Concept Clearance for a new Common Fund Program
SMaHT: Somatic Mosaicism Across Human Tissues

Summary

Program Goal:	*To systematically illuminate somatic variation and capture the role SM play in the establishment of the personal genome underlying biological processes of human health*
IC Director(s) or Senior IC Staff:	*Walter Koroshetz (NINDS), Rick Woychik (NIEHS), Eric Green (NHGRI) Roger Little (NIDA), Joshua Gordon (NIMH)*
Proposed Initiatives:	*1. Conduct a systematic investigation of somatic variants in select tissues from diverse human donors* *2. Develop innovative tools that optimize identification of somatic variants* *3. Create data and analysis toolkits using an open, FAIR workbench*

Background Information

Somatic mutations occur widely throughout the human lifespan as a response to environmental insults, dysregulation of DNA repair processes, activation of transposable elements, expansion and contraction of simple repeats, and other factors that affect the genome and epigenome, leading to clonal and stochastic somatic mosaicism. These changes build upon the ancestral DNA variation and the private, *de novo* variants that occur during gametogenesis to create a personal genome built from the cumulation of these events. Most current genetics and genomics studies focus on DNA variation that is captured from blood, which is used as a proxy for germline variation. However, mounting evidence illustrates the contribution of somatic variation to fetal development, disease processes, and aging. These pivotal studies underscore the importance of somatic variation in biological processes, yet only capture a small sampling of the breadth, depth, tissue specificity, drivers and consequences of somatic variation on health and development. The Somatic Mosaicism across Human Tissues (SMaHT) Common Fund program would be the first program to systematically document somatic variation and develop new molecular and data-driven tools to capture the role somatic variation plays in the establishment and function of the personal genome.

SMaHT Phase 1 Initiatives (Years 1-5)

The first phase of the SMaHT Program is designed around three initiatives:

1. Collection of anatomical, genomic, and transcriptomic data across multiple organ systems
2. Optimization and creation of molecular tools and data analysis pipelines to increase the sensitivity and specificity of detection of low-frequency variants (e.g. SNV, structural variants, and mobile DNA) throughout the genome, including repetitive sequences (e.g. the “dark genome”)
3. Development of a user-friendly, FAIR data workbench that integrates with current systems and easily interfaces with multiple genome browsers

The goal of Initiative 1 is to comprehensively catalog somatic variation in a core set of tissues in approximately 75 individuals, prioritizing opportunities that maximize the understanding of the genomic diversity and prevalence of somatic mosaicism. A fundamental component of SMaHT is the need to obtain

a core set of tissues (10-15 tissues from all 3 germ layers) from each participant in order to curate tissue-specific differences, the effects of environmental exposures, and contributions of micro-environments to these mutational processes. Key deliverables from this aim include: a curated biorepository of well-characterized tissues; a reference catalog of tissue-specific variants; and standard operating procedures for tissue collection, sequencing, collection of metadata, and construction of data dictionaries. These studies will yield a deeper understanding of variant location, frequency, and tissue specificity; variant clonality underlying the processes involved in cell lineages and cell fate; and the types and extent of somatic variation identified in these core tissues.

The goals of Initiative 2 are to: 1. Optimize current 'omics tools and data analysis pipelines and 2. Develop SMaHT-specific tools focused on increasing sensitivity of detection and specificity of the pipelines to detect multiple types of variations (e.g. SNVs, structural variants, mobile DNA), while reducing technical and biological noise. Key deliverables from this aim include robust detection of low frequency somatic variants in repetitive regions and variants in small sample sizes, as well as development of tools and pipelines for integrated multi-omics analyses. The SMaHT program will drive the creation of sequencing tools and data analysis pipelines that will benefit all types of genomics research, including the ability for increased accuracy of variation detection, as well as the sensitivity and specificity to discover all kinds of variants in small samples.

The goal of Initiative 3 is to create an open, FAIR data workbench that fully integrates with current genome browsers. This workbench will be interoperable with current platforms and facilitate analysis of SM alongside genomic variation. Key deliverables are: rapid access and sharing of biospecimens, protocols, datasets, and analysis pipelines; tools to analyze intra-individual and inter-individual somatic variation; and an integrated data workbench that facilitates harmonization of SMaHT tools with related programs. The SMaHT program will advance our understanding of the drivers and consequences of somatic mosaicism and integrate the personal genome as one facet of the biological processes underlying human health.

SMaHT Outputs and Outcomes

The SMaHT program builds on large-scale programs designed to capture genetic variation and understand the role of these variants in health and disease. Emerging studies highlight the role of somatic mutations in fetal development, aging, and disease. However, current programs do not typically consider the role of somatic mutations in these processes. Successful completion of the SMaHT Program will provide the biomedical research community with a somatic mutation scaffold; tools and data analysis pipelines that refine genomic and transcriptomic sequencing; and an integrated workbench to seamlessly analyze somatic variation alongside the reference genome. These tools and resources will ultimately provide the opportunity for systematic analysis of somatic variation in a multitude of genetics and genomics studies.