Genetic Data

Data size necessitate cloud

- GDPR vs. HIPAA, Common rule (institutional IRB variability)
  - What is anonymization for genetic data: no ability to recontact, to combine with other sources of data;
  - Interpretation of “harm” vs. inappropriate use/violation of DUA;
  - Social stigma;
  - Landscape of re-identifiability – consents and risks to privacy;
  - Balance between discovery and protection of privacy – federal regulation;
  - Vulnerable populations: benefits (representation), risk (obvious). Partnerships with group/populations;
  - Incidental findings and return of results – need for best practices (CLIA-certification)

Existing resources/maximization

- Variety of existing cloud-based resources
- COST - “Sandbox” environment (develop/test prototypes in small data, scalability, assess vulnerabilities)
- Approaches for piping data/summaries from local cluster to cloud – compatability with GDPR

Genetic Data

- Reproducibility and Standards
  - Harmonization of approaches to harmonizing (notebook, github)
  - External audit/crowd-source – readability/code-review, safety (trade-off)
  - Infrastructure, incl. during peer and proposal review
  - Funding and resources – life of data?
  - What are responsibilities of data owners, researchers, holders?
  - Crystal-gazing: how to craft a future-looking consent as discovery accelerates?

Next steps

Compliance around GDPR and regulatory standards;

Best practices: (a) code sharing with protections; (b) incidental findings;