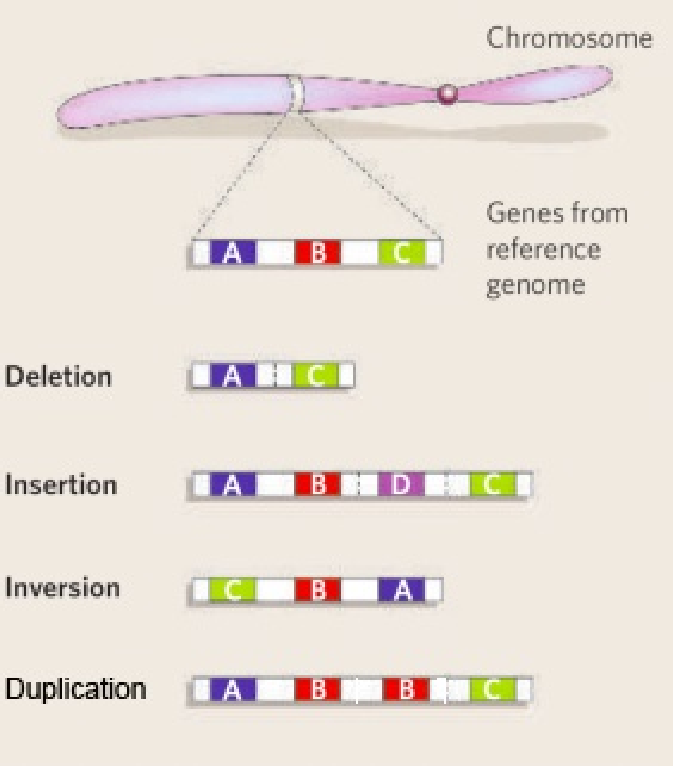
Your Genes and Your Health

http://bio84.stanford.edu/

Structural Variants in the Human Genome



Doug Brutlag. Professor Emeritus of
Biochemistry & Medicine (by courtesy)
Stanford University School of Medicine

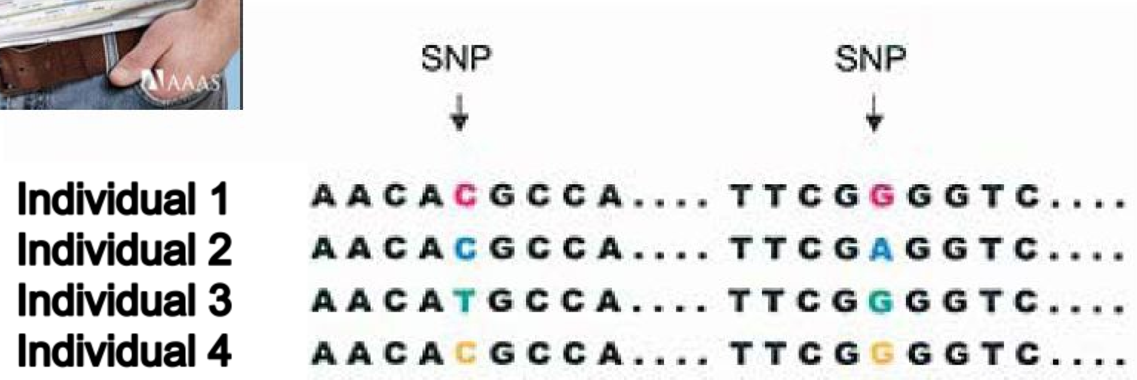
Human Genetic Variation
2007 Scientific Breakthrough of the Year

2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007

"It's all about me!"

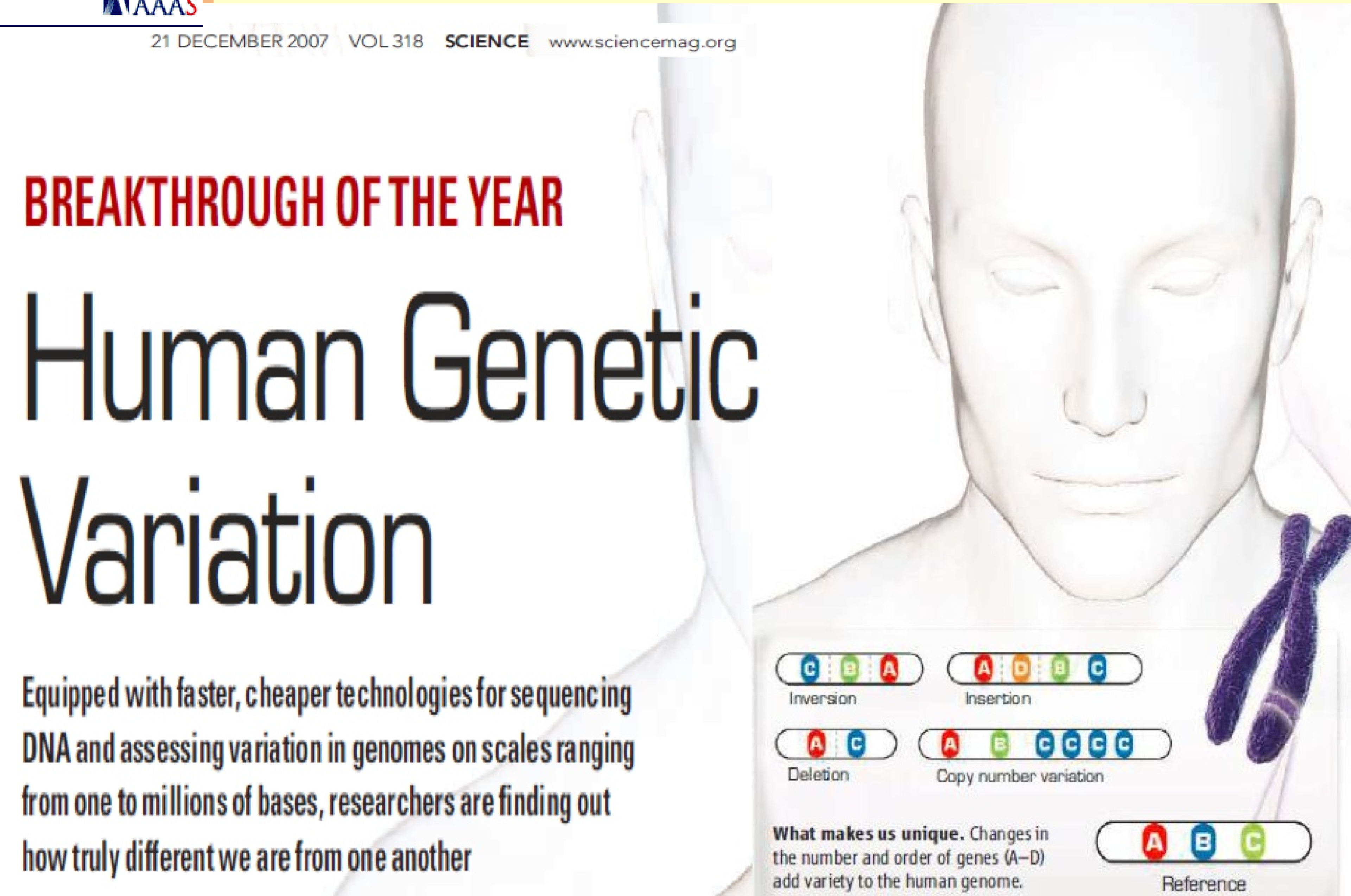
Single Nucleotide Polymorphisms (SNPs)



Individual 1	AACACGCCA....	TTCGGGGTC....
Individual 2	AACACGCCA....	TTCGAGGTC....
Individual 3	AACATGCCA....	TTCGGGGTC....
Individual 4	AACACGCCA....	TTCGGGGTC....

Science
AAAS

Structural Genomic Variants



21 DECEMBER 2007 VOL 318 SCIENCE www.sciencemag.org

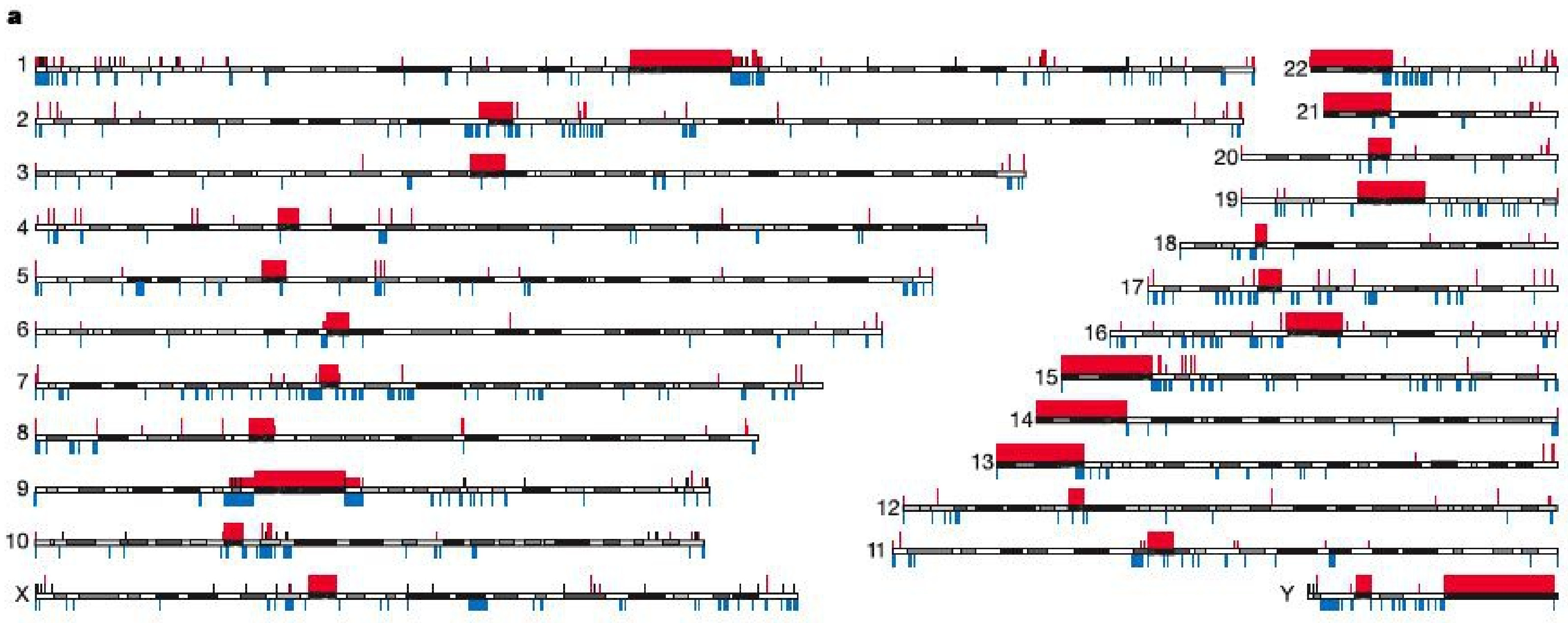
BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

Duplications and Deletions in the Human Genome



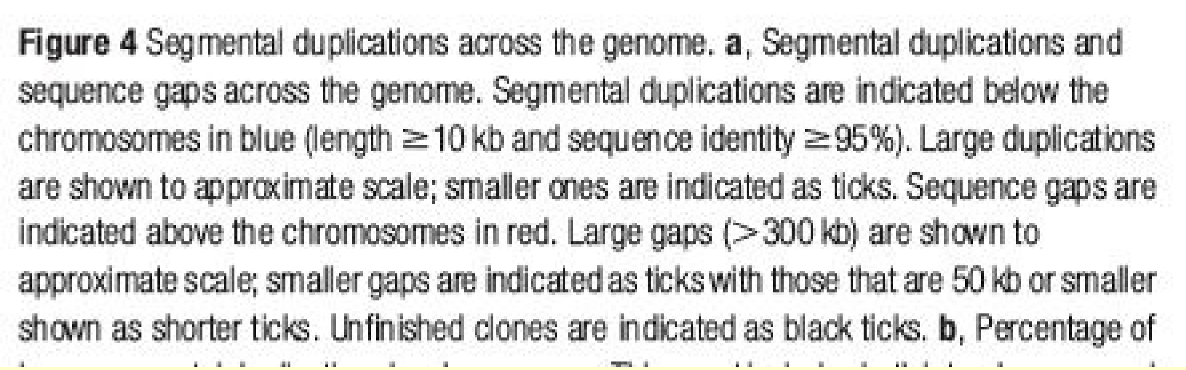
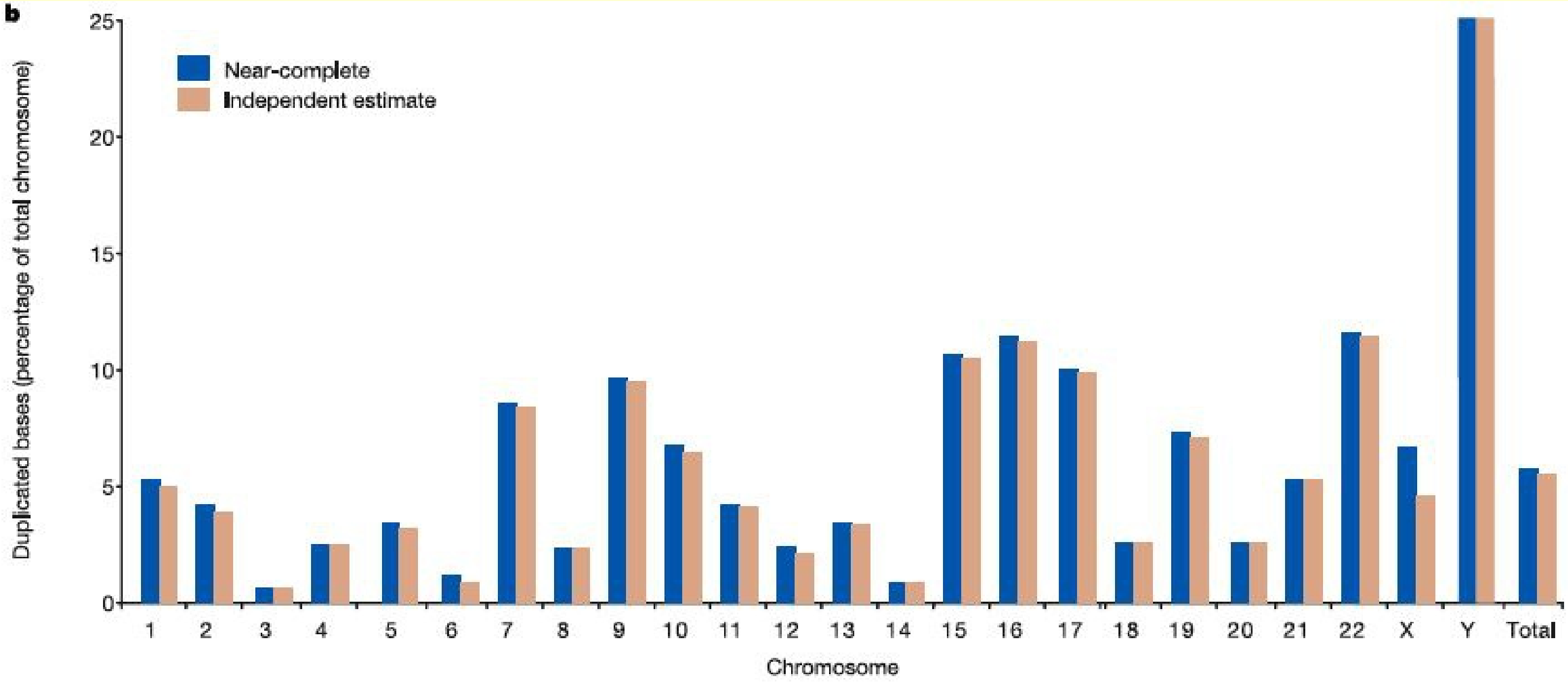


Figure 4 Segmental duplications across the genome. **a**, Segmental duplications and sequence gaps across the genome. Segmental duplications are indicated below the chromosomes in blue (length ≥10 kb and sequence identity ≥95%). Large duplications are shown to approximate scale; smaller ones are indicated as ticks. Sequence gaps are indicated above the chromosomes in red. Large gaps (>300 kb) are shown to approximate scale; smaller gaps are indicated as ticks with those that are 50 kb or smaller shown as shorter ticks. Unfinished clones are indicated as black ticks. **b**, Percentage of

Percentage of Chromosomes Duplicated



The Spectrum of Variations in the Human Genome

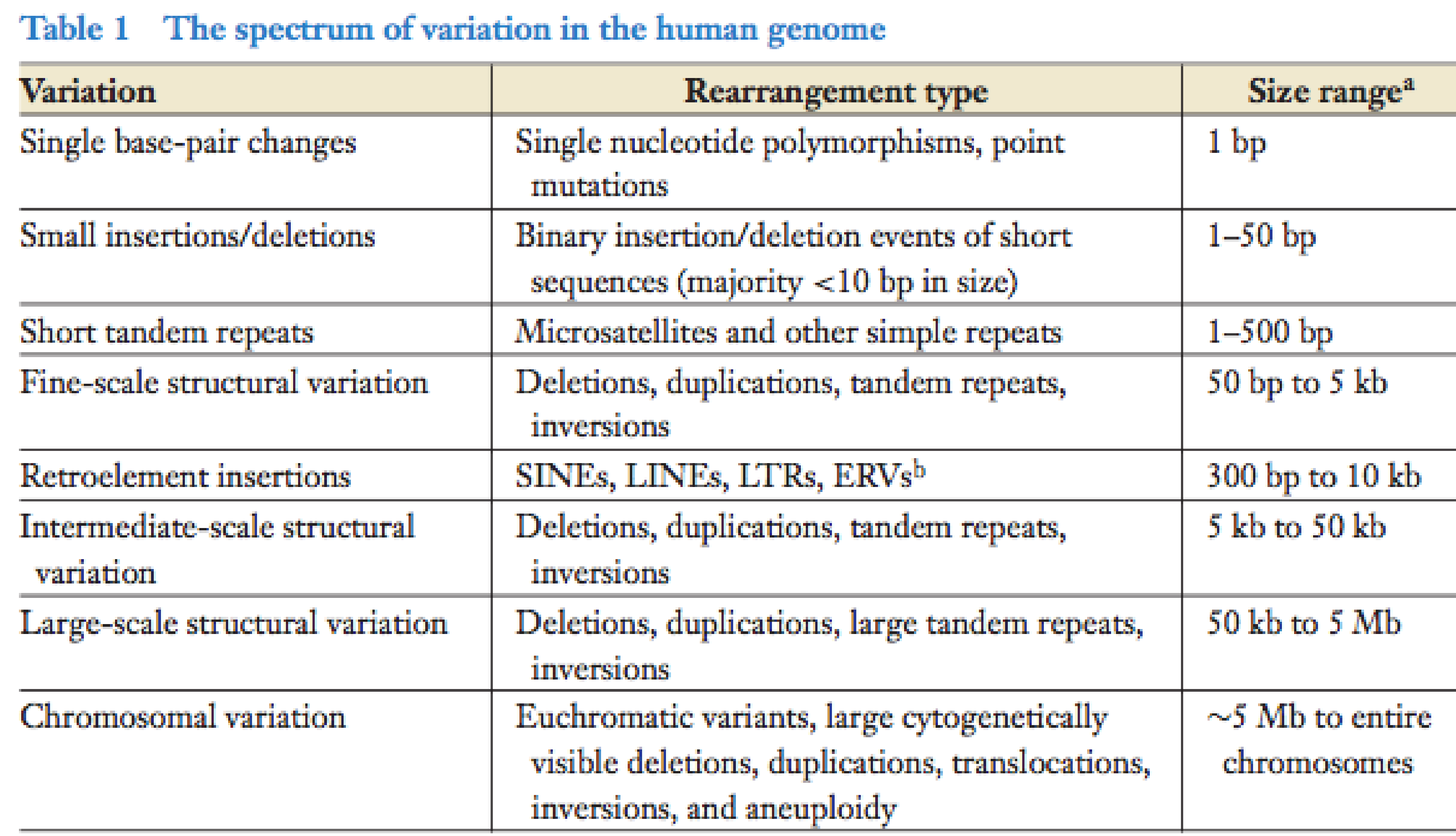
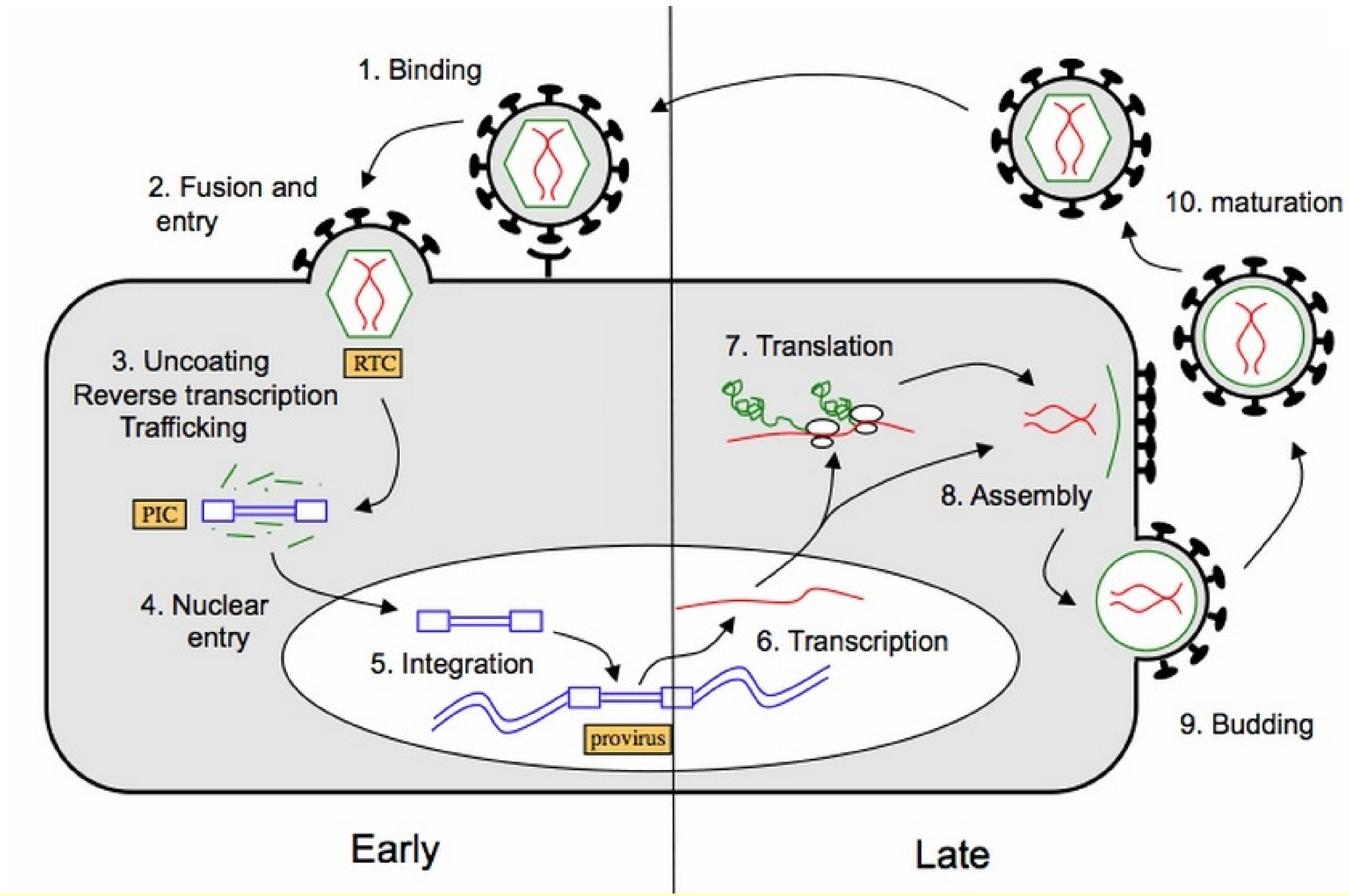


Table 1 The spectrum of variation in the human genome

Variation	Rearrangement type	Size range[a]
Single base-pair changes	Single nucleotide polymorphisms, point mutations	1 bp
Small insertions/deletions	Binary insertion/deletion events of short sequences (majority <10 bp in size)	1–50 bp
Short tandem repeats	Microsatellites and other simple repeats	1–500 bp
Fine-scale structural variation	Deletions, duplications, tandem repeats, inversions	50 bp to 5 kb
Retroelement insertions	SINEs, LINEs, LTRs, ERVs[b]	300 bp to 10 kb
Intermediate-scale structural variation	Deletions, duplications, tandem repeats, inversions	5 kb to 50 kb
Large-scale structural variation	Deletions, duplications, large tandem repeats, inversions	50 kb to 5 Mb
Chromosomal variation	Euchromatic variants, large cytogenetically visible deletions, duplications, translocations, inversions, and aneuploidy	~5 Mb to entire chromosomes

Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Retroviral Life Cycle

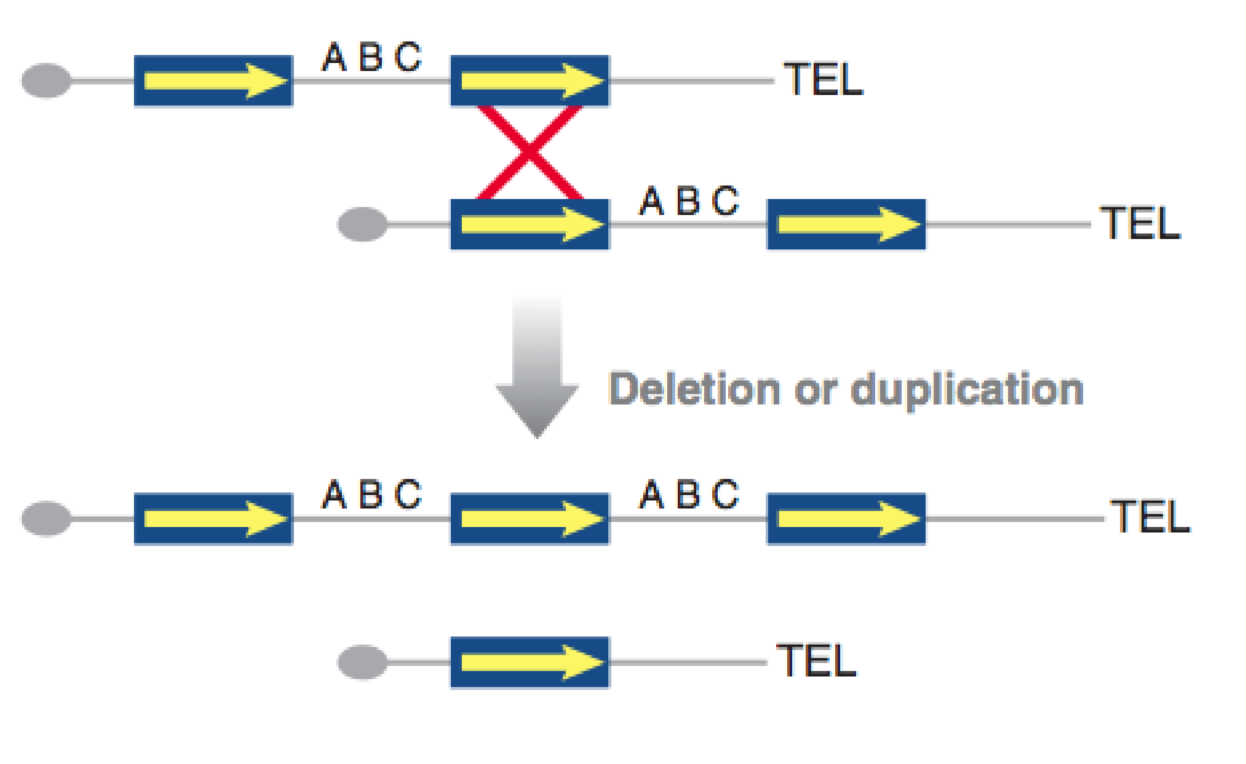


http://www.nimr.mrc.ac.uk/research/kate-bishop/

Repeated Elements in the Human Genome ERVs, LINES, SINES and ALUs

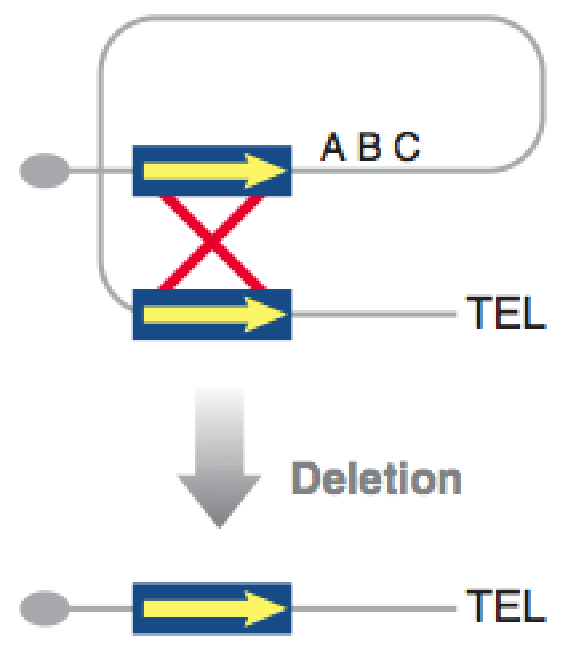
- ERVs-Endogenous Retroviruses
  - 10,000 base long RNA genome
  - Converted to DNA and integrate into genome with help of RNA reverse transcriptase and integrase enzymes and long tandem repeats (LTRs)
  - Transcribed into RNA and produce virus (HIV)
- LINES-Long interspersed nuclear elements
  - About 868,000 in human genome
  - 6,500 base pairs long including LTRs
  - Encode reverse transcriptase and integrase
  - Copy-paste mechanism to insert elsewhere
- SINES-Short Interspersed nuclear elements
  - Millions in human genome
  - 100-400 bases long
  - Often contain RNA polymerase III promoters but no genes
- ALUs- The most common SINE
  - 1,500,000 copies = 11% of human genome
  - 280 base pairs in length
  - Contain an RNA Polymerase III promoter, two Alu sites
  - Appear to evolve from 7S RNA signal recognition particle

Unequal Crossing Over Leads to Duplication and Deletion



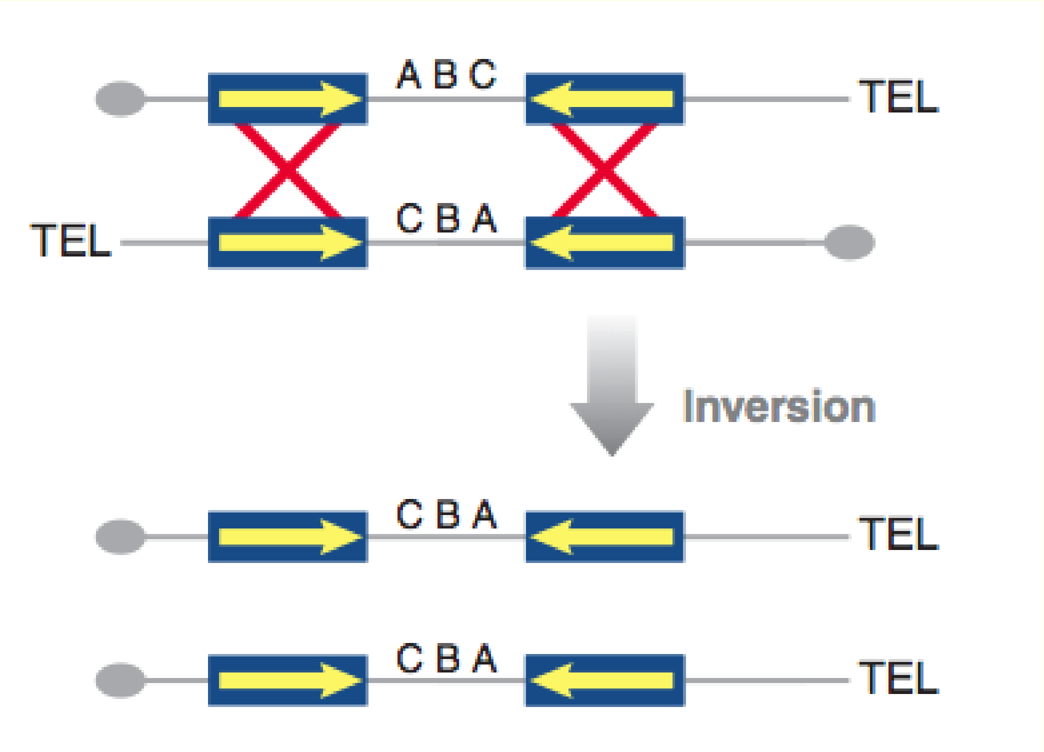
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Intra-Chromosomal Crossing Over Leads to Deletion



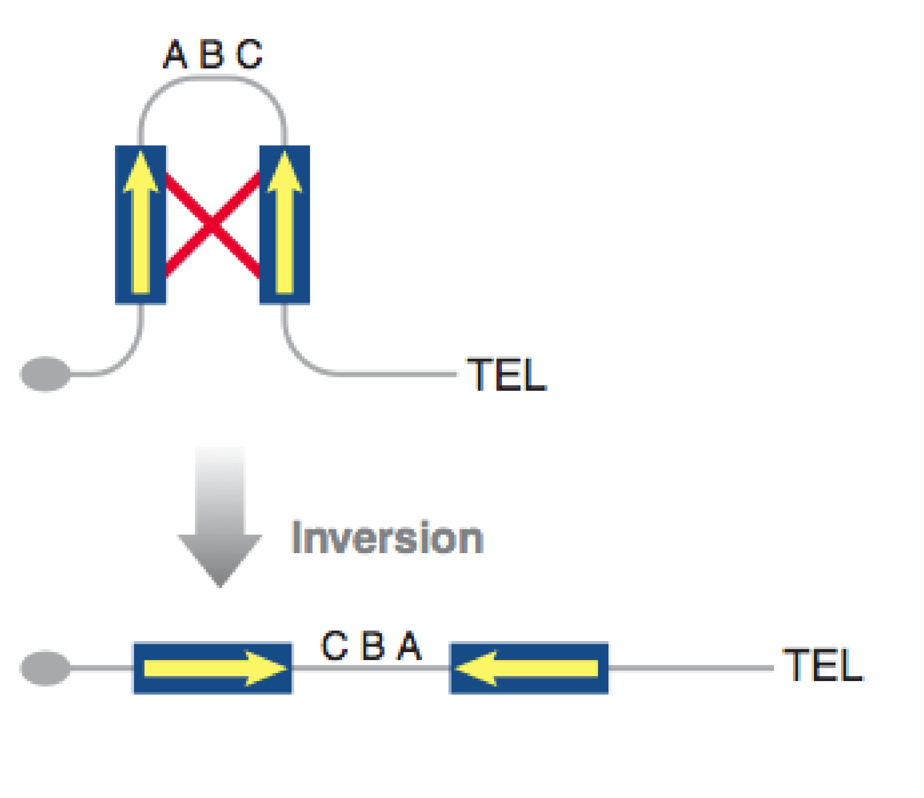
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Inter-Chromosomal Crossing Over Leads to Inversion



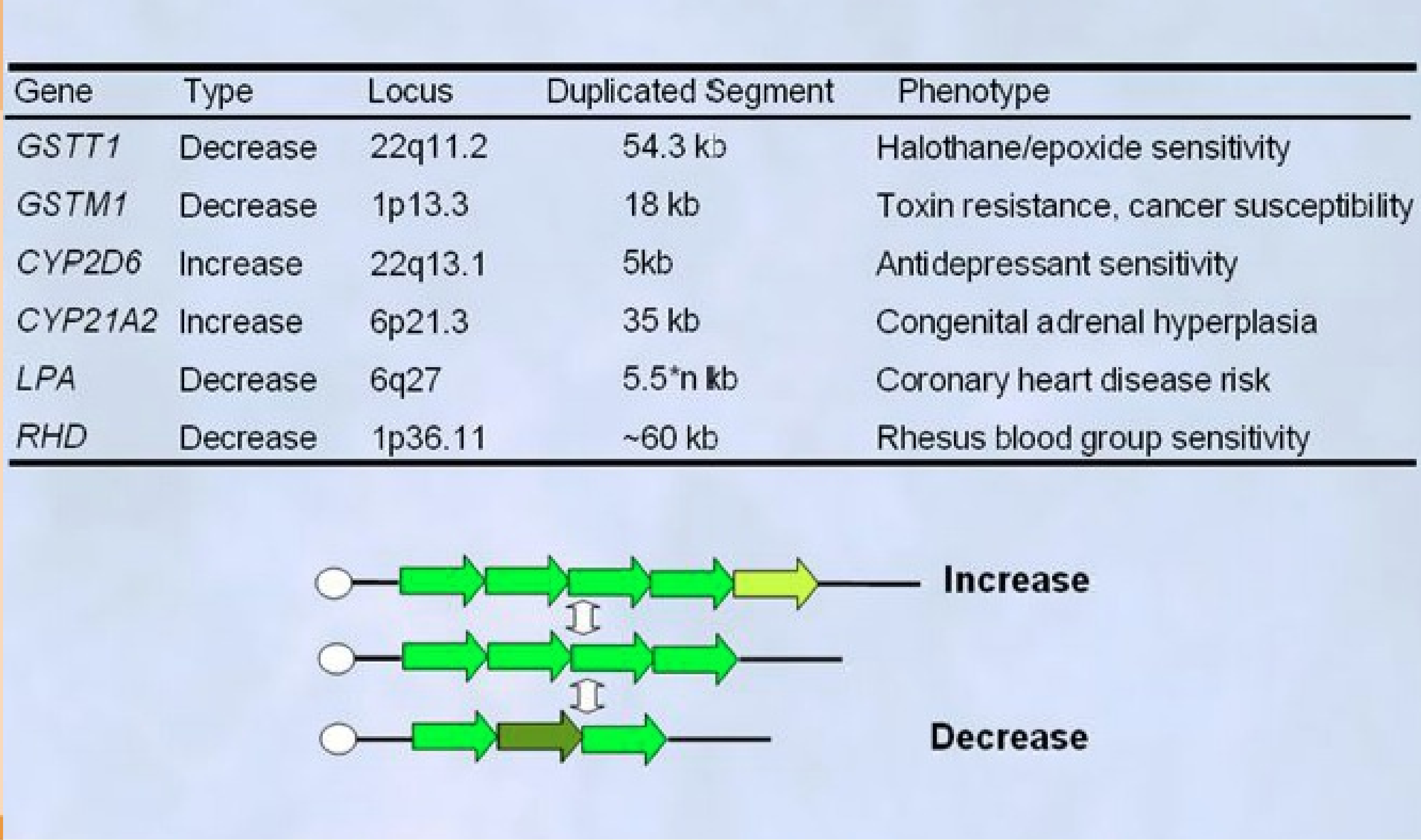
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Intra-Chromosomal Crossing Over Can Also Lead to Inversion



Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Copy Number Variation and Disease 2002

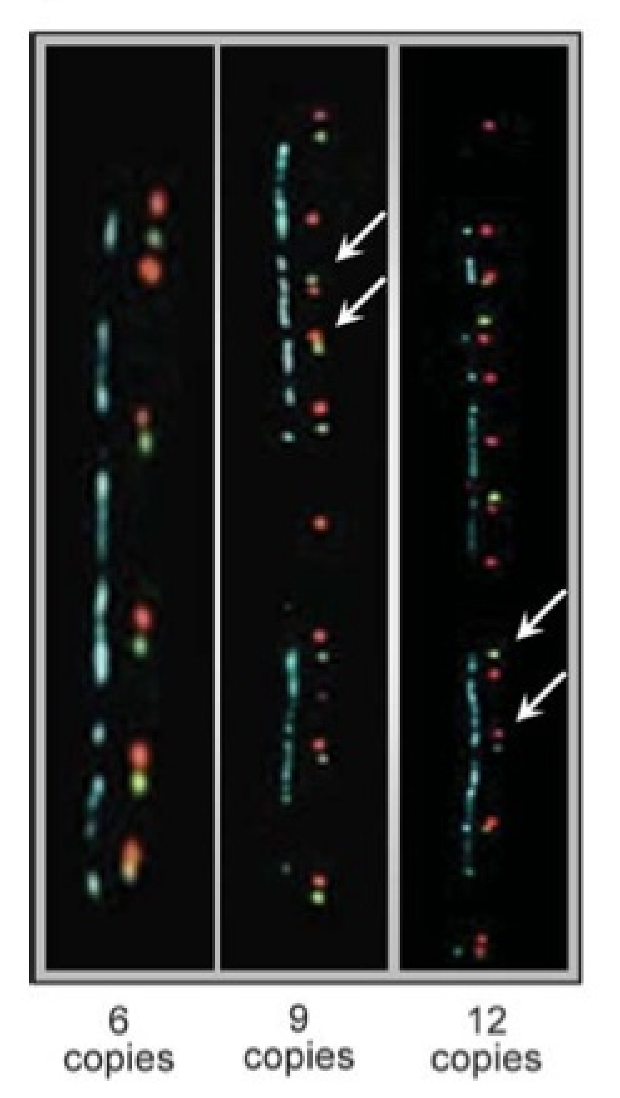


Gene	Type	Locus	Duplicated Segment	Phenotype
GSTT1	Decrease	22q11.2	54.3 kb	Halothane/epoxide sensitivity
GSTM1	Decrease	1p13.3	18 kb	Toxin resistance, cancer susceptibility
CYP2D6	Increase	22q13.1	5kb	Antidepressant sensitivity
CYP21A2	Increase	6p21.3	35 kb	Congenital adrenal hyperplasia
LPA	Decrease	6q27	5.5*n kb	Coronary heart disease risk
RHD	Decrease	1p36.11	~60 kb	Rhesus blood group sensitivity

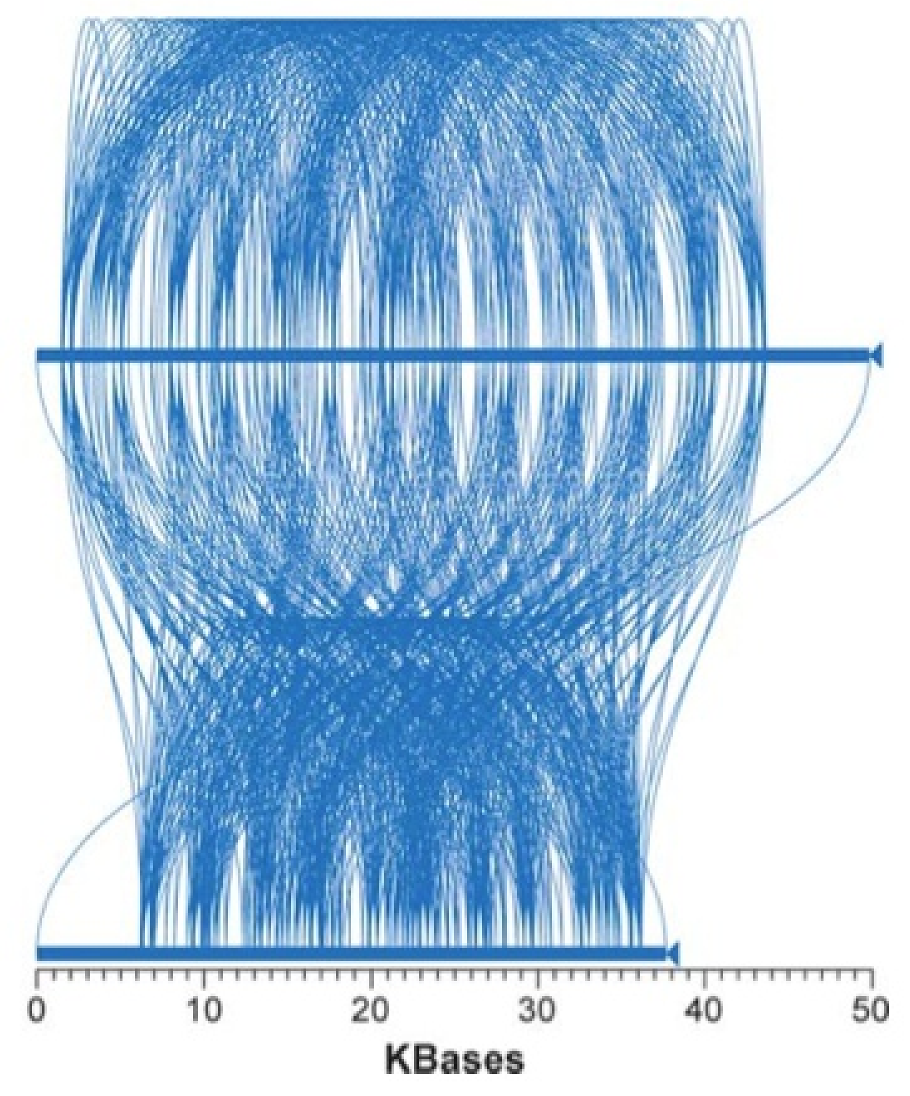
Henry Stewart Talks: Evan Eichler

Variations in α-Amylase Gene Tandem Repeat Arrays

FISH on DNA

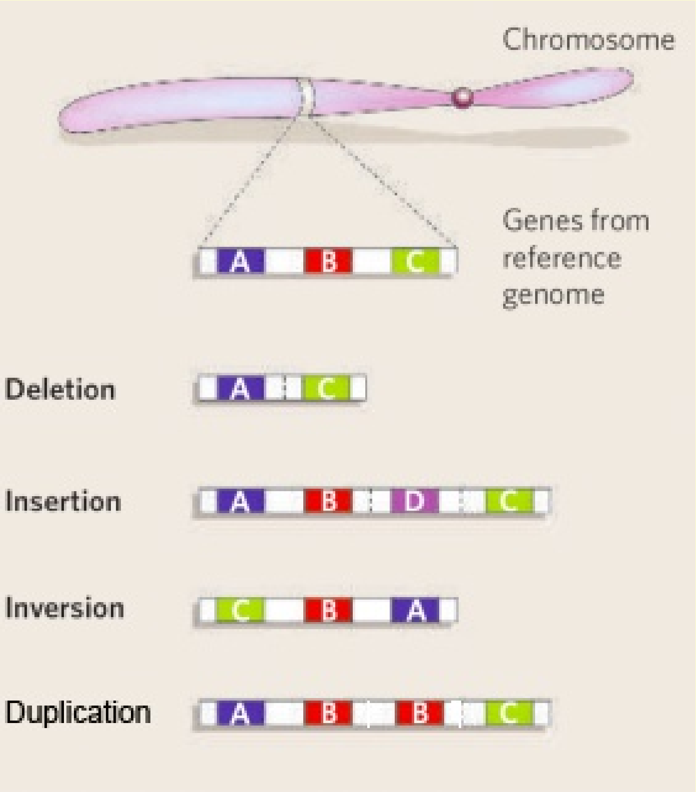


8 or 12 tandem repeats 4 kb long



Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Mapping Structural Variation in Humans >1 kb segments



- Structural Variations are Common 40% of the genome

-Structural Variations are involved in phenotype variation and disease

- Until recently most methods for detection were low resolution (>50 kb)

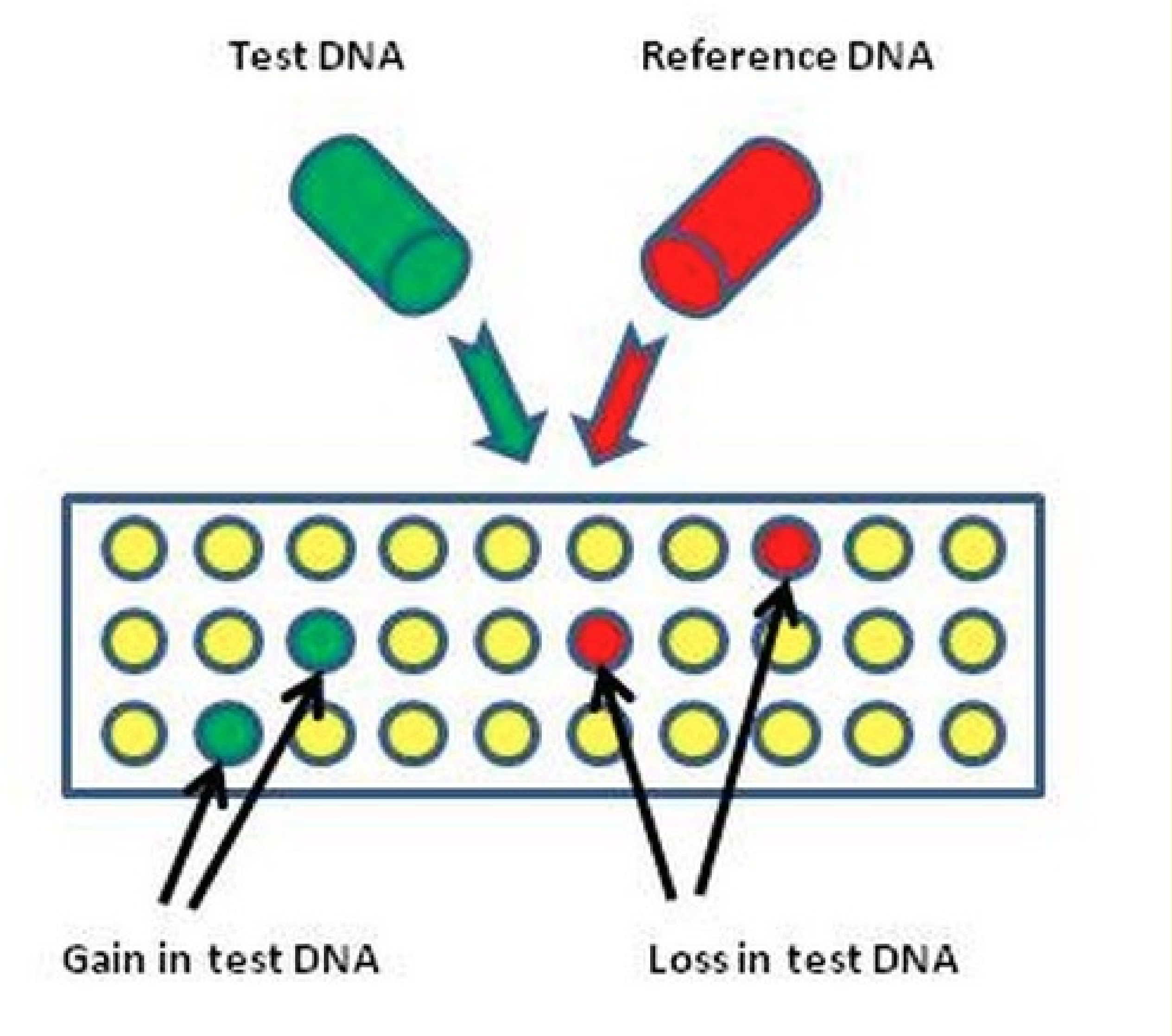
Courtesy of Mike Snyder

Why Study Structural Variation?

- They are common in “normal” human genomes and they are a major cause of phenotypic variation
- They are common in certain diseases, particularly cancers, behavioral disorders and neurodegenerative diseases
- They are now also showing up in rarer diseases and common behavioral disorders such as autism, schizophrenia, attention deficit, learning disabilities and many other neurological disorders

Courtesy of Mike Snyder

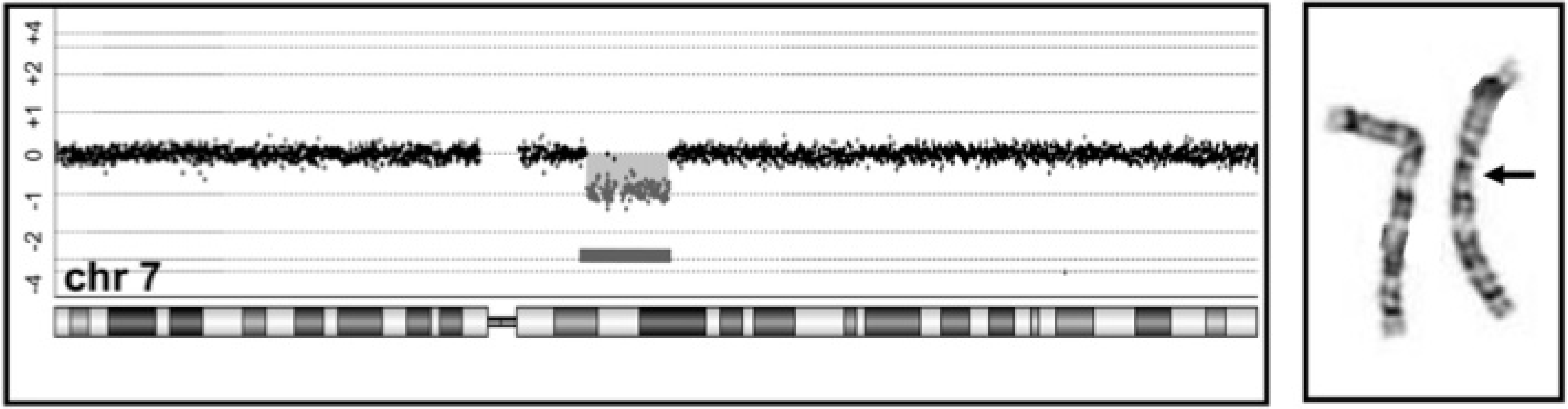
Comparative Genomics Hybridization (CGH)



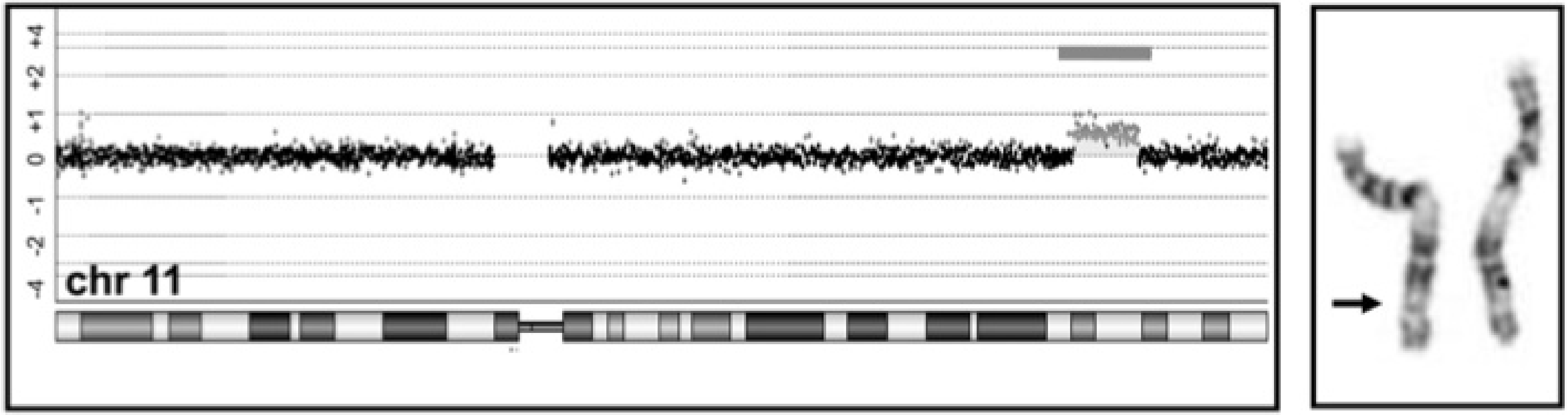
http://www.ncbi.nlm.nih.gov/dbvar/content/overview/

Detection of Duplications and Deletions Using Chromosomal Micro-Arrays

A 10.9 Mbase deletion at 7q11 in Williams-Beuren Syndrome

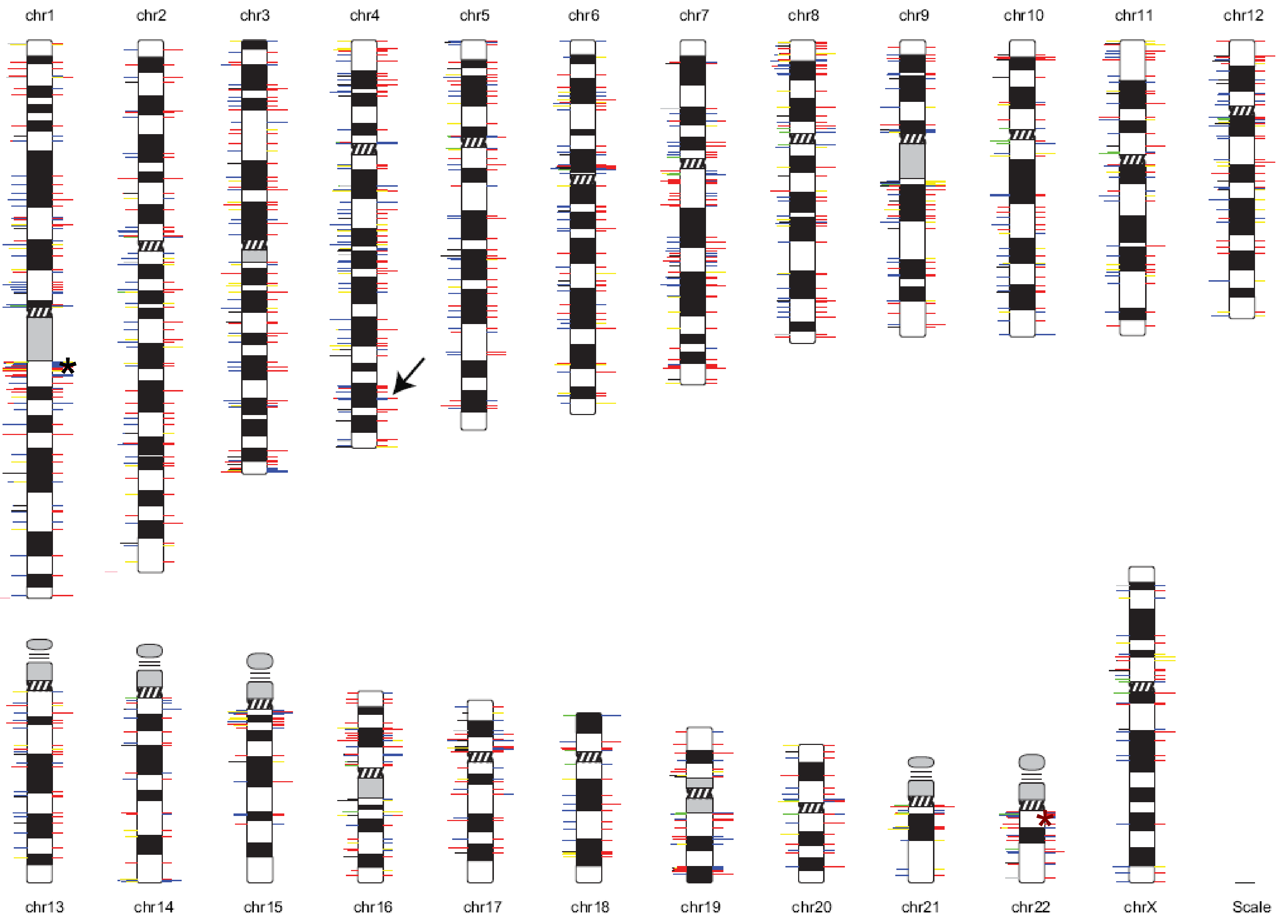


B 7.2 Mbase duplication in 11q



Miller et al., (2010) American Journal of Human Genetics 86, 749-764.

~1,000 Structural Variants



Courtesy of Mike Snyder

© Doug Brutlag 2016

Paired End Mapping

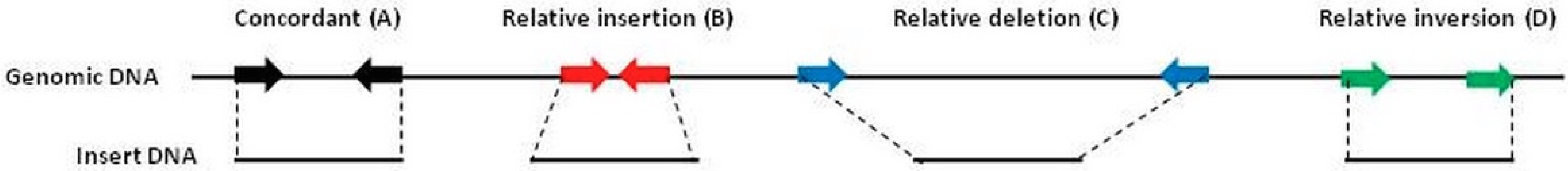
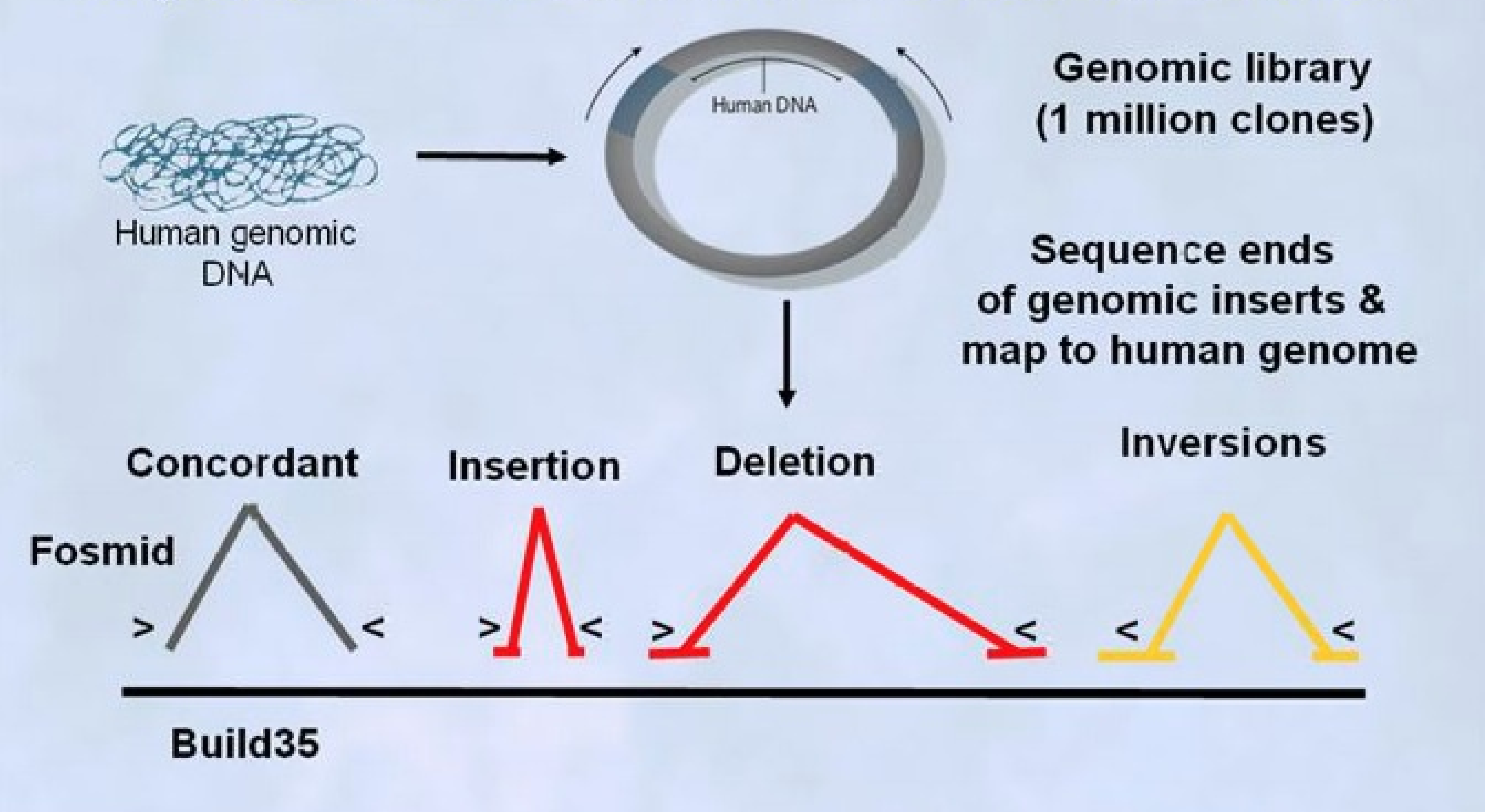


Figure 5: Paired-end mapping (PEM).

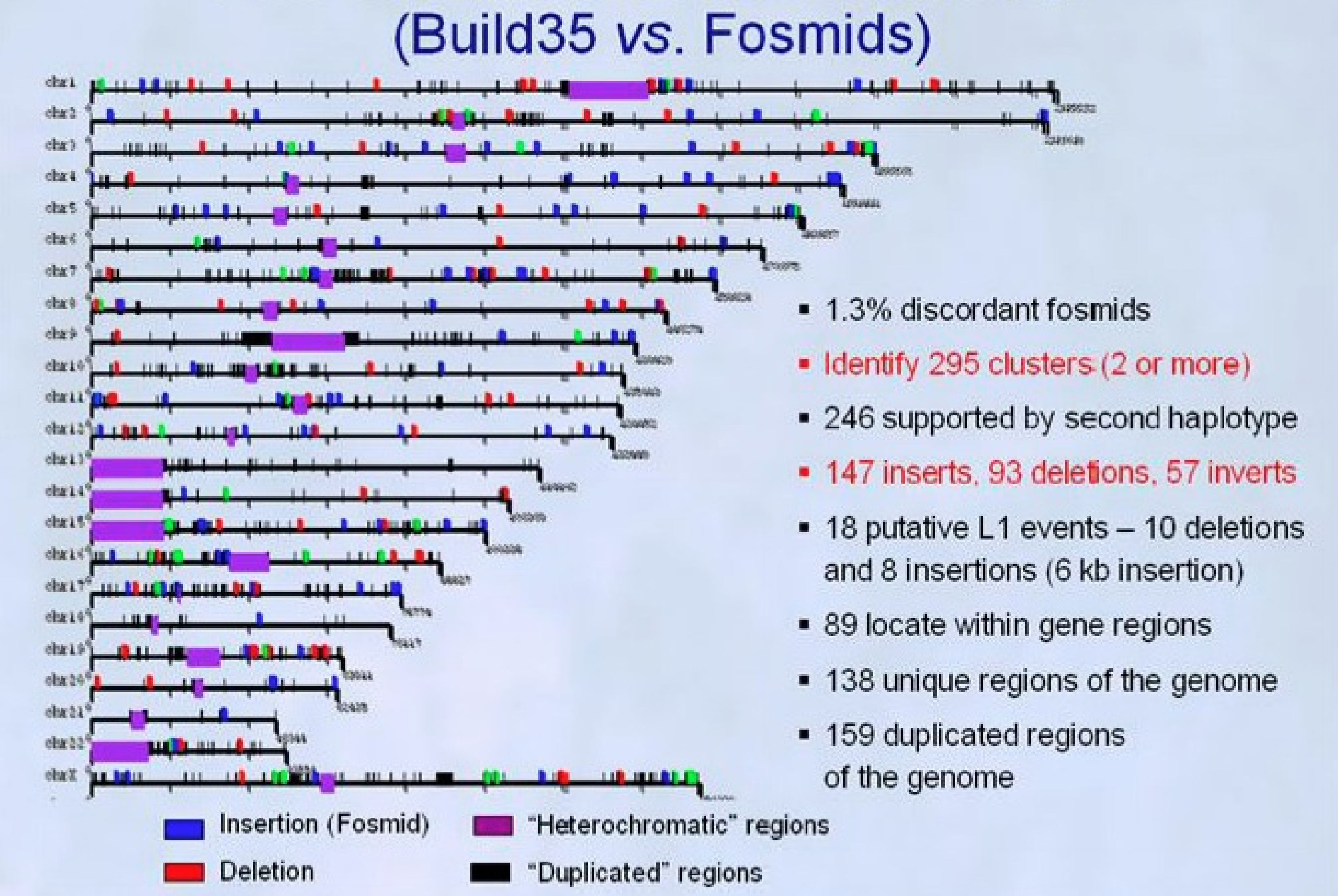
http://www.ncbi.nlm.nih.gov/dbvar/content/overview/

Sequence Base Resolution of Structural Variation



Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

Fine Scale Structural Variation in a Genome



(Build35 *vs.* Fosmids)

- 1.3% discordant fosmids
- Identify 295 clusters (2 or more)
- 246 supported by second haplotype
- 147 inserts, 93 deletions, 57 inverts
- 18 putative L1 events – 10 deletions and 8 insertions (6 kb insertion)
- 89 locate within gene regions
- 138 unique regions of the genome
- 159 duplicated regions of the genome

🟦 Insertion (Fosmid)

🟪 "Heterochromatic" regions

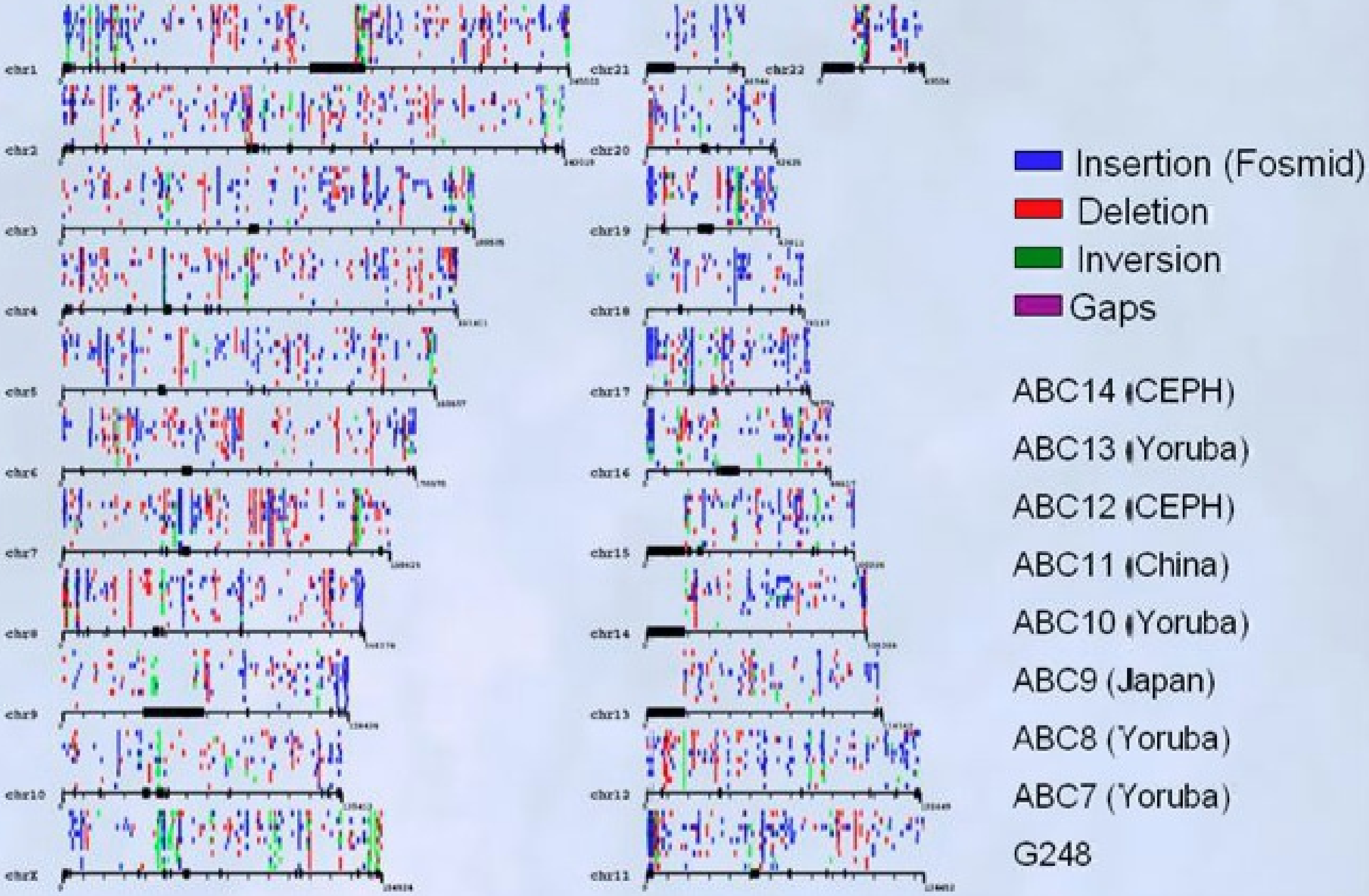
🟥 Deletion

⬛ "Duplicated" regions

Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

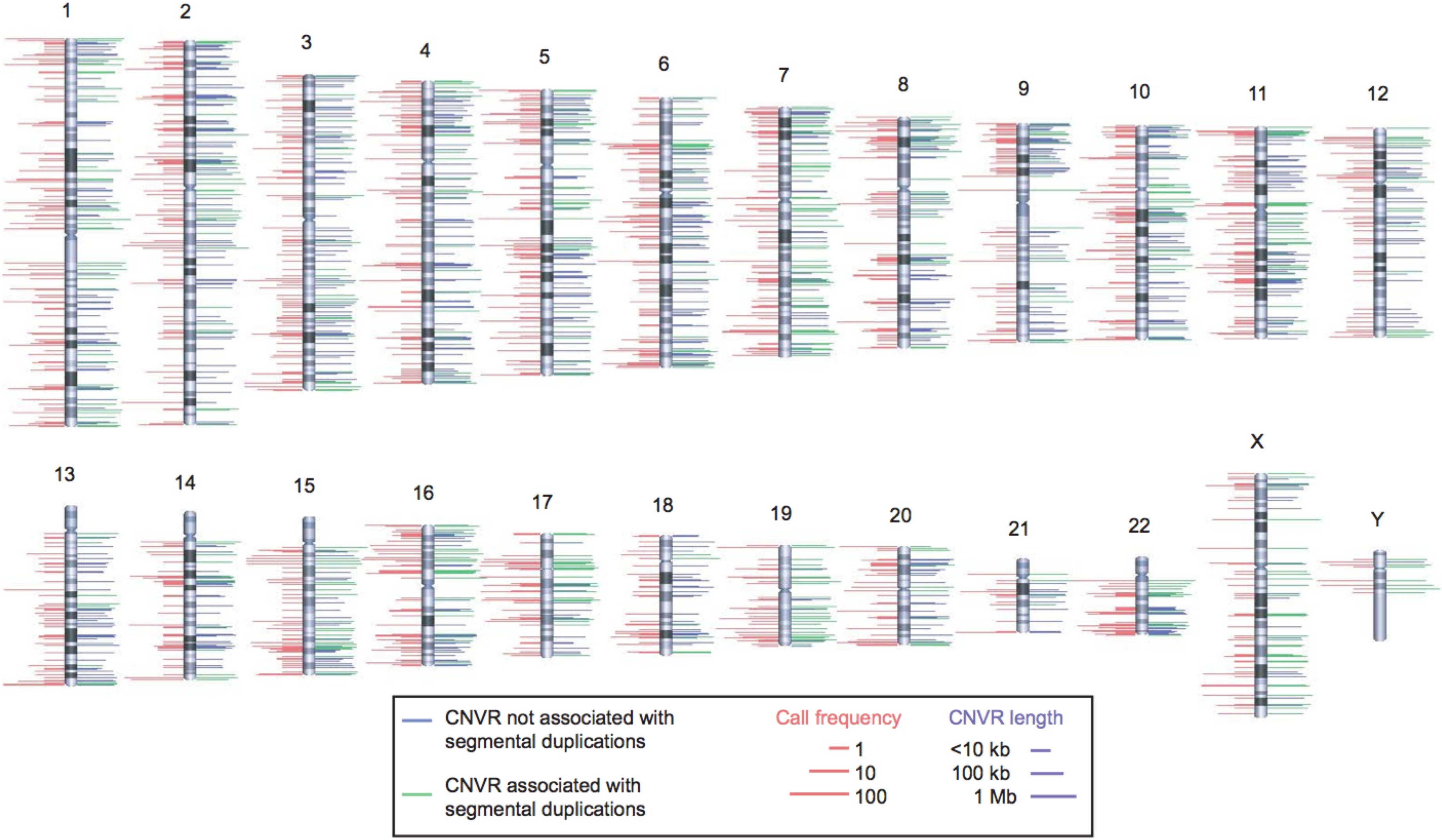
HENRY
STEWART
TALKS

Fine Scale Structural Variation for Eight Genomes



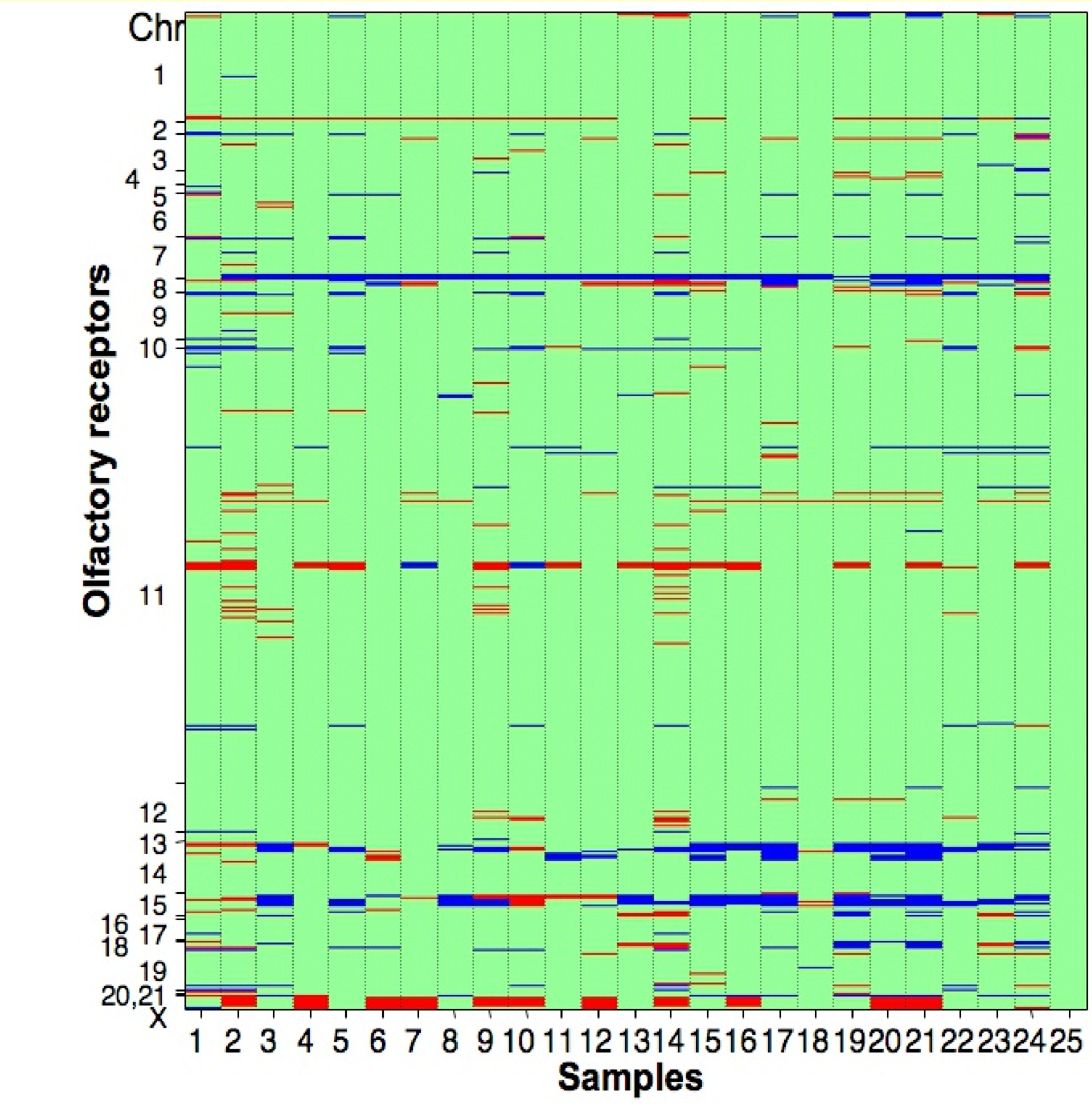
Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

Genomics Distribution of CNV Regions



Heterogeneity in Olfactory Receptor Genes

(Examined 851 Olfactory Receptor Loci)



🟥 Gain

🟦 Loss

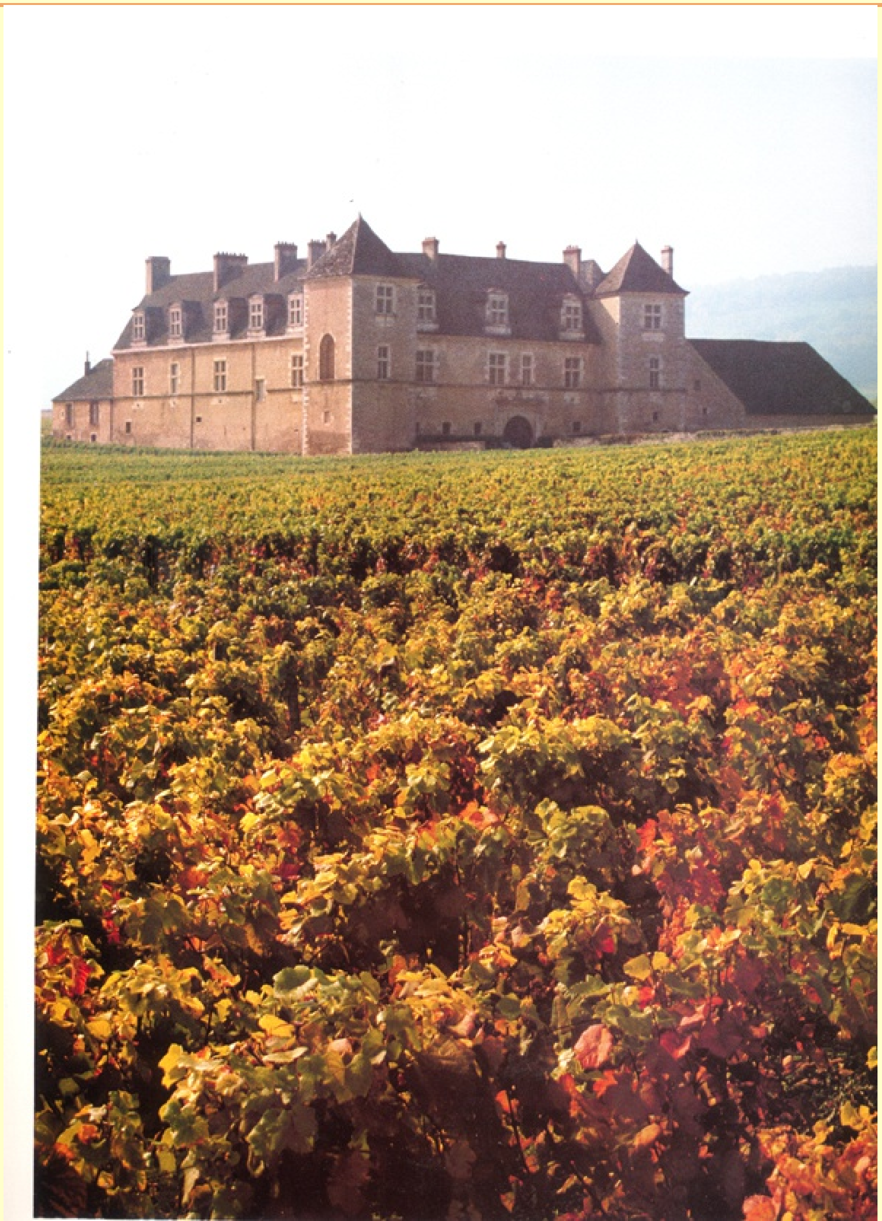
🟩 No change

Relative to the Reference Genome

CNVs affect:
93 duplicated genes
151 deleted genes

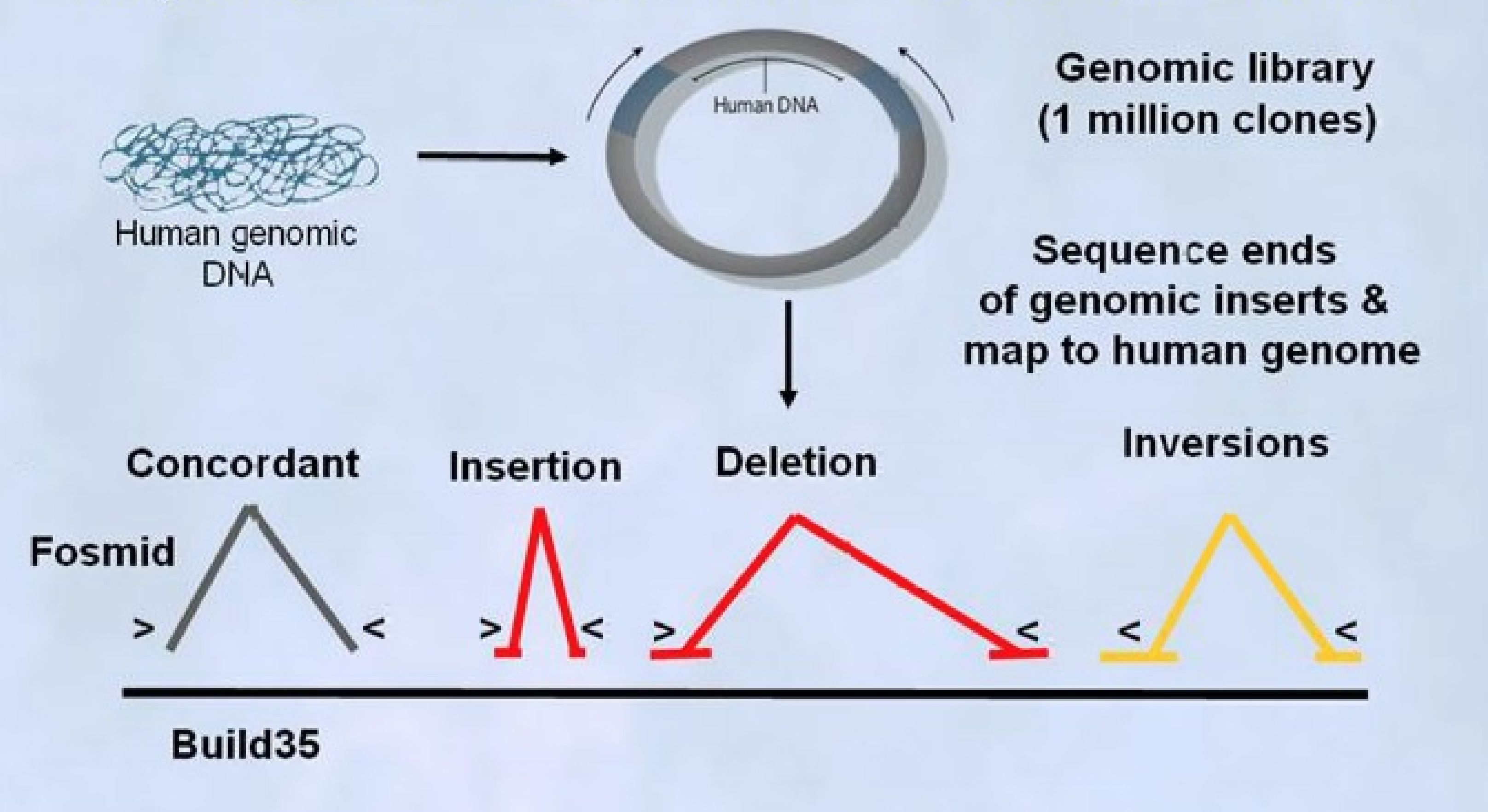
Courtesy of Mike Snyder Chair of Genetics

Clos Vougeot in Bourgogne



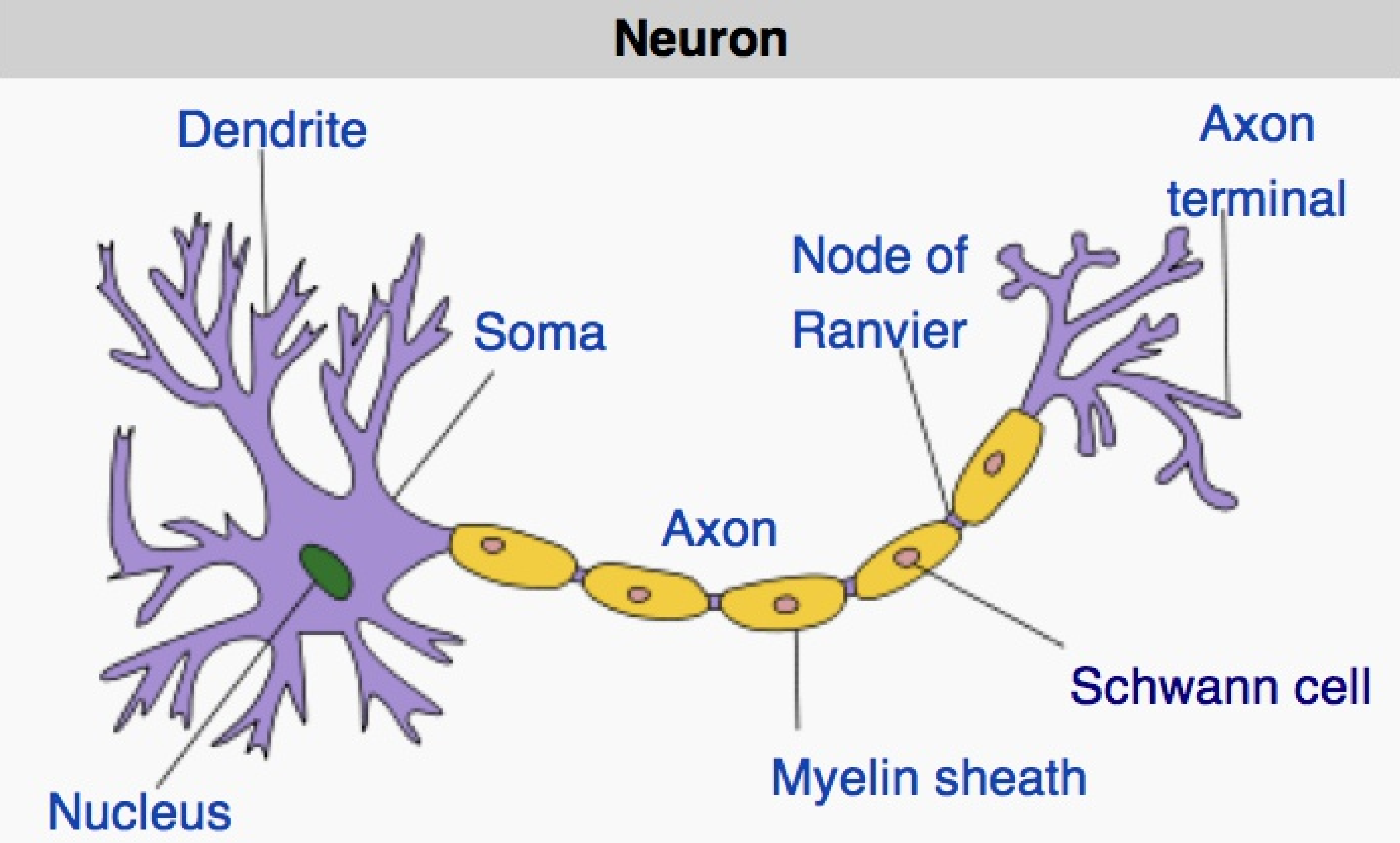
Chef d’Ordre de la Confrerie des Chevalier du Tastevins

Sequence Base Resolution of Structural Variation



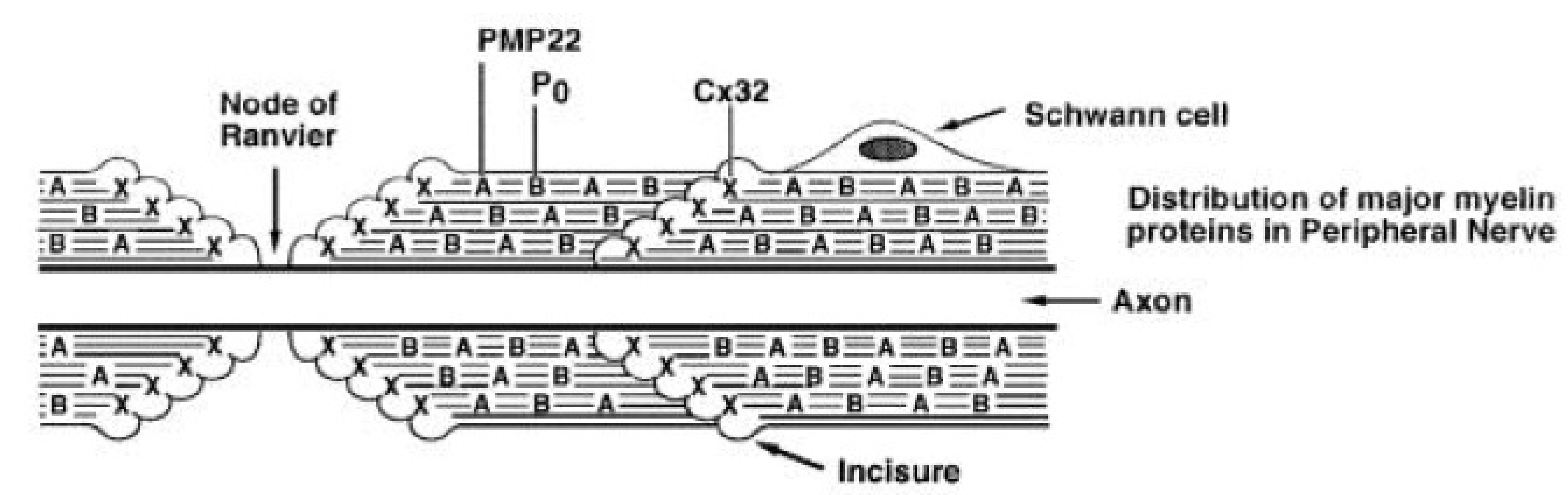
Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

Charcot Marie Tooth Hereditary Peripheral Neuropathy (CMT1) Caused by Abnormal Myelination of Long Axons



© Doug Brutlag 2016

Charcot Marie Tooth Hereditary Peripheral Neuropathy (CMT1) Caused by Abnormal Myelination of Long Axons



Distribution of major myelin proteins in Peripheral Nerve

Charcot-Marie Tooth Hereditary Neuropathy (CMT1) Disease Genes

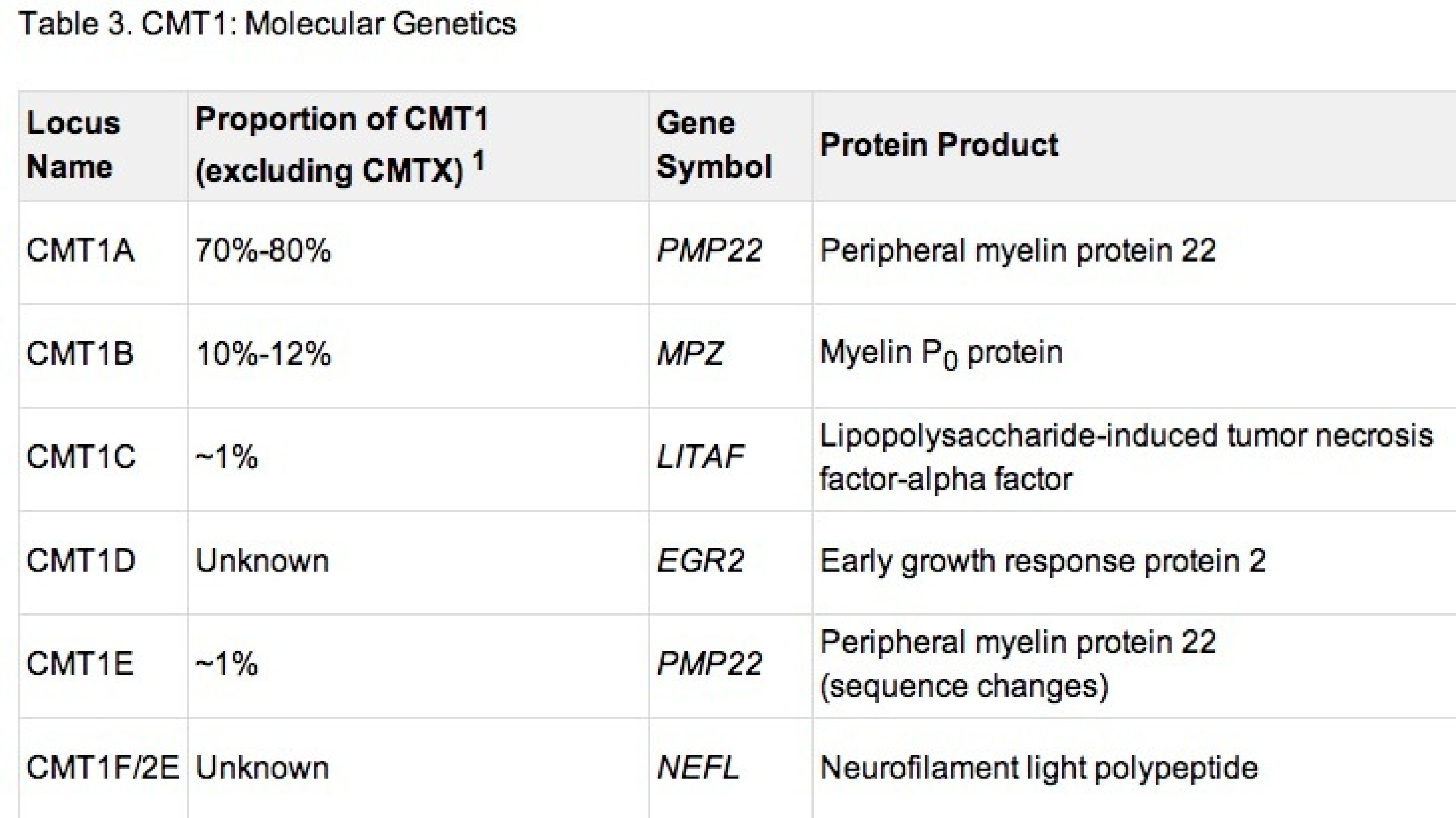
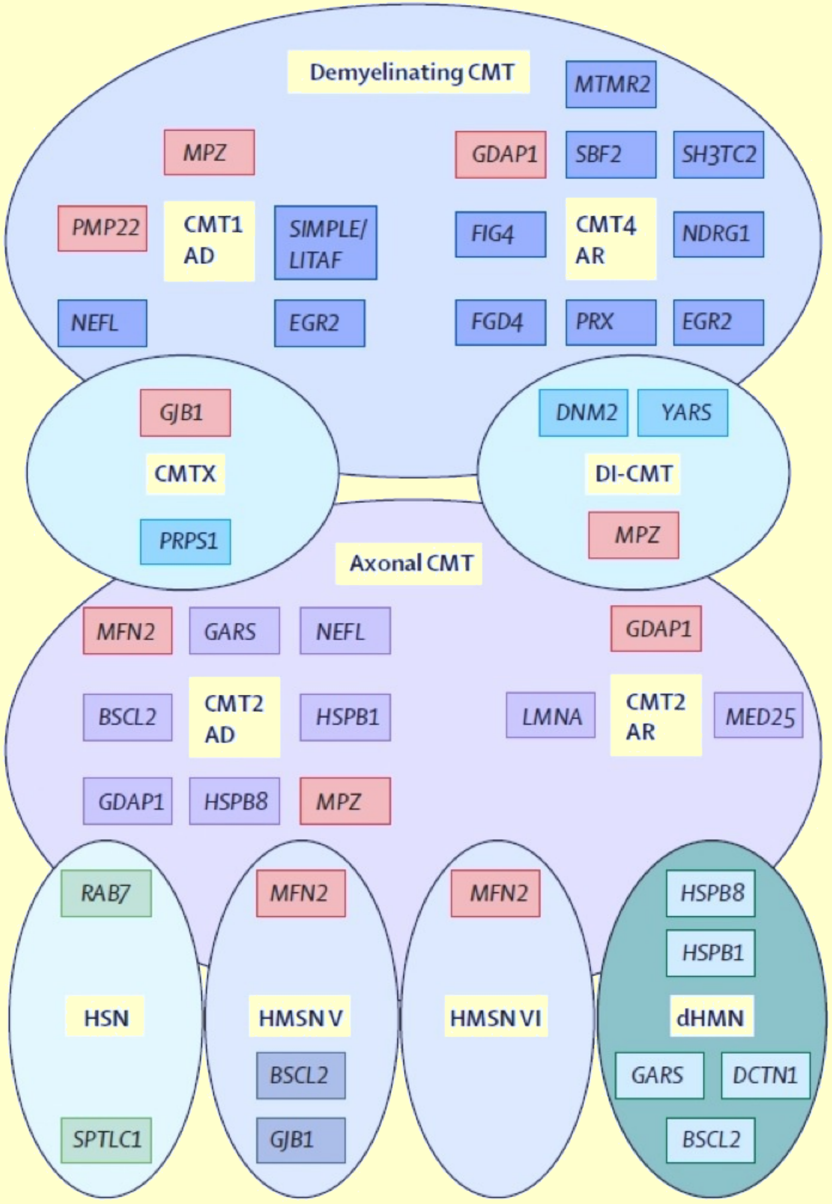


Table 3. CMT1: Molecular Genetics

Locus Name	Proportion of CMT1 (excluding CMTX) [1]	Gene Symbol	Protein Product
CMT1A	70%-80%	*PMP22*	Peripheral myelin protein 22
CMT1B	10%-12%	*MPZ*	Myelin P_0 protein
CMT1C	~1%	*LITAF*	Lipopolysaccharide-induced tumor necrosis factor-alpha factor
CMT1D	Unknown	*EGR2*	Early growth response protein 2
CMT1E	~1%	*PMP22*	Peripheral myelin protein 22 (sequence changes)
CMT1F/2E	Unknown	*NEFL*	Neurofilament light polypeptide

Hereditary Neuropathy Disease Genes

http://www.ncbi.nlm.nih.gov/books/NBK1358/



Schwann Cell

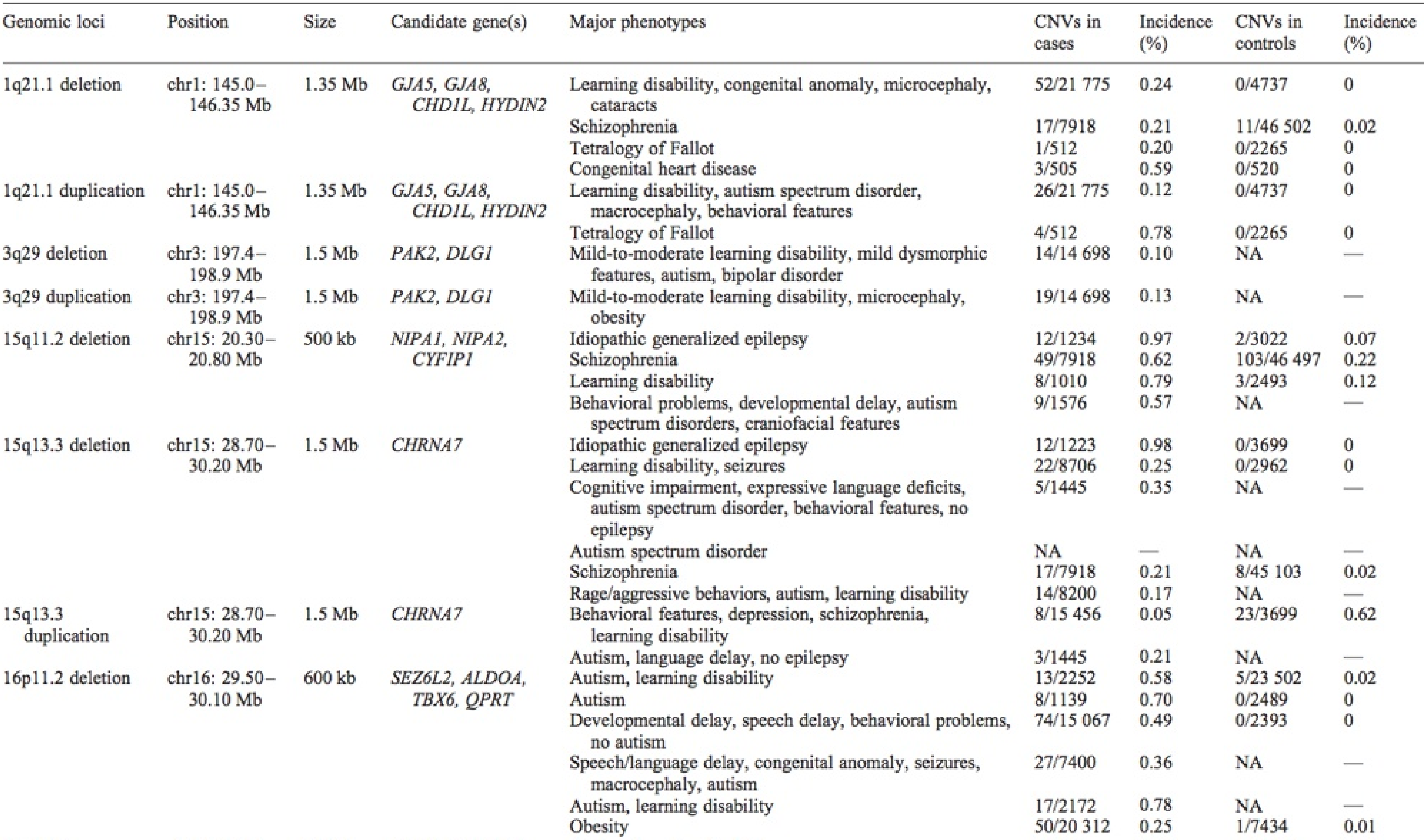
Attachment proteins

Axon proteins

Axon surface proteins

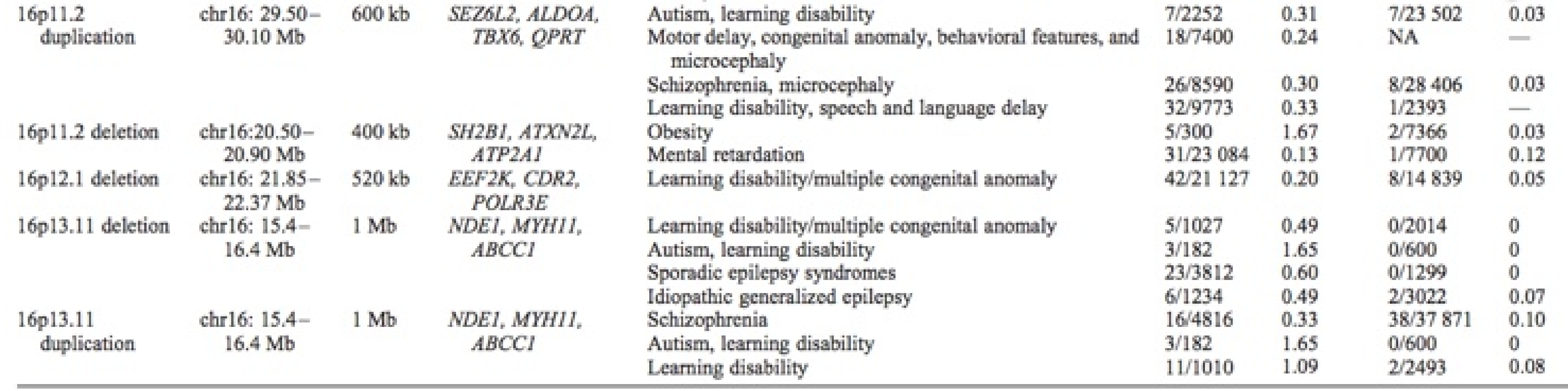
Gene Reviews
http://www.ncbi.nlm.nih.gov/books/NBK1358/

Clinical Features Associated with Structural Variations



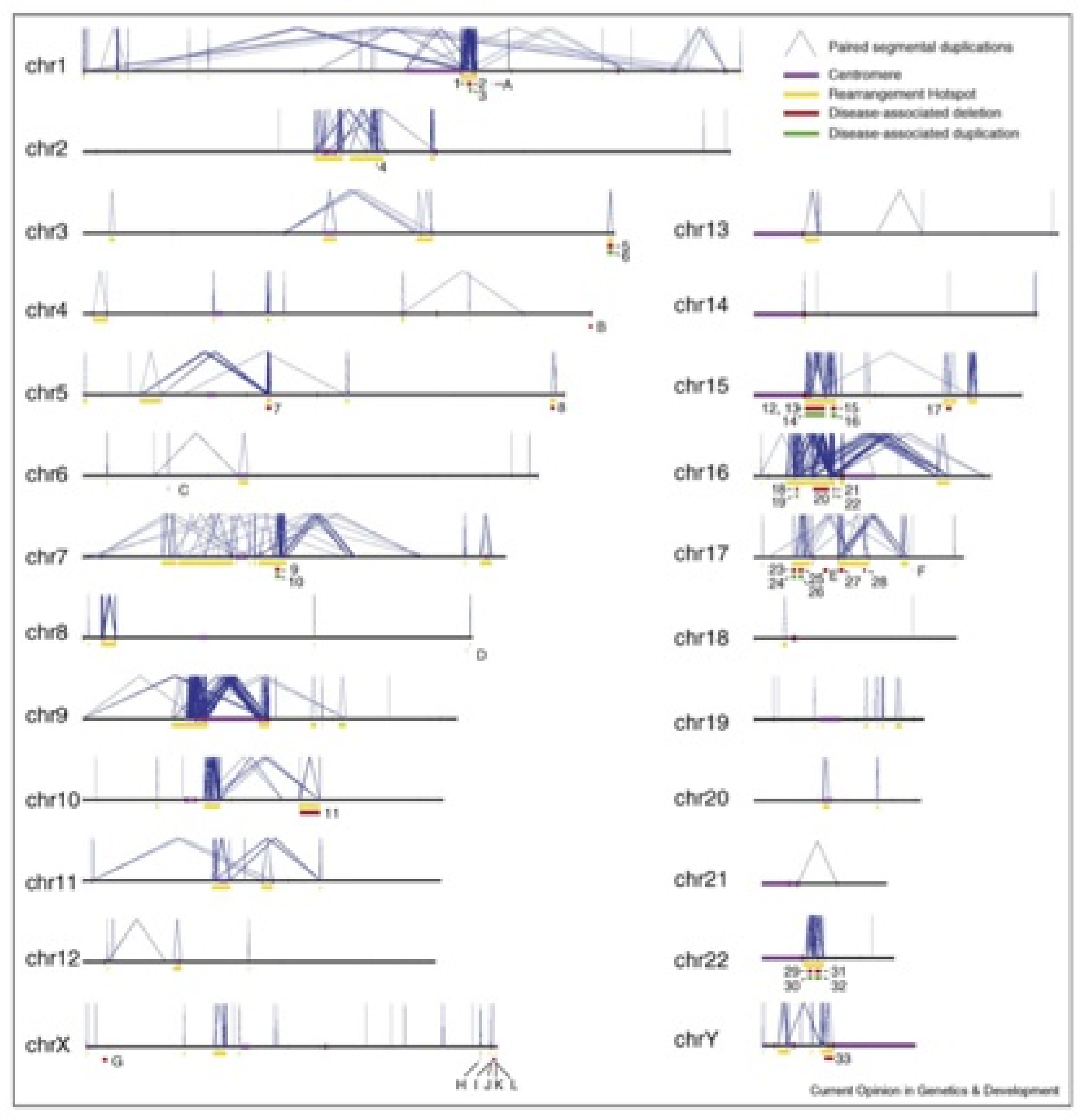
Genomic loci	Position	Size	Candidate gene(s)	Major phenotypes	CNVs in cases	Incidence (%)	CNVs in controls	Incidence (%)
1q21.1 deletion	chr1: 145.0–146.35 Mb	1.35 Mb	*GJA5, GJA8, CHD1L, HYDIN2*	Learning disability, congenital anomaly, microcephaly, cataracts	52/21 775	0.24	0/4737	0
				Schizophrenia	17/7918	0.21	11/46 502	0.02
				Tetralogy of Fallot	1/512	0.20	0/2265	0
				Congenital heart disease	3/505	0.59	0/520	0
1q21.1 duplication	chr1: 145.0–146.35 Mb	1.35 Mb	*GJA5, GJA8, CHD1L, HYDIN2*	Learning disability, autism spectrum disorder, macrocephaly, behavioral features	26/21 775	0.12	0/4737	0
				Tetralogy of Fallot	4/512	0.78	0/2265	0
3q29 deletion	chr3: 197.4–198.9 Mb	1.5 Mb	*PAK2, DLG1*	Mild-to-moderate learning disability, mild dysmorphic features, autism, bipolar disorder	14/14 698	0.10	NA	—
3q29 duplication	chr3: 197.4–198.9 Mb	1.5 Mb	*PAK2, DLG1*	Mild-to-moderate learning disability, microcephaly, obesity	19/14 698	0.13	NA	—
15q11.2 deletion	chr15: 20.30–20.80 Mb	500 kb	*NIPA1, NIPA2, CYFIP1*	Idiopathic generalized epilepsy	12/1234	0.97	2/3022	0.07
				Schizophrenia	49/7918	0.62	103/46 497	0.22
				Learning disability	8/1010	0.79	3/2493	0.12
				Behavioral problems, developmental delay, autism spectrum disorders, craniofacial features	9/1576	0.57	NA	—
15q13.3 deletion	chr15: 28.70–30.20 Mb	1.5 Mb	*CHRNA7*	Idiopathic generalized epilepsy	12/1223	0.98	0/3699	0
				Learning disability, seizures	22/8706	0.25	0/2962	0
				Cognitive impairment, expressive language deficits, autism spectrum disorder, behavioral features, no epilepsy	5/1445	0.35	NA	—
				Autism spectrum disorder	NA	—	NA	—
				Schizophrenia	17/7918	0.21	8/45 103	0.02
				Rage/aggressive behaviors, autism, learning disability	14/8200	0.17	NA	—
15q13.3 duplication	chr15: 28.70–30.20 Mb	1.5 Mb	*CHRNA7*	Behavioral features, depression, schizophrenia, learning disability	8/15 456	0.05	23/3699	0.62
				Autism, language delay, no epilepsy	3/1445	0.21	NA	—
16p11.2 deletion	chr16: 29.50–30.10 Mb	600 kb	*SEZ6L2, ALDOA, TBX6, QPRT*	Autism, learning disability	13/2252	0.58	5/23 502	0.02
				Autism	8/1139	0.70	0/2489	0
				Developmental delay, speech delay, behavioral problems, no autism	74/15 067	0.49	0/2393	0
				Speech/language delay, congenital anomaly, seizures, macrocephaly, autism	27/7400	0.36	NA	—
				Autism, learning disability	17/2172	0.78	NA	—
				Obesity	50/20 312	0.25	1/7434	0.01

Clinical Features Associated with Structural Variations (Continued)



16p11.2 duplication	chr16: 29.50–30.10 Mb	600 kb	*SEZ6L2, ALDOA, TBX6, QPRT*	Autism, learning disability	7/2252	0.31	7/23 502	0.03
				Motor delay, congenital anomaly, behavioral features, and microcephaly	18/7400	0.24	NA	—
				Schizophrenia, microcephaly	26/8590	0.30	8/28 406	0.03
				Learning disability, speech and language delay	32/9773	0.33	1/2393	—
16p11.2 deletion	chr16:20.50–20.90 Mb	400 kb	*SH2B1, ATXN2L, ATP2A1*	Obesity	5/300	1.67	2/7366	0.03
				Mental retardation	31/23 084	0.13	1/7700	0.12
16p12.1 deletion	chr16: 21.85–22.37 Mb	520 kb	*EEF2K, CDR2, POLR3E*	Learning disability/multiple congenital anomaly	42/21 127	0.20	8/14 839	0.05
16p13.11 deletion	chr16: 15.4–16.4 Mb	1 Mb	*NDE1, MYH11, ABCC1*	Learning disability/multiple congenital anomaly	5/1027	0.49	0/2014	0
				Autism, learning disability	3/182	1.65	0/600	0
				Sporadic epilepsy syndromes	23/3812	0.60	0/1299	0
				Idiopathic generalized epilepsy	6/1234	0.49	2/3022	0.07
16p13.11 duplication	chr16: 15.4–16.4 Mb	1 Mb	*NDE1, MYH11, ABCC1*	Schizophrenia	16/4816	0.33	38/37 871	0.10
				Autism, learning disability	3/182	1.65	0/600	0
				Learning disability	11/1010	1.09	2/2493	0.08

Rearrangement Hot Spots Associated with Disease



Current Opinion in Genetics & Development

© Doug Brutlag 2016

Structural Variations Resulting in Mendelian Inherited Disease

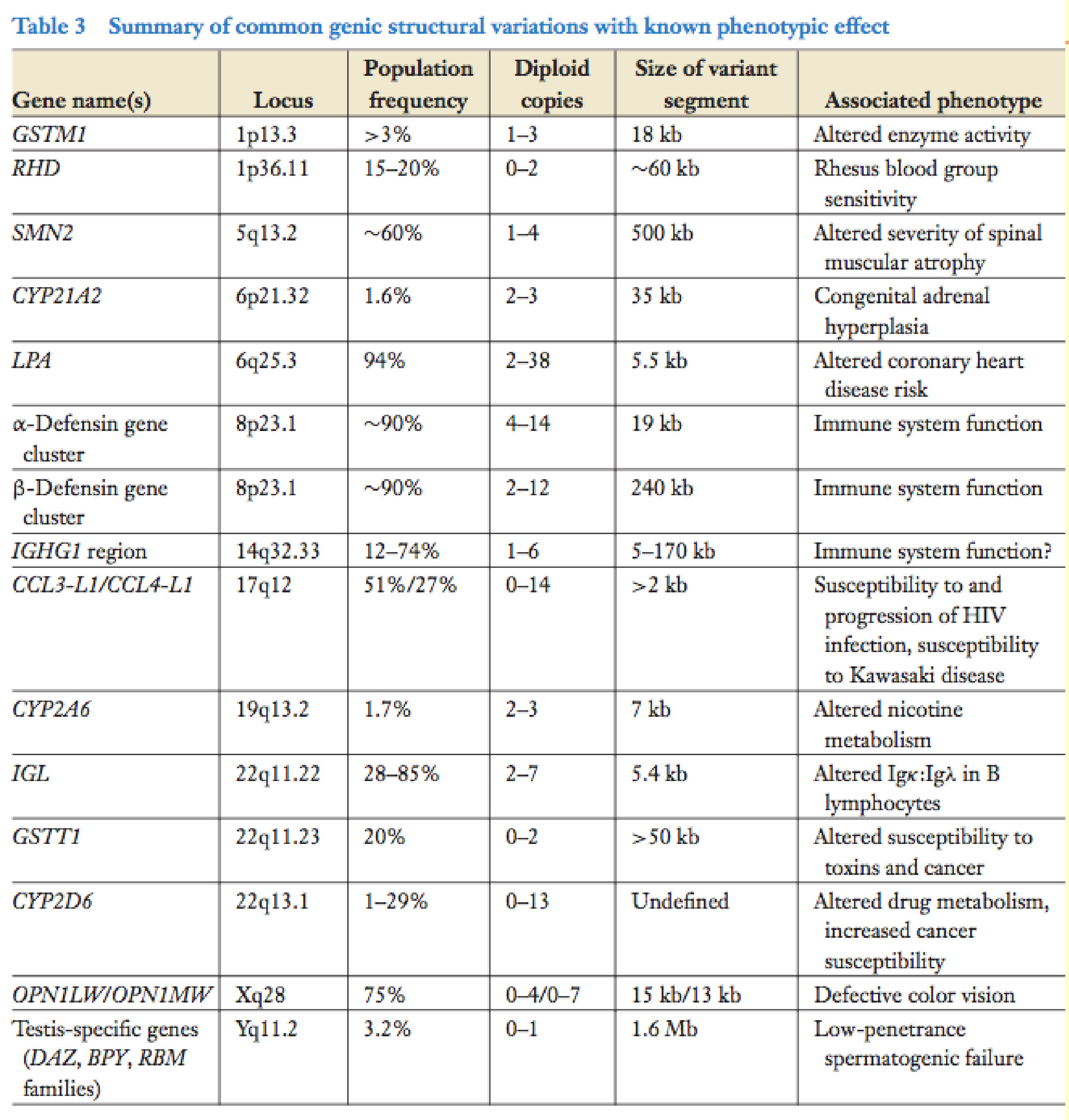


Table 3 Summary of common genic structural variations with known phenotypic effect

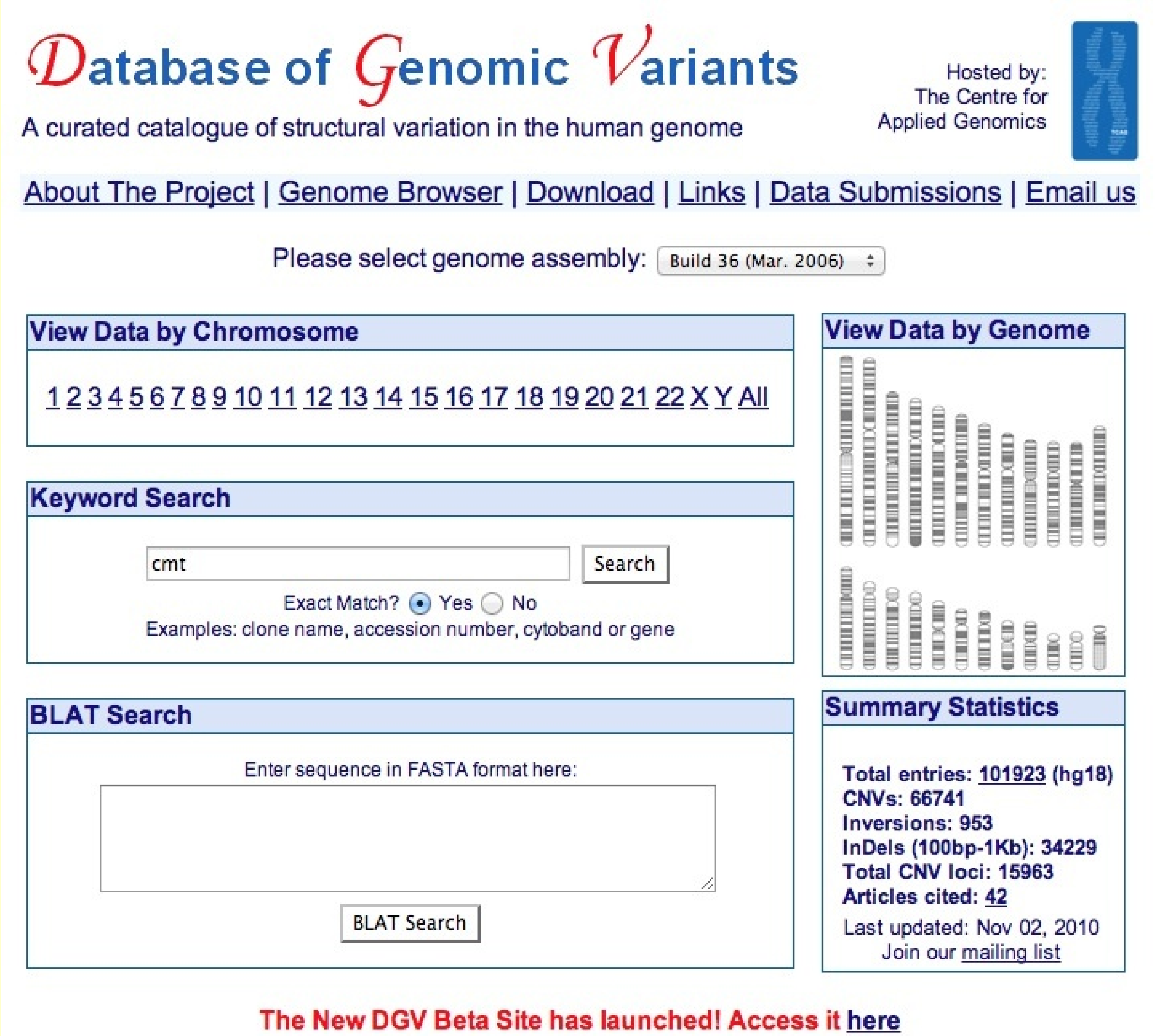
Gene name(s)	Locus	Population frequency	Diploid copies	Size of variant segment	Associated phenotype
GSTM1	1p13.3	>3%	1–3	18 kb	Altered enzyme activity
RHD	1p36.11	15–20%	0–2	~60 kb	Rhesus blood group sensitivity
SMN2	5q13.2	~60%	1–4	500 kb	Altered severity of spinal muscular atrophy
CYP21A2	6p21.32	1.6%	2–3	35 kb	Congenital adrenal hyperplasia
LPA	6q25.3	94%	2–38	5.5 kb	Altered coronary heart disease risk
α-Defensin gene cluster	8p23.1	~90%	4–14	19 kb	Immune system function
β-Defensin gene cluster	8p23.1	~90%	2–12	240 kb	Immune system function
IGHG1 region	14q32.33	12–74%	1–6	5–170 kb	Immune system function?
CCL3-L1/CCL4-L1	17q12	51%/27%	0–14	>2 kb	Susceptibility to and progression of HIV infection, susceptibility to Kawasaki disease
CYP2A6	19q13.2	1.7%	2–3	7 kb	Altered nicotine metabolism
IGL	22q11.22	28–85%	2–7	5.4 kb	Altered Igκ:Igλ in B lymphocytes
GSTT1	22q11.23	20%	0–2	>50 kb	Altered susceptibility to toxins and cancer
CYP2D6	22q13.1	1–29%	0–13	Undefined	Altered drug metabolism, increased cancer susceptibility
OPN1LW/OPN1MW	Xq28	75%	0–4/0–7	15 kb/13 kb	Defective color vision
Testis-specific genes (*DAZ, BPY, RBM* families)	Yq11.2	3.2%	0–1	1.6 Mb	Low-penetrance spermatogenic failure

Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

http://projects.tcag.ca/variation/



Database of Genomic Variants

A curated catalogue of structural variation in the human genome

Hosted by:
The Centre for
Applied Genomics

Please select genome assembly: Build 36 (Mar. 2006)

View Data by Chromosome

View Data by Genome

Keyword Search

cmt

Exact Match? ● Yes ○ No

Examples: clone name, accession number, cytoband or gene

BLAT Search

Enter sequence in FASTA format here:

Summary Statistics

Total entries: 101923 (hg18)
CNVs: 66741
Inversions: 953
InDels (100bp-1Kb): 34229
Total CNV loci: 15963
Articles cited: 42

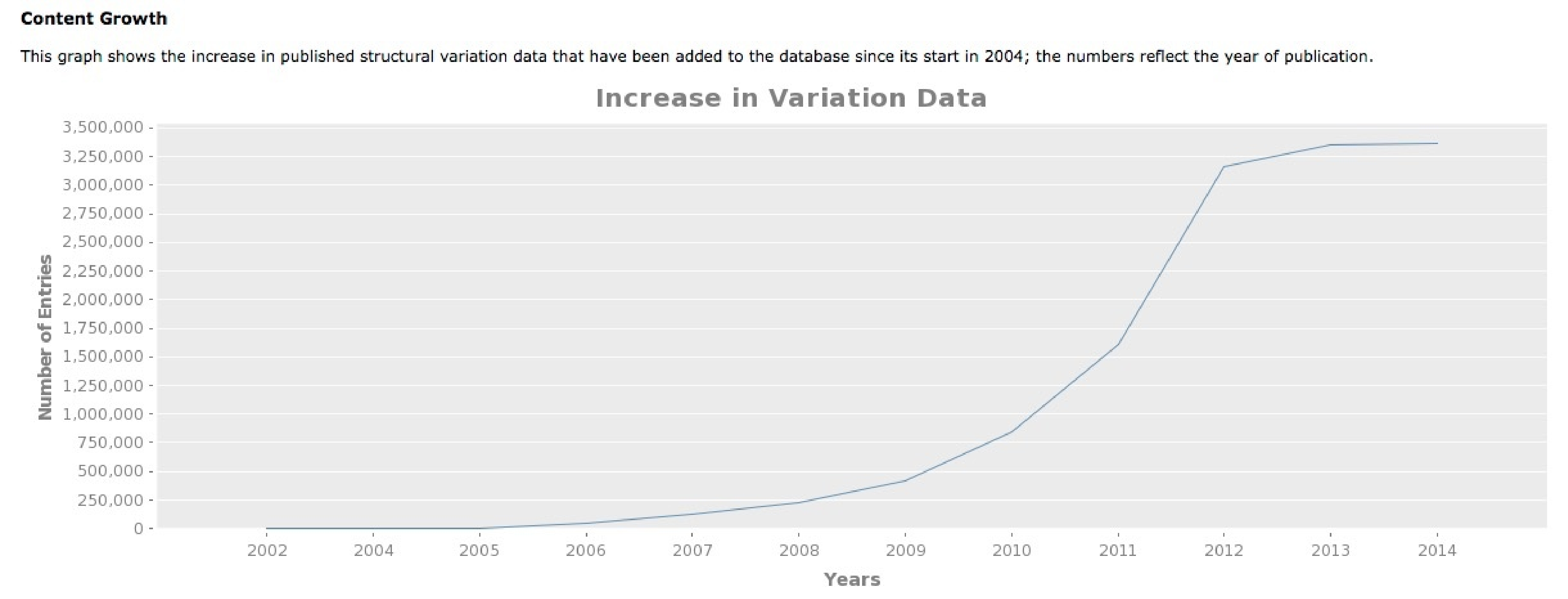
Last updated: Nov 02, 2010
Join our mailing list

The New DGV Beta Site has launched! Access it here

Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

http://projects.tcag.ca/variation/



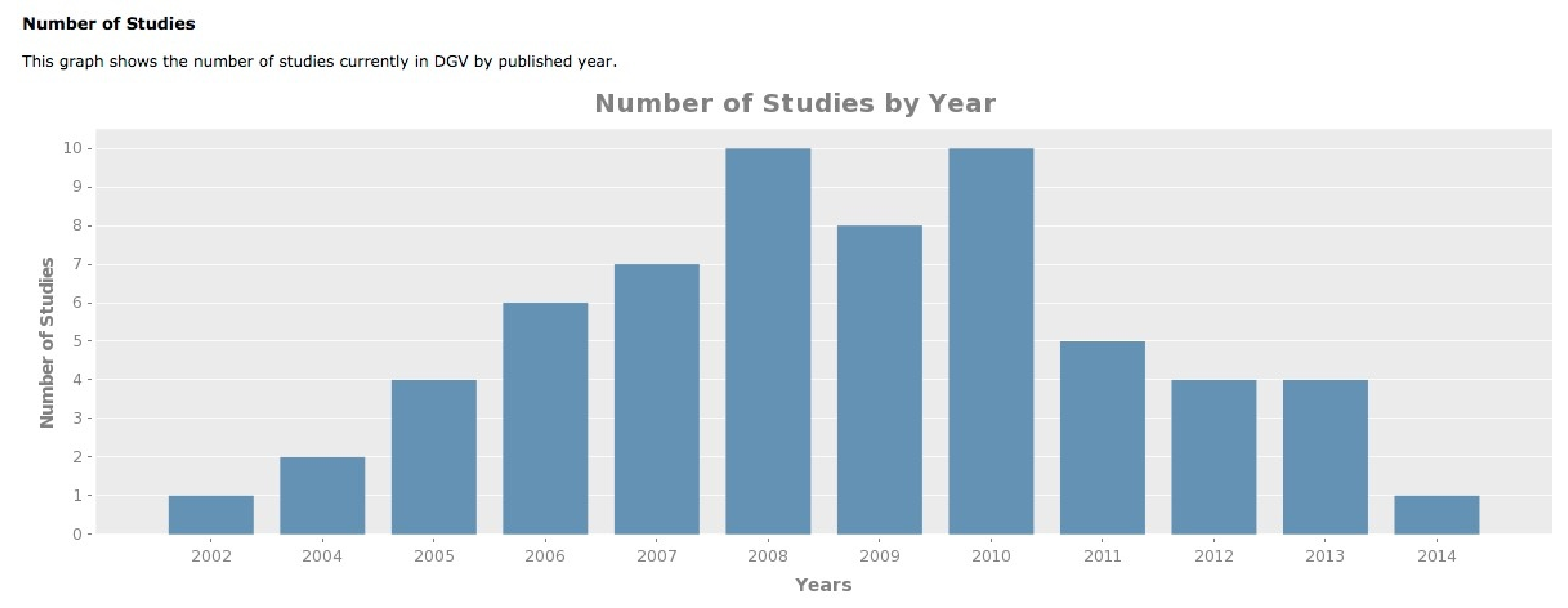
Content Growth

This graph shows the increase in published structural variation data that have been added to the database since its start in 2004; the numbers reflect the year of publication.

Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

http://projects.tcag.ca/variation/



Number of Studies

This graph shows the number of studies currently in DGV by published year.

Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

http://projects.tcag.ca/variation/