Cancer genomics has revealed that there are often thousands of mutations per genome, but only a small fraction of them are in coding regions. Yet, almost all of the known driver mutations in cancer are in coding regions. Many approaches have been developed to identify drivers and prioritize non-coding mutations. But currently, the relative impact of coding and non-coding mutations, especially how to prioritize them uniformly, is largely unknown.

In this proposal, we will develop systems-level, computational models to prioritize and rank non-coding and coding mutations in similar terms. These models will rank the impact of mutations causing cancer in terms of their underlying genomic (at the nucleotide-level) and epigenetic alterations and recurrence in cancer cohorts. We will then experimentally assay the phenotypes produced by these mutations on different scales: molecular activity and cellular phenotypes. Our proposed research will produce as products: a data resource of prioritized cancer mutations and iteratively novel and refined models for prioritizing them.

In addition to the relative impact between coding and non-coding, our proposal will allow us to ascertain the relative impact of mutations on multiple scales. Is it the case that a mutation prioritized to give a strong impact in terms of its molecular "endo-phenotype" will also have a strong effect on cellular phenotype, and this, in turn, will be associated with the organismal phenotypes of contracting cancer?

We will focus our analysis on prostate cancer, a disease that our group has studied extensively in the past. In addition to our genome-wide mutation prioritization, we will conduct focused investigations related to the sub-networks involving TP53 and RB, two genes that are particularly important for a subtype of prostate cancer, Neuroendocrine Prostate Cancer (NEPC).

**AIM 1 - Computational prioritization of coding and non-coding somatic mutations.** Here, we will computationally prioritize mutations on a number of scales. First, we will look for mutations that score highly in terms of putative molecular functional impact. This will be ascertained by features of the genomic sequence and epigenetic context, including whether or not they break motifs, create loss of function events, preserve interfaces in protein structures, are associated with allelic activity, or hit genes or regulators that are highly conserved or are at the center of networks. We will then take the orthogonal perspective of finding mutations that recur in large cancer cohorts and thus are under positive selection to cause cancer in humans. We do this by developing a whole-genome burdening formalism integrating many known genomic covariates of mutational rate (e.g., replication timing). We will also utilize deep learning algorithms to prioritize variants and integrate all scores predicted by algorithmic and mathematical methods into a combined prioritization. Finally, from looking at the results of the experiments in Aim 2 and 3, we will iteratively update our model over the course of the grant to make it more accurate, with the goal of producing a practical and usable prioritization scheme.

**AIM 2 - High-throughput *in vitro* quantification of molecular phenotypes of ~1500 non-coding and ~1000 coding mutations.** Based on the prioritization above, we will select ~200 coding and ~300 non-coding mutations per year and subject them to a number of high-throughput in vitro assays in the RWPE-1 cell line (a match to normal prostate tissue) to look at their molecular activity. In total, we will examine ~1000 coding and ~1500 non-coding mutations. We will take advantage of our Clone-seq pipeline to generate these mutant clones on a large-scale. Furthermore, we will assay the non-coding mutations using eSTARR-Seq and Promoter-seq to quantify their effect on enhancer and promoter activities. We will also assay the coding mutations using our high-throughput protein-protein interactome-screening methodology, INtegrated PrOtein INteractome perTurbation screening (InPOINT). This pipeline combines five different functional assays to examine experimentally the impact of hundreds of coding variants on protein stability and specific protein-protein interactions. From this we will be able to rank the pool of ~2500 variants in terms of molecular activity and then pass this list to the next aim.

**AIM 3 - Medium-throughput *in vivo* evaluation of cellular phenotypes, plus select detailed validation in prostate organoids.** We will further select 120 high-impact mutations, from those top-ranked in aim 2, for investigation of two potential cellular phenotypes related to cancer: growth and cell invasion (which is related to metastasis). The mutations will be introduced into RWPE-1 cells through CRISPR-Cas9 knockin mutagenesis. We then will select the top 10 coding and non-coding mutations and evaluate them in a more realistic context – organoids derived from normal prostate samples. We will further investigate the mechanisms through which mutations lead to cancer, focusing particularly on P53 and RB1 subnetworks, which are known to be key to NEPC. We will test alterations in transcript levels (for non-coding mutations), protein stability and interactions (for coding mutations), and selected functional and phenotypic assays in gene-edited and isogenic control organoids. Our large-scale phenotyping efforts will involve substantial interactions with two NCI-sponsored centers that are headed by co-investigators of this grant, which will allow us to probe mutational impact in progressively more complex contexts: one is the U54 Cancer Systems Biology Center at Yale (led by A Levchenko), and the other is the Prostate Cancer SPORE Center at Cornell (led by M Rubin).