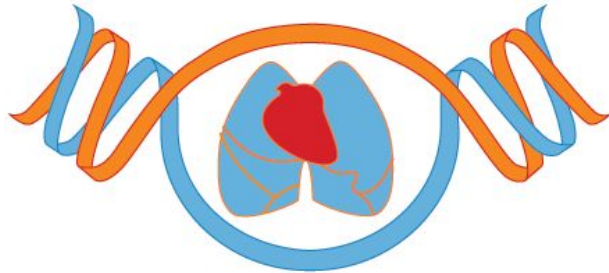


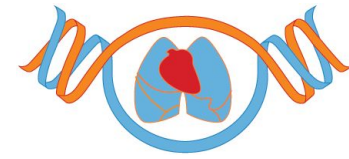
# TOPMed

## 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

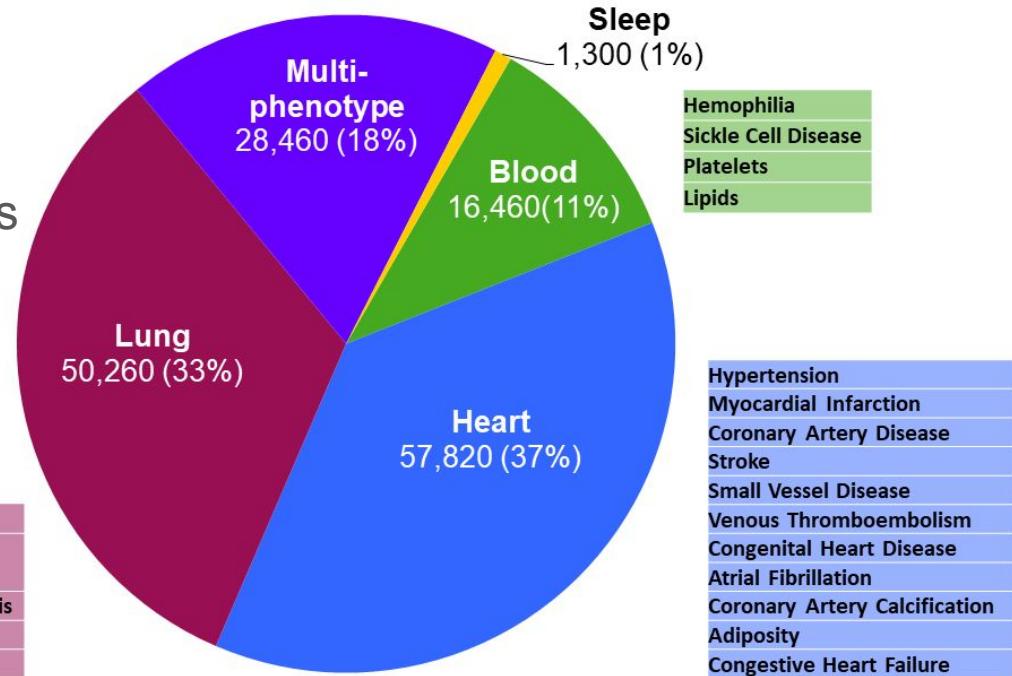
- TOPMed Projects and Studies
- WGS and Omics methods documentation
- Data access instructions
- Publications and Abstracts



# TOPMed: Participating Studies

Phase 1-6: ~155K Study Participants

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

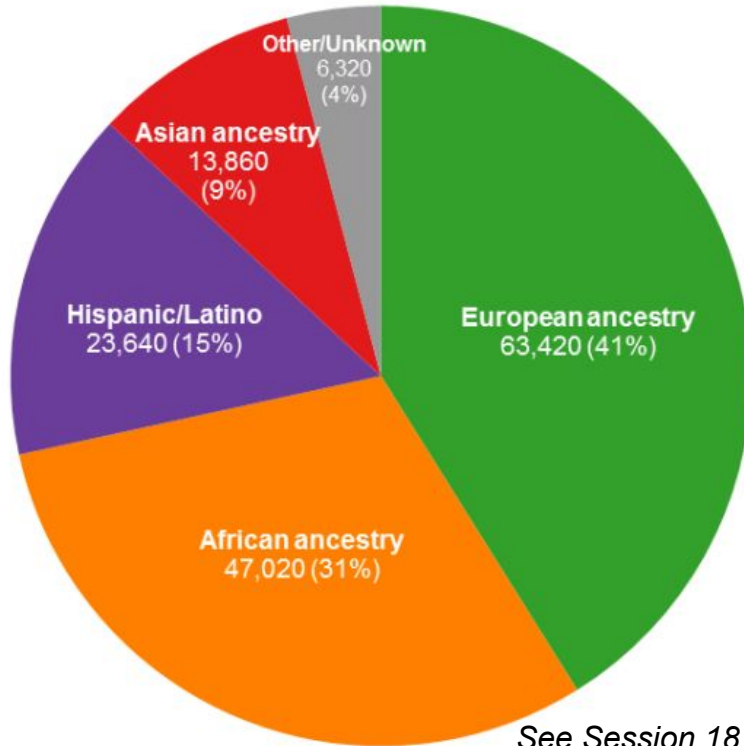


More info: [www.nhlbiwgs.org/](http://www.nhlbiwgs.org/), “About TOPMed”; <https://www.nhlbiwgs.org/group/project-studies>

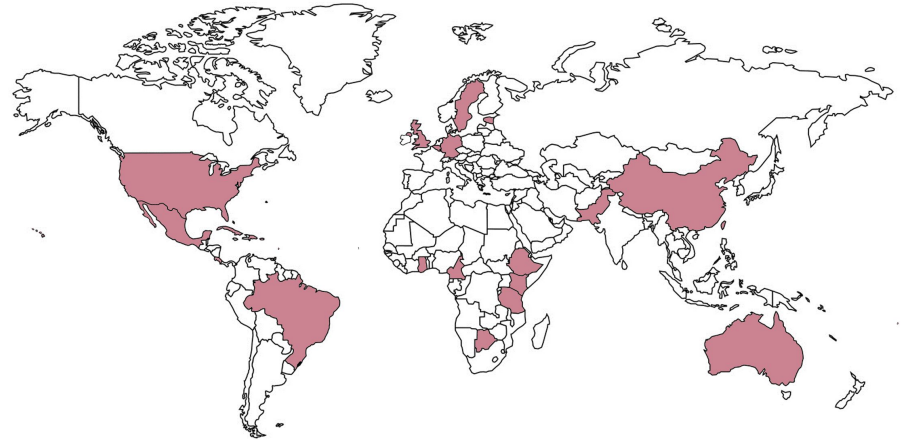


# TOPMed: Participant Diversity

Phases 1-6 (~155K Participants)

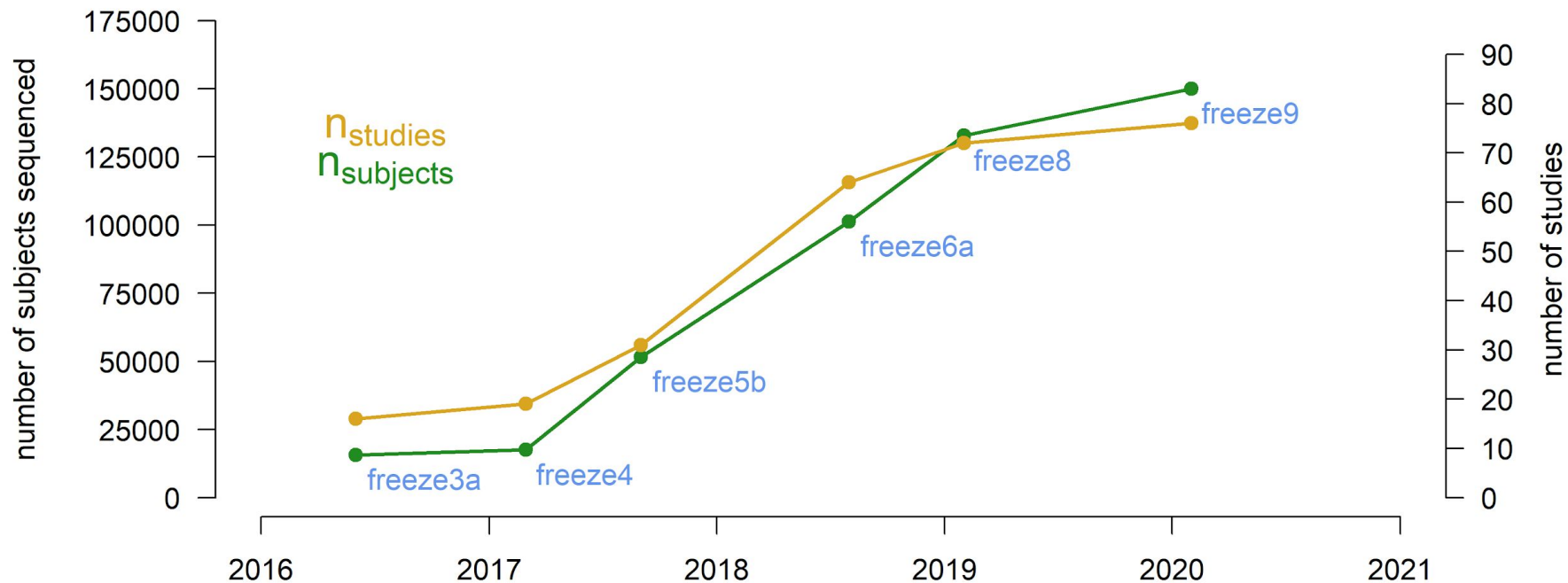


- Geographic, racial/ethnic, and genetic diversity
- ~60% non-European ancestry participants
- 18 countries represented

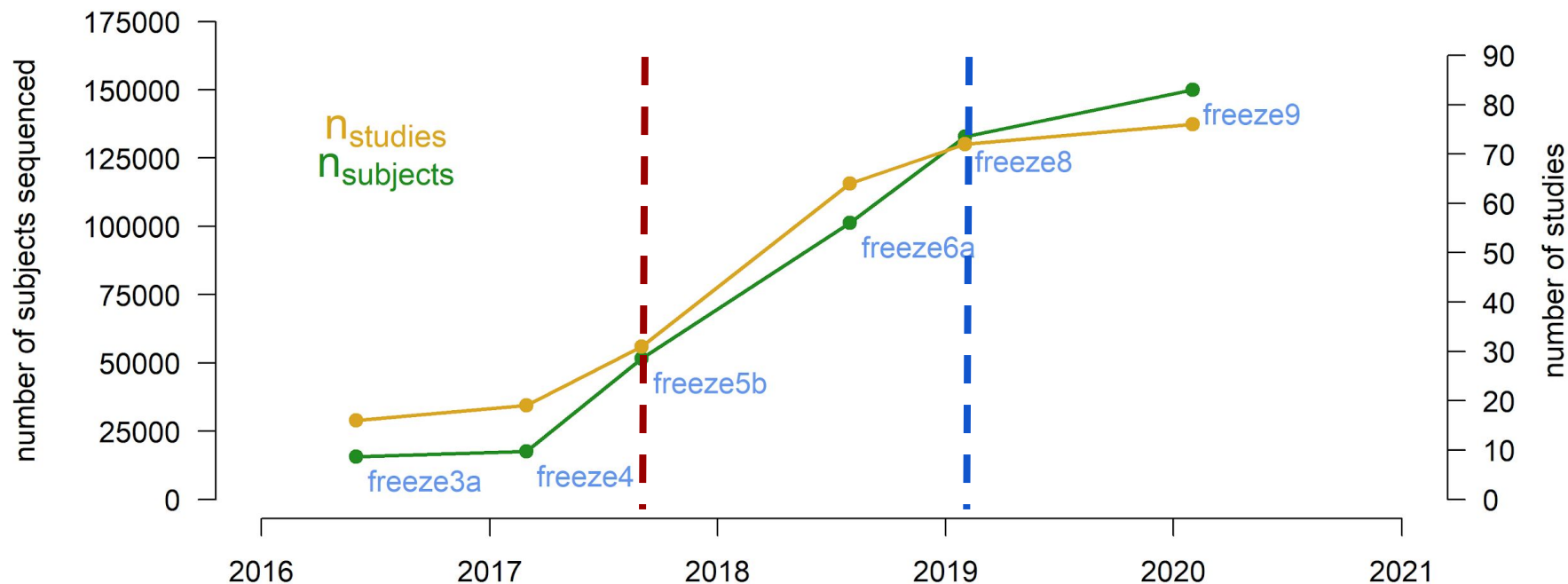


*See Session 18, Alyna Khan presenting “Guidelines on the use and reporting of race, ethnicity, and ancestry in the NHLBI Trans-Omics for Precision Medicine (TOPMed) program”*

# TOPMed: WGS Data Production



# 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**



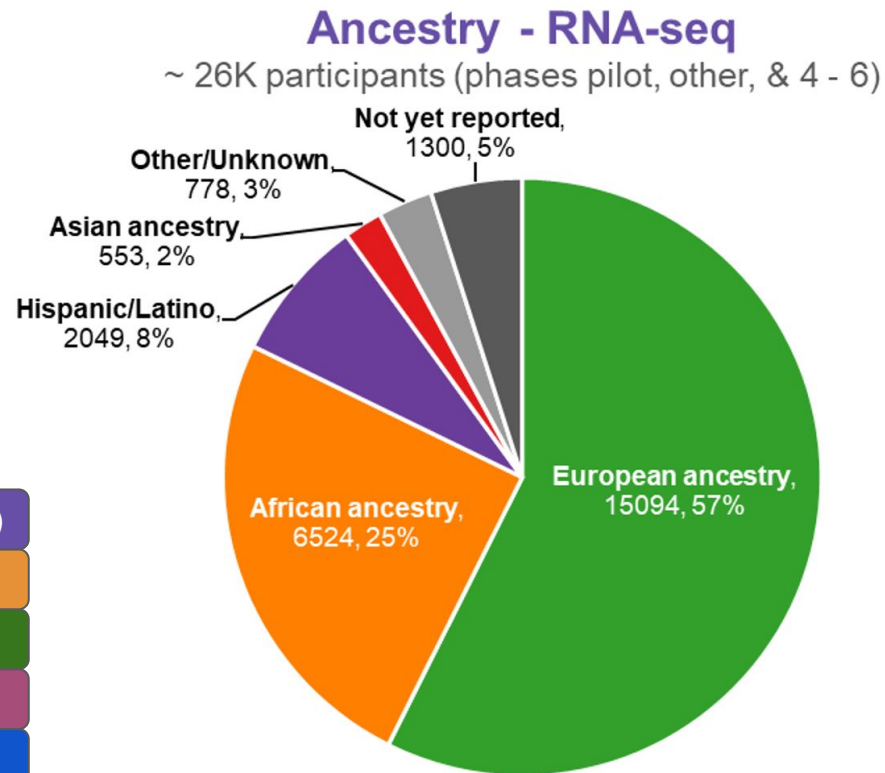
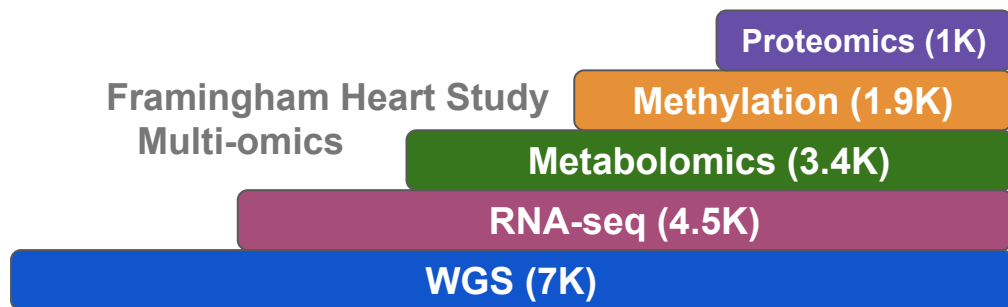
More info: <https://www.nhlbiwgs.org/topmed-data-access-scientific-community>



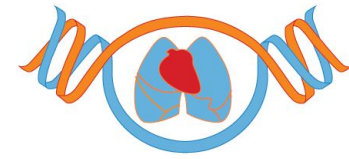
# Multi-omics Integration



- 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**

 Adrienne M. Stilp, Leslie S. Emery, Jai G. Broome, Erin J. Buth, Alyna T. Khan, Cecelia A. Laurie, Fei Fei Wang, Quenna Wong, Dongquan Chen, Catherine M. D'Augustine, Nancy L. Heard-Costa, Chancellor R. Hohensee, William Craig Johnson, Lucia D. Juarez, Jingmin Liu, Karen M. Mutalik, Laura M. Raffield, Kerri L. Wiggins, Paul S. de Vries, Tanika N. Kelly, Charles Kooperberg, Pradeep Natarajan, Gina M. Peloso, Patricia A. Peyser, Alex P. Reiner, Donna K. Arnett, Stella Aslibekyan, Kathleen C. Barnes, Lawrence F. Bielak, Joshua C. Bis, Brian E. Cade, Ming-Huei Chen, Adolfo Correa, L. Adrienne Cupples, Mariza de Andrade, Patrick T. Ellinor, Myriam Fornage, Nora Franceschini, Weiniu Gan, Santhi K. Ganesh, Jan Graffelman, Megan L. Grove, Xiuqing Guo, Nicola L. Hawley, Wan-Ling Hsu, Rebecca D. Jackson, Cashell E. Jaquish, Andrew D. Johnson, Sharon LR Kardia, Shannon Kelly, Jiwon Lee, Rasika A. Mathias, Stephen T. McGarvey, Braxton D. Mitchell, May E. Montasser, Alanna C. Morrison, Kari E. North, Seyed Mehdi Nouraei, Elizabeth C. Oelsner, Nathan Pankratz, Stephen S. Rich, Jerome I. Rotter, Jennifer A. Smith, Kent D. Taylor, Ramachandran S. Vasan, Daniel E. Weeks, Scott T. Weiss, Carla G. Wilson, Lisa R. Yanek, Bruce M. Psaty, Susan R. Heckbert, Cathy C. Laurie

**doi:** <https://doi.org/10.1101/2020.06.18.146423>



More info: [www.nhlbiwgs.org/dcc-pheno](http://www.nhlbiwgs.org/dcc-pheno)





# DCC Phenotype Tagging

dbGaP Advanced Search ?

C0005890

Common Data Elements IS UMLS

C0005890

Show All Filters

**Study (16)**

Sort By: Alphabetical

- ☐ Atherosclerosis Risk in Communities (ARIC) Cohort (phs000280.v7.p1) (6)
- ☐ CARDIA Cohort (phs000285.v3.p2) (13)
- ☐ Cardiovascular Health Study (CHS) Cohort (phs000287.v7.p1) (5)
- ☐ Framingham Cohort (phs000007.v31.p12) (62)
- ☐ Genetic Epidemiology Network of Arteriopathy

**ID Variables**

☐ Show

**Common Data Elements (3)**

Sort By: Alphabetical clear

- ☐ LOINC (70)
- ☐ PhenX (70)
- ☒ UMLS (119)

Studies (0) Phenotype Datasets (0) Variables (119) Molecular Datasets (0) Analyses (0) Documents (0)

Save Results Save Query

**1 Height in centimeters**

Name htcn  
Accession phv00122640.v1.p1  
Study Consent DS-HLBS-IRB-NPU --- Disease-specific (heart, lung, blood, and sleep disorders, irb, npu)  
Sample/Subject Count 2945  
Linked Document CARE CFS Data Dictionary, Part 1 (phd003364.1)  
Study NHLBI Cleveland Family Study (CFS) Candidate Gene Association Resource (CARE) (phs000287.v7.p1)  
Dataset CFS\_CARE\_Subject\_Phenotypes (pht001902.v1.p1)  
Embargo Release Date 2011-02-08  
Common Data Element [LOINC](#) [PhenX](#) [UMLS](#)

[Variable page](#) [Study page](#) [Dataset page](#)

**2 STANDING HEIGHT - CM**

Name STHT13  
Accession phv00099550.v1.p1  
Study Consent HMB-MDS --- Health/medical/biomedical (mds) , HMB-NPU-MDS --- Health/medical/biomedical specific (cardiovascular disease, mds) , DS-CVD-NPU-MDS --- Disease-specific (cardiovascular disease, mds)  
Sample/Subject Count 4894  
Linked Document Value Labels (phd002782.1)  
Study Cardiovascular Health Study (CHS) Cohort (phs000287.v7.p1)  
Dataset BASE1 (pht001450.v1.p1)  
Embargo Release Date 2010-08-06  
Common Data Element [LOINC](#) [PhenX](#) [UMLS](#)

[Variable page](#) [Study page](#) [Dataset page](#)

**3 STANDING HEIGHT - CM**

Name STHT13  
Accession phv00100382.v1.p1  
Study Consent HMB-MDS --- Health/medical/biomedical (mds) , HMB-NPU-MDS --- Health/medical/biomedical specific (cardiovascular disease, mds) , DS-CVD-NPU-MDS --- Disease-specific (cardiovascular disease, mds)

- DCC coordinated TOPMed effort to tag >16K dbGaP study variables
- Phenotype tags mapped to a standard ontology (UMLS)
- Results available through dbGaP searching tools

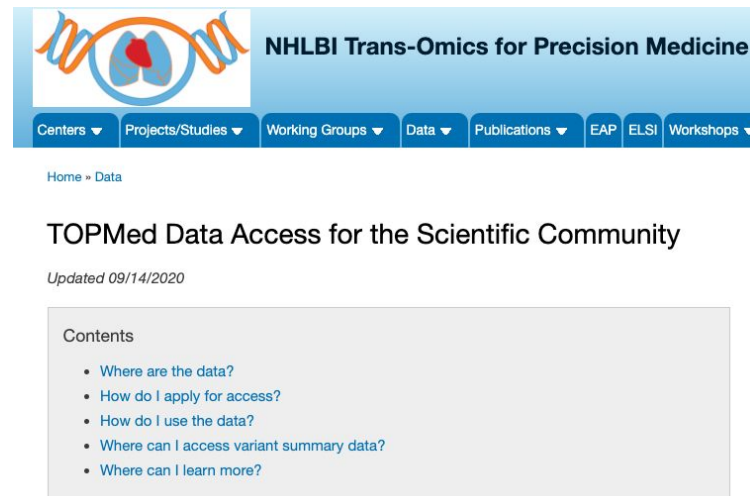


More info: [www.nhlbiwgs.org/dcc-phenotype-tagging-details](http://www.nhlbiwgs.org/dcc-phenotype-tagging-details)



# 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>



# 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



*Image: NHGRI Media Gallery*

# TOPMed Publications

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



More info: [www.nhlbiwqs.org/publications](http://www.nhlbiwqs.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



**NHLBI TOPMed: Genomic Summary Results for the Trans-Omics for Precision Medicine Program**

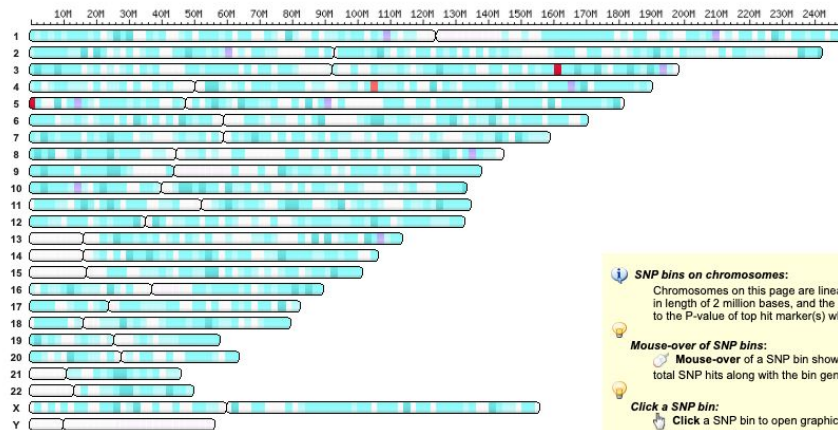
**dbGaP Study Accession:** phs001974.v1.p1

[Request Access](#)

**ANALYSIS:** A Multiethnic Genome-Wide Association Study Of Clonal Hematopoiesis Of Indeterminate Potential In 37 TOPMed Studies Using Freeze 8 Genomes

**METHOD:** Meta-analysis

**STUDY:** NHLBI TOPMed: Genomic Summary Results for the Trans-Omics for Precision Medicine Program (phs001974)



#### SNP bins on chromosomes:

Chromosomes on this page are linearly divided into **SNP bins** in length of 2 million bases, and the SNP bin color corresponds to the P-value of top hit marker(s) when applicable.



#### Mouse-over of SNP bins:

**Mouse-over** of a SNP bin shows the number of top SNP hits and of total SNP hits along with the bin genome co-ordinates.



#### Click a SNP bin:

**Click** a SNP bin to open graphic GaP Browser for details of genotype and phenotype for SNP markers as well as references/utility tools.



More info: [www.nhlbiwgs.org/topmed-genomic-summary-results-public](http://www.nhlbiwgs.org/topmed-genomic-summary-results-public)

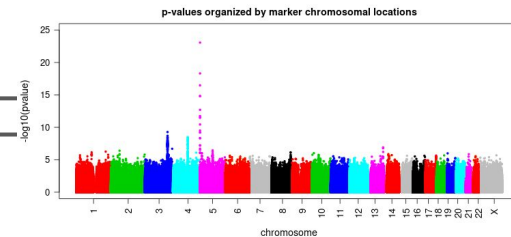
Article | Published: 14 October 2020

# 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



## NHLBI TOPMed: Genomic Summary Results for the Trans-Omics for Precision Medicine Program

dbGaP Study Accession: phs001974.v1.p1

[Request Access](#)

Study version history

[Study](#) [Phenotype Datasets](#) [Variables](#) [Molecular Datasets](#) [Analyses](#) [Documents](#)

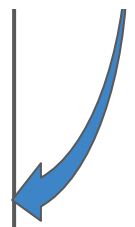
- [Browse all analyses within this study via Advanced Search](#)
- [List all analyses within this study](#)

### Analysis Name and Accession

**Analysis Name:** A Multiethnic Genome-Wide Association Study Of Clonal Hematopoiesis Of Indeterminate Potential In 37 TOPMed Studies Using Freeze 8 Genotypes

**Analysis Accession:** pha004972.1

[View association results in 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



Thank  
You!

## 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

