Genomics Data

Scale, Availability, Time (Future?), and Privacy

- **Scale**
  - Nucleic acid sequence (& other field-specific) databases [Nature DB list]
  - Sequences (+derivatives, i.e. variants, assemblies, etc?)
    - Used vs *deposited* (more feasible)
    - *Size* (comparison with other fields is straightforward) vs number of bases
    - [Muir *et al*, *Genome Biology* (2016)] & [Eisentstein, *Nature* (2015)]
- **Availability**
  - Ratio of publically available vs privately held data (ideas?)
  - Three levels of publicity: (1) public, (2) permission-requiring, (3) unpublished
  - Time needed to grant access (manual inspection of known databases)
- **Methods & Platforms**
  - Data processing platforms [TechCrunch]
  - Academic papers [Text mining of titles & keywords]
  - Academic tools/methods [Text mining of titles & keywords]
  - Tech startups/companies (?) [TechCrunch]