Genome Research - The Driver of Precision Medicine

Warren Kaplan
Head, Data Sciences Platform
Why Genomics
Nuclear fission
Five-dimensional energy landscapes

Seafloor spreading
The view from under the Arctic ice

Career prospects
Sequence creates new opportunities
Human Genome
~3.2 Billion letters: ~3 Gigabytes

Deoxyribonucleic Acid (DNA)

Chromosome
Sequencing Evolution/Revolution (per day)

1990: kiloBases

2000:

2010: GigaBases

2019: TeraBases
Early days: a DNA-sequencing lab in 1994
Why Genomes?
Source Code that defines who we are
Individualized genomic medicine
From prewomb to tomb

- Planning a baby
- Fetus
- Newborn
- Undiagnosed disease
- Disease prevention
- Infectious disease
- Cancer
- Molecular Dx, PGx
- Healthspan
- Molecular autopsy

-9m to 20
- 90+ to tomb
5M variants
2.8M good data quality
28,000 within genes
5,000 nonsilent variants
200 predicted damaging
100 matched for phenotype
50 checked for known disease DB
5 prioritised for review
1 candidate
Population Scale Genome Sequencing
Precision medicine in China
Genomics at scale is an engineering not a scientific endeavor

250 Gigabytes of Data per Genome
700 CPU Hours per Genome
> 5 million genetic variations per patient
scalable genome analysis platform
<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Genome Reference Bank</td>
<td>Garvan Institute of Medical Research</td>
<td>TRUE</td>
</tr>
<tr>
<td>Altruist</td>
<td>Sequencing.com</td>
<td>TRUE</td>
</tr>
<tr>
<td>AMPLab - 1000 Genomes Project</td>
<td>AMPLab UC Berkeley</td>
<td>TRUE</td>
</tr>
<tr>
<td>Australian Genomics Health Alliance - Germine</td>
<td>Australian Genomics Health Alliance</td>
<td>FALSE</td>
</tr>
<tr>
<td>Australian Genomics Health Alliance - Somatic</td>
<td>Australian Genomics Health Alliance</td>
<td>FALSE</td>
</tr>
<tr>
<td>Belgium Beacon</td>
<td>Belgian Medical Genomics Initiative</td>
<td>N/A</td>
</tr>
<tr>
<td>BioReference</td>
<td>BioReference Laboratories</td>
<td>FALSE</td>
</tr>
<tr>
<td>BRCA Exchange</td>
<td>BRCA Exchange</td>
<td>FALSE</td>
</tr>
<tr>
<td>Cafe CardioKit</td>
<td>University of Leicester</td>
<td>N/A</td>
</tr>
<tr>
<td>Cafe Varionome Central</td>
<td>University of Leicester</td>
<td>N/A</td>
</tr>
<tr>
<td>Cell Lines</td>
<td>Wellcome Trust Sanger Institute</td>
<td>FALSE</td>
</tr>
<tr>
<td>cosmic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dbGaP: Combined GRU Catalog and NHLBI Exome Sequence Project</td>
<td>National Center for Biotechnology Information</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Clinical Filtering

54 selected out of 1,139 patient records.

- Height (cm)
  - Min: 164.67
  - Max: 183.22

- Weight (kg)
  - Min: 28.00
  - Max: 100.00

- Systolic Blood Pressure (mmHg)
  - Min: 7.00
  - Max: 177.00

- Blood Glucose Level (mmol/L)
  - Min: 2.20
  - Max: 5.73

- Waist Circumference (cm)
  - Min: 72.43
  - Max: 90.09

<table>
<thead>
<tr>
<th>Location</th>
<th>Reference</th>
<th>Alternate</th>
<th>Type</th>
<th>consequences</th>
<th>clinvar</th>
<th>OncomerF</th>
<th>Allele Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>G</td>
<td>SNP</td>
<td>3_prime_3'UTR_variant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>G</td>
<td>SNP</td>
<td>3_prime_3'UTR_variant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>C</td>
<td>SNP</td>
<td>3_prime_3'UTR_variant</td>
<td></td>
<td>LikelyBenign</td>
<td></td>
</tr>
</tbody>
</table>
Web Notebooks for Bioinformatics Researchers
SYDNEY GENOMICS COLLABORATIVE

TRANSLATING GENOMICS RESEARCH INTO BETTER HEALTH OUTCOMES

NSW Government Health
Medical Genome Reference Bank

Provide integrated genomic and phenotypic reference from 4000+ healthy elderly Australians.

Genomes Sequenced

3000/4000
NSW Health Genomic Medical Research Grants

Empowering researchers to translate the potential of genomic research into better health outcomes.

Genomes Sequenced

3000/4000

1689/2295

Genomic Cancer Medicine Program

To expedite the translation of genomic discovery into improved health outcomes for cancer patients.

Genomes Sequenced

242/2000
Acute Care Genomics Flagship (2018 - 2020)
Cardiovascular Genetic Disorders Flagship (2018 - 2020)
chILDRANZ Flagship: Interstitial and Diffuse Lung Disease in Children (2018 - 2020)
KidGen 'HIDDEN' Renal Genetics Flagship (2018 - 2020)

Neuromuscular Disorders Flagship (2016 - 2018)
Mitochondrial Diseases Flagship (2016 - 2018)
Neurodevelopmental Disability Flagship (2016 - 2018)
Genetic Immunology Flagship (2016 - 2018)
KidGen Renal Genetics Flagship (2016 - 2018)
Government to spend millions on 'Mackenzie's mission' to increase access to genetic testing

7.30

Exclusive by medical reporter Sophie Scott and the Specialist Reporting Team's Rebecca Armitage

Updated 1 Mar 2018, 4:44pm
Study

The Medical Genome Reference Bank: a whole genome data resource of 4,000 healthy elderly individuals.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Alternative Stable ID</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGAS00001003511</td>
<td></td>
<td>Other</td>
</tr>
</tbody>
</table>
MEDICAL GENOME REFERENCE BANK
DATA ACCESS APPLICATION FORM
Genome Data Archiving and Sharing

Production Workflows

Patient Cohorts (DNA and Clinical Info)
Now for something different
The impact of health research funding in cancers with low survival rates
Ratio of Mortality to Incidence
Ratios of research funding to cancer incidence and mortality
Increased investment in Cancer Research leads to a reduced burden ($) and death. And vice versa
My Suggestion

- For Genetics Diseases we cannot compensate for the Billions of Dollars spent on cancer research
- A genome provides > 6 Billion data points. Each person has 5 million unique variants
- Supposing we said we would sequence genomes of everyone affected with a genetic/suspected genetic disease.
- This would NOT be a clinical test but rather a RESEARCH genome
- We would create a Committee and have researchers from around the world apply to access the genomes.
- We could have hundreds of researchers around the world using our genome data to do what they promised on their DAA.
So How Are We Going to Pay for the Genome Sequencing?

National Disability Insurance Scheme
NDIS support is for the benefit of the person with a disability. So the question we need to ask:

Would it be beneficial for persons with disabilities to know there are possibly thousands of researchers working on their data every day that is growing our understanding of the genetics of disability?
Acknowledgements

Data Sciences Platform:
Data Portals Team (Dmitry Degrave and Andre Hermanto)
Production Workflow Team (Joe Copty, Shyam Ravishankar, and Matthew Hobbs)
DICE Team (Derrick Lin and Manuel Sopena Ballesteros)