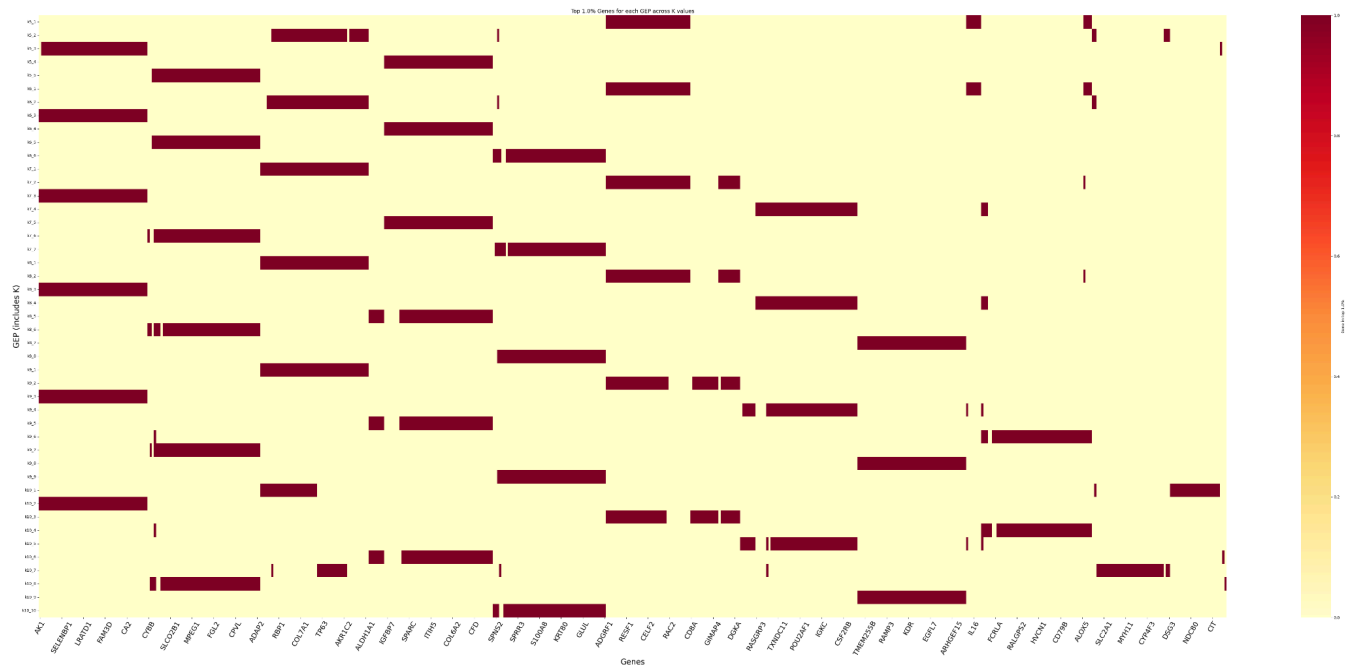# Visualizations on consensus results for Patient S1

Figure 1: Heatmap that shows for every GEP across all K values, which genes appear in the top 1% of that GEP's spectra scores (every 10 genes labeled in X axis for neatness)

Observations:
- Which genes repeatedly appear as top genes: Genes like AK1, SELENBP1, RBP1 and other genes which are repeatedly seen across multiple GEPs across different K values means that these genes consistently appear among the strongest genes for GEPs and likely indicate a robust and biologically stable marker.
- How genes are distributing across GEPs: genes that appear in only a few GEPs (like NDC80) are specific to particular programs while the genes present across many GEPs might be part of a metaGEP or broader signature
- Differences between Ks: when new genes appear at different Ks like NDC80, that could be indicative of program splitting or resolution dependent GEPs
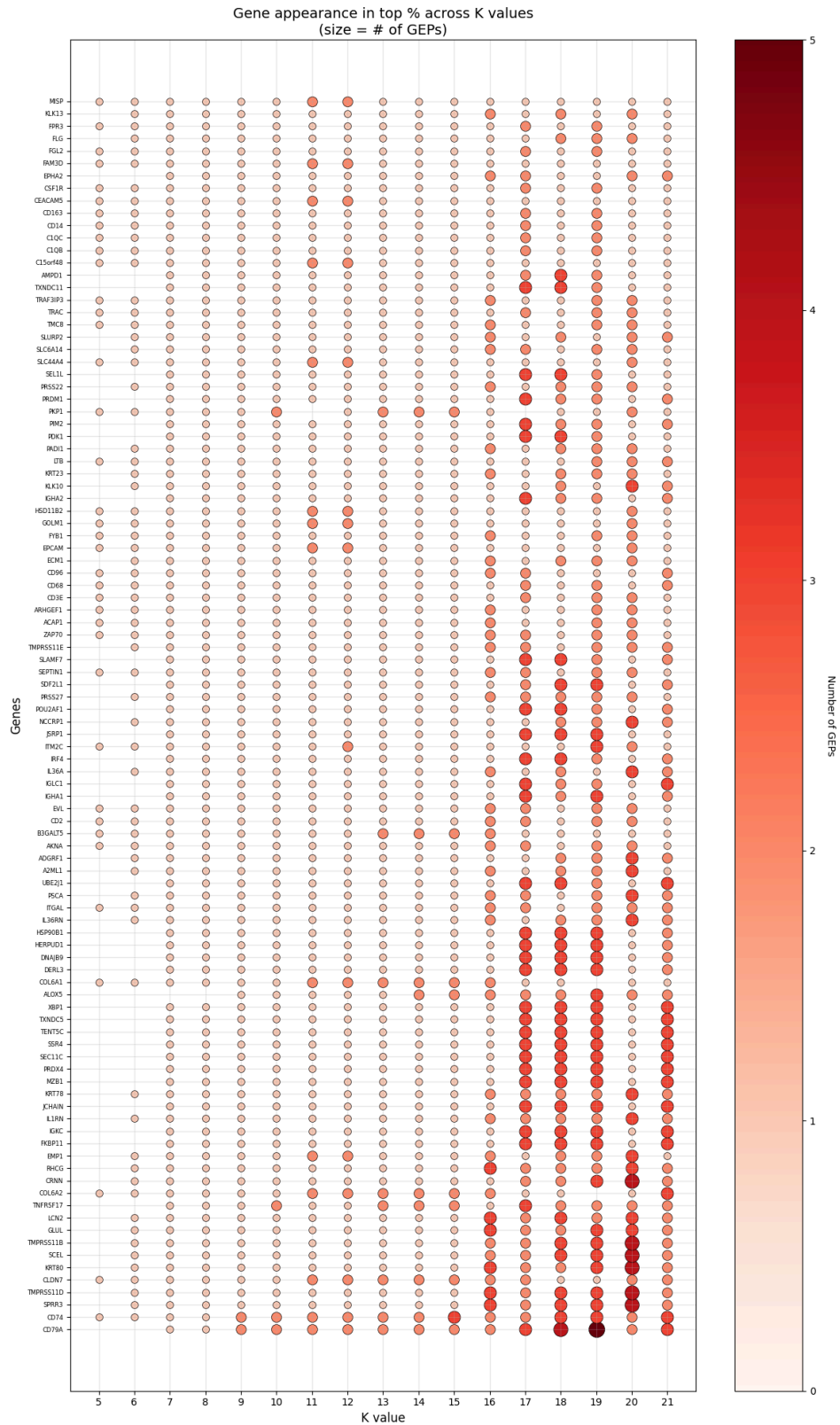
Figure 2: gene-by-K bubble heatmap showing how often each gene appears in the top 1% of GEPs for each value of K

Observations:

- The main takeaway from this figure is that the larger the bubbles are, the more often a particular gene is being repeated across multiple GEPs for a K value. This can be indicative of redundant GEPs being identified i.e. these GEPs are actually part of a broader GEP
- This trend can be observed in the later K values (especially from K=16 onwards) – where the bubbles become increasingly larger i,e, the genes seen in the y-axis are repeated several times in the top 1% of the 4/5/6 GEPs for that particular K value
- From this figure, it might be a takeaway that K=8 seems to be a good choice for a K value – here, there is no repetition amongst the top 1% of genes identified for each of the 8 GEPs
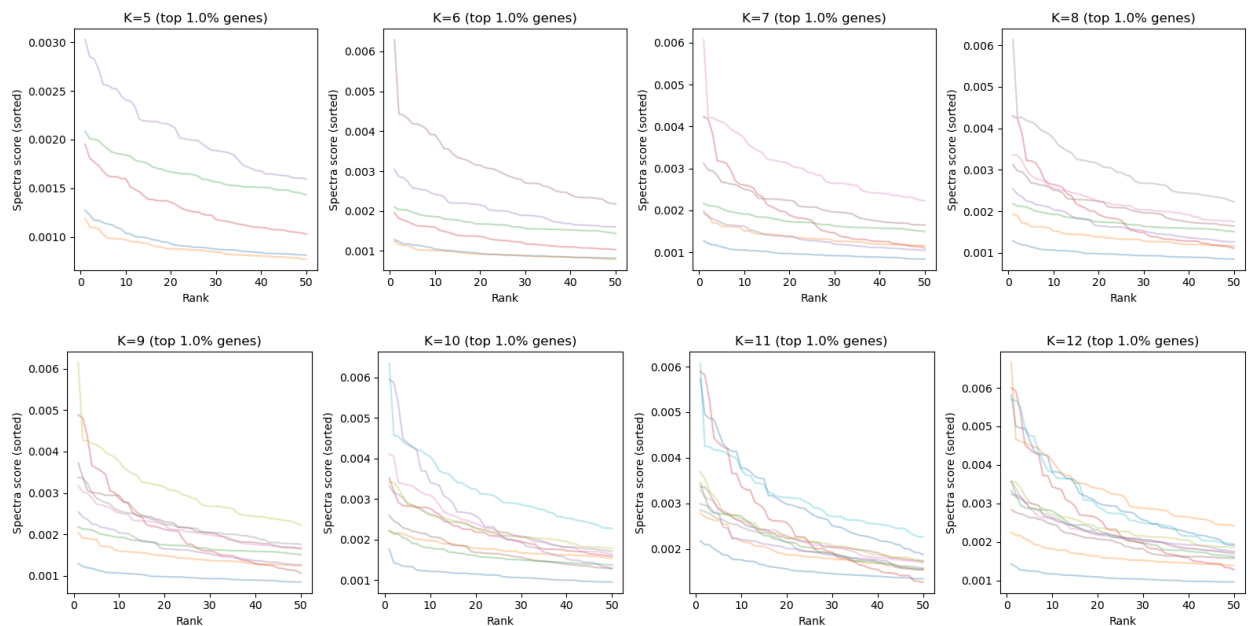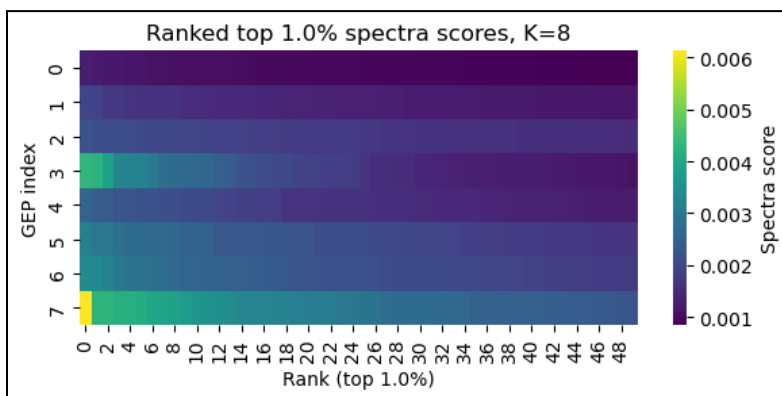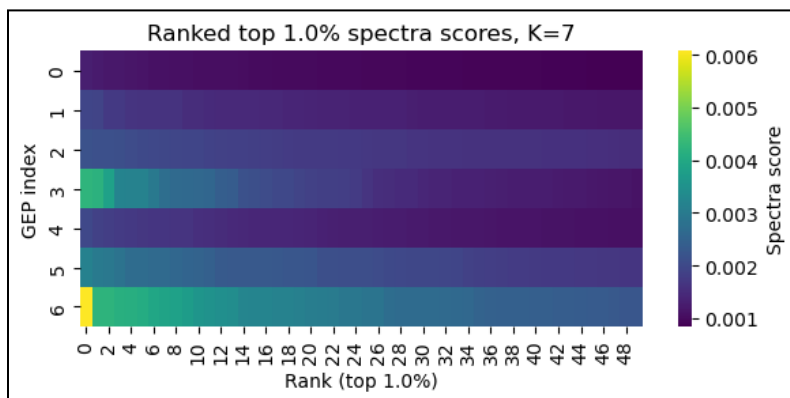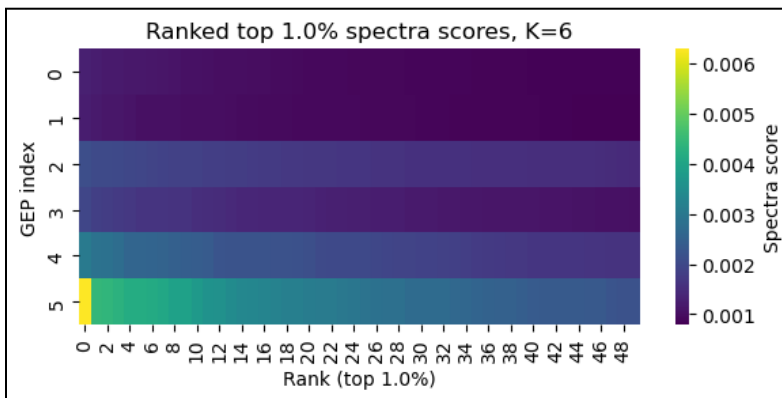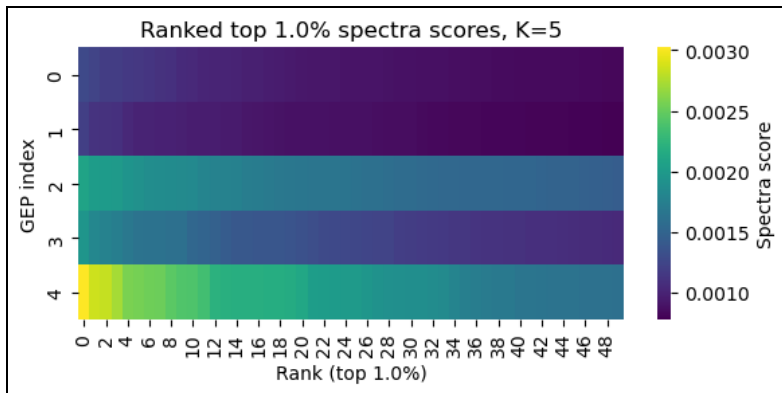

Figure 3: Ranked spectral line plots (top 1%)

Observations:

- Lines that show a steeper drop off in the spectral lines are indicative of the GEP containing only a few very strong genes i.e. it being a highly specific GEP. On the other hand, flatter curves are indicative of many moderately scoring genes so not as highly specific.
  - This can be noticed in K=6, 7,, 8… where there are steep curves in the spectral line that represents a GEP
- From K=9 onwards, we see curves that start to flatten and there are several overlapping curves like this – this shows that programs are splitting and we are losing peaks in the spectral vector lines
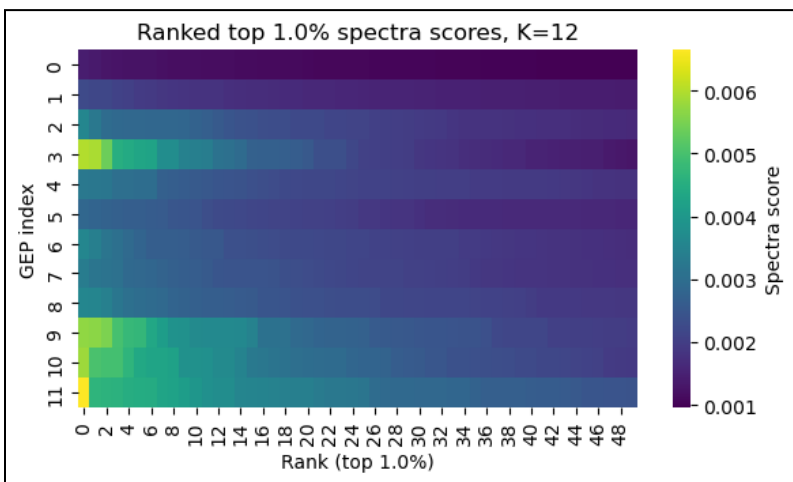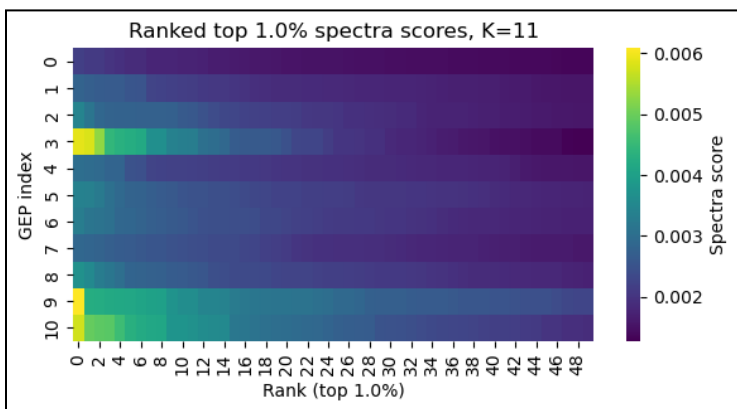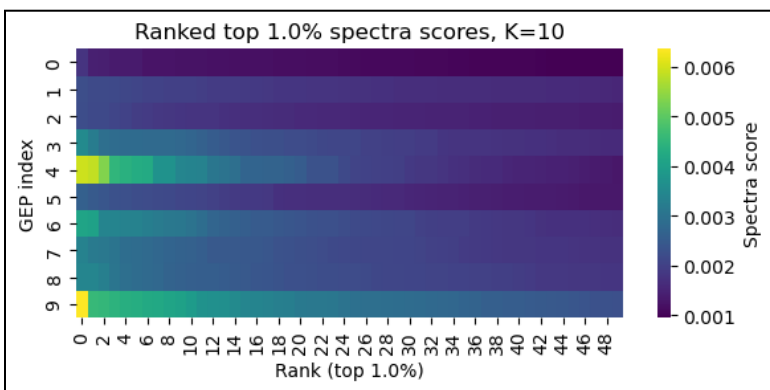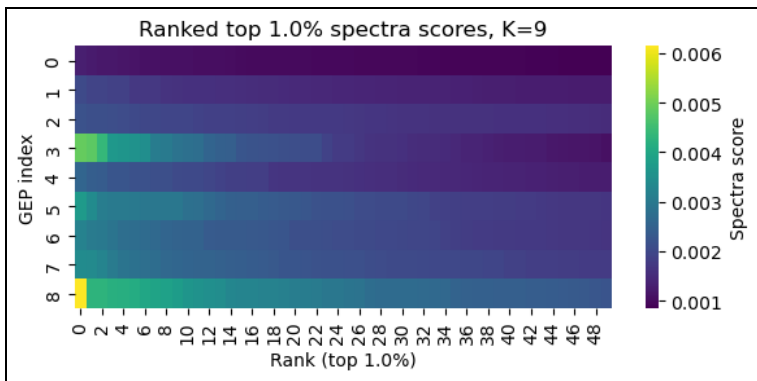
Ranked top 1.0% spectra scores, K=5

Ranked top 1.0% spectra scores, K=6

Ranked top 1.0% spectra scores, K=7

Ranked top 1.0% spectra scores, K=8

Ranked top 1.0% spectra scores, K=9

Ranked top 1.0% spectra scores, K=10

Ranked top 1.0% spectra scores, K=11

Ranked top 1.0% spectra scores, K=12

Figure 4: Ranked heatmap of top 1% genes

- Can be compared to spectral line plots in Figure 3 – some GEPs clearly have stronger top genes
- More intense rows means that we are discovering finer programs i.e. represents splitting of GEPs into finer GEPs