



Garvan Institute
of Medical Research

Genome Research - The Driver of Precision Medicine

Warren Kaplan

Head, Data Sciences Platform





A diagram illustrating the application of genomics in three key areas of research. At the bottom is a wide orange rectangle labeled 'GENOMICS'. Above it are three vertical rectangles: a green one on the left labeled 'Cancer', a red one in the middle labeled 'Healthy Ageing', and a purple one on the right labeled 'Inflammatory Diseases'. All rectangles have thin black borders.

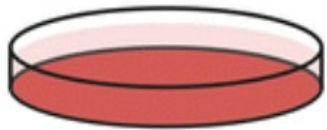
Cancer

Healthy Ageing

Inflammatory Diseases

GENOMICS

Why Genomics



15 February 2001

nature

£5.45 €6.25 ¥1154 (US\$) (US\$) (US\$)

www.nature.com

the human genome

Nucléar fission

Five-dimensional
energy landscapes

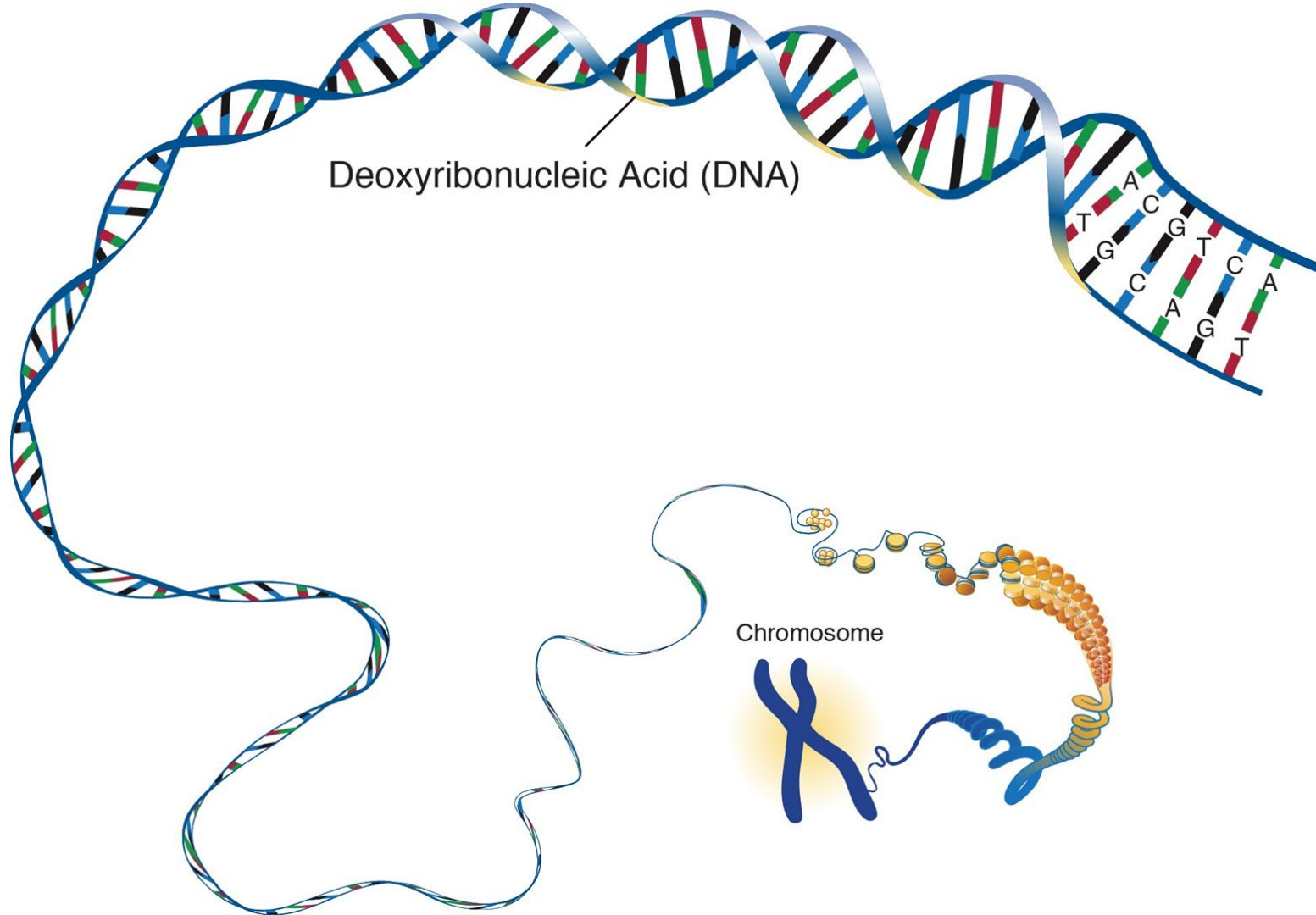
Seafloor spreading

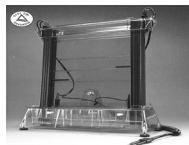
The view from under
the Arctic ice

Career prospects

Sequence creates new
opportunities

naturejobs
genomics special





1990



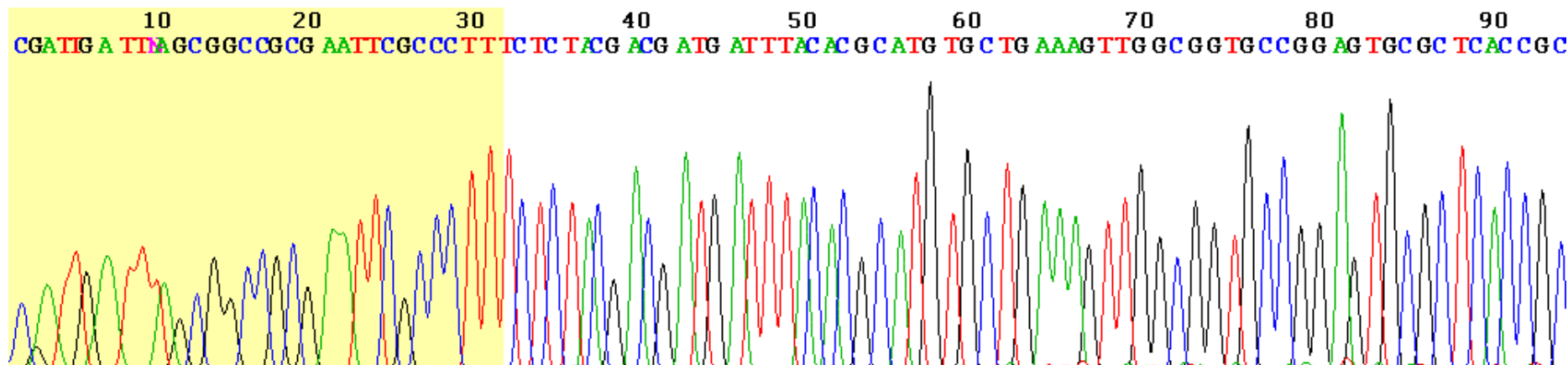
2019



2010

2000





Early days: a DNA-sequencing lab in 1994



0 5,000 10,000 15,000 20,000 25,000 30,000 35,000 40,000



ctgA

ctgA:9591..9639 (50 b)

Go

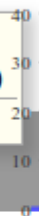
9,600

9,625



volvox-sorted SNPs/Coverage

a*	0	0%
T	21	100% (10 +, 11 -)
Total	21	



volvox-sorted.bam





Why Genomes?

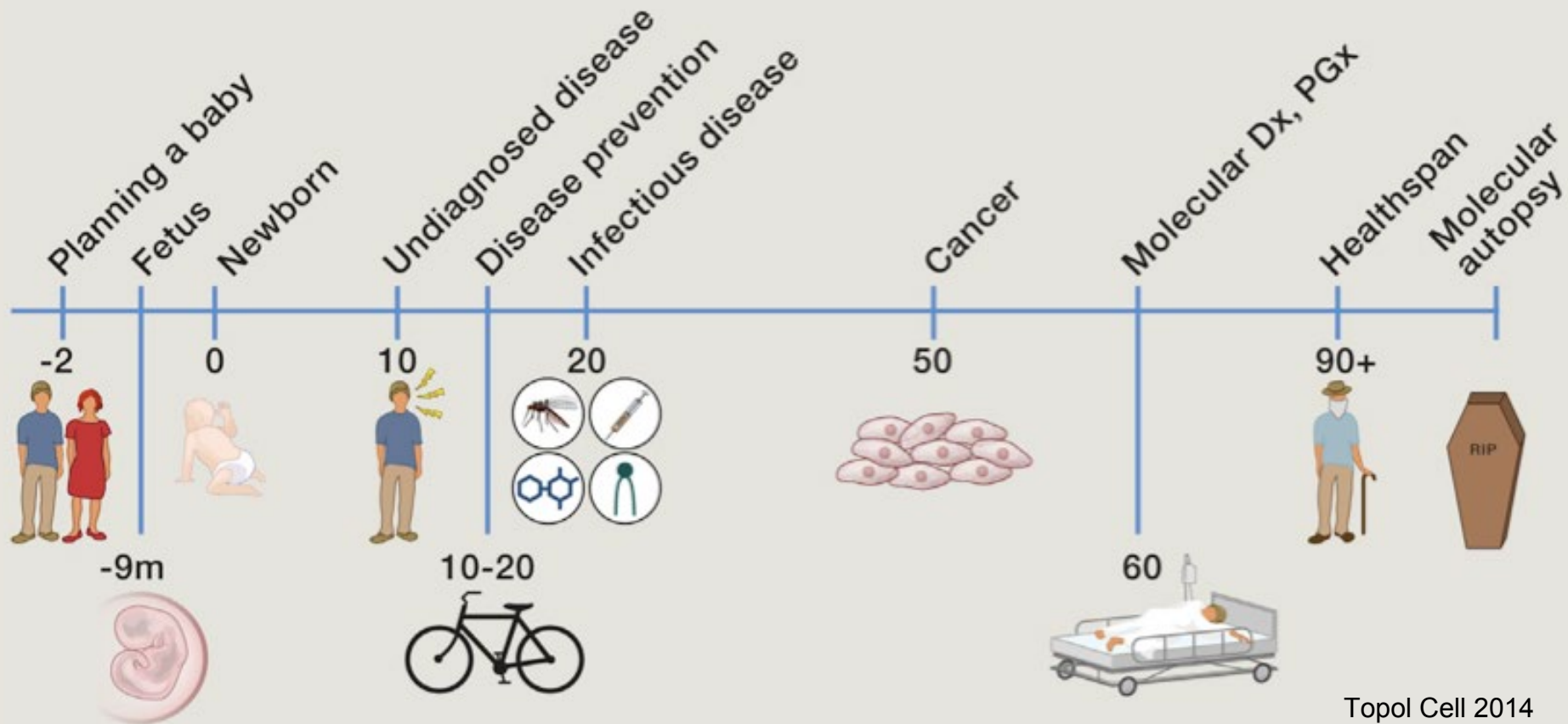
Source Code that defines who we are

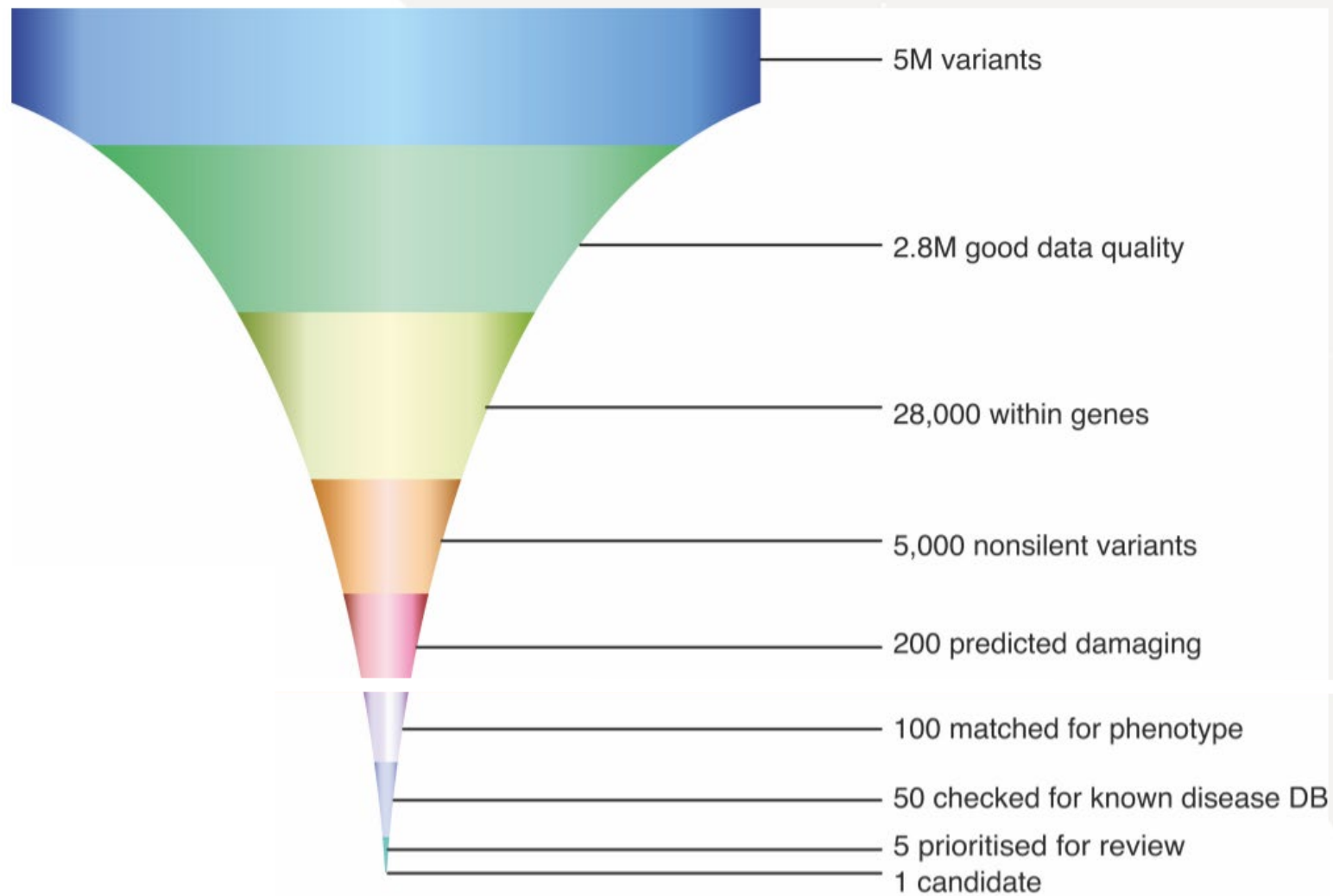




Individualized genomic medicine

From prewomb to tomb



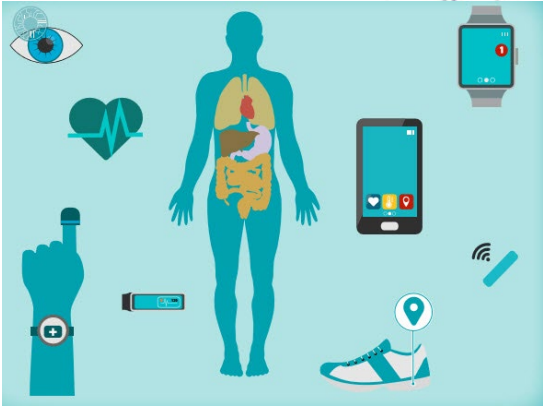
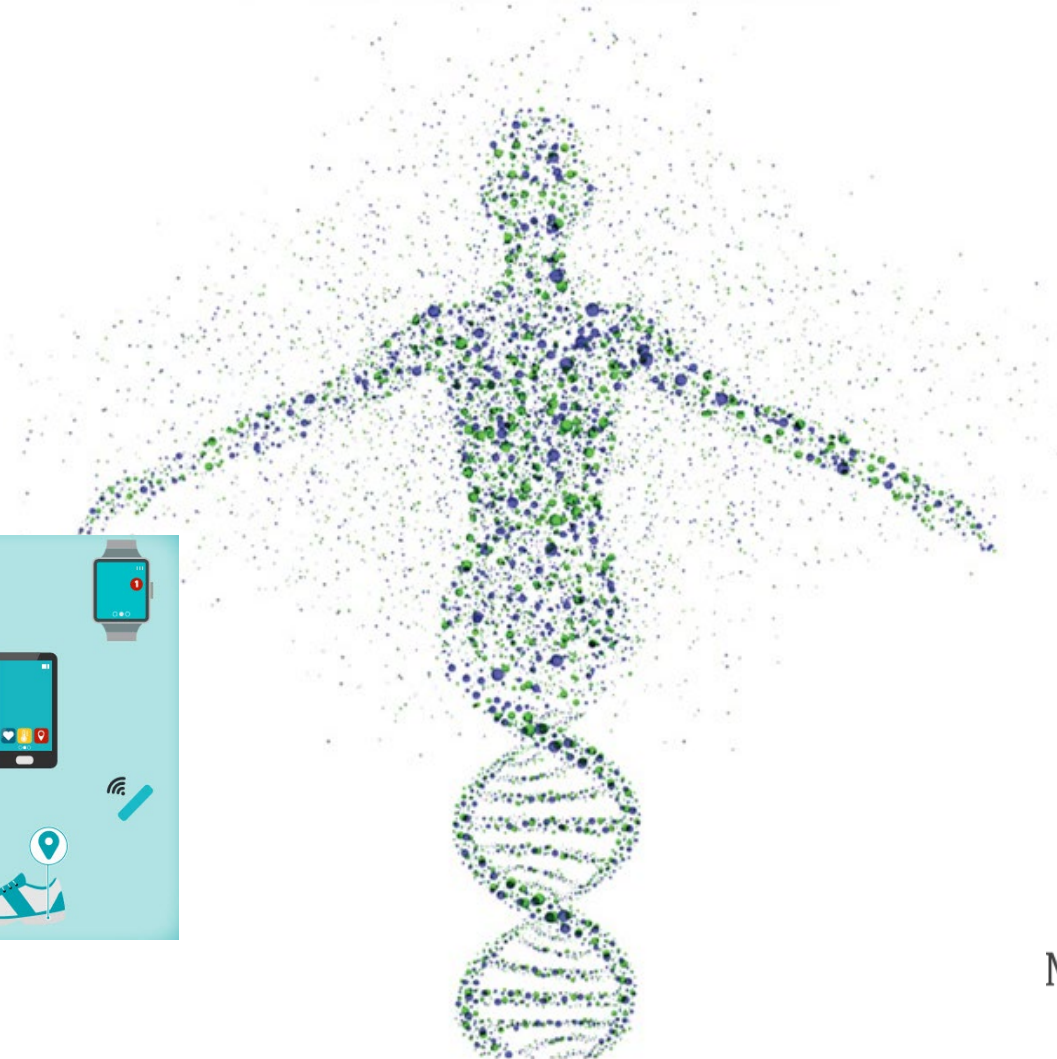




Garvan Institute
of Medical Research

Population Scale Genome Sequencing





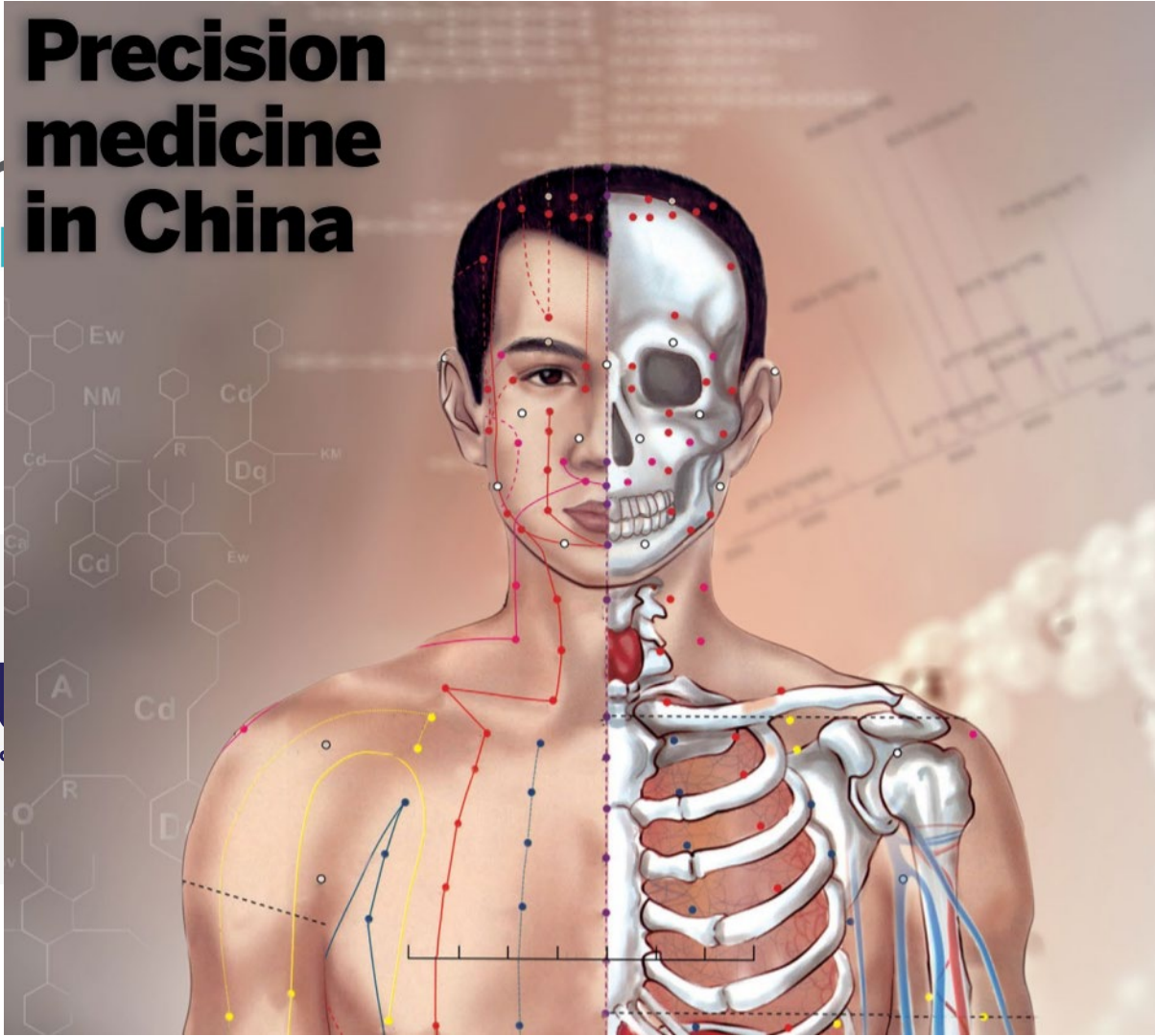
My Health Record



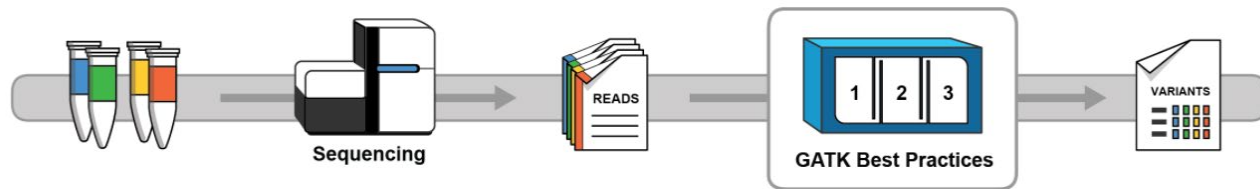
Genom
engl

Precision medicine in China

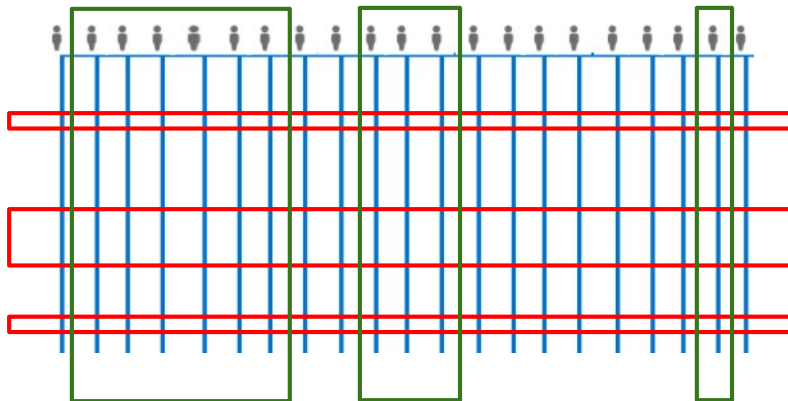
All of U
THE FUTURE OF HEALTH BEC



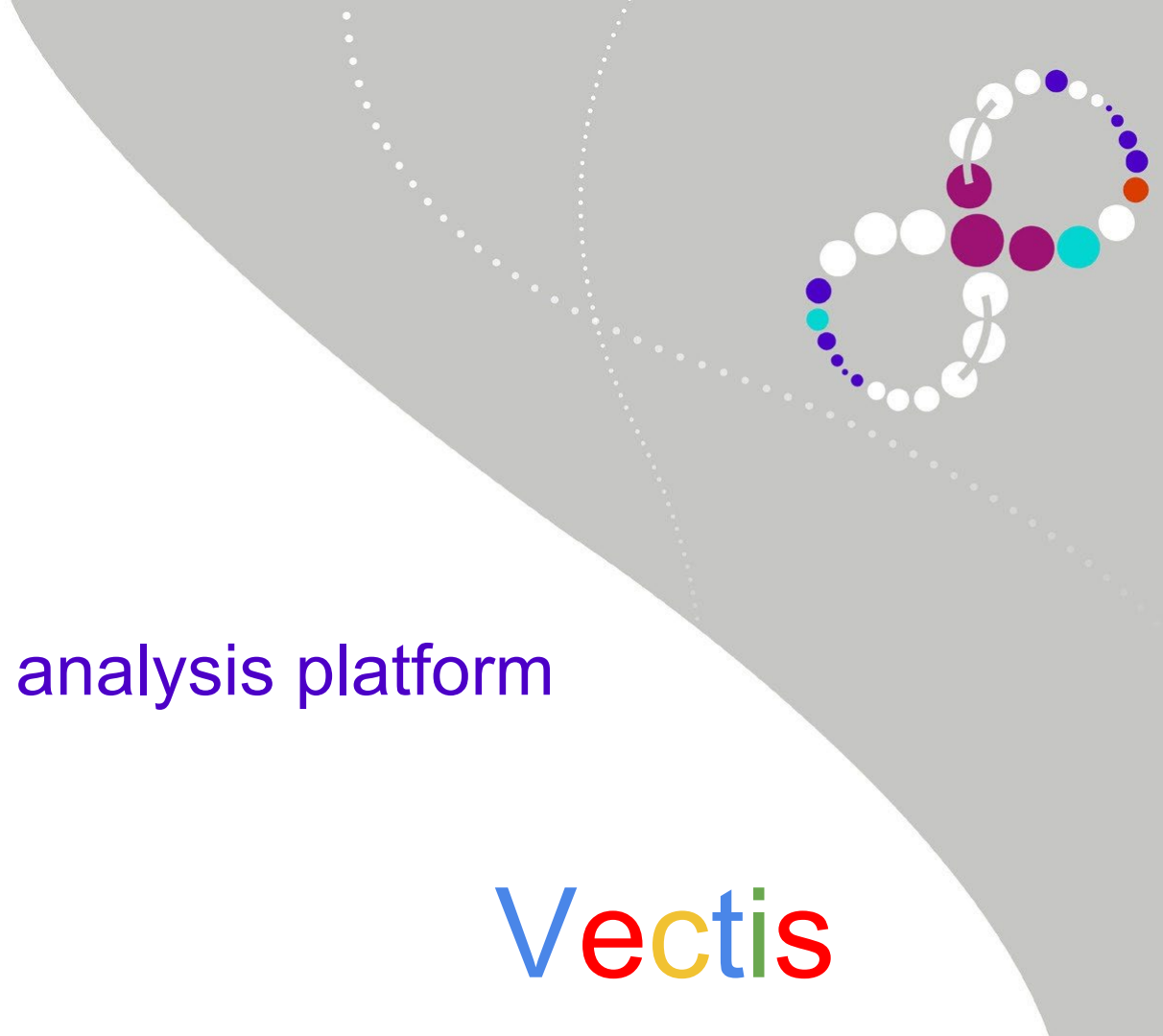
Genomics at scale is an engineering not a scientific endeavor



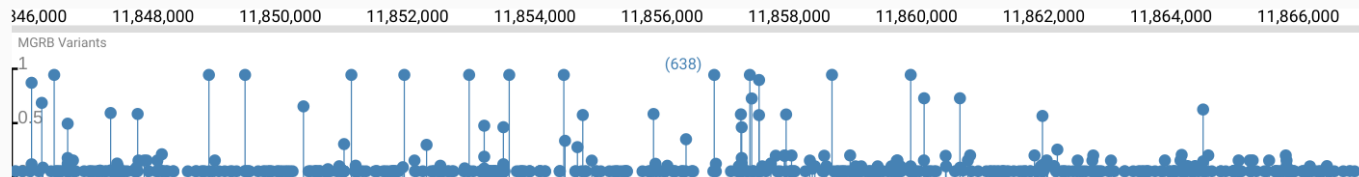
250 Gigabytes of Data per
Genome
700 CPU Hours per Genome
> 5 million genetic variations per
patient



scalable genome analysis platform



Vectis



Filter



Location	Reference	Alternate	Type	consequences	clinvar	GnomadAF	Allele Freq
1:11845837	G	A	SNP	3_prime_UTR_variant	UncertainSignificance		
1:11845946	C	T	SNP	3_prime_UTR_variant	LikelyBenign		



204 / 84146035

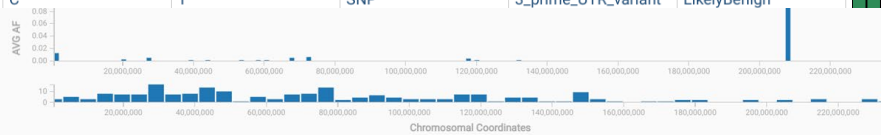
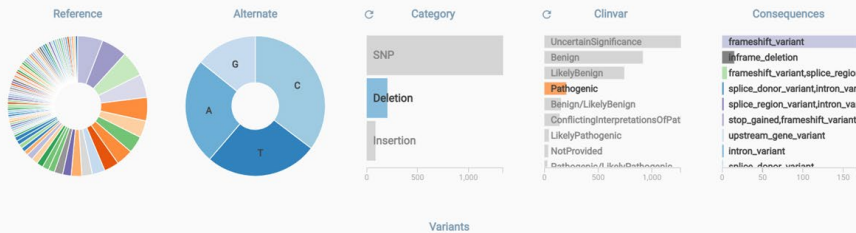
6 Charts

- ☐ Binned AF
- ☐ PolyPhen
- ☐ Sift
- ☐ Top 100 Genes
- ☐ Eigen
- ☐ Gnomad AF

Show SQL

Download Variants

Give feedback



Variant	Type	AF	gnomadAF	RSID	clinvar	consequences
1:19228996:GC:G	Deletion	5.2891e-4	6.4889e-5	rs387906314	Pathogenic	frameshift_variant
1:9305316:AC:A	Deletion	1.7575e-4	9.6874e-5	rs398122818	Pathogenic	frameshift_variant



Global Alliance for Genomics & Health

HG19

▼

2:45895>G

Q

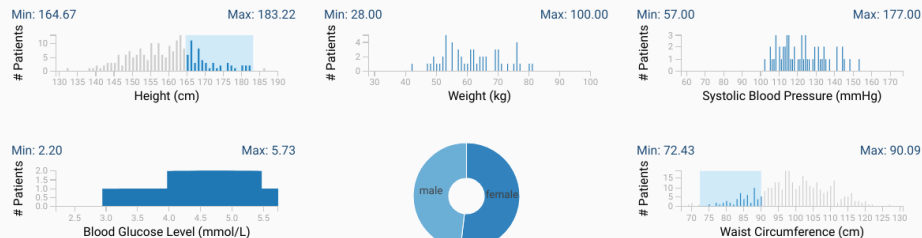
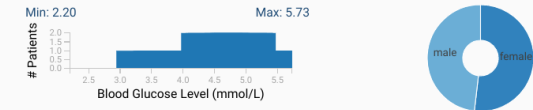
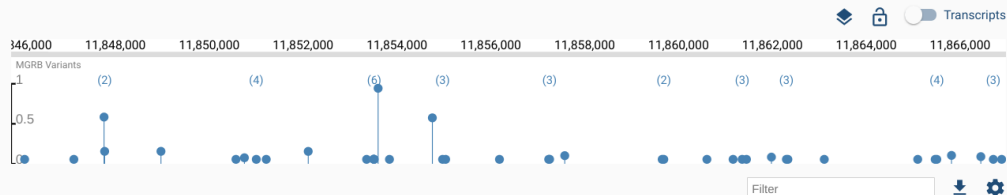
Examples: [22:46546565>A](#) or [2:45895>G](#)

Please note that position is 1-based and only SNP variants are supported

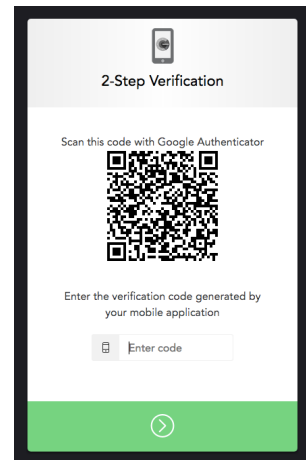
Name	Organization	Response
Medical Genome Reference Bank	Garvan Institute of Medical Research	TRUE
Altruist	Sequencing.com	TRUE
AMPLab - 1000 Genomes Project	AMPLab, UC Berkeley	TRUE
Australian Genomics Health Alliance - Germline	Australian Genomics Health Alliance	FALSE
Australian Genomics Health Alliance - Somatic	Australian Genomics Health Alliance	FALSE
Belgium Beacon	Belgian Medical Genomics Initiative	N/A
BioReference	BioReference Laboratories	FALSE
BRCA Exchange	BRCA Exchange	FALSE
Cafe CardioKit	University of Leicester	N/A
Cafe Variome Central	University of Leicester	N/A
Cell Lines	Wellcome Trust Sanger Institute	FALSE
cosmic		FAILED
dbGaP: Combined GRU Catalog and NHLBI Exome Sequence Project	National Center for Biotechnology Information	N/A

Clinical Filtering

54 selected out of 1,139 patient records.



Location	Reference	Alternate	Type	consequences	clinvar	GnomadAF	Allele Freq
1:11846067	G	C	SNP	3_prime_UTR_variant			
1:11847116	G	A	SNP	3_prime_UTR_variant			
1:11847759	C	T	SNP	3_prime_UTR_variant	LikelyBenign		

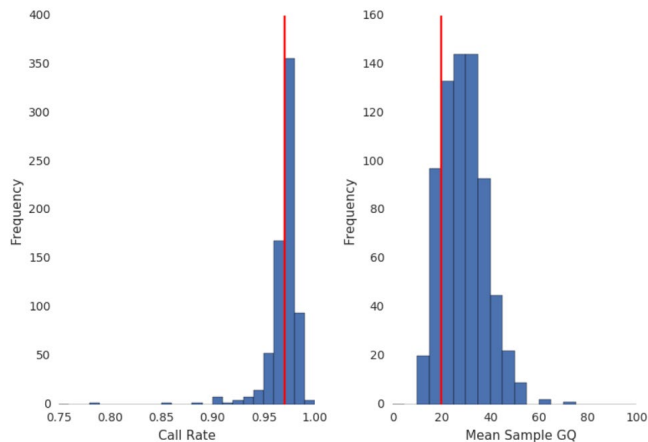


Web Notebooks for Bioinformatics Researchers

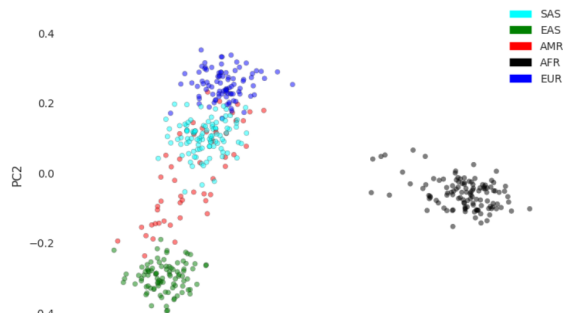
HailTutorial (Python)

Detached View Code Permissions Run All Clear

```
16 plt.axvline(20, color = 'r')
17
18 plt.tight_layout()
19 plt.show()
20 display()
```



```
1 pca_table = vds_pca.samples_keytable().to_pandas()
2 colors = {'AFR': 'black', 'AMR': 'red', 'EAS': 'green', 'EUR': 'blue', 'SAS': 'cyan'}
3 plt.clf()
4 plt.scatter(pca_table["sa.pca.PC1"], pca_table["sa.pca.PC2"], c = pca_table["sa.pheno.SuperPopulation"])
5 plt.xlabel("PC1")
6 plt.ylabel("PC2")
7 legend_entries = [mpatches.Patch(color=c, label=pheno) for pheno, c in colors.items()]
8 plt.legend(handles=legend_entries)
9 plt.show()
10 display()
```





SYDNEY GENOMICS COLLABORATIVE

TRANSLATING GENOMICS RESEARCH INTO BETTER HEALTH OUTCOMES



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



1689/2295

Genomic Cancer Medicine Program

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

Genomes Sequenced



242/2000

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



1689/2295

Genomic Cancer Medicine Program

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

Genomes Sequenced



242/2000

2015 Cohorts

Schizophrenia



Metastatic Melanoma



Mitochondrial Disease



Congenital Heart Disease



2016 Cohorts

Mendelian Disorders



Epilepsy



Rare Disease



Blinding Retinal Dystrophy



Genetic Disorders of Bone



Inherited Cardiomyopathies



2017 Cohorts

Bipolar Disorder



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)



Print



Email



Facebook



Twitter



More

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.

Study ID	Alternative Stable ID	Type
EGAS000001003511		Other

The image shows a title card with a white background. It features a large, thin black outline that forms a wide, shallow U-shape at the top and two vertical lines extending downwards. Inside this outline is a smaller, rounded rectangle with a thin black border. The text "MEDICAL GENOME REFERENCE BANK" and "DATA ACCESS APPLICATION FORM" is centered within this rounded rectangle in a black, sans-serif font.

MEDICAL GENOME REFERENCE BANK
DATA ACCESS APPLICATION FORM

Patient Cohorts (DNA
and Clinical Info)



Australian Genomics
Health Alliance



SYDNEY GENOMICS COLLABORATIVE



Patient Cohorts (DNA
and Clinical Info)



Production Workflows



Garvan
Institute



NCI



Azure



Google Cloud



Patient Cohorts (DNA
and Clinical Info)



Australian Genomics
Health Alliance



SYDNEY GENOMICS COLLABORATIVE

Genome Data Archiving and Sharing



Production Workflows



Patient Cohorts (DNA and Clinical Info)



Data Access Committee Support



Genome Data Archiving and Sharing



Production Workflows



Patient Cohorts (DNA and Clinical Info)



Genome Data Analysis



Data Access Committee Support



Genome Data Archiving and Sharing



Production Workflows



Patient Cohorts (DNA and Clinical Info)





Garvan Institute
of Medical Research

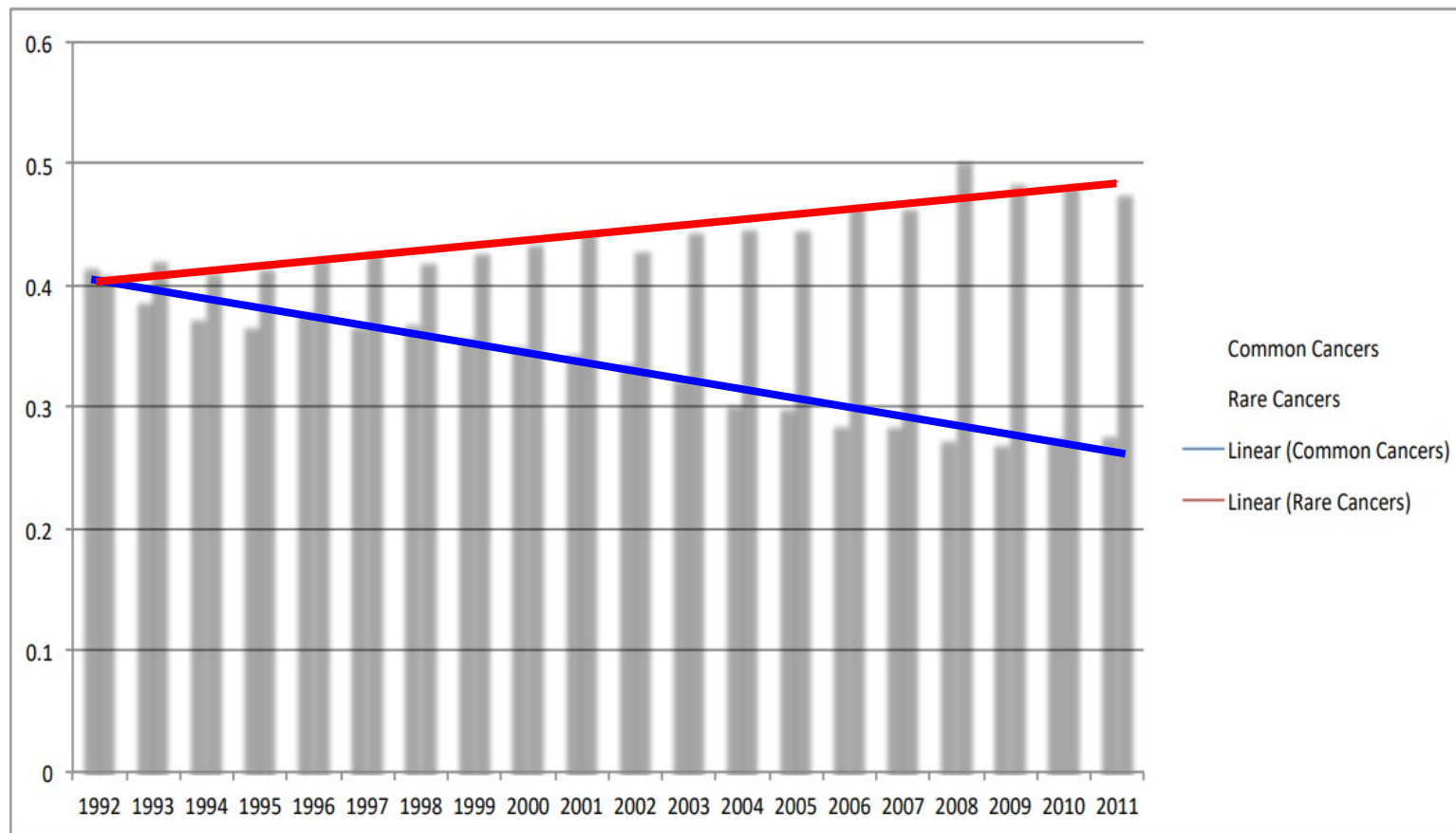
Now for something different



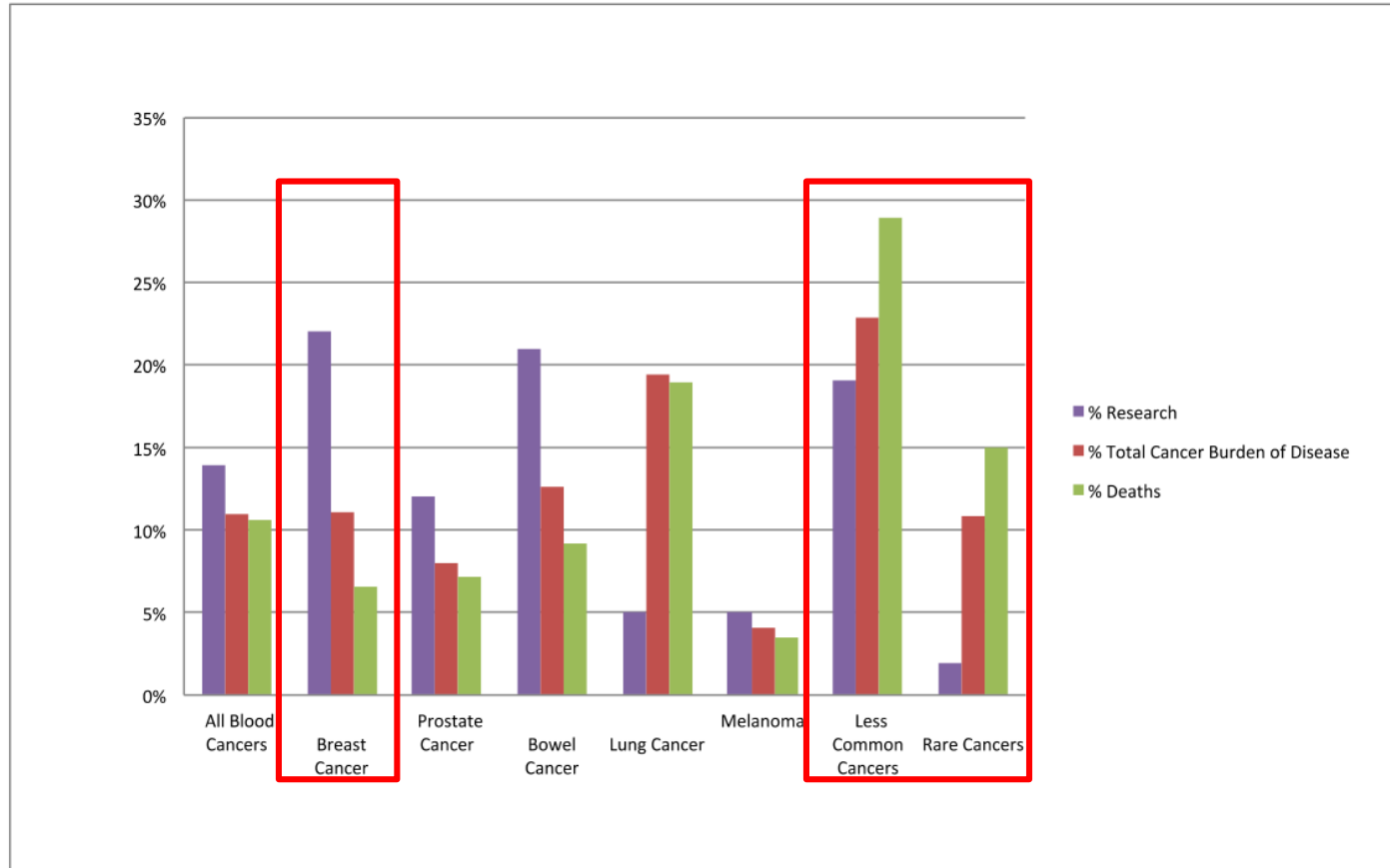
Garvan senate submission 2017

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

- 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

So How Can I Suggest Something So Ridiculous?

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 Copt, Shyam Ravishankar, and Matthew Hobbs)

DICE Team (Derrick Lin and Manuel Sopena Ballesteros)

