# InCRIMP: a versatile computational model for the integrative analysis of multi-omics data

Adrian Salavaty ( @mania_abbas; asalavaty@ccia.org.au) and Mark Pinese
Children's Cancer Institute, Lowy Cancer Centre, UNSW Sydney, Kensington, NSW, Australia

The code will be deployed on the GitHub: github.com/asalavaty/InCRIMP

## Introduction

Despite recent advancements in precision medicine, for most patients a targeted treatment cannot be identified. High-throughput studies have aimed to address our imperfect understanding of cancer biology through unbiased discovery of cancer risk and driver genes based on single omics profiles. As genes work in concert to drive cancer, we hypothesise that an **integrative** approach that considers **multiple** molecular data, in the context of multi-gene pathways, will yield the best understanding of cancer biology. Here we present InCRIMP (**In**tegratomic **C**ancer **R**isk **I**nfluential **M**odule **P**rioritization) which integrates multiple molecular measurements and state-of-the-art network analysis to achieve comprehensive molecular dissection of cancer cohorts, and unlock the true potential of molecular profiling to understand the risk genes and drivers of cancer.
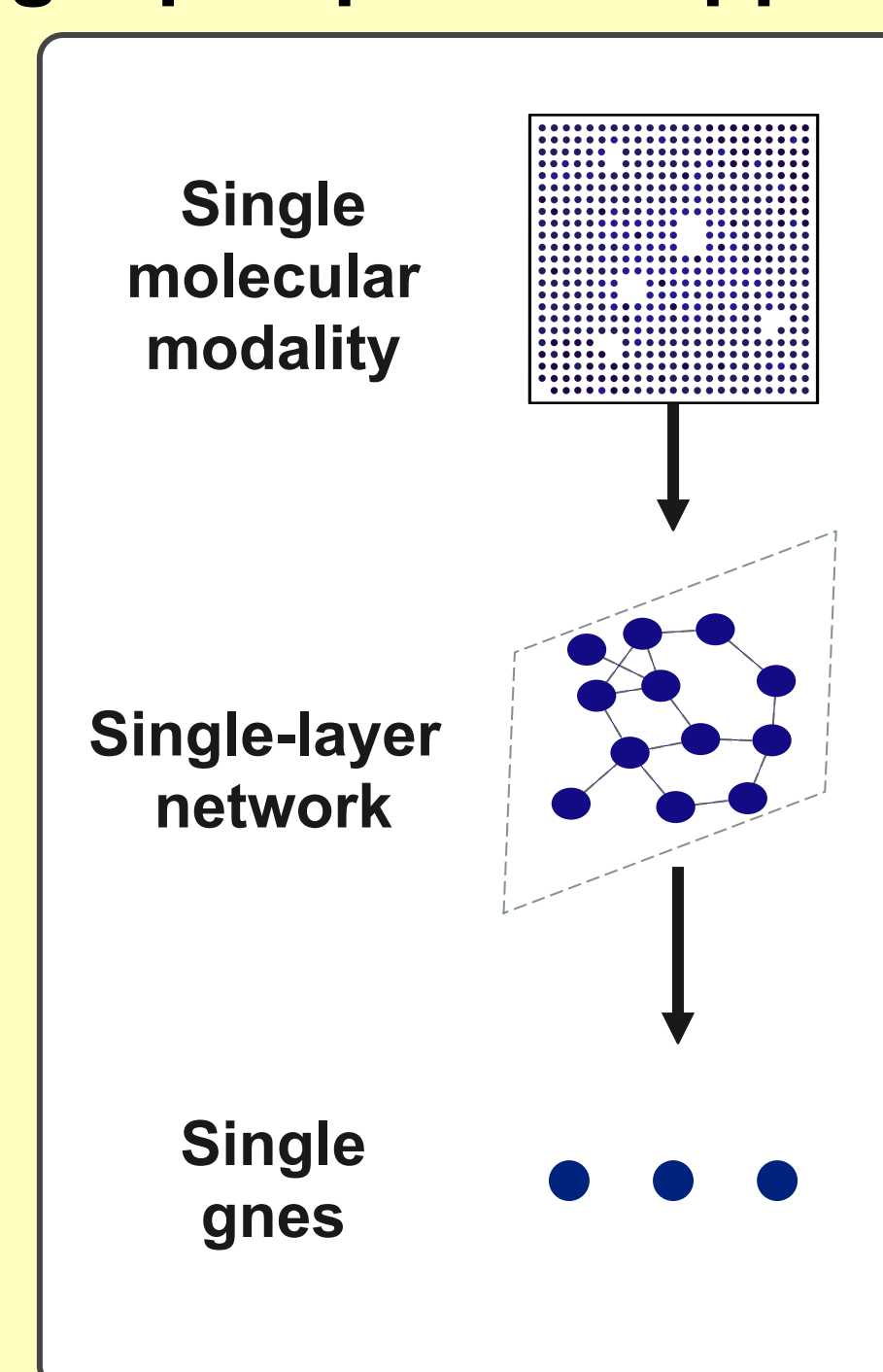


## Methods

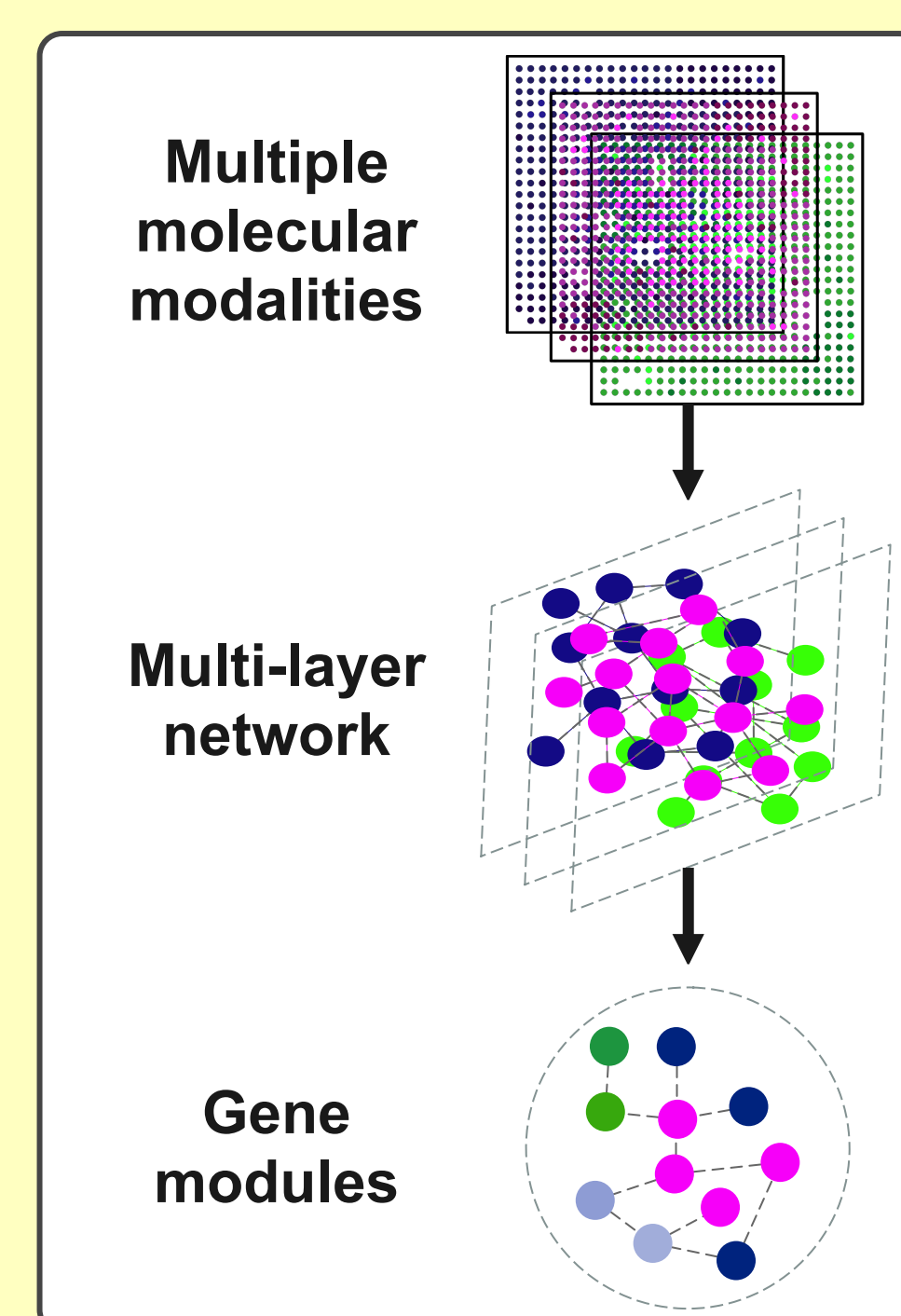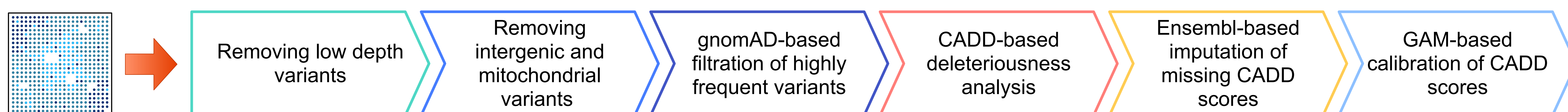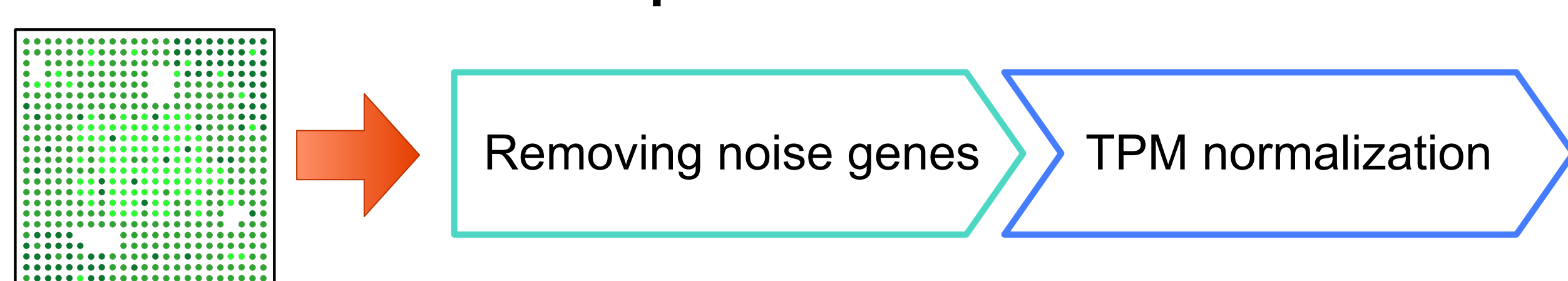**Pre-processing, filtration, and normalization of raw data**. SNV data were filtered and a deleteriousness score was assigned to each SNV based on CADD scores. The transcriptomic data was TPM-normalized after filtering out the noise genes. The PPI data was obtained from the STRING database.

### Germline and Somatic SNV Data



Removing low depth variants → Removing intergenic and mitochondrial variants → gnomAD-based filtration of highly frequent variants → CADD-based deleteriousness analysis → Ensembl-based imputation of missing CADD scores → GAM-based calibration of CADD scores

### Cancer and Normal Expression Data

Removing noise genes → TPM normalization

### PPIs

STRING

**Association Analysis and Un-weighted Multi-layer Network Reconstruction**. Two multi-layer networks were reconstructed; a three-layer multi-omics Risk network and a a three-layer multi-omics driver network.
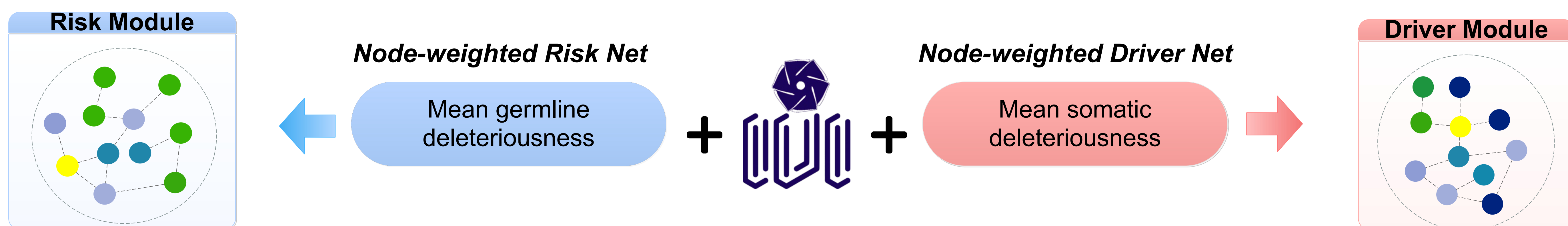
### Risk Network

| 01 | Germline Gene co-Deleteriousness |
| 02 | Normal Gene co-Expression |
| 03 | Protein-Protein Interaction (PPI) |

### Driver Network

| 01 | Somatic Gene co-Deleteriousness |
| 02 | Cancer Gene co-Expression |
| 03 | Protein-Protein Interaction (PPI) |

**Functional module identification**. Initially primitive risk and driver scores were calculated and assigned as the node weights. Then, functional modules were identified based on the Leiden algorithm. Lastly, final node scores were calculated by integrating the primitive scores and node mean neighborhood scores.



**Risk Module** ← *Node-weighted Risk Net* (Mean germline deleteriousness) + + *Node-weighted Driver Net* (Mean somatic deleteriousness) → **Driver Module**

## Results  *(An example based on Pediatric Lymphoma data)*



**Potential mechanism of involvement in lymphoma development**

- Apoptosis and Cell Cycle Regulation
- Cell growth, proliferation, and survival
- Cell migration and invasion
- Modulation of several signaling pathways
- No known mechanism of involvement

### Association with Lymphoma

| Genes | Association |
|---|---|
| IGF1, IGF1R, PRKCD, BCAR1, IRS1, PTPN1, NF2 | Linked to lymphoma development |
| PIK3CA, TP53 | Lymphoma driver and potential drug target |
| RCC1L, ABCF2, USP5, AAMP | No known link to lymphoma |

### First-ranked Driver Module of Pediatric Lymphoma

BCAR1, PTPN1, IRS1, PIK3CA, IGF1R, IGF1, PRKCD
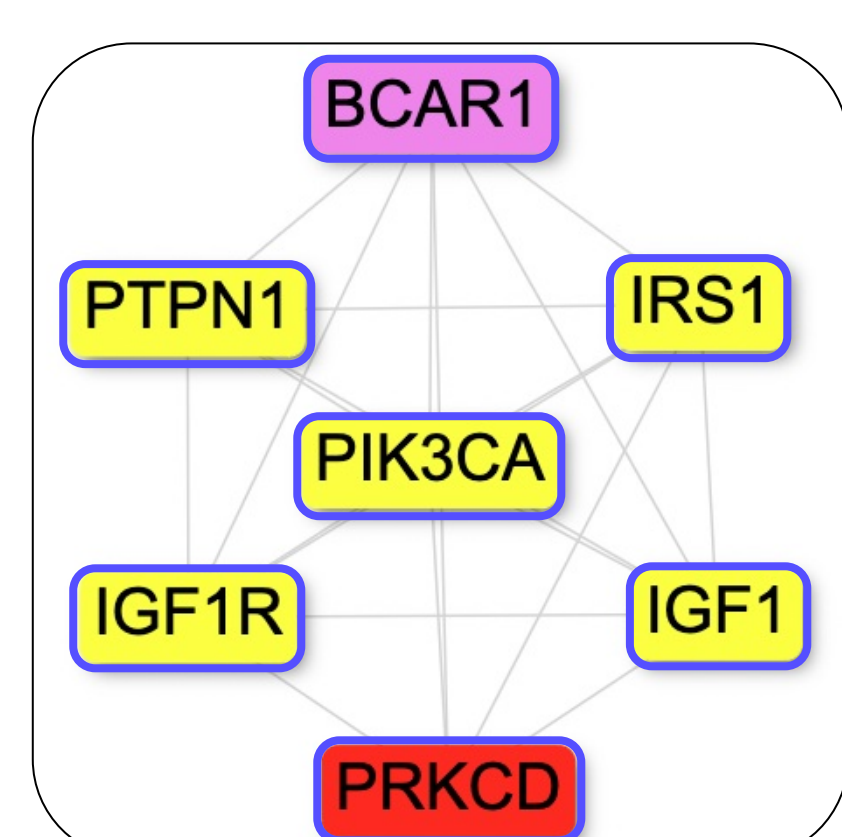
### First-ranked Risk Module of Pediatric Lymphoma

ABCF2, AAMP, RCC1L, USP5, TP53, NF2
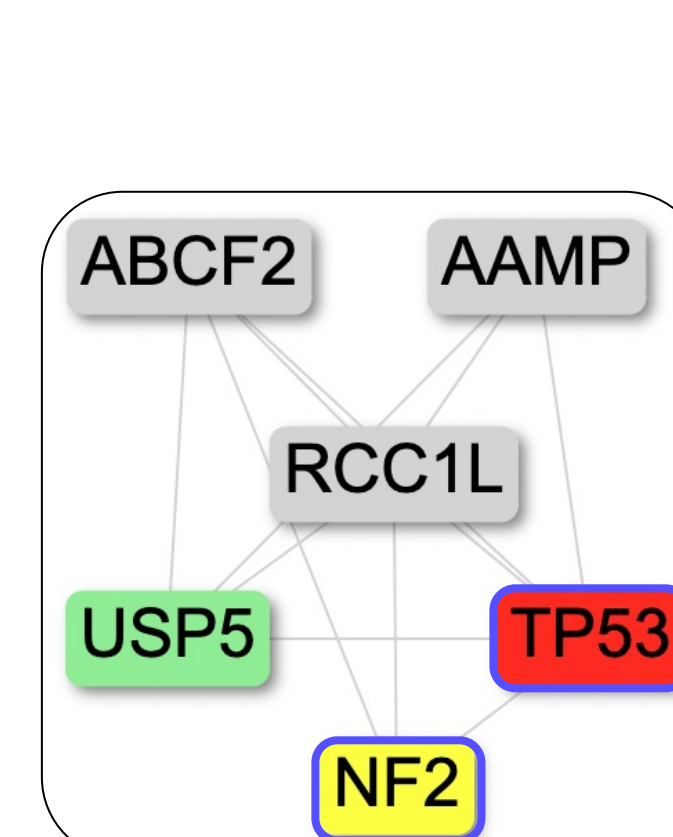
### Top 5 Candidate Single Driver and Risk Genes of Pediatric Lymphoma

| Rank | Risk | Driver |
|---|---|---|
| 1 | ATM | TP53 |
| 2 | TP53 | PIK3CA |
| 3 | CDK1 | ATM |
| 4 | NOL6 | CDK1 |
| 5 | CREBBP | SRC |

## Conclusion

InCRIMP has integrated multiple molecular data types in cancer to recapitulate known cancer biology, and drive the discovery of new cancer driver and risk gene networks and modules.