Molecular basis of genetic variation

Trygve Bakken
Department of Neurosciences
University of California, San Diego

presented by
Thomas Nichols, PhD
Department of Statistics & Warwick Manufacturing Group
University of Warwick

OHBM 2011 - Introduction to Imaging Genetics - 26 June, 2011
Outline

• What is DNA?

• Transcription & Translation (DNA → RNA → proteins)

• Genetic Recombination

• Linkage Disequilibrium & Polymorphisms
Genomic Era

- Sequenced ~6400 species
- 130 mammalian sequencing projects
- $1000 human genome
- Synthetic genomes
History of Deoxyribonucleic acid - DNA

• 1865, Mendel: Theory of heredity published
• 1869, Miescher: First isolated DNA in pus
• 1928, Griffith: Suggests that DNA carries genetic information
• 1952, Hersey & Chase: DNA is genetic material of a virus
• 1953, Crick & Watson: Suggest double-helix structure
• 1958, Nirenberg, Mathaei, Ochoa: DNA synthesised
• 1973, Cohen, Chang, Boyer: DNA transferred between organisms
1953 Watson and Crick model of DNA

1. Two complementary strands
2. Four nucleotide bases: ATGC
3. Sugar-phosphate backbone
Watson-Crick base pairing

- Bases: Adenine, Thymine, Guanine, Cytosine
- 2 or 3 hydrogen bonds link complementary bases
• Nucleotide is asymmetrical – C3 hydroxyl, C5 triphosphate

• 5' → 3' strand orientation
DNA packing – Nucleosomes

- ~3.2 billion bases
  ~2 meters long by $2.5 \times 10^{-9}$ meters wide
- ~150 bases wound around histones form nucleosomes
- 10-fold compression
DNA packing – Fibers

- Nucleosomes arranged into 30 nm fibers
- 50-fold compression
- Fibers fold to form ‘fractal globules’ at the scale of millions of bases
- Maximally dense, but still easy to fold and unfold
DNA packing – Chromosomes

- 8000-fold compression of DNA during cell division
- 22 pairs of autosomes and 2 sex chromosomes = 46 total
Central dogma of molecular biology

“once (sequential) information has passed into protein it cannot get out again”
– Crick, 1958

DNA → DNA
DNA → RNA
RNA → Protein

Replication
Transcription
Translation

From genes to proteins

- Genes are segments of DNA that often contain information to make proteins.
- Coding regions of our 20–25,000 genes compose ~1% of our genome.
- 98% of our DNA has unknown function.

NCBI, A Science Primer
Transcription of DNA into RNA

Ribonucleic acid (RNA), double helix like DNA, except

DNA bases: Adenine (A) Thymine (T) Guanine (G) Cytosine (C) A – T G – C
RNA bases: Adenine (A) Uracil (U) Guanine (G) Cytosine (C) A – U G – C

1. Initiation: transcription factors mediate RNA polymerase (RNAP) binding

2. Elongation: RNAP reads template DNA strand and forms messenger RNA (pre-mRNA)

3. Termination: single-stranded pre-mRNA is released
Splicing messenger RNA (mRNA)

- Exons contain information to make protein
- Spliceosome removes introns from pre-mRNA
- Alternate splicing generates protein diversity

Li 2007

Lee, Molecular Biology
Overview of protein translation

• mRNA is transported out of nucleus
• mRNA is read in 3-base units (codons) that specify amino acids
• Chain of amino acids = polypeptide
• Polypeptides fold to form proteins
Ribosomes translate mRNA

- Ribosome = 65% RNA + 35% protein

- Transfer RNA (tRNA) binds mRNA codon and adds amino acid to growing peptide chain
Genetic code specifies rules to make proteins

- Genetic code is unambiguous and redundant
- Code is highly conserved across all organisms
Molecular diversity of 20 amino acids

- Size
- Charge
- Hydrophobicity

Adapted from Taylor 1986
Secondary protein structure

- Completely specified by amino acid sequence
- Folded into conformation of lowest energy

α-helix

β-sheet

Kavraki 2007
Structure of an excitatory neurotransmitter receptor (GluA2)

- Primary structure
  - Amino acid sequence

- Secondary structure
  - Local folding

- Tertiary structure
  - 3D geometry of individual subunits

- Quaternary structure
  - 3D arrangement of subunits

Sobolevsky 2009
DNA replication is the third general mechanism of information transfer.

- Similar mechanism to transcription but with 2 strands
- Leading strand - DNA replication **continuous**, $5' \rightarrow 3'$
- Lagging strand - replication **discontinuous**, fragments ligated
- DNA polymerase has high fidelity (<1 error per $10^6$ bases)
Mitosis separates replicated chromosomes before cell division

- Replicated chromosomes are divided into 2 diploid daughter cells
- Nondisjunction $\rightarrow$ aneuploidy $\rightarrow$ genomic mosaicism
  - *chr. fails to split* $\rightarrow$ *wrong # of chr.* $\rightarrow$ different cells, different DNA!
- E.g. 4% of healthy neurons have an abnormal number of chr 21
  
  Rehen, JoN, 2005
Meiosis generates haploid cells

- DNA replication and 2 rounds of cell division
- Homologous recombination creates genetic diversity
- 4 haploid daughter cells (sperm or eggs) have a unique set of chromosomes with DNA from both parents
Genetic recombination during meiosis

- Homologous chromosomes align
- Double-stranded DNA break
- DNA from one chromosome ‘crosses over’ and invades sister chromosome
- Enzymes resolve ‘Holliday junction’
Recombination hot spots

• Recombination rate varies across the genome
• Most crossover events cluster into short regions
• DNA sequence motifs contribute to locations of hot spots

Kong 2002

chr 3
Single nucleotide polymorphisms (SNPs)

- DNA sequence variation at a single nucleotide
- Any 2 human individuals will differ at about 1 out of 1000 bases = 3–4 million differences
- SNPs vary between human populations based on ancestry
Recombination hot spots lead to linkage disequilibrium.

ACGATCGATGCACGATCGATCGTAGCTAGCCGTATCGTAGCTACGTAGC  Reference Sequence
ACGATCATTGCATCGATCGTAGCTAGCCGTATCGTAGCTACGTAGC  Person A

Mutation!

10 generations

Recombination hot spots
Hot spots lead to linkage disequilibrium

100–1000s of generations
Linked SNPs are proxies for a causative allele

**Linkage Disequilibrium:** Non-random association of alleles. They are seen on the same chromosome more frequently than you would expect by chance.

**Haplotype Block:** A combination of genetic variants which are transmitted together.
Visualization of haplotype blocks

- Red = regions of strong linkage; white = little or no linkage
- Haplotype blocks are on average 5–20 kilobases long
## Summary of some molecular forms of genetic variation

<table>
<thead>
<tr>
<th>Category</th>
<th>Molecular process</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Replication</td>
<td>Mutations</td>
<td>Single nucleotide polymorphisms (SNPs)</td>
</tr>
<tr>
<td>Mitosis/Meiosis</td>
<td>Nondisjunction</td>
<td>Aneuploidy</td>
</tr>
<tr>
<td>Meiosis</td>
<td>Recombination</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>DNA Transcription</td>
<td>Alternative splicing</td>
<td>Protein isoforms</td>
</tr>
</tbody>
</table>
Acknowledgements

Trygve Bakken!

Nicholas Schork
Scripps Genomic Medicine
The Scripps Research Institute

Charles Stevens
Molecular Neurobiology Lab
Salk Institute

Funding:
Scripps Genomic Medicine
NIH Neuroplasticity of Aging training grant

Anders Dale
Multimodal Imaging Lab
UC San Diego