Privacy in the genomic era

The issue

Current approaches

- the do nothing / “share the love” (aka the cowboy approach)
- full public access (1000G)
- dbGAP (full access after application, transit encryption)
- “walled garden” (100K Genomes, UK)

A solution?

*[1984] radical new idea in cryptography: prove something was true **without disclosing anything about it** - Prof. Shafi Goldwasser (MIT)*

*Part of the idea behind **homomorphic encryption**, is to bring encryption to the places where it's increasingly needed most.*

source: Encryption pioneer aims to end our data dilemma with cryptography's holy grail

Terminology

homomorphism - Greek: “same”; “form” or “shape”

homomorphic encryption - the encryption method

ciphertext - the encrypted data

plaintext - the original data

orthoginal rotation - transformation in which underlying/latent variables remain unrelated to each other, aid in identification of simple structure/pattern/solution

Homomorphic encryption (HE)

- allows computation on encrypted data
  - **without access to the secret key**

- data remains confidential (encrypted) while being processed
  - enabling useful tasks to be accomplished with data residing in *“untrusted”* environments

[Paper] *Private Genomes and Public SNPs: Homomorphic Encryption of Genotypes and Phenotypes for Shared Quantitative Genetics*

Richard Mott, Christian Fischer, Pjotr Prins and Robert William Davies
Genetics (Early online April 23, 2020)
https://doi.org/10.1534/genetics.120.303153

Software: https://github.com/encryption4genetics
Data (mouse): https://rdr.ucl.ac.uk/articles/Mouse_Platelet_Dataset/11907687

[Paper] Public Genomes and Private SNPs

- not the first time this has been proposed (Further reading)
- limitations with prior HE approaches
    - loss of information, making only “simple” analyses possbile
        - i.e. no retention of population structure
    - most are slower than analysis on unencrypted data

[Paper] Public Genomes and Private SNPs

- retention of layers of information that can be analysed
    1. unchanged likelihood of quantitative trait data
    2. LD
    3. association between variants and phenotypes
    4. h^2 (heritability)
- if data encrypted with HE is on a cloud that becomes compromised, stolen ciphertext **should** be valueless
- possible to share and analyse federated independently-transformed ciphertexts
    - with a few caveats…

[Paper] Public Genomes and Private SNPs - limitations

- **orthagonal rotation**
    - logistic regression is a no go on HE data (**not homomorphic**)
    - method is not **properly** secure
        - private variants not securely encrypted
    - cannot handle missing data -> imputation required prior
    - can only analyse subsets if encoded separately
        - i.e. can’t analyse sex if not split prior
        - though can include sex as covar and account in modelling

[Paper] Public Genomes and Private SNPs - limitations

- **mixed-model transformation**
    - more secure BUT
    - lose the ability to perform more complex analses on encrypted data
        - no variance components or h^2, loss of LD info

Challenge: break the encryption!

Such a move - towards the idea that an allele's effects are public property ***whilst an individual's genotypes are private*** *- is more important than the encryption mechanism used to attain it.*

Further reading

- Armknecht, F., Boyd, C., Carr, C., *et al*. (2015) **A Guide to Fully HomomorphicEncryption** (link)
- Kim, M., & Lauter, K. (2015). **Private genome analysis through homomorphic encryption.** BMC medical informatics and decision making, 15 Suppl 5(Suppl 5), S3. https://doi.org/10.1186/1472-6947-15-S5-S3
- Çetin, G.S., Chen, H., Laine, K. *et al*. Private queries on encrypted genomic data. BMC Med Genomics 10, 45 (2017). https://doi.org/10.1186/s12920-017-0276-z
- Bonte, C., Makri, E., Ardeshirdavani, A., *et al*. (2018). **Towards practical privacy-preserving genome-wide association study.** BMC bioinformatics, 19(1), 537. https://doi.org/10.1186/s12859-018-2541-3
- Park, S., Kim, M., Seo, S., *et al*. (2019). **A secure SNP panel scheme using homomorphically encrypted K-mers without SNP calling on the user side.** BMC genomics, 20(Suppl 2), 188. https://doi.org/10.1186/s12864-019-5473-z
- Mamo, N., Martin, G.M., Desira, M. *et al*. **Dwarna: a blockchain solution for dynamic consent in biobanking.** Eur J Hum Genet 28, 609–626 (2020). https://doi.org/10.1038/s41431-019-0560-9