RELATED WORK:

This interaction-based approach to cancer biology has recently been adopted in research: a combination of biological networks and summary statistics that quantify each gene's association with cancer has aided in the discovery of novel cancer driver genes (e.g. Horn et al., 2018; Leiserson et al., 2015; Reyna et al., 2018). Nodes represent genes, and edges represent relationships between adjacent genes in those networks. There are a plethora of biological networks that are derived from various sources and span various scales. Co-expression networks (Willsey et al., 2013), co-dependency networks (e.g. AchillesNet; Li et al., 2018), and co-evolution networks (Niu et al., 2018) are all good examples. Co-expression networks (Willsey et al., 2013), co-dependency networks (e.g., AchillesNet; Li et al., 2018), co-evolution networks (Niu et al., 2017), metabolic pathways (Kanehisa et al., 2017), and protein–protein interaction (PPI) networks are some of the most well-known instances (Lage et al., 2007; Li et al., 2017; Szklarczyk et al., 2019). PPI networks, in particular, are an intriguing representation of gene interactions since they often integrate data from various data sources, tissues, and molecular processes at various scales. Those PPI networks, on the other hand, are far from complete, and our understanding of them is skewed toward well-studied genes (Horn et al., 2018). Superimposing scores on the nodes is a popular starting point for methods that use networks to describe molecular relationships. These scores reflect the gene's marginal association with the disease of interest. The MutSig P-value (Lawrence et al., 2014) is a meta-P-value that describes whether there is a statistically significant difference in I mutational burden, (ii) mutation clustering, and (iii) functional effect of mutations in a gene between healthy and cancer tissues. A variety of methods have been developed to evaluate gene scores in conjunction with network information in order to distinguish altered gene subnetworks within the original network. They can be divided into clustering methods (e.g., Jia et al., 2011; Rossin et al., 2011) and methods that use network diffusion or network propagation (e.g., Cowen et al., 2017; Reyna et al., 2018) to detect altered subnetworks. Both approaches are based on the idea that genes that influence the same phenotype interact in a network. Network propagation approaches, in particular, have been effective in identifying novel cancer driver genes (Hristov et al., 2020; Leiserson et al., 2015; Reyna et al., 2018; Ruffalo et al., 2015; Vandin et al., 2011, 2012). Network propagation methods, on the other hand, take advantage of the flow of information between genes along paths, with the longer the paths, the more information is diluted. This makes detecting cancer genes that aren't on short paths between other cancer genes more difficult. NetSig is another solution that has worked well that does not rely on this assumption. It detects cancer genes solely based on a network's local neighbourhood of genes (Horn et al., 2018). While the aforementioned methods have shown great promise in a variety of biomedical applications (Cowen et al., 2017), including the discovery of novel cancer genes, they take an unsupervised approach to gene identification. However, there is knowledge about well-known cancer genes (e.g., Sondka et al., 2018), a significant layer of additional information that has only been used in a few methods for the prediction of cancer driver genes, namely Bayesian modelling (Sanchez-Garcia et al., 2014) and unsupervised network propagation, to the best of our knowledge (Hristov et al., 2020). We present a novel method for supervised classification of cancer genes, based on cancer gene annotations from the COSMIC database's Cancer Gene Census (CGC) (Sondka et al., 2018).