Structure and Analysis of Genetic Variation

Sven Cichon, PhD

Institute of Neuroscience and Medicine (INM-1), Research Center Jülich

Institute of Human Genetics, University of Bonn
Genome variation visible under the microscope already....
....but it gets enormous at the submicroscopic level

Sequence variation

- Single nucleotide
  - Base change – substitution – point mutation
  - Insertion-deletions ("indels")
  - SNPs – tagSNPs

2 bp to 1,000 bp
- Microsatellites, minisatellites
- Indels
- Inversions
- Di-, tri-, tetranucleotide repeats
- VNTRs

1 kb to submicroscopic
- Copy number variants (CNVs)
- Segmental duplications
- Inversions, translocations
- CNV regions (CNVRs)
- Microdeletions, microduplications

Structural variation

- Microscopic to subchromosomal
  - Segmental aneusomy
  - Chromosomal deletions – losses
  - Chromosomal insertions – gains
  - Chromosomal inversions
  - Intrachromosomal translocations
  - Chromosomal abnormality
  - Heteromorphisms
  - Fragile sites

Whole chromosomal to whole genome
- Interchromosomal translocations
- Ring chromosomes, isochromosomes
- Marker chromosomes
- Aneuploidy
- Aneusomy

Molecular genetic detection

Cytogenetic detection

Scherer et al., 2007
Scherer et al., 2007

Sequence variation

- Single nucleotide
  - Base change – substitution – point mutation
  - Insertion-deletions (“indels”)
  - SNPs – tagSNPs

Structural variation

- Di-, tri-, tetranucleotide repeats
- VNTRs

1 kb to submicroscopic
- Copy number variants (CNVs)
- Segmental duplications
- Inversions, translocations
- CNV regions (CNVRs)
- Microdeletions, microduplications

Microscopic to subchromosomal
- Segmental aneusomy
- Chromosomal deletions – losses
- Chromosomal insertions – gains
- Chromosomal inversions
- Intrachromosomal translocations
- Chromosomal abnormality
- Heteromorphisms
- Fragile sites

Whole chromosomal to whole genome
- Interchromosomal translocations
- Ring chromosomes, isochromosomes
- Marker chromosomes
- Aneuploidy
- Aneusomy

Molecular genetic detection

Scherer et al., 2007
Single nucleotide polymorphism (SNP)

C-allele: 70% frequency

C = major allele

T-allele: 30% frequency

T = minor allele
**How many SNPs in the human genome?**

<table>
<thead>
<tr>
<th>minor allele frequency</th>
<th>No. of SNPs (Mio.)</th>
<th>SNP density (1 SNP/bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.0</td>
<td>290</td>
</tr>
<tr>
<td>5</td>
<td>7.1</td>
<td>450</td>
</tr>
<tr>
<td>10</td>
<td>5.3</td>
<td>600</td>
</tr>
<tr>
<td>20</td>
<td>3.3</td>
<td>960</td>
</tr>
<tr>
<td>30</td>
<td>2.0</td>
<td>1,570</td>
</tr>
<tr>
<td>40</td>
<td>0.97</td>
<td>3,280</td>
</tr>
</tbody>
</table>

Based on mutation rate and population size it can be assumed that every base pair of the human genome exists in a mutated form in at least several individuals.
Fate of new mutations

The diagram illustrates the fate of new mutations over time (generations). The x-axis represents time in generations, and the y-axis shows the frequency of mutations. Different lines represent the progression of various mutations, with some reaching 100% frequency while others do not. The red dashed line indicates a significant event or transition point in the process.
„connected“ SNPs - „linkage disequilibrium“ (LD)
The block structure of the human genome

- **Block 1**
  - Hot-Spot 1

- **Block 2**
  - Hot-Spot 2

- **Block 3**

**SNPs**

**Frequent Haplotypes (htSNPs indicated)**
Tagging (or “tag”) SNPs in Haplotype Blocks

Linkage Disequilibrium

Physical Distance

European population

African population

Block 1

Block 2
How many SNPs needed to cover most of the common variation?

~300,000 tag SNPs needed to cover common variation genome-wide in Europeans at 95% level.
The HapMap Projects identifies all the tag SNPs in the genome

Goals:

- Define patterns of genetic variation across human genome in different populations (CEU, CHB, JPT, YRI), is being extended in phase 3
- Guide selection of SNPs efficiently to “tag” common variants
- Public release of all data (assays, genotypes)
THE HAPMAP PROJECT

Chapter and verse on human genetic variation
Tag SNPs from HapMap enable systematic candidate gene studies.
And if you don’t have good candidate genes for your brain imaging phenotype? Screen the whole genome (GWAS)
And if you don’t have good candidate genes for your brain imaging phenotype? Screen the whole genome using **SNP arrays**
SNP-Arrays

Scan of an individual’s DNA with an array harbouring a genome wide set of 550,000 tag SNP markers (Illumina)
2 bp to 1,000 bp
- Microsatellites, minisatellites
  → Indels
- Inversions
- Di-, tri-, tetranucleotide repeats
- VNTRs

Microscopic to subchromosomal
- Inversions, translocations
  → CNV regions (CNVRs)
- Microdeletions, microduplications

Whole chromosomal to whole genome
- Interchromosomal translocations
- Ring chromosomes, isochromosomes
- Marker chromosomes
  → Aneuploidy
  → Aneusomy

Scherer et al., 2007
5-HTTLPR (serotonin-transporter-linked polymorphic region): a degenerate repeat polymorphism
Sequence variation:
- Single nucleotide
  - Base change – substitution – point mutation
  - Insertion-deletions ("indels")
  - SNPs – tagSNPs

2 bp to 1,000 bp
- Microsatellites, minisatellites
  - Indels
  - Inversions
  - Di-, tri-, tetranucleotide repeats
  - VNTRs

1 kb to submicroscopic
- Copy number variants (CNVs)
- Segmental duplications
  - Inversions, translocations
  - CNV regions (CNVRs)
  - Microdeletions, microduplications

Structural variation:
- Chromosomal insertions – gains
- Chromosomal inversions
- Intrachromosomal translocations
- Chromosomal abnormality
  - Heteromorphisms
  - Fragile sites

Whole chromosomal to whole genome
- Interchromosomal translocations
- Ring chromosomes, isochromosomes
- Marker chromosomes
  - Aneuploidy
  - Aneusomy

Molecular genetic detection

Scherer et al., 2007
Copy number variants (CNVs) - Definition

- Variable presence or absence of a DNA sequence >1000 bp

- Different types of CNVs:
  - Deletions
  - Duplications/Triplications
  - Insertions

Diagram:

- Normal: A B C D E F G H
- Deletion: A B C {F G H}
- Duplication: A B C {D E} {D E} F G H
### Copy number variants (CNVs) - Definition

#### A) Deletions

<table>
<thead>
<tr>
<th>Copy number</th>
<th>i)</th>
<th>ii)</th>
<th>iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><img src="image" alt="Deletion" /></td>
<td><img src="image" alt="Deletion" /></td>
<td><img src="image" alt="Deletion" /></td>
</tr>
<tr>
<td>1</td>
<td><img src="image" alt="Deletion" /></td>
<td><img src="image" alt="Deletion" /></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td><img src="image" alt="Deletion" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### B) Duplications

<table>
<thead>
<tr>
<th>Copy number</th>
<th>i)</th>
<th>ii)</th>
<th>iii)</th>
<th>vi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><img src="image" alt="Duplication" /></td>
<td><img src="image" alt="Duplication" /></td>
<td><img src="image" alt="Duplication" /></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Duplication" /></td>
<td><img src="image" alt="Duplication" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><img src="image" alt="Duplication" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wain *et al.*, 2009
Larger CNVs (>50-100 kb) can be detected on SNP arrays.
Comparative genome hybridisation (CGH) is more sensitive

Array with reference sequence (e.g. BAC, oligo)

Fetal sample
Reference sample
Mix

Duplicate Gain
Deletion Loss
CNVs reported in the **Database of Genomic Variants (DGV)**: n=15,963 (freeze Nov 02, 2010)
Molecular mechanisms leading to CNV phenotype

A) gene dosage

B) gene interruption

C) gene fusion

Lupski & Stankiewicz, 2005
How CNVs occur – non-allelic homologous recombination (NAHR)

LCR – Low Copy Repeat (highly similar DNA sequences)

Malhotra & Sebat, 2012
New sequencing technologies may help to identify variants relevant for imaging phenotypes not detected so far

Next-Generation-Sequencing / Third-Generation-Sequencing

Enable identification of phenotype-relevant rare SNPs or CNVs/Indels
Methylation of chromatin or DNA – a mechanism relevant to brain imaging phenotypes?
Chromatin Methylation

DNA methylation and chromatin modifications are mediating epigenetic phenomena in animals and plants.
Promoter Methylation

- Unmethylated
- Methylated

CpG Island → Gene Expression

CpG Island → Gene Expression Repressed
Chemical pollutants, dietary components, temperature changes and other external stresses can have long-lasting effects on development, metabolism and health, sometimes (shown for plants) even in subsequent generations.

**Methylation changes**
Methylation detection: bisulfite sequencing
Summary

- The genome is highly variable.
- Especially common SNP variants as well as large structural variants (CNVs) can be tested using array-based technologies.
- The field is moving to whole-genome sequencing which allows also detection of rare SNPs and small CNVs/InDels
- Investigation of DNA-methylation is rapidly evolving and may help understand molecular mechanisms relevant for imaging phenotypes (although not a heritable component!)