Genomic Data Integration (GeDI)

Professor Kerina Jones

SAIL DATABANK

Swansea University Prifysgol Abertawe

Swansea University Medical School
Health Informatics Group

Rationale for GeDI

Huge interest and value in genomic data research and precision medicine

Debate – anonymization of genetic data - Identifiable/de-identified/anonymised

- GDPR Article 9: special category data, along with health, political views, race, etc.
- Unique ≠ identifiable
- How different is genomic data anyway?
  - Genomic data (largely) persists
  - Sensitive/discrimination
  - Partly unknown
  - Impact on kin - heritable conditions, paternity
- But not exclusively …
- Hearts and minds swayed by opinion
- Need for clearer data governance
- To date: majority of genomic data is collected for research

SAIL DATABANK

Institute of Life Science
College of Medicine

Are genomic data special?

- Laissez faire vs 'genetic exceptionalism'
  - Open genetic data advocates
  - Special rules & regulations
- Survey of genetic data uses:
  - Forensics
  - DTCs
  - Precision medicine
  - Health research/secondary use
  - Genealogical use
- But – maybe the jury is still out -

What — If Anything — Is Special about "Genetic Privacy"?

Jacqueline JL Chin* and Alastair V Campbell*

INTRODUCTION

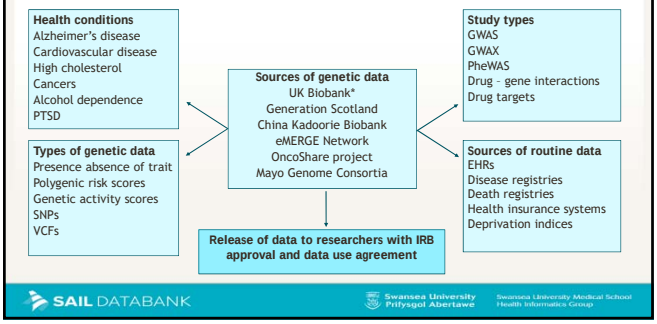
The term "genetic privacy" is one that should not be taken at face value. This is because it carries an unexamined assumption of genetic exceptionalism — the idea that there is something exceptional about genetic knowledge that warrants special ethical attention and protection in law or policy — and genetic exceptionalism *could* be false. Not only that, certain harms may be incurred by false beliefs in genetic exceptionalism, and the belief that there is a special category of privacy captured by the term "genetic privacy". In this chapter, we shall

- Public - DNA information different to other medical information: Yes 52%, No/unsure 48%
- Largest differences in opinion between genetic health professionals and the public

SAIL DATABANK

Institute of Life Science
College of Medicine

Studies using genetic data and health records



SAIL DATABANK — Swansea University Prifysgol Abertawe — Swansea University Medical School Health Informatics Group

Mini-case studies: genetic and health data in data safe havens

Sax Institute
45 & up: 260,000 people over
Whole genomes and derived data
Hybrid model with Gavarn
Remote access for researchers
Sample collection/resourcing

SAIL
PsyCymru
Welsh cohort with psychosis
Polygenic risk scores and CNVs
Access by project team only
Cross-disciplinary work and security assurance

DPUK
Dementia research: >2M people
Sequence, exomes, summary data on APOE4 genes
Remote access to named researchers
Technical issues and rationalising public benefit

IC/ES
Cohort: ~2,200 children with neurodevelopment disorder
VCF files
Access by IC/ES analyst only
Data governance and technical

100K
People with rare conditions and family members
Whole genomes and derived data
Remote access for researchers
Not known

BC Generations
30,000 people aged 35-69
Data type determined by research need
Data release. But if linked, then access is only via safe haven
Lengthy set up approvals

SAIL DATABANK — Swansea University Prifysgol Abertawe — Swansea University Medical School Health Informatics Group

Models of genomic data access

- Held by/released to researchers : mainly via biobanking or research studies
- Accessed within a data safe haven : mainly via data centres
- Publically available on a website
  - Personal Genome Project

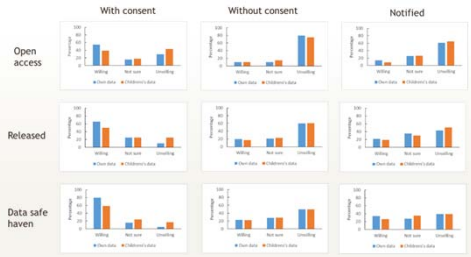
Can be with/without linked data

Can be identifiable/de-identified/anonymised

- How data are collected, risk appetite, access restrictions, approvals, purpose, technical capacity, social acceptability …

SAIL DATABANK — Swansea University Prifysgol Abertawe — Swansea University Medical School Health Informatics Group

Willingness for the reuse of genomic data collected for research



SAIL DATABANK — Swansea University Prifysgol Abertawe — Swansea University Medical School Health Informatics Group

Suggested wording – starting to see results!

For all research studies using personal data if want to option to share for linkage
Plan from the outset

For the Participant Information Sheet -
'The data you provide to us is important and may be useful to other research studies. So that your data can be used anonymously for further research in the public interest, an NHS organisation will replace your identifying details with a unique anonymous code. This will enable your data to be linked to routinely-collected data, including your health records. The data can then be used for research in anonymous form in a secure environment, such as the Secure Anonymised Information Linkage Databank (SAIL).'

For the consent form -
'I understand that my personal data will be de-identified by an NHS organisation so that it can be used in anonymous form for further research in the public interest.'

SAIL DATABANK — Swansea University Prifysgol Abertawe — Swansea University Medical School Health Informatics Group

Modelling: risk-control-utility

Metadata/Data
Presence/absence of a trait \
Monogenic risk score
Polygenic risk scores Link
SNPs
VCF files /

Partial gene sequence*
Full single gene sequence*
Multiple gene sequences*
Full genome sequence*

*+/- annotations

Risk factors:
Sensitive/stigmatising conditions
Rare conditions
Data granularity
Data extent
Linkage to other datasets
Other factors .....

Controls within a data safe haven – options and combinations determined on a project-by-project basis

1. Remote access as per usual (SAIL) including results release following scrutiny
2. Data use contingent on data provider permission
3. Access by analyst employed by SAIL/DSH only
4. Access within safe room only
5. Conditions for any data outs contingent on specific approvals
6. Suppression, aggregation, masking, differential privacy, etc., retaining maximum utility
7. Federated access i.e. genetic data not brought into DSH

Red – Amber – Green type model:
Green: business as usual
Amber: plus additional safeguards
Red: modify research question or no go

Summary

- Evidence to inform data governance for working with genetic data is essential
- Data protection legislation categorises genetic data as for other health data
- Currently the majority of genetic data is with research studies
- Our work indicates public support providing data are handled properly – data safe havens preferred
- General move from data release to data in safe haven
- Current state of play in research studies – missed opportunities (not just for genomic data)
- Importance will further increase as genetic data becomes more mainstream/in EHRs

Further work

- **How/what/who to store:** beyond VCFs further processing is likely to be needed before can analyse with health records
- **Pipelines:** governance and technical models for secure data transfer, management and linkage
- **Analysis:** how to prepare data and work with phenotypic and genomic data, software and skill sets
- **Pathfinders:** exemplar projects to test and consolidate processes
- **Stakeholders:** working with data controllers/providers, researchers, etc.
- **Working as part of Genomics Partnership Wales:** genomics for precision medicine strategy implementation; Wales Gene Park; Welsh research infrastructure; biobanking community, etc.
- **Further public and stakeholder engagement:** on genomic data research and precision medicine – just opened the box
- **Advising:** on the ethical, legal and societal issues in integrating genomic data with health records
- **Education and capacity building:** sharing what we've learned

SAIL interim policy

'Genomic data (derived from the DNA sequence of an individual) can be more sensitive than general health data, bringing with them specific issues were they to be disclosed. Furthermore, due to their inherent uniqueness, genomic data derivatives may pose additional risks when linked to other datasets in SAIL. To mitigate these risks, we have set out the following rules in relation to use of genomic data derivatives in SAIL:

1. *Project-based (non-routinely collected) genomic data may be deposited into SAIL providing all relevant approvals and permissions have been obtained, and an appropriate consent model has been used.*
2. *Routinely-collected genomic data are brought into SAIL as part of standard data feeds.*
3. *A member of the IGRP with genomics knowledge will sit on the committee when considering requests for access to genomic data.*
4. *If deemed particularly sensitive, some types of genomic data will require additional protective measures before release of results. This will be discussed with researchers during data scrutiny.*
5. *All the standard SAIL & SAIL Gateway operating policies and procedures will apply to these data in addition to the specific points described.'*

Finalise with reference to GA4GH: Framework for Responsible Sharing of Genomic and Health-Related Data

Questions?

The GeDI study is funded by the MRC as part of the Wales Genomic Medicine Centre
For further information, please contact Kerina Jones: k.h.jones@swansea.ac.uk

Are genomic data special?

> Our survey of the range of current uses of genetic information and of the likely future expansion of these uses illustrates that this is an area that needs careful ethical analysis; and that current safeguards to protect privacy, confidentiality and to prevent harmful applications of the information gained may be inadequate. The sheer volume of potential genetic information, the speed with which its scope is increasing and the potential for complex combinations of this information with other data have (as we have seen) led some writers to argue that we have to have special measures in place to deal with this type of information. However, we do not consider that such genetic exceptionalism is either warranted or helpful in describing and protecting the ethical values involved. The issues are the classical ones of social ethics: the rights and corresponding responsibilities of individuals; the need for clear legal and regulatory protection of privacy and personal information; and the challenge of finding a balance between the protection of the rights of individuals and the promotion of the public good. The rapid progress of genomics does serve to highlight — sometimes in a dramatic way — these core ethical issues. But it does not in any way change their character: the protection of privacy, and a nuanced understanding of what it entails, crosses the entire range of informational sources in contemporary society. There is no special realm called "genetic privacy".

Public workshops

- To gain views on socially acceptable approaches for accessing genetic + health data in anonymised form
- Focused on how (not if): open access, release to researchers, within a data safe haven
- Genetic data collected for research
- Series of 8 public workshops (N = 116)
- Men 54 : women 62
  - 6th formers
  - Further education
  - SU staff and student group
  - Business and NHS group
  - Consumer Panel
  - Science festival
  - Grand round
  - University of the 3rd age (U3A)