The focus of my research is on developing computational methods that accelerate the clarity and utility of omics data in biomedical science. I have used this focus to link genetic and molecular variation to phenotype in both natural and engineered cellular systems and approach these topics through the lens of computational biology, machine learning and advanced data integration. A growing body of work in the biomedical sciences generates and analyzes omics data; my work contributes to these efforts by focusing on trans-omics: the integration of different omics data types to bring mechanistic insights to the multi-scale nature of cellular processes.

My research program will develop and apply trans-omics methods, rooted in the basic sciences, to advance genomic medicine and make biopharmaceutical development more efficient. To achieve this, I will leverage my prior work, which borrows strength across genomics, transcriptomics, ribosome profiling, proteomics, structural genomics, metabolomics and phenotype variability data. Integration of multi-omics data, such as these, has made it possible to bridge the genotype-phenotype gap and tie cellular processes together mechanistically. For example, my research has uncovered striking ties between variants that co-occur in three-dimensional protein space and their downstream, molecular and clinical phenotypes[1]. It has also revealed a predictive structure in the way cells respond to drugs[1,2] and genetic engineering[3,4]. Broadly speaking, trans-omics provides a robust platform to better understand how cells achieve regulation at multiple scales of complexity and which genetic and molecular variants influence this process.

My early career research was amongst the first to discover the unique challenges of "piecing" together multiple layers of biology using multiple omics data types. These efforts led to 30 publications, many are in top-tier journals, like *Nature Biotechnology*, *Nature Communications*, and *JACS*, and a wide array of invited talks at venues like Protein Engineering Global Summit and the INBRE and NCGR Symposium on Integrative Omics . My research has been funded by NCI, Swiss National Science Foundation, NVIDIA and Novo Nordisk Foundation.

The next phase of my research continues to advance trans-omics within the biomedical sciences by placing emphasis on improving clinical diagnosis options and making the development of therapeutic proteins more efficient. To this end, I am currently conducting translational research with clinical oncology applications through the Altman Clinical and Translational Research Institute and the NCI-designated Cancer Therapeutics Training program. For advances in high-throughput omics technologies to empower patients, integrative methods must be developed to augment mutation panels and address bottlenecks in the production of biologics. To address specific aspects of clinical and drug development research, I collaborate with physician-scientists at Moores Cancer Center and La Jolla Institute for Immunology, who provide genomic and transcriptomic data for cancer patients in clinical trials. I also work closely with drug pharmaceutical companies like Celgene to apply machine learning and data integration of large-scale drug-treated profiles to further drug development pipelines. In this way, I facilitate trans-omics for translational research to advance a genomic-based cancer treatment vision.

FUTURE RESEARCH PLAN

My future line of trans-omics based research will address the quality and efficiency of drug action, production and response in cancer research through three specific aims: 1) Variant interpretation, 2) Therapeutic design and 3) Drug response features. This program will advance the integration and interpretation of diverse data types, ranging from structural biology and imaging data to single cell and bulk multi-omics data.

1. Integrative Systems for Variant Interpretation. The majority of cancer variants are unspecified, even in common oncogenes. Identifying disease-relevant variants would improve clinical diagnosis and diminish the futility of drug treatments.

There is compelling evidence which suggests that significant driver cancer mutations tend to co-occur in functionally-relevant clusters within protein three-dimensional space[1,2,5]. To expand on this concept, I will develop multi-scale modeling methods to identify variant clusters that associate with specific molecular and clinical phenotypes. Specifically, these methods will: 1) Delineate variant-induced changes in the structure and function of proteins, 2) Characterize perturbations in biomolecular networks resulting from changes in protein functionality, 3) Infer and predict how network dysregulation influences changes in cell physiology. To address this multi-layer challenge, I will expand upon a method, VTerra, which is currently under development. VTerra performs omics data-driven, knowledge-based inference modeling to identify variants that functionally converge with specific cellular states/phenotypes (e.g., up/down-regulated pathways, cell line drug sensitivity, or changes in cell viability upon genetic knock-down of key pathway genes). It does this by integrating publicly available multi-omic data from the Cancer Cell Line Encyclopedia (CCLE), The Cancer Genome Atlas (TCGA), Project Achilles and the Protein Data Bank. Expanding on this method, I will rank-order variant clusters based on their potential impact on protein functionality by integrating structural biology and first-principles modeling. Second, I will characterize molecular pathway changes within cohorts with common variant clusters. Web-based visualization tools, similar to those I have previously developed[1], will be designed to interactively explore multi-omics and single cell data and the molecular and temporal patterns within individual samples or cohorts. Third, I will link protein and network-level changes to specific phenotypes by implementing multivariate feature selection and inference modeling as well as tools that process and interpret single-cell sequencing data for the characterization of cell population heterogeneity. Through this stream, my work will provide a multi-scale view of variant interpretation to determine truly actionable mutations and advance drug efficiency.

2. Omics-guided Therapeutic Design. The genetic manipulation of host organisms to produce recombinant proteins requires control over multiple layers of biology to ensure high yield, stability and quality of therapeutic extracts.

To address this multi-layer challenge, I will develop knowledge-based approaches to characterize expression, protein translation and post-translational protein modification of heterologous peptides. First, machine learning and feature selection methods will be developed to determine critical quality attributes that explain discrepancies in expression and solubility. Expanding on prior work,[4] I will leverage data generated by the Human Protein Atlas, in which 45,000 human-derived fragments are expressed in *E. coli* under uniform conditions to identify differentiating factors between human and *E. coli* proteins that explain high or low expression/solubility. Second, I will focus on the proper translation of recombinant peptides. Building upon prior work,[6,7] I will identify features that influence cotranslational folding of proteins, such as ribosome pausing and codon usage. I will collect and generate transcriptomics, ribosome profiling and proteomics data in human cell lines and *Chinese Hamster Ovary* (CHO) engineered strains producing human proteins, like Interferon beta-1a, Glucagon-like Peptide 1 and Immunoglobulin G. Methods will be developed and deployed to identify key molecular attributes that differentiate endogenous and heterologous protein translation. Third, I will systematically characterize post-translational modifications by expanding on a recent study[8] and by leveraging genome-wide single and double knockdown data.[9] Through this stream, my work will provide insights into the "rules" governing interoperability between engineered and host systems, to advance the quality and efficiency of peptide production.

3. Predictive Features in Drug Response. Chromosomal aberrations dictate a cancer cell's genetic dependencies and molecular state through oncogene amplification on extra-chromosomal DNA (eccDNA). Understanding the features and mechanisms underlying pathway-driven drug response and evasion via eccDNA will aid in future efforts to combat chemotherapeutic drug resistance.

To address this challenge, I will expand upon prior work that identifies and visualizes pathway-driven cellular responses to stress, drug treatment and genetic manipulation[1–3,8,10]. I will apply these approaches to systems inherently disrupted by eccDNA. Specifically, this aim will: 1) Investigate how eccDNA influences pathway-driven cancers and 2) Determine the dynamics of eccDNA during drug treatment. First, I will expand upon preliminary work, which has uncovered a systematic method for predicting the presence of eccDNA in cancer cell lines. Presence of eccDNA can be predicted from publically-available copy number variation, RNAseq and fusion data. Many of these predictions have been validated via publically-available karyotype data as well as Fluorescence in situ hybridization (FISH). Using FISH, I quantify the presence of eccDNA and locate the amplified genes in eccDNA or homogeneous staining regions (HSRs). To link eccDNA-mediated cell states with pathway-driven cancer programs, I will use advanced data integration methods to interlace cytogenetic data (FISH images) with high-throughput omics data (single cell sequencing, Pacbio sequencing, transcriptomics, and proteomics) and molecular and clinical phenotype data. This will pinpoint key genomic and molecular features that coordinate eccDNA-mediated states. Second, I will perform adaptive laboratory evolution on eccDNA-presenting cancer cell lines by administering drugs that target genes localized on eccDNA amplicons or HSRs. Comprehensive molecular imaging, together with single-cell and Pacbio sequencing will be used to track dynamic changes in eccDNA counts and potential translocations to chromosomes. Through this stream, my work will provide insights into how cells with chromosomal aberrations like eccDNA respond and evade therapies to advance our ability to design and administer more effective treatments.

Obtaining Extramural Funding. My research program is in line with the goals of NIH (NCI, NHGRI, and NIGMS). Aspects of what I propose also apply to basic science funding agencies, such as NSF. I also plan to apply for funding through private agencies, like the Chan-Zuckerberg Initiative.

Conclusion. Through my research, I innovate trans-omics methods for cancer research by advancing the quality and efficiency of drug action, production and response. What makes my research portfolio unique is how it develops integrative computational and multi-omic data systems and then uses them to address: 1) variant interpretation to increase drug efficiency, 2) therapeutic design to increase biopharmaceutical production efficiency, 3) drug response to improve treatment efficiency. I am passionate about working toward a world where everyone - not just those with common variants - has access to precise and accurate targeted cancer therapies.

REFERENCES

1. Brunk, E. *et al. Nature Biotechnology* (2018) [Pubmed 29457794]
2. Mih, N., *et al. PLoS Comput. Biol.* (2016) [2nd and Co-corresponding Author] [Pubmed 27467583]
3. Brunk, E. *et al. Cell Systems* (2016) [Pubmed 27211860]
4. Sastry, A. *et al. Bioinformatics* (2017) [Last and Co-corresponding Author] [Pubmed 28398465]
5. Brunk, E. *et al. BMC Syst. Biol.* (2016) [Pubmed 26969117]
6. Ebrahim*, A., Brunk*, E., *et al. Nat. Communications* (2016) [Pubmed 27782110]
7. Latif, H. *et al. Biotechniques.* (2015) [Pubmed 26054770]
8. Brunk, E. *et al. Proc. Natl. Acad. Sci. U. S. A.* (2018) [Pubmed 30301795]
9. Monk*, J. M., Llyod*, C., Brunk*, E., *et al. Nat. Biotechnology* (2017) [Pubmed 29020004]
10. Mih, N. *et al. Bioinformatics* (2017) [Pubmed 29444205]