Genetics Primer

Brian J Bennett, PhD
Research Leader
Obesity and Metabolism Unit
Western Human Nutrition Research Center
USDA-Agricultural Research Service
Davis, CA
Goals

• Provide basic information about genetics
  • Refresher for some students

• Re-Familiarize vocabulary and concepts
  • Independent assortment
  • Genetic structure
  • Rare genetic variants
  • Common genetic variants
Outline

• Overview of Genetics
  • Traits
  • Patterns of inheritance
    • Mendel’s pea experiment

• Molecular Genetics
  • Genetic Structure and Organization
  • Genetic Variation

• Examples Linking Genetic Variation to Phenotypic Variation
  • Mendelian disorders and height
  • Association and complex traits
Basic Genetics Terminology

Genetics  The study of heredity and variation in living organisms.

Genome  The complete DNA sequence, containing all of the genetic information, within a cell or organism.

DNA  Deoxyribonucleic acid.  Molecule that codes for RNAs and proteins.

Trait  A phenotypic trait is an obvious, observable, and measurable characteristic.

Phenome  The set of all phenotypes expressed by an organism.
Traits
An observable characteristic of an organism
- Physical
- Behavioral
- Many naturally occurring (not specifically induced)
Traits

Qualitative or discrete traits:
- disease (often dichotomous; assessed by diagnosis): Huntington’s disease, obesity, hypertension
- serological status (seropositive or seronegative)

Quantitative or continuous traits:
- height, weight, body mass index, blood pressure
- assessed by measurement
Mendelian trait

A trait which is influenced by a single gene producing a clear pattern of dominant or recessive inheritance within families.

Examples: cystic fibrosis, sickle cell anemia, hemophilia

Human Height
Complex trait

A trait which is influenced by multiple genes and their interactions with each other and with the environment.

Examples: autism, diabetes mellitus, schizophrenia, HDL cholesterol levels, height, BMI
Single gene disorders or complex traits

Mendelian or monogenic

Complex or multifactorial

Disease gene

Modifying gene

Clinical Phenotype

Susceptibility genes

Environmental factors

Clinical Phenotype

Colombo et al. Cardiovascular Ultrasound 2008
Basic Genetics Terminology

**Genetics** - The study of heredity and variation in living organisms
Linking observable traits to patterns of inheritance
# Mendel’s Pea Plants

**Seed surface:**
- Round = R
- Wrinkled = r

**Seed color:**
- Green = G
- Yellow = g

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Round</th>
<th>Wrinkled</th>
<th>Green</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>RR or Rr</td>
<td>rr</td>
<td>GG or Gg</td>
<td>gg</td>
</tr>
</tbody>
</table>
Experimental cross: Round green $\times$ Wrinkled yellow

**Parental plants**  
$RRGG \times rrgg$

**Gametes**  
$RG \times rg$

**$F1$**  
$RrGg \times RrGg$

**Gametes**  
$RG \times Rg \times rG \times rg$
Mendel’s Law of Independent Assortment

<table>
<thead>
<tr>
<th></th>
<th>RG</th>
<th>Rg</th>
<th>rG</th>
<th>rg</th>
</tr>
</thead>
<tbody>
<tr>
<td>RG</td>
<td>RRGG</td>
<td>RRGg</td>
<td>RrGG</td>
<td>RrGg</td>
</tr>
<tr>
<td>Rg</td>
<td>RRGg</td>
<td>RRgg</td>
<td>RrGg</td>
<td>Rrgg</td>
</tr>
<tr>
<td>rG</td>
<td>RrGG</td>
<td>RrGg</td>
<td>rrGG</td>
<td>rrGg</td>
</tr>
<tr>
<td>rg</td>
<td>RrGg</td>
<td>Rrgg</td>
<td>rrGg</td>
<td>rrgg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Green</td>
<td>9</td>
<td>315</td>
</tr>
<tr>
<td>Round Yellow</td>
<td>3</td>
<td>108</td>
</tr>
<tr>
<td>Wrinkled Green</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>Wrinkled Yellow</td>
<td>1</td>
<td>32</td>
</tr>
</tbody>
</table>
Patterns of inheritance
Patterns of inheritance

- *Single allele dominant*
- *Single allele recessive*
- *X-linked*
Single allele dominant

Father (Normal)

Father (aa) 

a

a

Normal

aa

Normal

aa

Mother (Diseased)

Mother (Aa)

a

a

Diseased

Aa

Diseased

Aa
Single allele recessive

Father (Carrier)

Aa

A

a

Aa

 Carrier

aa

Normal

Mother (Carrier)

Aa

a

AA

Diseased

A

Aa

Carrier
Outline

• Overview of Genetics
  • Traits
  • Patterns of inheritance
    • Mendel’s pea experiment

• Molecular Genetics
  • Genetic Structure and Organization
  • Genetic Variation

• Examples Linking Genetic Variation to Phenotypic Variation
  • Mendelian disorders and height
  • Association and complex traits
Chromosome - Unit of heredity so genome can be replicated and transmitted to daughter cells during mitosis or gametes during meiosis.
- Single, linear DNA molecule associated with proteins (although mitochondrial circular).

Packaged as chromatin
Chromosome Dynamics During Meiosis

Germ cells only; not somatic cells

**MEIOSIS**

- **Homologous chromosomes**
  - Diploid 2n (2n2c)
- **DNA Replication and Recombination**
  - (2n4c)
- **Zygotene-pairing/synapsis**
- **Pachytene-crossing over**
- **Diplotene-separate**
- **Chiasma**
- **Recombinant chromatids**

**Prophase I**

- **Meiotic Division 1**
  - Sister chromatids
  - Reduction division
  - Cell Division 1
    - n (but 1n2c)

**Meiotic Division 2**

- Segregation division
- Cell Division 2
  - n (1n1c)

**Haploid gametes**

Fertilization: \( n + n = 2n \) (Diploidy restored in somatic cells, 2n2c prior to S phase)
Karyotype - Chromosome constitution of an individual.
- Photomicrograph of chromosomes from an individual arranged according to standard classification.
- Mammals are diploid organisms with heterogametic sex determination (XX females and XY males); sex chromosomes vs autosomes

![G-bandng](image1.png)
![Spectral karyotype (SKY) - Mouse](image2.png)
![Ideogram - Mouse](image3.png)

Male mouse 40, XY

Chromosome painting by FISH
## More Basic Genetics Terminology

| **Gene** | DNA sequence that codes for a specific product  
|          | --Alternative splicing and isoforms increase diversity. |
| **Nomenclature** | No standard set of rules so confusing. Synonyms.  
|          | Named after gene function (*Sod1, Acly*), size of gene product (*p53, p21*), or phenotype.  
|          | Mouse/human: gene *Brg1/BRG1*; protein *BRG1* |
| **Allele** | One of two or more alternative forms of a gene or DNA sequence at a specific chromosomal location.  
<p>|          | Numerous alleles may exist for a particular gene in a population, but an organism can only have 2 alleles. |</p>
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus</td>
<td>-Position of a gene or DNA sequence on a chromosome.</td>
</tr>
<tr>
<td>Linkage</td>
<td>-Genes or loci on the same chromosome show linkage if they have a tendency to be transmitted together through meiosis.</td>
</tr>
<tr>
<td>Linkage Disequilibrium (LD)</td>
<td>-Tendency of specific combinations of alleles at 2 or more linked loci to occur together on the same chromosome more frequently than would be expected by chance.</td>
</tr>
<tr>
<td>Haplotype</td>
<td>-A group of alleles from closely linked loci, usually inherited as a unit.</td>
</tr>
</tbody>
</table>
Human gene

Figure 1.2. Schematic of structure and expression of an idealized human gene. (From Gelehrter and Collins, 1990.)
Alleles and genotypes

Gene loci

P
a
B

Dominant allele

P
a
b

Recessive allele

Genotype:

PP
aa
Bb

Homozygous for the dominant allele
Homozygous for the recessive allele
Heterozygous
Classes of genetic variants

- Single nucleotide variant
- Insertion–deletion variant
- Block substitution
- Inversion variant
- Copy number variant

Kelly A. Frazer, Sarah S. Murray, Nicholas J. Schork & Eric J. Topol
Nature Reviews Genetics 10, 241-251 (April 2009)
# Types of SNPs

<table>
<thead>
<tr>
<th>Type of variant</th>
<th>Location</th>
<th>Functional effect</th>
<th>Frequency in genome</th>
<th>Predicted RR of phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense</td>
<td>Coding sequence</td>
<td>Premature termination of AA sequence</td>
<td>Very low</td>
<td>Very high</td>
</tr>
<tr>
<td>Non-synonymous (Missense)</td>
<td>Coding sequence</td>
<td>Changes an AA in resulting protein</td>
<td>Low</td>
<td>Low to very high, depending on location</td>
</tr>
<tr>
<td>Synonymous (Sense)</td>
<td>Coding sequence</td>
<td>Doesn’t change AA, but may alter splicing</td>
<td>Medium</td>
<td>Low to high</td>
</tr>
<tr>
<td>Promoter / regulatory region</td>
<td>Promoter, 5’ UTR, 3’ UTR</td>
<td>Can affect level, location, timing of gene expression</td>
<td>Low to medium</td>
<td>Low to high</td>
</tr>
<tr>
<td>Intronic</td>
<td>Non-coding regions</td>
<td>Might affect expression or mRNA stability</td>
<td>Medium</td>
<td>Very low</td>
</tr>
</tbody>
</table>
Haplotypes

SNPs  A/G
(Single Nucleotide Polymorphisms)

Haplotypes
A set of closely linked genetic markers present on one chromosome which tend to be inherited together
SNP Mapping to Identify Haplotypes

SNP
- single nucleotide polymorphism
- minor allele frequency (MAF) >1-5%
- rare variants <1%
- occur de novo 1 per 10 Mb each generation
- Indels also relatively common
## Classification of variation

<table>
<thead>
<tr>
<th>Variant class</th>
<th>Minor allele frequency</th>
<th>Implications for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very common</td>
<td>Between 5 and 50%</td>
<td>Amenable to association analysis using current genome-wide association methods</td>
</tr>
<tr>
<td>Less common</td>
<td>Between 1 and 5%</td>
<td>Amenable to association analysis using variants catalogued in the <a href="https://www.internationalgenome.org">1000 Genomes Project</a></td>
</tr>
<tr>
<td>Rare (but not private)</td>
<td>Less than 1% but still polymorphic in one or more major human populations</td>
<td>Amenable to framework of extreme phenotype resequencing, as well as co-segregation in families</td>
</tr>
<tr>
<td>Private</td>
<td>Restricted to probands and immediate relatives</td>
<td>Difficult to analyse except through co-segregation in families. As linkage evidence will (by definition) be modest, discovery would be limited to the most recognizable of variants</td>
</tr>
</tbody>
</table>
Genotype  -Allele composition of one or more genes/DNA sequences (e.g., SNPs) in an individual.

Phenotype  -Specific characteristics of an individual that results from the allele composition of a particular gene(s).
-Phenome is collective phenotype that results from all genotypes in the genome.
Outline

• Overview of Genetics
  • Traits
  • Patterns of inheritance
    • Mendel’s pea experiment
• Molecular Genetics
  • Genetic Structure and Organization
  • Genetic Variation

• Examples Linking Genetic Variation to Phenotypic Variation
  • Mendelian disorders and height
  • Association and complex traits
- Controlled by a single locus and shows a simple Mendelian inheritance pattern (e.g., agouti, albino, ABO blood type).

- Controlled by two or more loci and may be modified by the environment.
- Usually a continuous trait where the phenotype is quantitative in nature.
- Polygenic inheritance often explained by Mendelian inheritance at multiple loci.

[e.g., height, blood pressure]
Rare highly penetrant variants

Common variants

Distribution of rare variants and common variants on a normal distribution curve

Nat. Rev. Endocrinol. doi:10.1038/nrendo.2012.251
Discovery of rare and common variants for human height

Nat. Rev. Endocrinol. doi:10.1038/nrendo.2012.251
Mendelian inheritance and height

• Achondroplasia is a form of short-limbed dwarfism
• The average height is 131 centimeters (4 feet, 4 inches), males and 124 centimeters (4 feet, 1 inch), females
• Condition occurs in 1 in 15,000 to 40,000 newborns
• Achondroplasia is inherited in an autosomal dominant pattern,
  • one copy of the altered gene in each cell is sufficient to cause the disorder.
  • ~80 % achondroplasia patients have average-size parents;
    • these cases result from new mutations in the FGFR3 gene.
Inheritance and Achondroplasia

Fig. 1 Pedigrees of achondroplasia families. The asterisk on the right denotes homozygous achondroplasia.

All three markers which showed significant linkage with achondroplasia map in chromosome 4p16.3 (Fig. 2).

**Multipoint linkage analysis**

The achondroplasia locus was mapped with respect to the following markers: telomere–IDUA–(3 cM)–D4S412–(6 cM)–HOX7–(4 cM)–D4S431–centromere using multipoint linkage analysis (see Methodology). The most likely location for the achondroplasia locus is between D4S412 and the 4p telomere (Fig. 3).
Locus associated with Achondroplasia

Figure 2: Physical and genetic map of chromosome 4 markers. The genetic distances are in centiMorgans. Marker D4S111 is the intragenic marker of IDUA.

Receptor (FGFR3) has been localized in the achondroplasia gene region and is, by virtue of its pattern of expression, a potential candidate. Recent studies demonstrated expression of the murine Fgfr3 in several structures of the brain, cochlea, lens and cartilage. The highest level of the Fgfr3 expression outside of the nervous system was found in the cartilage rudiments of developing bone. Furthermore, Fgfr3 is expressed exclusively in resting cartilage during endochondral ossification.

Future studies will involve the identification of closer flanking markers. The achondroplasia gene region has been extensively mapped both genetically and physically.
Molecular Genetics of achondroplasia

The common FGFR3 mutations causing achondroplasia both result in Gly380Arg amino acid substitutions
Distribution of rare variants and common variants on a normal distribution curve

Nat. Rev. Endocrinol. doi:10.1038/nrendo.2012.251
Genetic variation ➔ Phenotypic variation

https://www.23andme.Gom/Cen101/snps
Association study 101

SNP = SINGLE NUCLEOTIDE POLYMORPHISM

10 million SNPs
Finding SNPs associated with phenotype

...AGATCGGCAGATTGATA
...AGATCGGCAGATTGATAC
...AGATCGGCAGATTGATA
...AGATCGGCAGATTGATA
...AGATCGGCAGATTGATA
...AGATCTGCAGATTTGATA
...AGATCTGCAGATTTGATA
...AGATCTGCAGATTTGATA
...AGATCTGCAGATTTGATA
...AGATCTGCAGATTTGATA

Height (m)

2.10
2.00
1.95
1.90
1.85
1.75
1.70
1.65
1.63
1.60
1.55
1.50
1.45
1.30
1.25
Null hypothesis

\[ H_0: \ [\text{Phenotype}] \perp [\text{SNP}] \]

\[ y = \mu + \epsilon \]
Alternative hypothesis

$H_1: [\text{Phenotype}] \sim [\text{SNP}]$

$y = \mu + X\beta + e$

<table>
<thead>
<tr>
<th>Height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.10</td>
</tr>
<tr>
<td>2.00</td>
</tr>
<tr>
<td>1.95</td>
</tr>
<tr>
<td>1.90</td>
</tr>
<tr>
<td>1.85</td>
</tr>
<tr>
<td>1.75</td>
</tr>
<tr>
<td>1.70</td>
</tr>
<tr>
<td>1.65</td>
</tr>
<tr>
<td>1.63</td>
</tr>
<tr>
<td>1.60</td>
</tr>
<tr>
<td>1.55</td>
</tr>
<tr>
<td>1.50</td>
</tr>
<tr>
<td>1.45</td>
</tr>
<tr>
<td>1.30</td>
</tr>
<tr>
<td>1.25</td>
</tr>
</tbody>
</table>
These associations are mapped across the genome

Hundreds of genes associated with height
Associations are often named for the closest gene
Summary

• Basic genetic terminology
• Use of height as an example for
  • Terminology
  • Mendelian inheritance
  • Common variation