Advances in Genetics:
From Base Pairs to Bedside

Pediatric Grand Rounds
Regina Margarita Zambrano, MD
Department of Pediatrics
Division of Genetics
LSUHSC and CHNOLA
January 18, 2012
Objectives

• To review the recent advances in Medical Genetics
• To discuss the benefits and limitations of recent technologic advances in genetic testing
• To recognize the value of family history and clinical examination in the evaluation of patients as well as ethical, legal and social implications
• To be aware of the complexity of current practice of Genetics requiring referral to specialized centers
• To recognize individualized medicine as our ultimate goal
Disclosure

Nothing to disclose
“We live in a revolutionary age. Our science has caught the spirit of the times, and more improvements have been made in all its branches in the last 20 years than have been made in a century before”

Benjamin Rush, 1791
Human Genome Project

- Conceived in the mid 1980’s
- Debated and argued
- Oct 1, 1990 start date
- Initial efforts focused on technology, maps & sequencing model organisms
- Ended in 2003
Whole Genome Sequences: Progress Report

Completed genomes as of January 17, 2012:
3065

www.genomesonline.org
Genomes Online Statistics

Completely Sequenced Genomes

Genome Projects according to Phylogenetic Groups

www.genomesonline.org
Human Genome Project

What have we learned about the human genome itself over the past decade?

How has the human sequence propelled our understanding of human biology?

What is the road ahead?
Understanding the structure of genomes

• In the early 2000 our knowledge of the contents of the human genome was surprisingly limited:

• Today the human genome is known to contain between 21,000 and 23,000 protein coding genes

• The human genome is more complex than imagined, regulatory elements outweigh coding regions
Worm vs. Human

Caenorhabditis elegans
- 20,470 protein-coding genes
  ~ 972 cells

Homo Sapiens (Ella Sapiens)
~ 22,000 protein-coding genes
~ 100 trillion cells
Generation of Complexity

• Only modest increase in gene number

• More sophisticated regulation of gene expression

• More sophisticated post-transcriptional regulation
“Variation is the Spice of Life”

- Since the early 1900’s Sir Archibald E. Garrod recognized the importance of human variation

- In 2007 Science Magazine recognized Human Genetic Variation as the Breakthrough of the year
Scales of DNA

• 1 base pair (bp): SNP

• 1,000 bp: size of a typical protein-coding region

• 100,000 bp (100 kb): size of a typical gene

• 3 million bp (3 Mb): minimum size visible with conventional cytogenetics

• 150 Mb: size of average chromosome

• 3,000 Mb (3 Gb): Size of the human genome
Scales of DNA

Chromosome 1: 246 Mb (2968 genes)

Human genome: 3 billion base pairs

Chromosome Y: 50 Mb (231 genes)
Genetic Variation

- Single nucleotide polymorphisms (SNPs)  
  1 bp

- Insertion/deletions (Indels)  
  few bp

- Short tandem repeats (STRs)  
  few bp

- Copy number variation (CNVs)  
  1 to 100s kb

- Cytogenetic deletions/insertions  
  >3 Mb

- Aneuploidy  
  >100 Mb
Copy number variants (CNVs)

- DNA segment $>$1kb with a variable copy number compared to the reference genome
- Microdeletions and microduplications
- Currently limited information regarding clinical significance: Pathological? Benign?

[Diagram of genetic changes: Inversion, Insertion, Deletion, Copy number variation, Reference sequence]
What is the extent of CNV between the genomes of normal individuals? 

Analyzed 20 normal individuals

~ 11 CNP differences between individuals

Average length of CNP was 465 kb

About 50% of identified CNP were recurrent in multiple individuals

CNP frequently located near regions responsible for neurodevelopmental disorders

Large-Scale Copy Number Polymorphism in the Human Genome  Sebat et al.
Science Vol 305, July 23, 2004
Copy number variation

The ultimate goal is to catalog all the CNV’s that can be examined for associations with phenotypes and interpreted in the clinical setting.

Database of Genomic Variants
A curated catalogue of structural variation in the human genome

http://projects.tcag.ca/variation/
The road ahead

• Affordable sequencing
• Understanding all the functional elements of the human genome
• Improve diagnosis of unexplained congenital disorders and identify therapeutic targets for genomic disorders
• Identification of susceptibility loci and functional validation studies of common diseases (polygenic disorders)
Role of Genes in Disease

Online Mendelian Inheritance in Man (OMIM®) lists 3,364 phenotype descriptions with known molecular bases (January 17, 2012).

Gene Tests lists 2,528 diseases with molecular tests, 2,273 clinically available (January 17, 2012).
Role of Genes in Disease

GeneTests: Growth of Laboratory Directory

Benefits of understanding Mendelian diseases?

• Elucidating gene function and regulation

• Recognizing normal and pathological pathways

• Development of effective diagnostic tools

• Development of therapeutic targets

• Contribute to understanding of the molecular genetic basis of common complex diseases
Traditional methods of finding Mendelian disease causative genes

- Linkage analysis
- Cytogenetics
- Compelling biology - candidate genes
- Animal models with similar phenotypes
New strategies of finding Mendelian disease causative genes

Molecular Karyotyping

1. BAC array
   4,200 BAC Clones

2. Oligonucleotide array
   135,000 Oligonucleotides

3. SNP array
   610,000 SNPs
Array Comparative Genomic Hybridization (aCGH)

- AKA chromosome microarray

- Developed as a research tool in cancer cytogenetics

- Revolutionized the diagnostic work-up of patients and facilitated identification of the molecular bases of many genetic disorders
aCGH
Pros and Cons of Arrays

• Pros
  – High detection rate 5-20%
  – Better mapping of aberrations
  – No need for cell culture
  – Shorter result time
  – Reveals unsuspected genomic imbalances
  – Detects genomic duplications

• Cons
  – Not able to identify balanced chromosomal rearrangements or inversions
  – Detection of imbalances of unclear significance
  – Costly
SNP Array

• What is a SNP?
  
  • A SNP is defined as a single base change in a DNA sequence that occurs in a significant proportion (more than 1 percent) of a large population.
SNP Array

- SNP arrays are comprised of oligonucleotides that correspond to SNPs along the human genome
- Used to: - detect polymorphisms within a population; - find disease susceptibility alleles; - find important variants in pharmacogenomics; - to detect LOH
- Allows the use of DNA sequence variation to identify copy number changes
- Identification of long stretches of homozygosity (consanguinity) and copy neutral genetic abnormalities like uniparental disomy
New strategies of finding Mendelian disease causative genes

• Next-generation sequencing:
  – Whole genome sequencing
  – Exome sequencing

• Considerations:
  – Exons (180,000) are about 1% of human genome or 30 Mb
  – About 85% of the disease causing mutations are found in coding regions or canonical splice sites
  – Difficult to interpret the functional consequences of variations in non-coding regions
WGS vs ES

<table>
<thead>
<tr>
<th>Whole Genome Sequencing</th>
<th>Exome Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determining the sequence of the entire human genome</td>
<td>Determining the sequence of the coding regions (exome) of our genome</td>
</tr>
</tbody>
</table>
Whole Genome and Exome Sequencing

- By sequencing the entire genome or exome of affected individuals, different groups have identified the molecular bases of Mendelian disorders.
Whole Genome and Exome Sequencing

- WGS and ES will identify a large number of sequence variants (Raw data is HUGE!)
- Using additional methods like linkage and bioinformatics, variants are prioritized and the clinically significant (causal) variant is identified.
- Recent AR and AD Mendelian disorder causal variants identified using ES and WGS data include: Miller syndrome, Metachondromatosis, Kabuki Syndrome, Schinzel-Gideon Syndrome, and the list is growing!
Tools used with ES/WGS

- At present, the sequencing of each new individual identifies ~402 novel variants introducing changes in the encoded protein
- Prioritize mutations that introduce truncations of the encoded protein
- Prioritize mutations that occur at highly conserved positions
- Consult SNP databases and use mutation prediction software
### Tools used with ES/WGS

Whole exome and whole genome sequencing. Bick, D; Dimmock, D. Curr Opin Pediatr 23:594-600.

<table>
<thead>
<tr>
<th>Has the variant been seen in a patient with a particular disease?</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMIM</td>
</tr>
<tr>
<td>HGMD</td>
</tr>
<tr>
<td>PubMed</td>
</tr>
<tr>
<td>Is the variant evolutionarily conserved?</td>
</tr>
<tr>
<td>phastCons</td>
</tr>
<tr>
<td>If it is a coding region variant will it result in a premature stop codon, a read-through of a stop codon, affect protein structure or function?</td>
</tr>
<tr>
<td>SIFT</td>
</tr>
<tr>
<td>PolyPhen-2</td>
</tr>
<tr>
<td>Has the variant been seen in humans before?</td>
</tr>
<tr>
<td>dbSNP</td>
</tr>
<tr>
<td>Does the variant affect RNA splicing?</td>
</tr>
<tr>
<td>GeneSplicer</td>
</tr>
<tr>
<td>In genes with several transcripts, does the variant affect a well documented transcript?</td>
</tr>
<tr>
<td>RefSeq</td>
</tr>
<tr>
<td>CCDS</td>
</tr>
<tr>
<td>Is there a region of homozygosity in consanguineous families to help find the location of a causal recessive disease?</td>
</tr>
<tr>
<td>BEAGLE</td>
</tr>
</tbody>
</table>

Table 1: Example of tools to analyze variants

<table>
<thead>
<tr>
<th>Has the variant been seen in a patient with a particular disease?</th>
<th><a href="http://www.ncbi.nlm.nih.gov/omim">http://www.ncbi.nlm.nih.gov/omim</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>OMIM</td>
<td><a href="http://www.hgmd.cf.ac.uk/ac/index.php">http://www.hgmd.cf.ac.uk/ac/index.php</a></td>
</tr>
<tr>
<td>Is the variant evolutionarily conserved?</td>
<td><a href="http://sift.jcvi.org">http://sift.jcvi.org</a></td>
</tr>
<tr>
<td>phastCons</td>
<td><a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a></td>
</tr>
<tr>
<td>If it is a coding region variant will it result in a premature stop codon, a read-through of a stop codon, affect protein structure or function?</td>
<td><a href="http://www.ncbi.nlm.nih.gov/projects/SNP/">http://www.ncbi.nlm.nih.gov/projects/SNP/</a></td>
</tr>
<tr>
<td>SIFT</td>
<td><a href="http://www.cbcb.umd.edu/software/GeneSplicer/">http://www.cbcb.umd.edu/software/GeneSplicer/</a></td>
</tr>
<tr>
<td>dbSNP</td>
<td><a href="http://faculty.washington.edu/browning/beagle/beagle.html">http://faculty.washington.edu/browning/beagle/beagle.html</a></td>
</tr>
</tbody>
</table>
Costs of genetic testing

Karyotype

FISH

$500-$600

$260
Costs of genetic testing

$1,000 - 2,000

SNP Array

$14,000

Exome Sequencing
Family History

The pedigree is one of the most useful tools in genetic diagnosis.
Genotype first diagnosis controversy

• New technologies have increased the number of patients with an established diagnosis.

• Recent publications advocate the "genotype first" approach to the evaluation of patients with disorders of development by primary care practitioners.

• The yield of abnormalities found after CMA is higher after the patient has undergone a thorough PE and FH (YL).

• I feel the family history and PE by an expert clinician should always precede testing in patients with disorders of development.

Case Report

- 14yo and 9yo Hispanic females
- DDM, microcephaly, and dysmorphic features
- Seen in 2005

Negative paternal and maternal family history and consanguinity was denied
Case Report

Suspected chromosomal anomaly. Karyotype was normal (2005)
Case Report

- 5 years later in 2010 returned to clinic and CMA was requested: dup 11p15.4 (278.4kb)
- Sister same abnormality
- Parental testing pending
NCBI Map Viewer

- 3 OMIM® annotated genes
  - NUP98
  - RHOG
  - STIM1
NUP98

- 920 amino acids; 97 kDa; contains repeated motifs (GLFG and FG) in N-term and a RNA binding motif in C-term
- Mediates nucleo-cytoplasmic transport of protein and RNA
- Chromosomal translocations involving NUP98 have been identified in patients with MDS, T-ALL, CML and AML
NUP98 and 29 partners and/or recurrent translocations. Editor 03/2002; last update 02/2010
NUP98

- Unclear what are the implications of duplications of *NUP98*

- Dual haploinsufficiency of *Nup98* and *Rae1* has been shown to result in premature separation of sister chromatids, leading to severe aneuploidy
RHOG

- Member of the RAS family of supergenes

- Encodes a GTP binding protein that acts in the pathway of signal transduction and plays a role in the regulation of cellular functions

- Unclear what are the implications of duplications of RHOG
STIM1

- 746 amino acids, 90% sequence identity to mouse Stim1, conserved from Drosophila to human
- Encodes a calcium sensor that conveys calcium load of the ER to store operated calcium channels at the plasma membrane
STIM1

• Mutations in STIM1 cause immune dysfunction with T-cell inactivation due to Ca entry defect

• It is not clear what is the effect of duplications of STIM1, but overexpression in HEK293 cells modestly enhanced calcium entry
Patient 254127: Chr11:3,637,741-4,238,493
Microcephaly and short stature

Our patients: Chr11:3,695,738-3,974,177
Microcephaly, short stature and dysmorphic features
Summary

- Duplications at 11p15.4 have not been reported in the literature and the information regarding clinic significance is still unclear.
- Both sisters have the duplication, it must have been inherited from a normal parent.
- Region is located near a known imprinted region (BWS region at 11p15.5).
# Evolution of Medicine

<table>
<thead>
<tr>
<th>19th Century</th>
<th>20th Century</th>
<th>21st Century</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treat symptoms</td>
<td>Treat diseases</td>
<td>Predict and Preempt symptoms &amp; disease</td>
</tr>
</tbody>
</table>

*Annual Meeting of the Roadmap Multidisciplinary Clinical Research Career Development Program*

Elias A. Zerhouni, M.D.

Director
Average Medicine vs. Individualized Medicine

• When it comes to the practice of medicine, we have been educated with the “classic case mentality”
• With the technologic advances we have learned of genetic variation giving each one of us our own “flavor of health or disease”
• With our increasing ability to identify and interpret genetic variations, a paradigm shift in medicine is occurring
One size fits all medicine or “Average Medicine”

Individualized medicine
Individualized medicine

• Individualized medicine is the use of information from a patient's genotype to:
  – Initiate a preventative measure against the development of a disease or condition, or
  – Select the most appropriate therapy for a disease or condition that is particularly suited to that patient
Pharmacogenomics

- Analyzes patient genotypes for cytochrome P450 (CYP) genes CYP2D6 and CYP2C19
- AmpliChip CYP450 Test is based on microarray technology
  - 2 CYP2C19 phenotypes
  - 4 CYP2D6 phenotypes
- It is a major step toward introducing personalized prescribing into the clinical environment
Obstacles to overcome

• Millions of genetic variant exist and identification of them all will take years
• Medication adverse reactions may depend on not only one variant but interacting variants
• Determination of such interactions is also going to take sometime
• Expense of testing
Mayo Clinic announced in December 2011 that it will start a pilot project to sequence the genomes of volunteers.

“The project will help managers at the clinic decide whether it makes sense to read and store a patient's whole genome early on, instead of ordering single genetic tests as and when the need arises.”
But...

- Who is going to store the information?
- How is it stored securely?
- Who has access?
- What are you going to do with information that you or the patient might not necessarily want to find out?
- There are some significant ethical and privacy issues and they are probably more difficult to solve than storing the information.
• In 2008 the Genetic Information Nondiscrimination act (GINA) was passed in the US
  – Prohibits insurers from using personal genetic information to determine eligibility or premiums
  – Prohibits an insurer from requiring a person to have a genetic test
  – Prohibits employers from using a person’s genetic information in making hiring, firing, job assignments decisions
  – Prohibits employers from requesting, purchasing, requiring personal or familial genetic information
Direct To Consumer Genetic testing

• When GINA was conceived the availability of genetic testing was limited
• By 2008 when the law was passed genetic testing was available for > 1,300 conditions
• Today we have available tests for > 2,500 disorders
• Most of these tests are offered in the clinical setting
• Some can be offered via DTC test kits
Questions?

• A bright future awaits

• Be mindful of Ethical, legal and Social implications

"You were to have inherited all this, son, but genetic screening has indicated you’re too big of a health risk."
Bio-Informatic Resources

- OMIM
  - http://www.omim.org

- Gene tests

- dbSNP

- UCSC Genome Browser
  - http://genome.ucsc.edu/

- Ensembl Genome Browser
  - http://www.ensembl.org/index.html

- DECIPHER
  - https://decipher.sanger.ac.uk/application/

- DGV
  - http://projects.tcag.ca/variation/