Seeing the Tree AND the Forest:
Understanding Individual and Population Variation in Biology, Medicine, and Society

Background

Individuals differ from each other in a myriad of ways, from their DNA to their behavior to their lifetime health. Understanding the underlying causes of this variation across individuals and populations is of the utmost importance for ultimately promoting the success of both the individual and the population within which they live. However, the directionality of cause and consequence is complex, and the pertinent factors that underlie why individuals are the way that they are span disciplines and cross the traditional boundaries of research. Human health, for example, is investigated by clinical researchers and healthcare professionals, basic and applied biologists, sociologists, statisticians, and more. While genetics clearly affects individual health outcomes, health cannot be fully understood without uncovering the cognitive processes of decision-making. Decision-making is mechanistically based in the structure and signaling of the brain, but family and friends, education, and socioeconomic status all play important and overlapping roles in health-related decisions and, therefore, health outcomes. Different populations—based in ethnic background, gender, sexual orientation, socioeconomic status, geographic location, and more—have differential access to education, work, and healthcare. These population-level metrics affect individual mental and physical health, thus shifting the state of the population. The interactions between individual and population are reciprocal, dynamic, and not well understood.

The recent explosion of interest in "Personalized Medicine" (often also referred to as Precision Medicine) has underscored the complexity of the causes and consequences individual and population variation. These issues are at the forefront of the new Dell Medical School, with its focus on Population Health, and healthcare in general, because there is fundamental responsibility in medicine to both the patient and the population. A promising and popular approach to understanding the causes of variation at the individual and population levels relies on technological advances in the collection, management, and analysis of large amounts of data (i.e., 'Big Data') (Fig. 1). For example, it is now feasible to sequence all of the genes expressed in the tumor of a patient with cancer. But it is not straightforward how information about tens of thousands of genes from a single individual (n=1) can be used to optimize treatment and care. Using conventional models, there is no statistical power in predicting outcomes for a single individual. Furthermore, it is a fallacy to assume that more data necessarily results in a more complete understanding or improved care. Patterns within the data may or may not be meaningful, and choices surrounding analysis and interpretation can have dire consequences for individuals and populations.

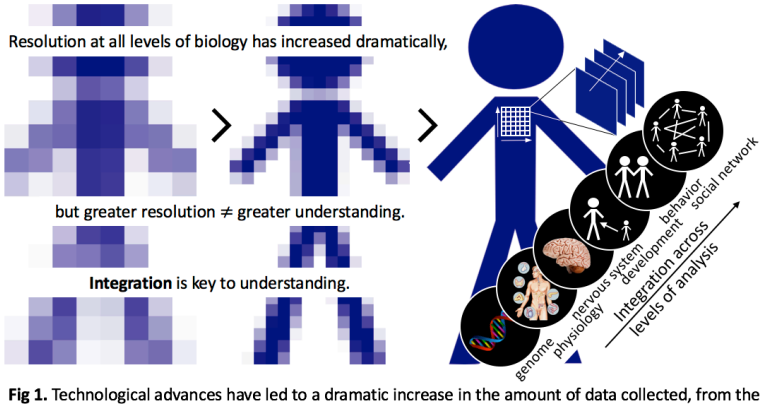

Fig 1. Technological advances have led to a dramatic increase in the amount of data collected, from the level of the genome to large social networks. But it cannot be assumed that more information equates to a greater understanding or improved outcomes for individuals or populations. Integrative approaches and collaborations across disciplines will be critical to gaining understanding and promoting success.

Although in many ways the questions and challenges associated with Personalized Medicine appear unique to human health, understanding the causes and consequences of individual and population variation also lies at the core of biological research. An individual's phenotype (i.e., the collection of observable characteristics) determines its fitness, or the amount of genetic material passed on to the next generation(s). Comparable measures of individual success in other disciplines could be human health, or personal, academic, or professional success. Natural selection acts on phenotypic variation within a population such that more successful, higher fitness phenotypes are more likely to persist. However, little is known about how this phenotypic variation develops or is maintained—and at which mechanistic level (e.g., genomic, neural, physiological)—within and across populations and species.

Multiple fields of research have embraced the gathering of Big Data, including in biology: genomics (e.g., measuring genome-wide expression in multiple brain regions), physiology (e.g., sensors that monitor cardiovascular activity continuously), behavior (e.g., tracking software to capture nuanced behavioral expression from video), and populations (e.g., social networks, population genetics). But just as Big Data in healthcare does not guarantee improved patient outcomes, unless the right questions can be asked of these biological data, sheer volume does not guarantee understanding. For example, the questions asked in most biological research considers a limited scope of cause and consequence: How does neural signaling affect behavior? How does behavior affect fitness? To understand how and why an individual succeeds requires mapping genomic, developmental, and neural mechanisms to phenotype to fitness, integrative data which are challenging to collect, analyze, and interpret. The current statistical models, for example, are not designed to incorporate the interacting effects of gene expression, development, and the environment given that development a) shapes gene expression, b) the environment within which an individual lives, and c) modulates gene by environment interactions.

The goal of this Pop-Up Institute is to transcend the traditional boundaries of research disciplines to address the causes and consequences of individual and population variation in biology, medicine, and society. It is clear that there are fundamental similarities among the questions being asked by natural, medical, and social scientists, but these similarities are often obscured by differences in language, methodology, and domain knowledge. Results are also applied to solve specialized problems. We propose to bring together experts from the Dell Medical School, the College of Natural Sciences, and the College of Liberal Arts to 1) identify the most promising questions about variation at the individual and population levels, 2) establish a unique and comprehensive research plan to answer these questions, and 3) develop solutions to shared problems that currently limit progress (e.g., analyzing and interpreting Big Data). By aligning similar research efforts across disciplines, this Pop-Up Institute will establish a lasting framework for novel collaborations to conduct groundbreaking, transformative research that has real-world, positive outcomes for humans and society.

Work Plan, Schedule (Fig. 2), and Products / Deliverables

Organization and preparation: Leading up to the Pop-Up Institute, the principal participants will advertise and generate campus-wide excitement for the Institute, in part by inviting one speaker per department (including Integrative Biology, Institute for Cellular and Molecular Biology Statistics and Data Science, Sociology, Nutrition, and the Dell Medical School) who conducts research on the Institute's theme. The principal participants have ample experience organizing symposia, workshops, and working groups of this kind, and they have been successful previously in generating corporate financial support.

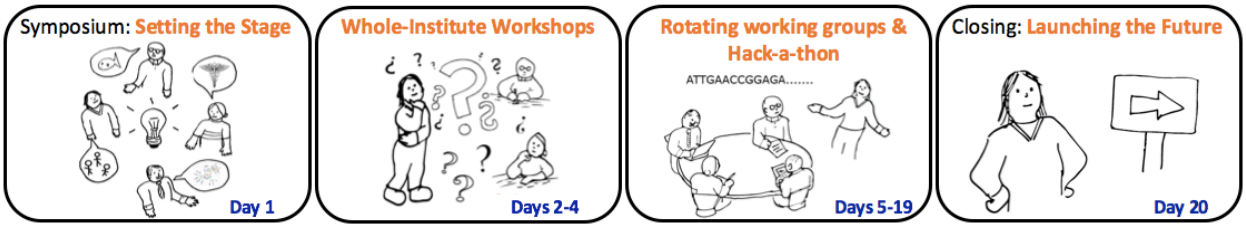

Fig. 2: Pop-Up Institute schedule: 4 weeks of collaborative, interdisciplinary research. Products listed below

Symposium: Setting the Stage: This event will be a special themed version of the Center for Computational Biology and Bioinformatics' annual Big Data in Biology Symposium. It will be open to the greater UT community and will include distinguished speakers from diverse disciplines, all of whom study the causes and consequences of individual and population variation. The symposium will generate awareness and excitement for the work to be completed by the Pop-Up Institute and begin the process of establishing a common knowledge base among Institute participants and attendees from across campus. All of the Institute participants will be present.

Products: In addition to the event itself, the symposium will introduce research questions and areas and build relationships across UT and beyond.

Whole-institute workshops: Following the symposium, all of the Institute participants will attend 2.5 days of workshops. The goal will be to establish a common language, discuss differences and identify similarities in knowledge and methodological and analytical approaches across disciplines, and identify the most promising directions for the field. Working groups will focus on these topics and the most promising research questions. Based on the previous experience of the principal participants, the working groups will include a provocateur, thinkers, and a writer with domain knowledge who can synthesize the discussion into text.

Products: During the workshops, participants will collectively write a high-impact review paper. Additional work will continue during the rotating working groups, as needed. The principal participants have successfully produced impactful review papers within this timeframe in interdisciplinary groups of this kind. The reviews will: establish the fundamental similarities in the diverse research questions across disciplines; define shared terminology; promote the use of evidence-based measures of individual success, including fitness, individual and community/population health, and personal, academic, and professional success; identify significant problems that limit progress, such as integrating and analyzing data across multiple sources and biological levels, Big Data, missing data, and interpreting outliers; propose solutions to these problems; and direct the field forward, including promoting collaborative enterprises.

Rotating working groups: The working groups will each be led by a principal participant who is most enthusiastic about the topic. The groups have the goal of the goal of establishing a unique and comprehensive research plan to answer the questions identified in the opening symposium and workshops and developing solutions to shared problems that currently limit progress. Each working group will have a different mix of participants during the course of the Institute, including participants from different fields or from the same/similar fields. Different members of the working groups will be present for different phases of the work, and there will be dedicated time for cross-talk among groups. One specialized, hands-on working group will include a 'hack-a-thon' in which publicly available datasets (e.g., genomic, transcriptomic, metabolomic) relevant to individual and population variation are analyzed. The goal of these analyses, which relate to health outcomes based on electronic medical records; measures of population health based on

demographic and geographic data; and/or effectiveness of cancer therapies based on genomic and transcriptomic data, will be to gain insight into some of the questions fundamental to the Institute.

Products: These efforts will result in grant proposals to pursue external funding to support innovative centers of excellence and transdisciplinary training programs to build on the progress of the Pop-Up Institute (e.g., National Institutes of Health (NIH) Big Data to Knowledge (BD2K); NIH: Genetics of Rare diseases; National Science Foundation Dimensions of Biodiversity). The meta-analyses conducted during the hack-a-thon will provide preliminary data for these grants. More generally, the rotating nature of these working groups will provide training opportunities for researchers, including interactions and cross-training between computational and non-computational researchers. The hack-a-thon model of analyzing publicly available datasets will also serve as a case study for how this kind of collaboration can proceed beyond the Pop-Up Institute and be applied to other themes.

Launching the Future: The goal of this closing event will be to highlight the accomplishments of the Institute and look forward to the future of the field. The work conducted at the Pop-Up Institute will 'launch the future' through the review papers that will inform the field and promote interdisciplinary, integrative research; grant proposals that can fund future research; and recruitment of faculty qualified to work in areas identified during the Institute to be critically important. All of the Institute participants will attend.

Products: Principal participants will focus on the future through strategic planning, including identifying research areas and candidates for joint recruitment across department and colleges.

Participants: The balance of small group work and full group events will make it feasible for even busy researchers to coordinate with team members and maintain outside commitments. In addition to the core group of principal participants, there is an extended group of expert researchers from across UT whose input and work would greatly contribute to the Institute. Although this list is not exhaustive, the following researchers work on topics related to individual and population variation: *College of Natural Sciences:* Drs. Lauren Meyers (Integrative Biology/Statistics and Data Science), Stephen Russell (Human Development and Family Sciences), Inderjit Dillon (Computer Science), and Pradeep Ravikumar (Computer Science); *Dell Medical School*: Esther Melamed (Neurology); *College of Liberal Arts*: Jennifer Beer (Psychology). From *outside of UT*: Suzanne Alonzo (University of California, Santa Cruz), Nicholas Christakis (Yale), Dustin Rubenstein (Columbia University), Al Hero (University of Michigan), Geoffrey Ginsberg (Duke University), and Yoav Gilad (University of Chicago). Finally, from *Austin-based healthcare*: Bill Rice (CPRIT and St. David's Healthcare) and Tate Erlinger (Seton Health Alliance).

Training and professional development will be central to the Pop-Up Institute by selecting 6 graduate students from programs across colleges, 5 graduate fellows supported by a new NIH BD2K training grant (of which principal participant Daniels is the PD), and 3 postdocs. Trainees will help organize, contribute intellectually, facilitate manuscript and grant writing, participate in the hack-a-thon, and engage the public through multiple media platforms.

Additional Products / Deliverables: Video and media— Multiple media platforms will be used to document and share the progress of each phase of the Institute. A short, high-quality video will also be produced to attract donors, as well as faculty and graduate student recruits.