**TF identification**

to look for gain-of-binding site events, we will perform a SNVs analysis in regulatory elements proceeding as in [35]. In summary: we will query the COSMIC Cancer Gene Census for TFs (termed CGC-TFs). For these factors, we will search the JASPAR and TRANSFAC motif databases for motifs that are bound by the cognate CGC-TFs to identify motif position weight matrices (PWMs). For each of the motif PWMs we will construct a position-specific scoring matrix (PSSM) and determine the threshold PSSM value false-positive rate. We will then analyze each SNV from the filtered COSMIC noncoding variant set that occurred in a regulatory element for its ability to modulate the motif PSSM score. For each SNV, we will generate in silico wildtype and mutant alleles, using the latest version of the human reference genome hg38, as the reference (wildtype) allele. Each pair of alleles will be scored against each CGC-TF PSSM to obtain a log-odds ratio score compared to a background of genomic nucleotide frequencies and only scores passing the CGC-TF-specific threshold will be retained. To enrich our dataset for events with high effect size, we will keep only pairs of CGC-TF motif scores where at least one score (wildtype or mutant) is log odds score over background ≥ 2. We will determine the delta value for each pair of PSSM scores by subtracting the mutant allele score from the wildtype score. The resulting distribution will reveal pairs of wildtype-mutant alleles from whole genome-sequenced samples with potential gain-of-binding site events, in which the mutant allele score is higher than the wildtype allele score for a particular CGC-TF. Potential loss-of-binding events will be obtained through wildtype allele score greater than the mutant allele score.

**Proteomics (incorporated into MR Analysis):**

MRs are nodes in the network (genes, proteins, or complexes) that regulate, simultaneously, the TFs in our network. We will identify MRs by mapping pathways in which the identified TFs participate. To integrate proteomic data of the studied tumor, we will use the Master Regulators with Context Genes with Weights pipeline [29900117] with the manually curated TRANSPATH database [47]. The modified pipeline will use the proteomics expression data as the context to preferentially choose pathways enriched with abundant protein levels.

**Addition/Removal of Mutated Pathways**

Mutations are a critical factor in the development of cancer. We will use the Ensembl Variant Effect Predictor (VEP) to provide information regarding the location and consequence of each SNV [27268795]. Mutations that lie in protein coding regions will be filtered to include only those that modify the protein amino acid sequence (missense, splice-site variation, frameshift, deletion).

A literature curation of the gene variants that cause protein transcript alterations will be done to explore their impact on TNBC tumorigenesis.If a gene variant has gain of function activities that affect a specific TNBC pathway, the interaction and pathway will be added to the network manually as appropriate, since they are not included in signaling databases.

\*\*\*For those mutations without relevant TNBC literature, we will conduct in-silico perturbation experiments via Ingenuity Molecule Activity Predictor. By gathering mutant protein to downstream target relationships, we will compare observed target gene expression levels to the output of knock-ins and knockouts to provide insight to the impact of the mutation on the protein’s function.\*\*

**Modification of Upstream Regulation due to Kinase Mutations**

Mutations that modify the protein transcript may cause profound kinase signaling changes within a cell. We will use the Ensembl VEP ProteinSeqs plugin to obtain reference (wild-type) and mutated protein transcripts. The ReKINect software will predict and classify kinase modifying mutations in the mutated transcripts [26388441]. ReKINect employs a database of all known human kinase domains, 111 SH2 domains, and 149,838 phosphorylation sites and will predict the functional impact of the modified protein transcript. If mutations affecting protein kinase function are identified, KINSpect, using the ReKINect database, will be used to provide the downstream targets of the mutated kinase [26388442]. In order to determine kinase signaling changes upstream of mutated proteins, we will use KinomeXplorer to identify atypical phosphorylating species for the mutant transcript compared to the wild type [24874572].

These kinase signaling modifications will be added to the overall network.