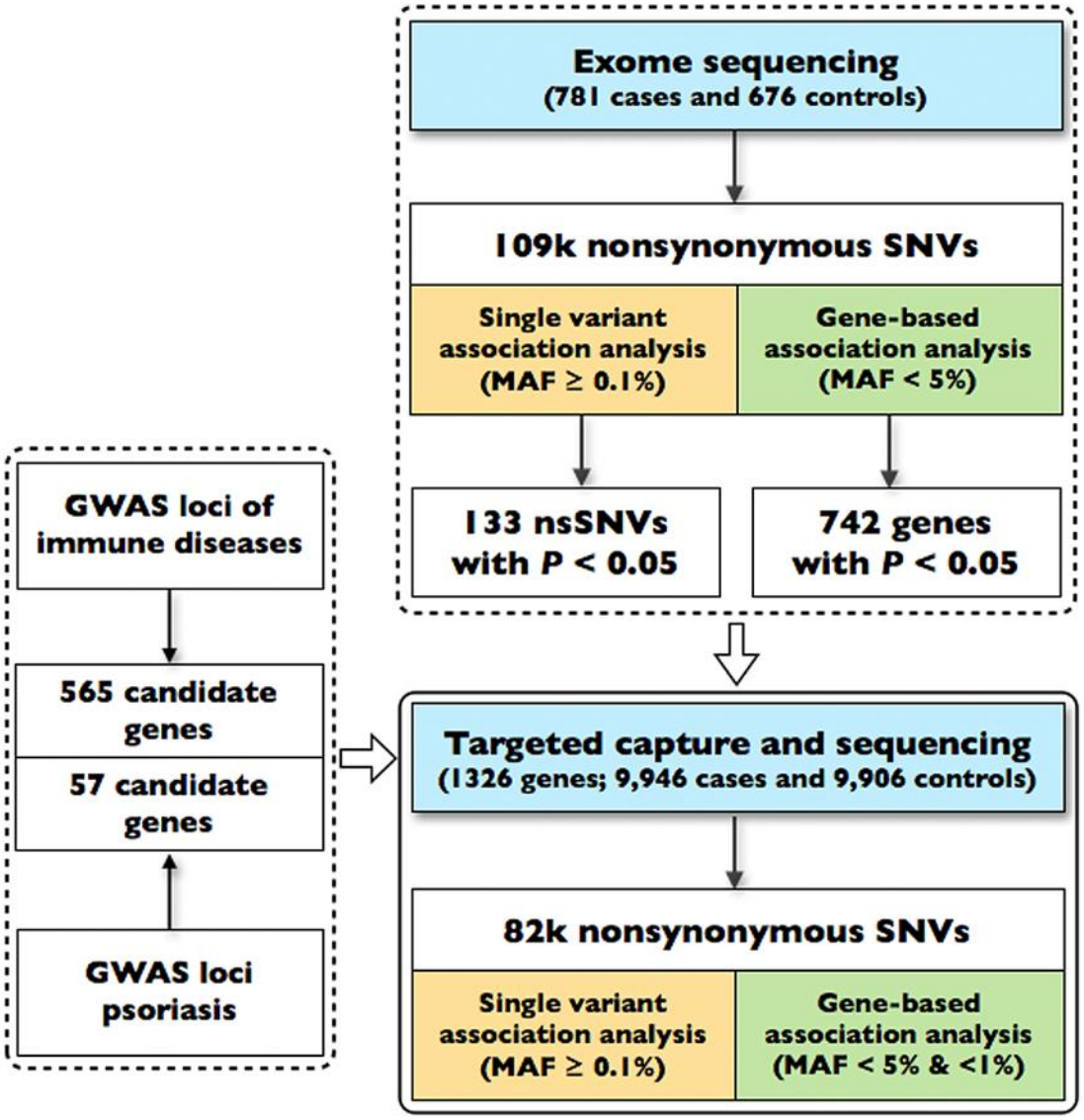


Genetic research of psoriasis

- Linkage analysis: PSORS1 through PSORS10

- GWAS study: Identified over 20 susceptibility loci for psoriasis, but less than 20% of disease variance is explained

Design of psoriasis project



Association analysis of rare missense SNVs

Chr	Gene	SNVs (hg18)	Allele	Function	Stages*	Sequencing				Combined			
						MAF		*P* value	OR	MAF		*P* value	OR (95%CI)
						Cases	Controls			Cases	Controls		
1	*TARBP1*	chr1:232631799	A/G	c.G2857A p.V953M	i	0	0.00149	2.14E-01	0.00	0.00028	0.00157	1.28E-05	0.18(0.08-0.43)
					ii	0.00031	0.00158	3.81E-05	0.19				
1	*TARBP1*	chr1:232649343	A/G	c.G1963A p.G655R	i	0	0.00149	2.15E-01	0.00	0.00047	0.00183	4.19E-05	0.26(0.13-0.52)
					ii	0.00051	0.00185	1.16E-04	0.28				
19	*FUT2*	chr19:53898296	T/C	c.C271T p.R91W	i	0.00418	0	3.18E-02	NA	0.00192	0.00053	5.51E-05	3.61(1.85-7.04)
					ii	0.00175	0.00057	6.32E-04	3.08				
19	*FUT2*	chr19:53898541	G/C	c.C516G p.D172E	i	0.00199	0.00074	3.73E-01	2.69	0.00066	0.00010	2.92E-03	6.92(1.57-30.45)
					ii	0.00056	0.00005	3.99E-03	10.95				

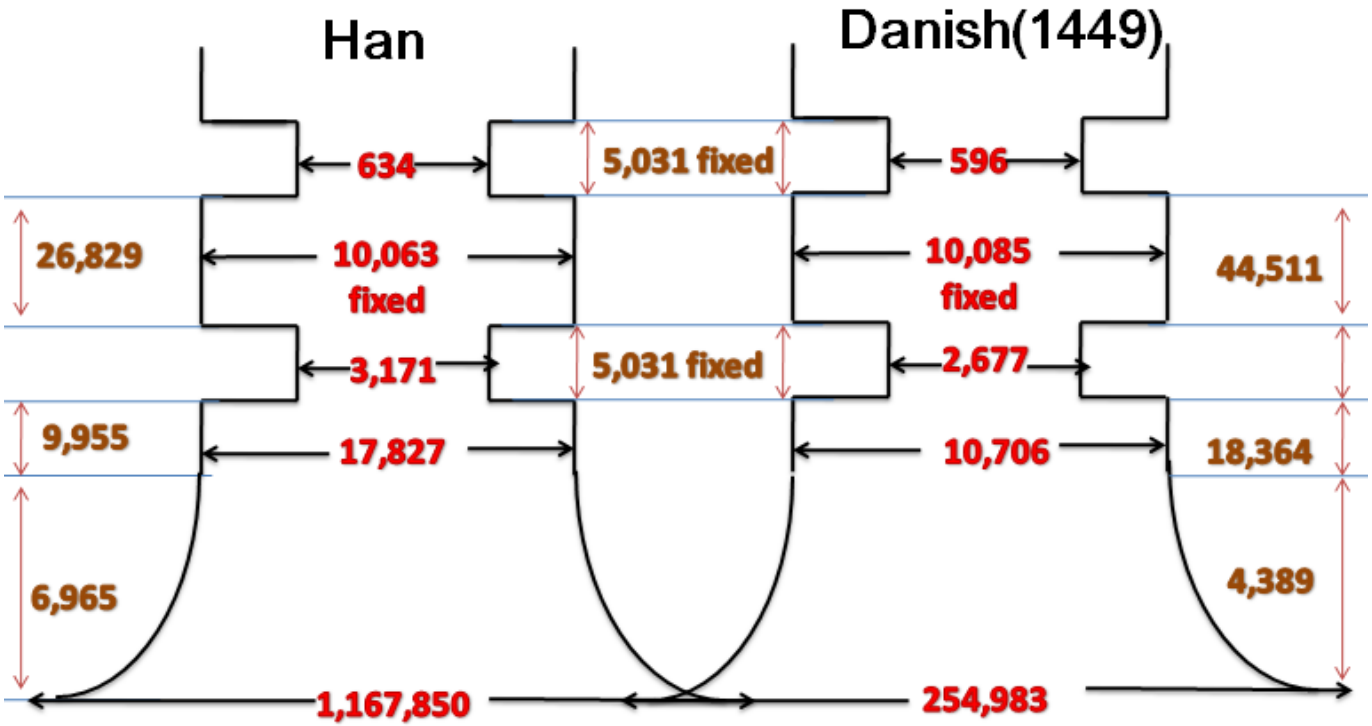
(a) Distribution of SNVs with different MAF of exome sequencing data

MAF	#of SNVs	Proportion
Singleton	203,596	37.15%
Doubleton	55,436	10.12%
(0,0.005)	346,322	63.20%
[0.005,0.01)	27,003	4.93%
[0.01,0.05)	53,274	9.72%
[0.05,0.1)	25,295	4.62%
[0.1,0.2)	31,881	5.82%
[0.2,0.3)	22,996	4.20%
[0.3,0.4)	20,755	3.79%
[0.4,0.5]	20,452	3.73%
All	547,978	100.00%

(b) Distribution of SNVs with different MAF of target sequencing data

MAF	#of SNVs	Proportion
Singleton	61,193	13.44%
Doubleton	40,453	8.88%
(0,0.005)	439,080	96.43%
[0.005,0.01)	2,894	0.64%
[0.01,0.05)	5,296	1.16%
[0.05,0.1)	1,934	0.42%
[0.1,0.2)	2,296	0.50%
[0.2,0.3)	1,430	0.31%
[0.3,0.4)	1,273	0.28%
[0.4,0.5]	1,111	0.24%
All	455,314	100.00%

Prediction of demographic history in 1500 exome of Han and Danish

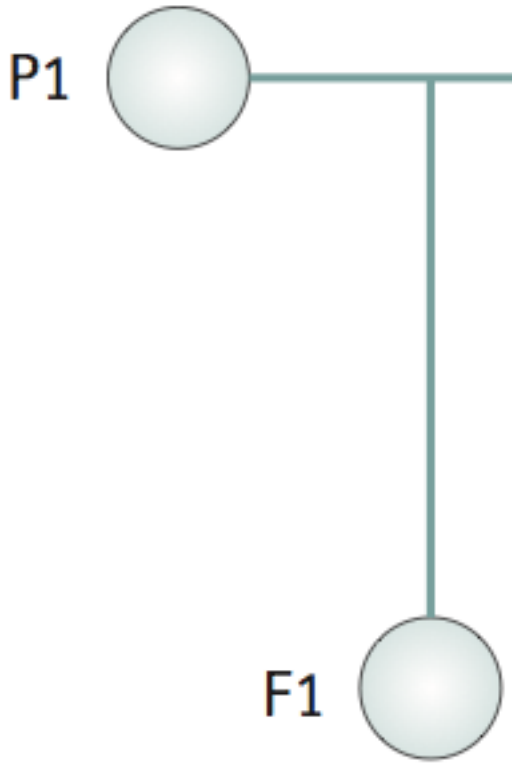


Best Likelihood through 1000 times of optimizations: -3850.54

Best Likelihood through 1000 times of optimizations: -4007.94

Puny Model

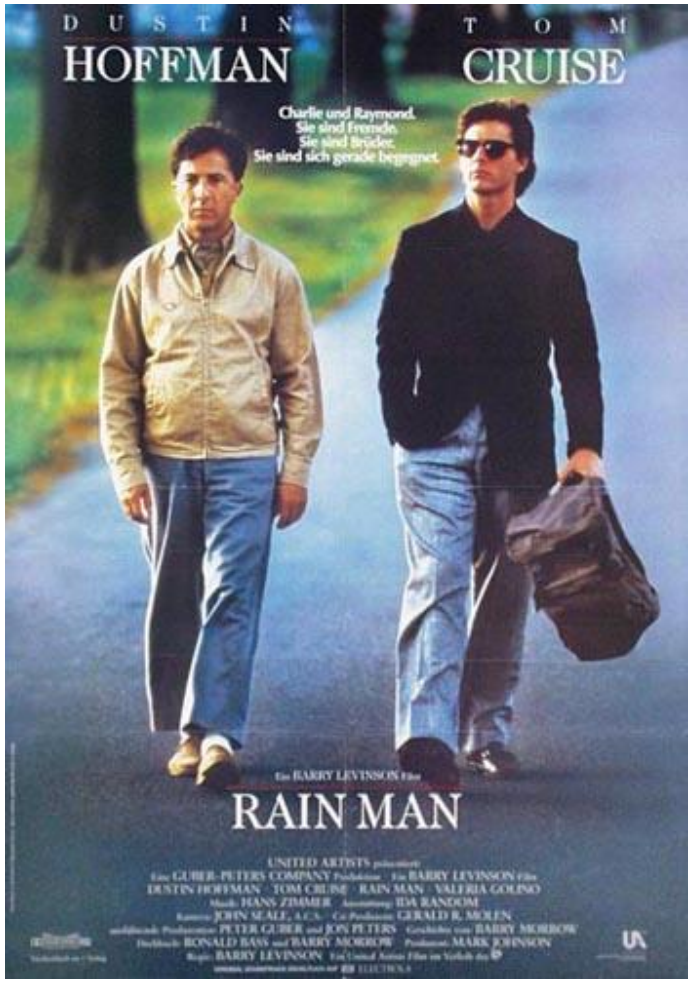
Novo mutation originated every generation



- 74 *de novo* SNVs, 3 *de novo* indels and 0.02 *de novo* CNVs per genome
- 1 *de novo* mutation per exome
- 1 out of 20,563 protein-coding genes are hit by a *de novo* mutation per generation

J. A. Veltman, H. G. Brunner, ***Nature Reviews Genetics*** 13, 565 (2012).

Explore origin of rare variants by autism study

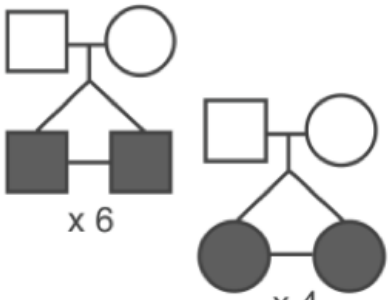


Prevalence: 1/55 children in US

Causation: environment, epigenetic factors and genetic components

De novo mutation plays an important role in Autism Spectrum Disorders

Workflow of identical twins autism genomes project



10 quad families, monozygotic twins concordant for autism diagnosis

high-coverage (~40x), paired-end whole genome sequencing with Illumina

alignment + variant calling

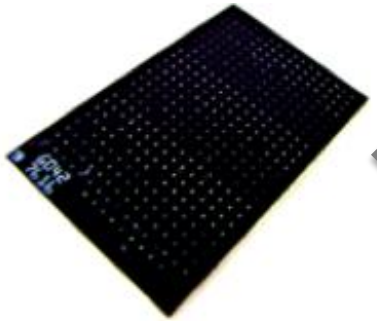
alignment with BWA and variant calling with GATK

forestDNM

DNM calling with forestDNM, a custom machine learning tool

putative germline mutations

668 germline DNMs in 10 families

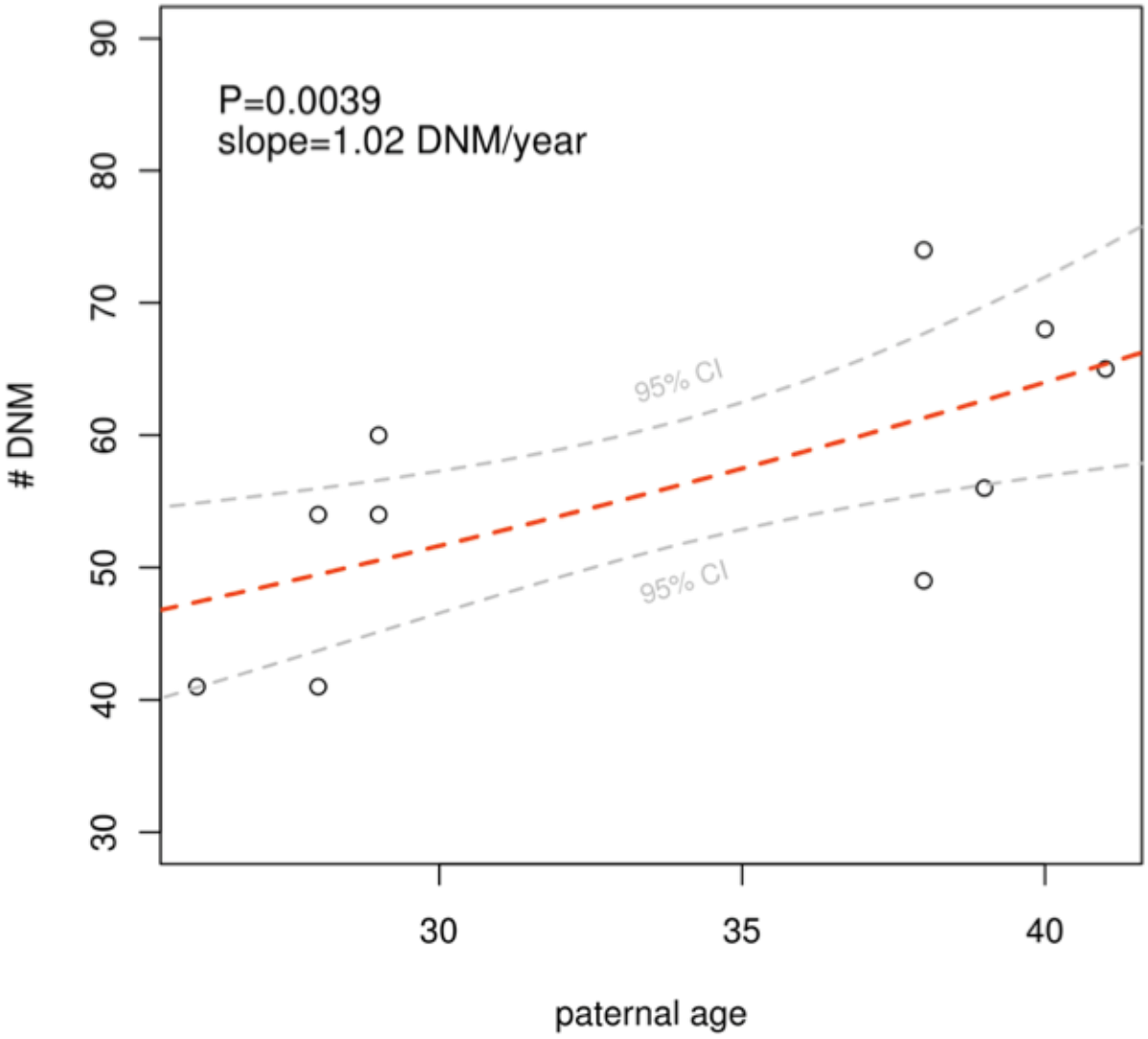


DNM validation with Sanger sequencing and Sequenom genotyping

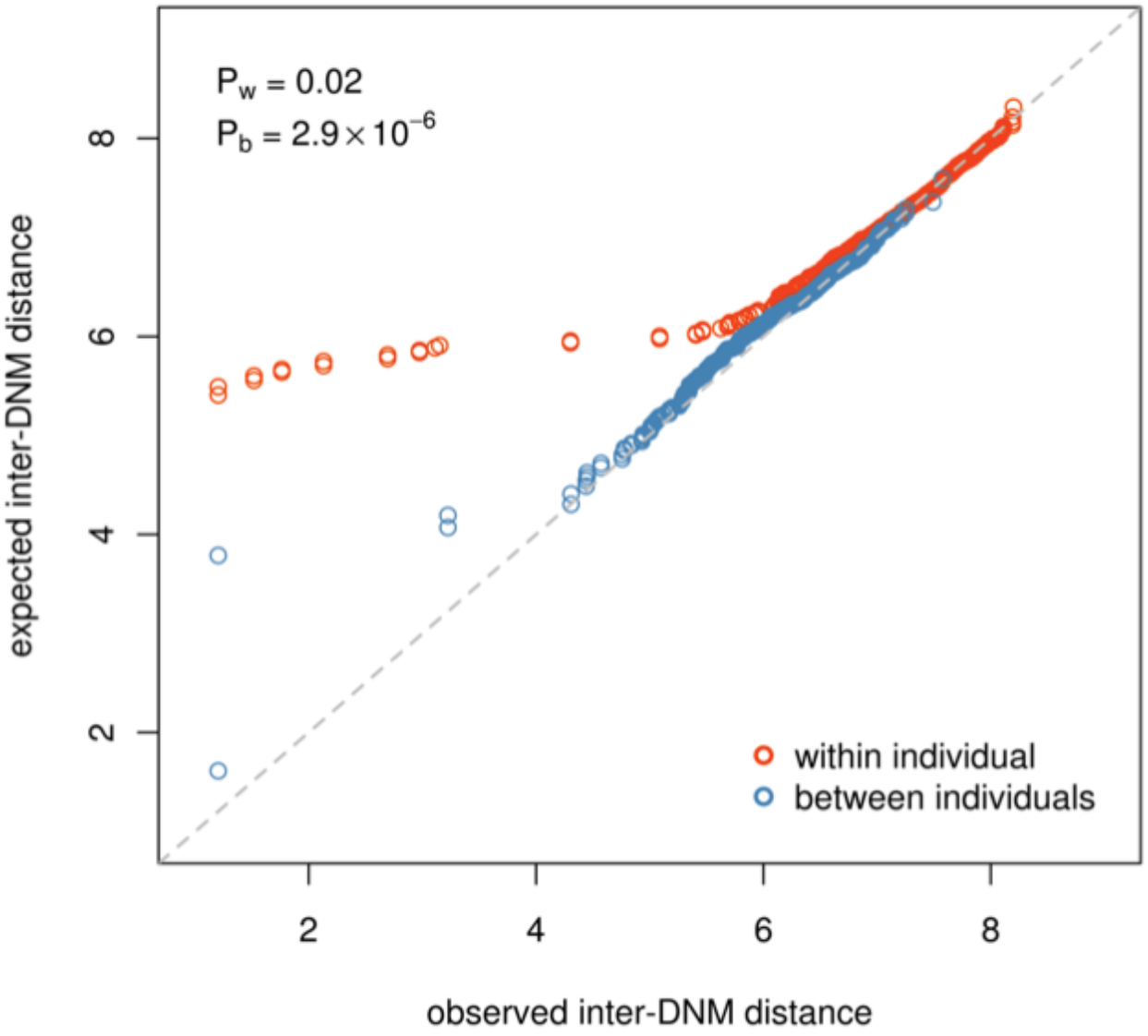
validated germline mutations

581 germline DNMs in 10 families

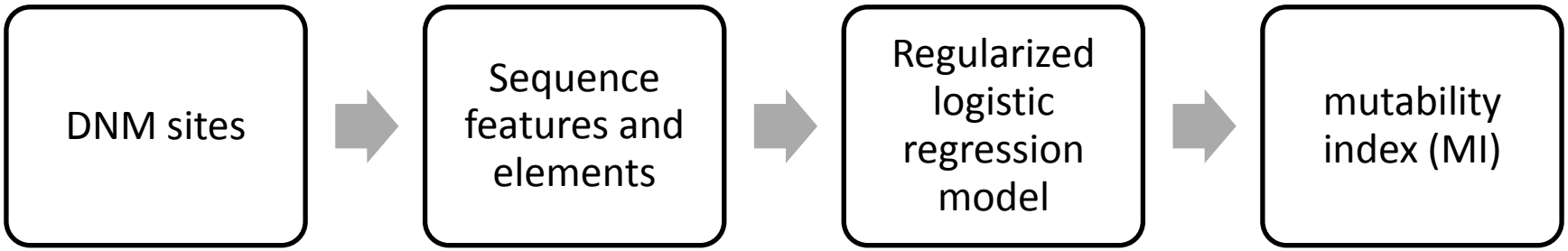
Paternal age and genome-wide mutation rates



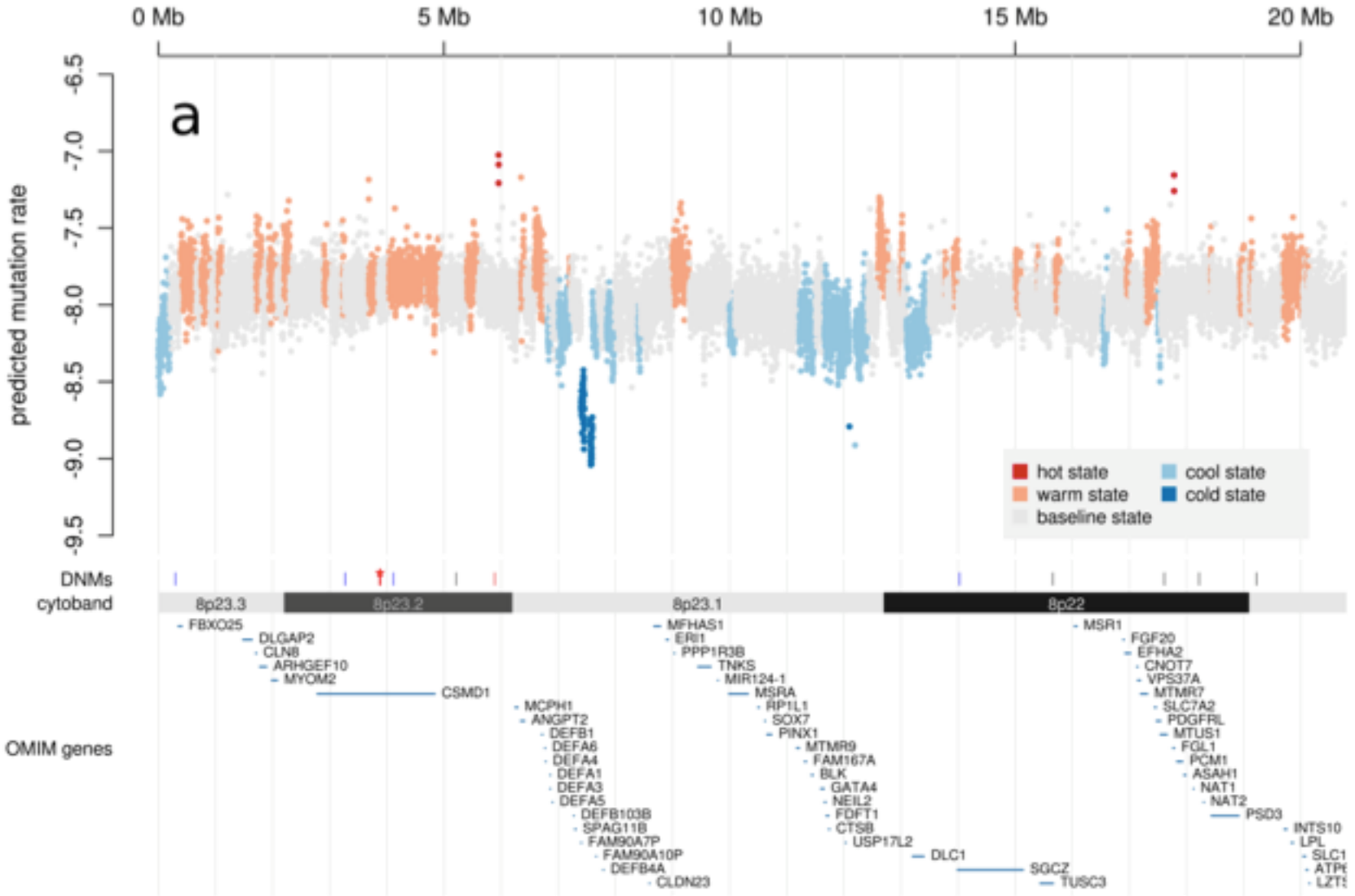
Non-random distribution of DNMs in the genome

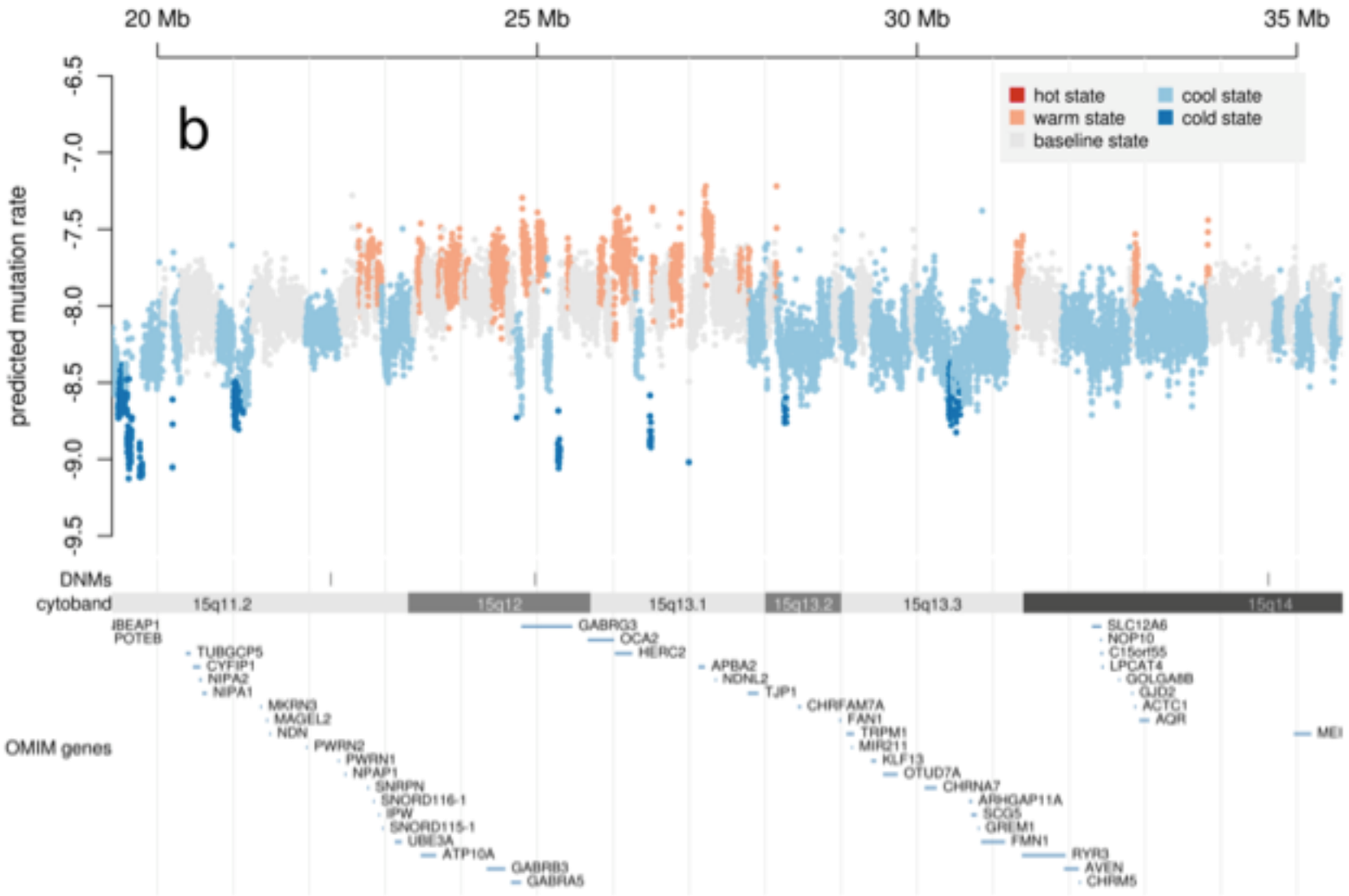


Determining Intrinsic Properties of the Genome that Influence Mutability

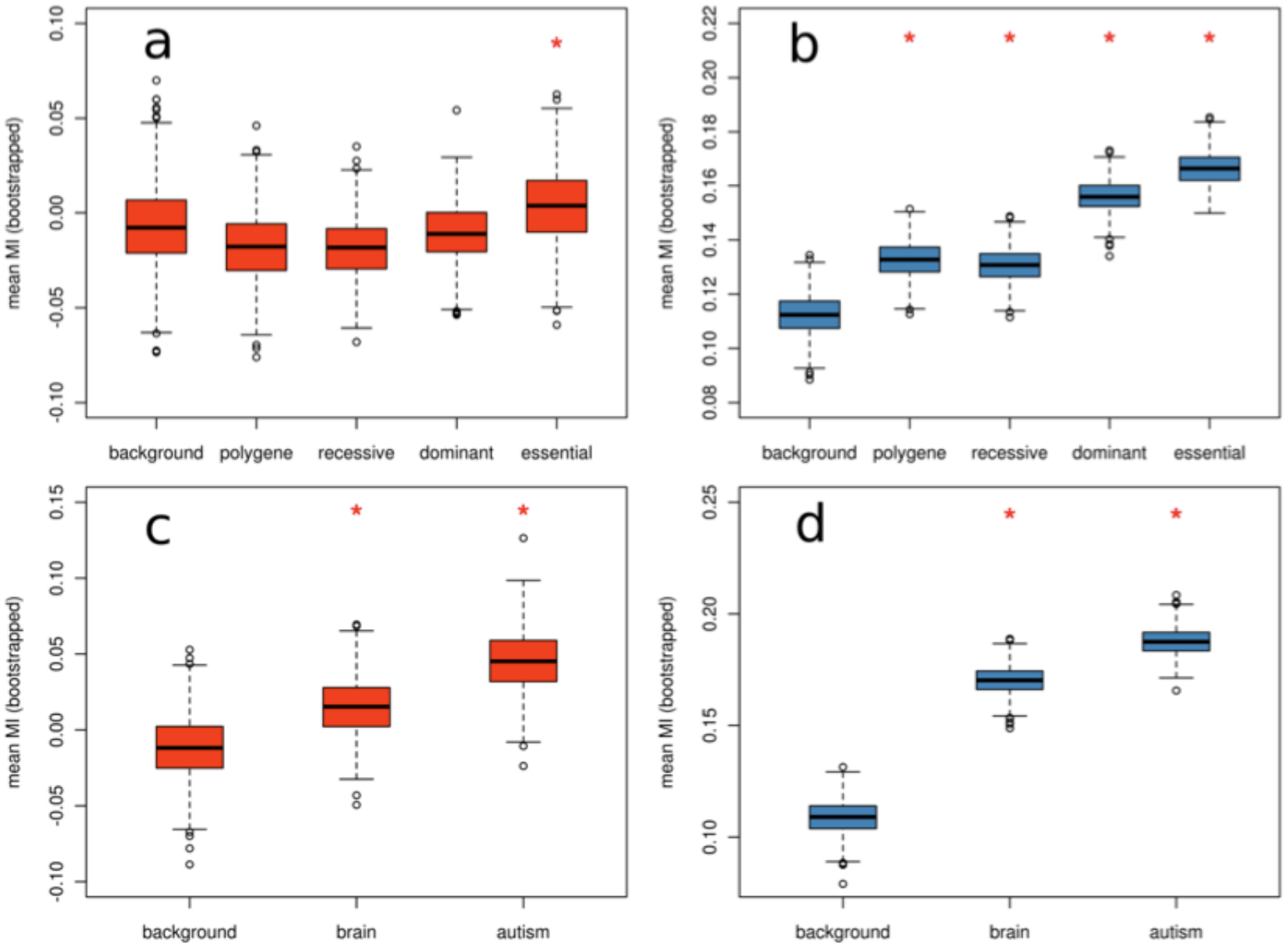


Landscape of mutability in the genome

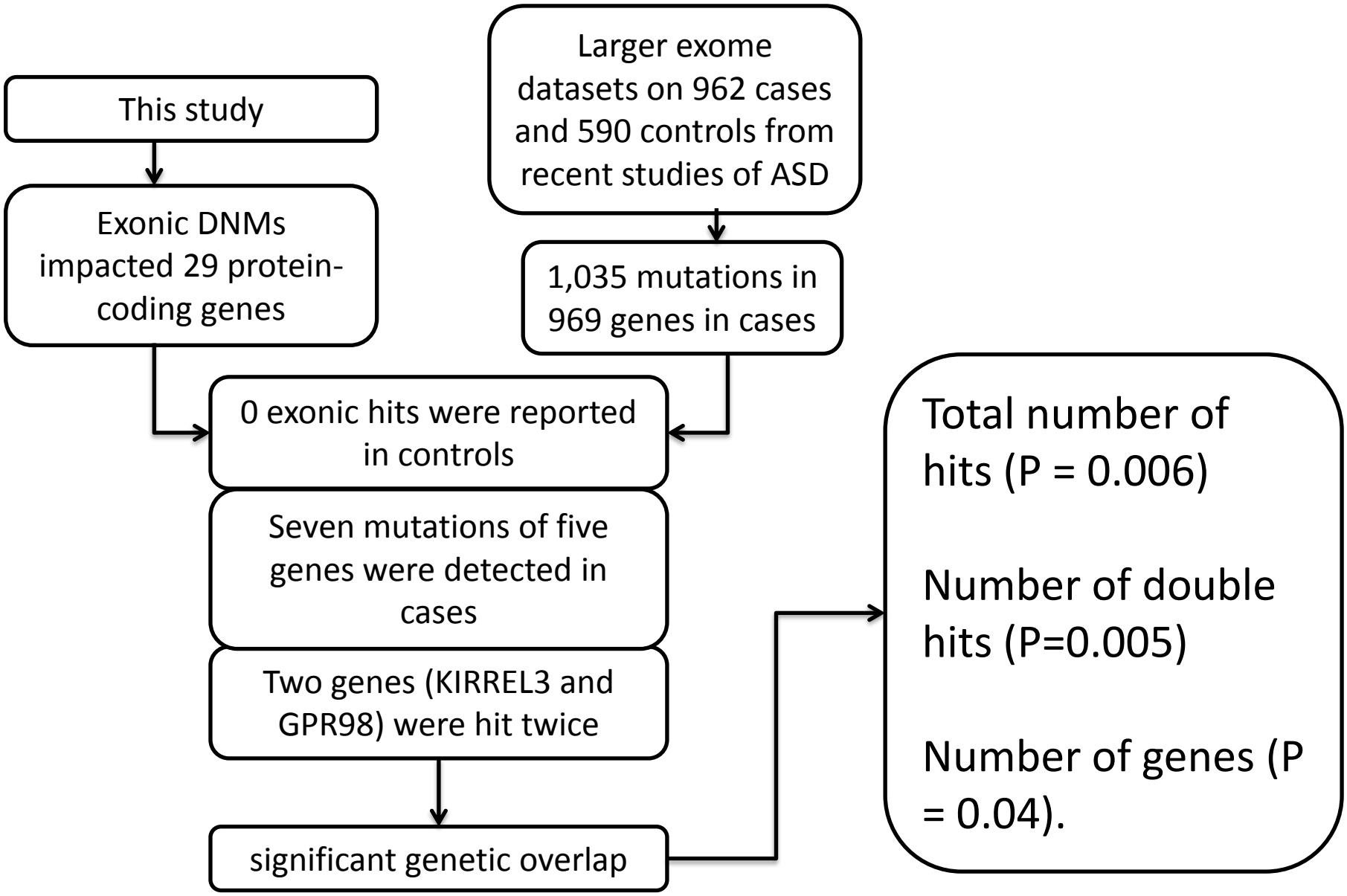



20 Mb
25 Mb
30 Mb
35 Mb
b
hot state
warm state
baseline state
cool state
cold state
predicted mutation rate
-6.5
-7.0
-7.5
-8.0
-8.5
-9.0
-9.5
DNMs
cytoband
15q11.2
15q12
15q13.1
15q13.2
15q13.3
15q14
OMIM genes
BEAP1
POTEB
TUBGCP5
CYFIP1
NIPA2
NIPA1
MKRN3
MAGEL2
NDN
PWRN2
PWRN1
NPAP1
SNRPN
SNORD116-1
IPW
SNORD115-1
UBE3A
ATP10A
GABRB3
GABRA5
GABRG3
OCA2
HERC2
APBA2
NDNL2
TJP1
CHRFAM7A
FAN1
TRPM1
MIR211
KLF13
OTUD7A
CHRNA7
ARHGAP11A
SCG5
GREM1
FMN1
RYR3
AVEN
CHRM5
SLC12A6
NOP10
C15orf55
LPCAT4
GOLGA8B
GJD2
ACTC1
AQR
MEI

Disease Genes are Characterized by High Mutability



Exonic Mutations in MZ Twins are Significantly Associated with ASD



Whole Genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151, 1431-1442 (2012).

Acknowledgement

Thank you!

jinxin@genomics.cn