The question

- 10 years since draft HGP
- 2 years since the “Year of the GWAS”
- Very little impact on clinical medicine
- But, sequencing is getting cheaper
- The number of genomes is set to rise
- What does a consultation look like in 5 years?

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost Estimate</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>$300,000,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2001</td>
<td>$100,000,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2007</td>
<td>$10,000,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2008</td>
<td>$2,000,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2008</td>
<td>$1,000,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2008</td>
<td>$500,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2008</td>
<td>$250,000</td>
<td>Sanger (ABI)</td>
</tr>
<tr>
<td>2009</td>
<td>$50,000</td>
<td>Sanger (ABI)</td>
</tr>
</tbody>
</table>

The idea

What if everybody’s genome was available in their medical record?
Patient zero

- 40 year old male in good health presents to his doctor with his whole genome
- No symptoms
- Exercises regularly
- Takes no medication
- Family history of aortic aneurysm
- Family history of sudden death

Clinical examination

- Normal appearing male
- Comfortable at rest
- HS 1,2+0
- No murmurs, rubs or gallops
- Chest clear, abdomen nad
- Musculoskeletal, neuropsych examinations grossly normal
- Afebrile
- HR 60pm, BP 128/80

Electrocardiogram

[Heart diagram and ECG trace]
**Echocardiography**

**Exercise test**

<table>
<thead>
<tr>
<th>Test Date</th>
<th>03/04/2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name (First Last)</td>
<td>MDX</td>
</tr>
<tr>
<td>Resting syst BP</td>
<td>120</td>
</tr>
<tr>
<td>Resting diast BP</td>
<td>80</td>
</tr>
<tr>
<td>Age (years)</td>
<td>45</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>70</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>170</td>
</tr>
<tr>
<td>Exercise mode</td>
<td>Bike</td>
</tr>
<tr>
<td>Duration</td>
<td>30 min</td>
</tr>
</tbody>
</table>

| Muscle not to scale |

| WBC | 4.9 |
| Total bill | 0.5 |
| Hb | 15.7 |
| Platelets | 147 |
| Na | 143 |
| K | 4.0 |
| BUN | 20 |
| Cr | 1.2 |
| eGFR | LDL 156 |
| Ca | 9.4 |
| Fasting glucose | 93 |
| hsCRP | <0.2 |
| Lp(a) | 114 |

**Lab tests panel**

**Parsing 6,000,000,000 data points**

When one base pair change can turn this into this.
Rare/novel algorithm

- What does it mean for a variant to be associated with disease?
  - Cosegregation in a large kindred?
  - Early stop in key gene
    - in one proband? in several individuals?
  - Splice site mutation?
  - Novel mutation?
  - Not seen in ?how many? controls
    - What’s a control?

Rare variant databases

- Human Genome Mutation database
  - Public/professional
- Human Variome project
- Human genome variation database
- Private databases

http://www.hgmd.cf.ac.uk/docs/oth_mut.html
Accessed 3/11/2010
GVS (SeattleSNPs) within transcript 40,287
SIFT (JCVI) Coding 17049 (6329 novel)
mitochondrial variants
Non-synonymous n=8286
Predicted damaging n=2124
Premature Stop n=140
Mendelian disease associated n=114

Polygenic disease – what we have now

Algorithms for entirely novel variants
Published Genome-Wide Associations through 3/2010, 779 published GWA at p<5x10^{-8} for 148 traits.

NHGRI GWA Catalog
www.genome.gov/GWAStudies

2008: the year of the GWAS – time for celebration?

J Hypertens. 2004 Sep;22(9):1717-21.
Missing heritability

- Rare variants
- Structural variants
- Epigenetic phenomena
- Over-zealous bounding of FWER
- G-G interaction
- G-E interaction
- G-G-E interaction

Can we apply this to individual genomes?

One approach

Challenges in applying results of GWAS to individual genomes

- Theoretical
  - Not enough variance explained
- Practical
  - Most NCBI databases are catalogs
  - Although sharing and making data publicly available (despite ethical concerns) remains routine, journals have not traditionally insisted on sufficient data for genome interpretation (standard is ‘reproduce the expt’ but even that often not met)
  - Even the GWAS catalogs do not contain sufficient data
    - Genotype frequencies
    - Strand direction variable, rarely reported
    - Chromosomal position changes with each genome build
Existing SNP databases are limited in resource and content

- NHGRI GWAS Catalog
  - 2,387 SNPs → 321 diseases, curated from 509 PubMed
  - Odds Ratio, but no genotypes

- NHLBI GWAS Catalog
  - 52,546 SNPs → 87 diseases, curated from 119 PubMed
  - p_value, no OR

Ways to apply this for genomic medicine

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>a/(a+c)</td>
</tr>
<tr>
<td>Specificity</td>
<td>d/(d+b)</td>
</tr>
<tr>
<td>Prevalence</td>
<td>(a+b)/(a+b+c+d)</td>
</tr>
<tr>
<td>NPV</td>
<td>d/(d+c)</td>
</tr>
<tr>
<td>PPV</td>
<td>a/(a+b)</td>
</tr>
<tr>
<td>OR</td>
<td>ad/cb</td>
</tr>
<tr>
<td>OR (a/b) / (c/d)</td>
<td></td>
</tr>
<tr>
<td>RR</td>
<td>(a/(a+b)) / (c/(c+d))</td>
</tr>
<tr>
<td>LR+</td>
<td>sen/1-spec</td>
</tr>
<tr>
<td>LR-</td>
<td>1-sen/spec</td>
</tr>
</tbody>
</table>

For GWAS, most OR are in the range 1.3-1.6

Odds are....the effect will be exaggerated

- Two groups (n=100), two conditions
  - First group Y=80, N=20
  - Second group Y=20, N=80
  - First group is 4x more likely to be Y
  - However, OR=(80/20)/(20/80) = 16
  - This can be even more extreme
    - eg (90/10)/(10/90), OR=81!
  - Remember that for GWAS, most OR are in the range 1.3-1.6
The Likelihood is . . . you will at least account for test characteristics

- The LR is easily overlaid on the pre-probability to provide a post-test probability
- This helps with the "relative risk" problem

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre test probability</td>
<td>Prevalence</td>
</tr>
<tr>
<td>Pre test odds</td>
<td>Prev/1-prev</td>
</tr>
<tr>
<td>Post test odds</td>
<td>Pre-test odds x LR</td>
</tr>
<tr>
<td>Post test probability</td>
<td>Post test odds / post test odds +1</td>
</tr>
</tbody>
</table>

Riskogram methods and figure

- Pre test prob from various sources
  - Prevalence usually (matched to age, sex, ethnicity if possible)
  - Lifetime risk occasionally
- Mean LR when multiple studies for same SNP
  - Weighted mean (square root of sample size)
- Only one SNP per haplotype block (largest LR)
- Pre test odds multiplied by LRs cumulatively
  - Presented in decreasing order of studies, then sample size

Report card

Challenges

- Calls were made vs human reference sequence
  - Risk alleles in human reference sequence
- Winner's curse
  - Literature bias towards positive results
- Negative studies need to be included in algorithm
- Data for LR only available for 40% papers

What of “patient” zero?

- SQ feedback
  - PGx information welcome
  - Approach to personal and family screening
- Medical advice
  - Personal and family screening
  - CAD risk
    - ATP3+LP+LR+PGx + clinical judgement
    - Rx statin

Conclusion

- In the future, we will not be limited by the availability of genetic information
- For medicine to become “personalized” we will need to learn how to parse this data
Acknowledgements

Practical

Check for rare variants

- Is it in dbSNP?
  - e.g. Chr position 6: 160881127
  - http://gvs.gs.washington.edu/GVS
  - rs3798220 (LPA)
  - Yes

- Is it in dbSNP?
  - e.g. 6: 7528007
  - No

- Go to Sift http://sift.jcvi.org/
- Choose 2a (nonsynonymous SNP genome scale), build 36
- Enter: 6,7528007,1,G/A
- Tick gene name
- Choose Proceed to Sift results page
- Choose complete set view results

Same mutation in PolyPhen – polymorphism phenotyping

- http://genetics.bwh.harvard.edu/pph/
- Desmoplakin
- Uniprot ID = P15924
- Position: 1838
- AA1 = Arginine (arg, R), polar, +ve charge
- AA2 = Histidine (his, H), polar, neutral charge
Polyphen result

Private database lookup

- Cardiogenomics
- Myosin binding protein C
- R326Q

Published data on this mutation

Likelihood ratios and the riskogram

Likelihood ratio example – KLF11

LR = sens / 1-spec
LR = (a/a+c) / (1 – (d/d+b))
LR = (25/25+1358) / (1-(1445/1445+18))
LR = 0.02 / (1 - 0.99)
LR = 0.02 / 0.01
LR = 2