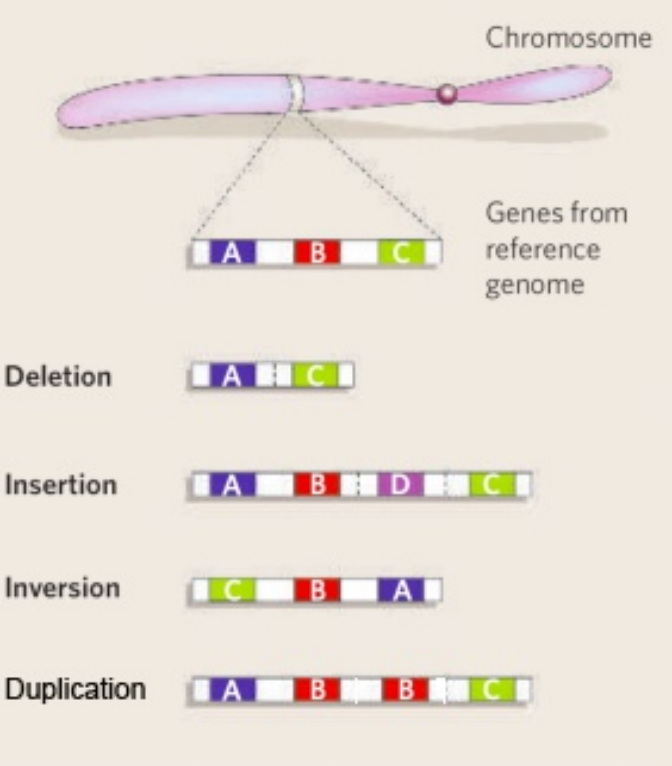
Genomics, Bioinformatics & Medicine

http://biochem158.stanford.edu/

Structural Variants in the Human Genome

http://biochem158.stanford.edu/Structural%20%20Variants.html



Doug Brutlag
Professor Emeritus of Biochemistry & Medicine
Stanford University School of Medicine

NIH Precision Medicine Initiative

http://www.nih.gov/precisionmedicine/

National Institutes of Health
Turning Discovery Into Health

Search

For Employees | Staff Directory | En Español

Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH

NIH Home > Research & Training

PRECISION MEDICINE INITIATIVE

Precision Medicine Initiative

- What are the near-term goals?
- What are the longer-term goals?
- How is it different?
- Who will participate?
- NIH Workshop

Precision Medicine Initiative

Far too many diseases do not have a proven means of prevention or effective treatments. We must gain better insights into the biology of these diseases to make a difference for the millions of Americans who suffer from them. Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While significant advances in precision medicine have been made for select cancers, the practice is not currently in use for most diseases. Many efforts are underway to help make precision medicine the norm rather than the exception. To accelerate the pace, President Obama has now unveiled the Precision Medicine Initiative – a bold new enterprise to revolutionize medicine and generate the scientific evidence needed to move the concept of precision medicine into every day clinical practice.

Email Updates

To sign up for updates please enter your e-mail address.

Submit

Related Links

- NEJM Perspective: A New Initiative on Precision Medicine
- White House Precision Medicine Web Page
- White House Fact Sheet: President Obama's Precision Medicine Initiative
- Precision Medicine Initiative and Cancer Research

Genomics England

http://www.genomicsengland.co.uk/

Genomics england

Home | About Genomics England | The 100,000 Genomes Project | GeCIP | Library and resources | News | Contact us

Genomics England, with the consent of participants and the support of the public, is creating a lasting legacy for patients, the NHS and the UK economy through the sequencing of 100,000 genomes: the 100,000 Genomes Project.

Genomics England was set up by the Department of Health to deliver the 100,000 Genomes Project. Initially the focus will be on rare disease, cancer and infectious disease. The project is currently in its pilot phase and will be completed by the end of 2017.

Read more...

Duplications and Deletions in the Human Genome

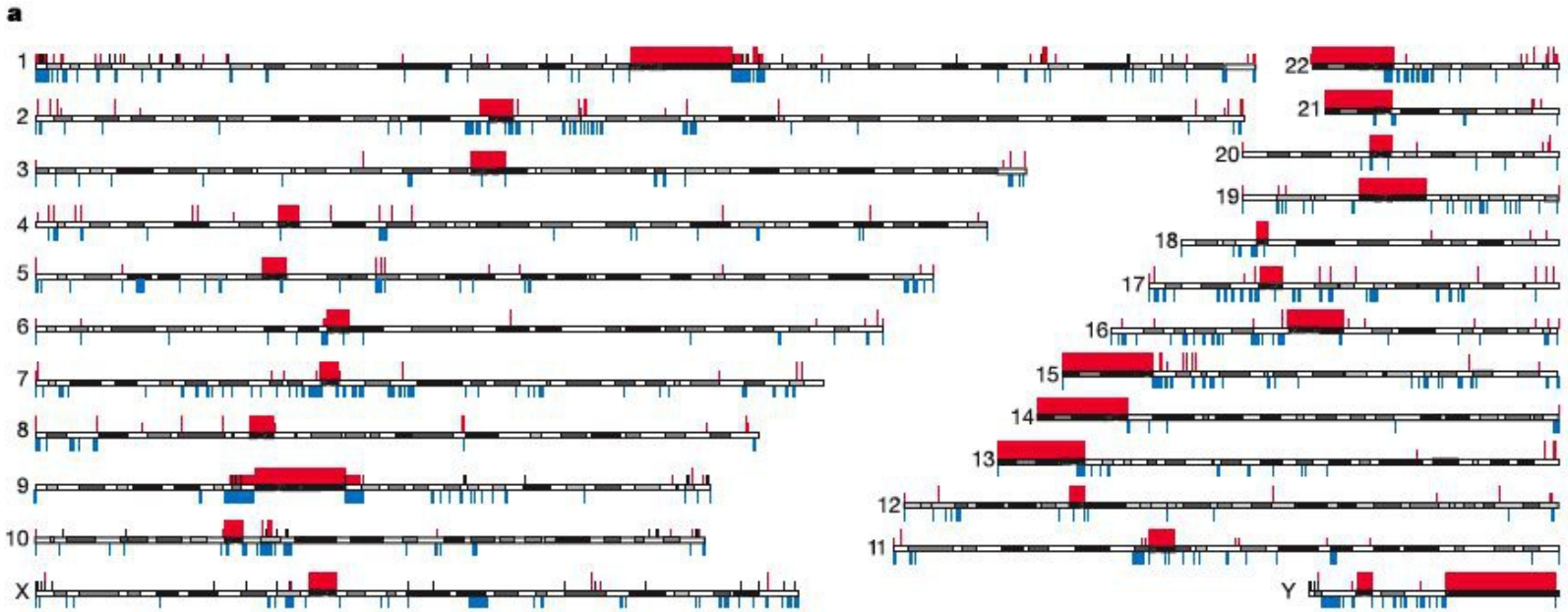
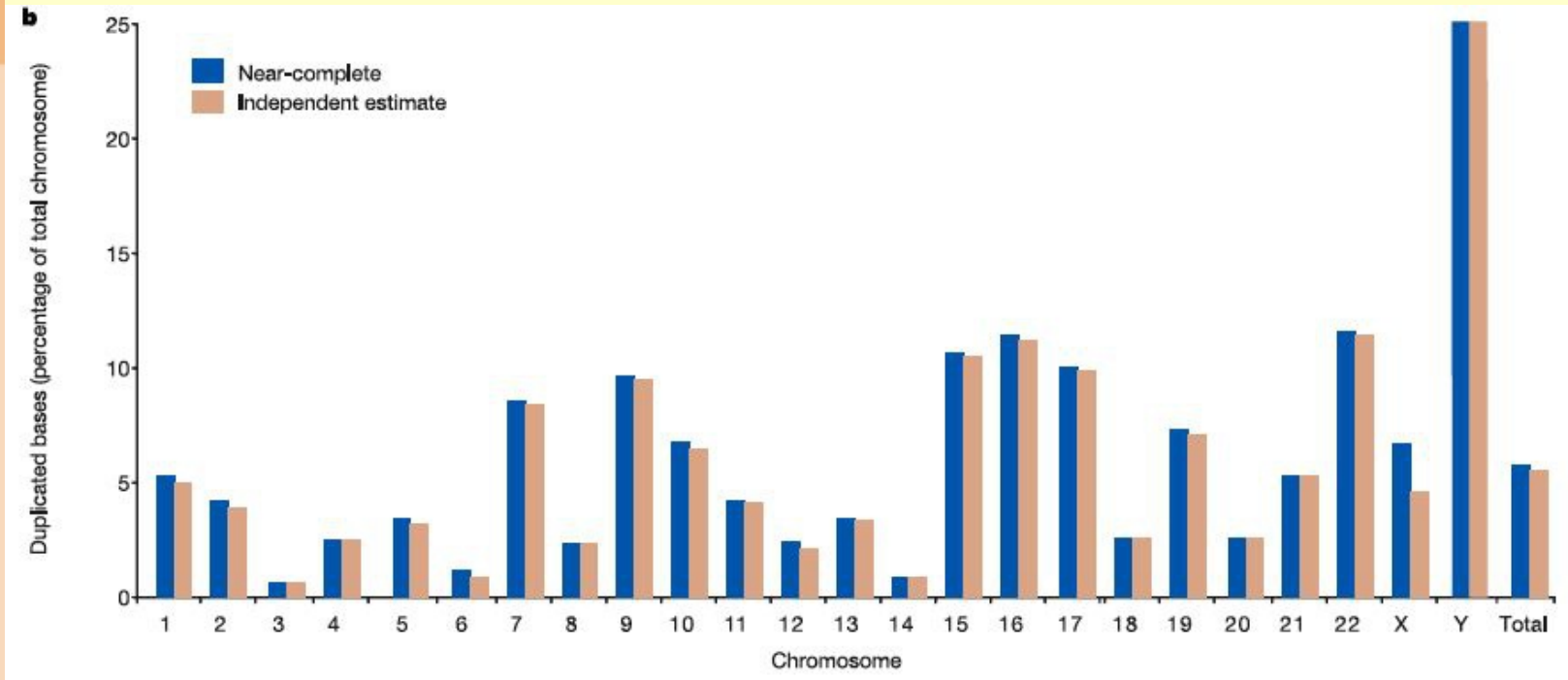


Figure 4 Segmental duplications across the genome. **a**, Segmental duplications and sequence gaps across the genome. Segmental duplications are indicated below the chromosomes in blue (length ≥10 kb and sequence identity ≥95%). Large duplications are shown to approximate scale; smaller ones are indicated 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, Annual. Rev. Genomics Hum. Genet. 2006. 7:407–42

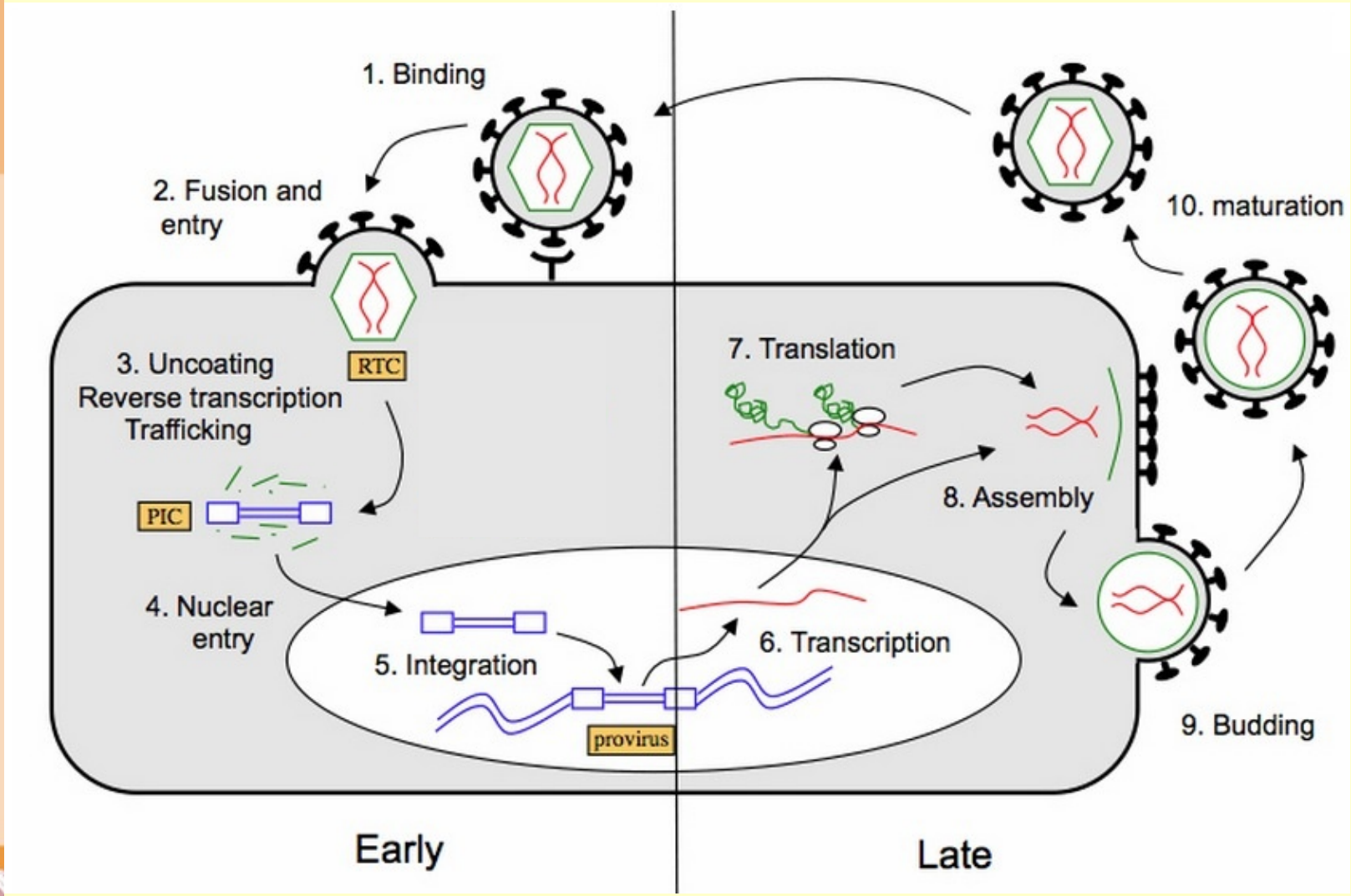
Repeated Elements in the Human Genome ERVs, LINES, SINES and ALUs

- 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 (example HIV)

Weiner Curr Opin Cell Biol, 2002 14 (3) 343-50

MRC National Institute for Medical Research

Retroviral Life Cycle



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

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

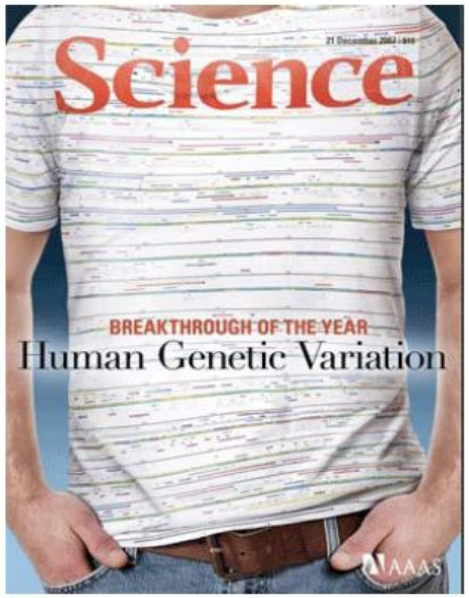
- ERVs-Endogenous Retroviruses
  - 10,000 base long RNA genome
  - Converted to DNA and integrate into genome with help of RNA reverse transcriptase and integrase enzymes and long tandem repeats (LTRs)
  - Transcribed into RNA and produce virus (HIV)
- LINES-Long Interspersed Nuclear Elements
  - About 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
  - 350 base pairs in length
  - Contain an RNA Polymerase III promoter, Alu site
  - Appear to evolve from 7S RNA signal recognition particle

Weiner Curr Opin Cell Biol, 2002 14 (3) 343-50

Human Genetic Variation
2007 Scientific Breakthrough of the Year

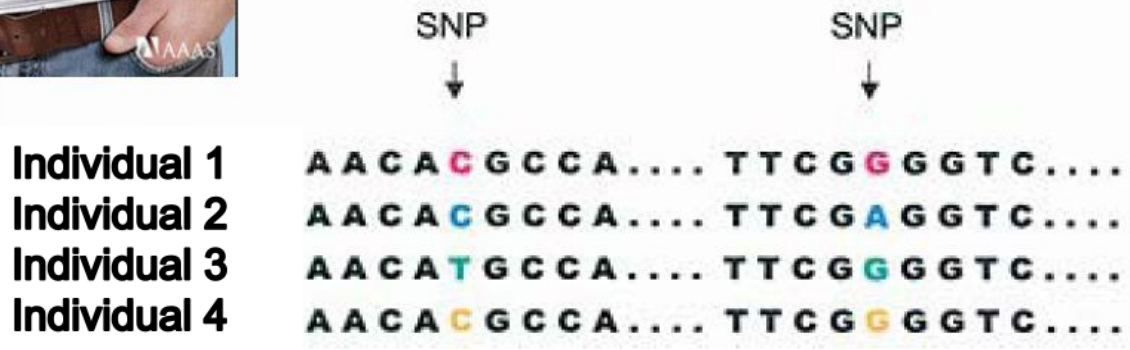
2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007



"It's all about me!"

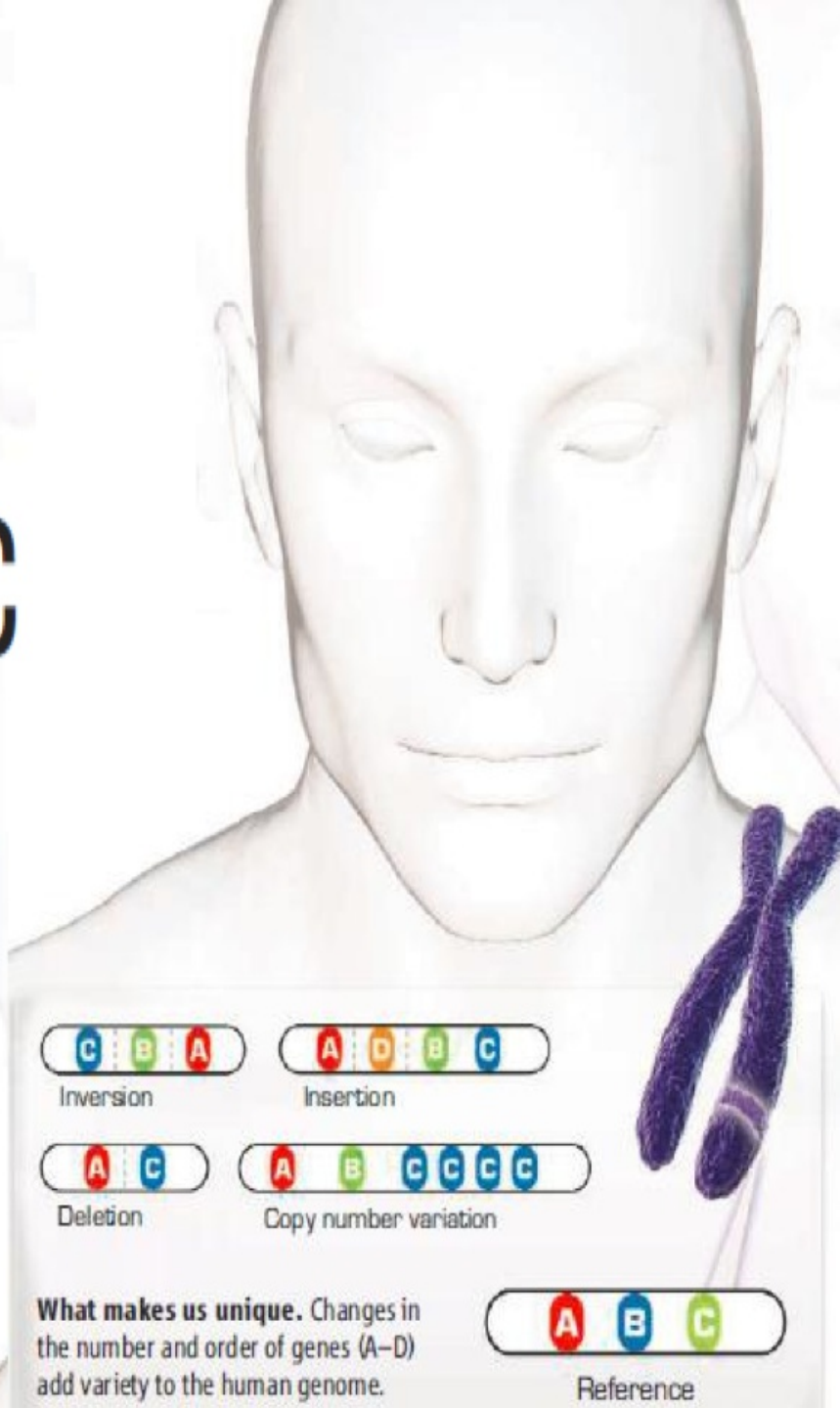
Simple Nucleotide Polymorphisms (SNPs)



BREAKTHROUGH OF THE YEAR

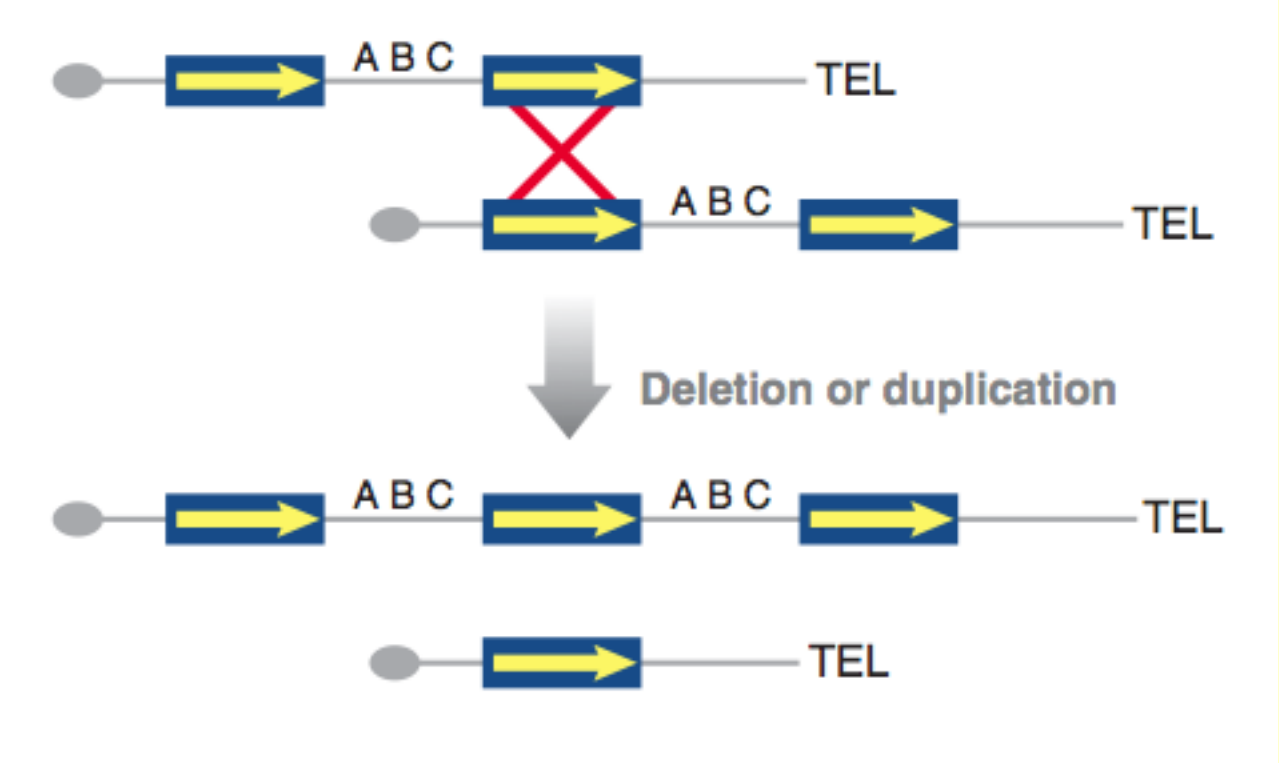
Human Genetic Variation

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



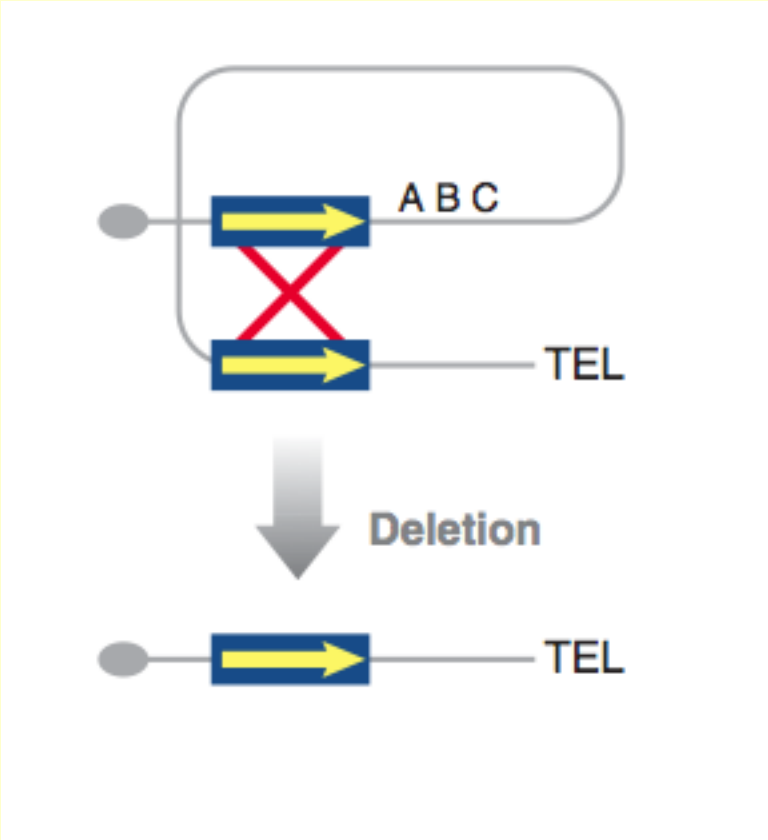
What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

Unequal Crossing Over Leads to Duplication and Deletion



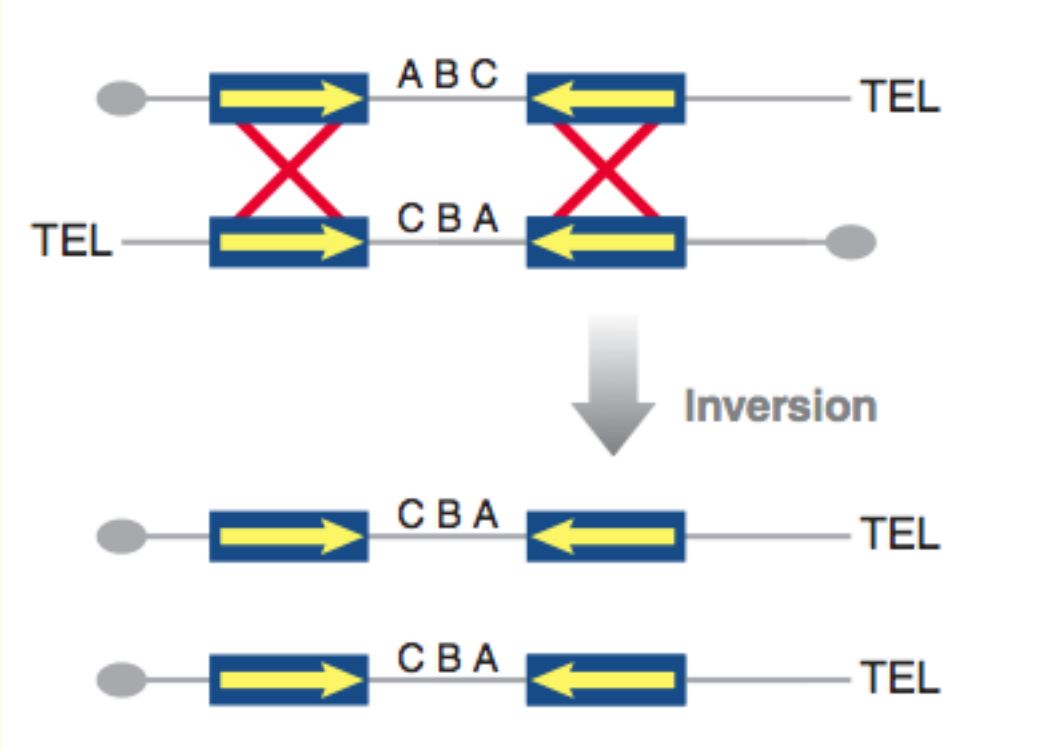
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Intra-Chromosomal Crossing Over Leads to Deletion



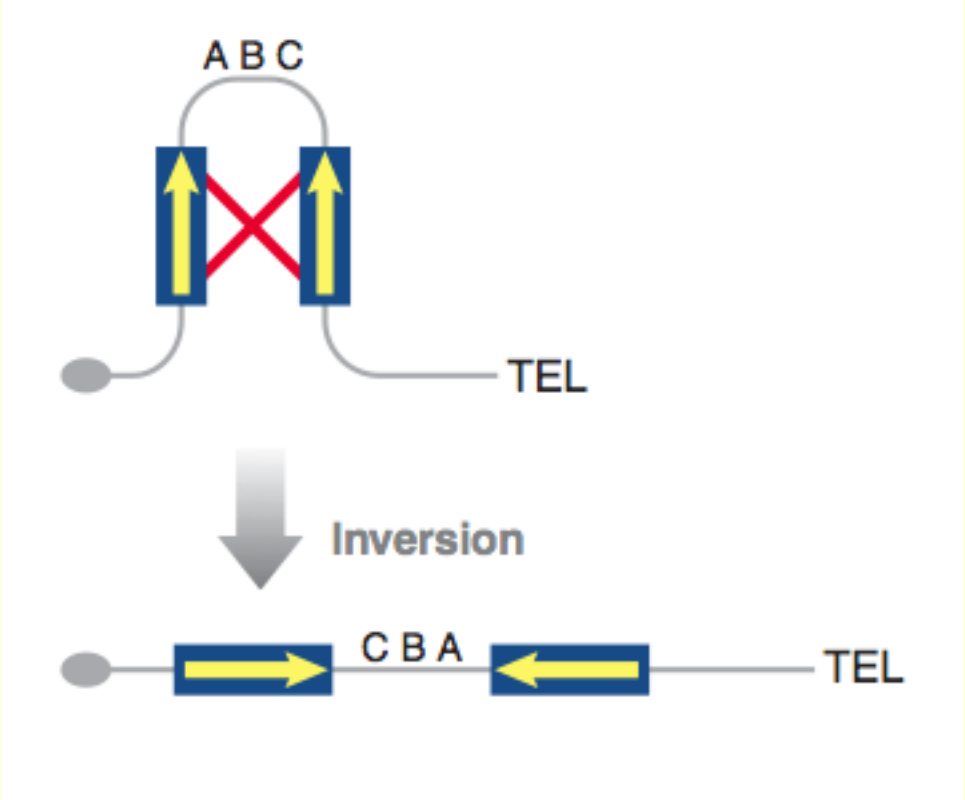
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Inter-Chromosomal Crossing Over Leads to Inversion



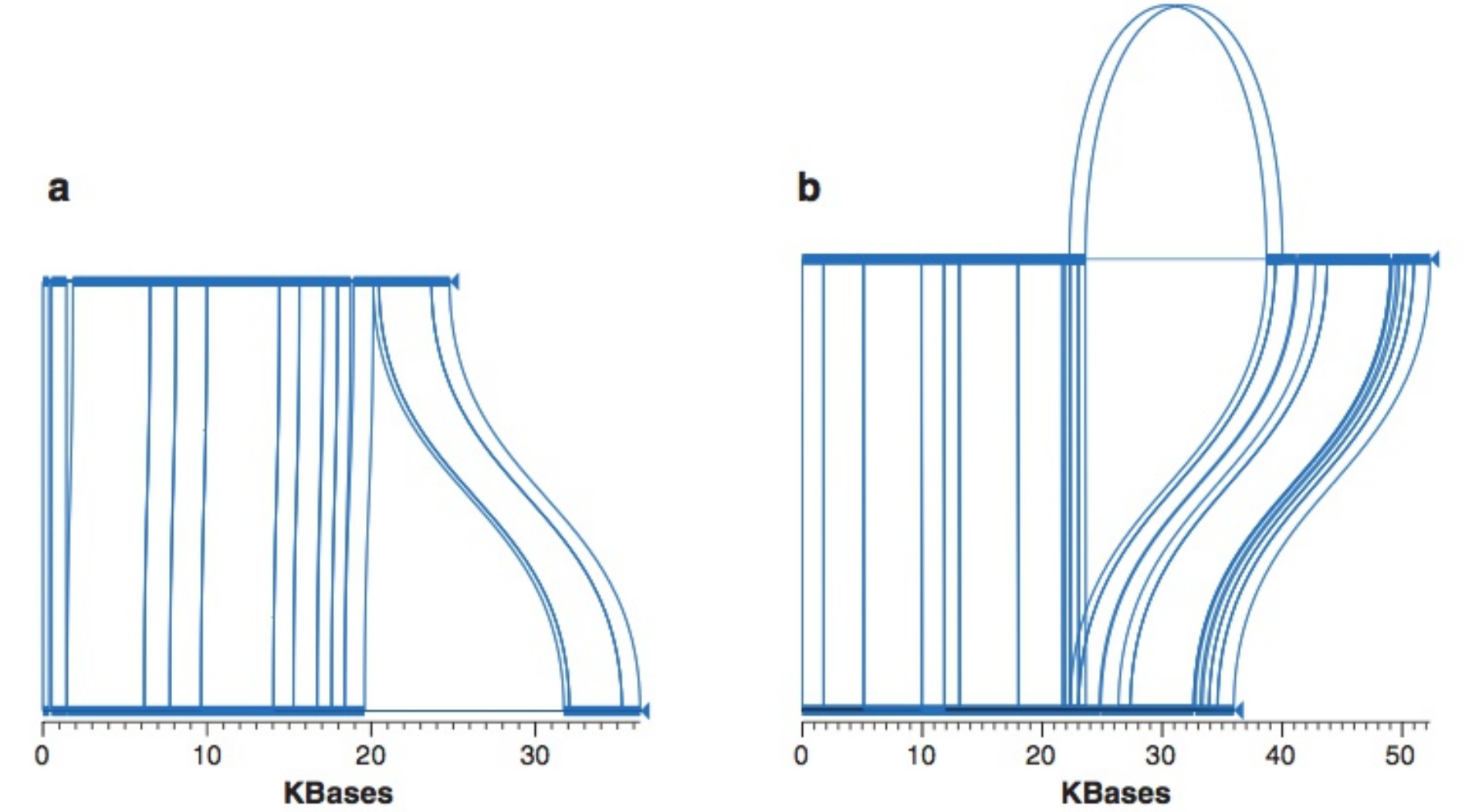
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Intra-Chromosomal Crossing Over Can Also Lead to Inversion



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

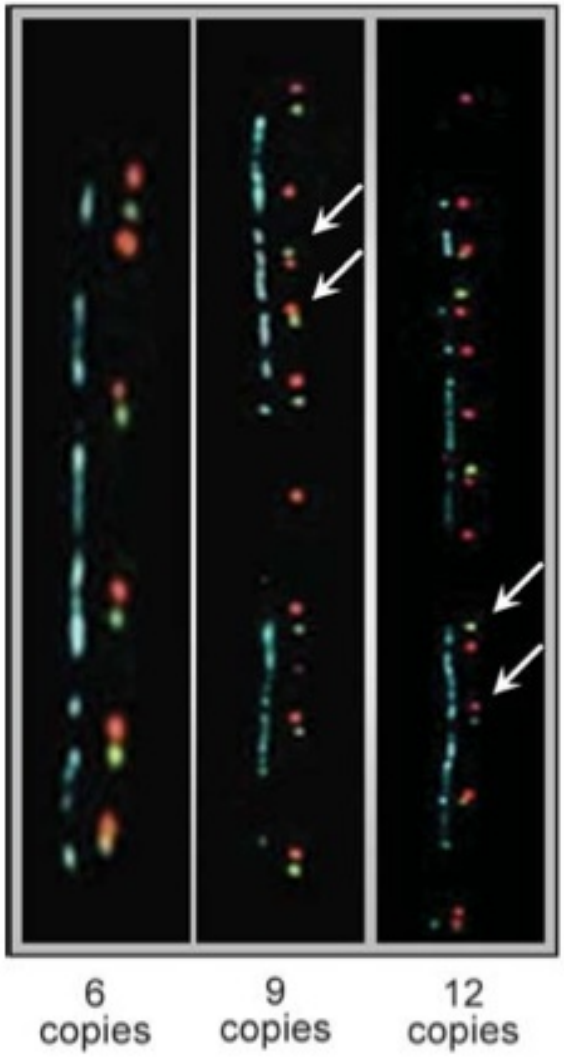
Deletions and Insertions at Repeat Sequences



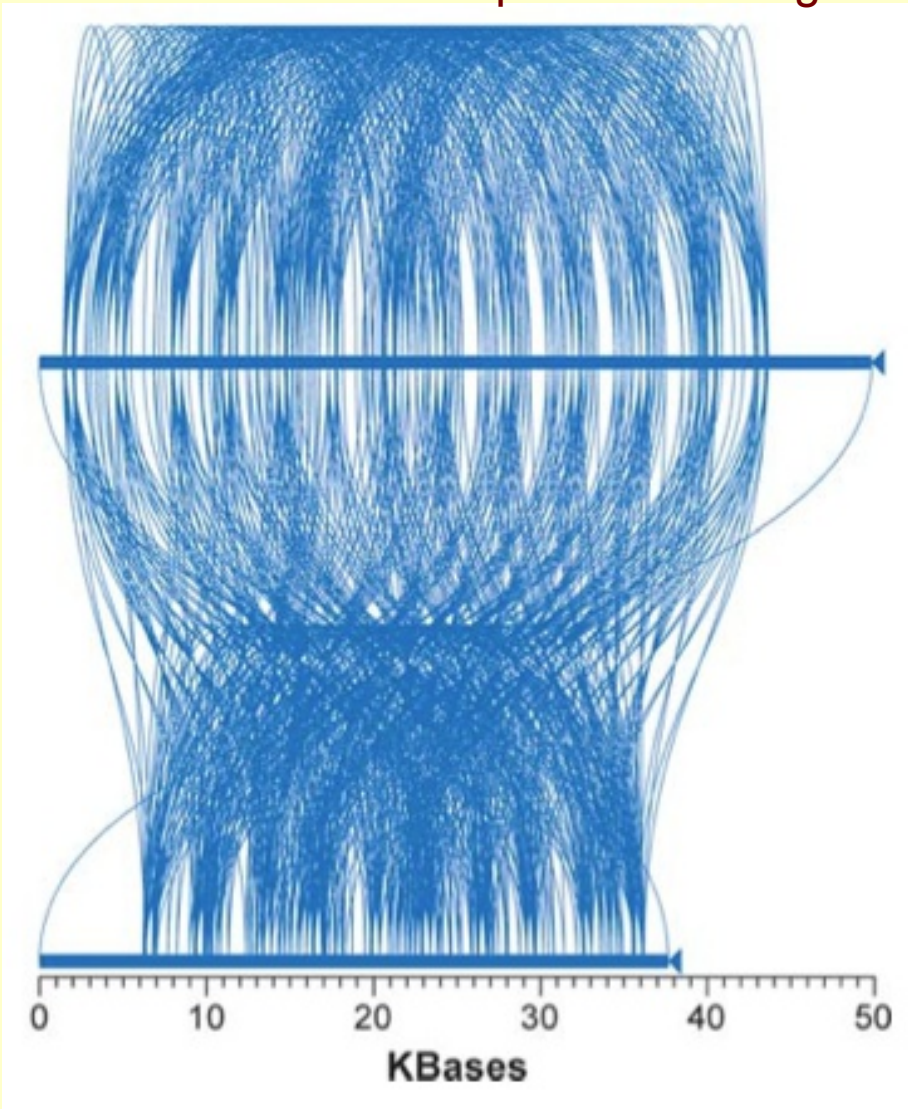
Sharp, Cheng & Eichler, Annu. Rev. Genomics Hum. Genet. 2006. 7:407–42

Variations in α-Amylase Gene Tandem Repeat Arrays

FISH on DNA

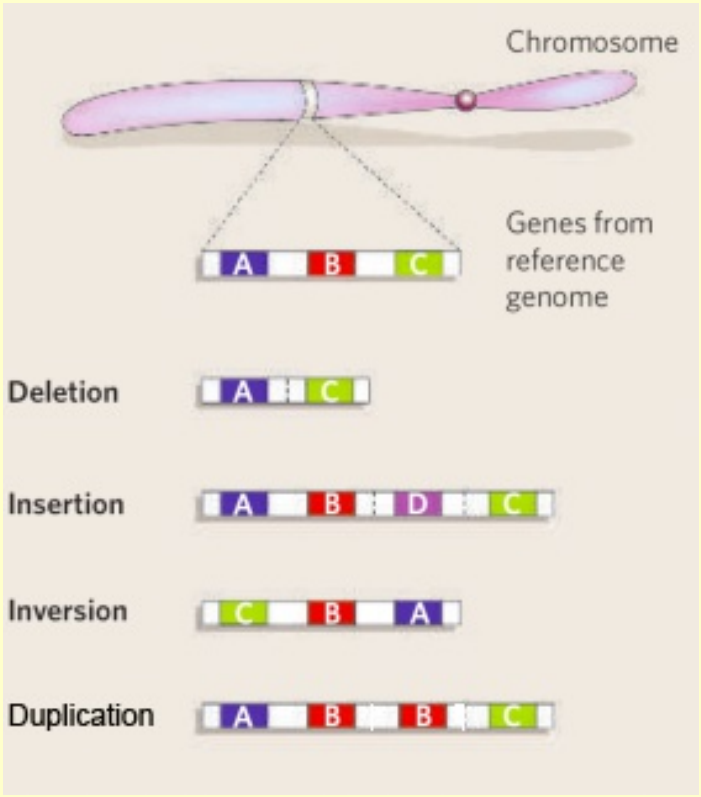


8 or 12 tandem repeats 4 kb long



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

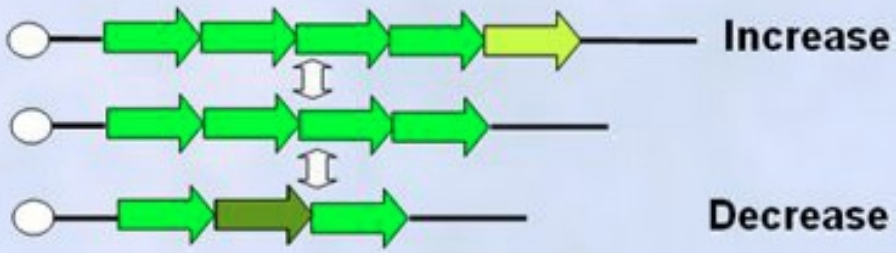
Why Study Structural Variation?

- 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 and neurodegenerative diseases
- They are now also showing up in rarer diseases and common behavioral diseases such as autism, schizophrenia, attention deficit, learning disabilities and many other neurological disorders

Courtesy of Mike Snyder

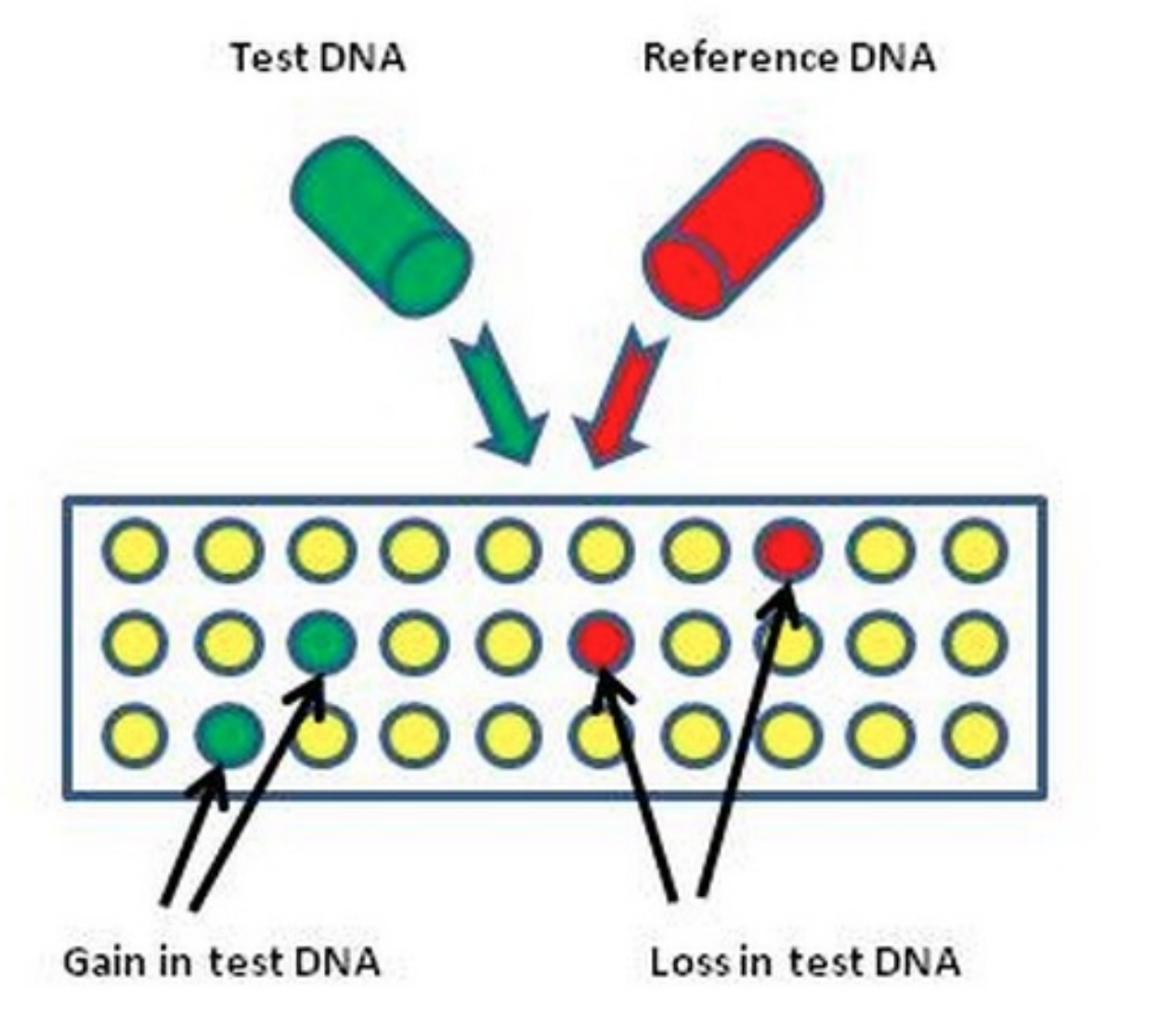
Copy Number Variation and Disease 2002

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



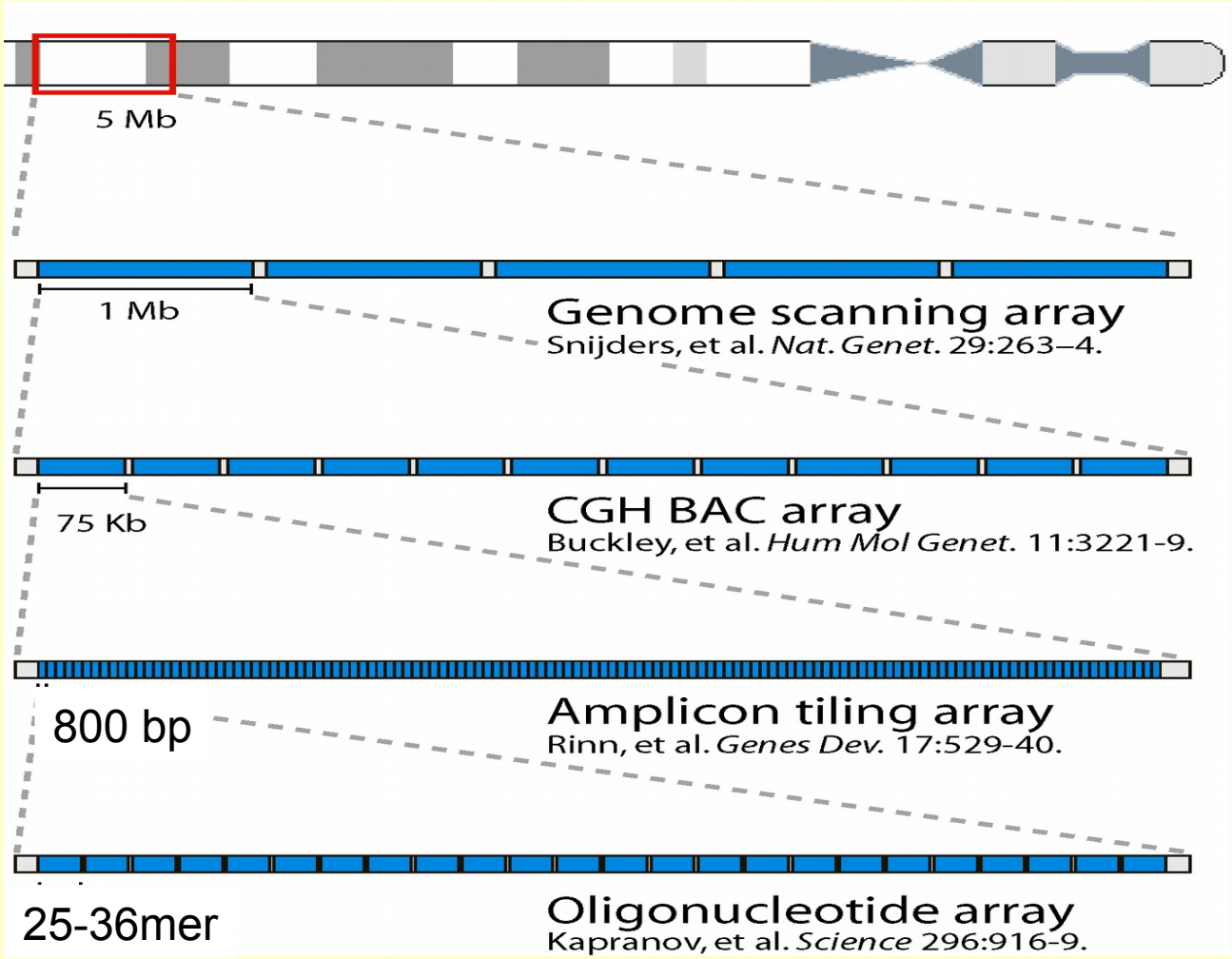
Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

Comparative Genomics Hybridization (CGH)



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

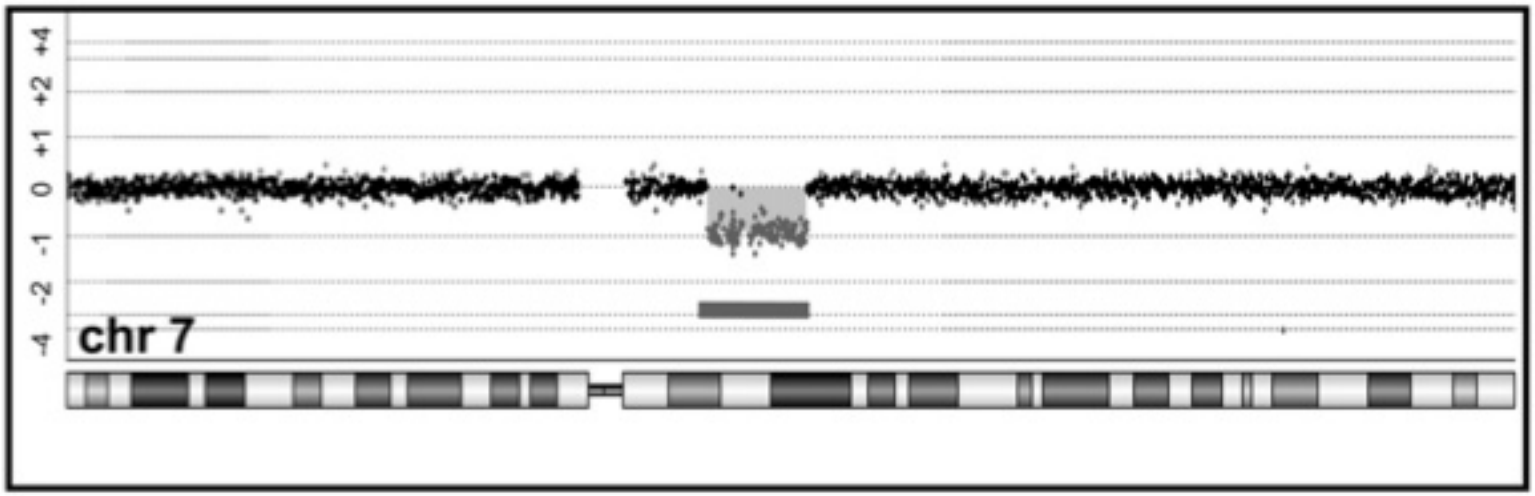
Comparative Micro Arrays (CMA) Using Genome Tiling Arrays

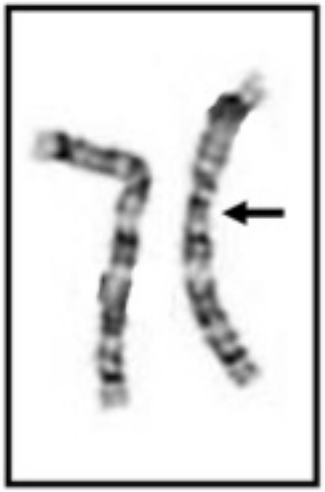


Courtesy of Mike Snyder

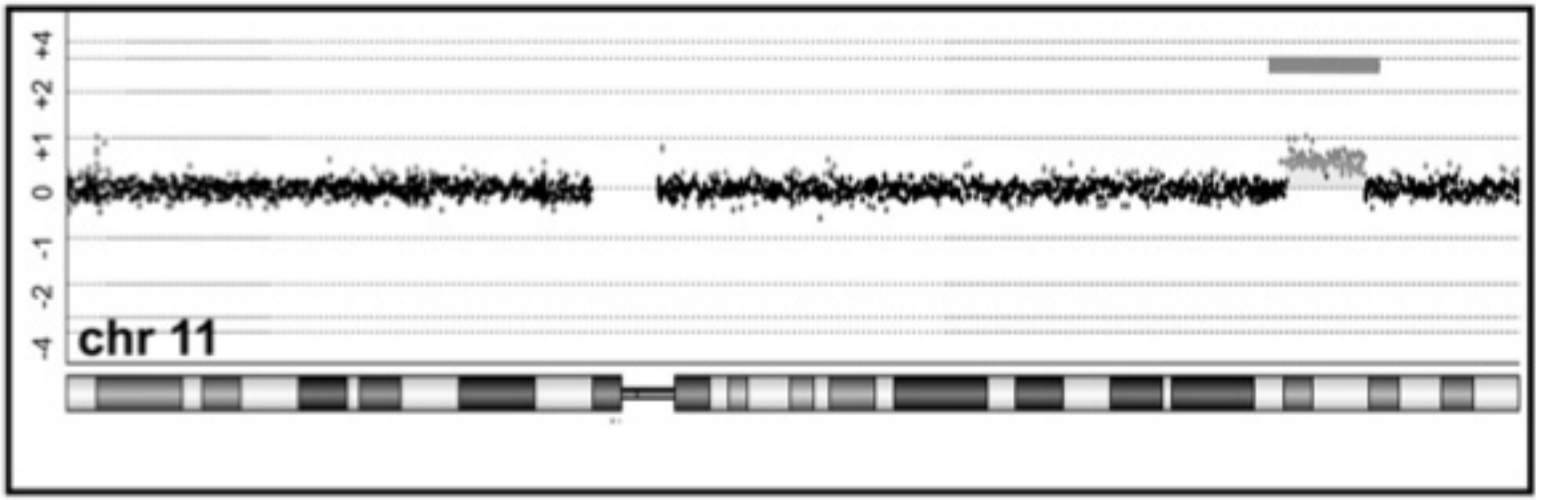
Detection of Duplications and Deletions Using Chromosomal Micro-Arrays

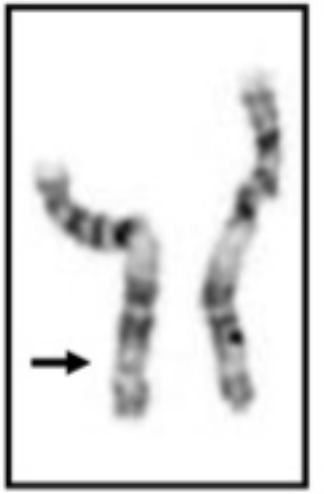
A 10.9 Mbase deletion at 7q11 in Williams-Beuren Syndrome





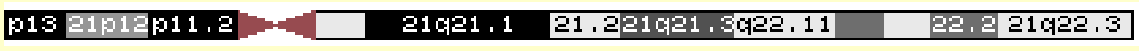
B 7.2 Mbase duplication in 11q

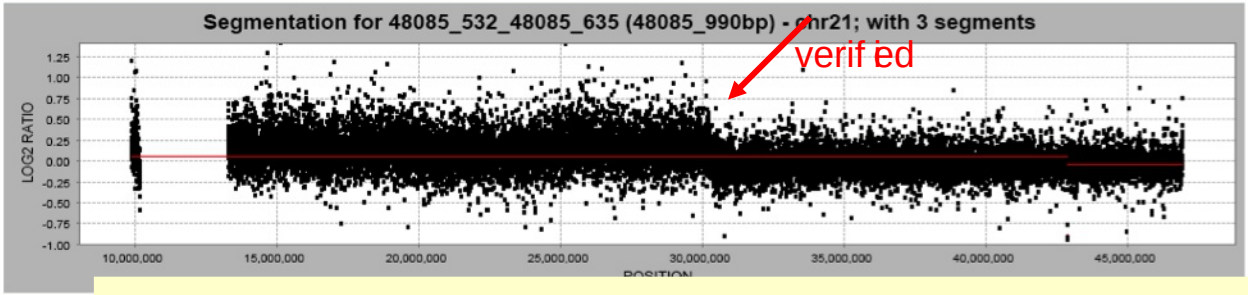


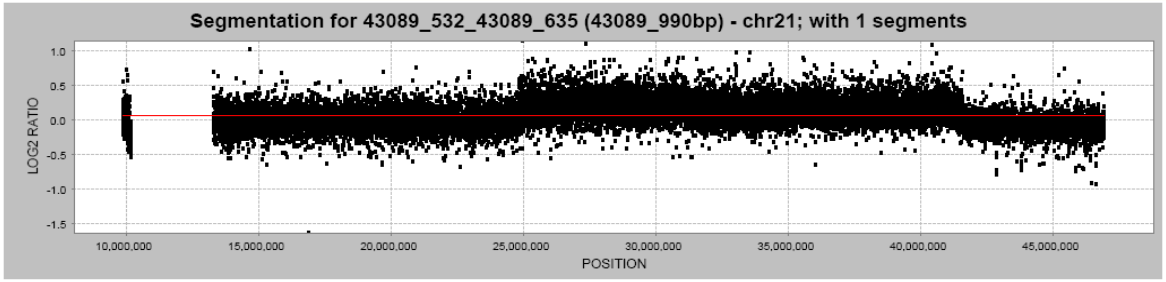


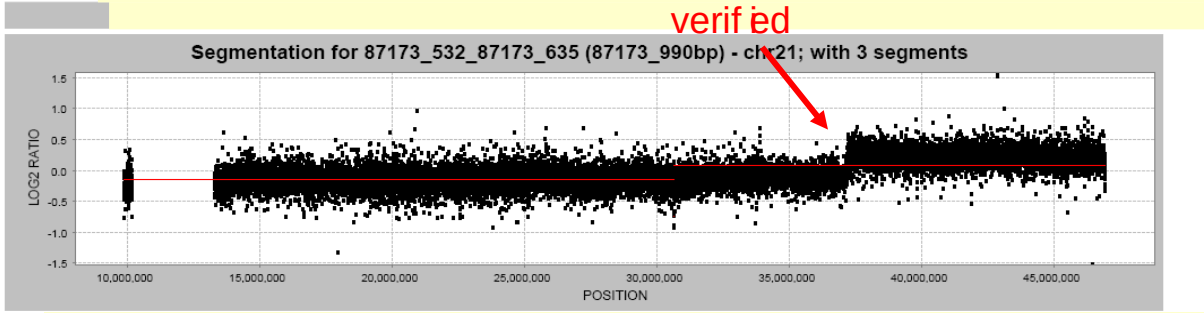
Miller et al., (2010) American Journal of Human Genetics 86, 749-764.

Mapping Breakpoints of Partial Trisomies of Chromosome 21









Courtesy of Mike Snyder

Paired End Mapping

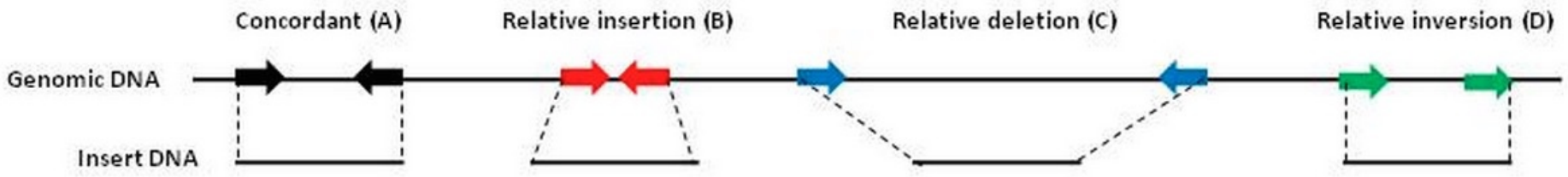
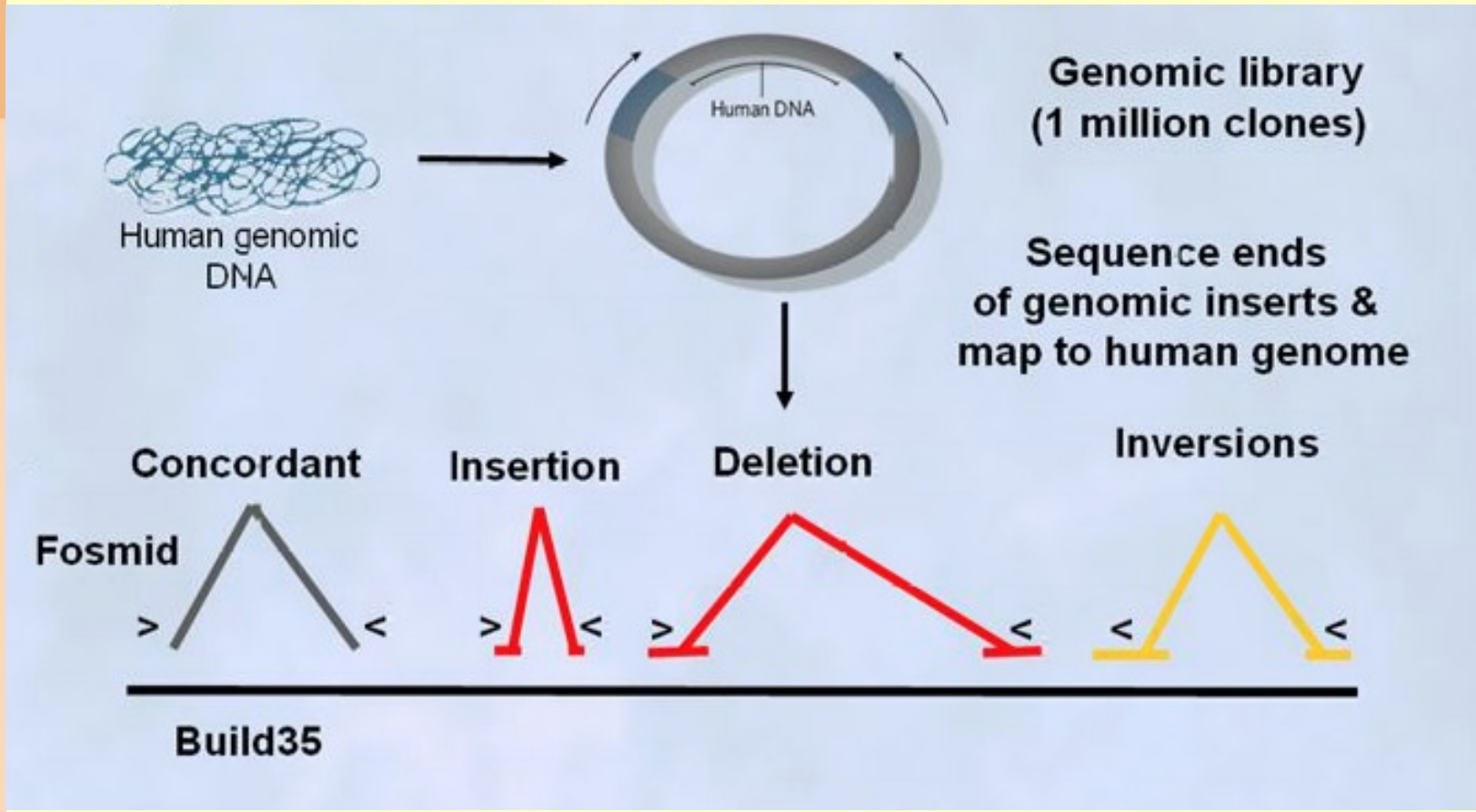


Figure 5: Paired-end mapping (PEM).

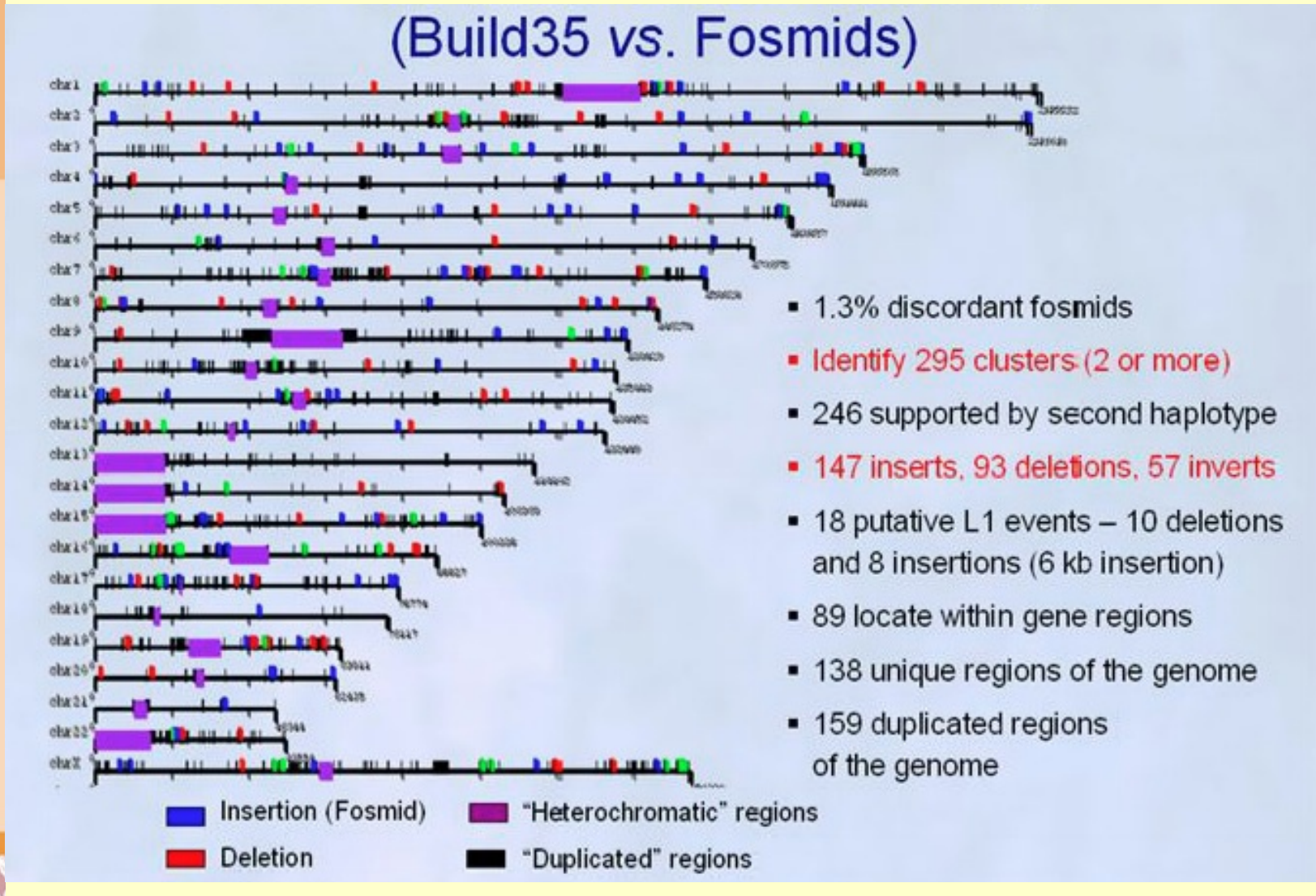
http://www.ncbi.nlm.nih.gov/dbvar/content/overview/

Sequence Base Resolution of Structural Variation



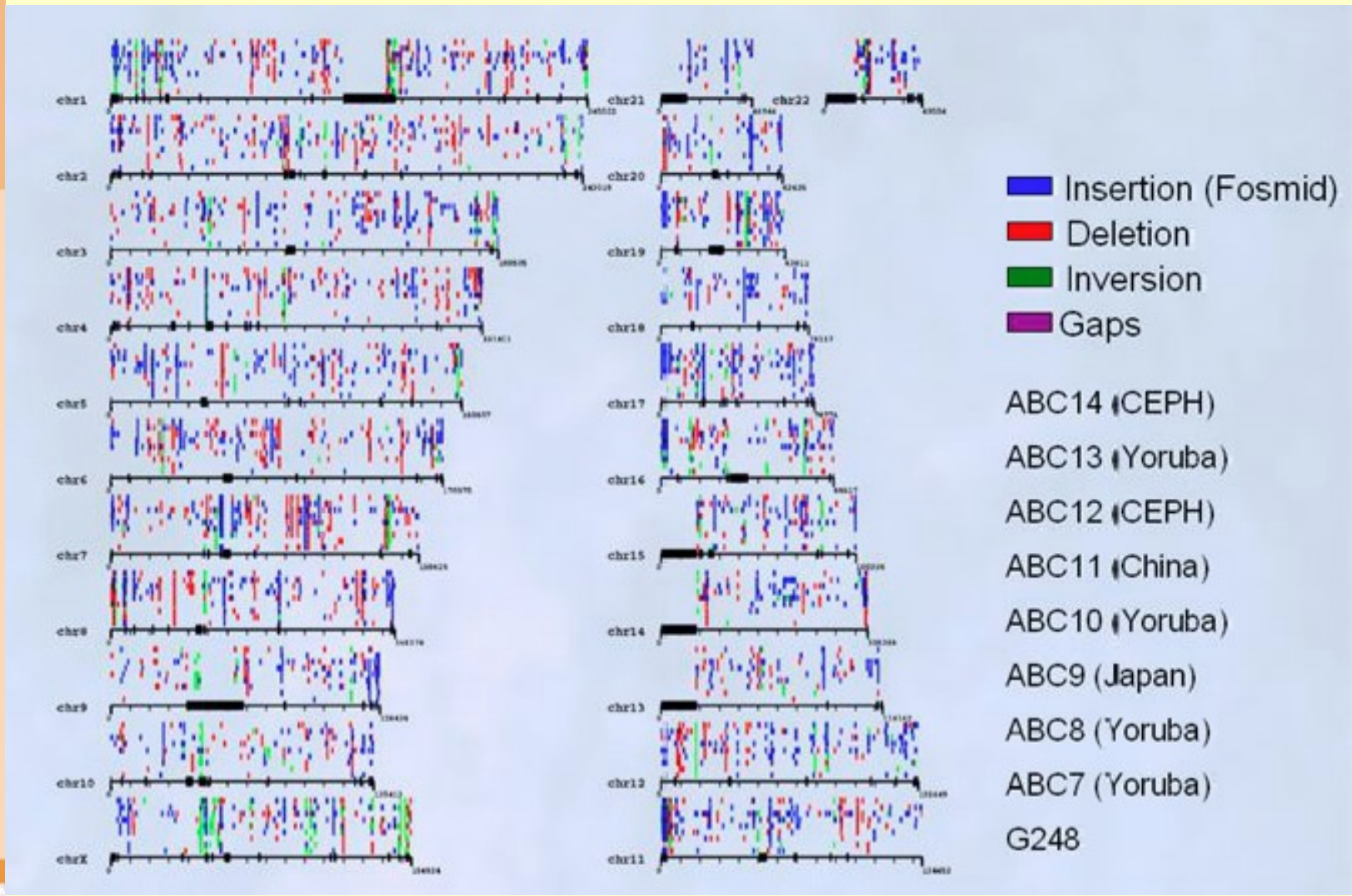
Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

Fine Scale Structural Variation for Eight Genomes



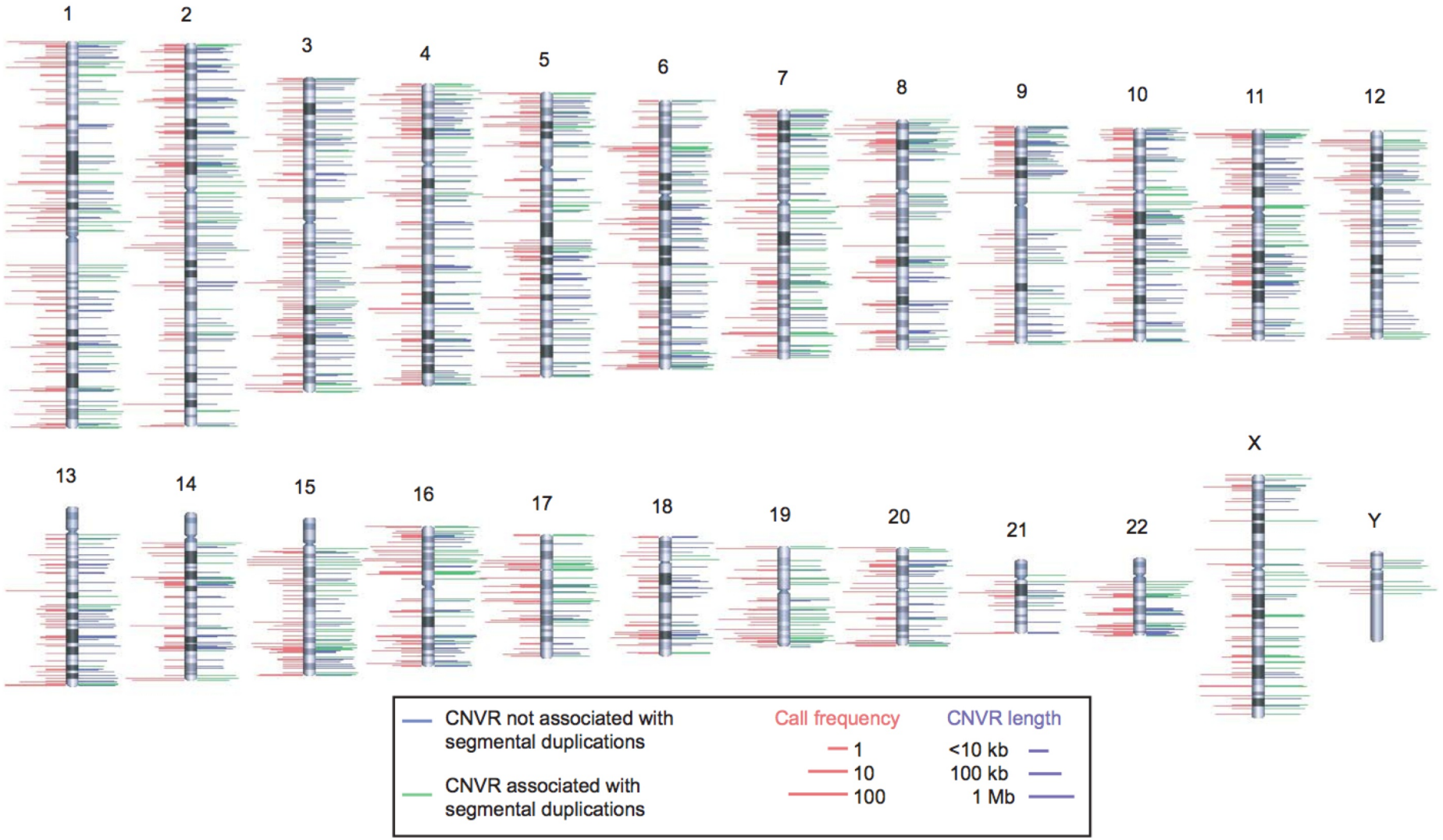
Henry Stewart Talks: Evan Eichler
http://hstalks.com/main/view_talk.php?t=1409&r=439&j=757&c=252

A Structural Variation Map of the Human Genome



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

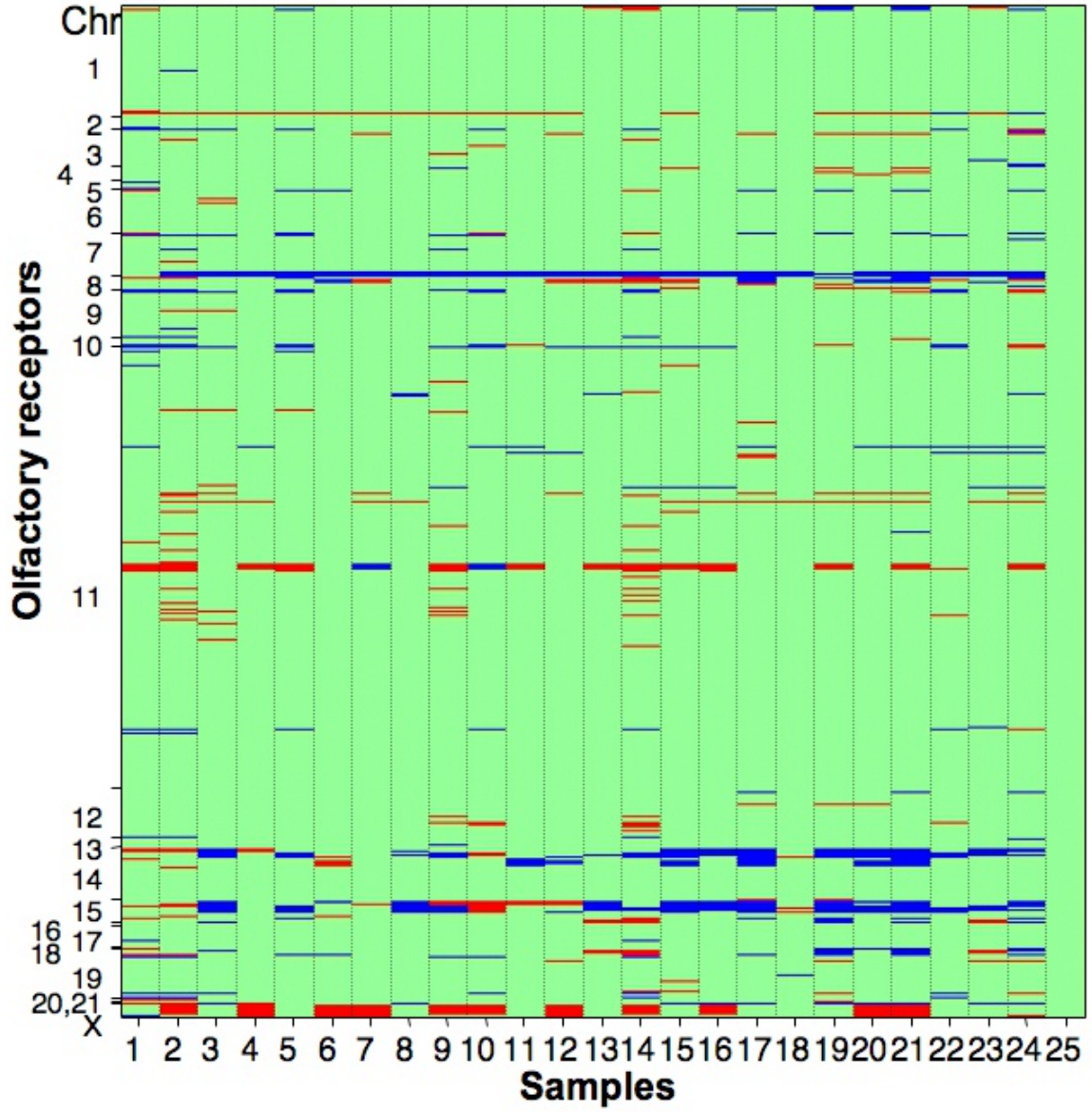
Genomics Distribution of CNV Regions



Courtesy of Mike Snyder

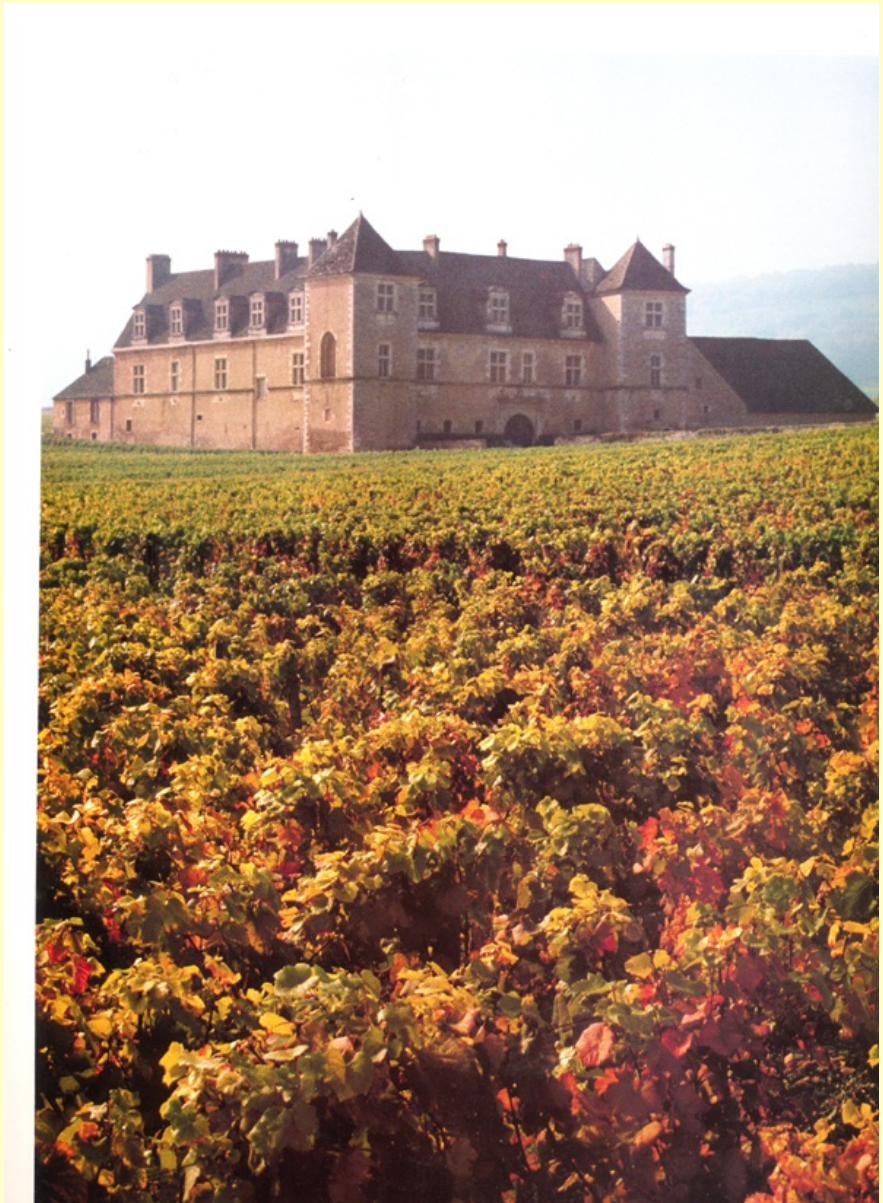
Heterogeneity in Olfactory Receptor Genes

(Examined 851 Olfactory Receptor Loci)



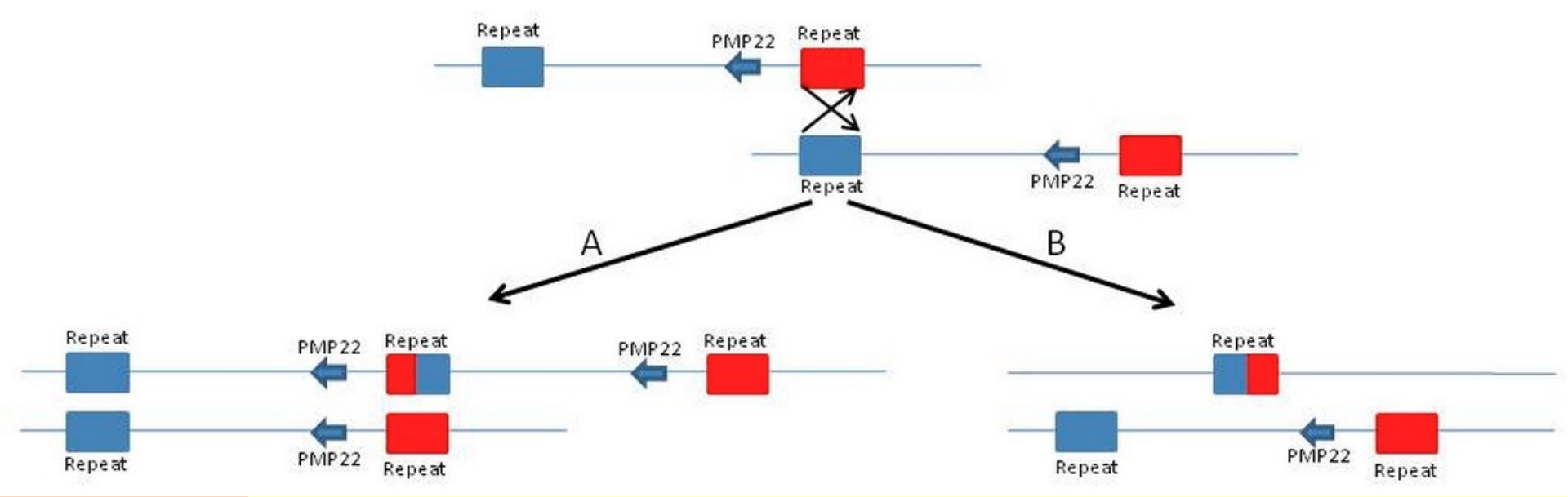
Hasin, ..Snyder, .. Lancet et al. PLoS Genet. 2008 Nov;4(11)

Clos Vougeot in Bourgogne



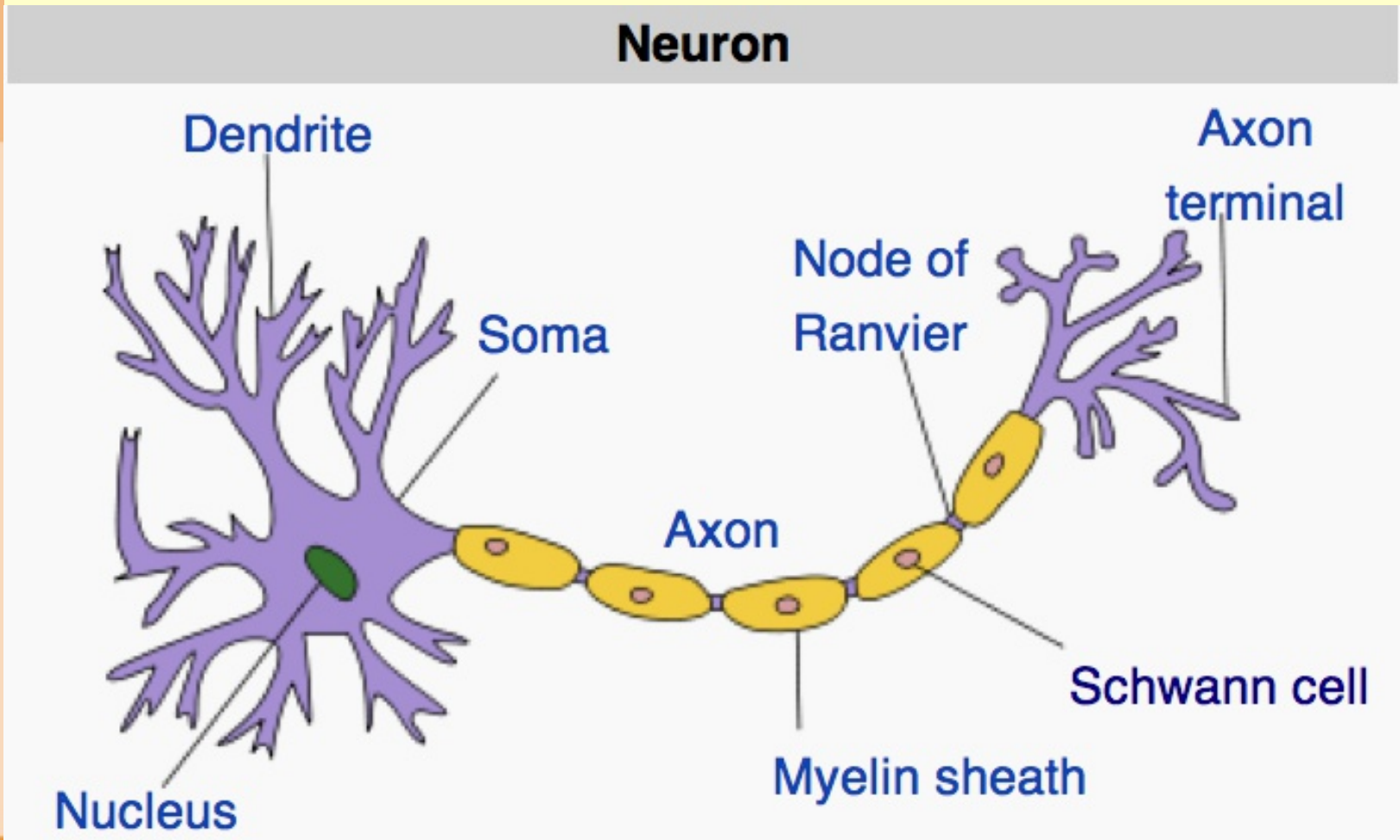
Chef d'Ordre de la Confrerie des Chevalier du Tastevins

Charcot-Marie-Tooth Hereditary Neuropathy (CMT1) Disease Results From CNV of PMP22 Gene in 17p11.2-12

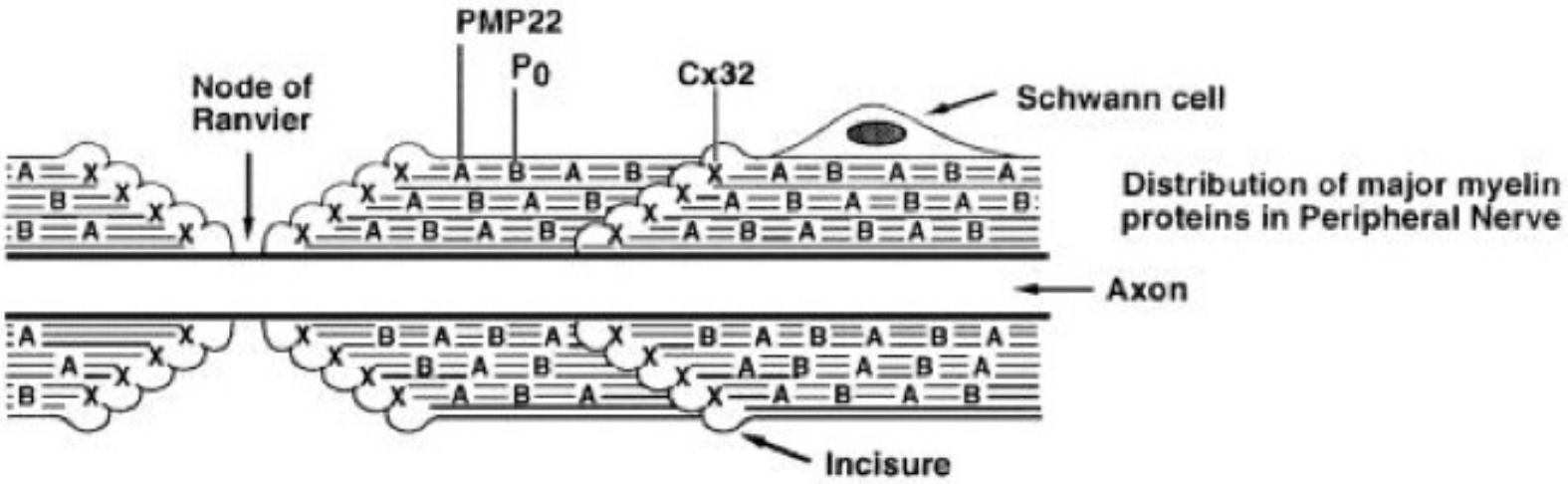


Peripheral Neuropathy, Yuen So, Medical Grand Rounds Jan 16

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



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



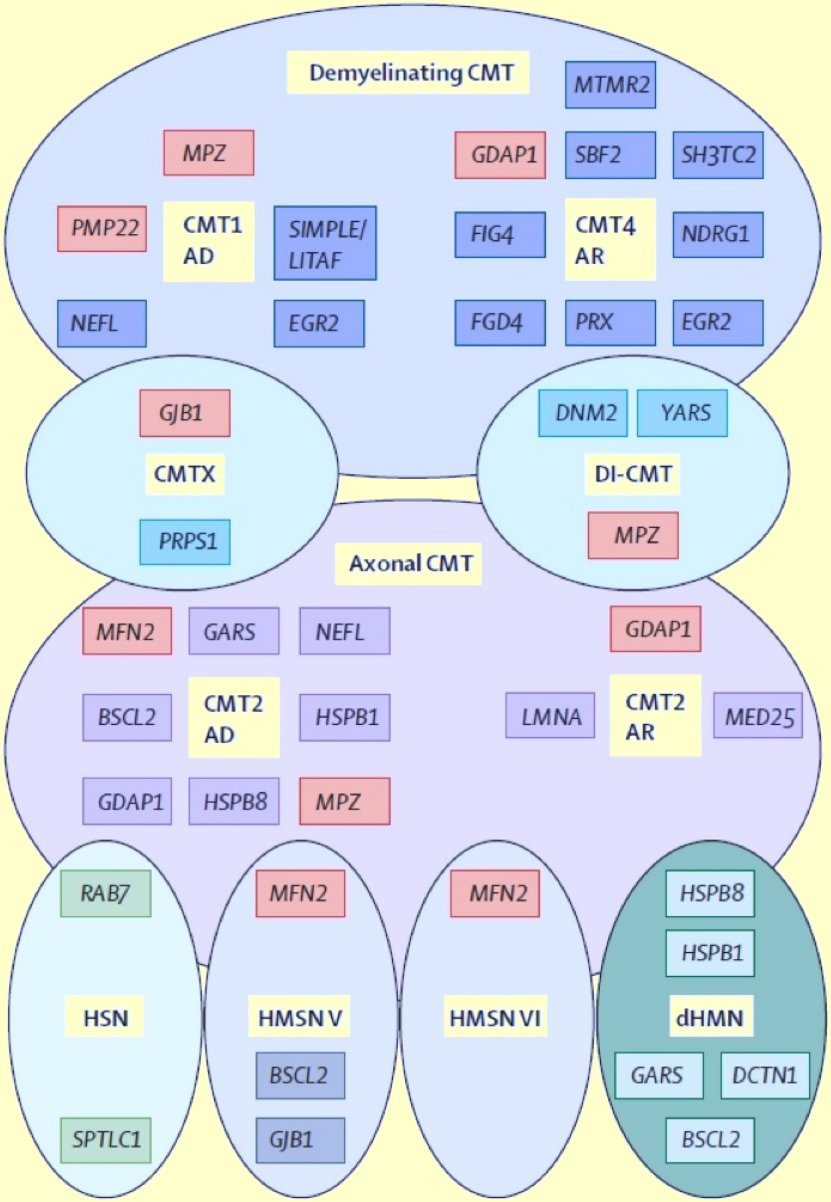
Charcot-Marie-Tooth Hereditary Neuropathy (CMT1) Disease Genes

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

CMT 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/

Structural Variations in Mendelian 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

Mendelian CNV mutations (Prof. Joris Veltman in Henry Stewart talks)

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

Behavioral Diseases 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

Behavioral Diseases Associated with Structural Variations (Cont.)

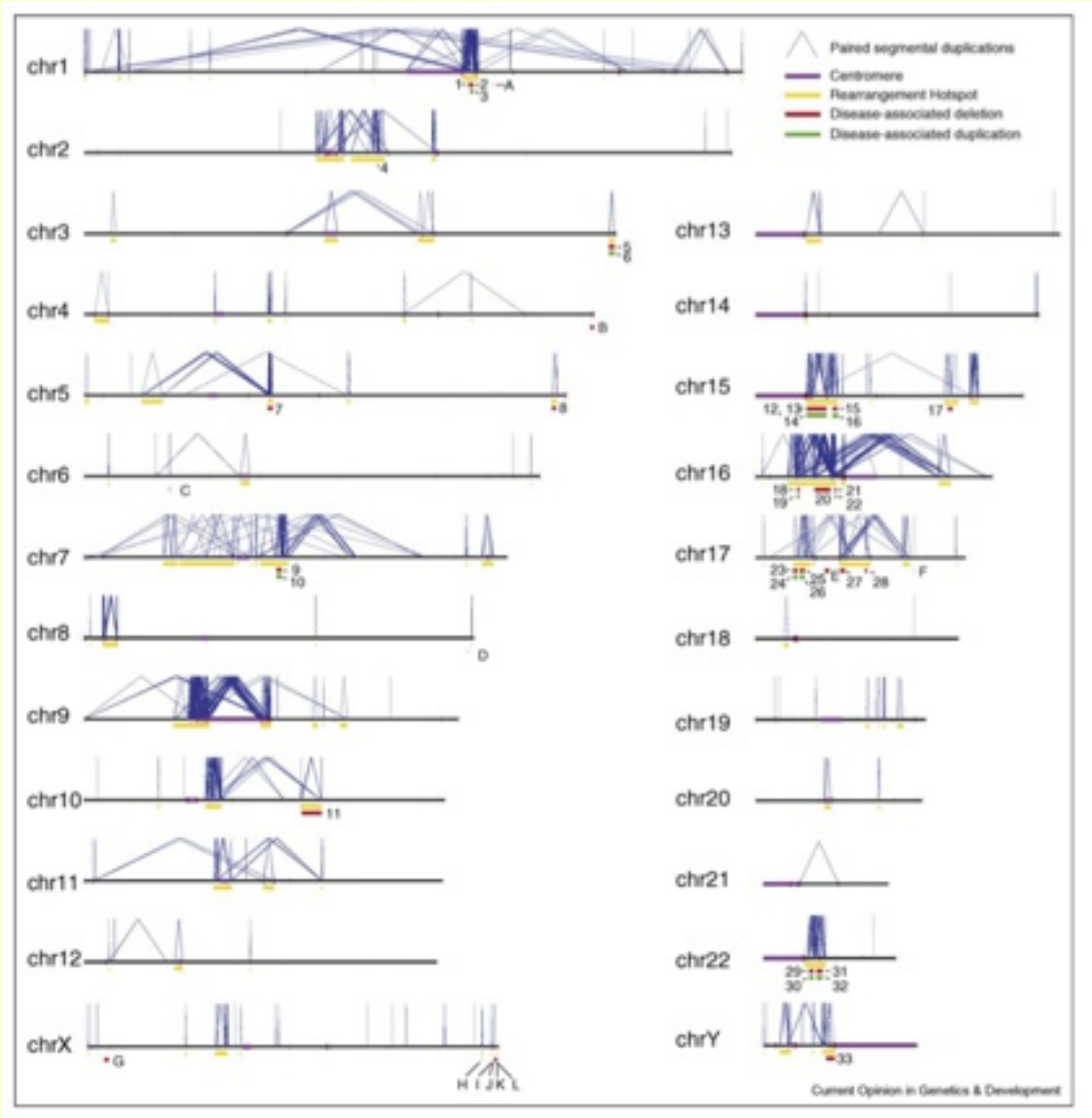
16p11.2 duplication	chr16: 29.50–30.10 Mb	600 kb	*SEZ6L2, ALDOA, TBX6, QPRT*	Autism, learning disability	7/2252	0.31	7/23 502	0.03
				Motor delay, congenital anomaly, behavioral features, and microcephaly	18/7400	0.24	NA	—
				Schizophrenia, microcephaly	26/8590	0.30	8/28 406	0.03
				Learning disability, 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

Inversions Lead to Instability & Disease

Table 2 Summary of polymorphic inversions that predispose to further rearrangements

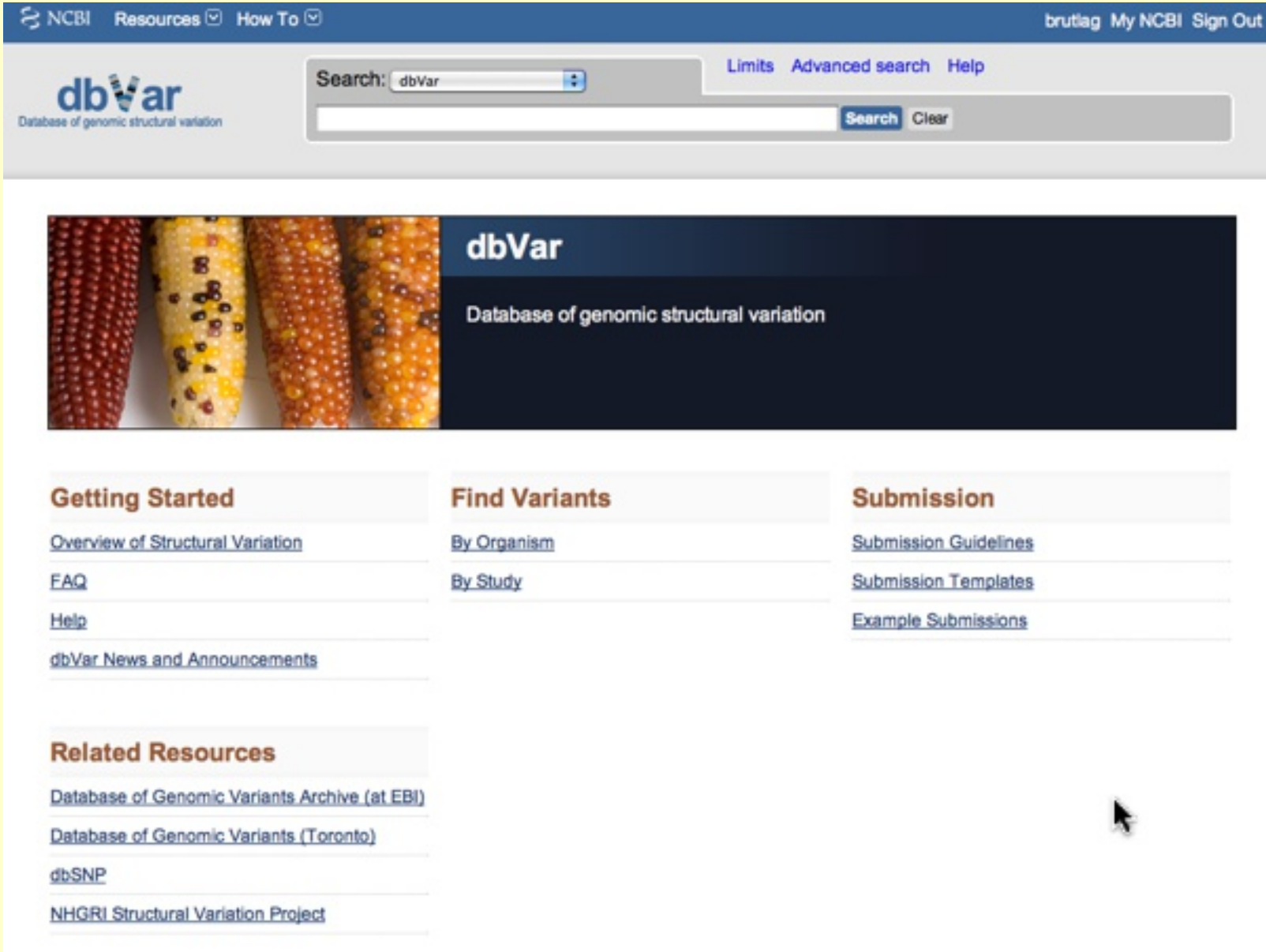
Locus	Cytogenetic location	Population frequency	Size of inversion region	Associated predisposition
OR genes	4p16	12%	~6 Mb	t(4;8)(p16;p23) translocation
Sotos syndrome critical region	5q35	Unknown	2.2 Mb	Deletion of SoS critical region
Williams-Beuren syndrome critical region	7q11.23	Unknown	1.6 Mb	Deletion of WBS critical region (and atypical WBS phenotype?)
OR genes	8p23	26%	4.7 Mb	inv dup(8p), +der(8)(pter-p23.1::p23.2-pter) and del(8)(p23.1;p23.2)
Angelman syndrome critical region	15q11-q13	9%	~4.5 Mb	Deletion of AS critical region
Proximal Yp	Yp11.2	33%	~4 Mb	*PRKX/PRKY* translocation (sex reversal)

Rearrangement Hot Spots Associated with Disease



dbVAR Database at NCBI

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



dbVAR Report on PMP22 Gene

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

dbVar
Database of genomic structural variation

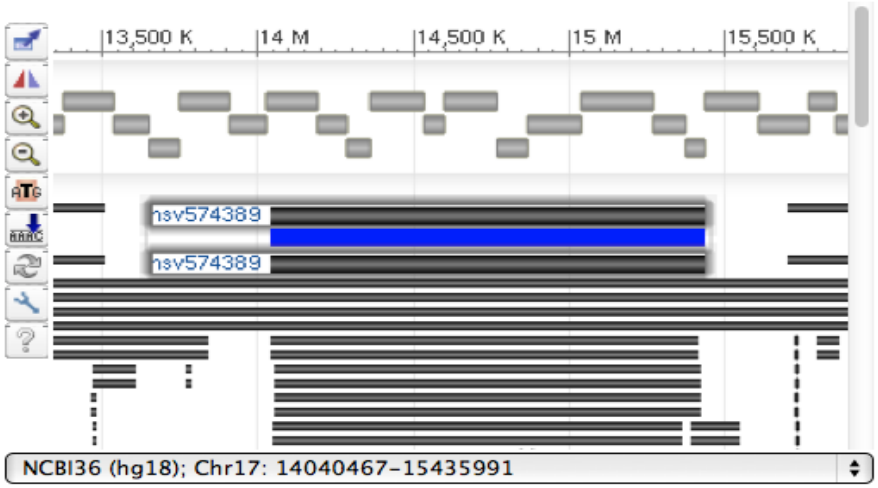
NCBI Resources How To

dbVar dbVar

Limits Advanced

Variant Information

- **Variant accession:** nsv574389
- **Organism:** Human
- **Study:** nstd54
- **Variant Type:** CNV
- **Method type:** SNP array
- **Validation:** Not tested
- **Genomic location:**
- **Submitted:** NCBI36 (hg18); Chr17: 14,040,467 - 15,435,991



Detailed Variant Placement Information

ID	Placement Type	Assembly	Placement	Start	Stop
NC_000017.9	Submitted Genomic	NCBI36 (hg18)	Chr17	14,040,467	15,435,991

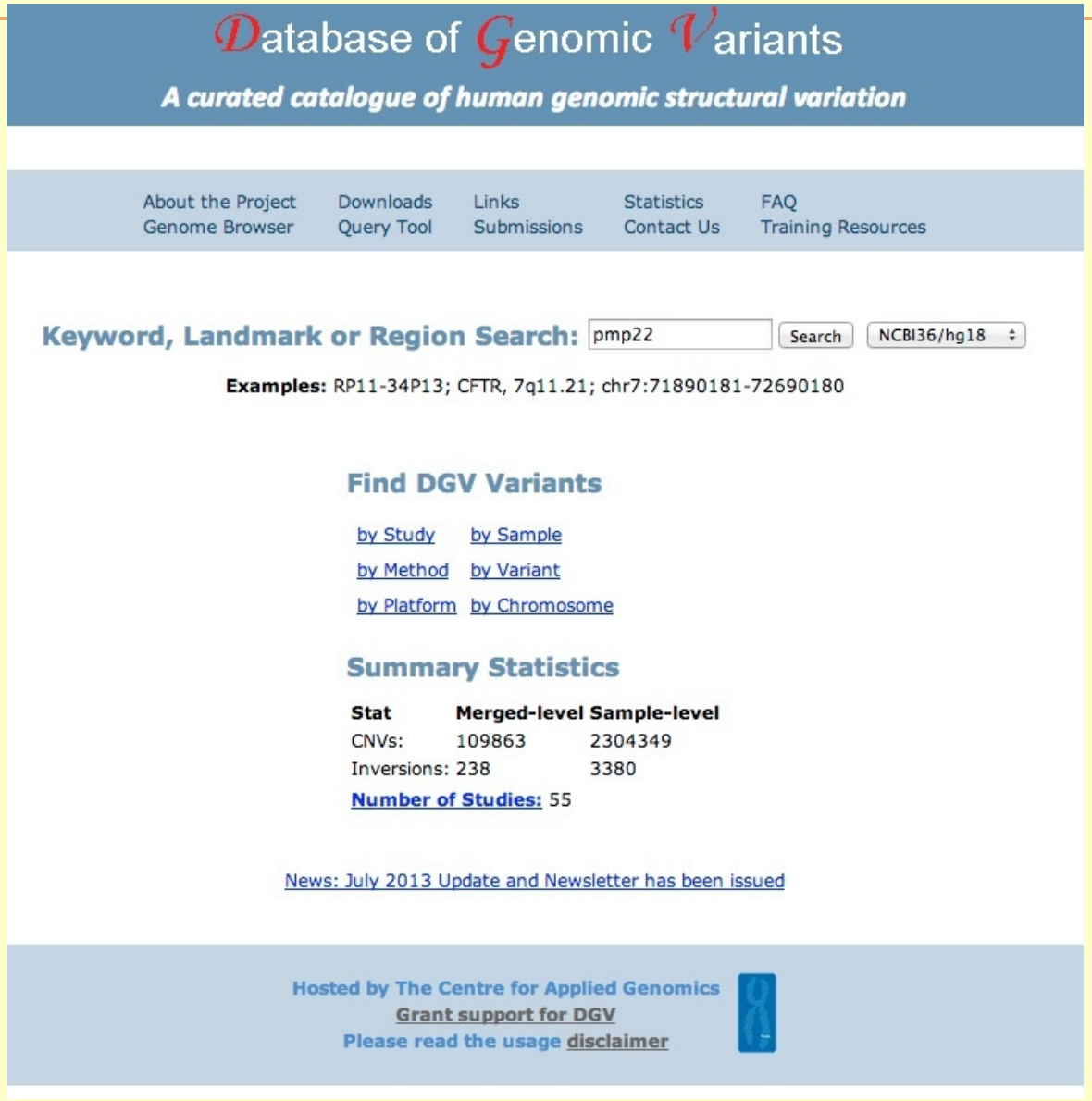
Supporting Variants

ID	Type	Allele Length	Sample ID	Subject Phenotype	Assembly	Placement	Start	Stop	Placement Type
nssv867002	Gain	1395524		Not reported	NCBI36 (hg18)	Chr17	14,040,467	15,435,991	Submitted Genomic

Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

http://dgv.tcag.ca/



Hosted by:
The Centre for
Applied Genomics

Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18): 100 kbp from chrX:153,058,413..153,158,412

Browser | Select Tracks | Custom Tracks | Preferences

Search

Landmark or Region:

chrX:153,058,413..153,1 Search

Examples: chr7:71890181..72690180, CFTR, AC108171.3, nsv529033.

Data Source

Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18)

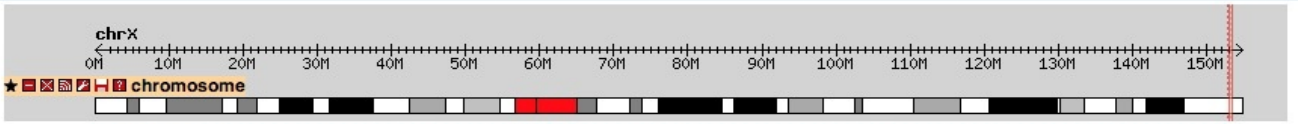
Scroll/Zoom: Show 100 kbp Flip

Filter variants

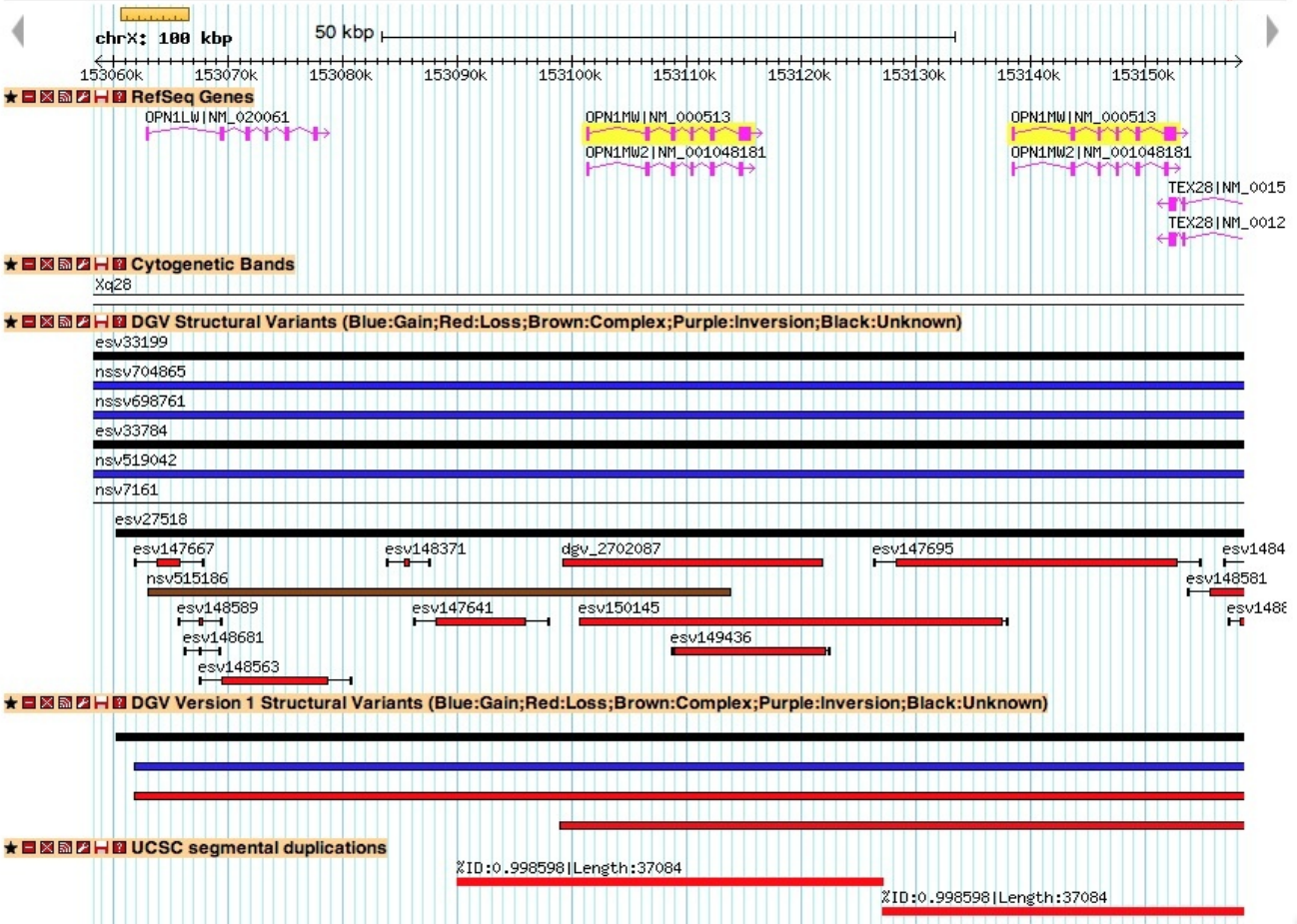
study = + -

Filter Reset

Overview



Details



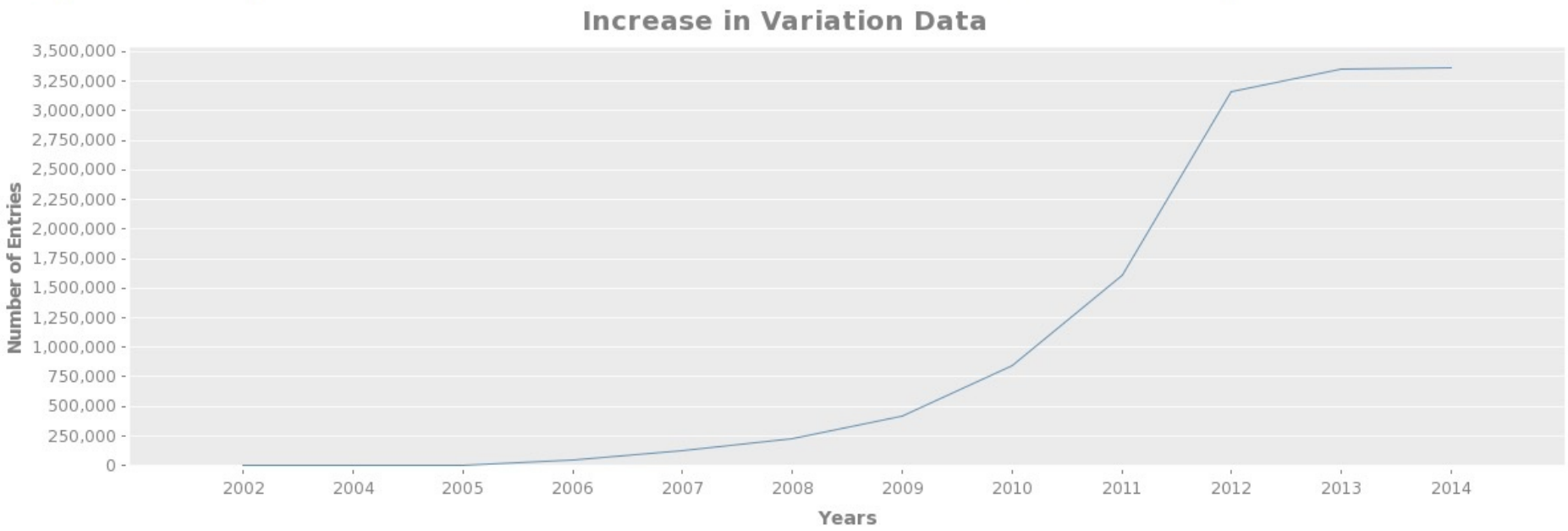
Hosted by:
The Centre for
Applied Genomics

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.



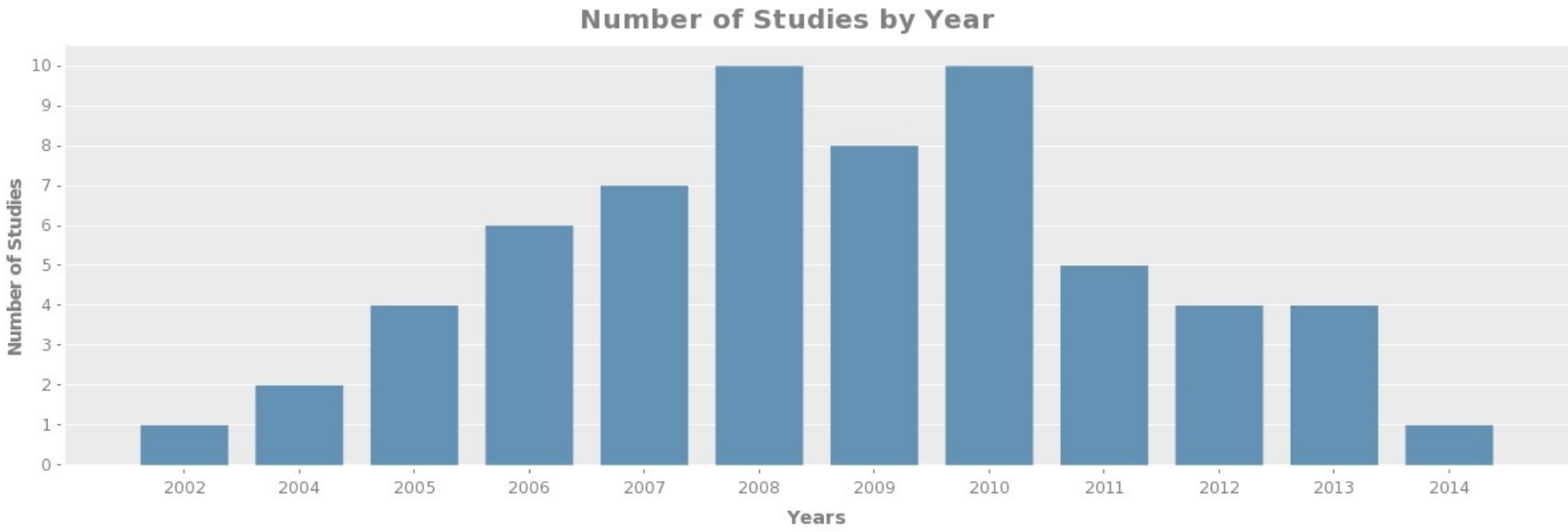
Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

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

Number of Studies

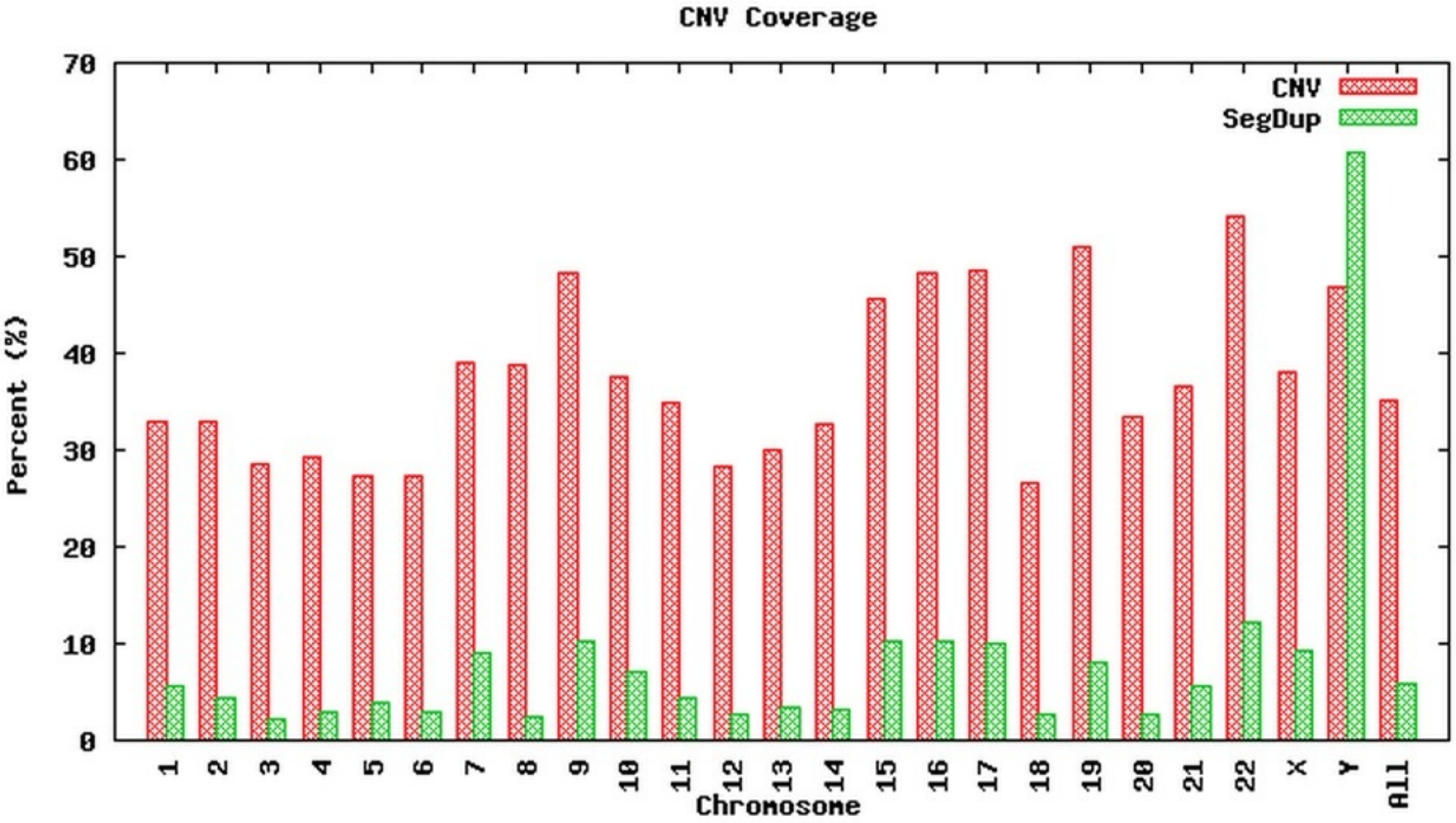
This graph shows the number of studies currently in DGV by published year.



Hosted by:
The Centre for
Applied Genomics

Database of Genomics Variants

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

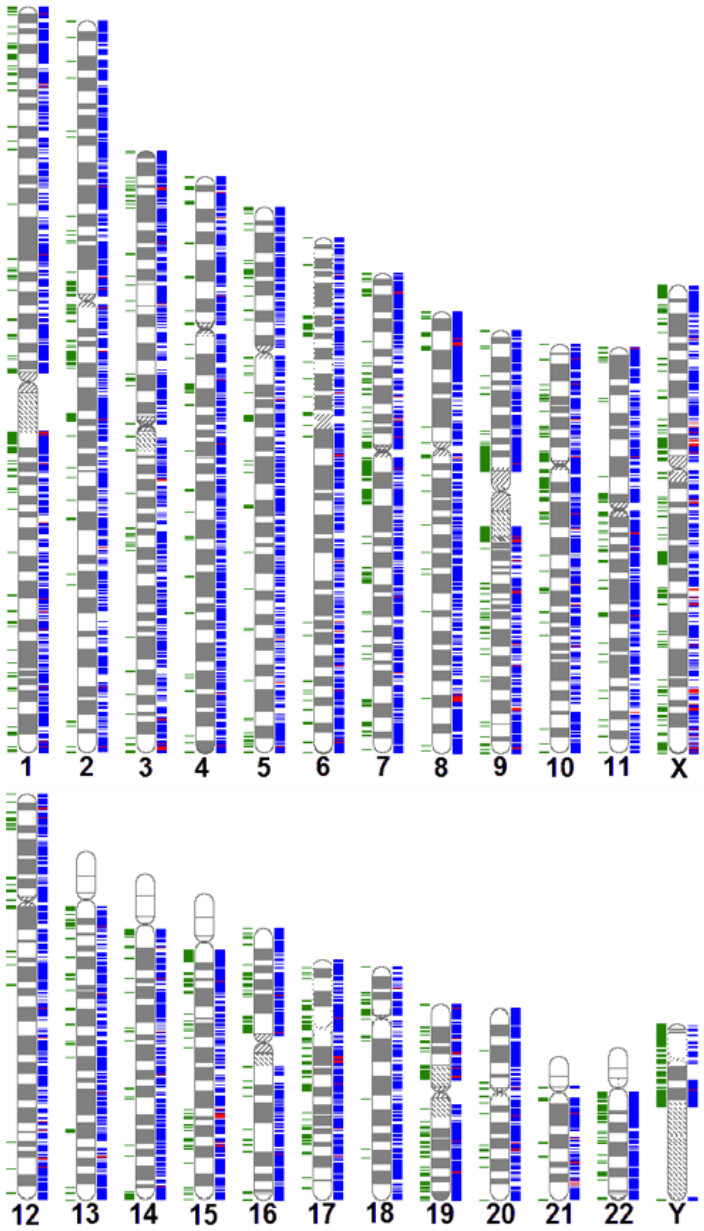


Hosted by:
The Centre for
Applied Genomics

Database of Genomic Variants

Genome-wide view of CNVs

Click on a cytoband to get a list of variants detected within that region



Legend: Blue bars indicate reported CNVs; Red bars indicate reported inversion breakpoints; Green bars to the left indicate segmental duplications.

Hosted by:
The Centre for
Applied Genomics

Showing 5 Mbp from chr17, positions 12,649,493 to 17,649,492

Instructions
Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change

Examples: chr7:71890181..72690180, CFTR, NM_030798.

[Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset]

Search

Landmark or Region:

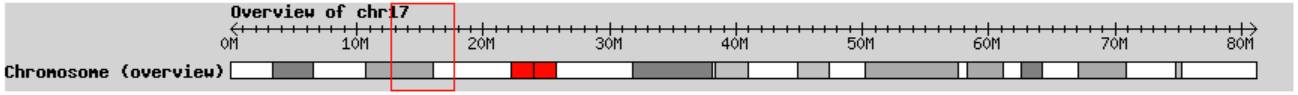
chr17:12649493..1764949 Search

Data Source

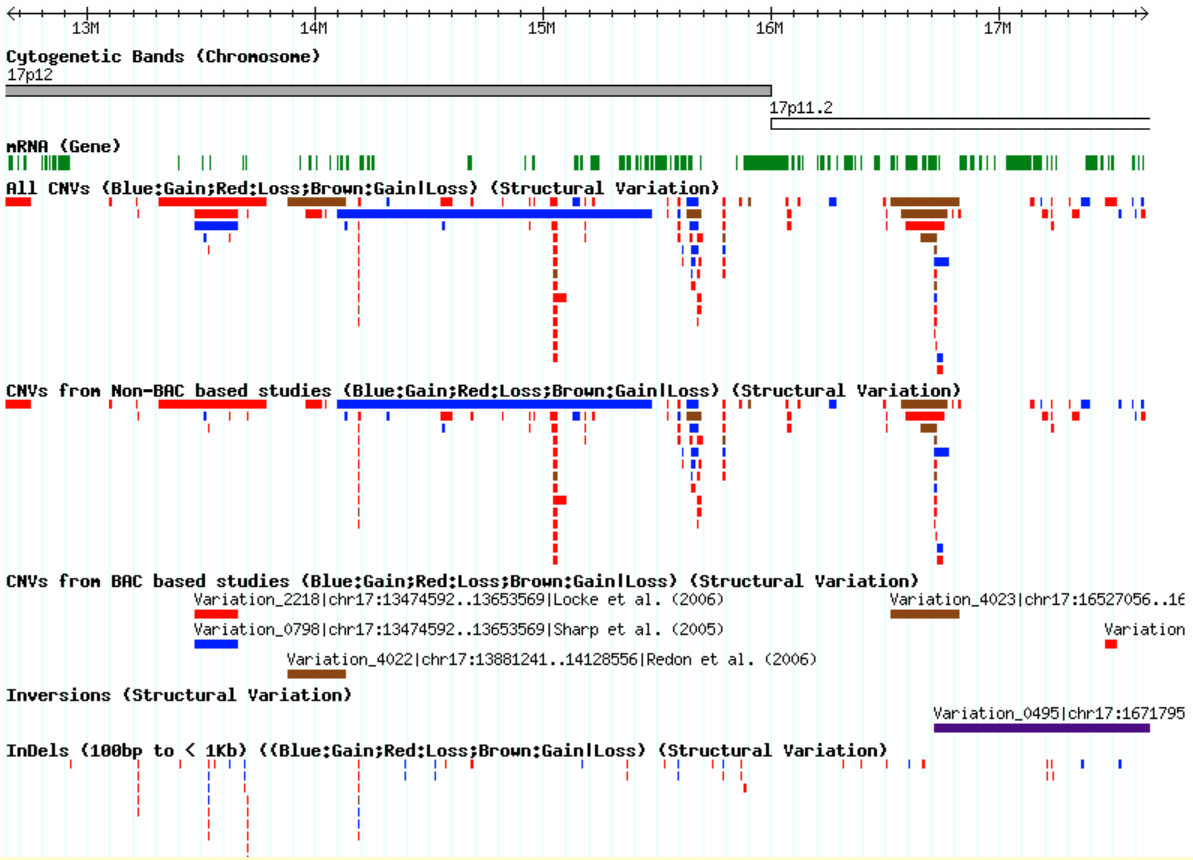
Genomic Variants in Human Genome (GRCh 37: Feb. 2009) (hg19)

Scroll/Zoom: Show 5 Mbp Flip

Overview



Details



NHGRI Structural Variation Project

http://www.ncbi.nlm.nih.gov/projects/genome/StructuralVariation/NHGRIStructuralVariation.shtml

NHGRI Structural Variation Project

The sequence-based Survey of Human Structural Variation aims to characterize common structural variants that are larger than SNPs, for example, multi-base insertions/deletions, inversions, translocations, and duplications. The approach entails sequencing the ends of fosmids and BACs from multiple individuals. This strategy can be efficiently scaled with current technology and is complementary to efforts to obtain human structural variation information by other technologies. more...

Fosmid library information

HapMap Identifier	Population	Library Name	Status	End sequences submitted to Trace	Full insert sequences submitted to GenBank	Reference
NA15510	N/A	WI2 (G248)	Complete	2,298,885	411	Tuzun et el., 2005
NA18517	Yoruba	ABC7	Complete	2,152,975	157	Kidd et al., 2008
NA18507	Yoruba	ABC8	Complete	3,888,476	226	Kidd et al., 2008
NA18956	Japan	ABC9	Complete	2,084,892	722	Kidd et al., 2008
NA19240	Yoruba	ABC10	Complete	2,121,489	509	Kidd et al., 2008
NA18555	China	ABC11	Complete	1,966,644	387	Kidd et al., 2008
NA12878	CEPH	ABC12	Complete	2,169,280	454	Kidd et al., 2008
NA19129	Yoruba	ABC13	Complete	2,057,345	368	Kidd et al., 2008
NA12156	CEPH	ABC14	Complete	2,089,193	351	Kidd et al., 2008
NA18552	China	COR02,COR2A	Complete	1,992,678	180	
NA18947	Japan	ABC16	Ongoing	1,546,191	202	
NA18564	China	ABC17	Ongoing	56,944		
NA10847	CEPH	ABC18	Ongoing	1,209,419	65	
NA18573	China	ABC19	Ongoing	43,351		
NA19102	Yoruba	ABC20	Ongoing	89,566		
NA11993	CEPH	ABC21	Ongoing	684,716	41	
NA11840	CEPH	ABC22	Ongoing	785,461	1	
NA18523	Yoruba	ABC23	Ongoing	1,544,982		
NA18502	Yoruba	ABC24	Ongoing	1,388,082	261	
NA11832	CEPH	ABC25	Ongoing	12,286		
NA18861	Yoruba	ABC26	Ongoing	14,559		
NA18942	Japan	ABC27	Ongoing	1,234,412	170	

NHGRI Structural Variation Clone Viewer

http://www.ncbi.nlm.nih.gov/projects/genome/StructuralVariation/NHGRIStructuralVariation.shtml

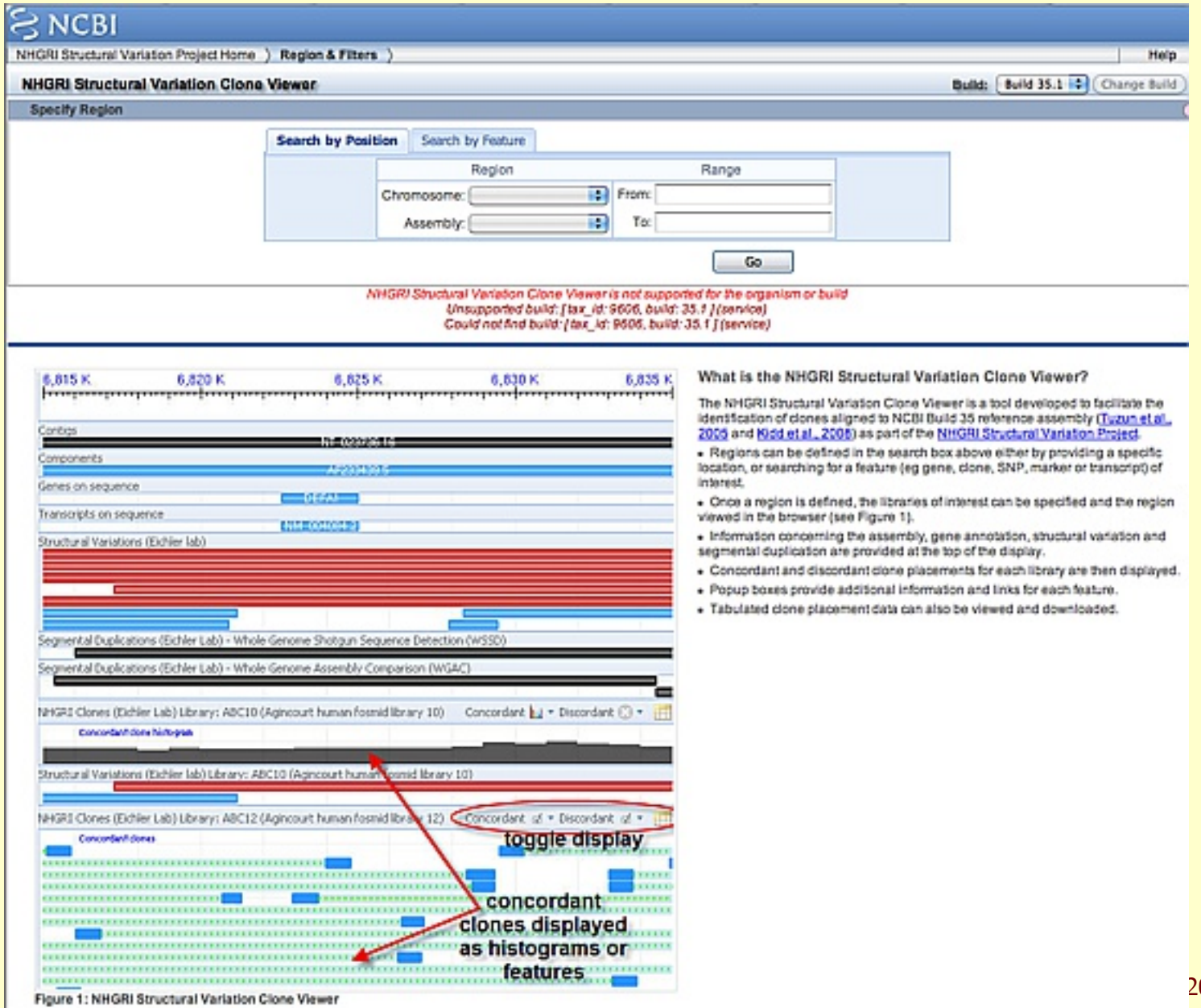
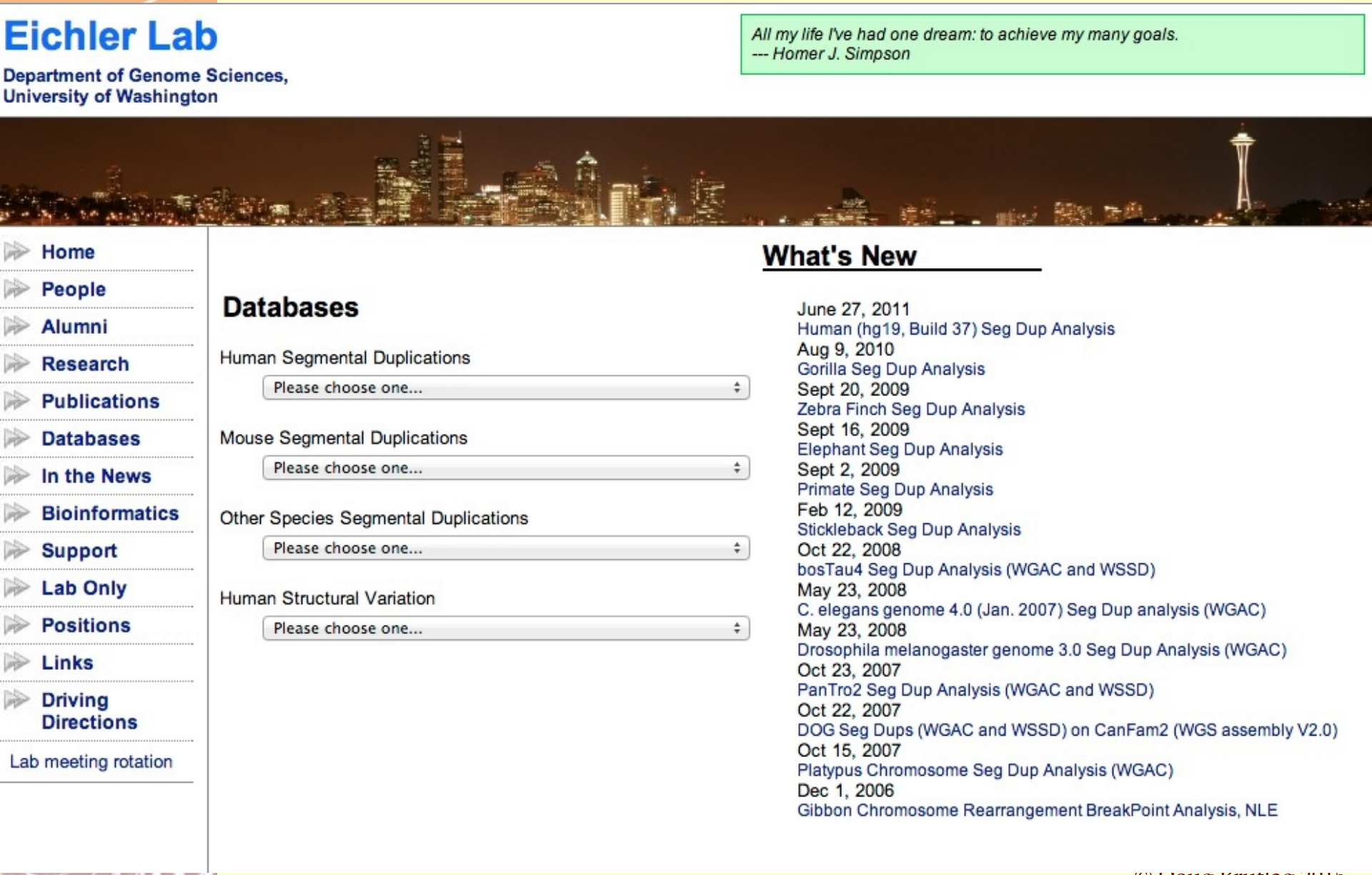


Figure 1: NHGRI Structural Variation Clone Viewer

Eichler Lab

http://eichlerlab.gs.washington.edu/database.html



Paired End Mapping (PEM)

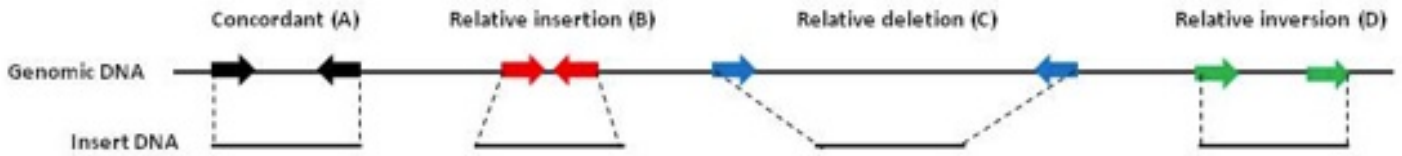


Figure 3: Paired-end mapping (PEM)

A library of known insert size e.g., 40kb fosmid sequences or 3kb DNA fragments is end sequenced and aligned to a genomic assembly.
(A) Ends that map at a similar distance and orientation to the genomic assembly are concordant and do not indicate any structural variation.
(B) Ends that map at a distance significantly less than the insert size on the genomic assembly indicate an insertion in the insert relative to the assembly.
(C) Ends that map at a distance significantly more than the insert size on the genomic assembly indicate an deletion in the insert relative to the assembly.
(D) Ends that map in the same orientation on the genomic assembly indicate an inversion relative to the assembly.

Copy Number Variation and Disease

Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans

Timothy J. Aitman[1], Rong Dong[1*], Timothy J. Vyse[2*], Penny J. Norsworthy[1*], Michelle D. Johnson[1], Jennifer Smith[3], Jonathan Mangion[1], Cheri Roberton-Lowe[1,2], Amy J. Marshall[1], Enrico Petretto[1], Matthew D. Hodges[1], Gurjeet Bhangal[3], Sheetal G. Patell[3], Kelly Sheehan-Rooney[3], Mark Duda[1,3], Paul R. Cook[1,3], David J. Evans[3], Jan Domin[3], Jonathan Flint[4], Joseph J. Boyle[5], Charles D. Pusey[3] & H. Terence Cook[5]

Nature, 2006

The Influence of *CCL3L1* Gene–Containing Segmental Duplications on HIV-1/AIDS Susceptibility

Enrique Gonzalez,[1*] Hemant Kulkarni,[1*] Hector Bolivar,[1*†] Andrea Mangano,[2*] Racquel Sanchez,[1‡] Gabriel Catano,[1‡] Robert J. Nibbs,[3‡] Barry I. Freedman,[4‡] Marlon P. Quinones,[1‡] Michael J. Bamshad,[5] Krishna K. Murthy,[6] Brad H. Rovin,[7] William Bradley,[8,9] Robert A. Clark,[1] Stephanie A. Anderson,[8,9] Robert J. O'Connell,[9,10] Brian K. Agan,[9,10] Seema S. Ahuja,[1] Rosa Bologna,[11] Luisa Sen,[2] Matthew J. Dolan,[9,10,12‡] Sunil K. Ahuja[1‡]

Science, 2005, **307**

A Chromosome 8 Gene-Cluster Polymorphism with Low Human Beta-Defensin 2 Gene Copy Number Predisposes to Crohn Disease of the Colon

Klaus Fellermann, Daniel E. Stange, Elke Schaeffeler, Hartmut Schmalzl, Jan Wehkamp, Charles L. Bevins, Walter Reinisch, Alexander Teml, Matthias Schwab, Peter Lichter, Bernhard Radlwimmer, and Eduard F. Stange

The American Journal of Human Genetics, 2006, **79**

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

Copy Number Variation and Disease 2008

Gene	Type	Duplicated Segment	Disease/Phenotype	
C4A/C4B	Decrease	32.8 kb	Lupus* (SLE)	Yang, 2007
DEFB4.103,104	Increase	310 kb	Psoriasis	Hollox, 2008
	Decrease		Crohn disease, IBD	Fellerman, 2006
CCL3L1	Decrease	64 kb	HIV susceptibility	Gonzalez, 2005
FCGR3B	Decrease	**	Glomeruloneprhitis	Aitman, 2006
				Fanciulli, 2008
IRGM	Deletion	**	Crohn disease	Parkes, 2007

**correspond to more ancient primate segmental duplications

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