Structural Variants in the Human Genome

Doug Brutlag
Professor Emeritus of Biochemistry & Medicine
Stanford University School of Medicine
2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007

“It’s all about me!”

Simple Nucleotide Polymorphisms (SNPs)

Individual 1

ACAGCCCA....

TTCGGGGTC....

Individual 2

ACAGCCCA....

TTCGAGGTC....

Individual 3

ACATGCCA....

TTCGGGGTC....

Individual 4

ACAGCCCA....

TTCGGGGTC....
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.
Henry Stewart Talks on Copy Number Variations

- Henry Stewart Talks  http://hstalks.com/
- Copy Number Variation
- Copy Number Variation by Prof. Stephen Scherer
- CNVs in human genomes by Prof. Chris Ponting
- The Future of CNVs: Sequence base resolution and links to human disease by Professor Evan Eichler – University of Washington
- You will need the Stanford name and password (stanford, member) in order to watch this course off campus.
1. Copy number variation (37 mins)
   Prof. Stephen W Scherer – Hospital for Sick Children and University of Toronto, Canada

2. Array comparative genomic hybridization to characterize copy number variation in the human genome (17 mins)
   Dr. Nigel Carter – The Wellcome Trust Sanger Institute, UK

3. CNVs in human genomes (32 mins)
   Prof. Chris Ponting – University of Oxford, UK

4. Gene copy number variation in human and primate evolution (32 mins)
   Prof. James Sikela – University of Colorado, Denver, USA

5. Population genetics of structural variation (26 mins)
   Dr. Don Conrad – Wellcome Trust Sanger Institute, Cambridge, UK

6. Genomic disorders: mechanisms for copy number variation and clinical implementation of high-resolution genome analysis (64 mins)
   Prof. James Lupski – Baylor College of Medicine, USA

7. Databases for CNV in control and disease populations (47 mins)
   Dr. Lars Feuk – Uppsala University, Sweden

**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 $\geq 10$ kb and sequence identity $\geq 95\%$). 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>
<thead>
<tr>
<th>Variation</th>
<th>Rearrangement type</th>
<th>Size rangea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single base-pair changes</td>
<td>Single nucleotide polymorphisms, point mutations</td>
<td>1 bp</td>
</tr>
<tr>
<td>Small insertions/deletions</td>
<td>Binary insertion/deletion events of short sequences (majority &lt;10 bp in size)</td>
<td>1–50 bp</td>
</tr>
<tr>
<td>Short tandem repeats</td>
<td>Microsatellites and other simple repeats</td>
<td>1–500 bp</td>
</tr>
<tr>
<td>Fine-scale structural variation</td>
<td>Deletions, duplications, tandem repeats, inversions</td>
<td>50 bp to 5 kb</td>
</tr>
<tr>
<td>Retroelement insertions</td>
<td>SINEs, LINEs, LTRs, ERVs</td>
<td>300 bp to 10 kb</td>
</tr>
<tr>
<td>Intermediate-scale structural variation</td>
<td>Deletions, duplications, tandem repeats, inversions</td>
<td>5 kb to 50 kb</td>
</tr>
<tr>
<td>Large-scale structural variation</td>
<td>Deletions, duplications, large tandem repeats, inversions</td>
<td>50 kb to 5 Mb</td>
</tr>
<tr>
<td>Chromosomal variation</td>
<td>Euchromatic variants, large cytogenetically visible deletions, duplications, translocations, inversions, and aneuploidy</td>
<td>~5 Mb to entire chromosomes</td>
</tr>
</tbody>
</table>
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
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
Unequal Crossing Over Leads to Duplication and Deletion
Intra-Chromosomal Crossing Over Leads to Deletion

Inter-Chromosomal Crossing Over Leads to Inversion

Intra-Chromosomal Crossing Over Can Also Lead to Inversion

Deletions and Insertions at Repeat Sequences
Variations in Tandem Repeat Arrays
Human $\alpha$-Amylase Gene Repeats

FISH on DNA
8 or 12 tandem repeats 4 kb long

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 disorders such as autism, schizophrenia, attention deficit, learning disabilities and many other neurological and behavioral disorders
### Copy Number Variation and Disease 2002

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

**Diagram:**
- **Increase**
  - Green arrows pointing to the right
- **Decrease**
  - Green arrows pointing to the left
Comparative Genomics Hybridization (CGH)
Comparative Micro Arrays (CMA) Using Genome Tiling Arrays

- **5 Mb**
- **1 Mb**
- **75 Kb**
- **800 bp**
- **25-36mer**

- **Genome scanning array**

- **CGH BAC array**

- **Amplicon tiling array**

- **Oligonucleotide array**

Courtesy of Mike Snyder
Detection of Duplications and Deletions Using Chromosomal Micro-Arrays

10.9 Mbase deletion at 7q11 in Williams-Beuren Syndrome

7.2 Mbase duplication in 11q

Mapping Breakpoints of Partial Trisomies of Chromosome 21

Segmentation for 48085_532_48085_635 (48085_990bp) - chr21; with 3 segments

Segmentation for 43089_532_43089_635 (43089_990bp) - chr21; with 1 segments

Segmentation for 87173_532_87173_635 (87173_990bp) - chr21; with 3 segments

Courtesy of Mike Snyder
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.
Sequence Base Resolution of Structural Variation

Gametic library (1 million clones) 
Sequence ends of genomic inserts and map to human genome

Concordant Insertion Deletion Inversions

Fosmid

> < > < > < < <

Build35

Henry Stewart Talks: Evan Eichler

© Doug Brutlag 2015
Fine Scale Structural Variation
8.8 x Coverage of a Human 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

- 89 locate within gene regions
- 138 unique regions of the genome
- 159 duplicated regions of the genome

Henry Stewart Talks: Evan Eichler

© Doug Brutlag 2015
A Structural Variation Map of Eight Human Genomes

Henry Stewart Talks: Evan Eichler

© Doug Brutlag 2015
Genomics Distribution of CNV Regions
Heterogeneity in Olfactory Receptor Genes
(Examined 851 Olfactory Receptor Loci)

CNVs affect:
- 93 duplicated genes
- 151 deleted genes

Clos Vougeot in Bourgogne
Chef d’Ordre de la Confrérie des Chevalier du Tastevins
Charcot-Marie-Tooth Hereditary Neuropathy (CMT1) Disease
Results From CNV of PMP22 Gene in 17p11.2-12

Figure 1: Charcot-Marie Tooth (CMT) disease
Unequal crossing over between two highly homologous repeats on chromosome 17p12 can result in (A) 3 copies of the PMP22 gene with the CMT1A phenotype or the reciprocal (B) and 1 copy of the PMP22 gene with the HNPP phenotype.

Peripheral Neuropathy, Yuen So, Medical Grand Rounds Jan 16, 2012

© Doug Brutlag 2015
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>
<thead>
<tr>
<th>Locus Name</th>
<th>Proportion of CMT1 (excluding CMTX)</th>
<th>Gene Symbol</th>
<th>Protein Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMT1A</td>
<td>70%-80%</td>
<td>PMP22</td>
<td>Peripheral myelin protein 22</td>
</tr>
<tr>
<td>CMT1B</td>
<td>10%-12%</td>
<td>MPZ</td>
<td>Myelin P₀ protein</td>
</tr>
<tr>
<td>CMT1C</td>
<td>~1%</td>
<td>LITAF</td>
<td>Lipopolysaccharide-induced tumor necrosis factor-alpha factor</td>
</tr>
<tr>
<td>CMT1D</td>
<td>Unknown</td>
<td>EGR2</td>
<td>Early growth response protein 2</td>
</tr>
<tr>
<td>CMT1E</td>
<td>~1%</td>
<td>PMP22</td>
<td>Peripheral myelin protein 22 (sequence changes)</td>
</tr>
<tr>
<td>CMT1F/2E</td>
<td>Unknown</td>
<td>NEFL</td>
<td>Neurofilament light polypeptide</td>
</tr>
</tbody>
</table>
CMT Hereditary Neuropathy Disease Genes

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

Schwann Cell proteins

Attachment proteins

Axon proteins

Axon surface proteins
### Structural Variations in Mendelian Disease

<table>
<thead>
<tr>
<th>Gene name(s)</th>
<th>Locus</th>
<th>Population frequency</th>
<th>Diploid copies</th>
<th>Size of variant segment</th>
<th>Associated phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSTM1</td>
<td>1p13.3</td>
<td>&gt;3%</td>
<td>1–3</td>
<td>18 kb</td>
<td>Altered enzyme activity</td>
</tr>
<tr>
<td>RHD</td>
<td>1p36.11</td>
<td>15–20%</td>
<td>0–2</td>
<td>~60 kb</td>
<td>Rhesus blood group sensitivity</td>
</tr>
<tr>
<td>SMN2</td>
<td>5q13.2</td>
<td>~60%</td>
<td>1–4</td>
<td>500 kb</td>
<td>Altered severity of spinal muscular atrophy</td>
</tr>
<tr>
<td>CYP21A2</td>
<td>6p21.32</td>
<td>1.6%</td>
<td>2–3</td>
<td>35 kb</td>
<td>Congenital adrenal hyperplasia</td>
</tr>
<tr>
<td>LPA</td>
<td>6q25.3</td>
<td>94%</td>
<td>2–38</td>
<td>5.5 kb</td>
<td>Altered coronary heart disease risk</td>
</tr>
<tr>
<td>α-Defensin gene</td>
<td>8p23.1</td>
<td>~90%</td>
<td>4–14</td>
<td>19 kb</td>
<td>Immune system function</td>
</tr>
<tr>
<td>β-Defensin gene</td>
<td>8p23.1</td>
<td>~90%</td>
<td>2–12</td>
<td>240 kb</td>
<td>Immune system function</td>
</tr>
<tr>
<td>IGHG1 region</td>
<td>1q32.33</td>
<td>12–74%</td>
<td>1–6</td>
<td>5–170 kb</td>
<td>Immune system function?</td>
</tr>
<tr>
<td>CCL3-L1/CCL4-L1</td>
<td>17q12</td>
<td>51%/27%</td>
<td>0–14</td>
<td>&gt;2 kb</td>
<td>Susceptibility to and progression of HIV infection, susceptibility to Kawasaki disease</td>
</tr>
<tr>
<td>CYP2A6</td>
<td>19q13.2</td>
<td>1.7%</td>
<td>2–3</td>
<td>7 kb</td>
<td>Altered nicotine metabolism</td>
</tr>
<tr>
<td>IGL</td>
<td>22q11.22</td>
<td>28–85%</td>
<td>2–7</td>
<td>5.4 kb</td>
<td>Altered Igκ-Igλ in B lymphocytes</td>
</tr>
<tr>
<td>GSTT1</td>
<td>22q11.23</td>
<td>20%</td>
<td>0–2</td>
<td>&gt;50 kb</td>
<td>Altered susceptibility to toxins and cancer</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>22q13.1</td>
<td>1–29%</td>
<td>0–13</td>
<td>Undefined</td>
<td>Altered drug metabolism, increased cancer susceptibility</td>
</tr>
<tr>
<td>OPN1LW1/OPN1MW</td>
<td>Xq28</td>
<td>75%</td>
<td>0–4/0–7</td>
<td>15 kb/13 kb</td>
<td>Defective color vision</td>
</tr>
<tr>
<td>Testis-specific genes (DAZ, BPP, RBM families)</td>
<td>Yq11.2</td>
<td>3.2%</td>
<td>0–1</td>
<td>1.6 Mb</td>
<td>Low-penetrance spermatogenic failure</td>
</tr>
</tbody>
</table>

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

### Table 1. Novel Recurrent Copy-Number Changes Associated with Intellectual Disability and Related Disorders.*

<table>
<thead>
<tr>
<th>Chromosome Region</th>
<th>Coordinates in Mb†</th>
<th>Deletion or Duplication Associated with Disorder</th>
<th>Selected References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1</td>
<td>Chromosome 1: 145.0–146.35</td>
<td>Deletion: intellectual disability, schizophrenia, multiple congenital anomalies; Duplication: intellectual disability, autism</td>
<td>Brunetti-Pierri et al.,* Mefford et al., International Schizophrenia Consortium, Stefansson et al., Greenway et al., Haldeman-Englert and Jewett.</td>
</tr>
<tr>
<td>3q29</td>
<td>Chromosome 3: 197.4–198.9</td>
<td>Deletion: intellectual disability; Duplication: intellectual disability</td>
<td>Ballif et al., Lisi et al., Willatt et al.</td>
</tr>
<tr>
<td>10q22-q23</td>
<td>Chromosome 10: 81.12–89.07</td>
<td>Deletion: intellectual disability</td>
<td>Balciuniene et al., van Bon et al.</td>
</tr>
<tr>
<td>15q11.2</td>
<td>Chromosome 15: 20.3–20.7</td>
<td>Deletion: intellectual disability, schizophrenia, epilepsy</td>
<td>Stefansson et al., de Kovel et al., Mefford et al., Burnside et al., Doornbos et al., Murthy et al., von der Lippe et al.</td>
</tr>
<tr>
<td>15q13.3</td>
<td>Chromosome 15: 28.7–30.2</td>
<td>Deletion: intellectual disability, epilepsy, schizophrenia, autism</td>
<td>Stefansson et al., Helbig et al., Sharp et al., van Bon et al., Ben-Shachar et al., Pagnamenta et al., Miller et al.</td>
</tr>
<tr>
<td>15q24</td>
<td>Chromosome 15: 72.2–73.8</td>
<td>Deletion: intellectual disability, autism</td>
<td>Andreix et al., Sharp et al., Mefford et al., El-Hattab et al.</td>
</tr>
<tr>
<td>16p11.2 (a)</td>
<td>Chromosome 16: 29.5–30.1</td>
<td>Deletion: intellectual disability, autism, obesity; Duplication: schizophrenia</td>
<td>Weiss et al., Battaglia et al., Bijlsma et al., Hempel et al., Shinawi et al., Jacquemont et al., Walters et al., McCarthy et al.</td>
</tr>
<tr>
<td>16p11.2 (b)</td>
<td>Chromosome 16: 28.7–29.0</td>
<td>Deletion: intellectual disability, obesity</td>
<td>Bachmann-Gagescu et al., Bochukova et al.</td>
</tr>
<tr>
<td>16p12</td>
<td>Chromosome 16: 21.8–22.4</td>
<td>Deletion: intellectual disability</td>
<td>Girirajan et al.</td>
</tr>
<tr>
<td>16p13.11</td>
<td>Chromosome 16: 15.4–16.4</td>
<td>Deletion: intellectual disability, epilepsy, autism, schizophrenia; Duplication: intellectual disability, ADHD, autism</td>
<td>de Kovel et al., Mefford et al., Heinzen et al., Williams et al., Ullmann et al., Kirov et al.</td>
</tr>
<tr>
<td>17q12</td>
<td>Chromosome 17: 31.8–33.3</td>
<td>Deletion: intellectual disability, autism, schizophrenia</td>
<td>Moreno-De-Luca et al., Loirat et al.</td>
</tr>
<tr>
<td>17q21.3</td>
<td>Chromosome 17: 41.0–41.7</td>
<td>Deletion: intellectual disability</td>
<td>Koolen et al., Shaw-Smith et al., Koolen et al.</td>
</tr>
</tbody>
</table>

* The listed recurrent deletions and duplications are those that have been reported since 2006. ADHD denotes attention deficit–hyperactivity disorder.
† The coordinates are based on the National Center for Biotechnology Information (NCBI) build 36.
Next Generation Sequencing to Identify Genes Associated with Learning Disability

<table>
<thead>
<tr>
<th>Study</th>
<th>Disorder</th>
<th>Presumed Inheritance</th>
<th>Type of Analysis</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ng et al.(^97)</td>
<td>Kabuki syndrome</td>
<td>De novo dominant</td>
<td>Multiple affected</td>
<td>MLL2</td>
</tr>
<tr>
<td>Hoischen et al.(^98)</td>
<td>Schinzel–Giedion syndrome</td>
<td>De novo dominant</td>
<td>Multiple affected</td>
<td>SETBP1</td>
</tr>
<tr>
<td>Vissers et al.(^99)</td>
<td>Nonsyndromic sporadic intellectual disability</td>
<td>De novo dominant</td>
<td>Trio</td>
<td>Multiple</td>
</tr>
<tr>
<td>Najmabadi et al.(^100)</td>
<td>Recessive intellectual disability</td>
<td>Autosomal recessive, consanguineous families</td>
<td>Targeted recessive</td>
<td>Multiple</td>
</tr>
<tr>
<td>Çalışkan et al.(^101)</td>
<td>Recessive intellectual disability</td>
<td>Autosomal recessive, consanguineous family</td>
<td>Recessive</td>
<td>TECR</td>
</tr>
<tr>
<td>O’Roak et al.(^102)</td>
<td>Autism</td>
<td>De novo dominant</td>
<td>Trio</td>
<td>FOXP1, GRIN2B, SCN1A, LAMC3</td>
</tr>
</tbody>
</table>
### Inversions Lead to Instability & Disease

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

dbVAR Report on PMP22 Gene


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

<table>
<thead>
<tr>
<th>ID</th>
<th>Placement Type</th>
<th>Assembly</th>
<th>Placement</th>
<th>Start</th>
<th>Stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_000017.9</td>
<td>Submitted Genomic</td>
<td>NCBI36 (hg18)</td>
<td>Chr17</td>
<td>14,040,467</td>
<td>15,435,991</td>
</tr>
</tbody>
</table>

Supporting Variants

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Allele Length</th>
<th>Sample ID</th>
<th>Subject Phenotype</th>
<th>Assembly</th>
<th>Placement</th>
<th>Start</th>
<th>Stop</th>
<th>Placement Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>nssv867002</td>
<td>Gain</td>
<td>1395524</td>
<td>Not reported</td>
<td>NCBI36 (hg18)</td>
<td>Chr17</td>
<td>14,040,467</td>
<td>15,435,991</td>
<td>Submitted Genomic</td>
<td></td>
</tr>
</tbody>
</table>
Homo sapiens chromosome 17, reference assembly, complete sequence

NCBI Reference Sequence: NC_000017.9

This sequence has been updated. See current version.

nc_000017.9 (78,774,742 bases)

nc_342,705 : 16,054,010 (2,711,306 bases shown, positive strand)

Tiling Path (Components)

nstd54: Cooper et al 2011 Structural Variation

nsyv574389

nsyv574389
Database of Genomics Variants

A curated catalogue of structural variation in the human genome

About The Project | Genome Browser | Download | Links | Data Submissions | Email us

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

View Data by Chromosome

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y All

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:

BLAT Search

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

The New DGV Beta Site has launched! Access it here
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.

Increase in Variation Data

Number of Entries

Years
Database of Genomics Variants
http://projects.tcag.ca/variation/
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.
Showing 5 Mbp from chr17, positions 12,649,493 to 17,649,492
NHGRI Structural Variation Project


### 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

<table>
<thead>
<tr>
<th>HapMap Identifier</th>
<th>Population</th>
<th>Library Name</th>
<th>Status</th>
<th>End sequences submitted to Trace</th>
<th>Full Insert sequences submitted to GenBank</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA15510</td>
<td>N/A</td>
<td>WI2 (G248)</td>
<td>Complete</td>
<td>2,298,885</td>
<td>322</td>
<td>Tuzun et al., 2005</td>
</tr>
<tr>
<td>NA18517</td>
<td>Yoruba</td>
<td>ABC7</td>
<td>Complete</td>
<td>2,152,975</td>
<td>115</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA18507</td>
<td>Yoruba</td>
<td>ABC8</td>
<td>Complete</td>
<td>3,888,476</td>
<td>169</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA18956</td>
<td>Japan</td>
<td>ABC9</td>
<td>Complete</td>
<td>2,084,892</td>
<td>651</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA19240</td>
<td>Yoruba</td>
<td>ABC10</td>
<td>Complete</td>
<td>2,121,489</td>
<td>385</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA18555</td>
<td>China</td>
<td>ABC11</td>
<td>Complete</td>
<td>1,966,644</td>
<td>313</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA18278</td>
<td>CEPH</td>
<td>ABC12</td>
<td>Complete</td>
<td>2,169,280</td>
<td>312</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA19129</td>
<td>Yoruba</td>
<td>ABC13</td>
<td>Complete</td>
<td>2,057,345</td>
<td>257</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA12156</td>
<td>CEPH</td>
<td>ABC14</td>
<td>Complete</td>
<td>2,089,193</td>
<td>206</td>
<td>Kidd et al., 2008</td>
</tr>
<tr>
<td>NA18552</td>
<td>China</td>
<td>JCVI*</td>
<td>Complete</td>
<td>1,992,678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18947</td>
<td>Japan</td>
<td>ABC16</td>
<td>Ongoing</td>
<td>1,546,191</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>NA18564</td>
<td>China</td>
<td>ABC17</td>
<td>Ongoing</td>
<td>56,944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA10847</td>
<td>CEPH</td>
<td>ABC18</td>
<td>Ongoing</td>
<td>1,209,419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18573</td>
<td>China</td>
<td>ABC19</td>
<td>Ongoing</td>
<td>43,351</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA19102</td>
<td>Yoruba</td>
<td>ABC20</td>
<td>Ongoing</td>
<td>89,566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA11993</td>
<td>CEPH</td>
<td>ABC21</td>
<td>Ongoing</td>
<td>684,716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA11840</td>
<td>CEPH</td>
<td>ABC22</td>
<td>Ongoing</td>
<td>785,461</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18523</td>
<td>Yoruba</td>
<td>ABC23</td>
<td>Ongoing</td>
<td>1,544,982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18502</td>
<td>Yoruba</td>
<td>ABC24</td>
<td>Ongoing</td>
<td>1,388,082</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>NA1832</td>
<td>CEPH</td>
<td>ABC25</td>
<td>Ongoing</td>
<td>12,286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18861</td>
<td>Yoruba</td>
<td>ABC26</td>
<td>Ongoing</td>
<td>14,559</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA18942</td>
<td>Japan</td>
<td>ABC27</td>
<td>Ongoing</td>
<td>1,234,412</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

* The JCVI library is comprised of 4 libraries: COR01, COR02, COR2A and COR03

© Doug Brutlag 2015
NHGRI Structural Variation Clone Viewer

Eichler Lab

http://eichlerlab.gs.washington.edu/database.html
## Copy Number Variation and Disease 2008

<table>
<thead>
<tr>
<th>Gene</th>
<th>Type</th>
<th>Duplicated Segment</th>
<th>Disease/Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4A/C4B</td>
<td>Decrease</td>
<td>32.8 kb</td>
<td>Lupus* (SLE)</td>
<td>Yang, 2007</td>
</tr>
<tr>
<td>DEFB4.103,104</td>
<td>Increase</td>
<td>310 kb</td>
<td>Psoriasis</td>
<td>Hollox, 2008</td>
</tr>
<tr>
<td></td>
<td>Decrease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL3L1</td>
<td>Decrease</td>
<td>64 kb</td>
<td>HIV susceptibility</td>
<td>Gonzalez, 2005</td>
</tr>
<tr>
<td>FCGR3B</td>
<td>Decrease</td>
<td>**</td>
<td>Glomerulonephritis</td>
<td>Aitman, 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fanciulli, 2008</td>
</tr>
<tr>
<td>IRGM</td>
<td>Deletion</td>
<td>**</td>
<td>Crohn disease</td>
<td>Parkes, 2007</td>
</tr>
</tbody>
</table>

**correspond to more ancient primate segmental duplications
Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans

Nature, 2006

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

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

The American Journal of Human Genetics, 2006, 79