**SIGNIFICANCE**

The complexity of genetic variation associated with cancer demands rigorous approaches that can study the effects of different types of variants. A wealth of annotated data are available due to advances in sequencing technologies and efforts by consortia like ENCODE and 1000 Genomes, what engenders the need for comprehensive computational, mathematical, and experimental methods and analyses. We will leverage our experience to develop a pipeline for prioritizing variants in cancer and assessing the disruption it incurs at multiple scales.

**Evaluation of Coding and Noncoding Variants**

Both coding and noncoding variants may vary in degree of impact on cancer development or protein formation and function. Historically, there has been a bias towards studying coding variants due to the functional significance of protein coding regions. However, as noncoding alterations constitute the majority of disease associated variants [1], further study of non-coding regions may also be critical to a better understanding of cancer biology. Accordingly, we will consider a combination of coding and noncoding variants.

**Variant Prioritization at Multiple Scales**

Effects of numerous genetic variants transcend the molecular level and propagate into the phenotype. However, the extents to which variant effects take place at the levels of molecular activity, cellular phenotype, and organismal phenotype are still unclear. The assumption that the impact of variants is consistent at all three levels needs to be examined. For that purpose, we plan to use a variety of pipelines, cell-based assays, CRISPR-Cas9-based methods, and realistic prostate organoids. We will also study the relationship between different mutations and tumor growth and invasiveness (see Aims 2-3).

**Variant Prioritization in Aggressive and Nonaggressive Cancers**

The integration of numerous clinical and genomic datasets has shown a promising ability to study different diseases including prostate cancer [2]. We think that studying the functional impact of variants through integrated, heterogeneous data will provide deeper insights into the genetic patterns underlying cancer. That might also help in (1) comparatively understanding the impact of variants in both aggressive and nonaggressive cancers and (2) assessing indolent and lethal variants using survival data.

**Coding Variant Prioritization through Protein Interactome Networks**

Many proteins carry out diverse functions through interacting with other proteins [3]. Recent studies have been conducted on genetic coding mutations in the context of the human interactome network [4-7], where on average, a functionally active protein interacts with >5 other protein partners. We will leverage our experience to deploy a novel approach that systematically uses several agnostic functional assays in parallel.This approach serves as a paradigm to prioritize coding variants and provides important insights into mutation mechanisms of interaction from a systems biology perspective.

**Studying Variant Functional Impact on Tumor Growth**

Recent studies \cite{26456849} \cite{23388632} suggest that certain passenger mutations described as “mini-drivers” may have a weak effect on tumor cell fitness and in turn promote or inhibit tumor growth. From a tumor fitness perspective, three categories can thus emerge: positively-selected driver variants, neutrally-selected passenger variants, and negatively-selected mini-driver variants. As there is a functional impact associated with any positively or negatively selected variant, we think that studying the interplaying effects of both weak positive and negative selection variants may also reveal valuable insights into tumor growth patterns.

**Application to prostate cancer**

In our work, we will focus on prostate cancer. Significant efforts have been made to study genetic and environmental causes of this cancer type, but major leaps forward are still needed to develop a more complete etiology of this disease that affects XXX million Men worldwide. Along with other major factors associated with prostate cancer such as the hormonal action of androgens and estrogens [8], more than 70 genetic susceptibility variants have been identified [9]. Suspected loci are continuously being discovered using GWAS studies [10] and genotyping arrays [11]. Such variants increase the predictability of the disease and have been associated with altering the expression levels of several genes.

In addition to prioritizing susceptible cancer variants, we will investigate the following related questions: (1) are noncoding variants as deleterious as coding ones w.r.t. prostate cancer incidence?, (2) do deleterious variants lead to the emergence of more deleterious ones in tumor cells?, (3) is there a fitness benefit for heterozygous v.s. homozygous mutation in tumor suppressor genes?, and (4) is there a relationship between mutations that lead to loss of heterozygosity in tumor cells?

**Innovation**

Our mathematical model, its multi-tiered cutting-edge biological validation in concert and each individually, and the real-time Bayesian update of the former with the latter are fresh, exciting contributions to the field.

* Mathematical model: We will combine the best insights from diverse variant prioritization approaches: including recurrence-based approaches, biochemical/biophysical approaches, and evolutionary approaches to rank the impact of both coding and noncoding variants.
* Bayesian update: The plan is for data from the biological validation steps to be rolled out in real time, so that our mathematical model will learn from early validation experiments to update parameters for variant prioritizations for subsequent validation experiments
* Multi-tiered validation: Our planned mix of high-throughput and low-throughput validation experiments covers both breadth and quality, both molecular and cellular endophenotypes, and assays both coding and noncoding mutation impacts.
* eSTARR-seq: this unique barcoding approach from Yu Lab allows direct quantification of enhancer activity, with 40-fold increase in sequencing efficiency compared with traditional STARR-seq
* InPOINT: this unique technology from Yu Lab directly examines the biochemical consequences of coding variants on protein stability and interactions
* CRISPR: This genomic editing breakthrough technology from Levchenko lab can build a cellular variants impact evaluation model to introduce targeted mutation in coding and noncoding regions from normal prostate cell lines, which will grow in prostate organoid to investigate tumor progression effect.
* Organoid technology: This technique, successfully deployed in Rubin lab differs from traditional cell culture by maintaining cancer cells in three-dimensional (3D) cultures. Benign and cancer cells that are grown in 3D retain cell-cell and cell-matrix interactions that more closely resemble those of the original tumor compared to cells grown in two dimensions on plastic