**HotNet2 (Nature)**

1. The authors filter out hypermutated samples from the initial sample set before applying the diffused heat model. However, since the heat diffusion model ranks the SNVs based on their frequency as well as the nature of interaction between two proteins, shouldn’t it still be alright to have the hypermutated samples in the study. Infact, they themselves suggest that HotNet2 is able to circumvent the influence of hypermutated genes in network weighing, by generating ‘star subnetworks’ from these hub nodes). By that accord, including hypermutated samples in this study should actually help pinpoint which networks are playing a regulatory role in an ultramutated case, and will actually be more informative (for ex., in NSCLC samples). Note: the resource they cite for exclusion of hypermutator samples excludes samples with >500 mutations.
2. One important assumption this paper seems to make is that genes with low expression in all tumor types are not of significance (they exclude these from the sample set). That is, only if there is an overexpression is there cause for investigation. Consistently low expression genes may still have important regulatory roles to play – for example, tumor suppressor genes can be expected to be downregulated in cancers, and excluding them from the study will probably result in more networks that are enriched in oncogenes, rather than disrupted networks that regulate growth and division in normal cells. Can a better alternative be to exclude just the housekeeping genes, instead of the consistently low expression genes? Or is there any additional benefit that can come from using ‘control datasets’ of normal cells’ RNA-Seq data to identify low expression genes that are normally expressed at higher levels?

**TRUP (Fernandez-Cuesta et al, Genome Biology)**

1. Why do the authors filter out predictions where both breakpoints lie in self-chain regions? Some papers point at the importance of short low-copy repeat regions (the self-chains) in inducing somatic variants, which in turn are associated with genomic disorders. Hence I am not clear as to why we would need to filter out predictions covering SC regions.
2. The authors present variable criteria for selection fusion points of interest. In the initial dataset with 30x coverage, they place a minimal requirement of 5 reads spanning the breakpoint (17% of the average number of reads!) so as to remove pipeline artifacts. Then in the cross-comparison with other tools, they relax it to atleast 2 supporting reads. Later, in the methods, they list 3 different read coverage requirements (defaults being 2, 3, 4) based on the number of discordant read pairs. Won’t this sort of arbitrary parametrization influence the number of fusion genes identified?

**Zodiac (Zhu et al)**

1. The heavy emphasis on the prior while predicting important networks from the user query means that we loose rare regulatory networks from the prediction. The use of 1448 samples across 11 different cancer types (with individual cohort sizes ranging from 20-250 samples) means that the final prior set built is extremely generalist. I wonder if a cohort-based ‘normalization’ would be able to make the final prior on networks more detailed and representative of the different cancers, or would the prior still be the same, being based solely on the major interactions learnt across all the cancers?
2. The authors suggest that using this BGM model to identify genetic interactions of individual cancer types requires more computation. Why isn’t it just sufficient to have a large enough sample set for querying against the prior? The authors suggest that the posterior network prediction does result in inference of new edges, giving ‘customized network inference’. I do not understand how this is different from the former.
3. The authors say that this resource can be used to generate a deeper understanding of the molecular mechanisms of cancer. But every cancer type has different regulatory mechanisms driving it, so is it sufficient to just look at the major pan-cancer networks are perturbed or should one also add annotations for cohort-specific genes and networks in the final results? I do not fully understand how a map explaining the interaction between methylation and copy number and genetic mutations in a specific set of known driver genes/networks can help uncover rare networks that may be cancer specific.