NEW YORK – A new $140 million project to survey the diversity of genomes of different cells inside a human body plans to commit millions to develop new technologies along the way.

Announced in May by the US National Institutes of Health and funded for at least five years, the Somatic Mosaicism across Human Tissues (SMaHT) program tries to bring together just about every sequencing technology out there to address the challenge of finding true genetic variation between cells of the same individual and working out what those changes mean for health and disease.

Participants in the program are expected to spend about 20 percent of their grant funds on "collaborative activities," which could include benchmarking methods, developing data visualization tools, or analyzing tissue slices by complementary methods.

The 14 technology development grantees will work with each other and five so-called genome characterization centers. Of the $140 million, at minimum $7 million, likely more, will go to tool developers: approximately $250,000 in direct costs for two years, plus varying amounts of indirect costs. Their projects will then be evaluated for utility, and funding could be renewed for another three years. The entire SMaHT network could also be renewed for another five years.

"That's a massive investment, astronomically large," said Phil Jones, a researcher at the UK’s Wellcome Sanger Institute, whose research over the last decade has helped show that somatic gene variants can be found in normal tissues throughout the human body. "Clearly, [NIH] thinks it’s very important, and I would concur."

"We need to understand how [these genetic differences] vary between individuals and between populations," he said.

The program’s concept is a little hard to grasp all at once. "This is the first systematic approach to categorize somatic and mosaic variants," said Fritz Sedlazeck, a researcher at Baylor College of Medicine, who is working with Tao Wu, also of Baylor, on strategies to identify structural variants and transposon activity as they relate to epigenetic modification.

"My own definition is that somatic mutations are unique to a tissue while mosaic variants are really rare within a tissue," he said.

A major goal of the program is to create a catalog of human somatic variation by sequencing samples from approximately a dozen sets of tissues, collected from 150 individuals. That will likely include samples from the brain, blood, skin, muscle, colon, spleen, uterus, vas deferens, ovaries, and testes. In addition, the project aims to produce tools to study somatic and mosaic variants and a data workbench that will integrate with existing databases and computational tools.
The project's work in identifying variants will be "very important for annotation resources," Sedlazeck said. It also represents a fresh scientific challenge.

"It's almost boring to call SNPs on an individual for germline. We can do it on large-scale clinical genomes, so that's exciting, but from a developer's point of view, it's getting boring, and we can't innovate much more." Working on SMaHT will allow him to "jump into a new universe of variants."

The researchers will face challenges at every level of the project. First, they must identify many, very rare variants. For most, the allele frequency will be below the limit of detection, given the error rates for standard long- and short-read sequencing approaches. Thus, methods like duplex sequencing will be needed to verify that the variants are real. Then, they must be assigned to a cell type. Here, single-cell and high-resolution spatial methods will come into play. Finally, the project will evaluate different clonal populations for phenotype, which will likely include the context of epigenetic or multiomic data.

To make matters more complicated, every tissue is different, and the researchers will "need to look at clones in an informed way," Jones said. "The technology you use and the way you apply it has to be done in a way that's informed by the architecture of the tissue. That's something we've learned doing this for a decade. You can't have a one-size-fits-all approach."

Alexej Abyzov, a researcher at the Mayo Clinic and one of the tools development grantees, is working on a method to detect mutations in cells that combines the benefits of sequencing colonies of cells and single-cell whole-genome sequencing. "What we are trying to do is use cultured cells for some limited time to get a few copies [of that genome], and then use amplification," he said. Starting with more than one copy of the target genome means errors in early rounds of amplification get diluted, while amplification will enable the study of cells that don't easily grow into large colonies for bulk sequencing.

In his early work for SMaHT, he plans to use single-cell whole-genome amplification from BioSkryb, a St. Jude Children's Research Hospital spinout based in Durham, North Carolina.

Another tool for rare variant detection is duplex-consensus sequencing, where both strands of a molecule are linked in the sequencing library preparation. Several duplex sequencing technologies have been funded, including a Tn5-based method — initially established by Sunney Xie's lab at China's Peking University — being refined by a team at Boston Children's Hospital led by Sangita Choudhury. They're looking at "making it much simpler" so that it could be done by labs that aren't primarily focused on genomics, she said.

Her hopes are that her assay can be used as a validation technology for when the genome centers make a variant call.

She also suggested that the assay could be adapted to work with Slide-seq, a spatial omics method invented by Fei Chen of the Broad Institute that is also the subject of a tools grant. The two labs haven't formalized a collaboration yet, Choudhury said, "but we're thinking how we can work together to bring Slide-seq and duplex sequencing together."

Chen's grant includes ATAC-seq (assay for transposase-accessible chromatin by sequencing) pioneer Jason Buenrostro and Gad Getz, also of the Broad.

"We're proposing to improve quality to have higher [genomic] resolution," for spatial analysis of DNA, Buenrostro said. "By increasing genomic resolution, we should be able to track these clones." The first Slide-seq assays were for spatial transcriptomics, and the technology has recently been adapted to provide information on copy number variants.
The grant will support ways to increase the recovery of genomic fragments to increase coverage, allowing detection of a single gene variant. They’ll also work on targeted capture of regions of interest.

In general, Slide-seq is "really good at identifying clustered cells that might share a feature, like a somatic mutation, and how that variant is causing cells to expand within a tissue, or how it might reprogram cells to alter tissue function," said Dan Landau, of Weill Cornell Medical School and the New York Genome Center.

He likened the human body to a "Raggedy Ann" doll, made of different fabric patches sewn together. "We are patches of clones that come together as one," he said.

Those clonal patches are, of course, made of individual cells, and Rahul Satija, of New York University and the NYGC, is a leader in the field of single-cell data analysis. He’s working with Landau to combine single-cell transcriptomics with many other layers of omics data, including histone modification, DNA methylation patterns, and protein expression.

"This grant will allow us to create a suite of tools to phenotype a specific genotype in the context of somatic mosaicism," Landau said. These types of tools will help determine the downstream results of somatic variants that help cells grow into larger clonal populations.

He pointed to a blood condition called clonal hematopoiesis of indeterminate potential. A largely incidental finding, a patient with CHIP may go to a doctor, and standard pathology testing will say nothing is wrong, he said, while a molecular assay might reveal that 10 to 20 percent of cells will have a mutation in the splice factor SF3B1, a variant also associated with some types of myelodysplastic syndrome, which can cause symptoms like fatigue and carries an increased risk for leukemia. "The phenotype has to be different," Landau said. "And yet we don't have any way of knowing what about these cells allows them to behave differently."

"Our ambition is to be able to link somatic genotypes with somatic phenotypes," he said.

Satija said he’s excited to be a "part of a whole" that will unlock the use of the tools he has developed. To do it alone, "we would need to reproduce a huge infrastructure to get the samples," he said. "Better for us to come in with unique added value and do it in collaboration with a larger network that is able to identify samples."

While the SMaHT grants have all been awarded, NIH has left the door open for additional researchers to step in. It is "likely to establish an associate member policy to facilitate collaborations with researchers who are not funded directly through the SMaHT program," NIH said in a statement.

Filed Under: Sequencing, Research Funding, Informatics, North America, Spatial Biology, NIH, variant classification, single-cell sequencing, single-cell genomics, targeted sequencing, NIH funding