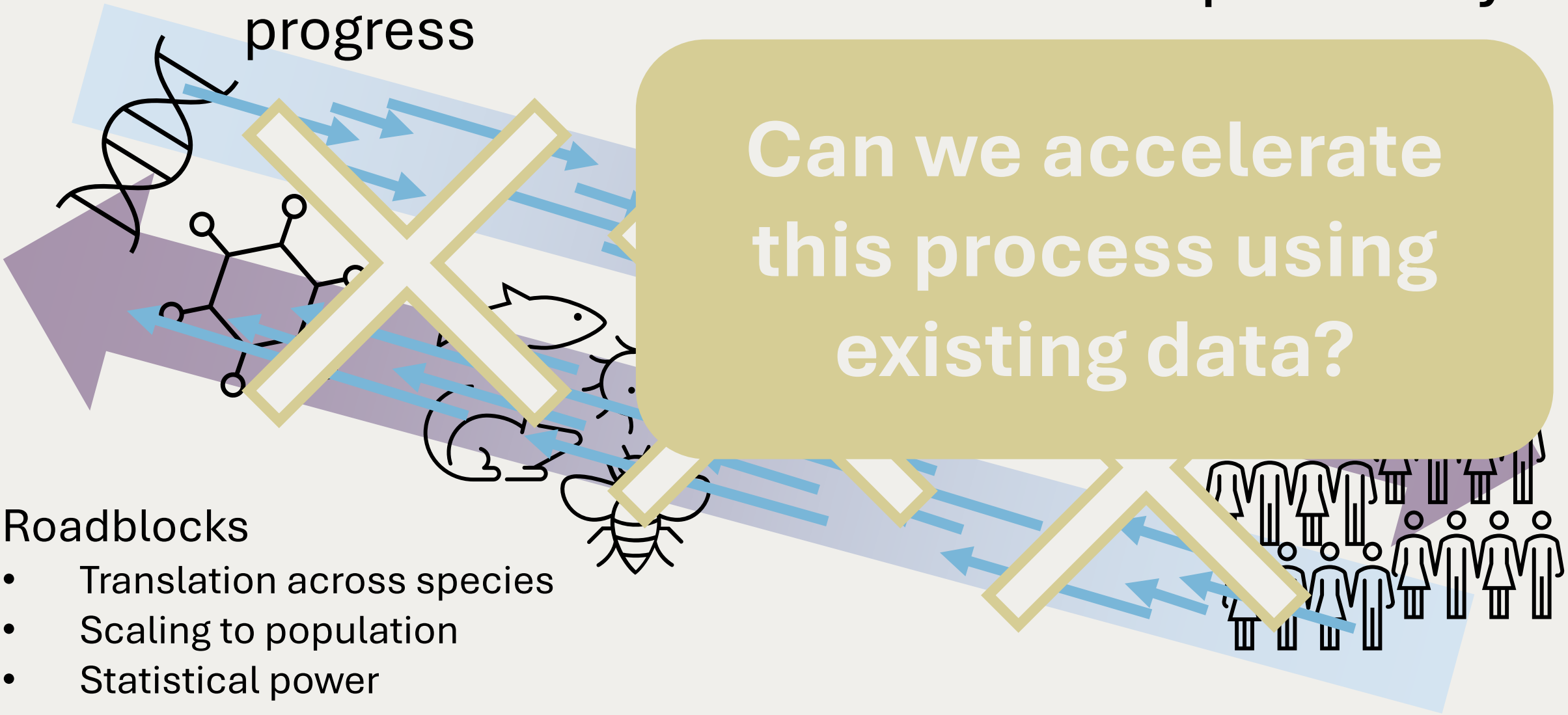
Use of Existing Data Collected by Other Disciplines

Eric T. Klopack
Department of Epidemiology and Biostatistics
School of Public Health Bloomington
Indiana University

RCCN Workshop: Multidisciplinary Research in Aging & Team Science
21 May 2026

Study-by-study progress

Translation and portability



Can we accelerate this process using existing data?

Roadblocks

- Translation across species
- Scaling to population
- Statistical power
- Replication

TraMA as an example

Motivation:
Develop a measure of biological aging using RNA

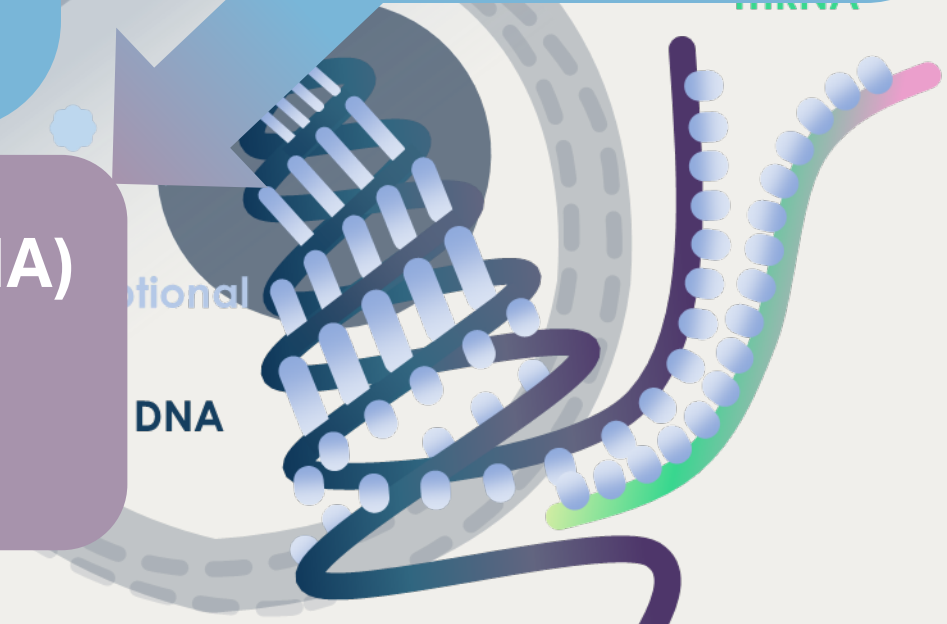
Health and Retirement Study 2016 Venous Blood Study

50/50 Training/Testing

Machine learning
predicting 4-year survival

Transcriptomic Measure of Aging (TraMA)

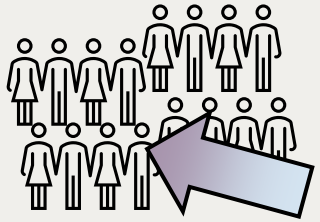
Index of 35 gene + chronological age

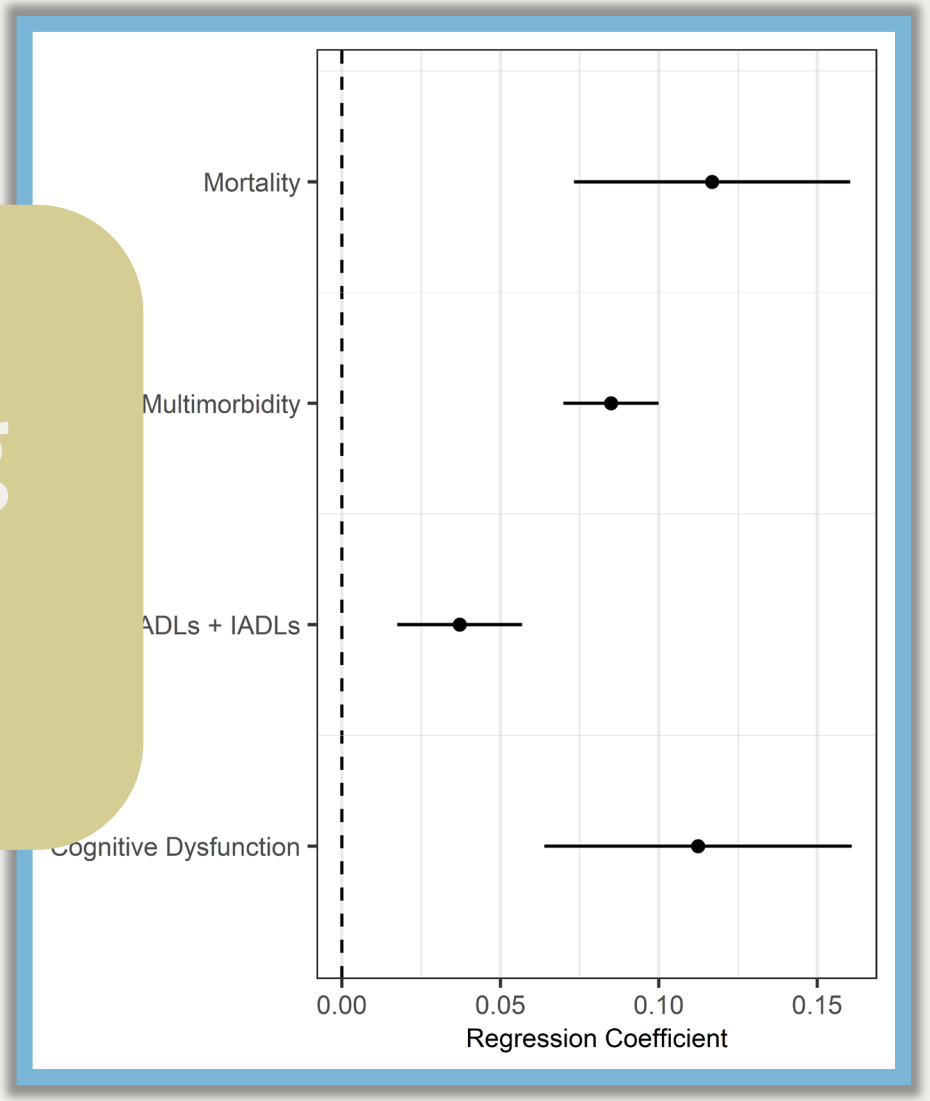


This measure works

We can use existing data for replication

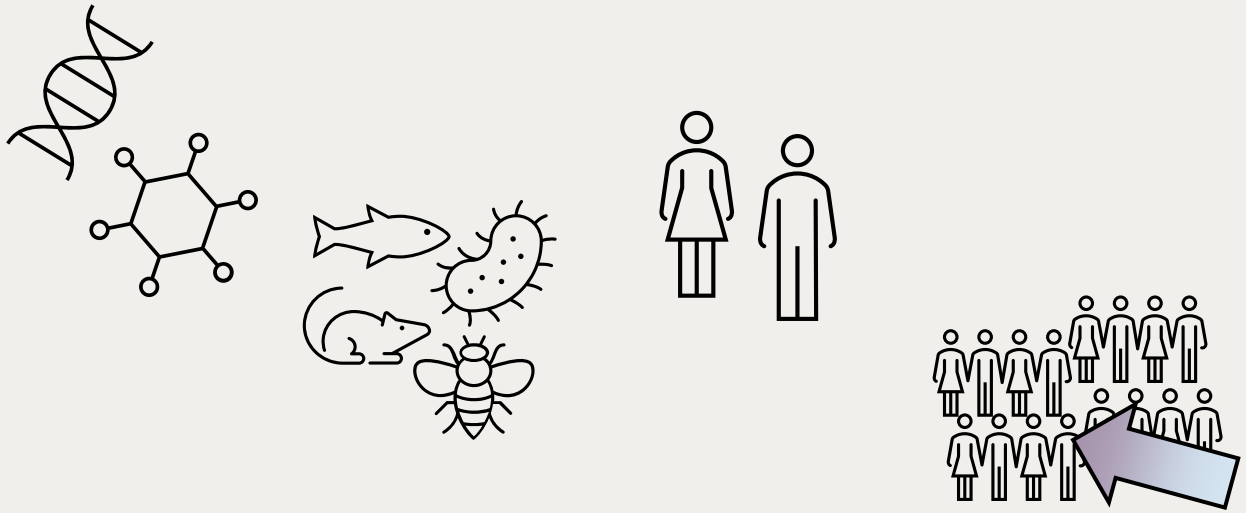


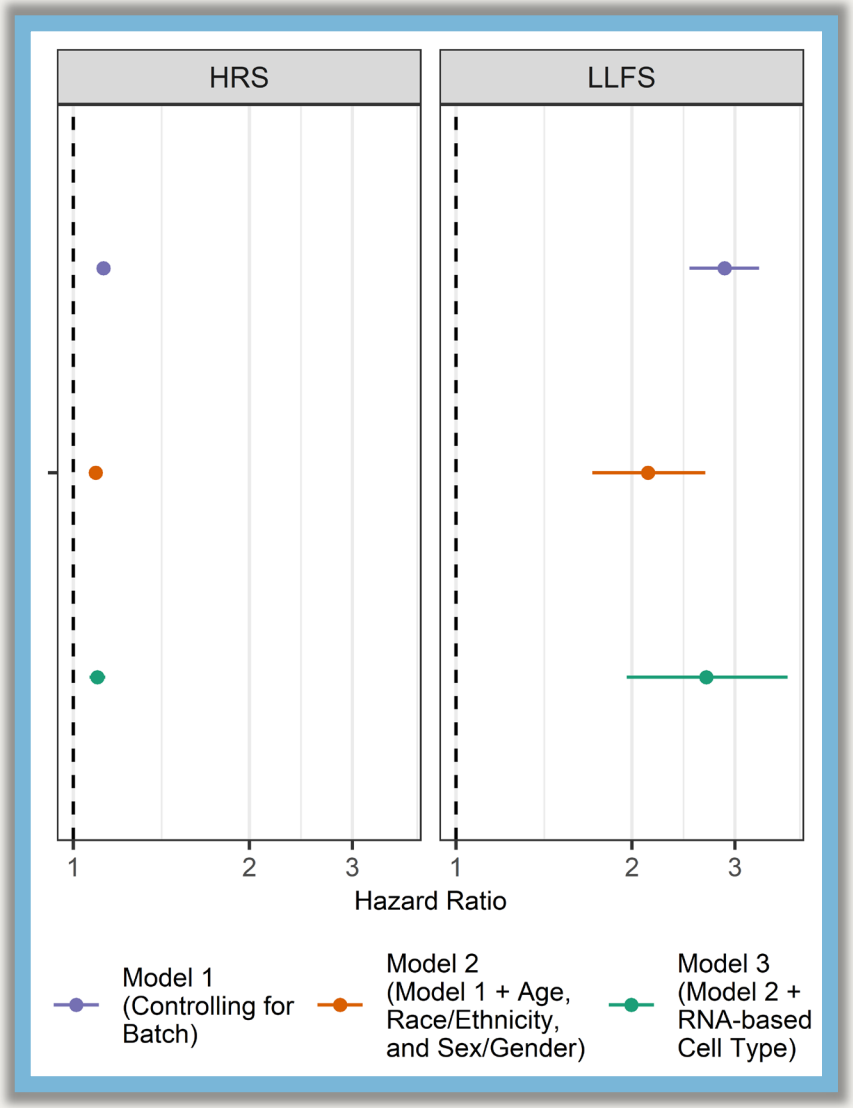




Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

This measure predicts mortality in the Long Life Families Study (LLFS)



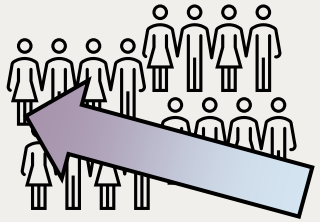


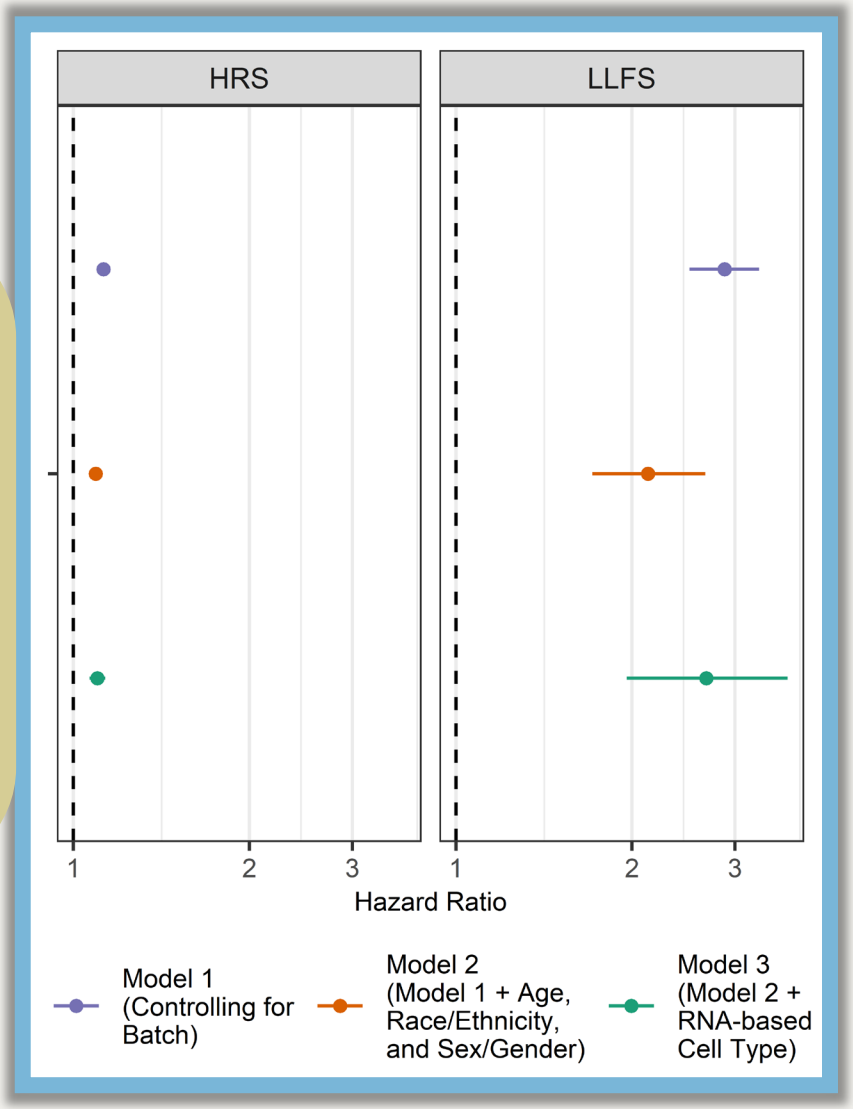
Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

This measure predicts

We can use existing data to show translation to clinical samples

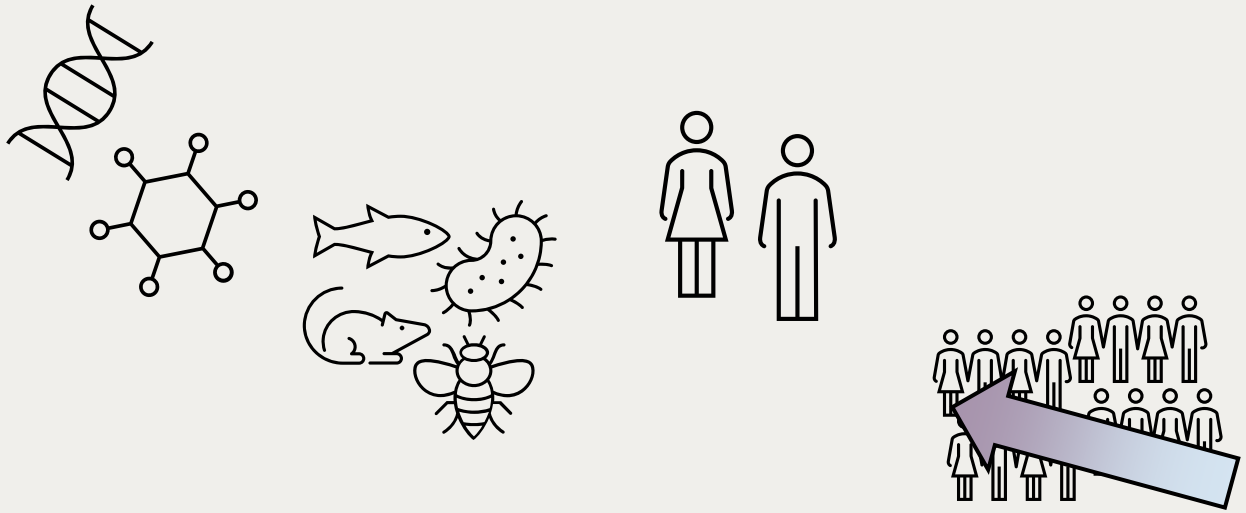


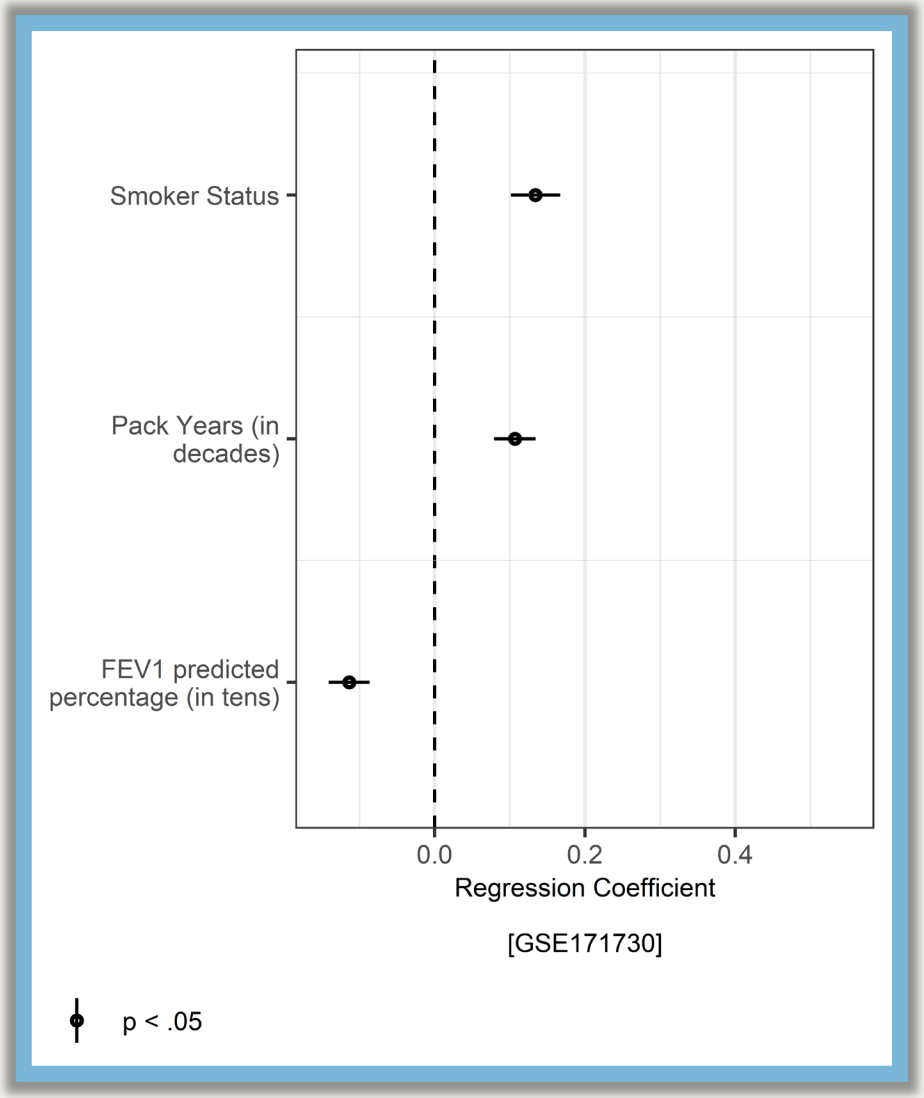




Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

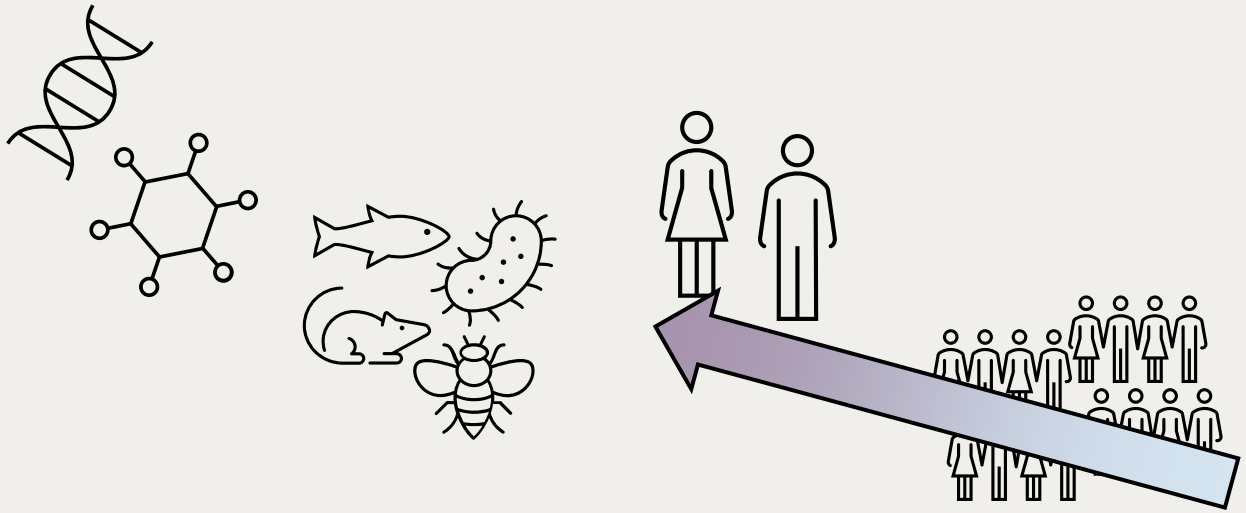
Using data from the **Gene Expression Omnibus (GEO):**
TraMA is associated with smoking, diabetes, IBD severity, sepsis, cancer, pediatric lupus

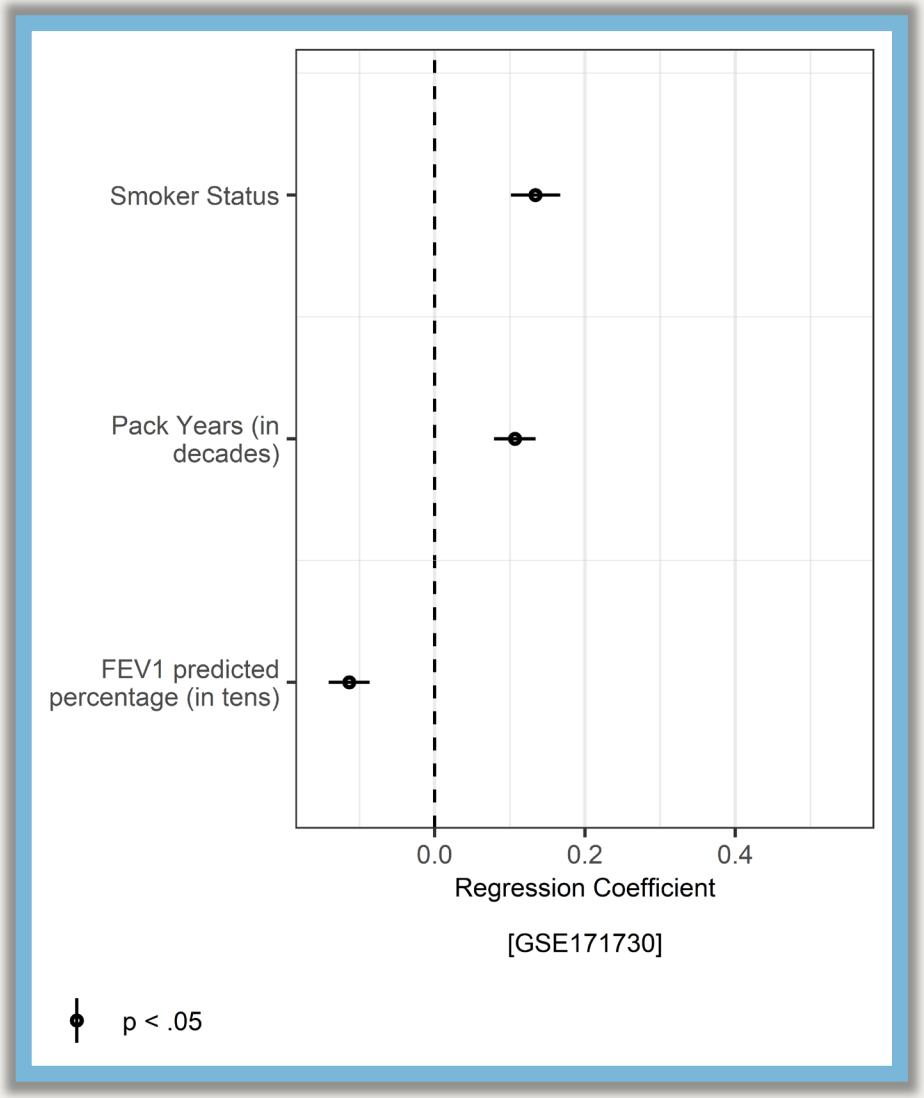




Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

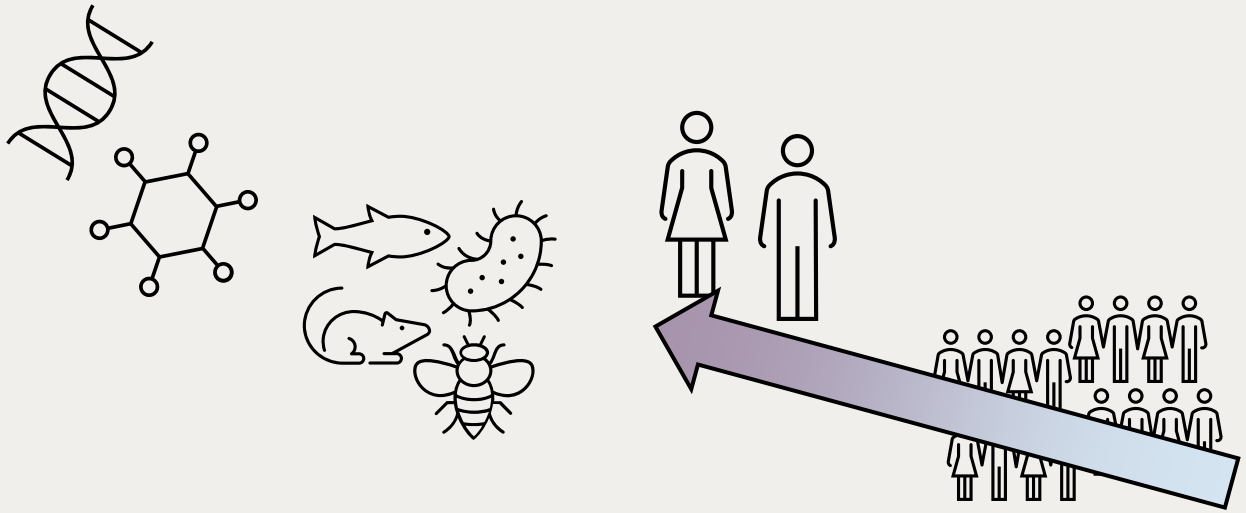
Using data from the **Gene Expression Omnibus (GEO):** TraMA is associated with smoking, diabetes, IBD severity, sepsis, cancer, pediatric lupus

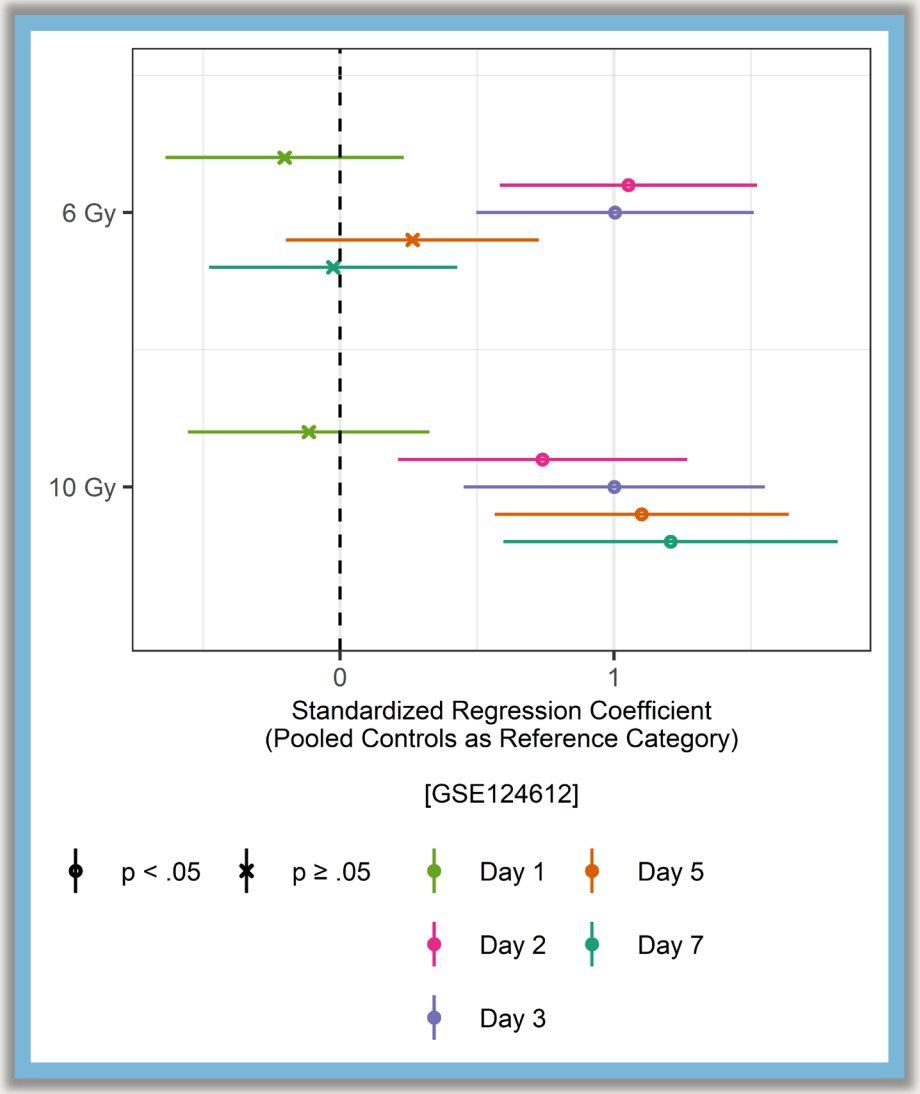




Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

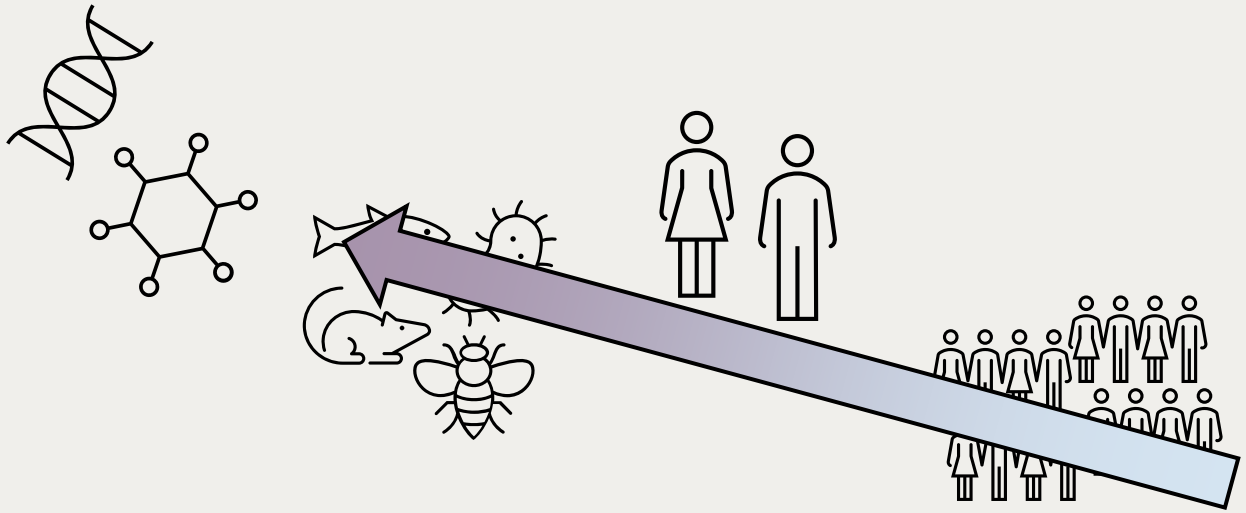
Using data from the **Gene Expression Omnibus (GEO):** TraMA is associated with radiation exposure in mice (Mus musculus)

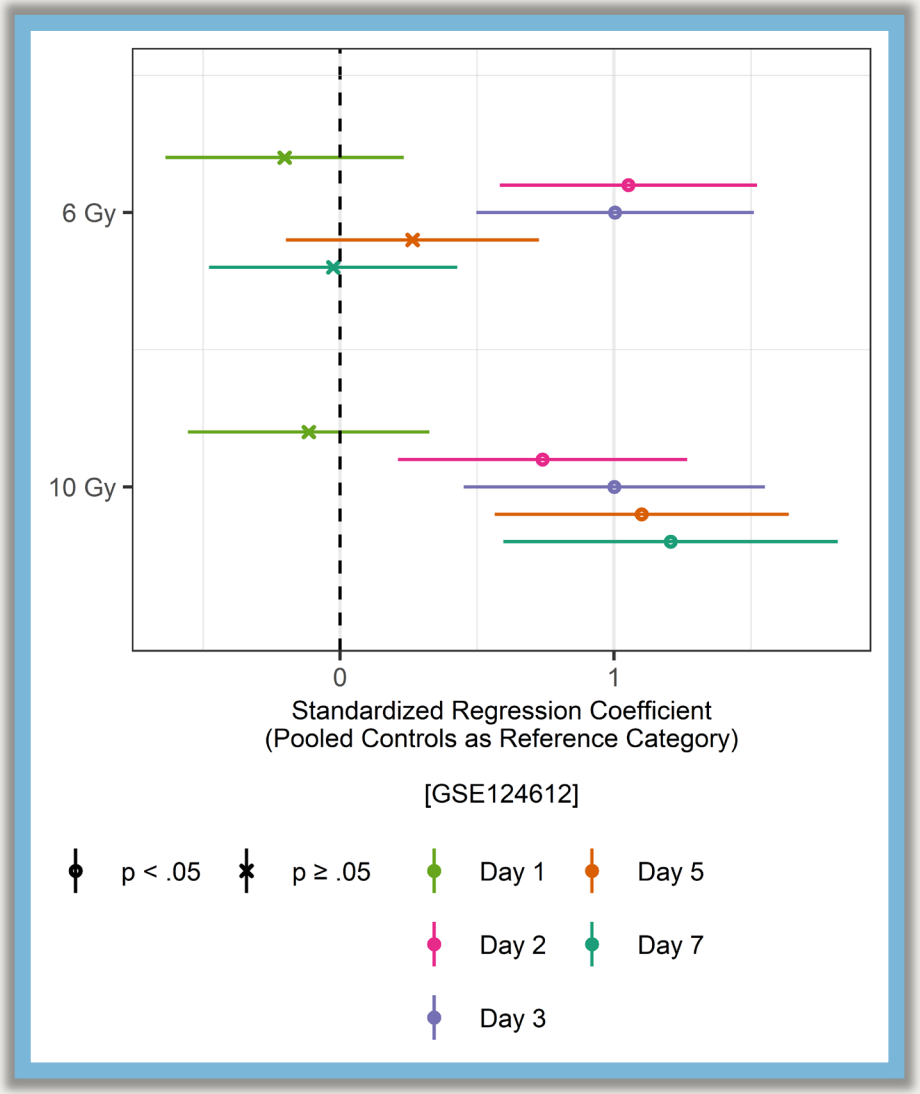




Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

Using data from the **Gene Expression Omnibus (GEO):** TraMA is associated with radiation exposure in mice (Mus musculus)



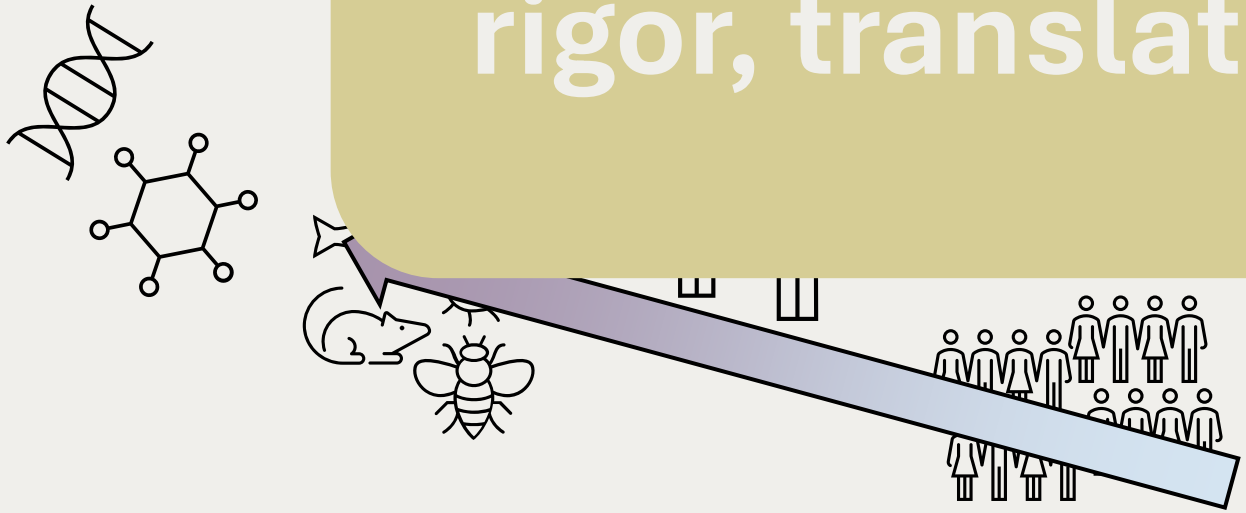


Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

Using data from the **Gene Expression Omnibus (GEO):**

TraMA is associated with radiation exposure

Expedite replication, rigor, translatability





Klopack ET, Seshadri G, Arpawong TE, Cole S, Thyagarajan B, Crimmins EM. Development of a novel transcriptomic measure of aging: Transcriptomic Mortality-risk Age (TraMA). *Aging*. 2025;17(6):1521-1543. doi:10.18632/aging.206272

Opportunities and possibilities

- Public data (Gene Expression Omnibus, National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site, The database of Genotypes and Phenotypes, population representative data)

HRS | HEALTH AND RETIREMENT STUDY

MIDUS
Midlife in the United States

Future of Families & Child Wellbeing Study
PRINCETON | COLUMBIA | NOTRE DAME

GEO
Gene Expression Omnibus

NIAGADS

Add Health
The National Longitudinal Study of Adolescent to Adult Health

dbGaP
GENOTYPES and PHENOTYPES

Opportunities and possibilities

- Public data
- Harmonized data

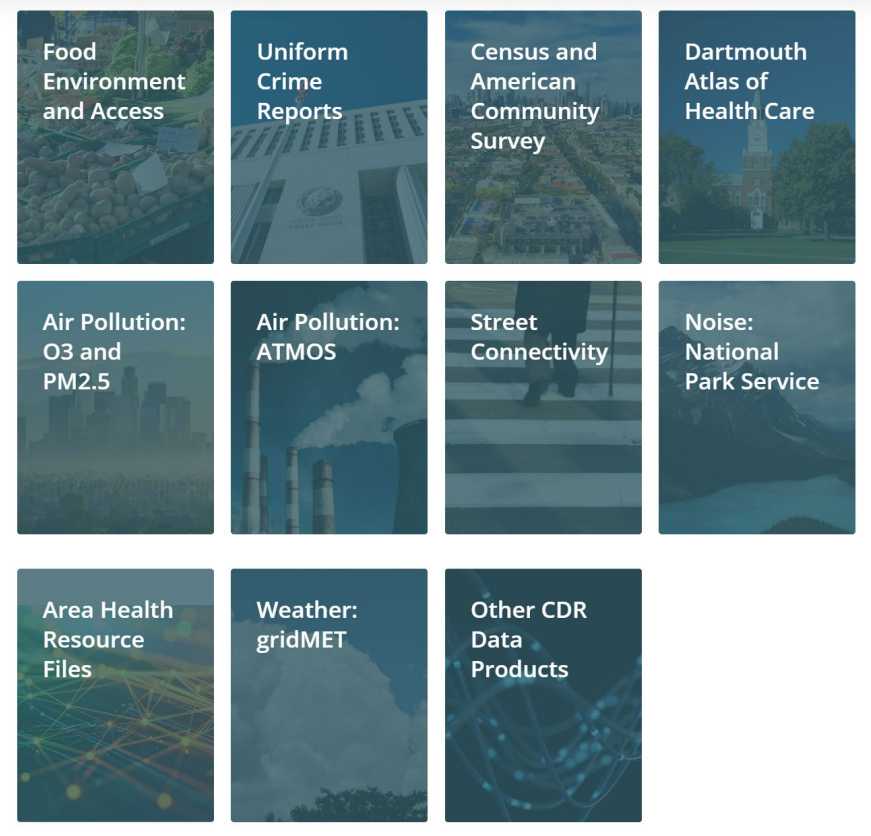
GATEWAY TO GLOBAL AGING DATA



Opportunities and possibilities

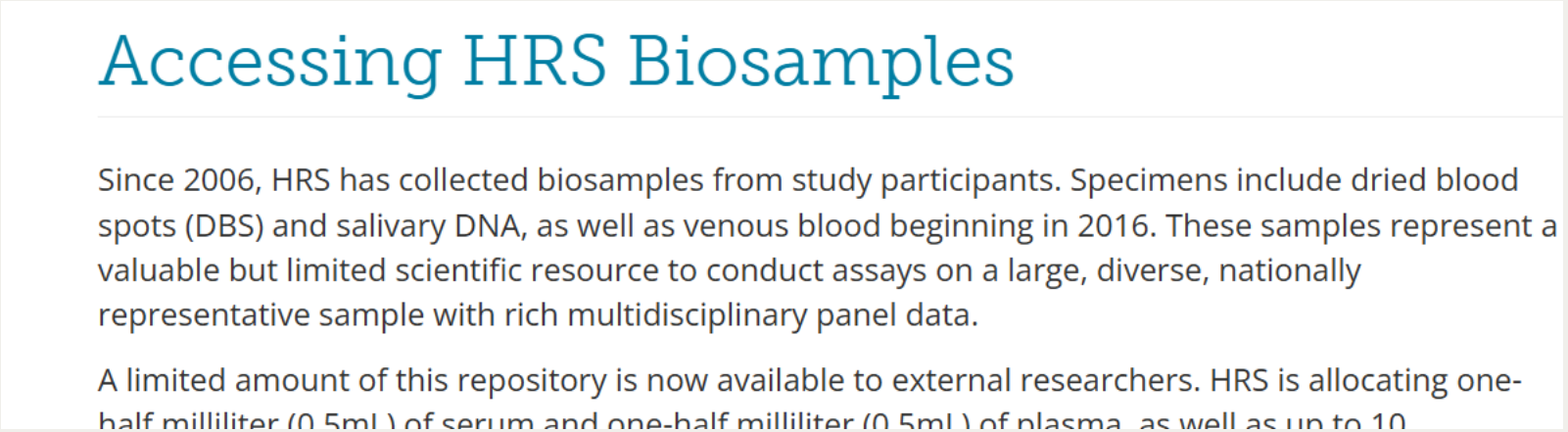
- Public data
- Harmonized data
- Administrative and other data linkages





Opportunities and possibilities

- Public data
- Harmonized data
- Administrative and other data linkages
- Biosamples



Accessing HRS Biosamples

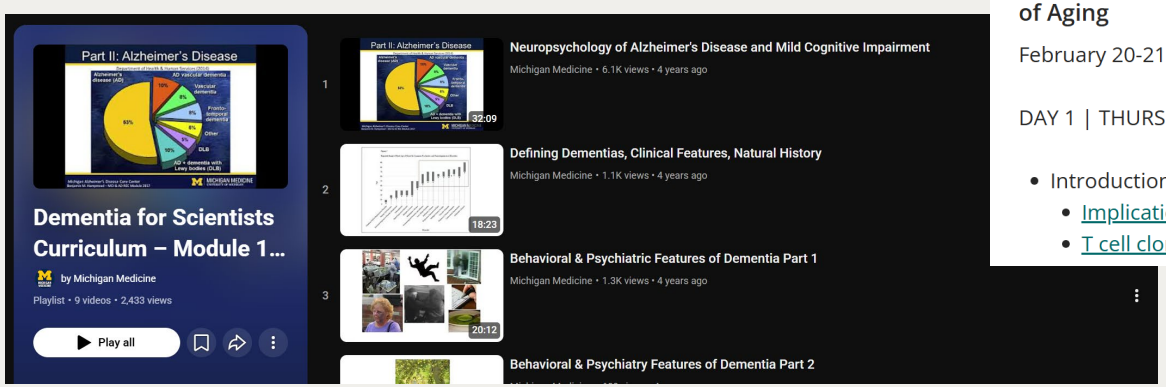
Since 2006, HRS has collected biosamples from study participants. Specimens include dried blood spots (DBS) and salivary DNA, as well as venous blood beginning in 2016. These samples represent a valuable but limited scientific resource to conduct assays on a large, diverse, nationally representative sample with rich multidisciplinary panel data.

A limited amount of this repository is now available to external researchers. HRS is allocating one-half milliliter (0.5mL) of serum and one-half milliliter (0.5mL) of plasma, as well as up to 10

Opportunities and possibilities

- Public data
- Harmonized data
- Administrative and other data linkages
- Biosamples
- Training and **experience**

Genomics for Social Scientists
NATIONAL INSTITUTE ON AGING

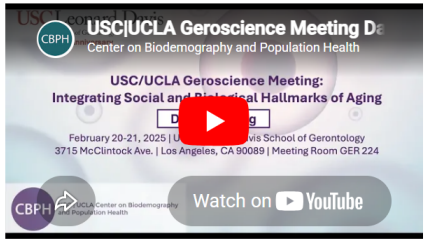


USC|UCLA Geroscience Meeting: Integrating Social and Biological Hallmarks of Aging

February 20-21, 2025

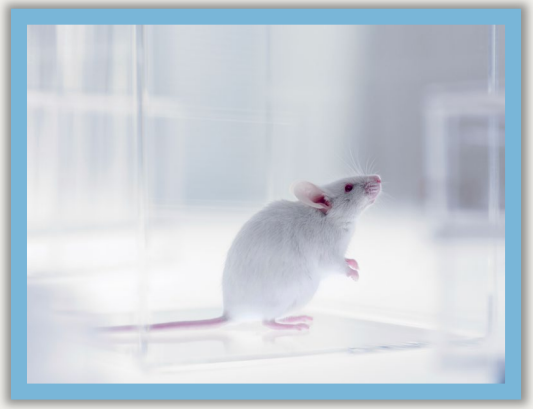
DAY 1 | THURSDAY, FEBRUARY 20, 2025

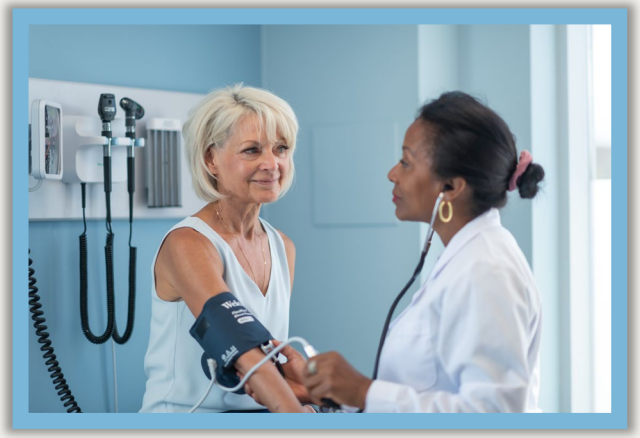
- Introduction (Eileen Crimmins)Novel Biomarkers of Aging
  - Implications of GLP-1s for population health and aging (Jennifer Dowd)
  - T cell clonality health outcomes in the Health and Retirement Study (Bharat

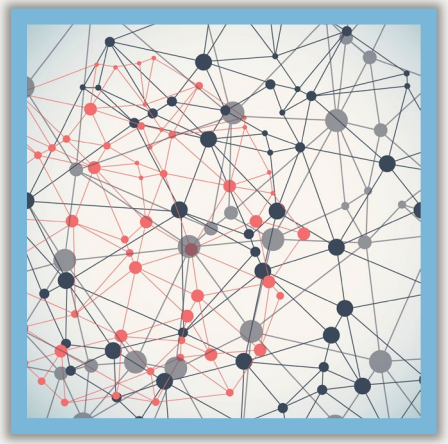


Potential Issues

- Multidisciplinary teams and training needed
- Communicating across discipline
- Types of data available (e.g., omics, harmonized psychosocial)
- Balancing open data and participant privacy
- Harmonization
- Big data complexity
- Time and effort









Public Data Speeds Science

- Data collected by many field for many purposes
- Rapidly produce evidence for translation and portability
- Improve rigor and reproducibility in a single paper
- Novel linkages and datasets
- **NIA + NIH success story**

THANK YOU!

COLLABORATORS

University of Southern California
Eileen M. Crimmins
Jennifer Ailshire
Jung Ki Kim

UCLA
Teresa Seeman
Steve Cole
Judith E. Carroll

University of Michigan
Jessica D. Faul
Colter Mitchel

University of Minnesota
Bharat Thyagarajan

University of Texas
Mateo Farina

Trinity College Dublin
Rose Anne Kenny
Cathal McCrory
Belinda Hernandez

Queens University Belfast
Amy Jayne McKnight
Claire Potter
Claire Hill

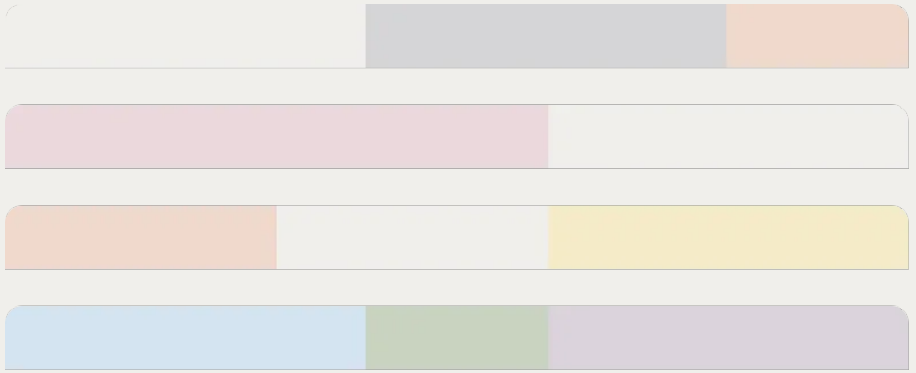
Max Planck Institute for Human Development
Qiao Wu

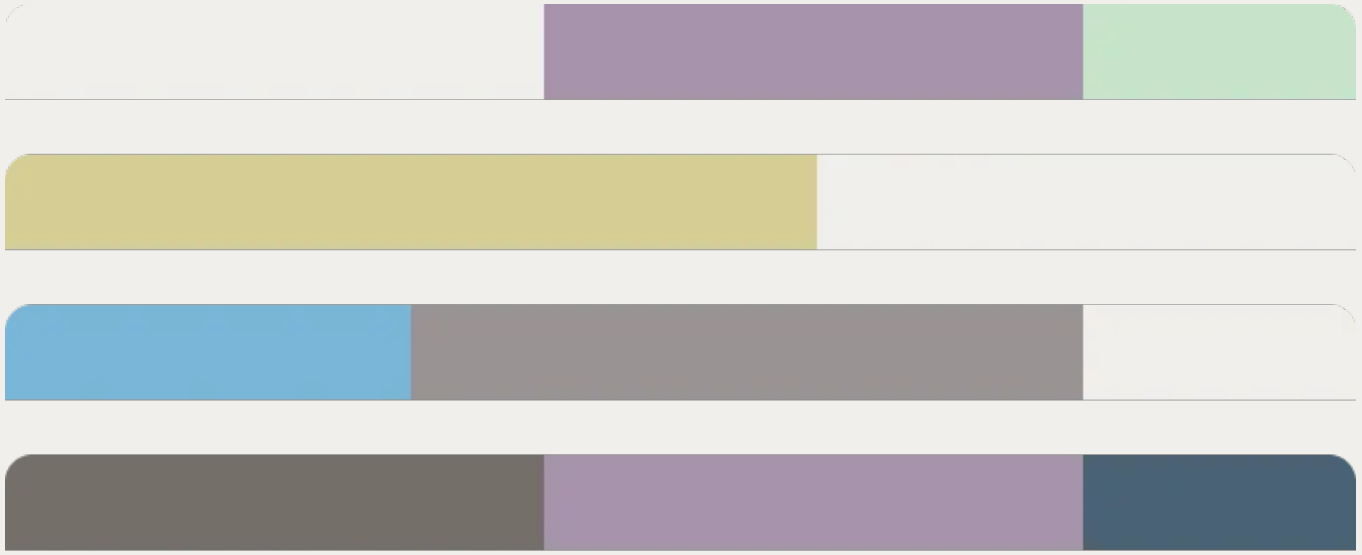
FUNDING

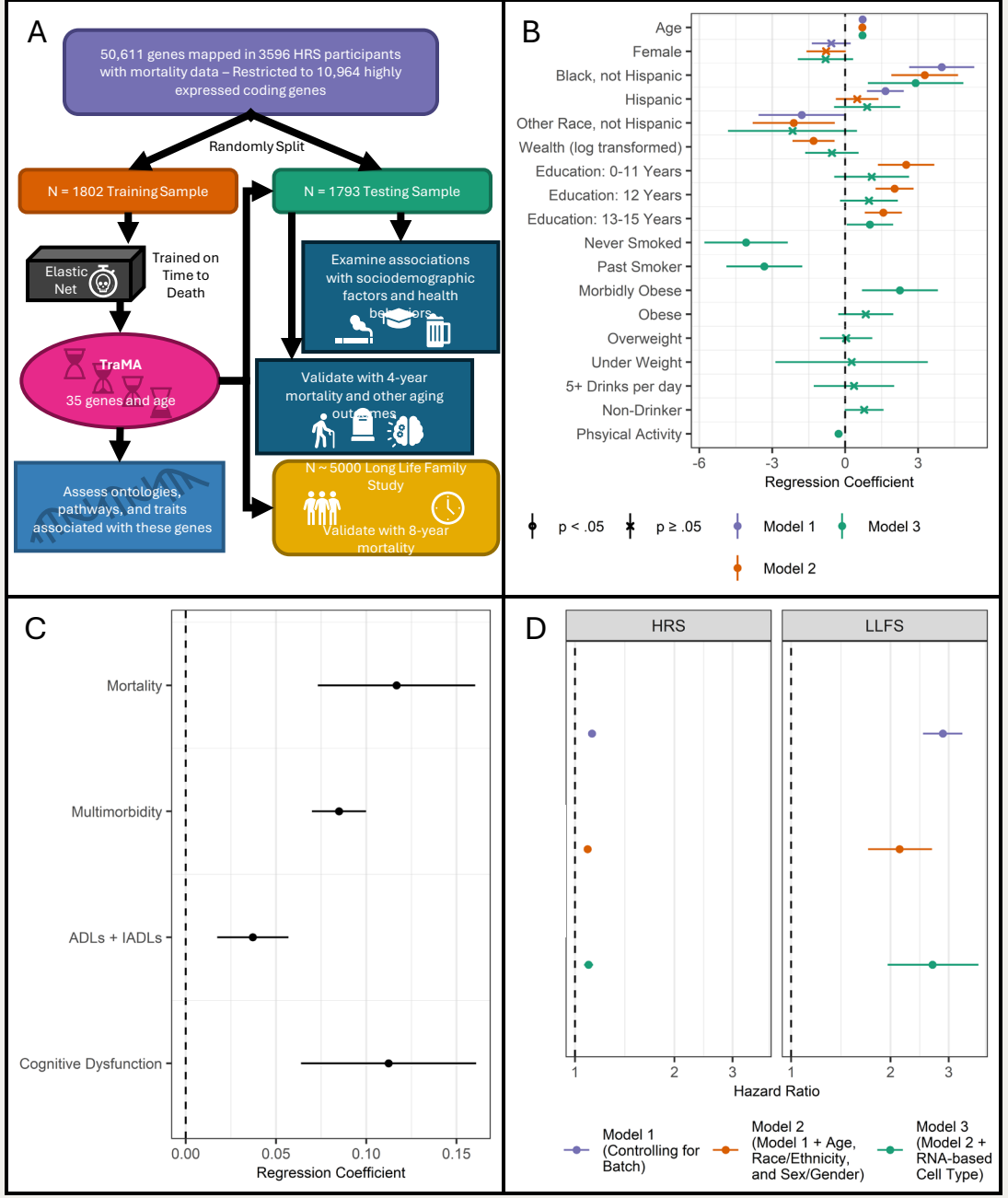
Support was provided by grants from the National Institute on Aging (T32AG000037-43, P30AG017265, and R01AG060110)

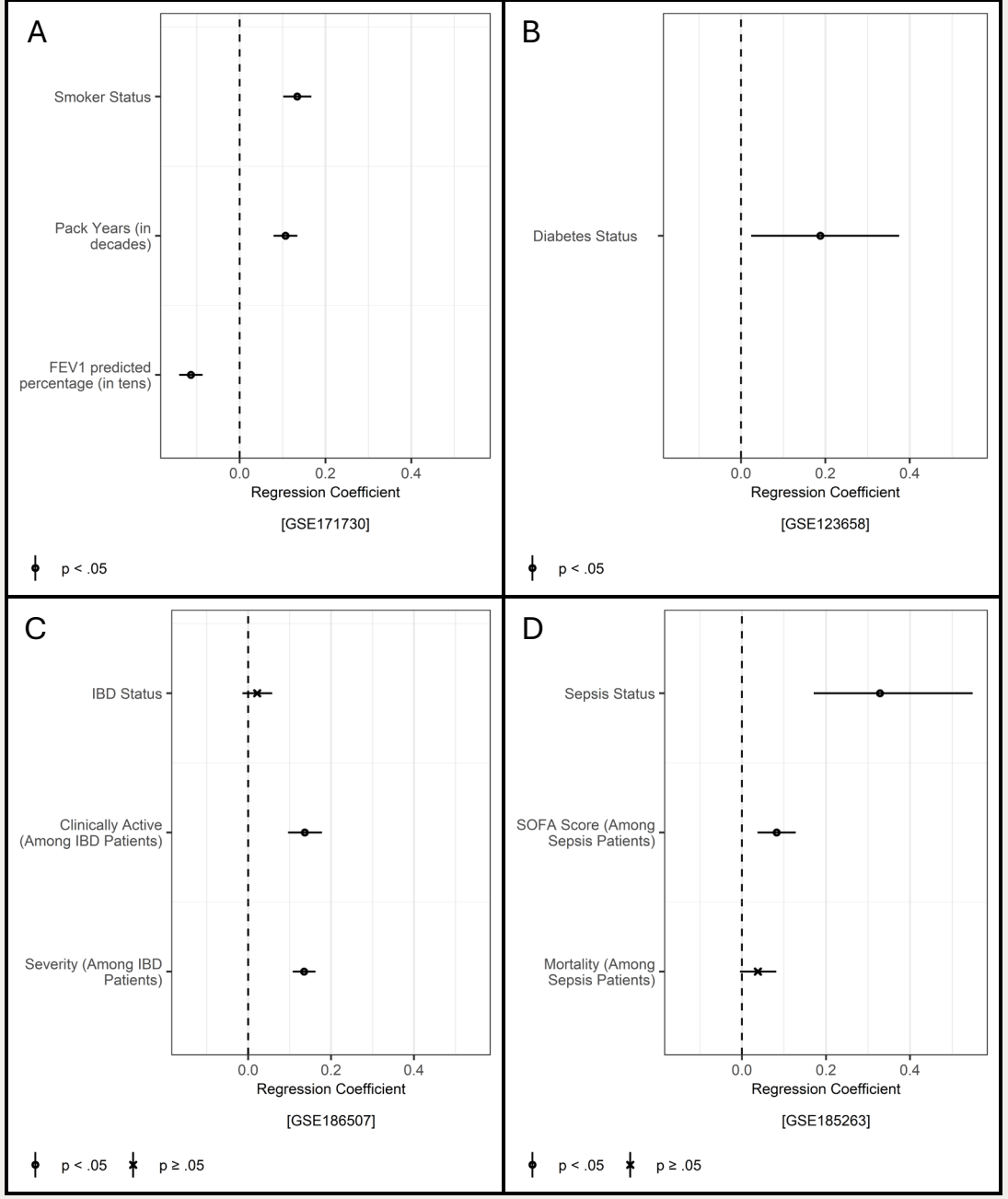
The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

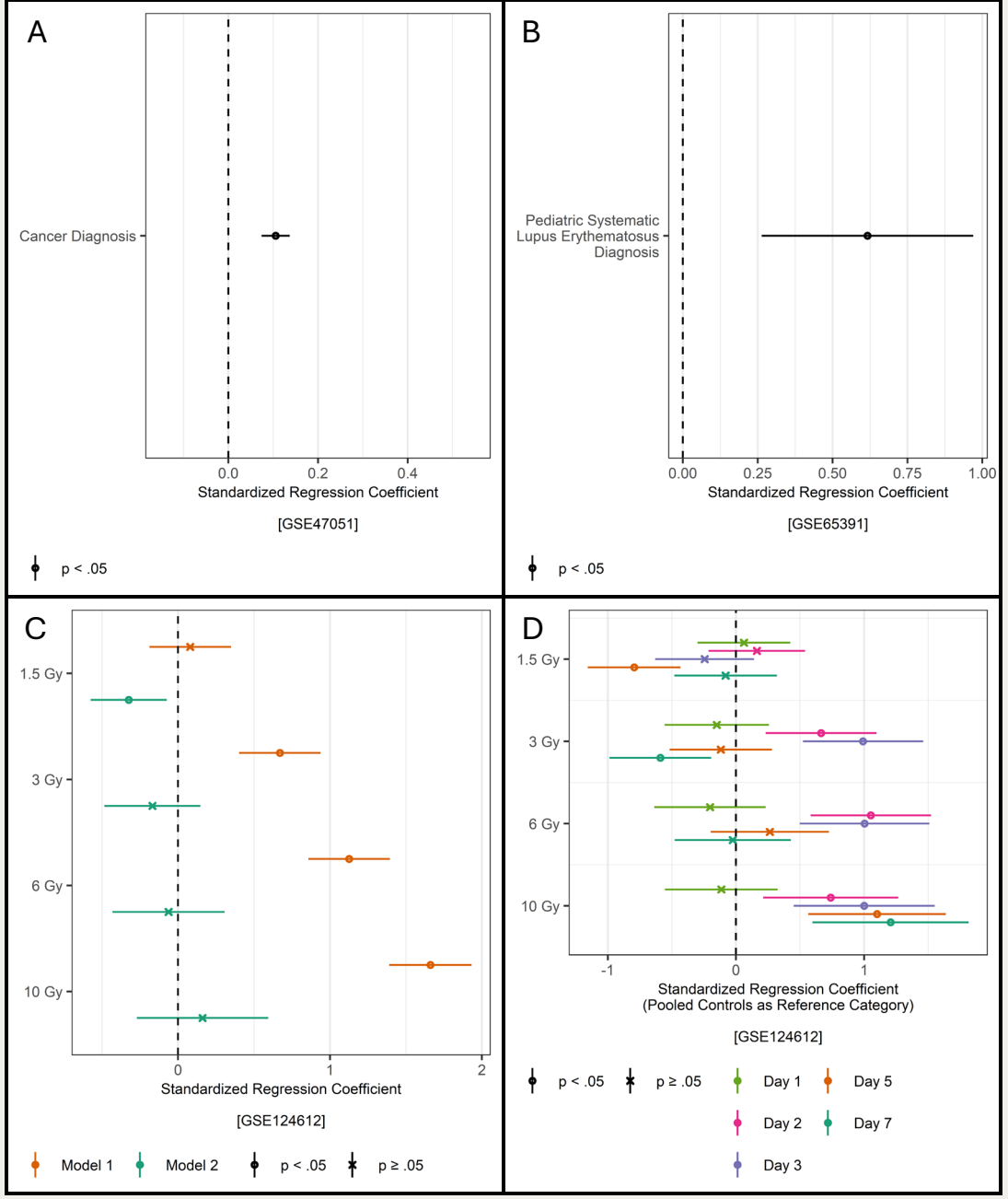


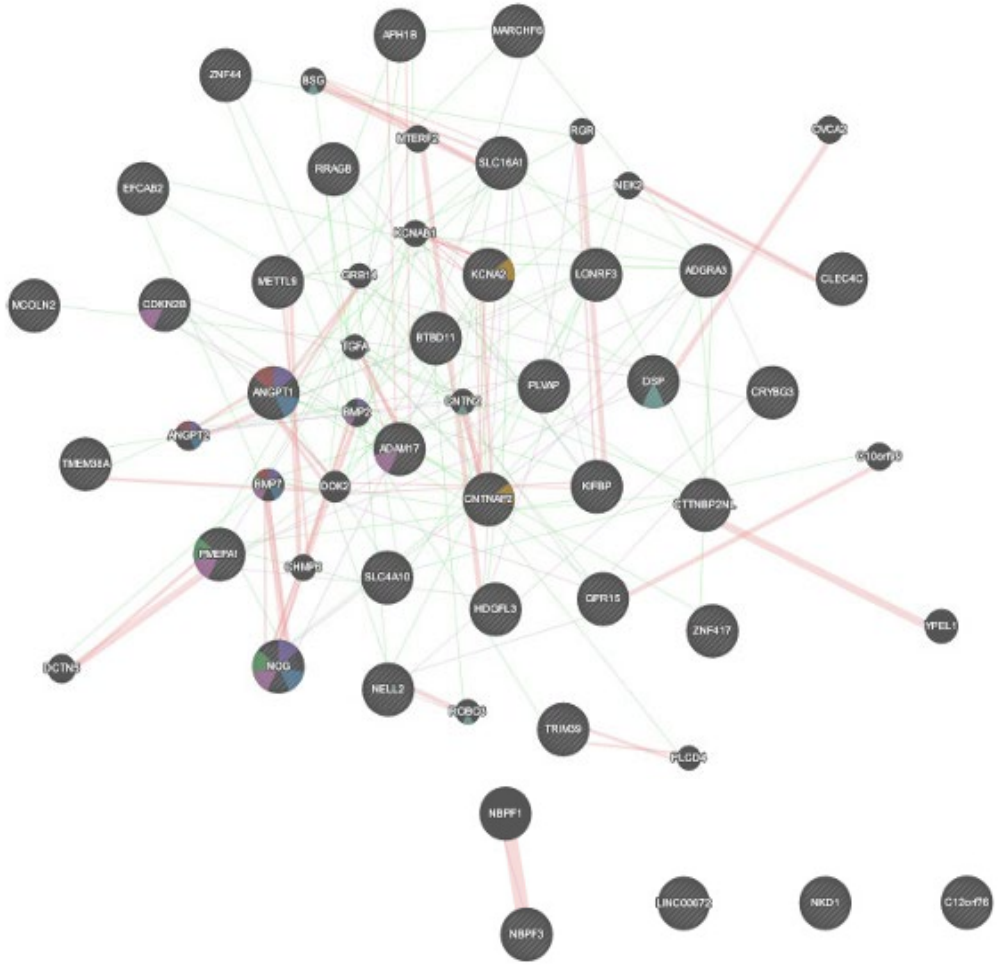





A
50,611 genes mapped in 3596 HRS participants with mortality data – Restricted to 10,964 highly expressed coding genes
Randomly Split
N = 1802 Training Sample
N = 1793 Testing Sample
Elastic Net
Trained on Time to Death
Examine associations with sociodemographic factors and health behaviors
TraMA
35 genes and age
Validate with 4-year mortality and other aging outcomes
Assess ontologies, pathways, and traits associated with these genes
N ~ 5000 Long Life Family Study
Validate with 8-year mortality
B
Age
Female
Black, not Hispanic
Hispanic
Other Race, not Hispanic
Wealth (log transformed)
Education: 0-11 Years
Education: 12 Years
Education: 13-15 Years
Never Smoked
Past Smoker
Morbidly Obese
Obese
Overweight
Under Weight
5+ Drinks per day
Non-Drinker
Phsyical Activity
-6
-3
0
3
Regression Coefficient
p < .05
p ≥ .05
Model 1
Model 3
Model 2
C
Mortality
Multimorbidity
ADLs + IADLs
Cognitive Dysfunction
0.00
0.05
0.10
0.15
Regression Coefficient
D
HRS
LLFS
1
2
3
Hazard Ratio
Model 1 (Controlling for Batch)
Model 2 (Model 1 + Age, Race/Ethnicity, and Sex/Gender)
Model 3 (Model 2 + RNA-based Cell Type)


A
Smoker Status
Pack Years (in decades)
FEV1 predicted percentage (in tens)
0.0
0.2
0.4
Regression Coefficient
[GSE171730]
p < .05
B
Diabetes Status
0.0
0.2
0.4
Regression Coefficient
[GSE123658]
p < .05
C
IBD Status
Clinically Active (Among IBD Patients)
Severity (Among IBD Patients)
0.0
0.2
0.4
Regression Coefficient
[GSE186507]
p < .05
p ≥ .05
D
Sepsis Status
SOFA Score (Among Sepsis Patients)
Mortality (Among Sepsis Patients)
0.0
0.2
0.4
Regression Coefficient
[GSE185263]
p < .05
p ≥ .05


A
Cancer Diagnosis
0.0
0.2
0.4
Standardized Regression Coefficient
[GSE47051]
p < .05
B
Pediatric Systematic Lupus Erythematosus Diagnosis
0.00
0.25
0.50
0.75
1.00
Standardized Regression Coefficient
[GSE65391]
p < .05
C
1.5 Gy
3 Gy
6 Gy
10 Gy
0
1
2
Standardized Regression Coefficient
[GSE124612]
Model 1
Model 2
p < .05
p ≥ .05
D
1.5 Gy
3 Gy
6 Gy
10 Gy
-1
0
1
Standardized Regression Coefficient
(Pooled Controls as Reference Category)
[GSE124612]
p < .05
p ≥ .05
Day 1
Day 2
Day 3
Day 5
Day 7

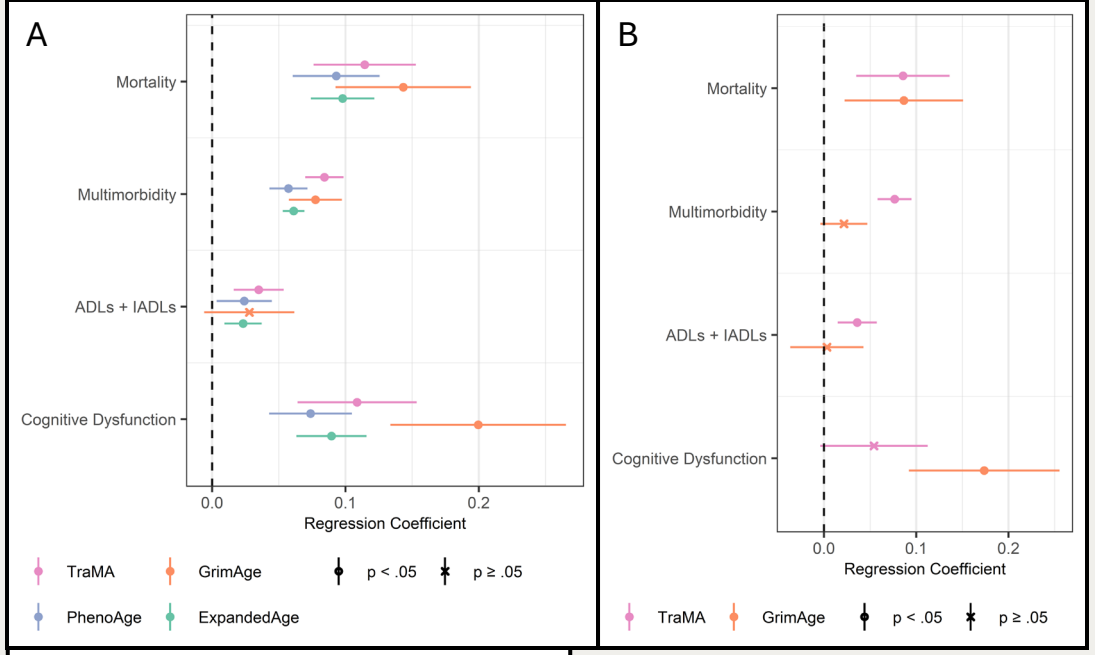


Networks

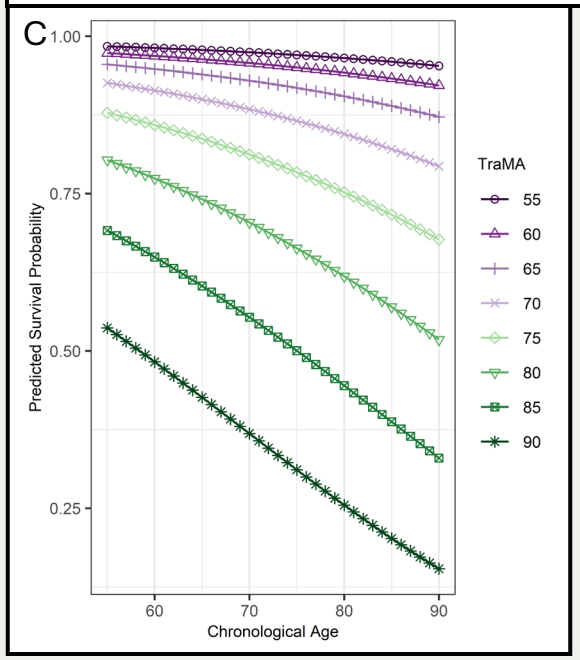
- Co-expression
- Physical Interactions
- Genetic Interactions

Functions

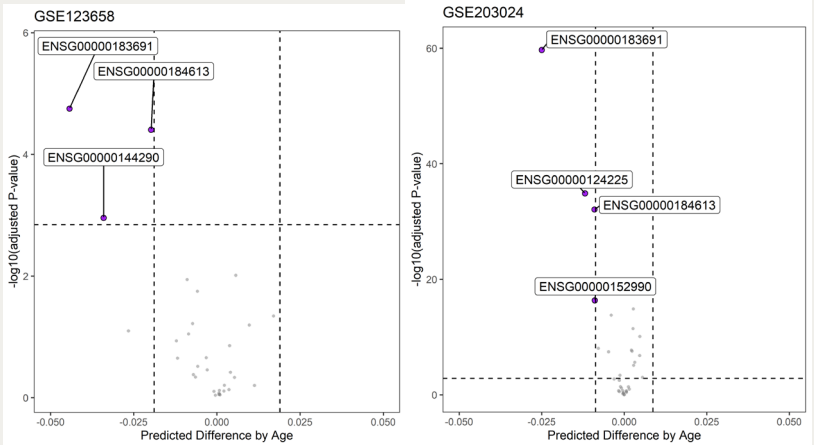
- nephron development
- main axon
- glomerulus development
- cell adhesion mediator activity
- regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
- regulation of pathway-restricted SMAD protein phosphorylation
- kidney vasculature development


A
Mortality
Multimorbidity
ADLs + IADLs
Cognitive Dysfunction
0.0
0.1
0.2
Regression Coefficient
TraMA
GrimAge
p < .05
p ≥ .05
PhenoAge
ExpandedAge

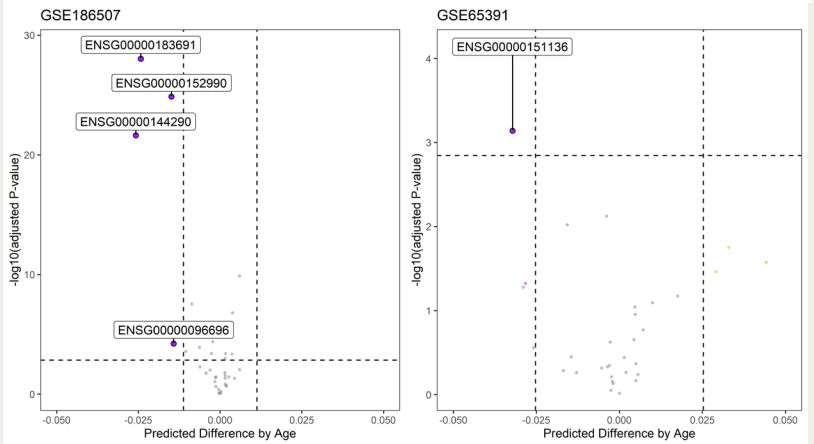
B
Mortality
Multimorbidity
ADLs + IADLs
Cognitive Dysfunction
0.0
0.1
0.2
Regression Coefficient
TraMA
GrimAge
p < .05
p ≥ .05


C
1.00
0.75
0.50
0.25
Predicted Survival Probability
60
70
80
90
Chronological Age
TraMA
55
60
65
70
75
80
85
90

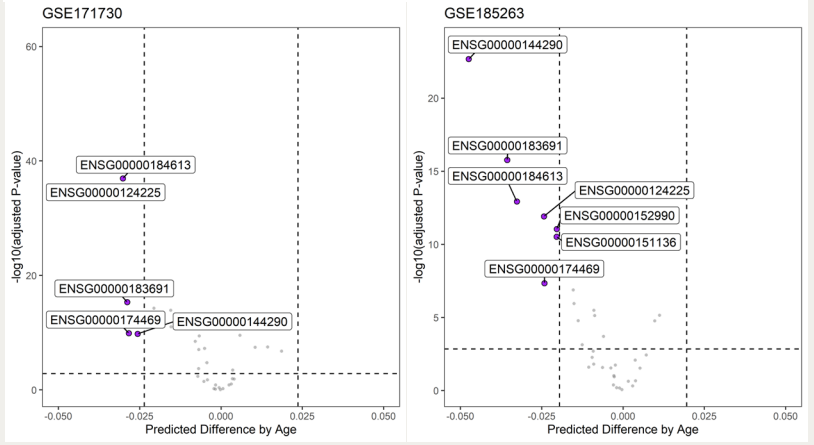

TPM
0.0
3.2
17
72
3.1e+2
1.3e+3
0.0
0.5
1.0
PLVAP
DSP
CLEC4C
GPR15
MCOLN2
BTBD11
LINC00672
NELL2
NOG
CNTNAP2
KCNA2
SLC4A10
PMEPA1
METTL9
MARCHF6
CDKN2B
ANGPT1
NKD1
LONRF3
NBPF3
EFCAB2
ZNF417
TMEM38A
CRYBG3
ZNF44
ADAM17
CTTNBP2NL
SLC16A1
ADGRA3
C12orf76
HDGFL3
APH1B
TRIM39
KIFBP
RRAGB
0
2
4
6
8
Whole Blood
Cells - EBV-transformed lymphocytes
Brain - Cerebellum
Brain - Cerebellar Hemisphere
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Brain - Putamen (basal ganglia)
Brain - Caudate (basal ganglia)
Brain - Hippocampus
Brain - Amygdala
Brain - Frontal Cortex (BA9)
Brain - Cortex
Brain - Nucleus accumbens (basal ganglia)
Brain - Hypothalamus
Brain - Anterior cingulate cortex (BA24)
Cells - Cultured fibroblasts
Muscle - Skeletal
Liver
Heart - Left Ventricle
Pancreas
Adrenal Gland
Stomach
Heart - Atrial Appendage
Kidney - Medulla
Kidney - Cortex
Minor Salivary Gland
Spleen
Small Intestine - Terminal Ileum
Colon - Transverse
Esophagus - Mucosa
Skin - Sun Exposed (Lower leg)
Skin - Not Sun Exposed (Suprapubic)
Vagina
Cervix - Ectocervix
Thyroid
Pituitary
Testis
Nerve - Tibial
Lung
Prostate
Cervix - Endocervix
Bladder
Breast - Mammary Tissue
Adipose - Visceral (Omentum)
Adipose - Subcutaneous
Ovary
Uterus
Fallopian Tube
Colon - Sigmoid
Esophagus - Muscularis
Esophagus - Gastroesophageal Junction
Artery - Aorta
Artery - Tibial
Artery - Coronary


GSE123658
ENSG00000183691
ENSG00000184613
ENSG00000144290
-log10(adjusted P-value)
Predicted Difference by Age

GSE203024
ENSG00000183691
ENSG00000124225
ENSG00000184613
ENSG00000152990
-log10(adjusted P-value)
Predicted Difference by Age


GSE186507
ENSG00000183691
ENSG00000152990
ENSG00000144290
ENSG00000096696
-log10(adjusted P-value)
Predicted Difference by Age

GSE65391
ENSG00000151136
-log10(adjusted P-value)
Predicted Difference by Age


GSE171730
ENSG00000184613
ENSG00000124225
ENSG00000183691
ENSG00000174469
ENSG00000144290
-log10(adjusted P-value)
Predicted Difference by Age

GSE185263
ENSG00000144290
ENSG00000183691
ENSG00000184613
ENSG00000124225
ENSG00000152990
ENSG00000151136
ENSG00000174469
-log10(adjusted P-value)
Predicted Difference by Age