Somatic Mosaicism across Human Tissues; Stage 2

Amy C. Lossie, PhD

SMaHT NIH Working Group Coordinator

Program Director

Division of Neuroscience and Behavior/NIDA

National Institutes of Health
Office of Strategic Coordination–The Common Fund

OSC (Common Fund) – Somatic Mosaicism across Human Tissues (SMaHT)

Concept Clearance: Stage 2 Common Fund Program

Title: Somatic Mosaicism across Human Tissues (SMaHT)

Objective: Answer three questions: What mutations matter? Why do they matter? How can we rigorously study them?

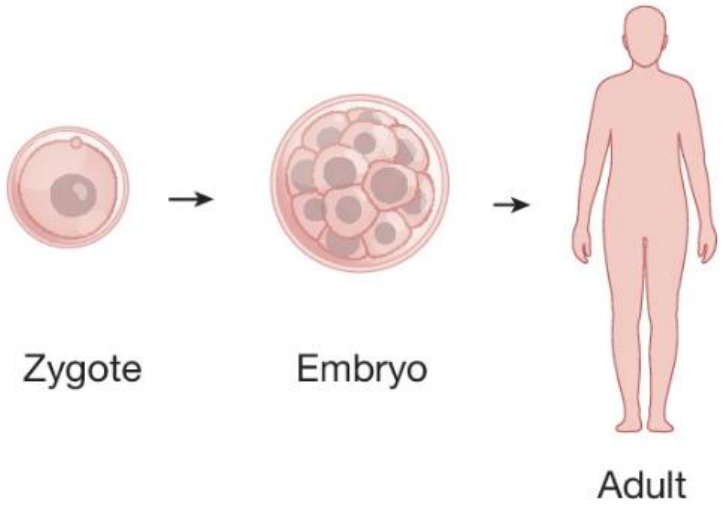
1. Tissue Procurement Center
2. Genome Characterization Centers
3. National Somatic Mosaicism Commons Data Analysis and Organizational Center
4. Experimental and Computational and Technology Development Projects

Funds Available and Anticipated Number of Awards: $40M/year; $200M over 5 years. Planned awards: one TPC (1); 3-5 GCCs; 1 NSMC-DAOC; 8-12 Experimental Technology Development projects and 3-4 Computational Technology Development Projects (contingent upon funding availability)

Program Duration: 5 years

Council Action: Vote for approval of the concept for SMaHT

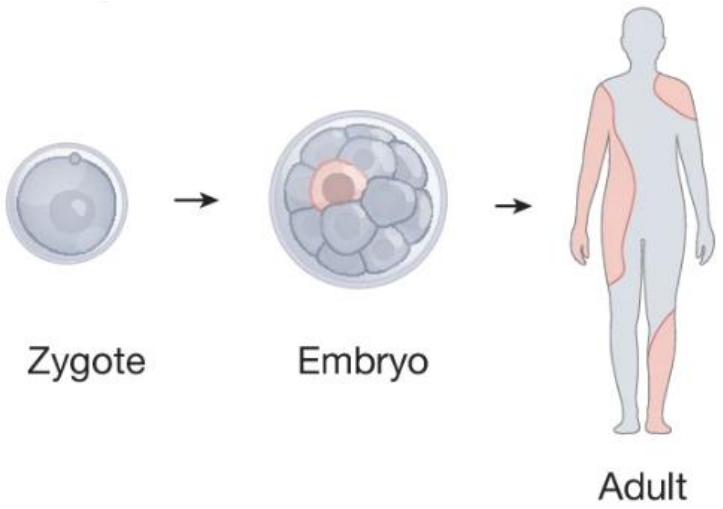
Inherited Mutations



Occur in all cells

Very easy to detect

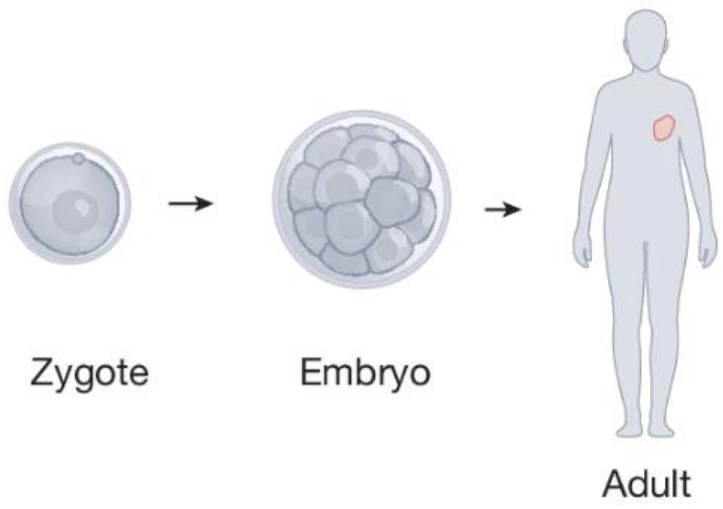
Early Somatic Mutations



Occur in many cells

Can detect

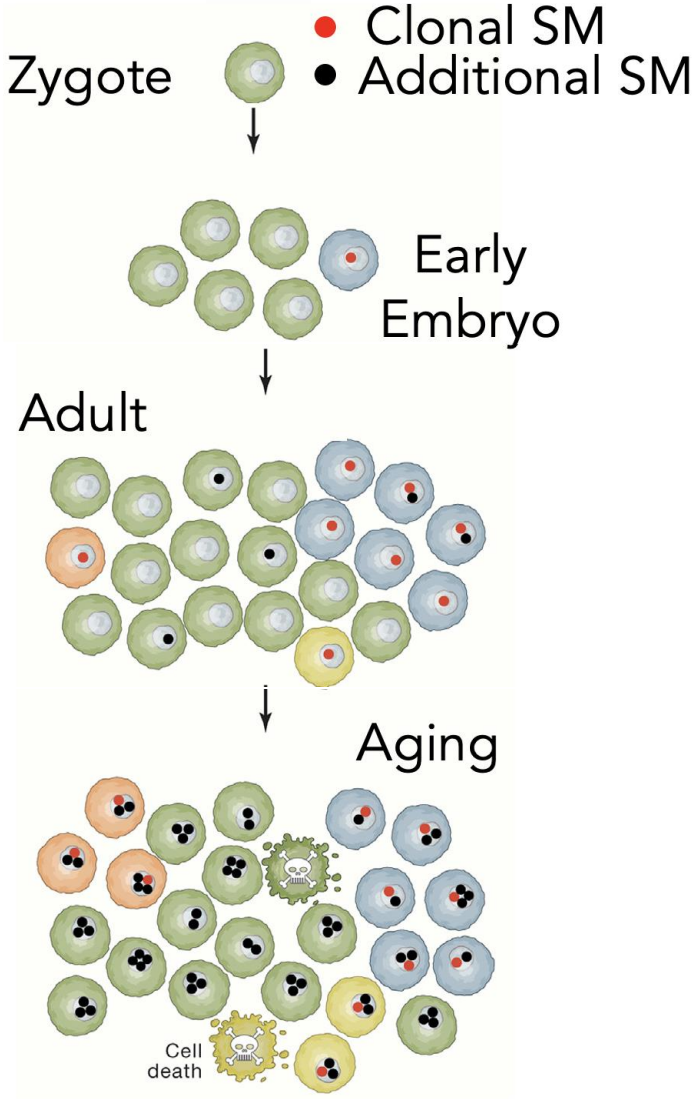
Late Somatic Mutations

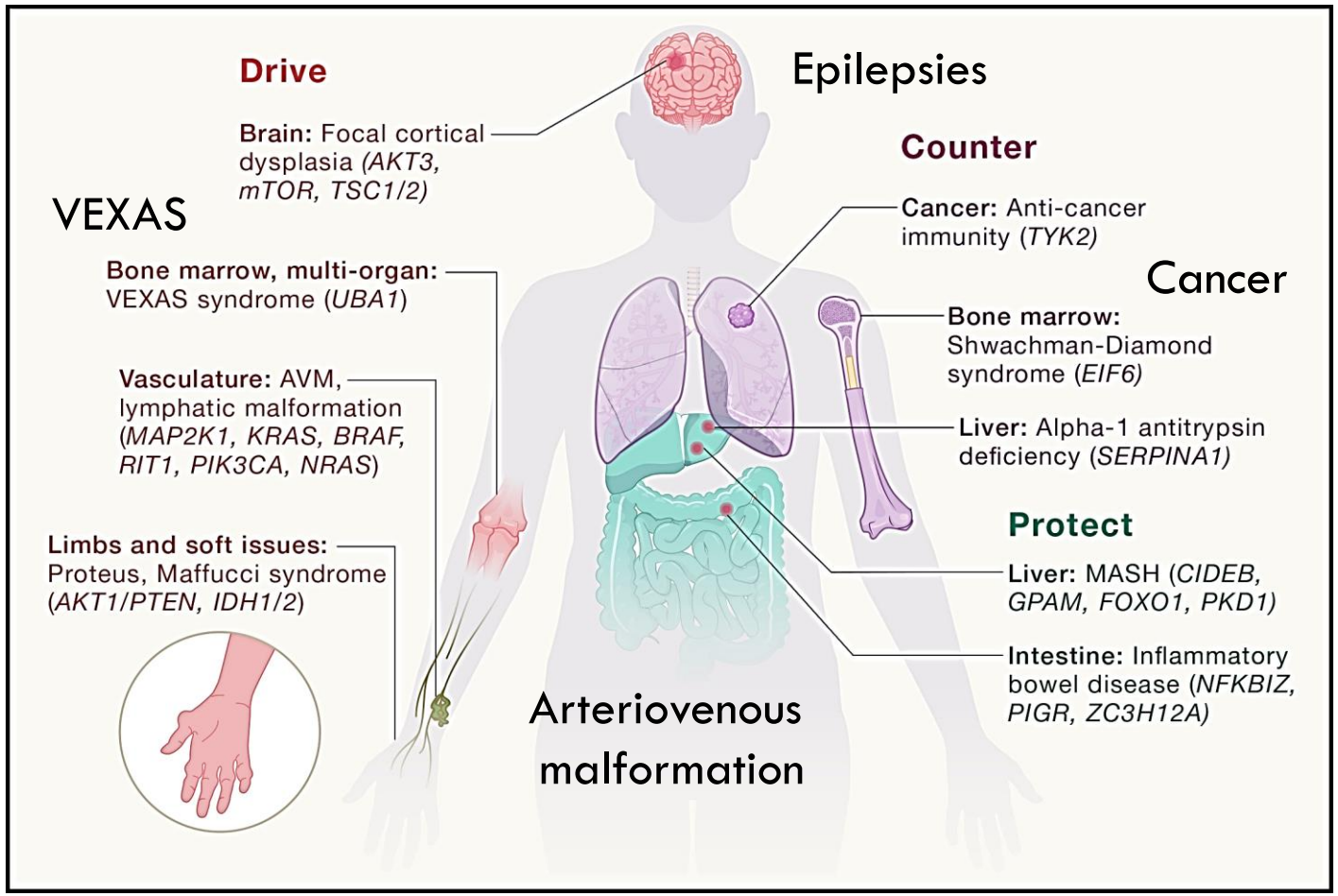


Occur in few cells

Hard to detect

Somatic Mutations Contribute to Health & Disease





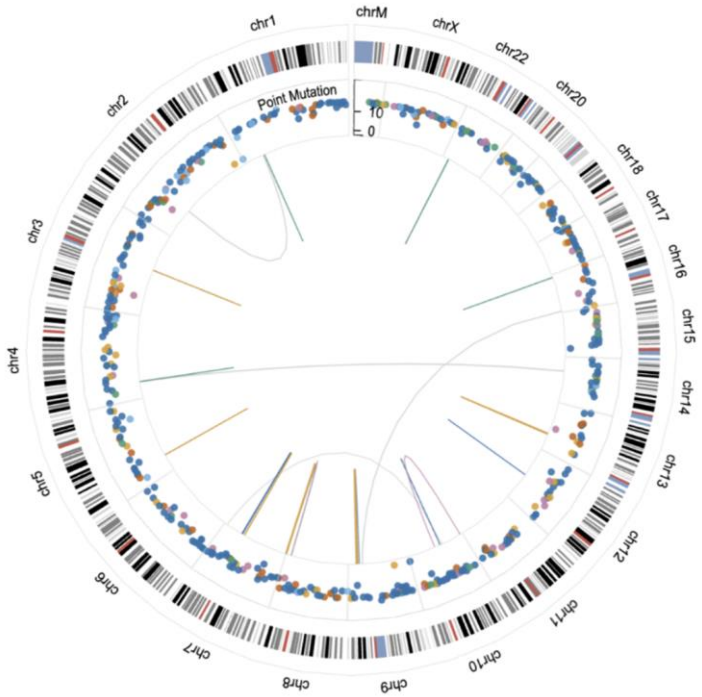
Somatic Mutations are Hard to Detect

- **Sensitivity:** low frequency (<1%) variants are hard to detect; Uneven distribution across tissues
- **Specificity:** many sources of technical variation provide significant background
- **Repetitive Regions:** CNVs, TEs etc. comprise ~45% of the genome, but need long-read
- Need high coverage or special methods

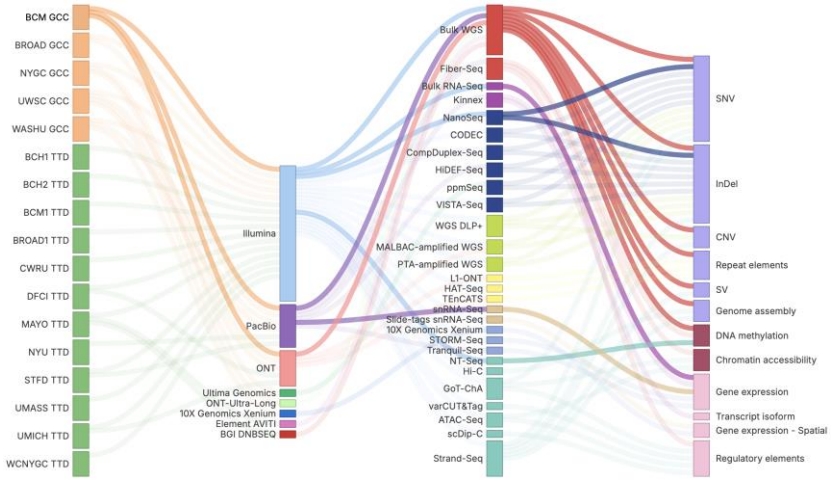


SMaHT Phase 1 Goals and Deliverables

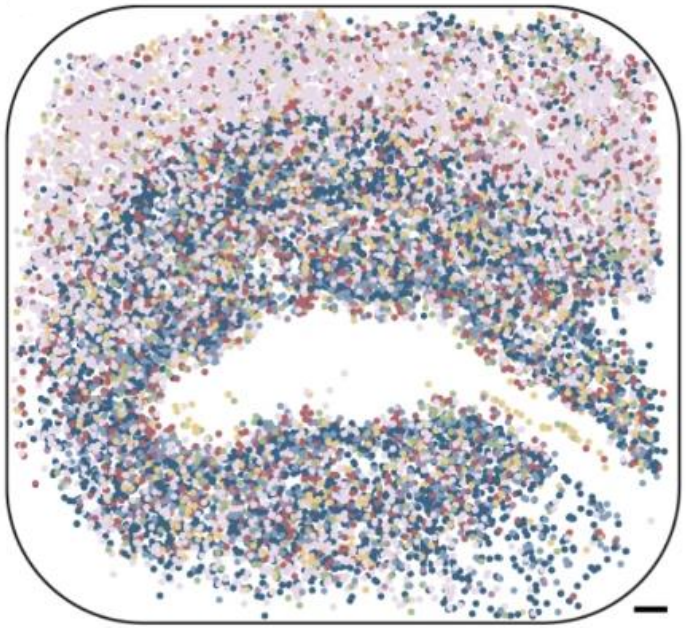
Systematically illuminate somatic variation and capture the role SM play in biological processes underlying cellular lineages, mutational signatures, mutation rates, and clonal expansions



Somatic Variant Catalog



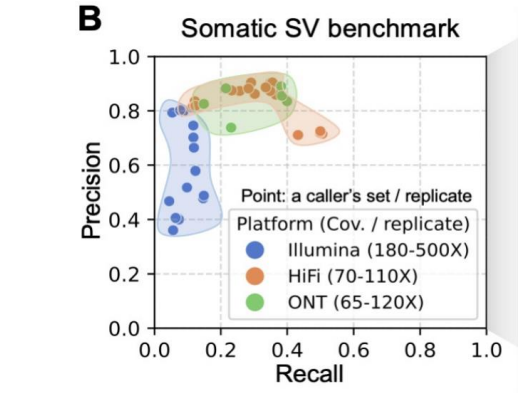
Data Workbench and Visualization Tools

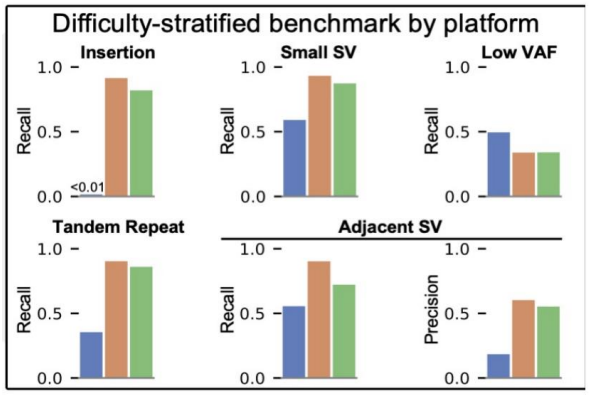


Sequencing Tools

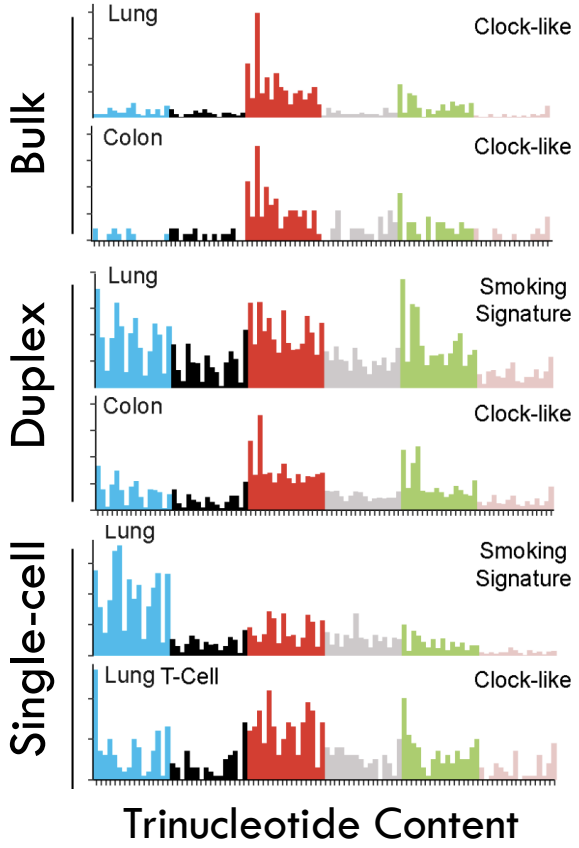
WGs Established Rigorous Methods to Detect all SM

Structural Variants Require Long-Read

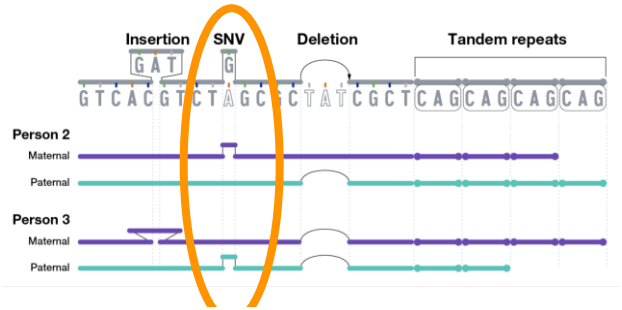


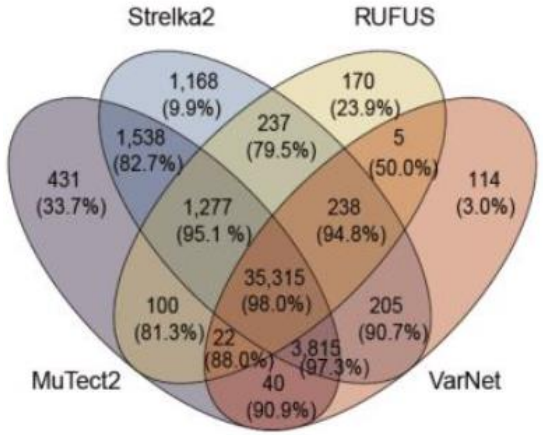


Duplex & Single Cell Studies Provide Refined Signatures

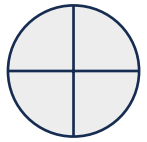
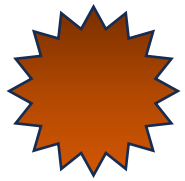


Multiple Callers Improve Rigor





First 25 Donors

	# Donors		
 Drugs Alcohol Tobacco	14	Male; 59 GERD	Male; 67	Female; 78
 Cancer, Heart disease	16	High Mutation Burden	Large Burden in Skin	Large Burden in Skin & Liver
 Toxins & Chemo	4; 2	High Transposon Count in Colon	Clonality: Different Mutation Burden across Cores	Clonality: Different Mutation Burden across Cores
 Radiation	3			

We're finding interesting somatic mutations…

How do we systematically determine what, if anything, these mutations are doing in the cell, tissue, organism?

SMaHT Stage 2: Interpreting Somatic Mutations



Community Feedback

Request for Information

- External Experts
- SMaHT Investigators
- SMaHT Program Consultants
- NIH WG members

Focused Interviews

- SMaHT Junior Investigators

Consistent Responses

Benchmarks: shared datasets + evaluation

FAIR data: accessible releases + portals

Training: clinical use + early careers

Function: validation + disease relevance

Stage 2 Priorities

Find → Confirm → Test → Interpret

- Expand healthy reference
- Add clinically relevant cohorts
- Link variants to cell state, tissue context, gene regulation, and disease
- Release reusable and reproducible workflows

SMaHT Stage 2: Which somatic mutations matter? Why do they matter? How can we rigorously study them?

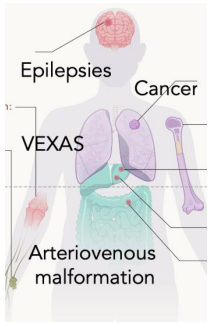
Organizational Center



Leadership
Governance
Training

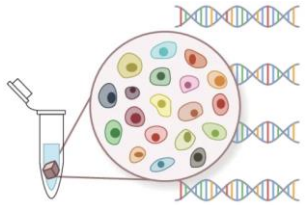
Healthy & Clinical Samples

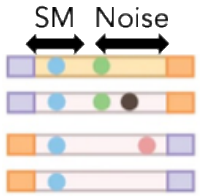




Rigorous sample collection
Extensive metadata
Biobank

Deep Sequencing

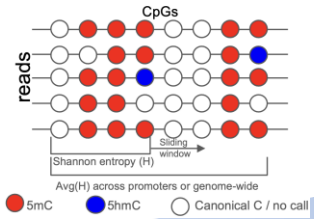




Data Generation
Tool Integration
Rigorous design
Reproducible data

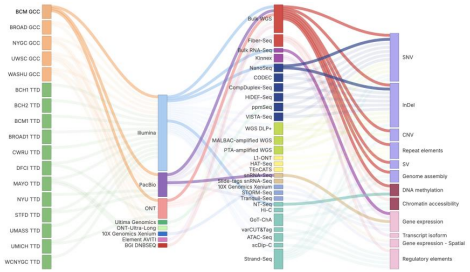
Tool Development





Fills gaps
Adds value
Solves problems

Multi-Omics Analysis



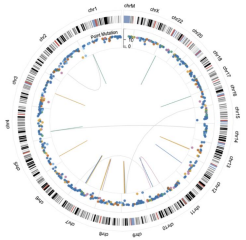
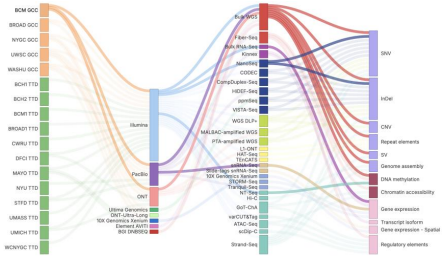
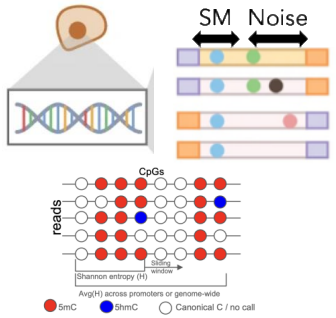
Integrates all data
Develops pipelines
Visualization portal
Sustainable resource
gnomAD/ClinVar-like resources

Portable Workflows



Cross-platform
Reusable
Plug & Play
Facilitates community use

Stage 2 Provides a Portable Platform to Test, Validate, and Determine Function of Somatic Mutations

Stage 1			gnomAD Genome Aggregation Database
Stage 2			ClinVar
Impact	Turns difficult detection problems into usable methods	Converts the SMaHT Portal into a Recomputable Scientific Commons	Employs coordinated, benchmarked methods to determine effects of a mutation

Common Fund Criteria

SMaHT Stage 2 provides a cross-tissue foundation that disease-specific studies are unlikely to build on their own.

Transformative	Moves somatic mosaicism from variant discovery to interpretation using a shared human tissue reference and evidence framework.
Catalytic	A five-year investment will make available reusable data, SOPs, benchmarks, workflows, tools, training, and a research framework for broad community use.
Goal-Driven	200 healthy donors, 3 or more disease/phenotype cohorts, versioned data releases, benchmarked reproducible workflows, and tools for hard-to-detect variants.
Synergistic	Serves many NIH missions by giving disease, aging, genomics, diagnostics, and data science communities a common reference and shared methods.
Novel	Fills a gap not addressed within NIH portfolio: a coordinated normal-tissue baseline, standardized methods, cross-tissue data, and functional validation.

When complete, researchers will be able to answer which somatic variants matter, why they matter, and how they should be reproducibly studied

Budget summary: $202.5M over five years

Initiative	FY28	FY29	FY30	FY31	FY32	Total
Initiative 1 – Tissue Procurement Center	$2.5M	$2.5M	$2.5M	$2.5M	$2.5M	$12.5M
Initiative 2 – Genome Characterization Centers	$16.5M	$16.5M	$16.5M	$16.5M	$16.5M	$82.5M
Initiative 3 – NSMC-Data Analysis & Org Center	$8M	$8M	$8M	$8M	$8M	$40M
Initiative 4 – Computational and Experimental Tech Dev	$13M	$13M	$13M	$13M	$13M	$65M
RMS - NIH staff salary/travel and NIH-organized workshops	$500K	$500K	$500K	$500K	$500K	$2.5M
TOTAL	**40.5M**	**40.5M**	**40.5M**	**40.5M**	**40.5M**	**$202.5M**

Budget component notes: Initiative 1 includes Tissue Procurement Center ($2.5M/year). Initiative 2 encompasses 3-5 Genome Characterization Centers ($16.5M/year). Initiative 3 includes one National Somatic Mosaicism Commons Data Analysis and Organizational Center award ($8M/year). Initiative 4 includes 8-12 experimental technology development awards ($8M/year) and 3-4 computational workflows ($5M/year).

SMaHT NIH WG Members

NIH Working Group members and Co-chairs providing input

Co-Chairs providing input

- Nora Volkow (NIDA)
- Roger Little (NIDA)
- Erin Ramos (NHGRI)
- Carolyn Hutter (NHGRI)
- Amy Adams (NINDS)
- Dan Gallahan (NCI)

Legend

NIH WG Leadership Team

NIH WG Members

OD
Ajay Pillai
Dena Procaccini
David Bollweg
Pamela Birriel

NCI
Ron Johnson
Wendy Wang
Linda Zane

NCATS
Passley Hargrove
Tiina Urv

NHGRI
Sarah Wheelan

NIA
Max Guo
Alison Yao

NIAMS
Yan Wang

NIDA
Amy C. Lossie
Mehrnoosh Ahmadi

NIEHS
Michelle Heacock

NINDS
Jill Morris

Budget Request: $202.5M over five years

SMaHT Stage 1 Highlights

- >100 papers – 1300 citations by non-SMaHT investigators
- Benchmarking paper package slated for Oct 15 publication in *Cell*
- P25 Flagship paper package under negotiations
- Incorporated duplex sequencing into core pipeline
  - detects mutations at 10^{-8}
- Benchmarking single-cell analysis in core pipeline
  - detects single molecule mutations
- Finding interesting mutational events in a subset of donors
- Finding evidence of exposures and life events through analysis of mutational signatures

Council Action:

Vote for approval of the concept for Somatic Mosaicism across Human Tissues (SMaHT): Stage 2

commonfund.nih.gov

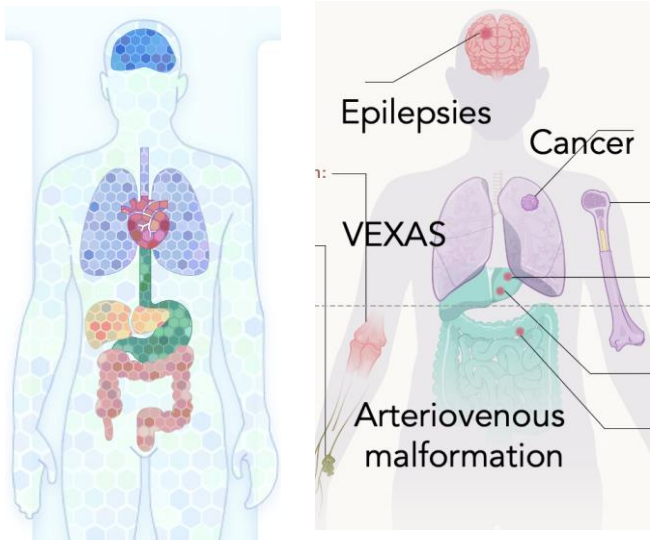
National Institutes of Health
Office of Strategic Coordination–The Common Fund

Initiative 1: Tissue Procurement Center

Purpose: 1. Create tissue banks from healthy populations and special sample cohorts; 2. Develop SOPs for tissue collection, distribution, banking, and long-term storage

Deliverables:

- Biorepository of well-characterized tissues from 200 healthy donors (Phase 1 + Stage 2)
- Biorepository of up to 10 special sample cohorts
- Standardized tissue collection, preservation, pathology, metadata schema, and distribution protocols that are interoperable with other programs (e.g. GTEx, dGTEx)



Tissue Banks

Initiative 1: Tissue Procurement and Genome Characterization Centers

Purpose: 1. Comprehensively catalog somatic variants in 50 new healthy donors and 3 cohorts of clinically-relevant samples and 2. Conduct a deep dive of somatic mutations in 25 healthy donors

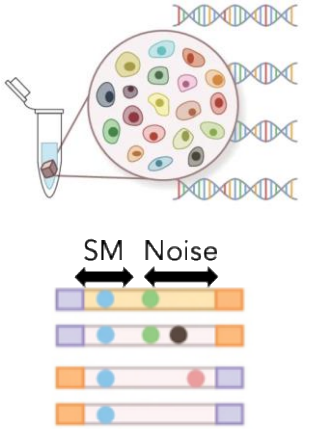
Deliverables:

- Functional and deep multi-omics analysis of a subset of healthy and clinically-relevant samples
- Sequencing Pipelines to characterize clinically-relevant samples at scale

Deeper Understanding of:

- Variant location, frequency, and tissue specificity
- Cell lineages and cell fate
- Variant accumulation in normal and clinically-relevant cells
- Types and extent of somatic variation in core tissues

Deep Sequencing



Data Generation

Tool Integration

Rigorous design

Reproducible data

Initiative 3: NSMC Data Analysis and Organizational Center

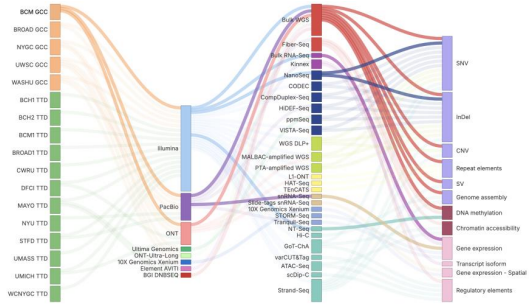
Purpose: 1. Create a stable data portal with visualization tools using GA4GH and FAIR standards; 2. Devise portable, reusable analysis pipelines for the NSMC and other researchers; 3. Provide leadership, governance, and training

Deliverables:

- Variant Catalog of healthy and clinically-relevant samples
- Data workbench with visualization tools that integrates with existing genomic data resources
- Universal pipelines other teams can adopt
- Pilot projects, community engagement

Scientific Advances:

- A better understanding of how timing, developmental trajectories, and mutational signatures of somatic mutations contribute to human health and disease processes
- Plug and play sequencing and analysis pipelines based on benchmarked samples, best practices, and vetted analysis tools modeled after gnomAD/ClinVar



GA4GH/FAIR Data Workbench and Visualization Tools

gnomAD/ClinVar

Portable Pipelines



Leadership

Governance

Training

Initiative 4: Experimental and Computational Technology Development Projects

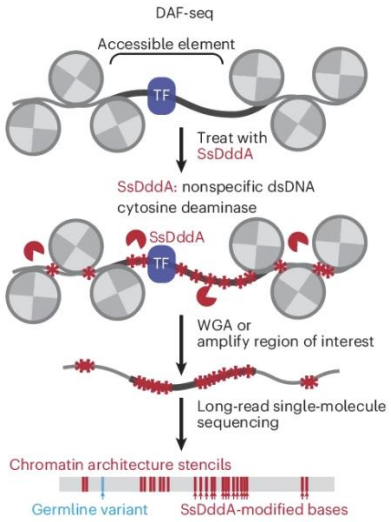
Purpose: Accelerate development, optimization and implementation of tools and data analysis pipelines for significantly improving sensitivity and specificity of variant detection, integrated multi-omics analysis, and spatial integration

Deliverables:

- Improved detection of mutations in challenging areas, low-frequency mutations and difficult samples
- Portable, cross-platform workflows

Scientific Advances:

- Increased accuracy of variant detection
- Analysis of structural variants, mobile DNA and repetitive DNA in small samples at low allele frequencies
- Integrate SNVs, structural variants, mobile DNA and repetitive DNA into germline genetic studies
- Agnostic analysis pipelines for increased community use



Precise Sequencing Tools



Portable Workflows