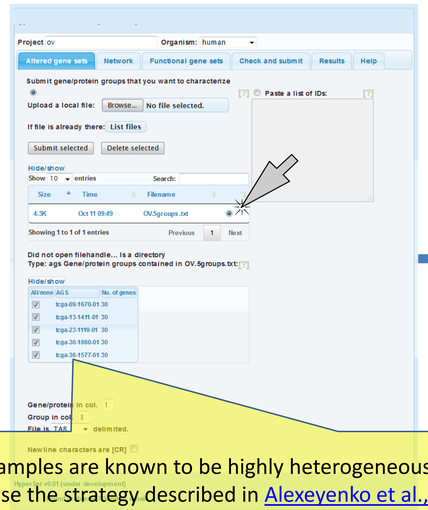
Exploration of transcriptome landscapes

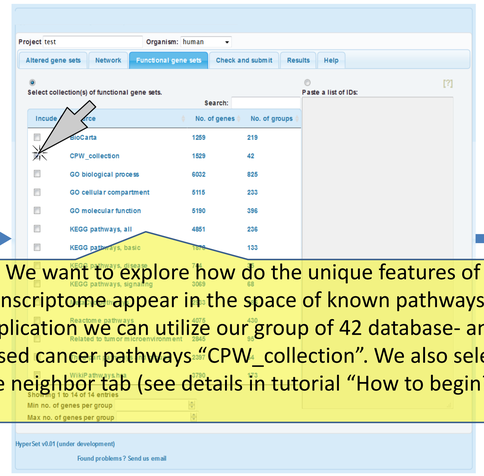
https://www.evinet.org



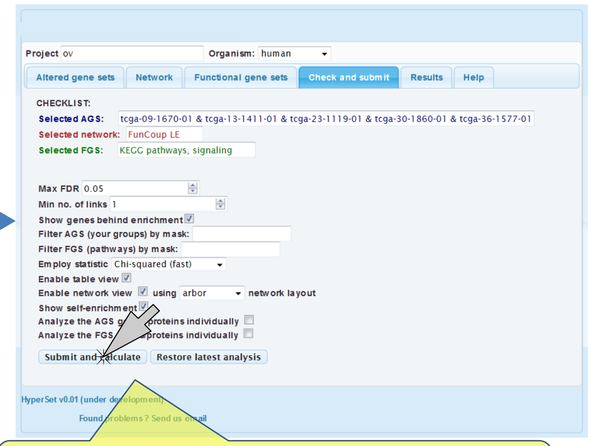
Cancer samples are known to be highly heterogeneous across patients. In order to characterize them, one can choose the strategy described in Alexeyenko et al., 2012: in each 'omics' sample, select top N genes by which it differs from the rest of the cohort. We can do that with e.g. following R code (see more details in tutorial "Using R package NEArender in data analysis pipelines"):

```
> library("NEArender");
> dim(Data$OV$GE$Agilent)
        [1] 17813   589
> m0 <- Data$OV$GE$Agilent[,sample(colnames(Data$OV$GE$Agilent), 5)];
> ov1 <- samples2ags(m0, method="top", Ntop=30);
> print.ags.list(ov1, File = "OV.5groups.txt");
```

Next, we submit the obtained text file with N=30 genes per sample using the first tab.

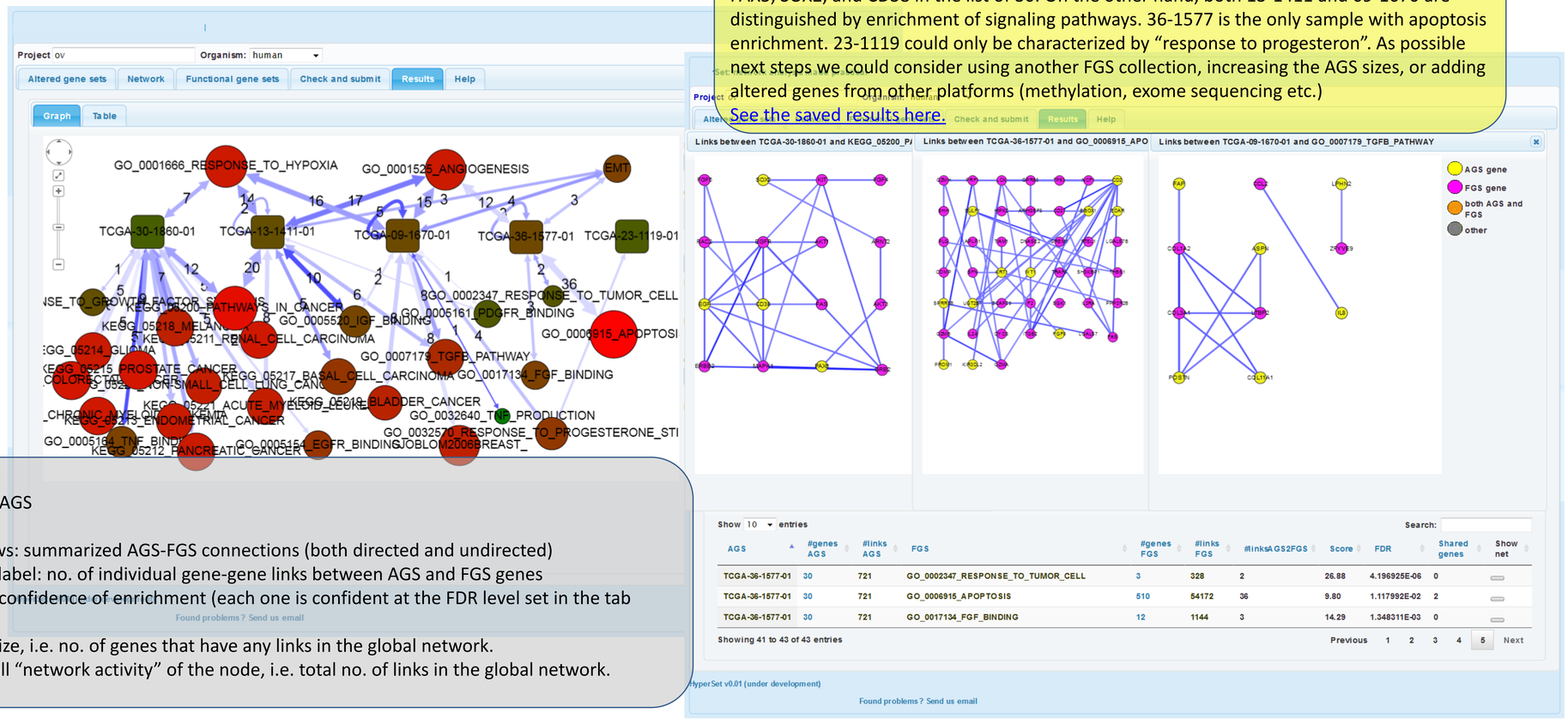


We want to explore how do the unique features of each transcriptome appear in the space of known pathways. For the cancer application we can utilize our group of 42 database- and publication-based cancer pathways "CPW_collection". We also select a network in the neighbor tab (see details in tutorial "How to begin?").



As was explained in tutorial "How to begin?", we overview the selected parameters and click "Submit".

The first observation we can make is that, indeed, the 5 transcriptomes differ in their pathway patterns. 30-1860 is enriched in connections to multiple cancer pathways (which pretty much say ditto given the pathway overlap); this is mostly due to the presence of EGF , PAX5, SOX2, and CD38 in the list of 30. On the other hand, both 13-1411 and 09-1670 are distinguished by enrichment of signaling pathways. 36-1577 is the only sample with apoptosis enrichment. 23-1119 could only be characterized by "response to progesteron". As possible next steps we could consider using another FGS collection, increasing the AGS sizes, or adding altered genes from other platforms (methylation, exome sequencing etc.)
See the saved results here.



Graph legend

Rounded rectangles: AGS
Circles: FGS
Double-headed arrows: summarized AGS-FGS connections (both directed and undirected)
Arrow thickness and label: no. of individual gene-gene links between AGS and FGS genes
Arrow transparency: confidence of enrichment (each one is confident at the FDR level set in the tab "Check and submit")
AGS/FGS size: node size, i.e. no. of genes that have any links in the global network.
AGS/FGS color: overall "network activity" of the node, i.e. total no. of links in the global network.