Overview of 4 potential routes for increase researcher access
Outline

• Overarching goals
• Open Access
• Streamlined access
• Researcher commons
• Central server
Goals

• Improve researcher access to datasets consistent with consents
  • Improves targeted research
  • Required for serendipitous studies

• Compliance with consents is not a zero-sum game with researcher access

• Research participants overwhelming want to see this happen
  • This fact is rediscovered multiple times
80/20 solution

• We should not let the 1% or 10% or even 20% of “edge cases” prevent a good solution for 80% or more of the data. Instead find a pragmatic, acceptable solution for the majority of the data, and special case the ones which don’t fit.

• Historical consents – always a bad idea to be held hostage to historical scenarios. Reconsenting is annoying by often feasible.

• Special cases – Drug addiction, sexual behaviour etc. should not dominate the debate – place into a special case scenario, and expect to handle these differently
Open Access

EBI is an Outstation of the European Molecular Biology Laboratory.
Open Access proposal

- Have anonymised identifiers for genotype and phenotype information
- Ensure the consent process informs the participant of risks
- Totally maximises the serendipity for reuse.
- Is *already* in widespread use for molecular-only studies (HapMap etc)
- ? Could be extended for other normal phenotypes?
- ? Is being extended for disease phenotypes (PGP)
Pros and Cons

• Pros
  • 0 headache in researcher access
  • Maximal reuse of data
  • Most likely to generate serendipitous discoveries

• Cons
  • Small but higher than other schemes risk of participant harm
  • Unknown risk of lack of participation, in particular by disadvantaged groups
  • Harder sell to local IRB boards and current practice
Streamline current system

- Sensible proposal for improving the current system
- 8 major points: consolidate DACs, share more language + terms, have broad consents
- Proposal to change policy (Homer et al): Release genotype numbers.
Pros and Cons

- Pros
  - Improves researcher access
  - Sets up future broad consents
  - Releasing genotype numbers (and therefore Pvalues) provides for broad reuse

- Cons
  - Potential, v. low risk of participation in a study via genotype number and genotypes
  - Perpetuates current system, providing less impetus for deeper reform
Researcher Commons

• In effect, pre-authorisation of researchers for broad consents
• Can be consistent with current broad consents (so no change in relationship between researcher and participant)
• Practical benefits for broad research use
• Needs a certification authority
• Can be internationalised
Pros and Cons

• Pros
  • Improves researcher access
  • Serendipitous research more easily achieved
  • Provides context for centralised (institutional or broader) systems

• Cons
  • Reputational risk of researchers changing the rules to suite them
  • Perpetuates current system, providing less impetus for deeper reform
Central server that provides analysis results

- Different levels:
  - Imputation
  - Stats models
  - Pvalue server
  - Curated variants

- Particular suggestion
  - Low level data kept private
  - People can use a cloud like infrastructure for flexibility
Pros and Cons

**Pros**
- Enables more research over datasets
- Might provide a mid-level access option (cloud not raw access)
- Heavy lifting happens once

**Cons**
- Centralising might cause I/O or people (helpdesk) bottleneck
- (Focused on high to mid level analysts – Pvalue lists just as useable)
Ewan Birney’s overview
These are not mutually incompatible

- Streamlining research is good
  - And sets up the way for Researcher Commons
  - Genotypes/Pvalues good
- Researcher commons is good
  - And will change the view of broad sets of data
- Open access is great, and already being used
  - And should continue – molecular/normal phenotypes?
- Central servers are worth trying
  - Not the only solution
Something old, something new

- Established resources
  - dbGap, EGA
  - UCSC, Ensembl, Entrez

- Emerging resources
  - BioSamples (EBI/NCBI)
  - PRIDE, ICGC Portal, CGHub, Galaxy …

- New resources
  - <your whizzy idea here>
These are not mutually incompatible

- Streamlining research is good
  - And sets up the way for Researcher Commons
  - Genotypes/Pvalues good
- Researcher commons is good
  - And will change the view of broad sets of data
- Open access is great, and already being used
  - And should continue – molecular/normal phenotypes?
- Central servers are worth trying
  - Not the only solution