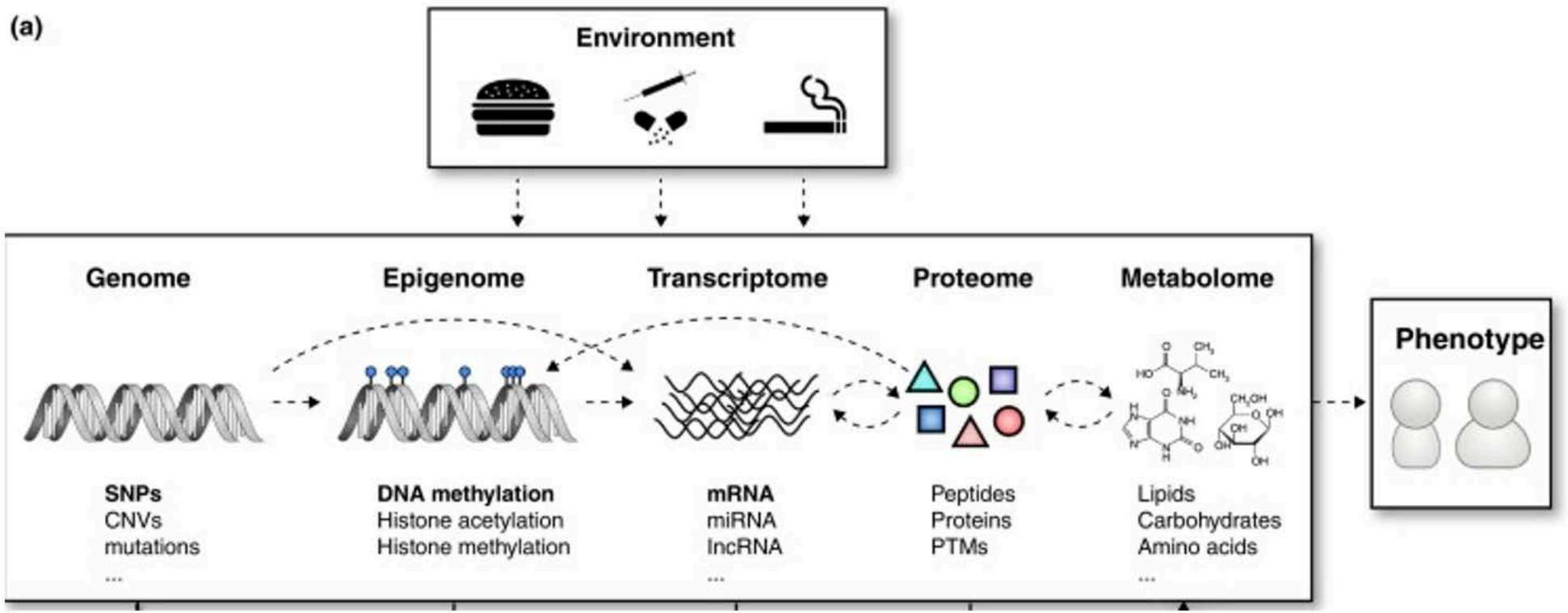
FHS OMIC RESOURCES
Nancy Heard-Costa

OVERVIEW

MODERN VIEW



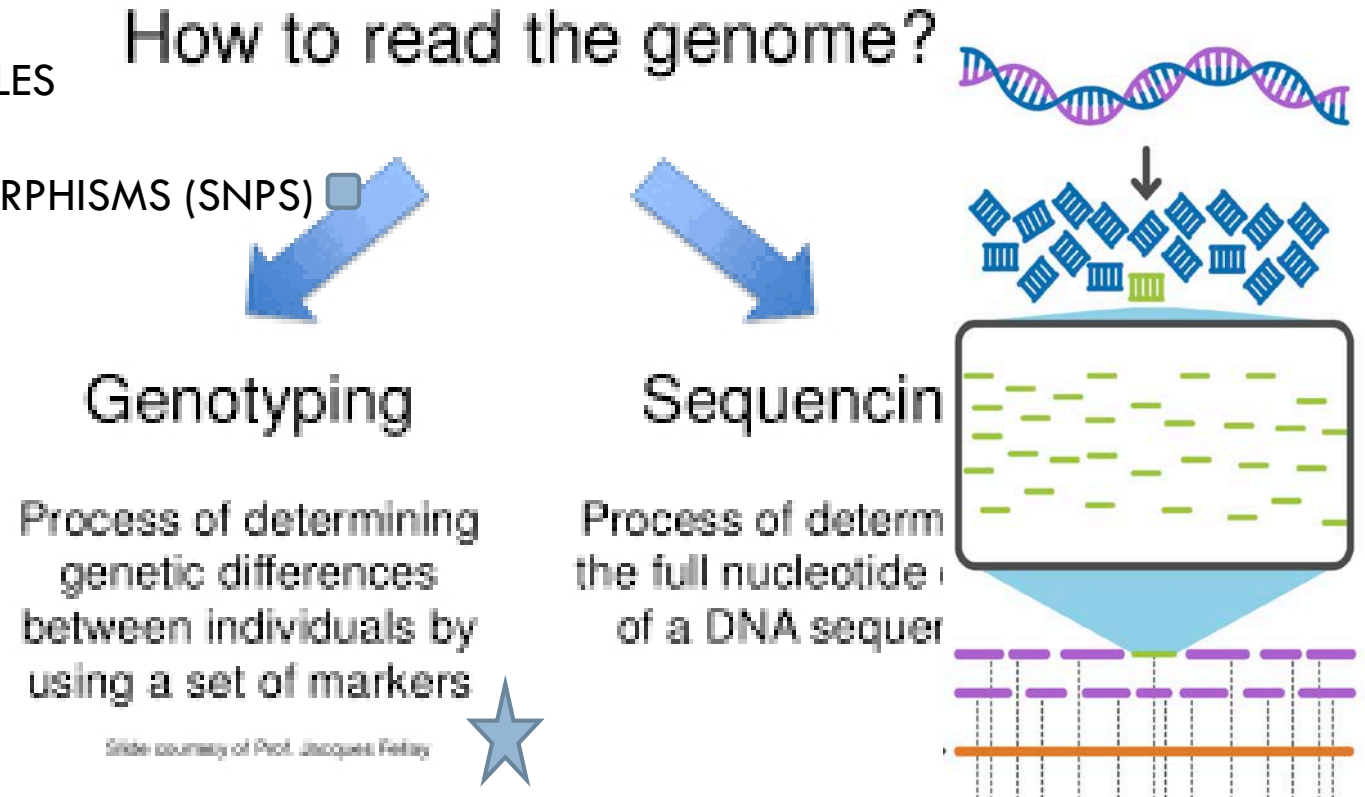
REPETITIVE DNA SEQUENCES - CODOMINANT, ABUNDANT AND MULTI-ALLELIC

VARIATION IN A SINGLE BASE PAIR

GENOMICS

MARKER EXAMPLES

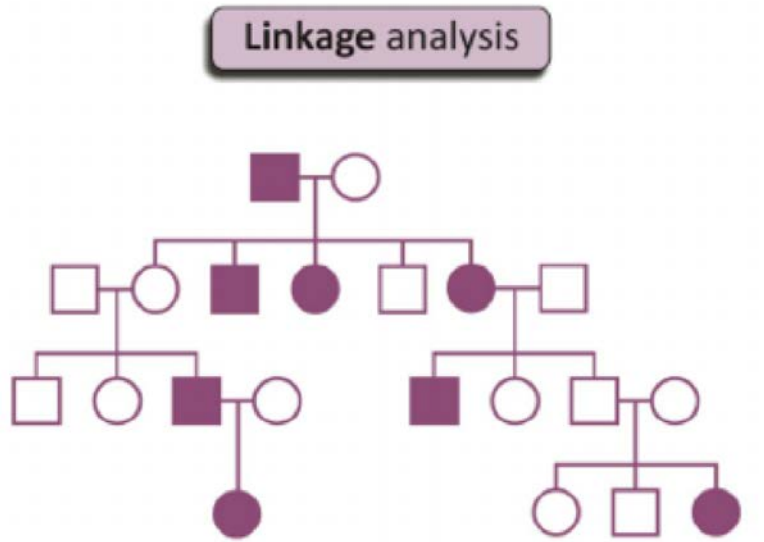
- MICROSATELLITES
- SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)
- IMPUTED GENOTYPES



NOT DIRECTLY ASSAYED – USE REF PANEL HAPLOTYPES AT DENSE SET OF SNPS TO IMPUTE INTO SAMPLES ASSAYED FOR A SUBSET OF SNPS

GENOMICS – A

- LINKAGE ANALYSIS
- ASSOCIATION ANALYSES
  - CANDIDATE GENE ANALYSIS
  - GWAS



Candidate Gene vs GWAS

More Information Online WWW.DIFFERENCEBETWEEN.COM

	Candidate Gene	GWAS
DEFINITION	Candidate gene approach is a technique that investigates the associations between genetic variation within pre-specified genes of interest and phenotypes or disease states	GWAS is a technique that investigates the associations between genetic variations of the entire genome with the disease states
TESTING	Genetic variation is analysed within pre-specified genes of interest	Genetic variation is analysed in the entire genome
STATISTICAL POWER	High	Low
PRIOR KNOWLEDGE OF THE GENES	Needed	Not needed
SELECTION OF GENES	Yes	No

GENOMICS -

- MICROSATELLITES- MARSHFIELD MARKERS
  - ~ 4100 non-OMNI cohorts, ~ 670 markers
- SNPs
  - 9274 non-OMNI : 550K markers
  - OMNI1 & OMNI2 : 628K markers
- IMPUTED
  - 8481 non-OMNI with 550K
  - three reference panels: HAPMAP, 1000G, HRC
- SEQUENCING
  - 7171 WGS across all cohorts

    - Listings above are a subset of all available

GWAS

- GENOME WIDE ASSOCIATION STUDY
- Straightforward Concept: to extensively genotype or haplotype large human population cohorts and *statistically link* specific allelic or other genotypic *variation* to epidemiological data on a disease/trait in the population

› PLoS Genet. 2009 Jun;5(6):e1000539. doi: 10.1371/journal.pgen.1000539. Epub 2009 Jun 26.

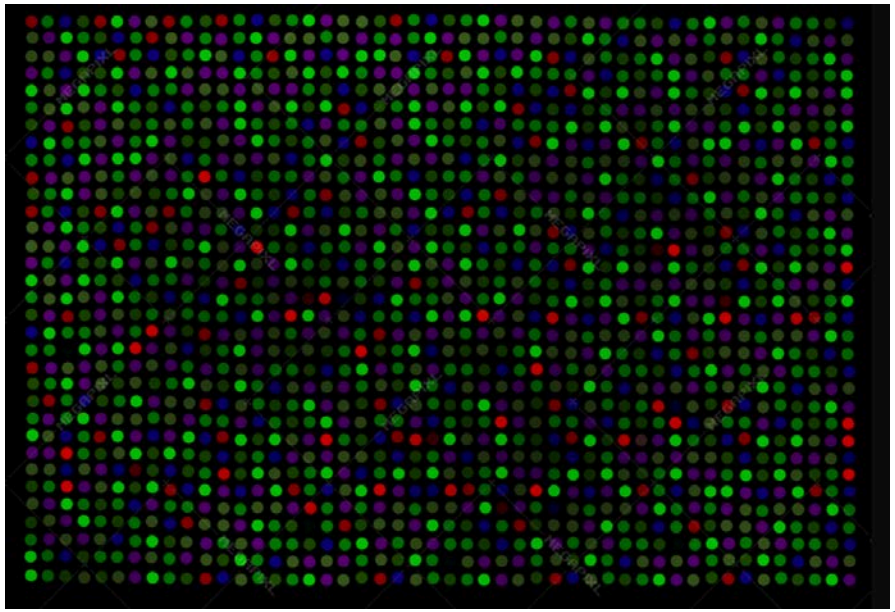
NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium

Nancy L Heard-Costa [1], M Carola Zillikens, Keri L Monda, Asa Johansson, Tamara B Harris, Mao Fu, Talin Haritunians, Mary F Feitosa, Thor Aspelund, Gudny Eiriksdottir, Melissa Garcia, Lenore J Launer, Albert V Smith, Braxton D Mitchell, Patrick F McArdle, Alan R Shuldiner, Suzette J Bielinski, Eric Boerwinkle, Fred Brancati, Ellen W Demerath, James S Pankow, Alice M Arnold, Yii-Der Ida Chen, Nicole L Glazer, Barbara McKnight, Bruce M Psaty, Jerome I Rotter, Najaf Amin, Harry Campbell, Ulf Gyllensten, Cristian Pattaro, Peter P Pramstaller, Igor Rudan, Maksim Struchalin, Veronique Vitart, Xiaoyi Gao, Aldi Kraja, Michael A Province, Qunyuan Zhang, Larry D Atwood, Josée Dupuis, Joel N Hirschhorn, Cashell E Jaquish, Christopher J O'Donnell, Ramachandran S Vasan, Charles C White, Yurii S Aulchenko, Karol Estrada, Albert Hofman, Fernando Rivadeneira, André G Uitterlinden, Jacqueline C M Witteman, Ben A Oostra, Robert C Kaplan, Vilmundur Gudnason, Jeffrey R O'Connell, Ingrid B Borecki, Cornelia M van Duijn, L Adrienne Cupples, Caroline S Fox, Kari E North

Affiliations + expand

PMID: 19557197 PMCID: PMC2695005 DOI: 10.1371/journal.pgen.1000539

-HER OMICS



EPIGENOMICS

- Epigenetics –alterations in gene expression without changing the DNA sequence
- essential for controlling normal development and homeostasis
- not all are permanent
- can respond to changes in behavior or environment
- Can vary between tissues
- Common type: DNA Methylation
  - in a gene promoter, typically acts to repress gene transcription.

EPIGENOMICS

DNA METHYLATION:
GEN2 Ex8/EX9
NOS Ex1/2
GEN3 Ex2

- Epigenetics –alterations in gene expression without chang[illegible]the DNA sequence
- essential for controlling normal development and homeostasis
- not all are permanent
- can respond to changes in behavior or environment
- Can vary between tissues
- Co[illegible]
  - i[illegible]cription.

A DNA methylation biomarker of alcohol consumption

C Liu, R E Marioni, [...] D Levy

Molecular Psychiatry **23**, 422–433 (2018) | Cite this article

13k Accesses | **122** Citations | **39** Altmetric | Metrics

TRANSCRIPTOMICS

- Transcriptome: the set of all RNA transcripts, from protein coding (mRNA) to non-coding RNA including microRNAs (miRNA) , small interfering RNA, and others

- investigate when and where genes are turned on or off in a tissue and quantify that expression
- Each gene may produce more than one variant of mRNA because of alternative splicing, RNA editing, or alternative transcription initiation and termination sites.
  - the transcriptome captures a level of complexity that the simple genome sequence does not

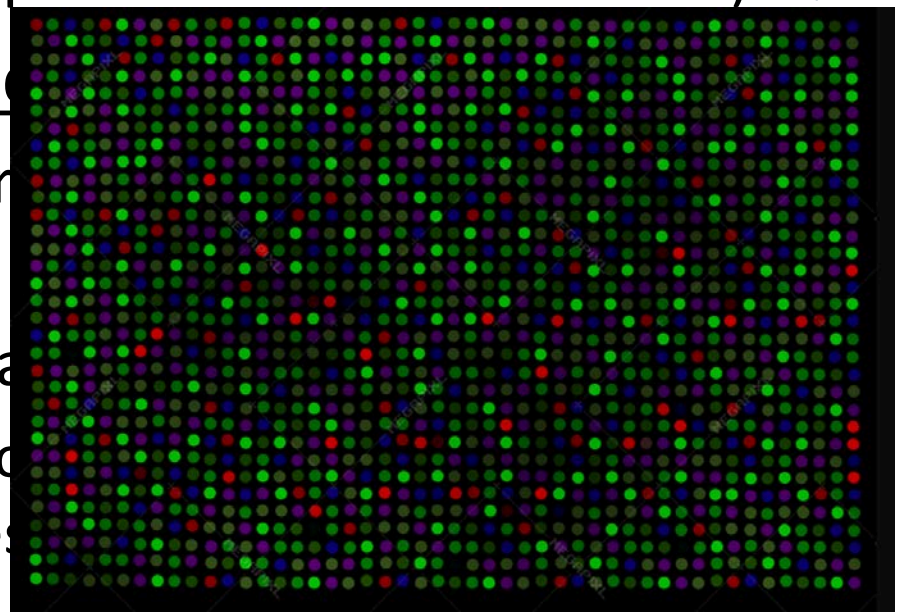
TRANSCRIPTOMICS - ASSAYS

- DNA microarrays/gene expression profiling
  - measure the relative activity(expression) of thousands of previously identified target genes (chips with probe sequences on solid substrate)
- High-throughput RNA sequencing, (RNA-Seq)
  - detects all transcripts in a sample including regulatory non-coding transcripts
    - Provides information on gene sequences as well as their expression level
    - can also identify disease-associated gene fusions, single nucleotide polymorphisms and even allele-specific expression

TRANSCRIPTOMICS - ASSA

TRANSCRIPTOME:
GEN2 Ex8/Ex9
NOS/GEN3 Ex2
OMNI1 Ex3/Ex4

- DNA microarrays/gene expression profiling
  - measure the relative activity(expression) of thousands of previously identified target genes (chips with probe sequences on solid substrate)
- High-throughput RNA sequencing, (RNA-Se
  - detects all transcripts in a sample including r
    transcripts



PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Whole Blood Gene Expression and Atrial Fibrillation: The Framingham Heart Study

Honghuang Lin, Xiaoyan Yin, Kathryn L. Lunetta, Josée Dupuis, David D. McManus, Steven A. Lubitz, Jared W. Magnani, Roby Joehanes, Peter J. Munson, Martin G. Larson, Daniel Levy, Patrick T. Ellinor, Emelia J. Benjamin

Published: May 7, 2014 • https://doi.org/10.1371/journal.pone.0096794

PROTEOMICS

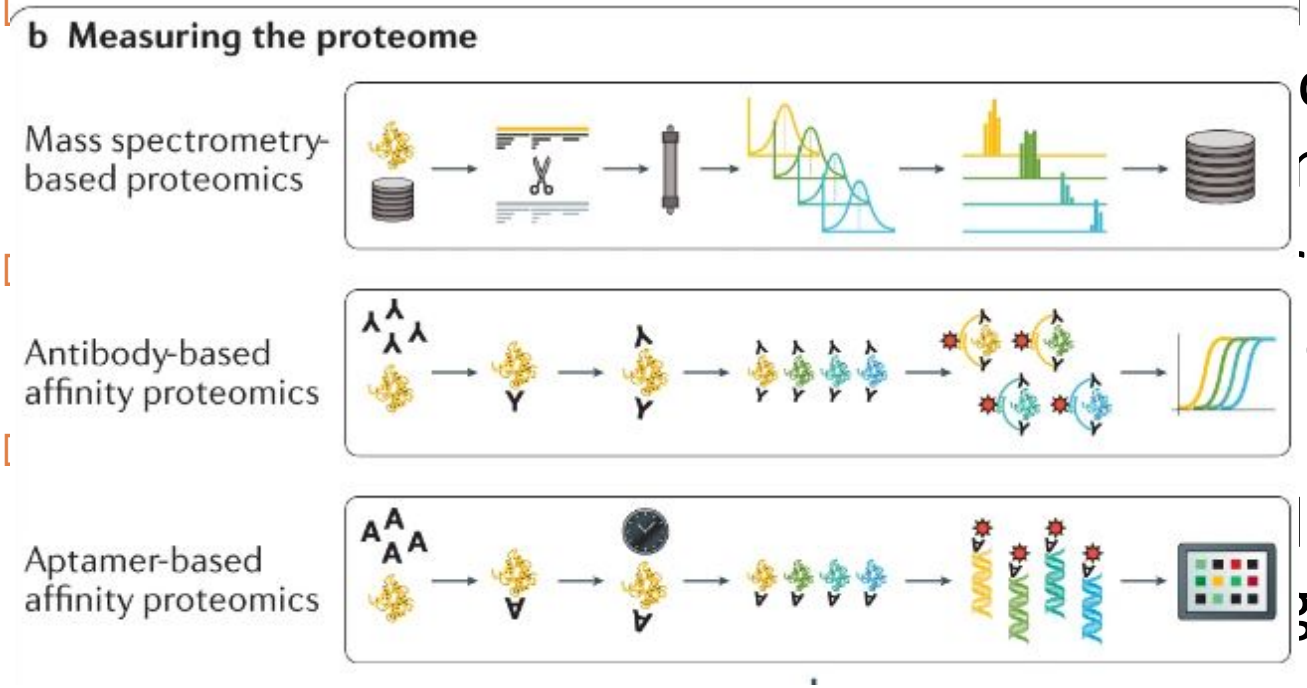
- the systematic identification and quantification of the complete complement of proteins (the proteome) of a biological system (cell, tissue, organ, biological fluid, or organism) at a specific point in time.
- Most biochemical reactions in a cell are regulated by highly specialized proteins, which are the prime mediators of the cellular phenotype.
- Thus the identification, quantitation and characterization of all proteins in a cell are of utmost importance to understand the molecular processes that mediate cellular physiology.

- There are various analytical methods

PROTEOMICS

PROTEOMICS:
GEN2 Ex5-8
GEN3 Ex1/Ex2

- the systematic identification and quantification of the complete of a biological system (cell, n) at a specific point in time.
- egulated by highly specialized of the cellular phenotype.
- characterization of all proteins erstand the molecular gy.



- There are various analytical methods

METABOLOMICS

- the comprehensive analysis of metabolites in a biological specimen
- Metabolites are the small molecule substrates, intermediates and products of cell metabolism.
- offers a more direct measure of the biochemical activities occurring within cells, for the closest possible representation of the molecular phenotype
- Pose significant analytical challenge to measure molecules with disparate physical properties
  - (e.g., polarity ranging from very water soluble organic acids to very nonpolar lipids)

METABOLOMICS

METABOLOMICS:
GEN2 Ex5, Ex8, Ex9
NOS Ex2
GEN3 Ex1/Ex2

- the comprehensive analysis of metabolites in a biologi[illegible] specimen
- Metabolites are the small molecule substrates, inte[illegible]mediates and products of cell metabolism

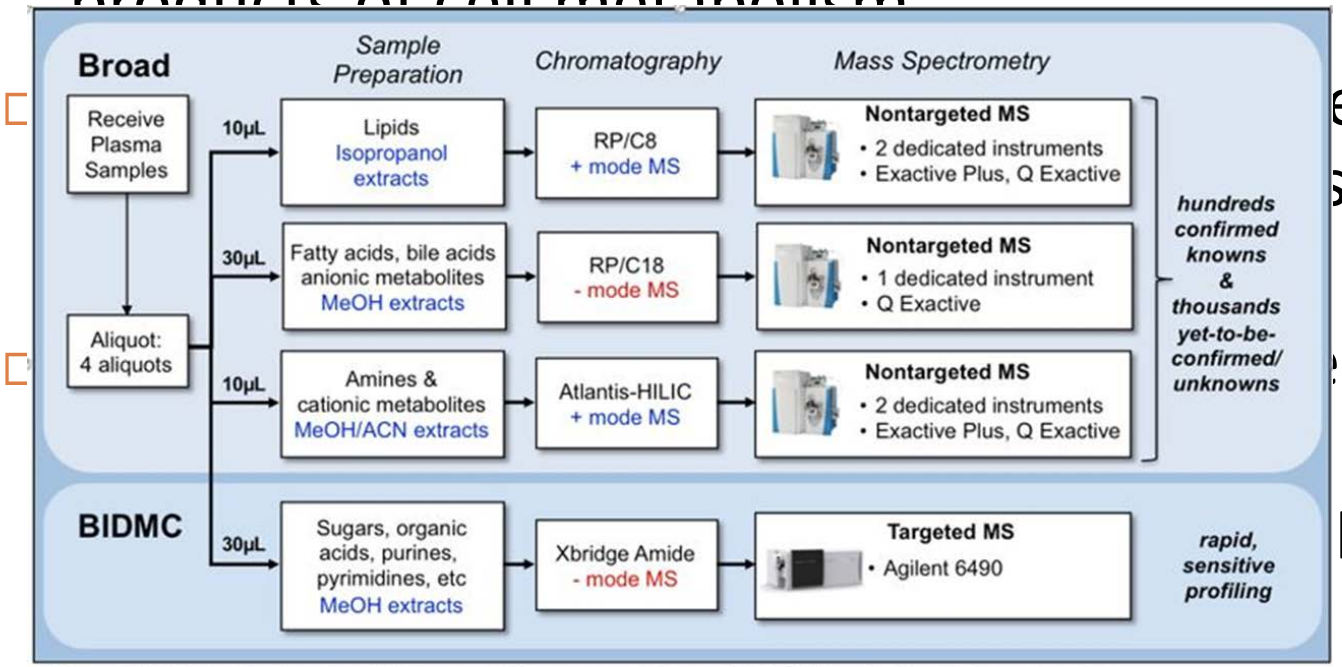
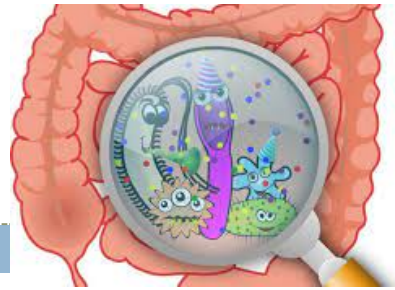


Figure 1. Metabolite profiling workflows and Broad-BIDMC platform integration

- [illegible]emical activities occurring [illegible]sentation of the molecular
- [illegible]easure molecules with
- [illegible]ble organic acids to very

MICROBIOME

- Determines the composition and function of a community of microorganisms in a particular location.
- The human gut is home to a variety of microbes, including bacteria, archaea, fungi microbial eukaryotes and viruses/phages
- microbiota have been considered the most important microecosystem living in symbiosis with the body
- crucial determinant of intestinal inflammation and as a key player in chronic inflammatory liver diseases



MICROBIOME

MICROBIOME:
NOS/GEN3/OMNI2 Ex3

- Determines the composition and function of a community of microorganisms in a particular location.
- The human gut is home to a variety of microbes, including bacteria, archaea, fungi microbial eukaryotes and viruses/phages
- microbiota have been considered the most important microecosystem

ammation and as a key player in

› Cell Host Microbe. 2020 Aug 12;28(2):245-257.e6. doi: 10.1016/j.chom.2020.05.013.
Epub 2020 Jun 15.

Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level

Douglas J Kenny [1], Damian R Plichta [2], Dmitry Shungin [3], Nitzan Koppel [4], A Brantley Hall [2], Beverly Fu [4], Ramachandran S Vasan [5], Stanley Y Shaw [6], Hera Vlamakis [7], Emily P Balskus [8], Ramnik J Xavier [9]

Affiliations + expand

PMID: 32544460 PMCID: PMC7435688 DOI: 10.1016/j.chom.2020.05.013

COMMON THEME OF ASSOCIATION

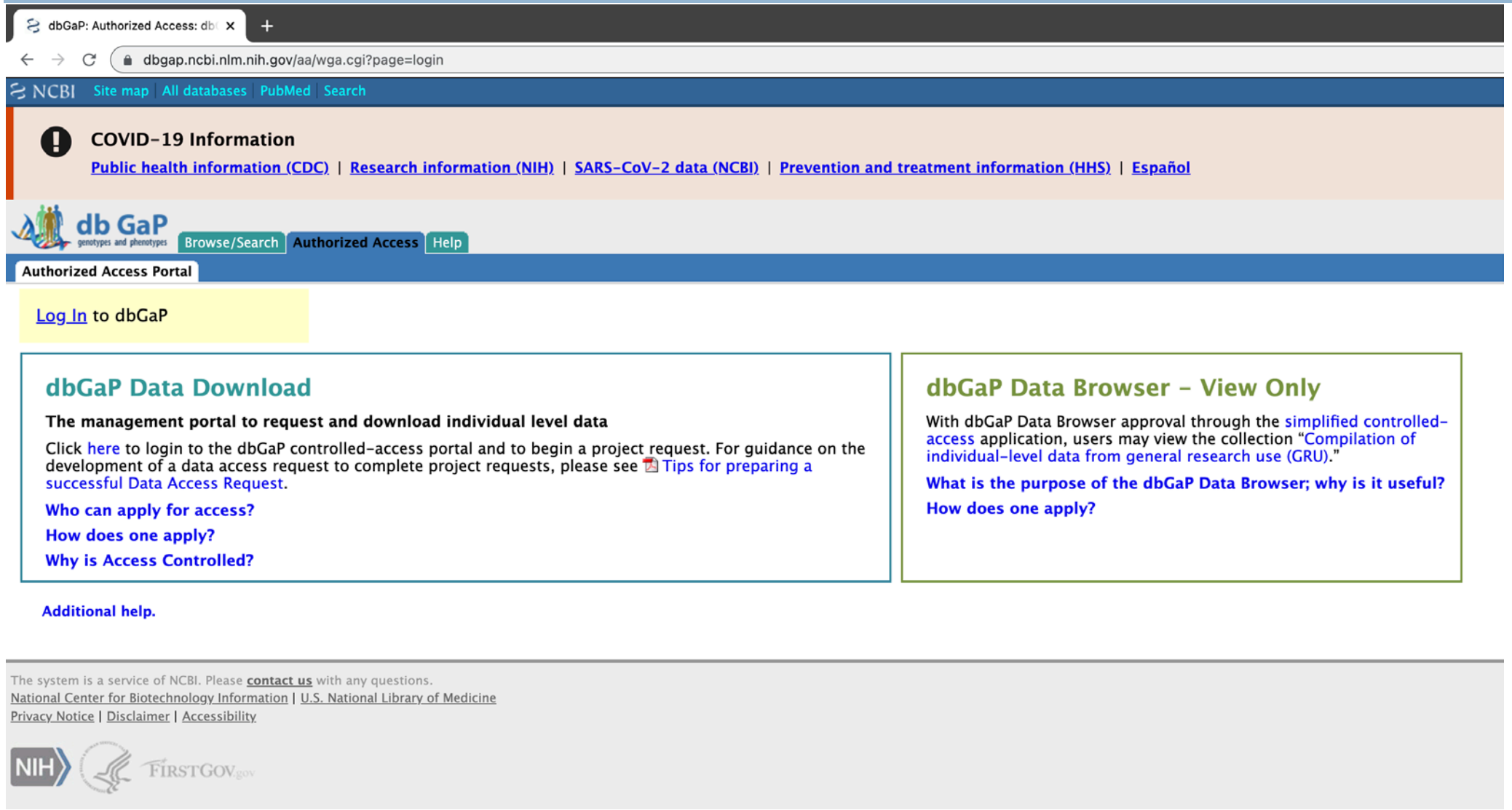
- statistically link *variation* to epidemiological data on a disease/trait
- Variation used depends on OMICS under study
  - Genome: genetic markers
  - Epigenome: DNA methylation markers
  - Transcriptome: gene expression
  - Proteome: proteins
  - Metabolome: metabolites
  - Microbiome: microbiota

MULTI-OMICS

- aka multiomics, integrative omics, "panomics" or 'pan-omics'
- integrated approaches that combine individual omics data, in a sequential or simultaneous manner, to understand the interplay of molecules
- enable a more comprehensive understanding of molecular changes contributing to normal development, cellular response, and disease.
- help assess the flow of information from one omics level to the other and thus help bridge the gap from genotype to phenotype

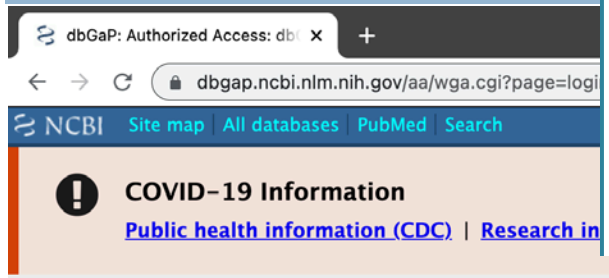
HOW TO ACCESS

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login



HOW TO ACCESS

https://dbgap.nc



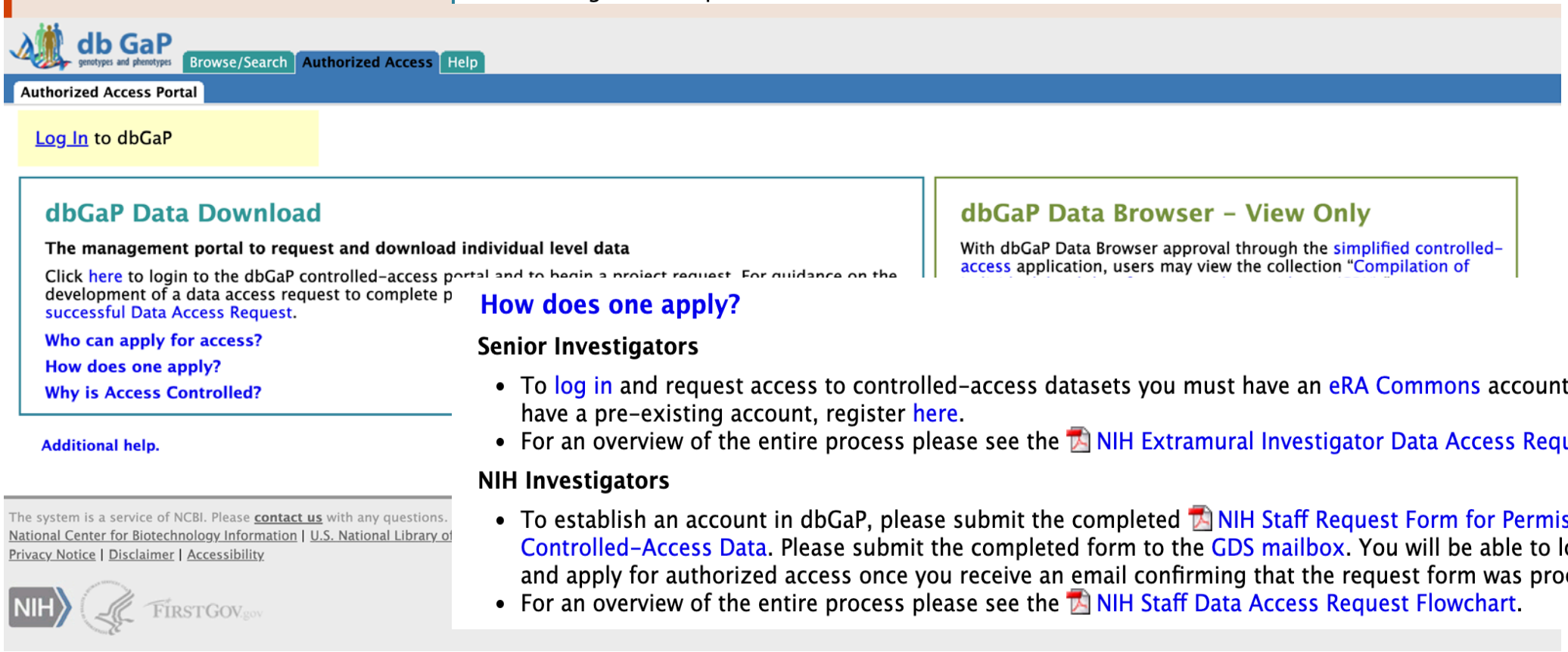
Who can apply for access?

Senior Investigators

- Extramural Investigators must be permanent employees of their institution at a level equivalent to a tenure-track professor or senior scientist with responsibilities that most likely include laboratory administration and oversight. Laboratory staff and trainees such as graduate students and postdoctoral fellows **are not** permitted to submit project requests.

NIH Investigators

- NIH Intramural Investigators must be tenure-track investigators, senior investigators, senior scientists, senior clinicians, or staff scientists.
- NIH extramural scientific staff must have administrative responsibility for the data; have substantial research involvement in the award that generated the data; or need access to carry out research unrelated to their portfolio management responsibilities.



How does one apply?

Senior Investigators

- To log in and request access to controlled-access datasets you must have an eRA Commons account. If you do not have a pre-existing account, register here.
- For an overview of the entire process please see the NIH Extramural Investigator Data Access Request Flowchart.

NIH Investigators

- To establish an account in dbGaP, please submit the completed NIH Staff Request Form for Permission to Access Controlled-Access Data. Please submit the completed form to the GDS mailbox. You will be able to log in to dbGaP and apply for authorized access once you receive an email confirming that the request form was processed.
- For an overview of the entire process please see the NIH Staff Data Access Request Flowchart.

Access Control -NIH Genomic Data

- Respect for, and protection of the interests of research participants are fundamental to NIH's stewardship of human genomic and associated phenotypic data.
- informed consent determines the appropriateness of sharing data
- FHS consent groups – data split into 2 at dbGaP

Consent group	Is IRB required?	Data Access Committee	Number of participants
Health/Medical/Biomedical (IRB, MDS)	Yes	National Heart, Lung, and Blood Institute DAC (nhlbigeneticdata@nhlbi.nih.gov)	13072
Health/Medical/Biomedical (IRB, NPU, MDS)	Yes	National Heart, Lung, and Blood Institute DAC (nhlbigeneticdata@nhlbi.nih.gov)	2072

- The third consent group, no genetic access, not included

```
The value ranges and observation number stated in the manual are based
on the original data set.  In some cases, observation may be deleted
due to participant consent form restrictions.  If observations have
been deleted from this data set, the ranges or observation number may
differ from those stated in this manual.
```

WHERE ARE THE OMICS DATA LOCATED

Summary level phenotypes for the Framingham Cohort study participants can be viewed at the top-level study page phs000007 Framingham Cohort.
Individual level phenotype data and molecular data for all Framingham top-level study and substudies are available by requesting Authorized Access to the Framingham Cohort study phs000007.

The Framingham Cohort is utilized in the following dbGaP substudies. To view genotypes, analysis, expression data, other molecular data, and derived variables collected in these substudies, please click on the following substudies below or in the "Substudies" box located on the right hand side of this top-level study page phs000007 Framingham Cohort.

Substudies

phs000282.v21.p13 : NHLBI Framingham Candidate Gene Association Resource (CARe)
phs000307.v17.p13 : NHLBI Framingham Heart Study Allelic Spectrum Project
phs000342.v20.p13 : NHLBI Framingham SNP Health Association Resource (SHARe)
phs000363.v19.p13 : NHLBI Framingham SABRe CVD
phs000401.v15.p13 : NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (FHS)
phs000651.v12.p13 : Building on GWAS: the U.S. CHARGE consortium - Sequencing (CHARGE-S): FHS
phs000724.v9.p13 : Framingham Offspring Exam 8 DNA Methylation Study
phs001610.v3.p13 : T2D-GENES Multi-Ethnic Exome Sequencing Study: Framingham Heart Study

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v32.p13

Appendix - key omics details

- Includes
  - list of FHS substudies (phs numbers)
  - Data descriptions inc. FHS cohort/exam context
  - Data location – FHS substudy number & data identifier(s)
  - Tallies – overall and by cohort overall
  - Tallies – cohort exam breakout

Appendix - key omics details

RESOURCE TYPE	RESOURCE	DESCRIPTION	COHORT/EXAM
METABOLOMICS	TOPMED_METABOLO	4 platforms : C8-pos Lipid, C18-neg, HILIC-pos, Amide-neg	Gen2 Ex8-9, NOS & Gen3 Ex2
METABOLOMICS	l_mtbgcms	Metabolomics - Risk Factor Study: Gas Chromatography/Mass Spec	Gen2 Ex8, Gen3 Ex1
METABOLOMICS	l_mtblcmhi	Central Metabolomics - HILIC - Installment 1,2	Gen2 Ex5
METABOLOMICS	l_mtbli	Metabolomics (Hilic - Installment 1,2,3)	Gen2 Ex5
METABOLOMICS	l_mtbllipi	Metabolomics - Lipid Platform - Installment 1,2,3	Gen2 Ex5

- Data descriptions inc. FHS co

RESOURCE TYPE	RESOURCE DESIGNATION	PHS	identifier
TRANSCRIPTOMICS	RTPCR	phs000342	l_rnatrans_ex08_1b_1164s
TRANSCRIPTOMICS	SABRE_EXPRESS	phs000363	phe000002
	CHARGES_WES_FREEZE5	phs000651	phg001094

data identifier(s)

RESOURCE TYPE	RESOURCE DESIGNATION	TOTAL ACROSS FHS	Cohort					
			Gen 1 Original	Gen 2 Offspring	Gen 2 New Offspring Spouse	Gen 3 Generation 3	Omni I	Omni 2
METABOLOMICS	TOPMED_METABOLO	3025	0	1741	61	1223	0	0
METABOLOMICS	l_mtbgcms	650	0	316	0	334	0	0
METABOLOMICS	l_mtblcmhi	2067	0	2067	0	0	0	0

RESOURCE TYPE	RESOURCE DESIGNATION	COHORT	OFFSPRING					NOS			GEN3			OMNI1			OMNI2	
		MIX*	EXAM5	EXAM6	EXAM7	EXAM8	EXAM9	EXAM1	EXAM2	EXAM3	EXAM1	EXAM2	EXAM3	EXAM1	EXAM3	EXAM4	EXAM1	EXAM3
METABOLOMICS	TOPMED_METABOLO					386	1355		61			1223						
METABOLOMICS	l_mtbgcms					316					334							
METABOLOMICS	l_mtblcmhi		2067															
METABOLOMICS	l_mtbli		2526															
METABOLOMICS	l_mtbllipi		2069															
METABOLOMICS	l_mtbnegam										998							
METABOLOMICS	l_mtbtarg										996							
METABOLOMICS	l_umtbl			386														

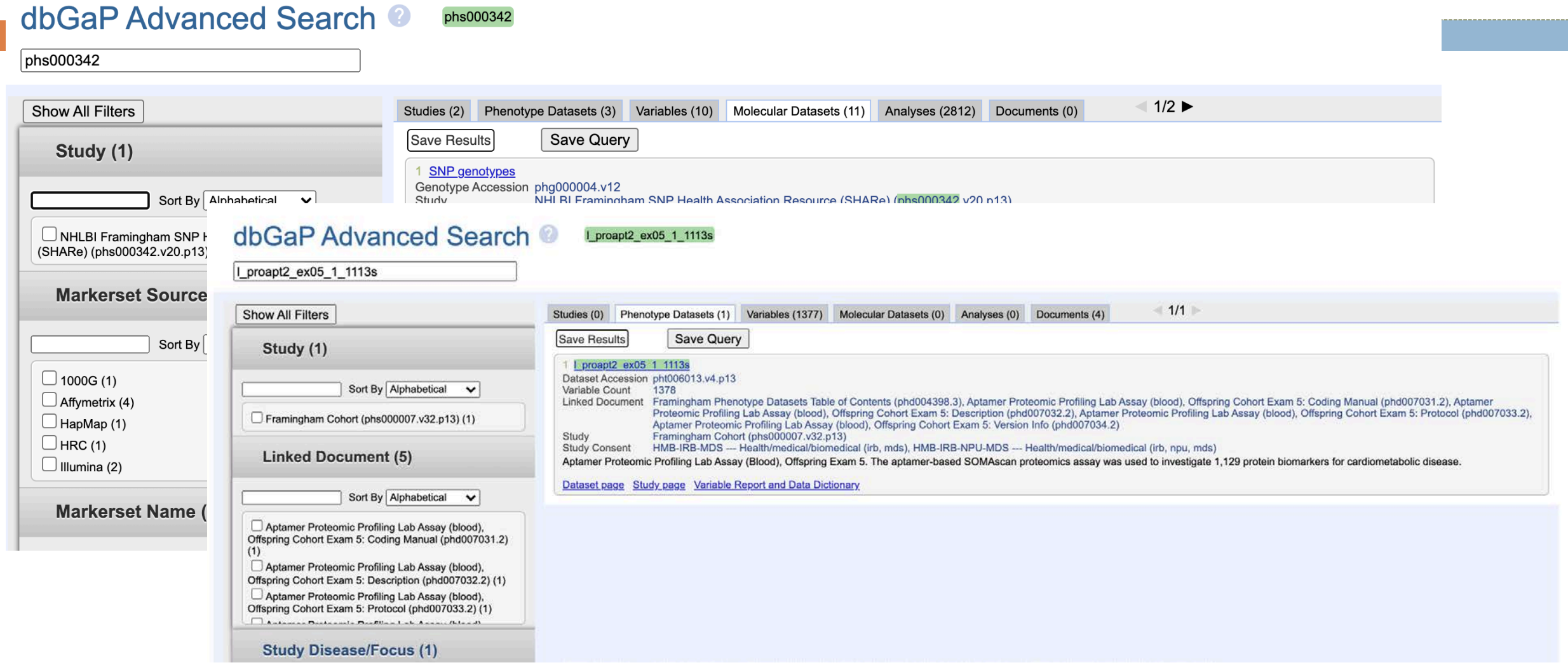
Searching dbGaP

https://www.ncbi.nlm.nih.gov/gap/advanced_search/

- Enter the phs number of study under question
- Has tabs reflecting different sources available
- Can type in the identifier for direct search

Searching dbGaP

https://www.ncbi.nlm.nih.gov/gap/advanced_search/



FHS SUBSTUDY SUMMARIES

Genomic File types

Data available includes the following (vary according to source &/or type available at time of deposit):

- Genotypes (matrix)- Set of text and binary files with SNP genotypes, pedigree, and gender information in PLINK format
- Genotypes (individual) set of Text files with genotypes, call scores, and allele intensities in a standard NCBI format; one file per sample (.ind files)
- Study sample information includes manifest of all genotype data files in the release
- SRA raw sequencing data (BAMs, fastQ)
- Binary raw files containing probe intensity data; one file per sample (.CEL files)
- Marker & sample quality metrics & reports including call rate, allele frequency, and others

CARe phs000282

- This substudy phs000282 Framingham CARe contains 50K genotypes from ~2,100 candidate genes across a range of cardiovascular, metabolic, and inflammatory syndromes, produced as part of NHLBI's Candidate Gene Association Resource (CARe) project.
- N = 7556

  (GaP accession : phg000076)

Allelic Spectrum Project phs000307

- **Allelic Spectrum Project :**deep coverage targeted re-sequencing and variant identification for 216 genes in the Framingham Heart Study (FHS) sample
- SNP/CNV derived from exome sequencing data
- Inc. genotype-calls-vcf, genotype-qc, marker-info, sample-info

(GaP accession: phg000175)

SHARe **phs000342**

- **phs000342 Framingham SHARe includes**:

1. Legacy set – bundle of historic FHS genotype files
2. 100K - Affymetrix 100K GWAS array
3. 550K Affymetrix genotypes, produced as part of NHLBI's SNP Health Association Resource (SHARe) project.
4. Imputed genotypes (derived from 550K data set)
5. OMNI5.0 - 4.3 million genotypes produced under Illumina's Fast Track Genotyping Service
6. AXIOM – Illumina population focused GWAS genotype array
7. Exome Chip

phs000342 Legacy genotypes

- microsatellite markers
- SNPs
- apoE genotype*
- in different limited subsets of Gen1 & Gen2 participants

- (**GaP accession: phg000005**)

- ***PLEASE USE **fhs_master_apoe_source** file in next release as includes all available APOE across all FHS cohorts

phs000342 100K

- **100K Genome-Wide Association Study and Results Browser**
- subset of 1345 adult participants (original cohort and offspring) of the largest 310 pedigrees in the FHS, many biologically related, was genotyped with the 100K Affymetrix GeneChip.

- Inc genotype-calls-matrixfmt, marker-info, raw-data-cel, sample-info

(**GaP accession: phg000005**)

phs000342 SHARE genotypes

- **Framingham Heart Study SHARe Genome-Wide Association Study.**
- In 2007, the FHS entered a new phase with the conduct of genotyping for the FHS SHARe project, for which genotyping was conducted using approximately 550,000 SNPs
- Includes:

  Affymetrix 500K mapping array (**GaP accession: phg000006)**

  (from 250K Nsp Array & 250K Sty array)

  Affymetrix 50K supplemental array (**GaP accession: phg000004**)

- In over 9200 participants from the three generations (including over 1500 families).
- Inc. genotype-calls-indfmt, genotype-calls-matrixfmt, genotype-qc, marker-info, sample-info

phs000342 Imputed SHARE genotypes

1. Imputed SNP genotypes using MACH software and HapMap phase2 reference panel with Share 550K genotypes

   inc. genotype-imputed-data, genotype-imputed-metrics, sample-info

   (GaP accession: phg000043)

2. Imputed SNP genotypes using MACH / minimac software and 1000G reference panel with Share 550K genotypes

   inc. genotype-imputed-data, genotype-imputed-metrics, genotype-qc, sample-info

   (GaP accession: phg000679)

3. Phased Imputed SNP genotypes using minimac and the Haplotype Reference Consortium reference panel with Share 550K genotypes

   inc. genotype-calls-vcf, genotype-qc, marker-info, sample-info

   (GaP accession: phg000835)

IMPUTATION LIMITED TO SUBSET OF FHS PARTICIPANTS WITH HIGH QUALITY SHARe GENOTYPING

phs000342 Omni5M genotypes

- **Illumina HumanOmni5M-4v1 Array.** In 2011, genotyping of approximately 4.3 million SNPs was conducted in subset of 2500 Framingham ***Offspring*** Cohort participants using the Illumina HumanOmni5M-4v1 array designed to target variation down to 1% minor allele frequency.
- Data available: genotype-BAF, genotype-calls-indfmt, genotype-calls-matrixfmt, marker-info, sample-info

  Contains data for B-allele freq analysis (BAFs, thetas,LogRRs)

- **GaP accession: phg000256**

phs000342 AXIOM genotypes

- **Affymetrix Axiom Genome-Wide BioBank array** configuration (628,088 SNP markers) with no customization
- Conducted in all consented OMNI1 and OMNI2 cohort participants with DNA available
- N= 845
- Contains data genotype-calls-matrixfmt, marker-info, raw-data-cel, sample-info
- **GaP accession: phg000617**

phs000342 Exome chip

- Exome Chip design:

  http://genome.sph.umich.edu/wiki/Exome_Chip_Design

- CHARGE cohorts, including Framingham, were jointly called in Houston

  http://depts.washington.edu/chargeco/wiki/ExomeChip

- Best Practices and Joint Calling of the HumanExome BeadChip:

  The CHARGE Consortium:

  https://doi.org/10.1371/journal.pone.0068095

- **N = 8153**

  (GaP accession: phg000618)

phs000401 HeartGO

- The NHLBI "Grand Opportunity" Exome Sequencing Project (GO-ESP), a signature project of the NHLBI Recovery Act investment, was designed to identify genetic variants in coding regions (exons) of the human genome (the "exome") that are associated with heart, lung and blood diseases.
- contains exome sequence data and harmonized phenotype variables, produced as part of NHLBI's GO-ESP project
- N= 464
- **(GaP accession: phg000682)**

CHARGES phs000651

- This substudy phs000651 Framingham CHARGE-S contains whole genome, whole exome, and targeted sequence data, produced as part of NHLBI's CHARGE-S project.
- The study has taken a two pronged approach to following-up GWAS:

1. First **regional capture targeted sequencing** was performed in genomic regions influencing 15 phenotypes to localize causal variants that are responsible for the GWAS signal ;
2. Second, Whole exome capture sequencing (WES) and **low-pass** whole genome sequencing (WGS) were completed for the cohort random sample and 7 phenotypes for which there were more than 3 GWAS signals in coding regions to detect novel rare and common variants.

CHARGES phs000651

- 1. **Regional capture targeted sequencing** was performed in genomic regions influencing 15 phenotypes to localize causal variants that are responsible for the GWAS signal ;
- Approximately 2 Mb of the genome was sequenced for the targeted loci.
- N = 1396

FIRST DEPOSIT:

(GaP accession: phg000380)

SECOND DEPOSIT: original plus + 300 selected for bone mineral density phenotype)

(GaP accession: phg000438)

CHARGES phs000651

2. **Whole exome capture sequencing (WES) and *low-pass* whole genome sequencing (WGS)** were completed for the cohort random sample and 7 phenotypes for which there were more than 3 GWAS signals in coding regions to detect novel rare and common variants.

There have been 5 releases of data (3 WES, 2 WGS)

WES Freeze3 **(GaP accession: phg000526); N = 850**

WES Freeze4 **(GaP accession: phg000689); N = 1271**

WES Freeze5 **(GaP accession: phg001094); N = 1702**

WGS Freeze1 **(GaP accession: phg000527); N = 320**

WGS Freeze3 **(GaP accession: phg000684); N = 855**

DNA Methylation phs000724

- Illumina Infinium® HumanMethylation450 BeadChip
- DNA methylation ***raw*** data in Gen2 ex8
  - **N= 2631**
  - inc marker-info, raw-data-idat, sample-info

    (GaP accession: phg000492)
- ***Processed*** *CpG data* from Gen2 ex8 plus Gen3 ex2
  - **N= 4151**
  - inc marker-info, methylation-data-matrixfmt, sample-info but no raw data

    (GaP accession: phg001091)

SABRe CVD phs000363 :

- Project 1: iTRAQ Px data set 135 case/control pairs iTRAQ is used in proteomics to study quantitative changes in the proteome
  - file name: l_protitraq_ex08_1_0736s
- Project 2: Immunoassays of 180 circulating protein biomarkers of atherosclerosis and metabolic syndrome in 7400 FHS participants.

  file name(s):

  l_mpimn01_2005_m_0692s, l_mpimn02_2005_m_0693s, l_mpimn03_2005_m_0694s, l_mpimn04_2005_m_0757s, l_mpimn05_2005_m_0758s, l_mpimn06_2005_m_0792s, l_mpimn07_2005_m_0802s, l_mpimn08_2005_m_0836s, l_mpimn09_2005_m_0850s, l_mpimn10_2005_m_0854s, l_mpimn11_2005_m_0855s, l_mpimn12_2005_m_0932s, l_mpimn13_2005_m_0931s, l_mpimn14_2005_m_0856s, l_mpimn15_2005_m_0974s, l_mpimn16_2005_m_0976s, l_mpimn17_2005_m_0977s

SABRe CVD phs000363 :

- Project 3: Gene expression profiling of WBC derived RNA to characterize the genomic signatures of atherosclerosis and metabolic syndrome in ~5600 FHS participants using Affymetrix Human Exon 1.0 ST expression microarray

  inc. expression-data-matrixfmt, marker-info, raw-data-cel, sample-info

  (GaP accession: phe000002)

- Project 4: MicroRNA profiling using custom set of oligonucleotide probes in WBC derived RNA in ~ 5700 FHS participants to characterize microRNA regulation of gene expression and the relations of microRNA to clinical traits and diseases. inc. expression-data-matrixfmt, marker-info, sample-info

  (GaP accession: phe000005)

phs000974 TOPMED - OVERALL

This substudy includes the following data:

1. ~ 30X WGS
2. Methylation Illumina 850K Infinium EPIC BeadChip
3. Metabolomics 4 platforms : C8-pos Lipid, C18-neg, HILIC-pos, Amide-neg
4. SOMAscan™ proteomic profiling
5. Three different RNASEQ batches

phs000974 TOPMED - WGS

- ~ 30X WGS was performed using DNA from blood, PCR-free library construction and Illumina HiSeq X technology.
- The Informatics Research Center conducts joint genotype calling across all samples available to produce genotype data "freezes."
- TOPMed data are being made available to the scientific community:
  - genotypes
  - read alignments via the Sequence Read Archive (SRA)
  - variant summary information via the Bravo variant server and dbSNP.

  N ~ 7170 across all FHS cohorts

(GaP accessions: phg001050, phg001282)

WGS Sample related info is described in:
https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000974/phs000974.v4.p3//manifest/Study_Report.phs000974.TOPMed_WGS_Framingham.v4.p3.MULTI.pdf

phs000974 TOPMED -Methylation

- Illumina 850K Infinium MethylationEPIC BeadChip
- Data processing steps at
- https://topmed.nhlbi.nih.gov/sites/default/files/TOPMed_Methylation_array_pipeline_COREyr3.pdf
- DNA methylation Gen2 ex9, NOS ex1-2, Gen3 ex2
  - **N= 1809**
  - BAMS, processed data
  - **(GaP accession: unassigned)**

phs000974 TOPMED - METABOLOMICS

- **(GaP accession: unassigned)**

N= 3025

(GaP accession: unassigned)

phs000974 TOPMED - METABOLOMICS

- (G

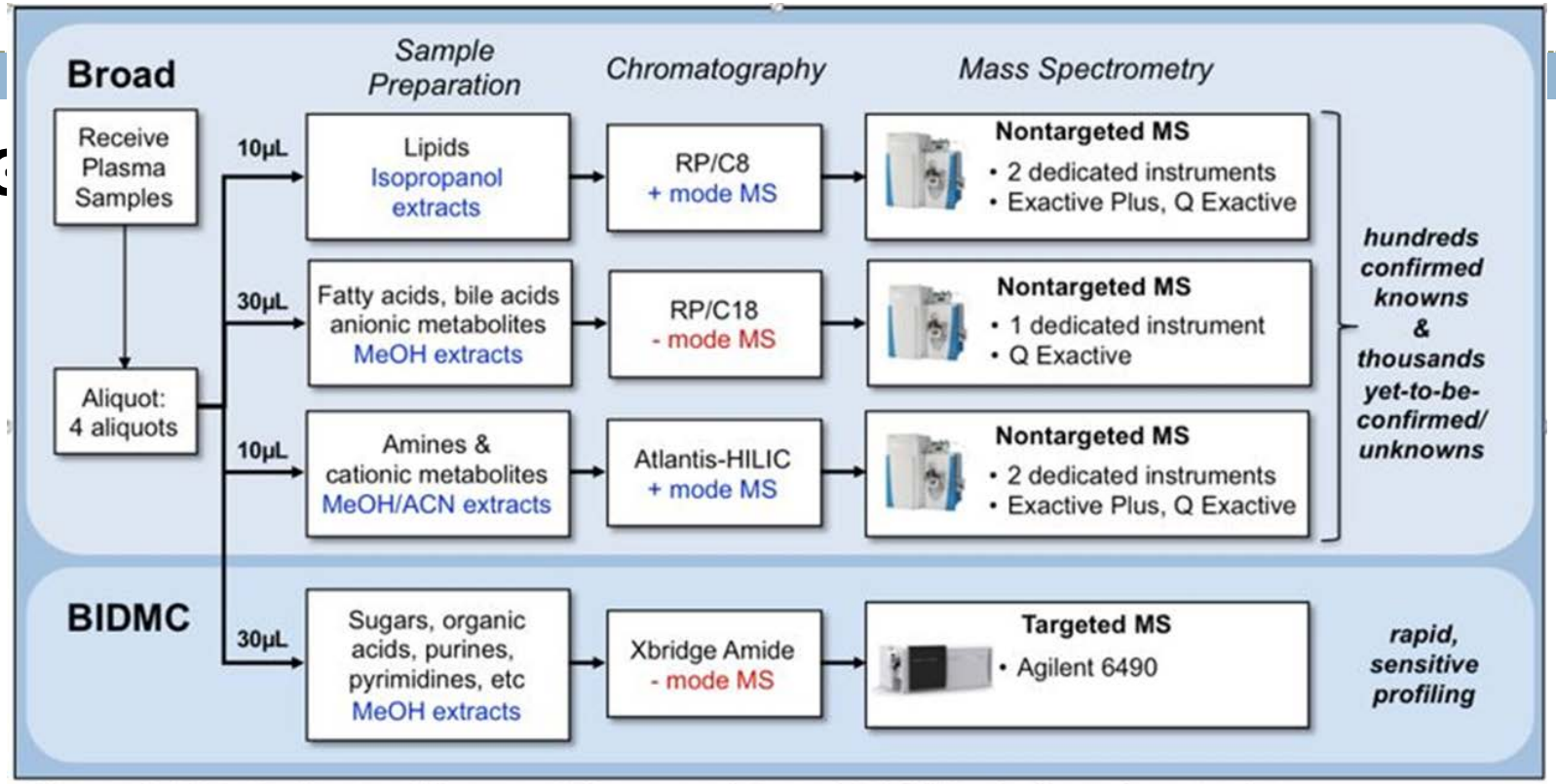


Figure 1. Metabolite profiling workflows and Broad-BIDMC platform integration

N= 3025

(GaP accession: unassigned)

phs000974 TOPMED - Proteomics

- The SOMAscan™ proteomic profiling platform is an aptamer-based technique that uses chemically modified single-stranded DNA aptamers to assay 1,129 proteins in an accurate, high throughput manner.
- In Gen3 Ex2 participants
  - **N= 900**
  - **(GaP accession: unassigned)**

phs000974 TOPMED - RNASEQ

3 Different sets

** NO OVERLAP**

Levy : 1505 Gen3 Ex2

FHS : 720 Gen2 Ex9

46 NOS Ex2

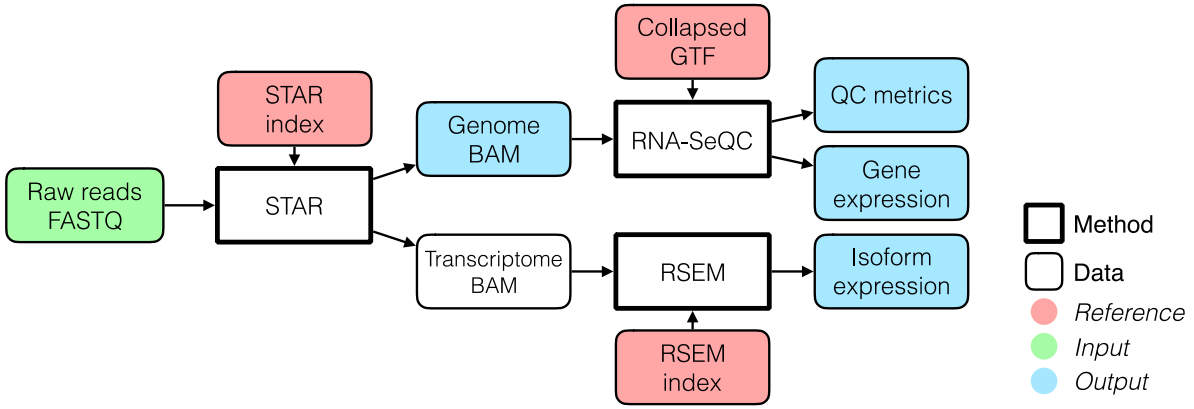
425 Gen3 Ex2

FHS2: 288 Gen2 Ex9

850 Gen3 Ex2

https://topmed.nhlbi.nih.gov/sites/default/files/TOPMed_RNAseq_pipeline_COREyr3.pdf

TOPMed RNA-seq pipeline V2



Alignment

ameters and wrapper scripts:
https://github.com/broadinstitute/gtex-pipeline

On the FireCloud platform:

namespace broadinstitute_gtex

STAR: Dobin et al., *Bioinformatics*, 2013
RSEM: Li et al., *Bioinformatics*, 2010
RNA-SeQC 2: Graubert et al., *in preparation*, 2019

UPCOMING FHS SUBSTUDIES

- phs002558 NHLBI Framingham Alzheimer's Disease Sequencing Project (ADSP)
- phs002559 NHLBI Framingham BRain Imaging, Cognition, Dementia and Next Generation GEenomics (**BRIDGET**)
- phs002660 Clinical and Genetic Correlates of the Gut Microbiome and Relation to Cardiometabolic Risk
- phs002661 Framingham Heart Study RNA sequencing in postmortem brain tissue

ADSP phs002558

- NHLBI Framingham Alzheimer's Disease Sequencing Project (ADSP)
- To identify novel genetic variation influencing AD risk and protection
- study design and sample selection were conducted to address issues of phenotypic heterogeneity and maximize statistical power.
- The study design includes 2 primary phases: a whole-genome sequencing (WGS) family-based study and a whole-exome sequencing (WES) case-control study.
- The WGS study targeted rarer variation through allelic segregation and linkage analyses in multiplex AD families.
- The WES case-control study targeted low-frequency coding variation in genes that contribute to AD risk or protection.
- The FHS ADSP WES samples are from the WES case-control study phase.

BRIDGET phs002559

- NHLBI Framingham BRain Imaging, Cognition, Dementia and Next Generation GEenomics **(BRIDGET)**
- novel methylation bisulfite sequencing technology, to help identify functional and disease relevant AD variants
- 198 FHS participants were selected to be at the two extremes of small vessel disease burden in the brain.:
  - ½ : low white matter hyperintensity (WMHI) burden and no MRI infarcts.
  - ½ : extensive WMHI burden (>1.5 SD above age & sex adjusted mean) and at least one MRI infarct.

inc: BAM, BisSNP vsf, methylation calls (Bismark)

RNA sequencing in post mortem brain tissue phs002661

- Framingham Heart Study RNA sequencing in post mortem brain tissue
- Large and small RNA sequencing of hypothalamus and nucleus accumbens from postmortem brain tissue samples from FHS participants
- Selected participant included:
    a. consistently obese (cases),
    b. consistently normal weight (controls) based on
    c. without respect to BMI and falling within neither case nor control definition

Microbiome phs002660

- Clinical and Genetic Correlates of the Gut Microbiome and Relation to Cardiometabolic Risk
- In subset of Generation 3/Omni 2 Cohorts at exam 3
- **Gut microbiome data from stool samples**
- **FASTq files** metagenome FASTQ data **from non-human 16S rRNA gene sequencing** from 624 FHS participants

- released in conjunction with PMID: **32544460**
- Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol