Model-free estimation of driver interactions across cancers

Yuelin Yao[1, 2], Abel Jansma[2, 3], Luigi Del Debbio[3], Sjoerd Beentjes[2, 4], Chris Ponting[2], Ava Khamseh[1, 2, 3]

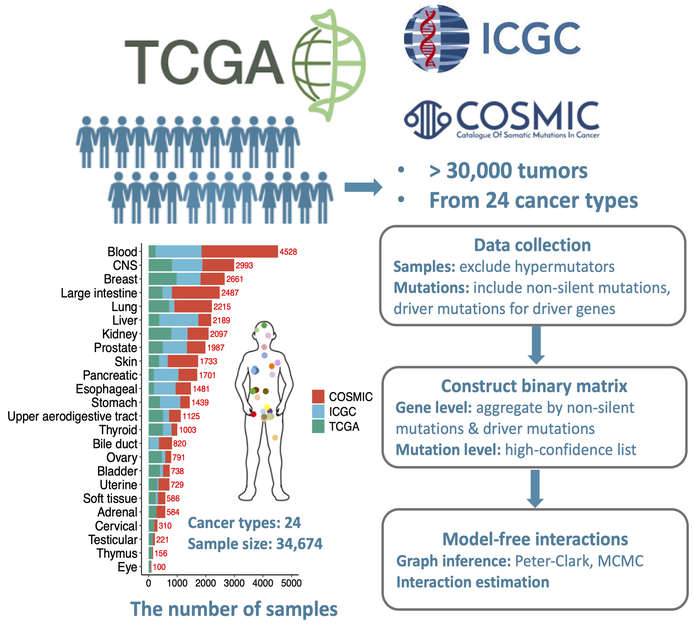
[1]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom.
[2]MRC Human Genetics Unit, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom.
[3]Higgs Centre for Theoretical Physics, School of Physics Astronomy, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom.
[4]School of Mathematics, University of Edinburgh Edinburgh EH9 3JZ, United Kingdom.

1 Motivation

- Large-scale cancer genomics programs (e.g., TCGA, ICGC) have provided an unprecedented opportunity to identify cancer driver genes and mutations.
- A single driver mutation or a single driver gene is usually not sufficient to induce oncogenesis, and how genetic interactions contribute to cancer phenotypes is still under investigation.
- To date, most cancer studies have mainly reported pairwise interactions at the gene level, and neglected the interactions at the mutation level.
- The lack of data on higher-order interactions should not be misunderstood as a proof that higher-order interactions do not exist or are less important.

2 Graphical summary

We apply a model-independent approach to estimate pairwise and higher-order interactions at both gene and site levels. Using data from 34,674 tumour samples across 24 cancer types, we are able to identify key interactions within and across cancer types, accounting for co-occurrence, mutual exclusivity as well as more complex higher-order relationships.



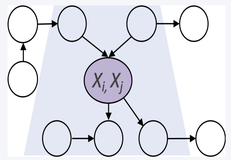
3 Model-free interactions

Definition

Let X be the mutational status of a set of driver genes or mutation sites (0, not mutated; 1, mutated). A pair of driver genes or mutations $\{X_i, X_j\} \in X$ has a pairwise interaction I_{ij}, where:

$$I_{ij} = \log \frac{p\left(X_i = 1, X_j = 1 \mid \underline{X} = 0\right) p\left(X_i = 0, X_j = 0 \mid \underline{X} = 0\right)}{p\left(X_i = 1, X_j = 0 \mid \underline{X} = 0\right) p\left(X_i = 0, X_j = 1 \mid \underline{X} = 0\right)}$$

- It is model-independent and can be directly estimated from observations [1].
- Condition on the Markov boundary, a minimal subset conditioned on which X_i and X_j become independent of other mutations.
- Markov boundary is identified by iterative MCMC method for causal discovery [2].
- Standard deviation is computed using bootstrap resampling.
- If two mutations or genes are independent, I_{ij}=0.



Markov boundary

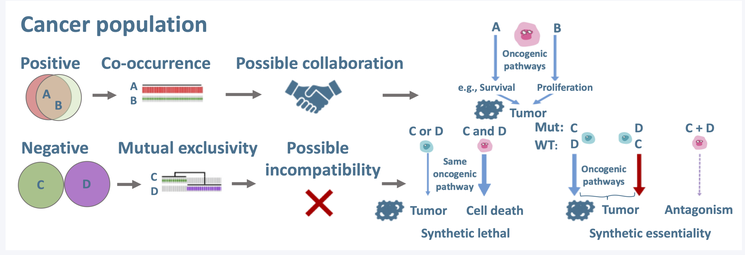
- It can be extended to higher-order interaction by taking n'th derivatives of $\log p(X)$:

$$I_{ijk} = \log \frac{p\left(1,1,X_k \mid \underline{X}\right) p\left(0,0,X_k \mid \underline{X}\right)}{p\left(0,1,X_k \mid \underline{X}\right) p\left(0,1,X_k \mid \underline{X}\right)}$$

$$= \log \left(\frac{p\left(1,1,1 \mid \underline{X}\right) p\left(0,0,1 \mid \underline{X}\right) p\left(0,1,0 \mid \underline{X}\right) p\left(1,0,0 \mid \underline{X}\right)}{p\left(0,0,0 \mid \underline{X}\right) p\left(0,1,1 \mid \underline{X}\right) p\left(1,1,0 \mid \underline{X}\right) p\left(1,0,1 \mid \underline{X}\right)} \right)$$

Interpretation

- Positive pairwise interaction: pairs of mutated genes or mutations are more likely to co-occur than being mutually exclusive in cancer population, indicative of two collaborating oncogenic pathways of hallmark features [3].
- Negative pairwise interaction: pairs of mutated genes or mutations are more likely to be mutually exclusive than being co-occurring, which might point to synthetic lethal or synthetic essentiality [3].



Schematic representation of pairwise genetic interactions

- Three-point interaction: mutations in a third gene affect the genetic interaction between the other two genes.

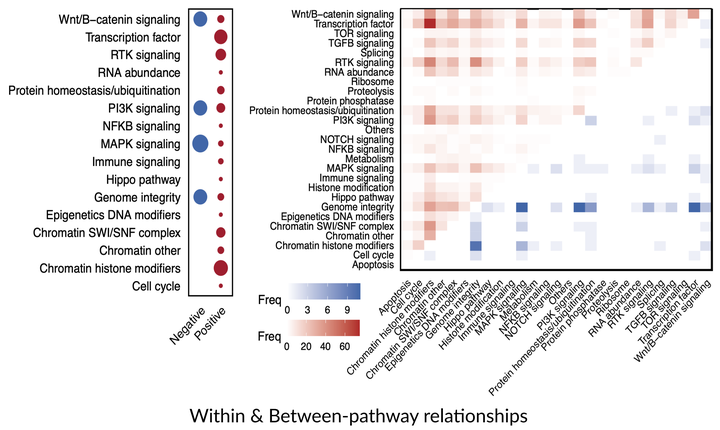
4 Interactions reveal underlying biology

Genetic interactions are context-dependent

- Genetic interactions are highly cancer-type specific.
- The sign of pairwise interaction might change across tissues, e.g., *PIK3CA* vs. *KRAS*, negative in lung cancer, but positive in large intestine cancer.
- The sign might also change depending on the aggregation approach or whether it is gene level or mutation level, e.g., in CNS cancer, *IDH1* R132H mutation particularly selects for *TP53* R273C mutation, compared to other *TP53* mutations.
- Higher-order interactions exist, e.g., *TP53*, *ATRX*, *IDH1* in CNS cancer.

Functionally relevant mechanisms of oncogenic pathways

- Highly connected genes are significantly involved in cellular processes and epigenetic changes.
- P53 and cell-cycle pathways are frequently co-altered across multiple cancer types.
- Genetic interactions inform pathway relationships.
- Genetic interactions significantly overlap with associations from the STRING database, which includes known physical and functional protein-protein interactions.



Within & Between-pathway relationships

5 Conclusion

Overall, these key interactions reflect the perturbation of cancer-relevant pathways or processes and might provide great benefits including:

- Gaining common mechanistic insights into the progression of cancer.
- Identifying prognostic or predictive interactions.
- Uncovering potential targeted and combination therapeutic opportunities.
- Stratifying tumour subtypes based on interactions.

References

1. Beentjes, S. V. & Khamseh, A. Higher-order interactions in statistical physics and machine learning: A model-independent solution to the inverse problem at equilibrium. *Physical Review E* **102,** 053314 (2020).
2. Kuipers, J. *et al.* Efficient sampling and structure learning of Bayesian networks. *Journal of Computational and Graphical Statistics*, 1–12 (2022).
3. El Tekle, G. *et al.* Co-occurrence and mutual exclusivity: what cross-cancer mutation patterns can tell us. *Trends in cancer* **7,** 823–836 (2021).