Supplementary material: Dynamic cancer drivers: A causal approach for cancer driver discovery based on bio-pathological trajectories

Andres M. Cifuentes-Bernal[1], Vu VH Pham[1], Xiaomei Li[1], Lin Liu[1], Jiuyong Li[1] and Thuc Duy Le[1*]

[1] UniSA STEM, University of South Australia, Adelaide, 5095 Mawson Lakes SA, Australia.

[1*] Correspondence author: thuc.le@unisa.edu.au

1 GO biological processes analysis

We performed a GO biological processes enrichment analysis to the single cell RNA sequencing data from NCBI GEO database, accession GSE75688 [Chung et al., 2017] to verify how significantly related are our discoveries to processes in cancer. Our analyses show a significant number of discovered drivers are strongly related relevant biological processes in cancer disease. Top ten enriched terms (for each trajectory) from Go Biological Processes are shown in Fig. 1 . For both dynamic drivers inferred sets (i.e from VIMtime(SC), and HER2time(SC)), top 10 enriched terms (ranked by p-value) are relevant to cancer (e.g. regulation of transcription and regulation of apoptotic process).

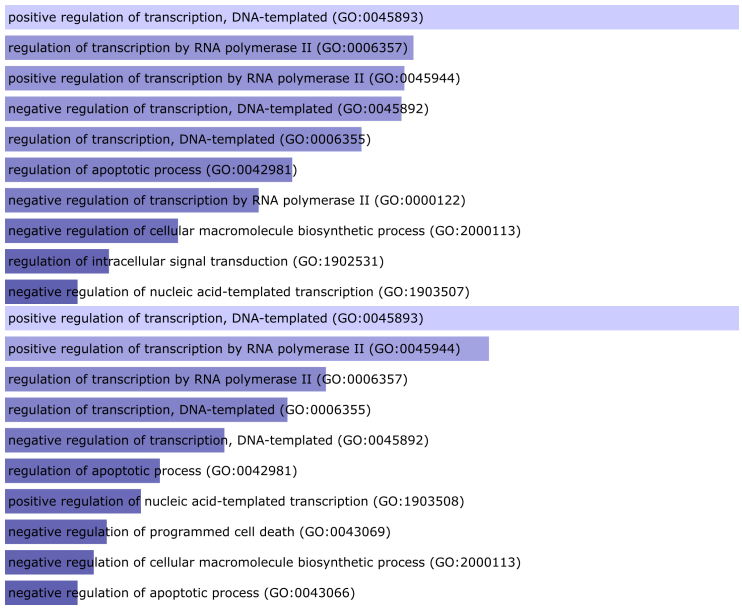

Figure 1: Top 10 Go biological terms 2021 (ranked by p-value) from our enrichment analysis for the gene set obtained from "VIMtime(SC)" (upper) and "HER2time(SC)" (lower). In both cases, enriched terms correspond to regulation of biological processes relevant to cancer progression. Bar length represents the significance of the term. Brightness is used as auxiliary visual for significance, the lower the p-value, the brighter the colour. Analysis performed by using Enrichr [Kuleshov et al., 2016]

2 Gene-disease association analysis (DisGeNet)

We performed an enrichment analysis to identify gene-disease associations (GDAs) in the inferred drivers from the GSE75688 single cell dataset [Chung et al., 2017]. We use DisGeNET [Piñero et al., 2016] as gene-disease associations database for this analysis. Our analyses show a significant number of discovered drivers are strongly related to cancer disease. The top ten enriched GDAs terms (ranked by p-value) for each of the analysed pseudotimes (i.e VIMtime(SC), and HER2time(SC)) are shown in Fig. 2. In both cases, all of top ten terms are related to cancer disease and cancer progression.

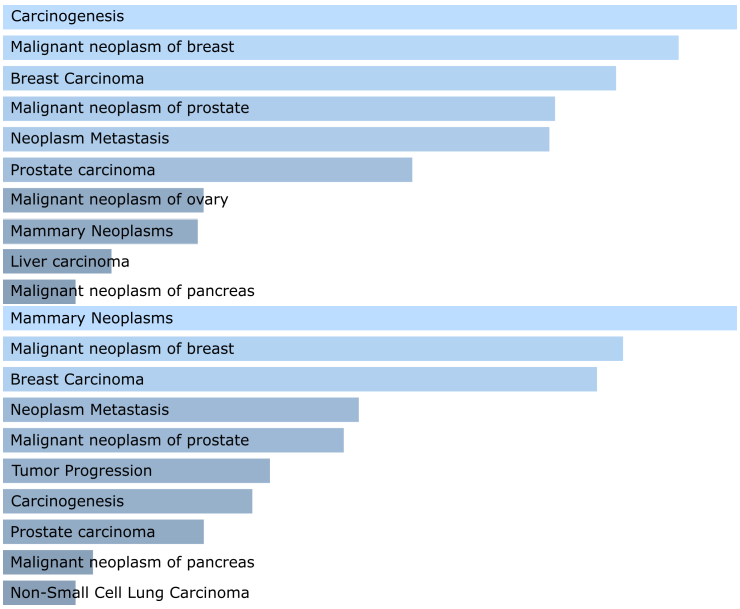

Figure 2: Top 10 GDAs terms (ranked by p-value) from our enrichment analysis for the gene set obtained from "VIMtime(SC)" (upper) and "HER2time(SC)" (lower) retrieved from DisGeNET. Bar length represents the significance of the term. Brightness is used as auxiliary visual for significance, the lower the p-value, the brighter the colour. Analysis performed by using Enrichr [Kuleshov et al., 2016]

References

[Chung et al., 2017] Chung, W., Eum, H. H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., Ryu, H. S., Kim, S., Lee, J. E., Park, Y. H., Kan, Z., Han, W., and Park, W.-Y. (2017). Single-cell rna-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications*, 8(1):15081.

[Kuleshov et al., 2016] Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., and Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44:W90–7.

[Piñero et al., 2016] Piñero, J., Bravo, l., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L. I. (2016). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, 45(D1):D833–D839.