TOPMed
Data overview and access

TOPMed Data Coordinating Center
University of Washington

October 26th, 2020
ASHG Ancillary Session
TOPMed Website - [www.nhlbiwgs.org](http://www.nhlbiwgs.org)

About TOPMed

*Updated 08/10/2020*

**Contents**

- Overview
- Study Characteristics
  - Study Designs
  - Participant Diversity
- Whole Genome Sequencing
- Resources for the Scientific Community

Central hub of information for the scientific community

- TOPMed Projects and Studies
- WGS and Omics methods documentation
- Data access instructions
- Publications and Abstracts
TOPMed: Participating Studies

- Studies added in yearly phases
- First phase in 2015
- Currently > 80 participating studies
- ~155K study participants
- Range of HLBS phenotypes

TOPMed: Participant Diversity

- Geographic, racial/ethnic, and genetic diversity
- ~60% non-European ancestry participants
- 18 countries represented

See Session 18, Alyna Khan presenting “Guidelines on the use and reporting of race, ethnicity, and ancestry in the NHLBI Trans-Omics for Precision Medicine (TOPMed) program”
TOPMed: WGS Data Production

Freeze 5b: 32 studies, 56K participants
All studies released on dbGaP

Freeze8: 72 studies, 138K participants
70/72 studies released on dbGaP

More info: https://www.nhlbiwgs.org/topmed-data-access-scientific-community
Multi-omics Integration

- A unique resource for integrative omics analysis and discovery in diverse participants
- Studies with multiple omics types on same participants (example: FHS)
TOPMed: Harmonized Phenotypes

- DCC harmonized >100 phenotype variables across dozens of TOPMed studies
  - Common covariates, demographics, inflammation, lipids, blood pressure, blood cell count, VTE, atherosclerosis
- Being made available through dbGaP and BioData Catalyst
- Publication describing robust, reproducible approach (currently in pre-print)

A system for phenotype harmonization in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program


doi: https://doi.org/10.1101/2020.06.18.146423

More info: www.nhlbiwgs.org/dcc-pheno
DCC Phenotype Tagging

DCC coordinated TOPMed effort to tag >16K dbGaP study variables

Phenotype tags mapped to a standard ontology (UMLS)

Results available through dbGaP searching tools

More info: www.nhlbiwgs.org/dcc-phenotype-tagging-details
How to access to TOPMed data

➔ Available through **study-specific accessions**
  ◆ phsXXXXXX
  ◆ All molecular data through TOPMed accession
  ◆ Phenotype data may be in TOPMed or pre-existing accessions

➔ Submit **dbGaP application** for access

➔ **Data Use Limitations (DULs)** vary by study

➔ dbGaP applications reviewed by **NHLBI Data Access Committee (DAC)**

More info: [https://www.nhlbiwgs.org/topmed-data-access-scientific-community](https://www.nhlbiwgs.org/topmed-data-access-scientific-community)
TOPMed Data Use Limitations

● Heterogeneity of DULs across TOPMed

● Compilation of diverse studies with unique histories, source populations, and informed consent processes

● Proposed research uses must align with DULs and participant consents

● Some studies further require
  ○ documentation of local IRB approval (-IRB)
  ○ letter of collaboration (-COL)

● Look for “Data Use Certification (DUC) Agreement” on dbGaP study pages
TOPMed Publications

- 48 publications and growing
  - Word cloud of abstracts
- 22 pre-prints in review

More info: [www.nhlbiwgs.org/publications](http://www.nhlbiwgs.org/publications)
TOPMed Genomic Summary Results (GSR)

- TOPMed GSR available through dbGaP phs0001974
  - e.g. association test results
- Controlled-access
  - numerous TOPMed studies are “sensitive” for GSR sharing under NIH Genomic Data Sharing policy
  - GRU consent
- Top hits publicly available on dbGaP Genome Browser

Inherited causes of clonal haematopoiesis in 97,691 whole genomes

Alexander G. Bick, Joshua S. Weinstock, [...] Pradeep Natarajan

Nature (2020) | Cite this article
4117 Accesses | 1 Citations | 186 Altmetric | Metrics

dbGaP Data Access Request to access full association results

View top hits in publicly available dbGaP Genome Browser
Acknowledgements

- TOPMed study investigators
- TOPMed study participants
- NIH/NHLBI R01HL-120393; U01HL-120393; HHSN268201800001I
- Figures in this slide deck
  - Ken Rice
  - Caitlin McHugh
  - Catherine Tong
  - Alyna Khan

TOPMed DCC

- Faculty leadership: Susanne May (PI), Bruce Weir, Ken Rice, Bruce Psaty, Tim Thornton
- Project management: Matt Conomos, Sarah Nelson, Ben Heavner, Quenna Wong, Catherine Tong
- Analysts: Caitlin McHugh, Stephanie Gogarten, Deepti Jain, Adrienne Stilp, Jen Brody, Josh Bis, Dave Levine
- Administrative: Michael Bowers, Jenn Purnell, Kate Wehr, David Beame, Addison Keely
- ELSI leadership: Malia Fullerton

Former TOPMed DCC members

- Senior leadership and expertise: Cathy Laurie, Cecelia Laurie, Susan Heckbert
- Phenotype harmonization: Leslie Emery, Jai Broome, Erin Buth, Alyna Khan, Fei Fei Wang
TOPMed DCC including former members