Understanding Genes & Mutations

John A A Phillips III
May 16, 2005
Learning Objectives

- Understand gene structure
- Become familiar with genetic & mutation databases
- Be able to find information on genetic variation & the phenotype you are studying in the GCRC
Gene Components & Structure

Promoter
Enhancer
5’
CAP Site
DNA
Exons
Introns
mRNA
AATAAAA
Poly A tail
AUG: Start Translation
UAG: Stop Translation
Transcribed
Translated
Start Transcription
Transcribed
AATAAAA polyadenylation signal
Genetic & Mutation Databases

OMIM

Human gene mutation database
http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html?
http://www.genet.sickkids.on.ca/cftr/

Genetic Associations
http://geneticassociationdb.nih.gov/

Single nucleotide polymorphisms (SNP)
OMIM Genetic Database

#219700
CYSTIC FIBROSIS; CF

Alternative titles; symbols
MUCOVISCIDOSIS

Gene map locus 7q31.2

TEXT

DESCRIPTION

A number sign (#) is used with this entry because the disorder is caused by mutations in the cystic fibrosis conductance regulator gene (CFTR; 602421), located on chromosome 7.

Formerly known as cystic fibrosis of the pancreas, this entity has increasingly been labeled simply 'cystic fibrosis.' Manifestations relate
There are currently **1383** mutations listed on this CFTR mutation database.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>missense mutation</td>
<td>558</td>
<td>40.35</td>
</tr>
<tr>
<td>frameshift</td>
<td>228</td>
<td>16.49</td>
</tr>
<tr>
<td>splicing</td>
<td>181</td>
<td>13.09</td>
</tr>
<tr>
<td>nonsense mutation</td>
<td>142</td>
<td>10.27</td>
</tr>
<tr>
<td>in frame in/del</td>
<td>28</td>
<td>2.02</td>
</tr>
<tr>
<td>large in/del</td>
<td>39</td>
<td>2.82</td>
</tr>
<tr>
<td>promoter mutation</td>
<td>8</td>
<td>0.58</td>
</tr>
<tr>
<td>sequence variation</td>
<td>199</td>
<td>14.39</td>
</tr>
</tbody>
</table>
### Genetic Association Database

**Search Results:**

- **Record found:** 26

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Disease</th>
<th>Chr</th>
<th>Ch Band</th>
<th>DNA Position</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLADRB1</td>
<td>Kaposi's sarcoma</td>
<td>0</td>
<td>p21.3</td>
<td>32004020</td>
<td></td>
</tr>
<tr>
<td>SLC11A1</td>
<td>Susceptibility to HIV-1</td>
<td>2</td>
<td>q36</td>
<td>214072421</td>
<td></td>
</tr>
<tr>
<td>ABCB1</td>
<td>HIV-1</td>
<td>7</td>
<td>q21.1</td>
<td>88777598</td>
<td></td>
</tr>
<tr>
<td>ABCB1</td>
<td>HIV-1</td>
<td>7</td>
<td>q21.1</td>
<td>98745209</td>
<td></td>
</tr>
<tr>
<td>SERPIN1</td>
<td>HIV</td>
<td>14</td>
<td>q32.1</td>
<td>92834750</td>
<td></td>
</tr>
<tr>
<td>HLAC</td>
<td>HIV-1 infection</td>
<td>6</td>
<td>p21.3</td>
<td>31244500</td>
<td>0.001</td>
</tr>
<tr>
<td>CCR5</td>
<td>HIV-1 propagation in vitro</td>
<td>3</td>
<td>q21</td>
<td>48372224</td>
<td></td>
</tr>
<tr>
<td>CCR5</td>
<td>HIV</td>
<td>3</td>
<td>q21</td>
<td>40372224</td>
<td></td>
</tr>
<tr>
<td>CCR2</td>
<td>HIV/SIV infection</td>
<td>3</td>
<td>q21</td>
<td>48366231</td>
<td></td>
</tr>
<tr>
<td>CCL5</td>
<td>HIV/SIV infection</td>
<td>17</td>
<td>q11.2-q12</td>
<td>34844046</td>
<td></td>
</tr>
<tr>
<td>CCR2</td>
<td>HIV/SIV infection</td>
<td>3</td>
<td>p21</td>
<td>48366231</td>
<td></td>
</tr>
<tr>
<td>CCR3</td>
<td>HIV/SIV infection</td>
<td>2</td>
<td>q21</td>
<td>137062976</td>
<td></td>
</tr>
<tr>
<td>CCR4</td>
<td>HIV/SIV infection</td>
<td>2</td>
<td>q21</td>
<td>127062976</td>
<td></td>
</tr>
</tbody>
</table>
Types of Point Mutations

- **Silent** (do not alter codons but ??)
- **Missense** (amino acid substitutions)
- **Nonsense** (premature termination)
- **Frameshift** (premature termination)
- **Splicing** (create or destroy splice sites; often cause frameshift & premature termination)
- **RNA processing** (capping, splicing, polyadenylation & editing)
Gene Components & Structure

- **Promoter**
- **Enhancer**
- **Start Transcription**
- **Exons**
- **Introns**
- **DNA**
- **AATAAAA** polyadenylation signal
- **CAP Site**
- **mRNA**
- **Transcribed**
- **Translated**
- **AUG: Start Translation**
- **UAG: Stop Translation**
- **Poly A tail**
There are currently **1383** mutations listed on this CFTR mutation database.

**Mutations by type.**

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>missense mutation</td>
<td>558</td>
<td>40.35</td>
</tr>
<tr>
<td>frameshift</td>
<td>228</td>
<td>16.49</td>
</tr>
<tr>
<td>splicing</td>
<td>181</td>
<td>13.09</td>
</tr>
<tr>
<td>nonsense mutation</td>
<td>142</td>
<td>10.27</td>
</tr>
<tr>
<td>in frame in/del</td>
<td>28</td>
<td>2.02</td>
</tr>
<tr>
<td>large in/del</td>
<td>39</td>
<td>2.82</td>
</tr>
<tr>
<td>promoter mutation</td>
<td>8</td>
<td>0.58</td>
</tr>
<tr>
<td>sequence variation</td>
<td>199</td>
<td>14.39</td>
</tr>
</tbody>
</table>
Genetic Code is Composed of Triplets

<table>
<thead>
<tr>
<th>First letter</th>
<th>Second letter</th>
<th>Third letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU (Phe)</td>
<td>UU (Ter)</td>
</tr>
<tr>
<td></td>
<td>UUG (Leu)</td>
<td>UG (Ter)</td>
</tr>
<tr>
<td>C</td>
<td>CUC (Leu)</td>
<td>CU (Ter)</td>
</tr>
<tr>
<td></td>
<td>CUA (Leu)</td>
<td>CA (Ter)</td>
</tr>
<tr>
<td></td>
<td>CUG (Leu)</td>
<td>CG (Ter)</td>
</tr>
<tr>
<td>A</td>
<td>AU (Ile)</td>
<td>AC (Thr)</td>
</tr>
<tr>
<td></td>
<td>AA (Met)</td>
<td>AG (Lys)</td>
</tr>
<tr>
<td></td>
<td>AG (Arg)</td>
<td>AG (Arg)</td>
</tr>
<tr>
<td>G</td>
<td>GU (Val)</td>
<td>GC (Ala)</td>
</tr>
<tr>
<td></td>
<td>GA (Asp)</td>
<td>GG (Gly)</td>
</tr>
</tbody>
</table>

Met + Start

Term

Term

Trp
Missense Mutations

- Changes in bases 1, 2 & 3 of codon are almost always, always & sometimes missense, respectively.
- Missense changes amino acid sequence ie Ala110Pro
- The amino acid substitution may be harmful or neutral.
### CFTR

**Nucleotide substitutions (missense / nonsense)**

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Codon</th>
<th>Nucleotide</th>
<th>Amino acid</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM010791</td>
<td>1</td>
<td>ATG-AGG</td>
<td>Met-Arg</td>
<td>Asthma</td>
<td>1</td>
</tr>
<tr>
<td>CM941965</td>
<td>1</td>
<td>ATGc-ATA</td>
<td>Met-Ile</td>
<td>Cystic fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>CM960274</td>
<td>1</td>
<td>ATGc-ATT</td>
<td>Met-Ile</td>
<td>Cystic fibrosis</td>
<td>3</td>
</tr>
<tr>
<td>CM930093</td>
<td>1</td>
<td>ATG-AAG</td>
<td>Met-Lys</td>
<td>Cystic fibrosis</td>
<td>4</td>
</tr>
<tr>
<td>CM930094</td>
<td>1</td>
<td>cATG-GTG</td>
<td>Met-Val</td>
<td>Cystic fibrosis</td>
<td>5</td>
</tr>
<tr>
<td>CM940232</td>
<td>2</td>
<td>gCAG-TAG</td>
<td>Gln-Term</td>
<td>Cystic fibrosis</td>
<td>6</td>
</tr>
<tr>
<td>CM940233</td>
<td>3</td>
<td>gAGG-TGG</td>
<td>Arg-Trp</td>
<td>Cystic fibrosis</td>
<td>6</td>
</tr>
<tr>
<td>CM930095</td>
<td>4</td>
<td>TCG-TAG</td>
<td>Ser-Term</td>
<td>Cystic fibrosis</td>
<td>7</td>
</tr>
<tr>
<td>CM970256</td>
<td>5</td>
<td>CCT-CTT</td>
<td>Pro-Leu</td>
<td>Cystic fibrosis</td>
<td>8</td>
</tr>
<tr>
<td>CM021068</td>
<td>6</td>
<td>tCTG-GTG</td>
<td>Leu-Val</td>
<td>Cystic fibrosis</td>
<td>9</td>
</tr>
<tr>
<td>CM031655</td>
<td>7</td>
<td>gGAA-TAA</td>
<td>Glu-Term</td>
<td>Cystic fibrosis</td>
<td>10</td>
</tr>
<tr>
<td>CM010792</td>
<td>11</td>
<td>cGTT-ATT</td>
<td>Val-Ile</td>
<td>Asthma</td>
<td>1</td>
</tr>
<tr>
<td>CM962449</td>
<td>13</td>
<td>TCC-TTC</td>
<td>Ser-Phe</td>
<td>Cystic fibrosis</td>
<td>11</td>
</tr>
</tbody>
</table>
Nonsense Mutations

- Change an amino acid to a premature termination codon (PTC) ie Lys132Ter or Stop
- Encode a truncated protein
- PTCs can cause exon skipping or nonsense mediated decay (NMD)
Missense & Nonsense Mutations & CF

**Compound heterozygote:** An individual who has two different abnormal alleles at a particular locus, one on each chromosome of a pair; usually refers to individuals affected with an autosomal recessive disorder.

**Phenotype:**
- Meconium ileus at birth
- Chronic obstructive lung disease with frequent infections
- Pancreatic insufficiency leading to poor weight gain

**Genotype (CFTR mutations):**
- $\Delta F508 / N1303K$
- $G542X / R553X$

**Samantha**

**Thomas**
### Frameshift mutation

An insertion or deletion involving a number of base pairs that is not a multiple of 3 & consequently disrupts the triplet reading frame, usually leading to the creation of a premature termination (stop) codon & a truncated protein product.

#### Small deletions

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Location/codon</th>
<th>Deletion</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD972120</td>
<td>13</td>
<td>TTGTC^TCCAAaCTTTTTTTCTCA</td>
<td>Congenital absence of vas deferens</td>
<td>1</td>
</tr>
<tr>
<td>CD951649</td>
<td>14</td>
<td>TGTCTC^GAAcTTTTTTCTCAG</td>
<td>Cystic fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>CD941630</td>
<td>15</td>
<td>TCCAAA^CTTTTTTTCTCAG_E1I1_GTGA</td>
<td>Cystic fibrosis</td>
<td>3</td>
</tr>
<tr>
<td>CD941996</td>
<td>26</td>
<td>TTGAGG^AAAGGatACAGACAG</td>
<td>Cystic fibrosis</td>
<td>4</td>
</tr>
<tr>
<td>CD962505</td>
<td>33</td>
<td>GCGCCTG^GAAAttcagacataccaaatCCCTCTCTG</td>
<td>Cystic fibrosis</td>
<td>5</td>
</tr>
<tr>
<td>CD920839</td>
<td>36</td>
<td>ATGTCA^GACatATACCAAAATC</td>
<td>Cystic fibrosis</td>
<td>6</td>
</tr>
<tr>
<td>CD985594</td>
<td>37</td>
<td>TGCAGAC^ATAtaCCAAATCCCT</td>
<td>Cystic fibrosis</td>
<td>7</td>
</tr>
<tr>
<td>CD931164</td>
<td>51</td>
<td>TCTATCT^GAAaATTGGAAAG</td>
<td>Cystic fibrosis</td>
<td>8</td>
</tr>
<tr>
<td>CD972121</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD930918</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD952063</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD983595</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD983595</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Frameshift Mutations

In/del of anything but \((3)^N\) bases alters reading frame & amino acid sequence

\begin{align*}
\text{mRNA} & \quad \text{Normal} \\
\text{Protein} & \\
\text{mRNA} & \quad \text{Variant} \\
\text{Protein} & \\
\end{align*}

- In/del of anything but \((3)^N\) bases alters reading frame & amino acid sequence
- Usually produces truncated protein
### CFTR

**Nucleotide substitutions (splicing)**

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>IVS</th>
<th>Donor/ Acceptor</th>
<th>Relative location</th>
<th>Substitution</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS920988</td>
<td>1</td>
<td>as</td>
<td>-13</td>
<td>C-G</td>
<td>Cystic fibrosis?</td>
</tr>
<tr>
<td>CS983586</td>
<td>1</td>
<td>ds</td>
<td>+1</td>
<td>G-T</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS941436</td>
<td>1</td>
<td>ds</td>
<td>+4</td>
<td>A-T</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS900849</td>
<td>2</td>
<td>as</td>
<td>-1</td>
<td>G-A</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS962489</td>
<td>2</td>
<td>as</td>
<td>-2</td>
<td>A-G</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS983587</td>
<td>2</td>
<td>as</td>
<td>-3</td>
<td>C-A</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS941437</td>
<td>2</td>
<td>as</td>
<td>-3</td>
<td>C-T</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS993379</td>
<td>2</td>
<td>as</td>
<td>-10</td>
<td>T-G</td>
<td>Cystic fibrosis?</td>
</tr>
<tr>
<td>CS971645</td>
<td>2</td>
<td>ds</td>
<td>+1</td>
<td>G-C</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS9003286</td>
<td>2</td>
<td>ds</td>
<td>+1</td>
<td>G-T</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS941989</td>
<td>2</td>
<td>ds</td>
<td>+2</td>
<td>T-C</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CS962490</td>
<td>2</td>
<td>ds</td>
<td>+9</td>
<td>A-T</td>
<td>Cystic fibrosis?</td>
</tr>
</tbody>
</table>

**Normal**

`---TCC gtgagtgga--`

**Mutant**

`---TCC gtgagcgga--`

**DS IVS3 +6 T-C**
Splicing Mutations

- Destroy/create a splice site, ISE or ESE
- Can cause read through, exon skips or activate cryptic splice sites
- Changing splicing alters RNA sequence & translation product
5’ & 3’ Splice & Branch Point Sequences

DNA

Exon  Intron  Exon  Intron  Exon

Transcription

mRNA

5' splice site  branch point  3' splice site

0.35  0.58  1.00  1.00  0.57  0.71  0.84  0.47  0.78  0.81  0.83  0.89  0.82  0.86  0.91  0.87  0.74  1.00  1.00  0.49

... AAG / GTAAGT .......CTR AY ......YYYYYYYYYYYYYNcAG / G ... 

C  G  10-50 bp from AG  T

0.38  0.39  0.22

exon  intron  exon
**Silent Mutations**

Do not alter the amino acid sequence i.e., Gly109Gly

Be careful, they may produce new splice sites or affect ESEs that regulate splice choice.
Single Nucleotide Polymorphisms (SNPs)

- Any single base polymorphic substitution (does not include deletions or insertions)
- Occur ~1300 bp, there are millions of SNPs
- Have a population frequency of at least 1%
- A SNP may or may not affect gene function
<table>
<thead>
<tr>
<th>Description</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 kb incl. ex. 14b-18</td>
<td>Cystic fibrosis</td>
<td>1</td>
</tr>
<tr>
<td>(mutation described at genomic DNA level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21080 bp incl. p. 2-3</td>
<td>Cystic fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>(mutation described at genomic DNA level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 bp nt. 852</td>
<td>Cystic fibrosis</td>
<td>2</td>
</tr>
<tr>
<td>(mutation described at genomic DNA level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 bp nt. 2991</td>
<td>Cystic fibrosis</td>
<td>4</td>
</tr>
<tr>
<td>(mutation described at genomic DNA level)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary of Mutations That Cause CF

[Diagram showing exon and mutation locations]
Learning Objectives

- Understand gene structure
- Become familiar with genetic & mutation databases
- Be able to find information on genetic variation & the phenotype you are studying in the GCRC
Genetic & Mutation Databases

OMIM

Human gene mutation database
http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html?
http://www.genet.sickkids.on.ca/cftr/

Genetic Associations
http://geneticassociationdb.nih.gov/

Single nucleotide polymorphisms (SNP)
Potential Contribution of Genetic Variation to GCRC Studies

- Mendelian (N=42)
- Complex Disease (N=64)
- Drug (N=149)
- Susceptibility (N=57)
- No Genetic (N=26)

Organ Systems Studied:
- Circulatory
- Digestive
- Endocrine
- Immune
- Infection
- Muscular
- Nervous
- Oncology
- Reproductive
- Respiratory
- Skeletal
- Urinary