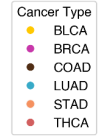
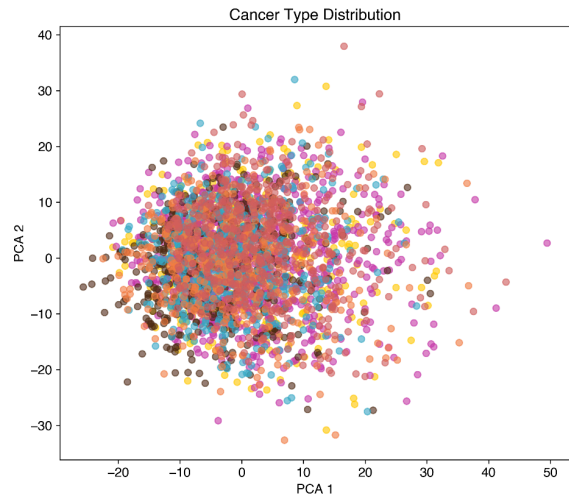
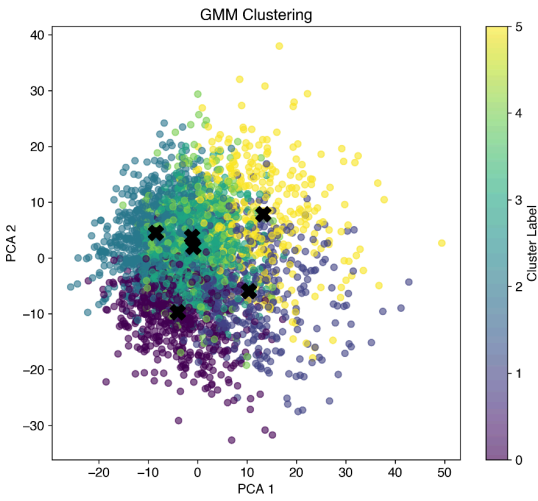
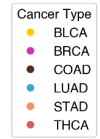
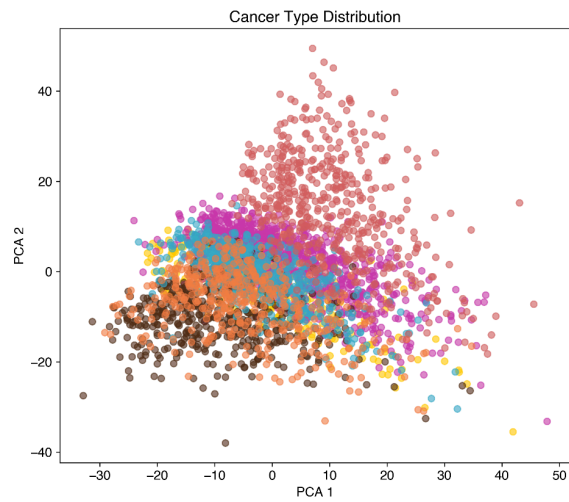
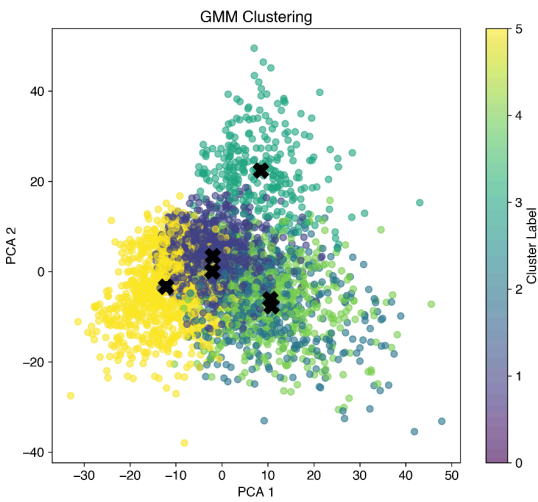


## Supplemental Figures

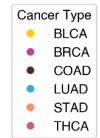
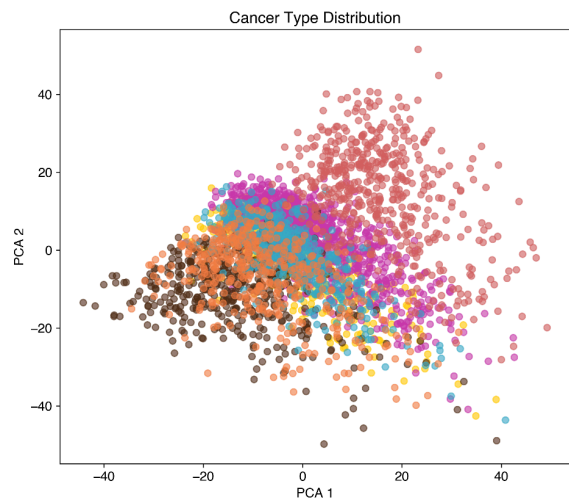
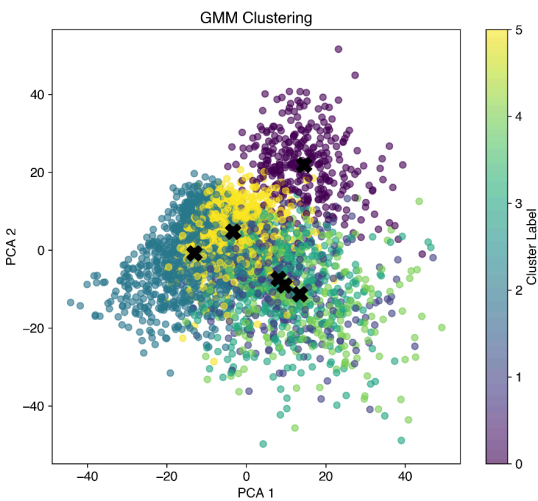
SC: 3 R: 3



SC: 3 R: 4

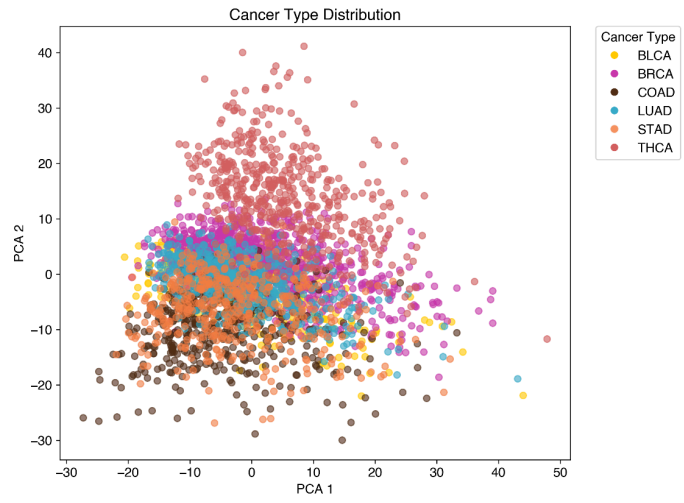
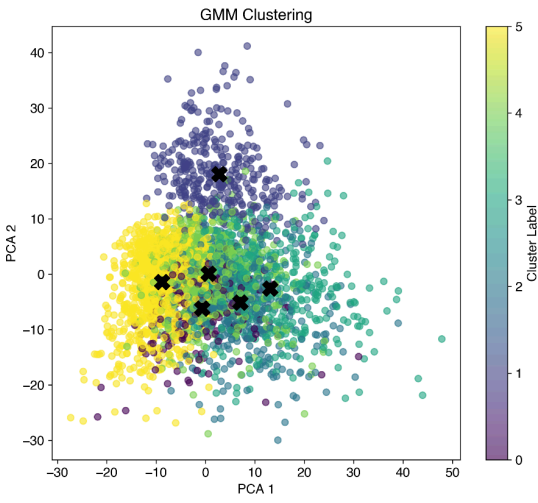


SC: 3 R: 5

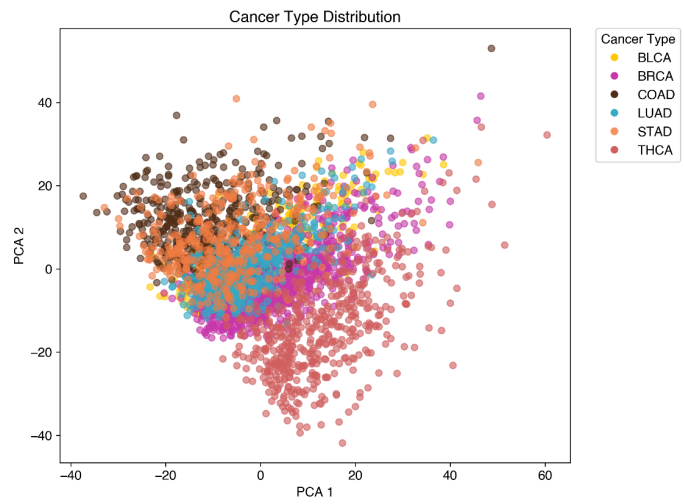
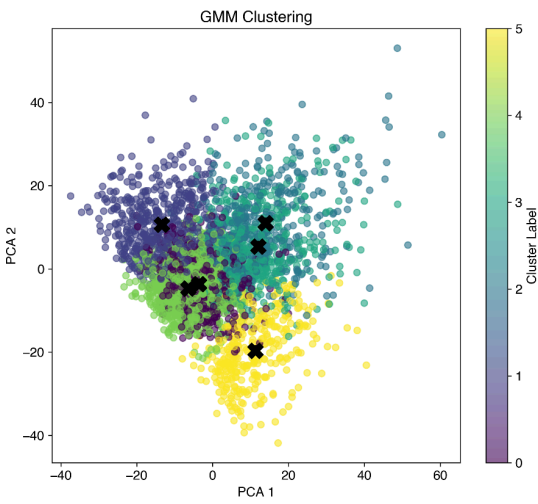


**Supplemental Figure 1:** Clustering performance with a fixed sample count of 3 and increasing sampling counts from 3 to 5 demonstrates progressive enhancement in cluster separation. With 3 repetitions, clustering lacks clear separation, as reflected by the Gaussian Mixture Model (GMM) clustering. Increasing the repetition count improves both cancer type and GMM clustering, resulting in greater separation and more distinct cluster formation.

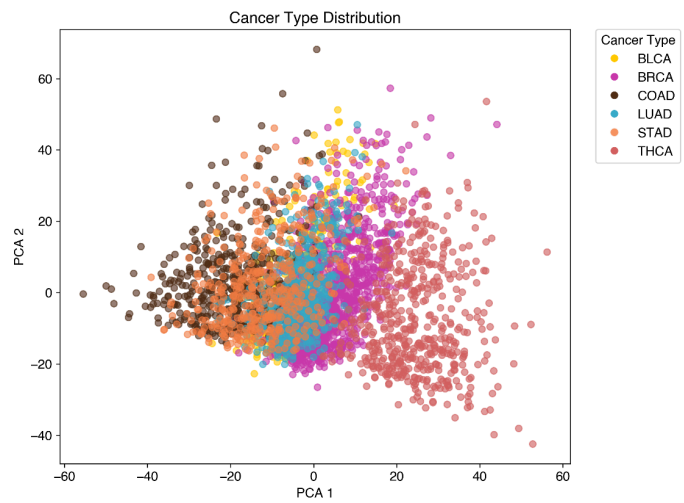
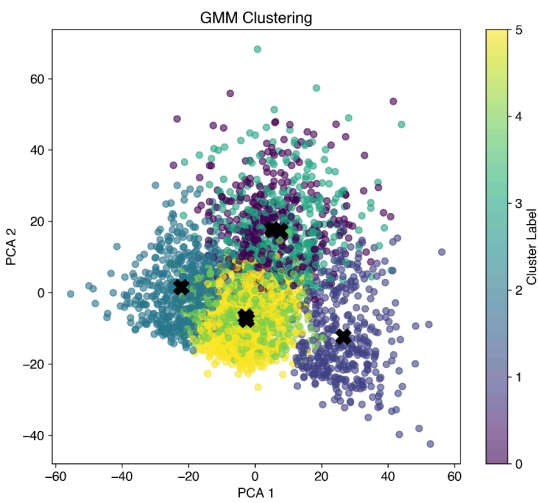
SC: 4 R: 3



SC: 4 R: 4

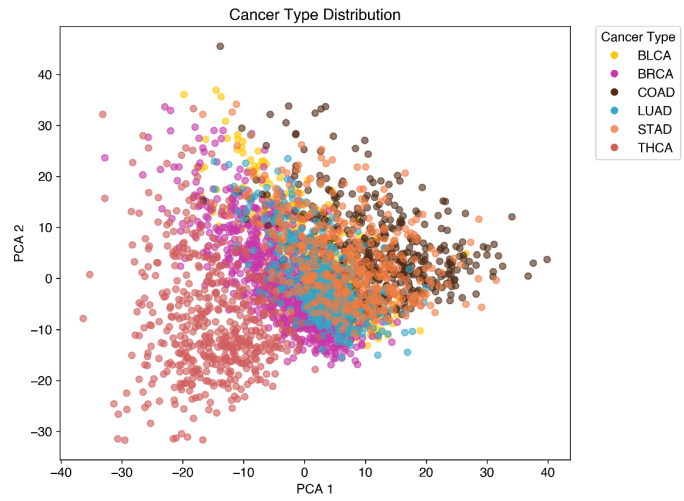
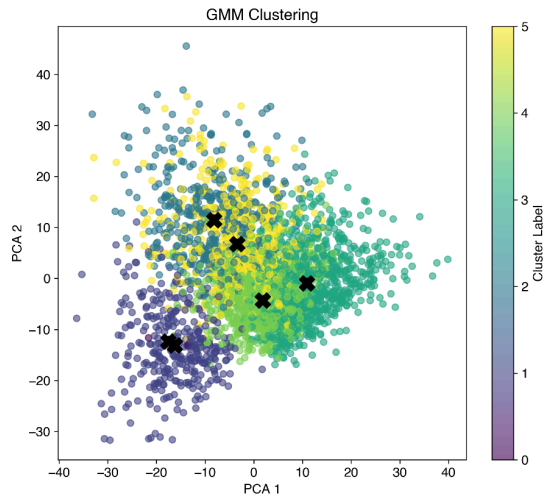


SC: 4 R: 5

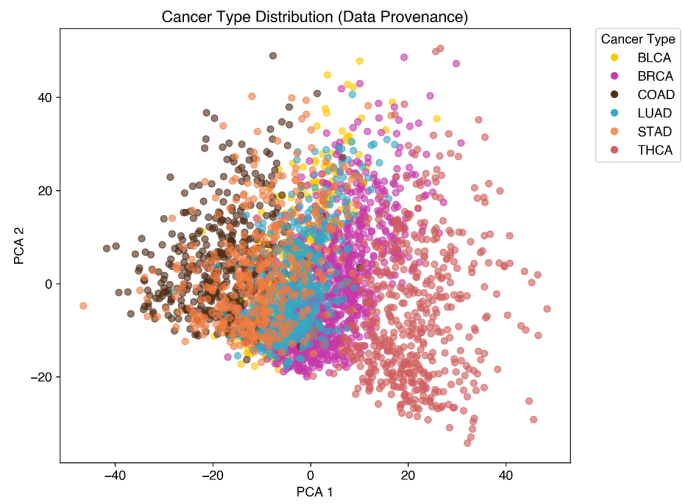
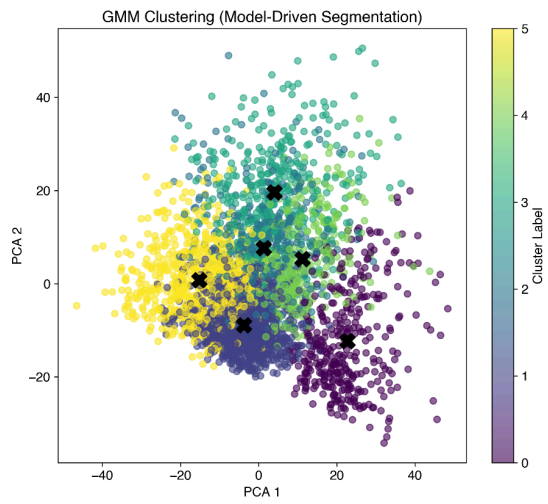


**Supplemental Figure 2:** Clustering performance with a fixed sample count of 4 and varying repetition counts between 3 and 5 shows enhanced clustering quality with increasing repetition. As repetition count increases, both Gaussian Mixture Model (GMM) clustering and cancer type clustering exhibit improved alignment and separation, indicating more robust cluster formation.

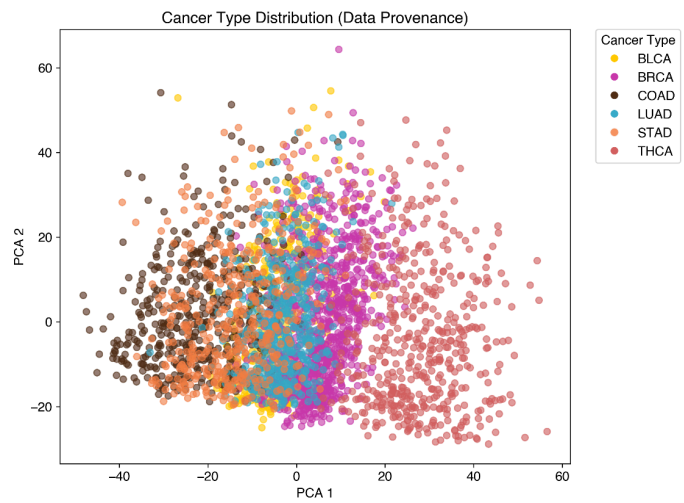
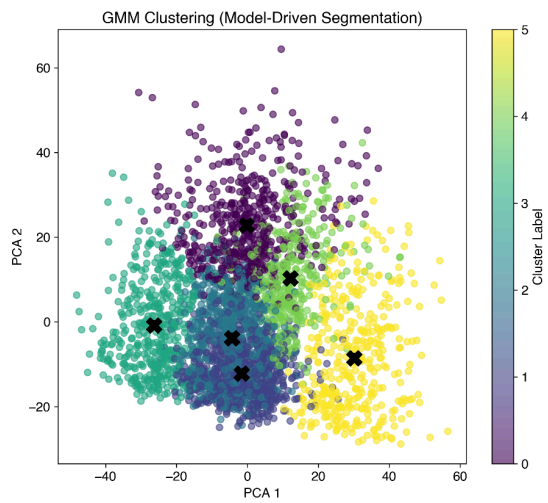
SC: 5 R: 3



SC: 5 R: 4

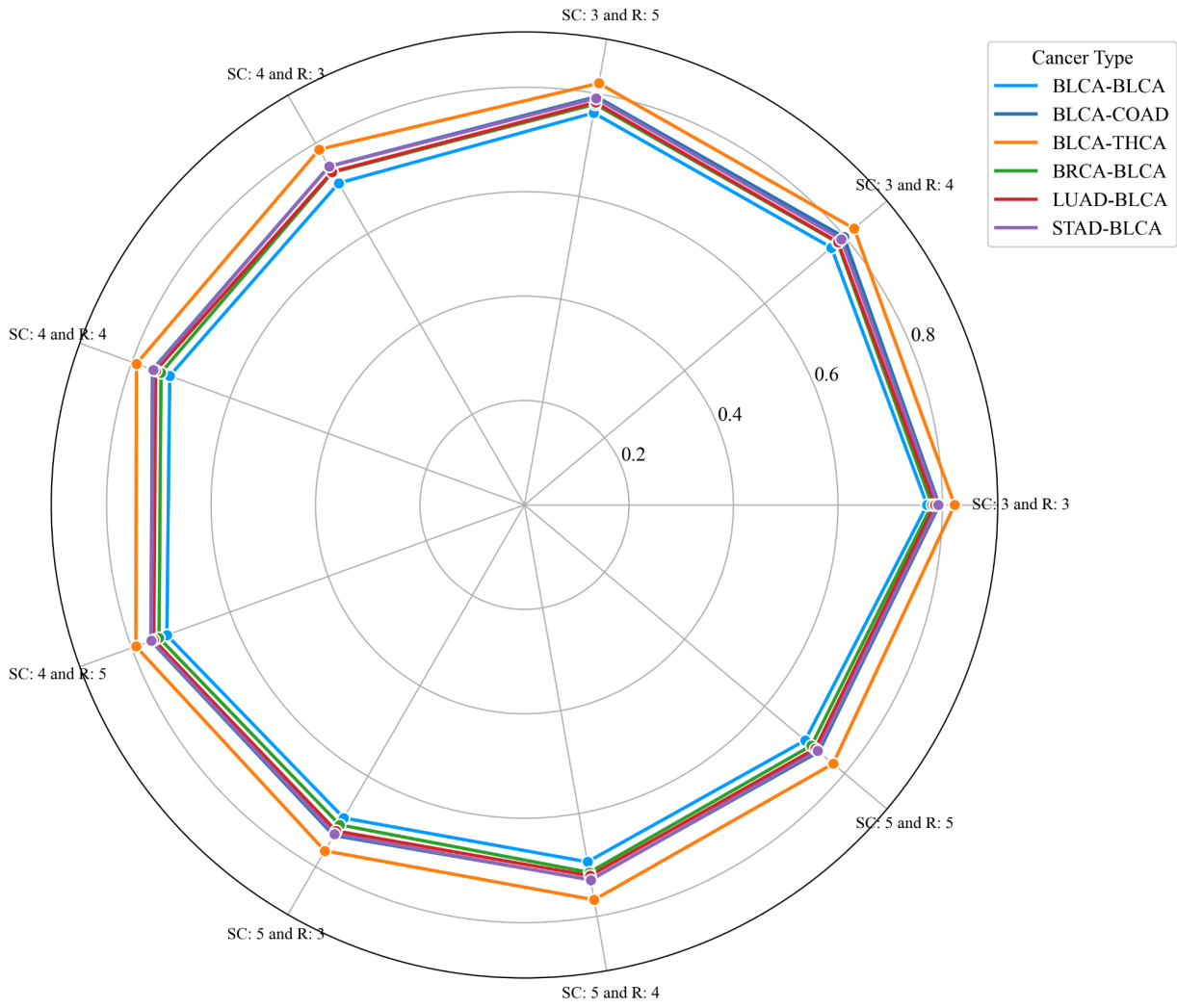


SC: 5 R: 5

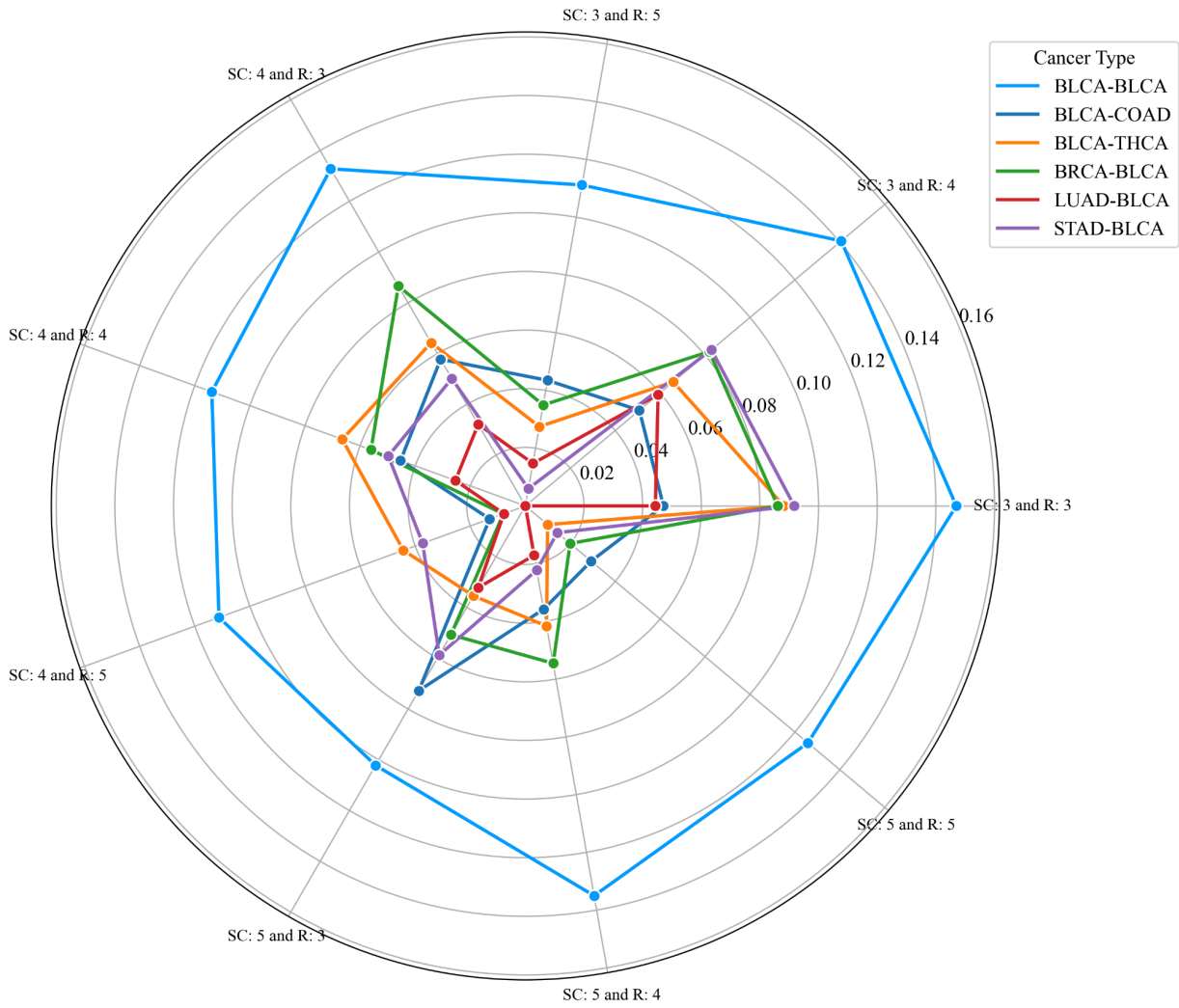


**Supplemental Figure 3:** Clustering performance with a fixed sample count of 5 and varying repetition counts between 3 and 5 demonstrates improved clustering with increasing repetition. Additionally, the Gaussian Mixture Model (GMM) unsupervised clustering aligns well with the cancer subtype clusters, indicating strong agreement between predicted and true cluster structures.

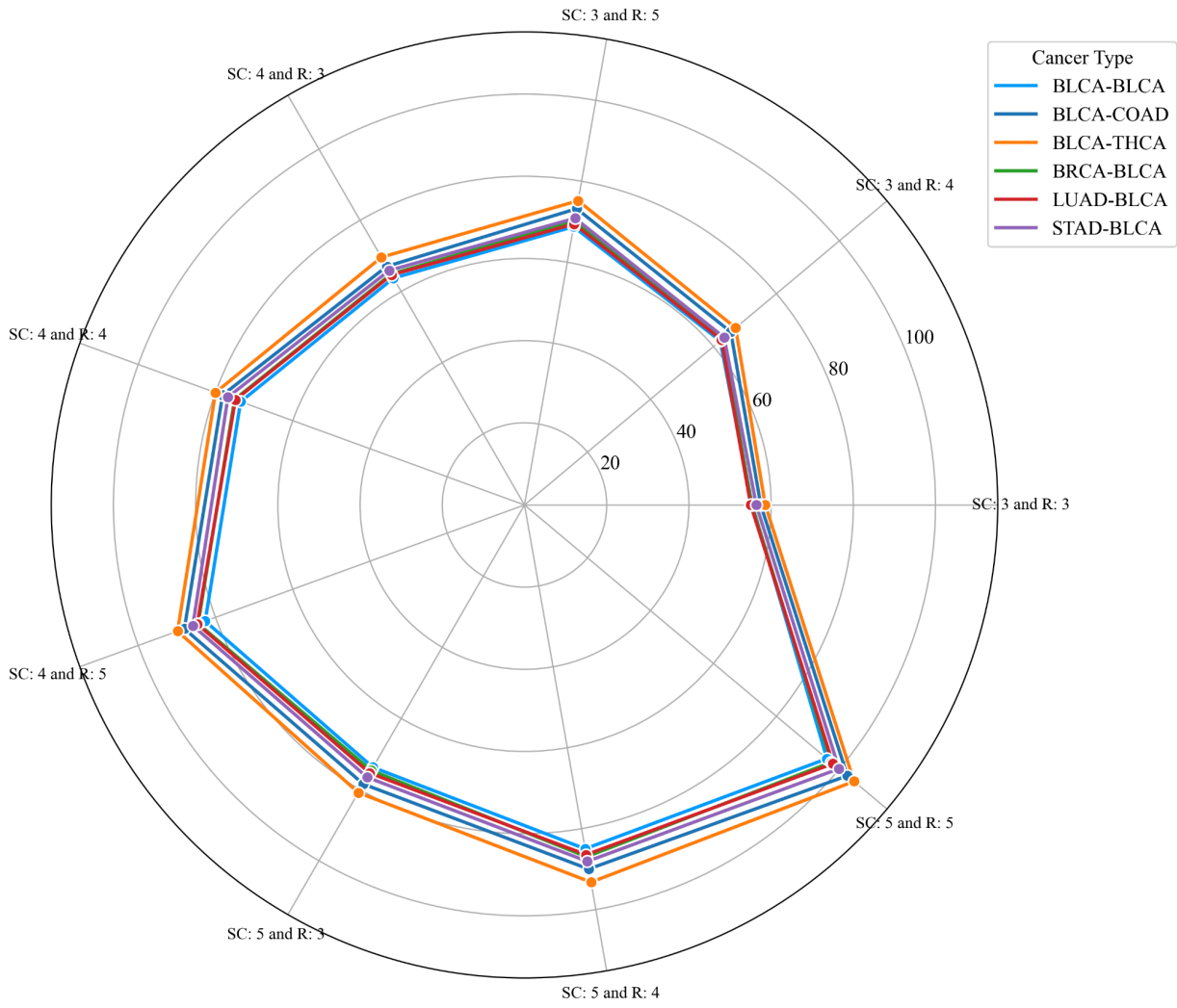
### BLCA - Cosine



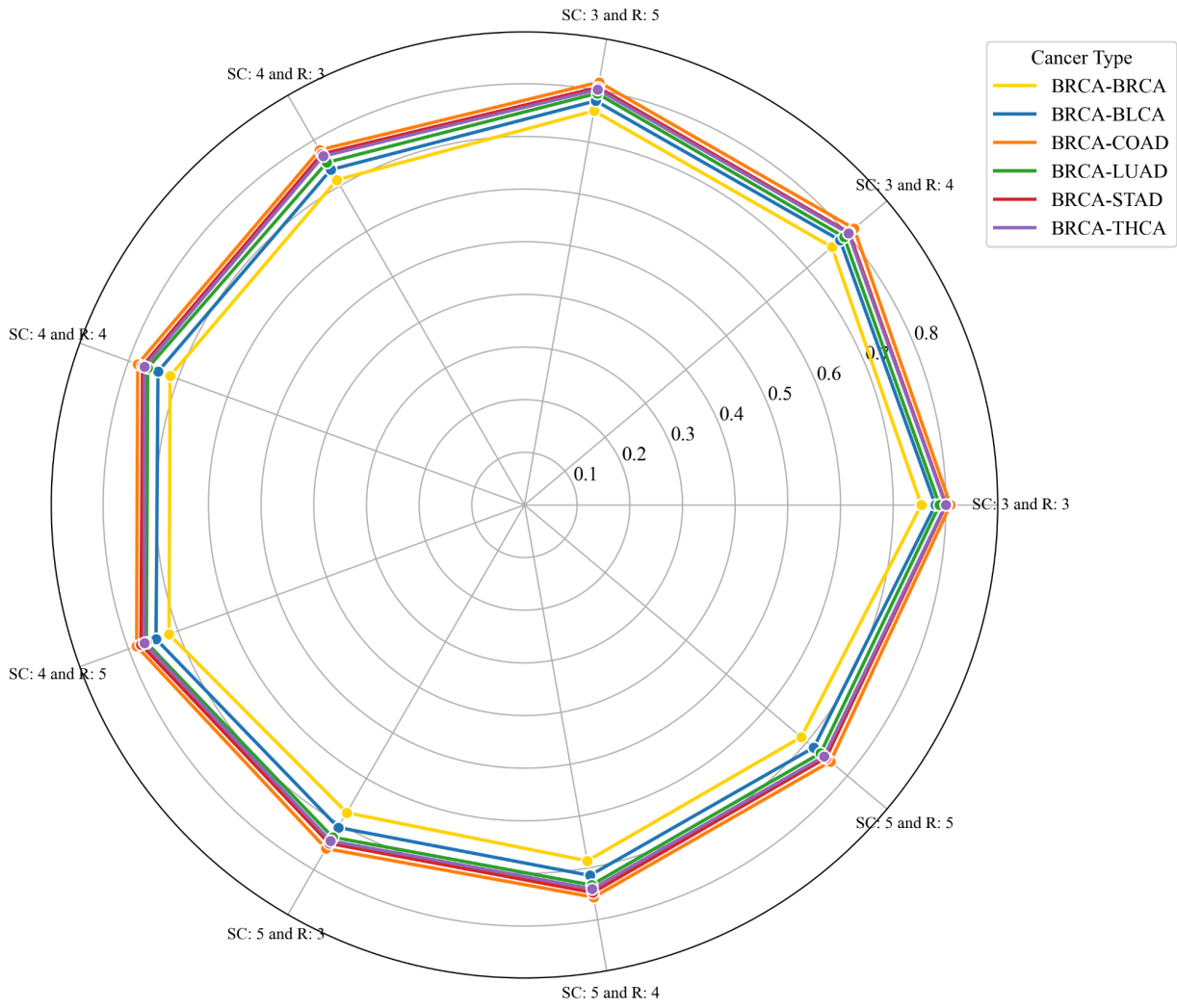
### BLCA - Dot Product



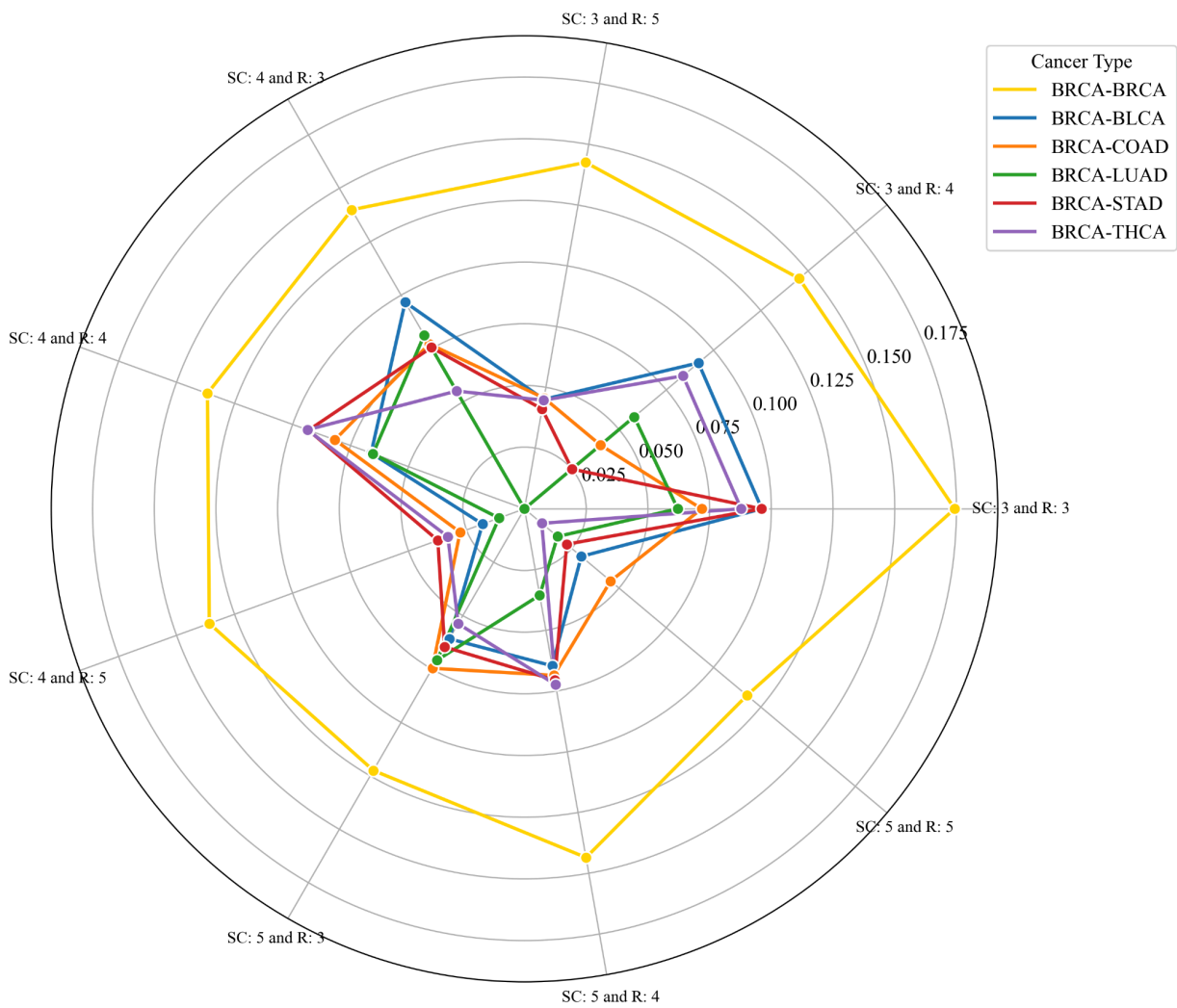
### BLCA - Euclidean Distance



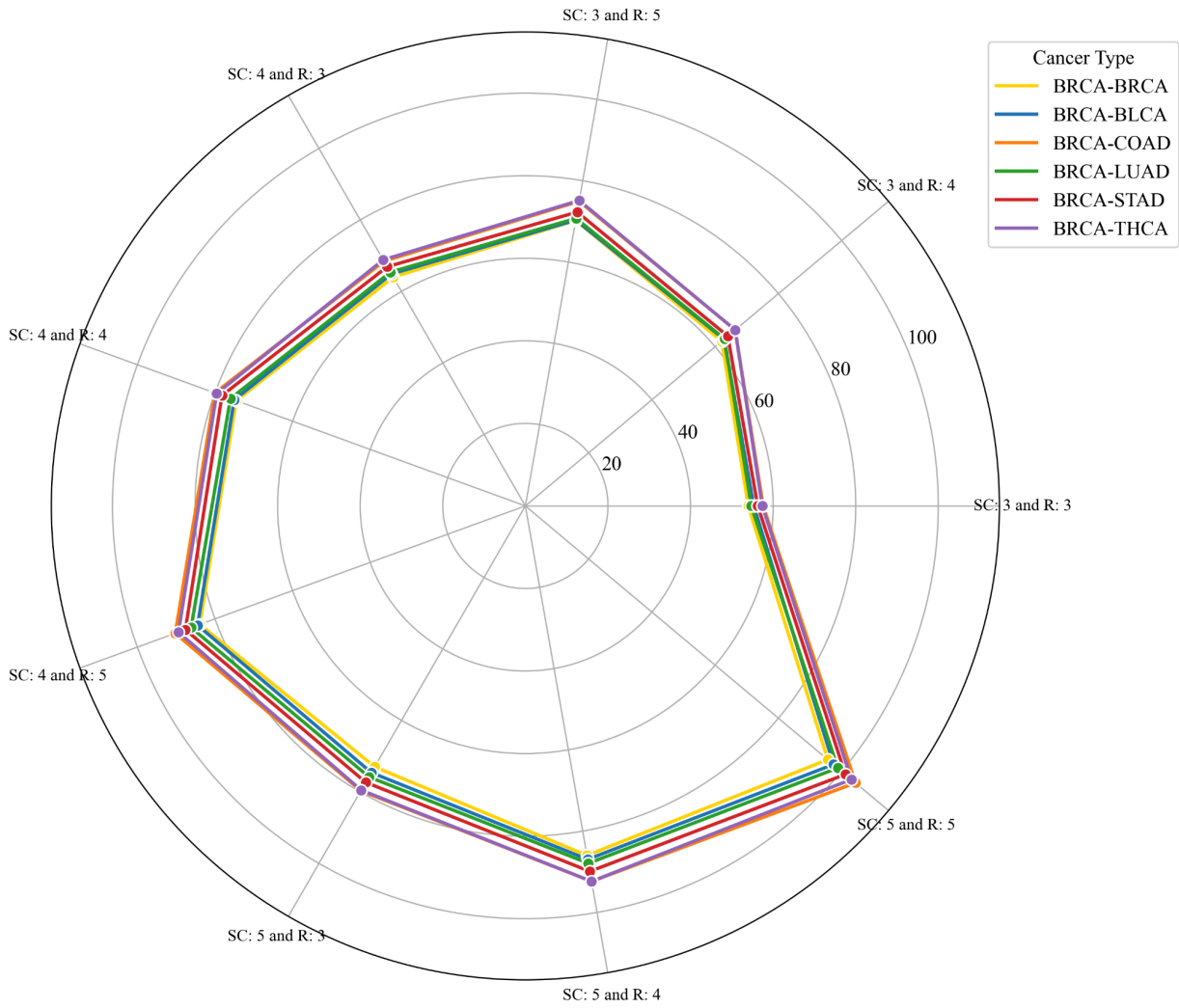
### BRCA - Cosine



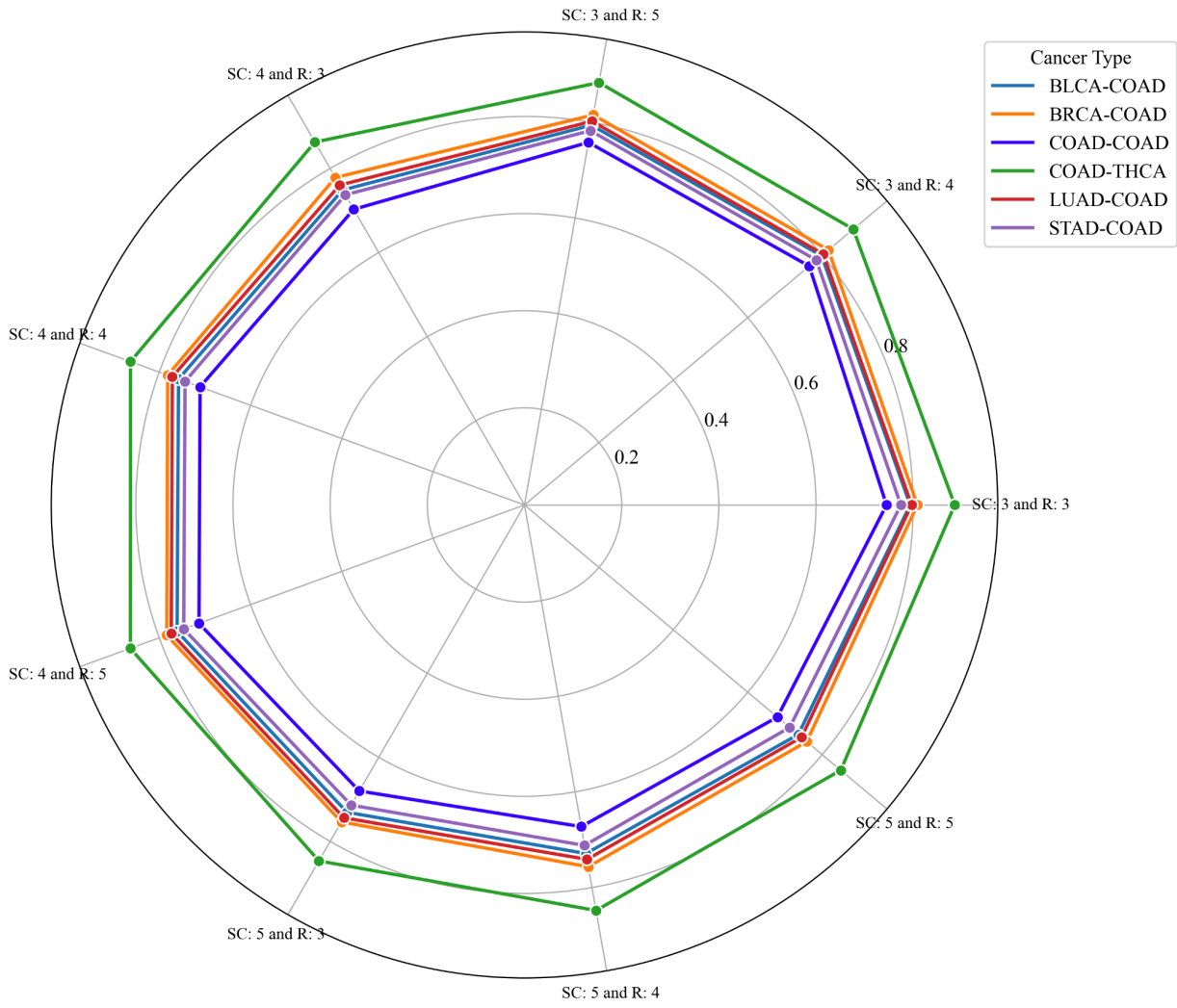
### BRCA - Dot Product



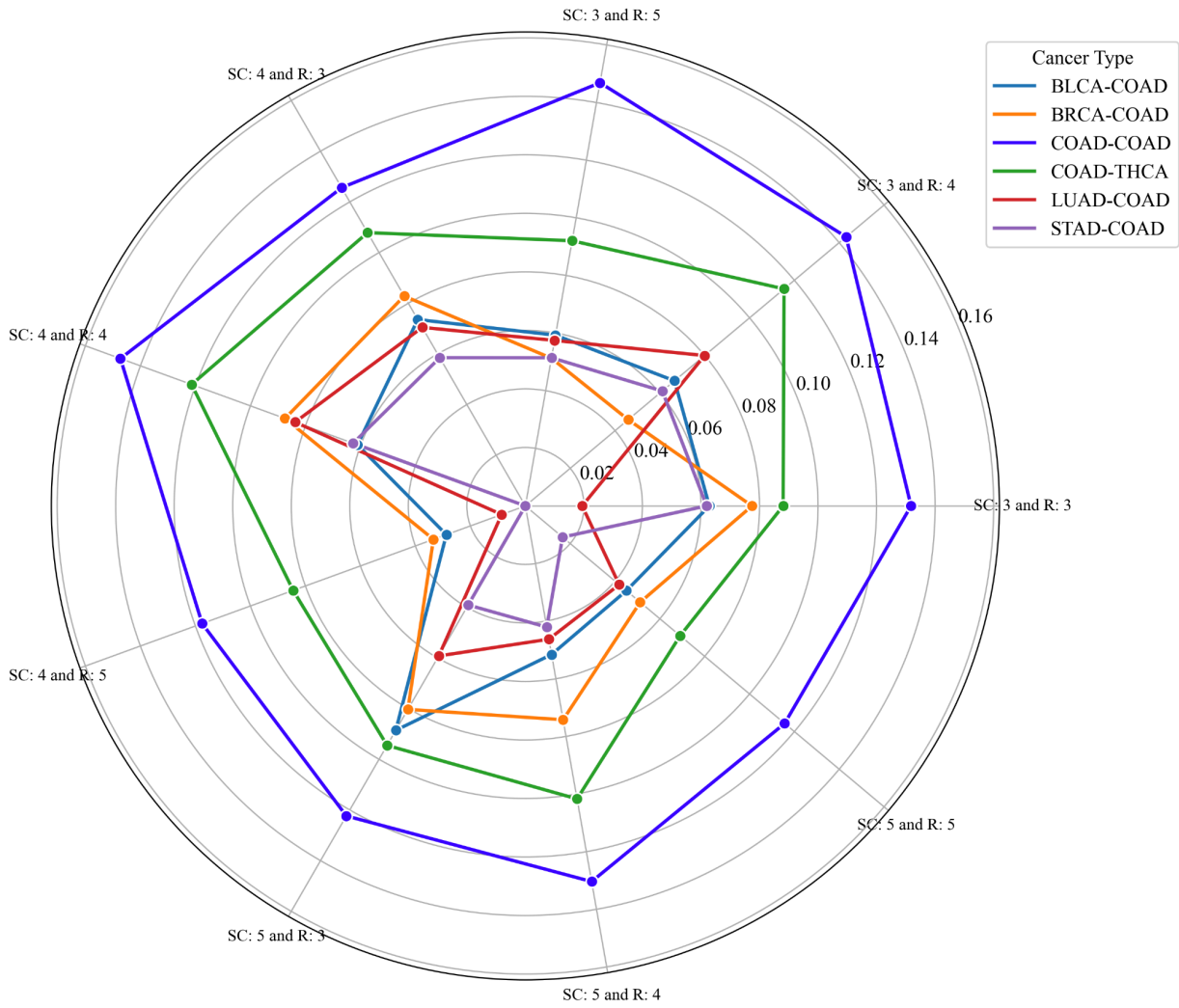
### BRCA - Euclidean Distance



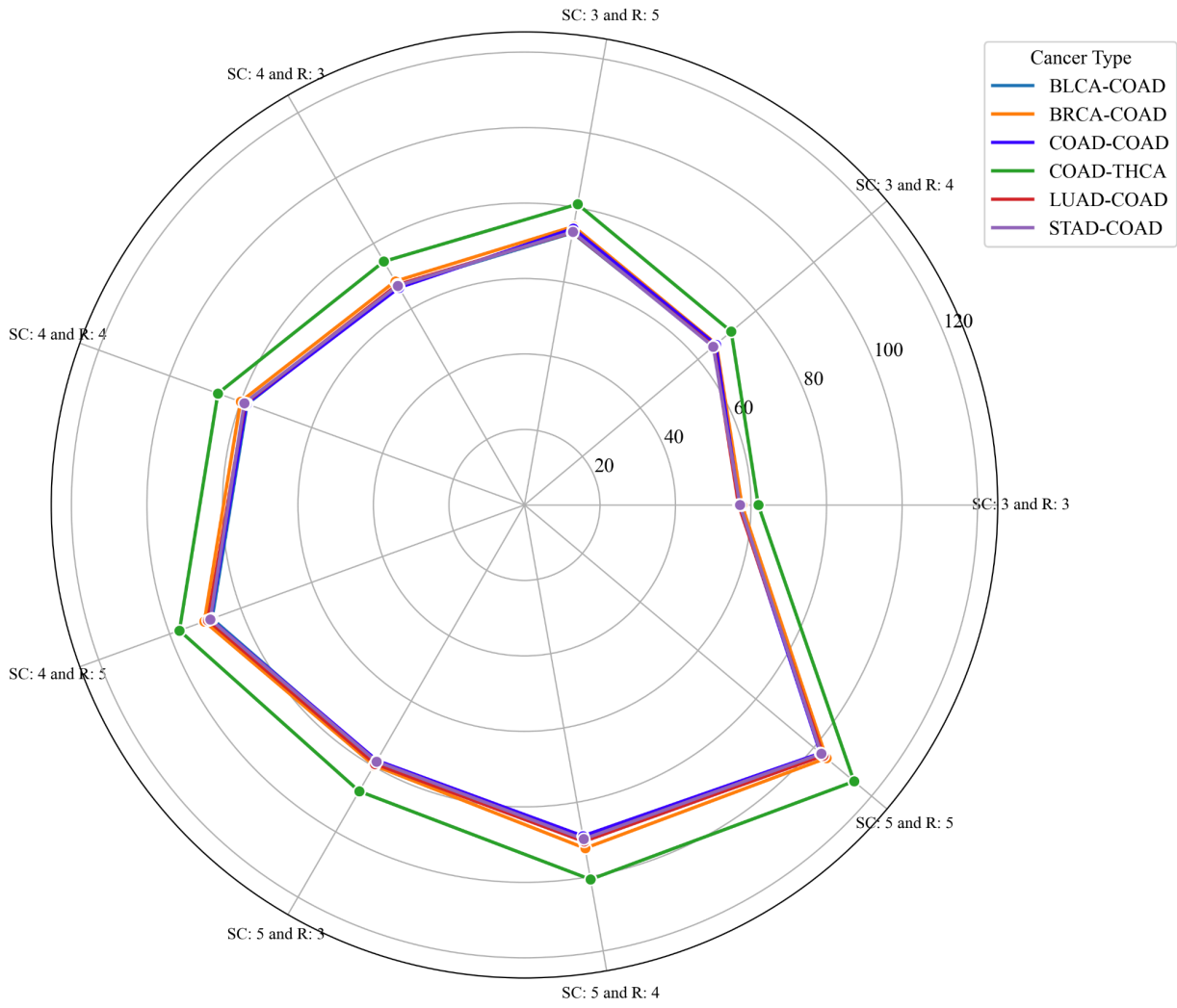
### COAD - Cosine



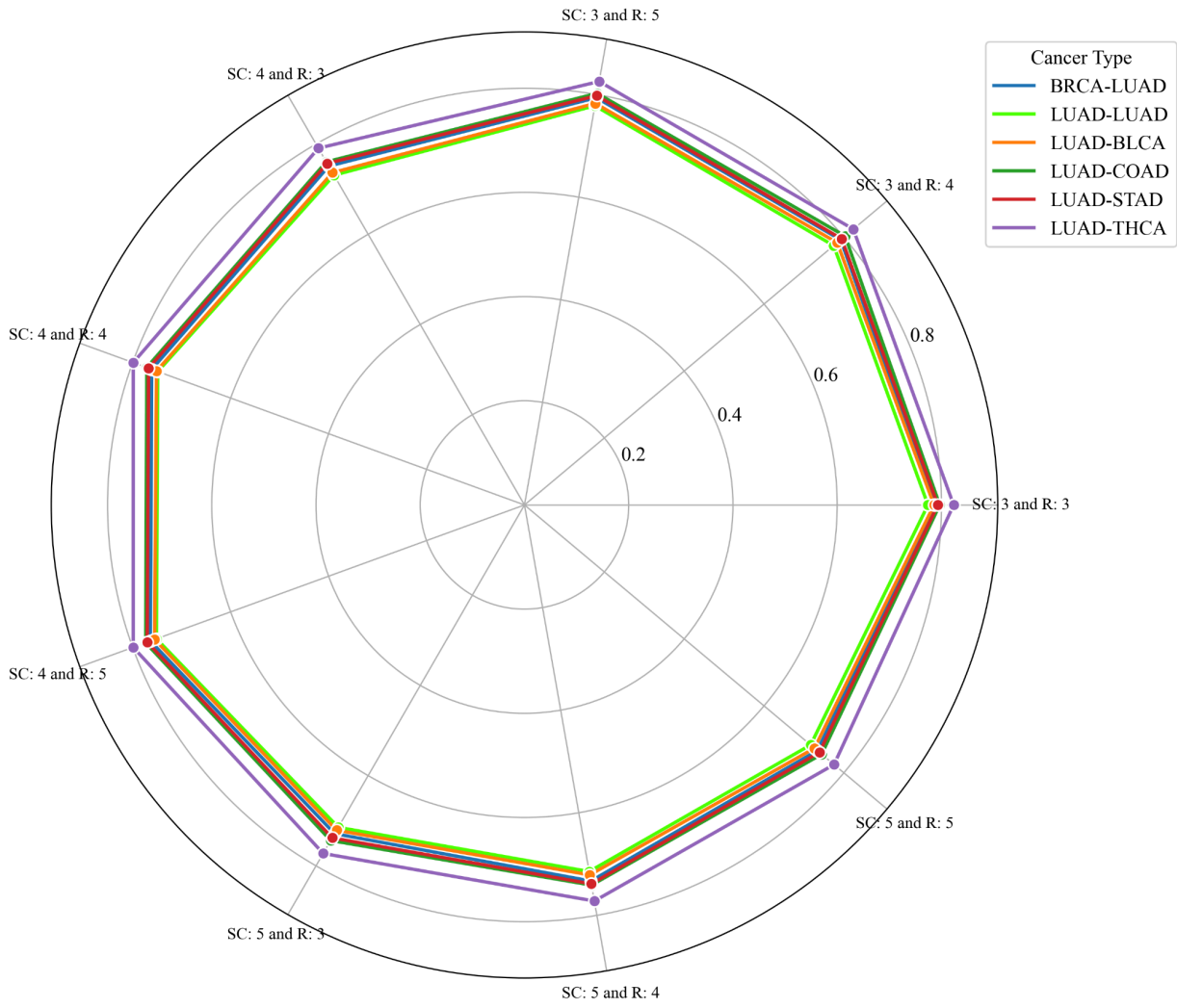
### COAD - Dot Product



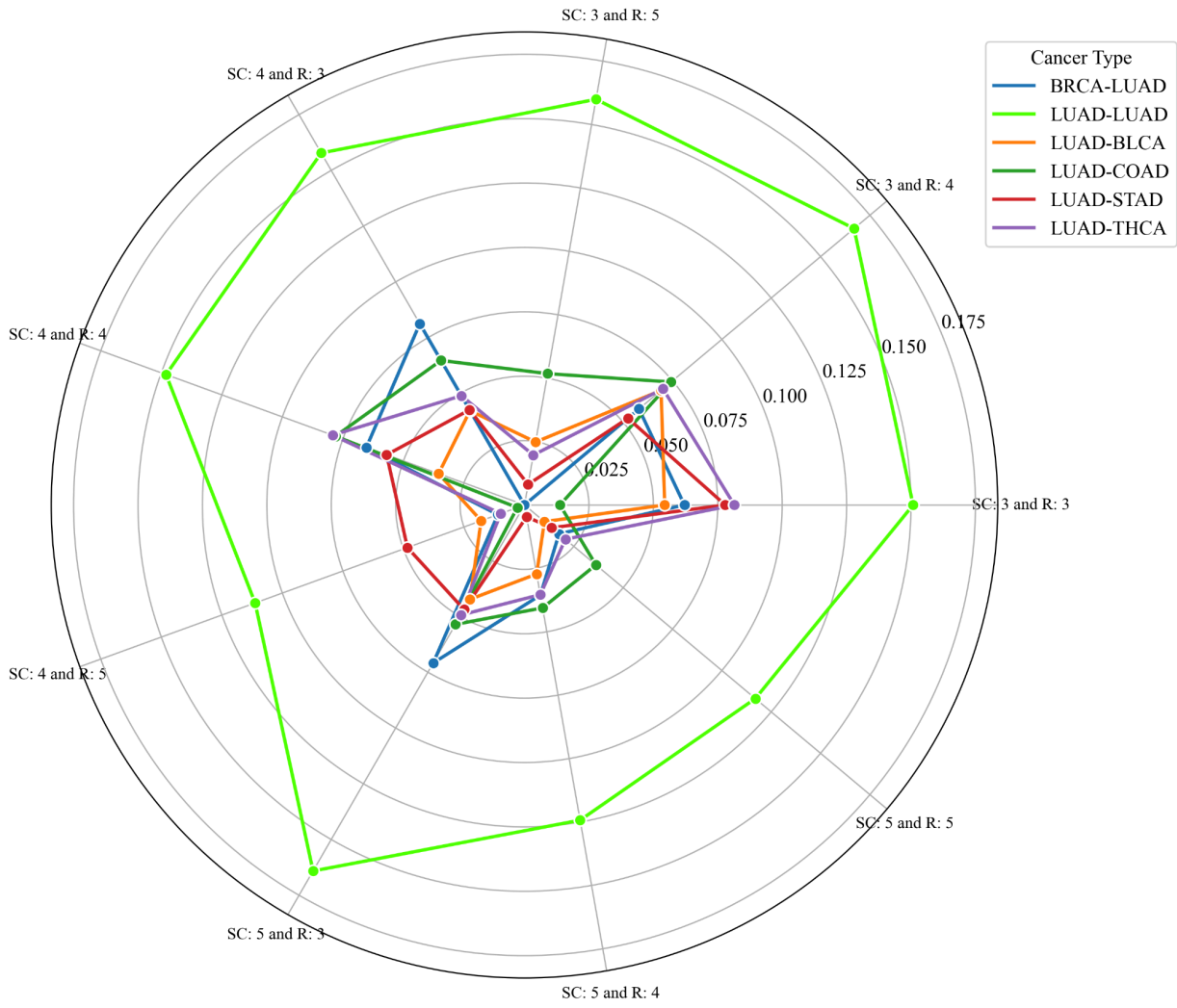
### COAD - Euclidean Distance



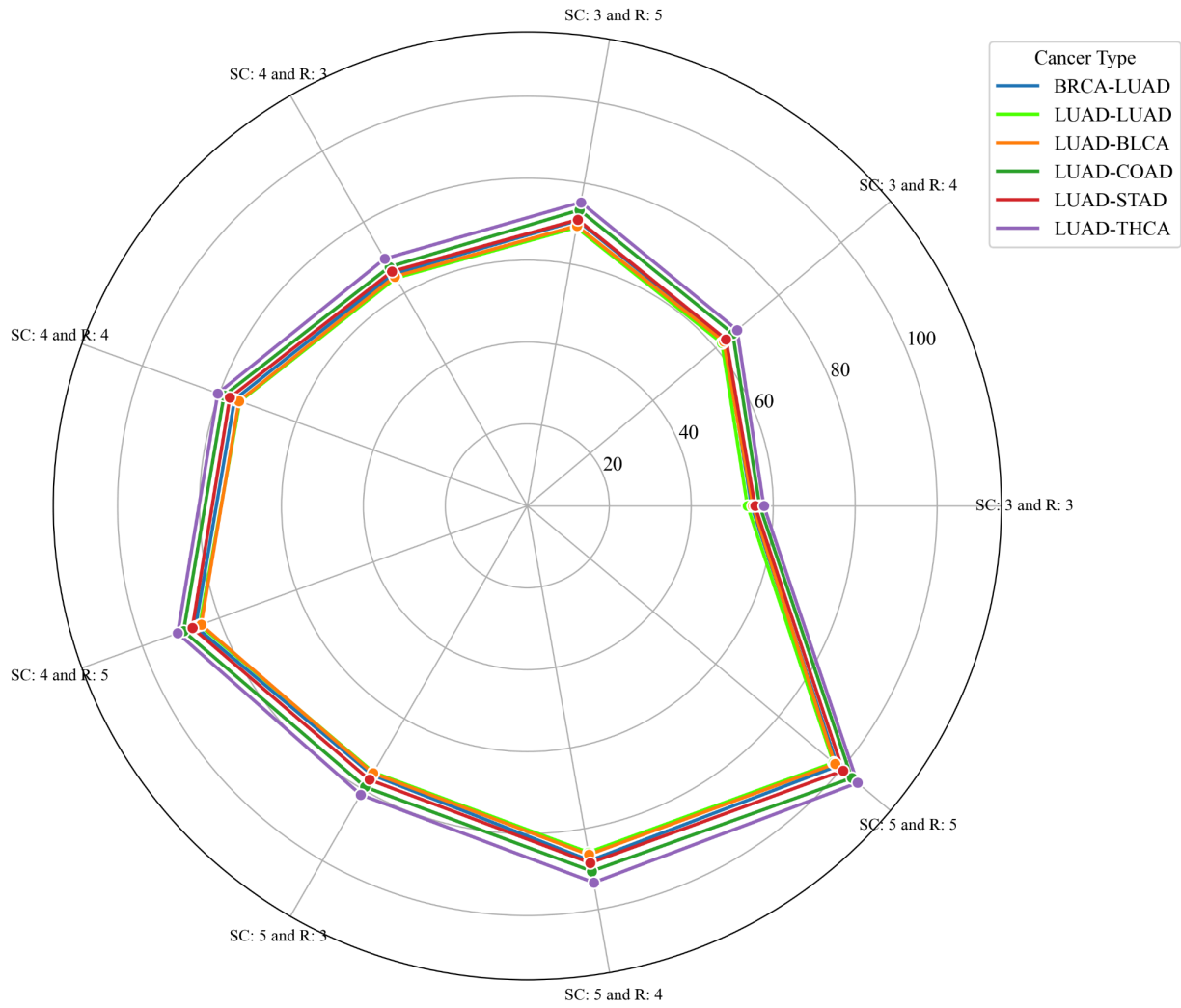
### LUAD - Cosine



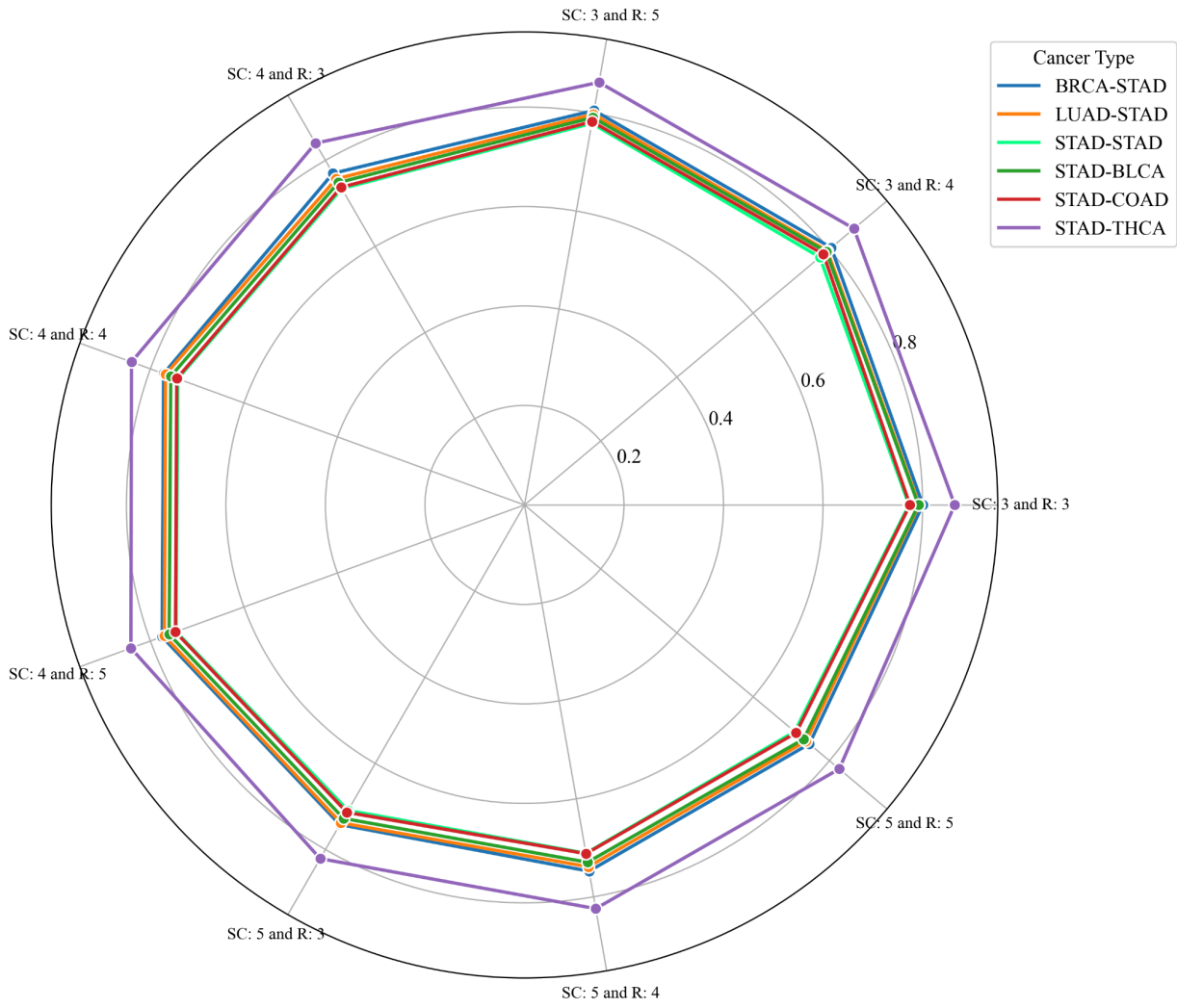
### LUAD - Dot Product



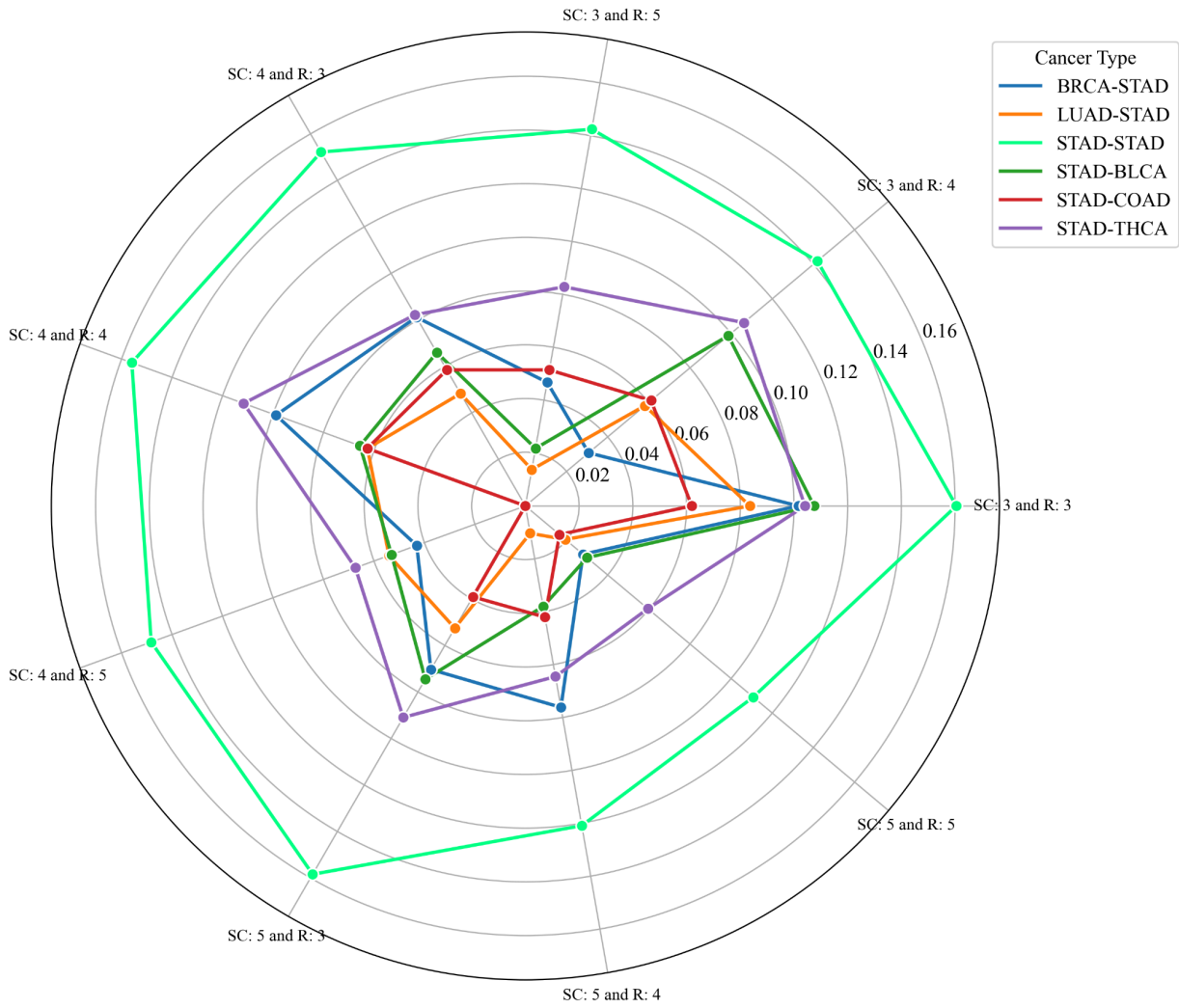
### LUAD - Euclidean Distance



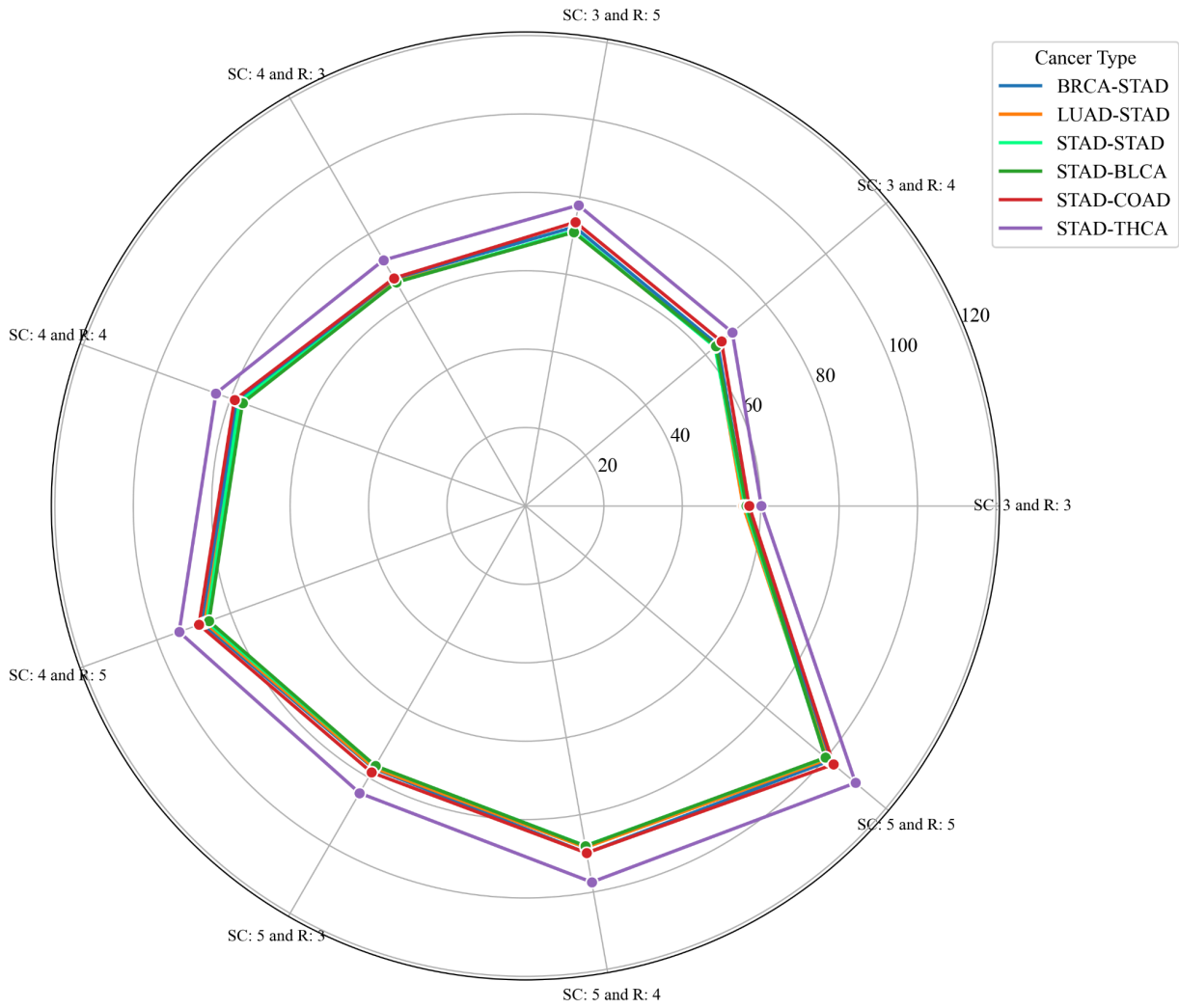
### STAD - Cosine



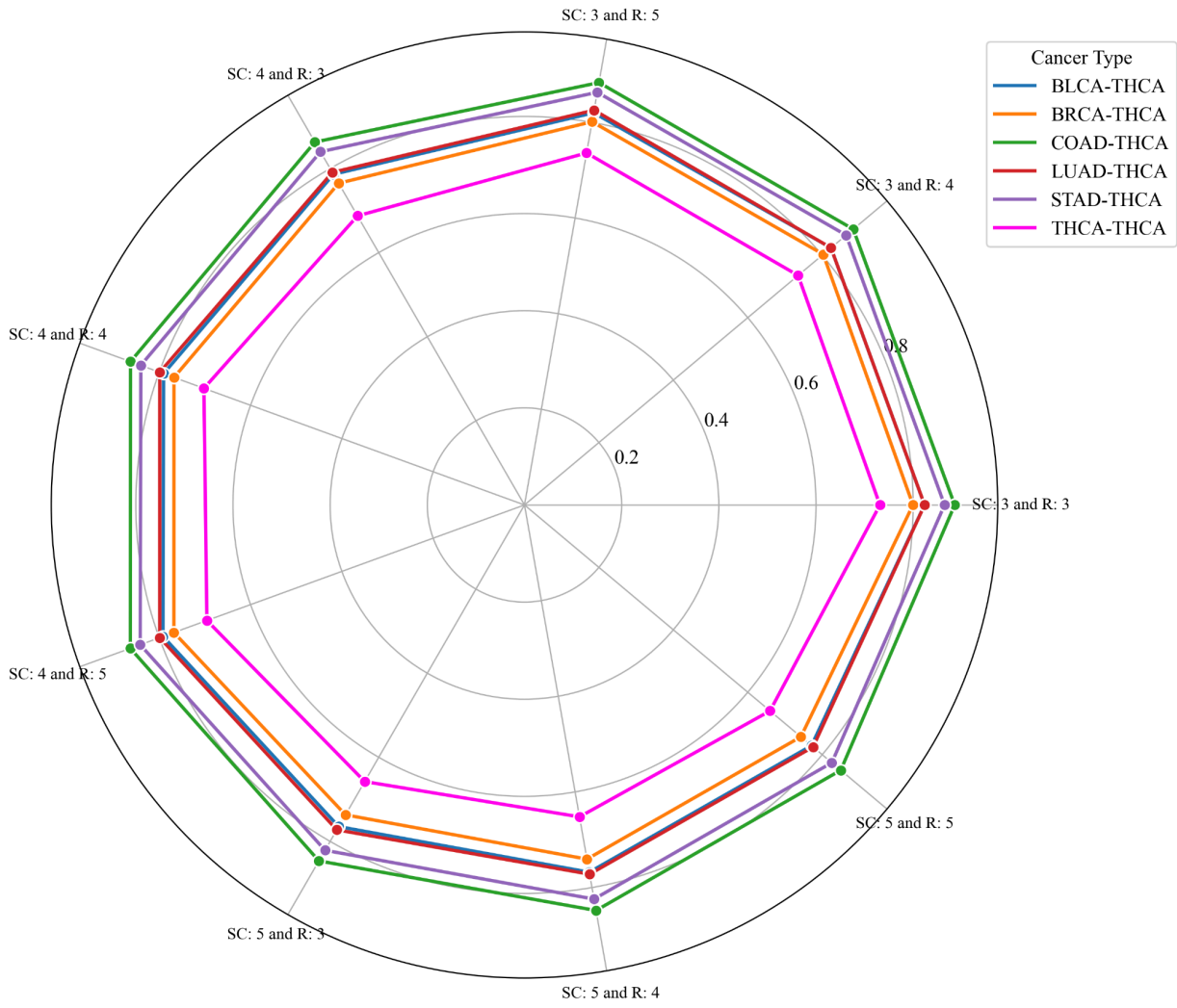
### STAD - Dot Product



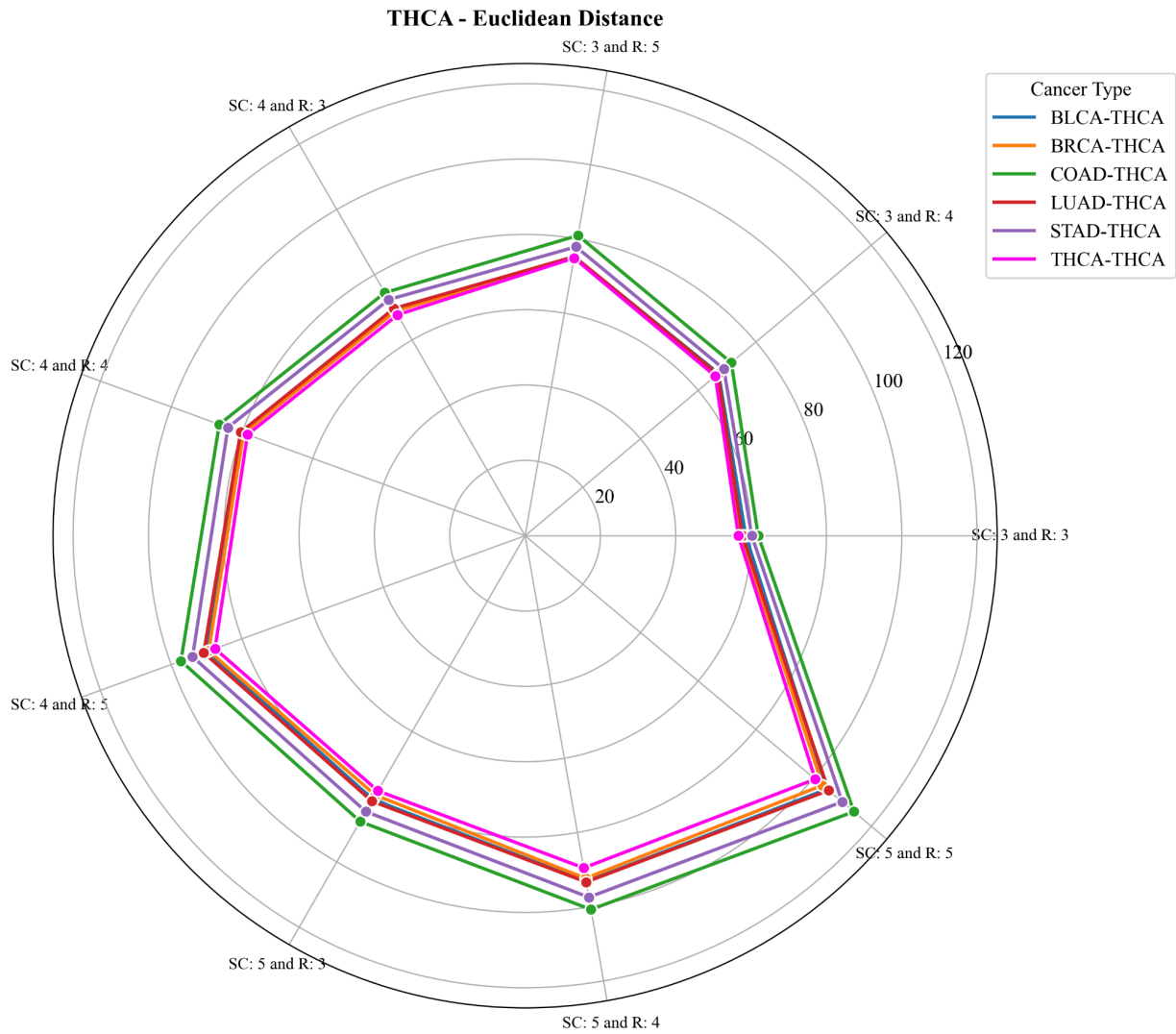
### STAD - Euclidean Distance



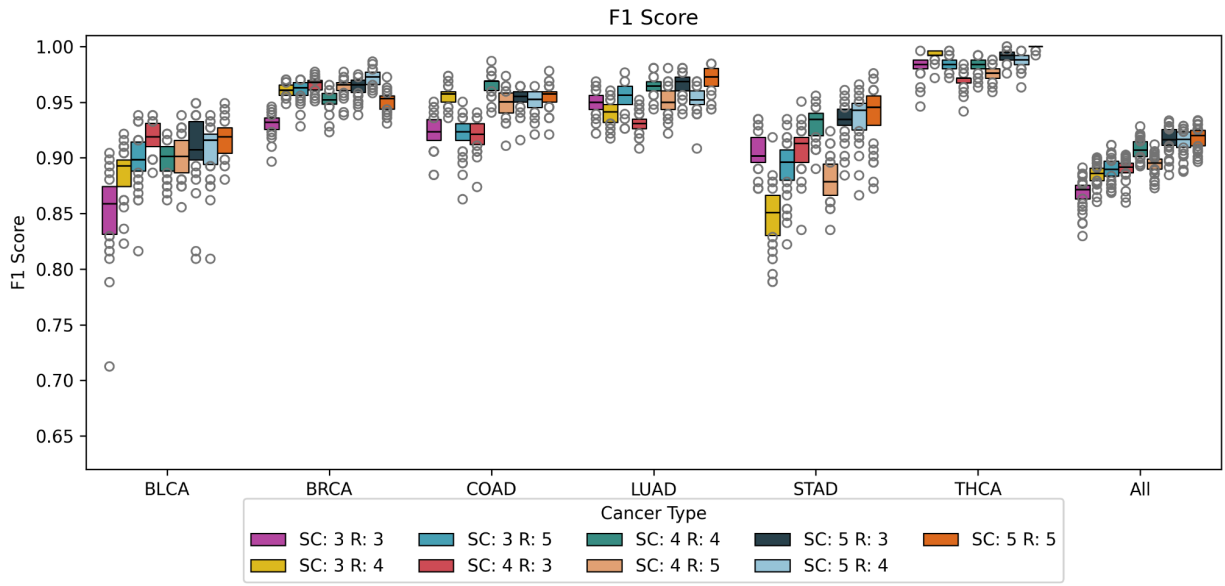
### THCA - Cosine



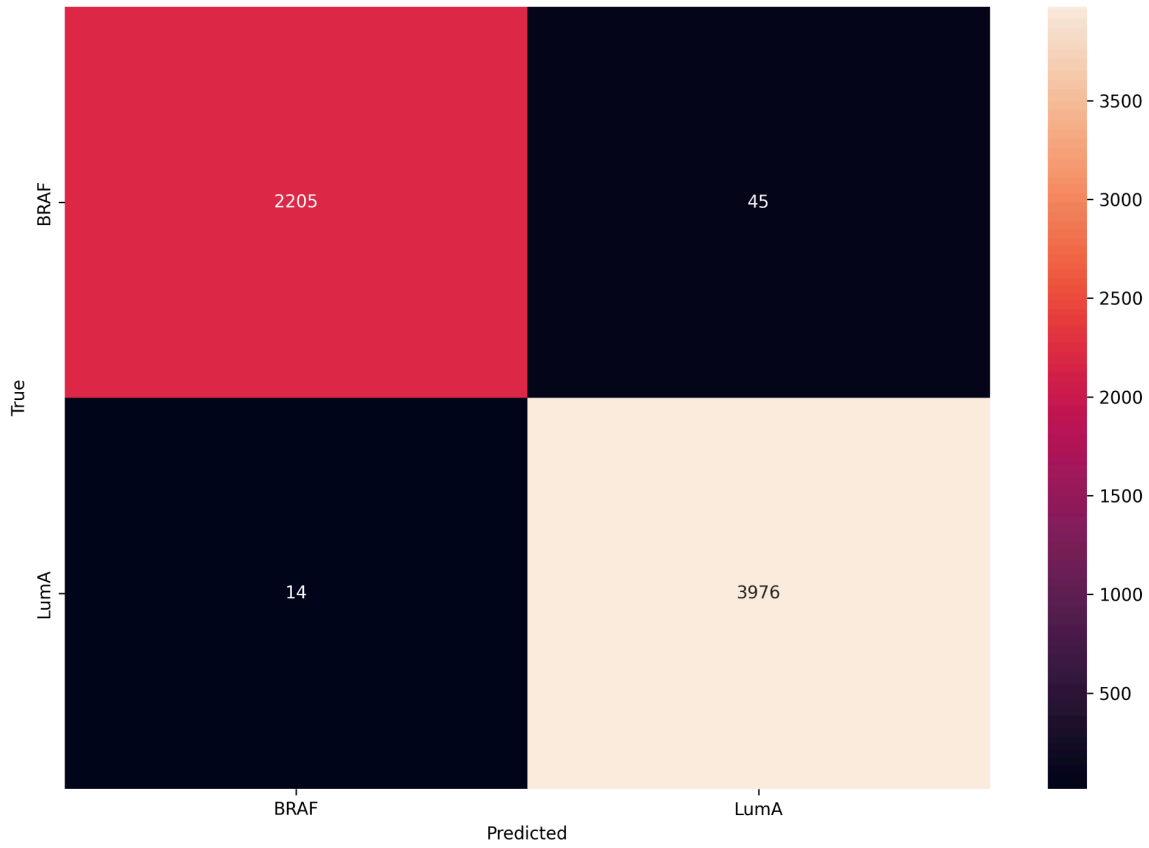




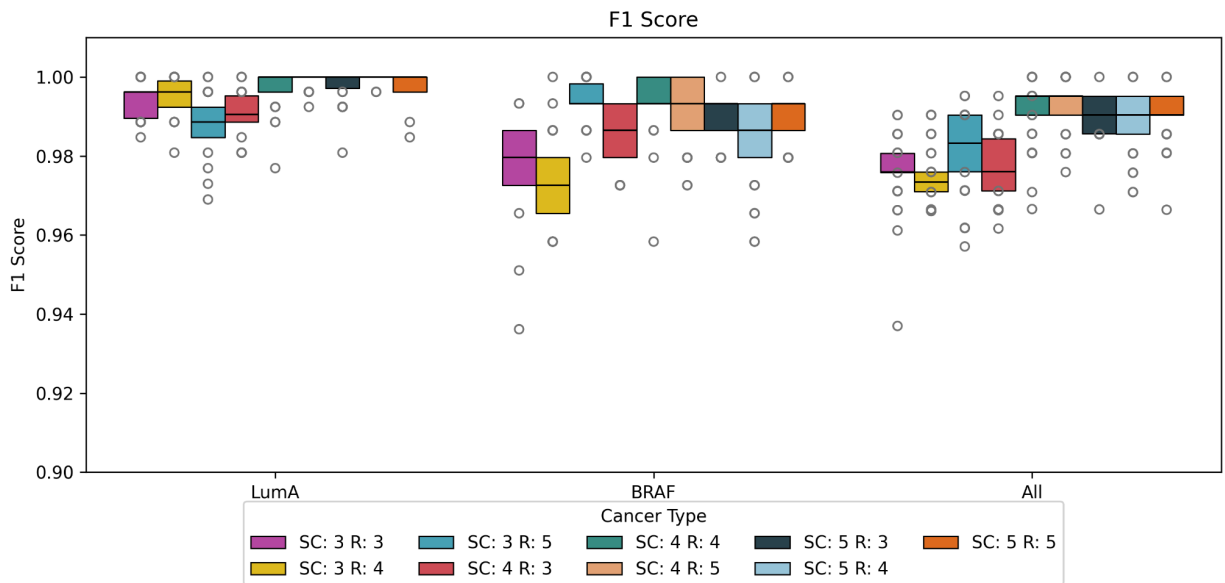
**Supplemental Figure 4:** Pairwise distance calculations between cancer types using Euclidean distance, cosine similarity, and dot product metrics demonstrate clear separation between intra- and inter-cancer distances. Cosine similarity and Euclidean distance are normalized between 0 and 1, whereas the dot product is only bounded at 0 with no defined upper limit.



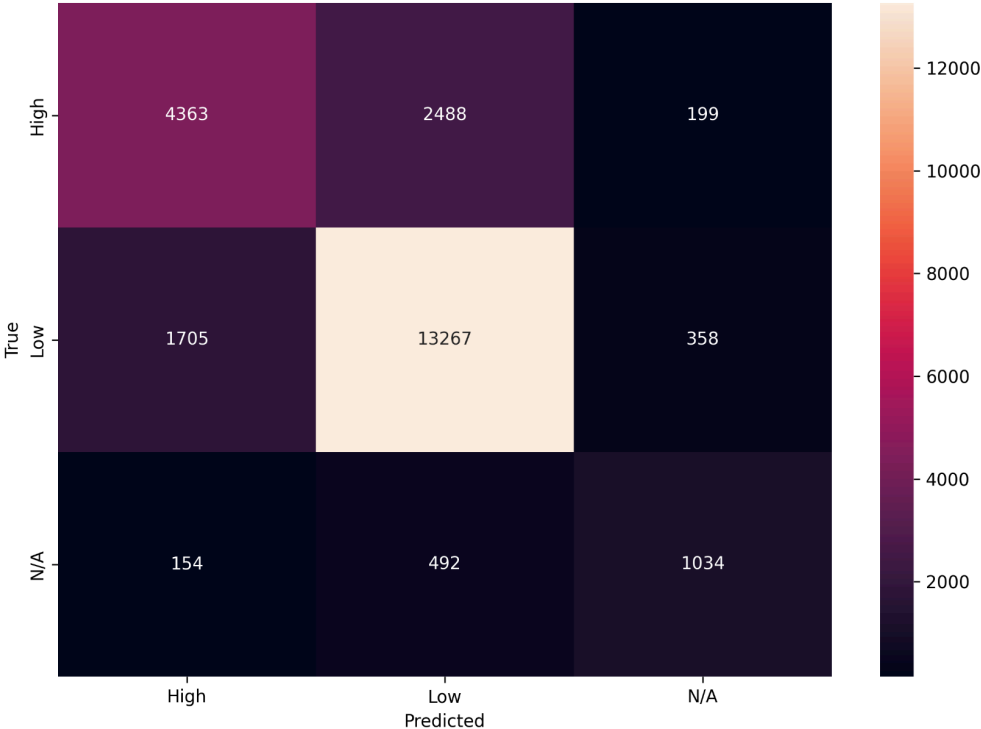
**Supplemental Figure 5:** F1 scores depicting classification performance across six different cancer types, evaluated over varying sample counts (SC) and repetitions (R), indicate that increasing either sample count or repetition enhances classification performance across all cancer types. The overall average F1 score across all cancers ranges between 0.90 and 0.95, demonstrating robust and consistent performance.



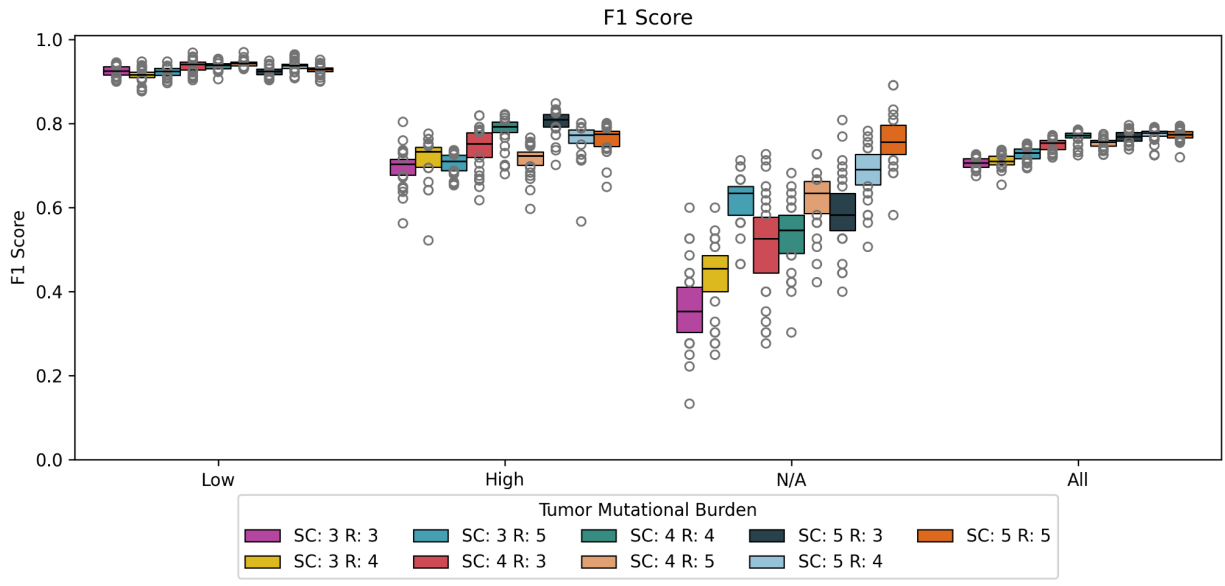
**Supplemental Figure 6:** Confusion matrix illustrating true versus predicted classifications for Luminal A (LumA) and BRAF cancer subtypes. Misclassifications are minimal, with a higher occurrence of BRAF samples being incorrectly classified as LumA.



**Supplemental Figure 7:** F1 scores for subtype classification of Luminal A (LumA) and BRAF compared with overall F1 performance across varying sample counts (SC) and repeats (R), ranging from 3 to 5. Performance remains consistently high, with F1 scores  $\geq 0.9$  across all conditions.



**Supplemental Figure 8:** Confusion matrix depicting true versus predicted classifications of Low, High, and N/A tumor mutational burden (TMB). Elevated misclassifications are observed for the N/A class, predominantly predicted as Low TMB. Additionally, minor misclassifications are noted for the N/A class when predicted as High TMB.



**Supplemental Figure 9:** F1 scores for Low, High, and N/A tumor mutational burden (TMB) classes, along with an overall average across all classes, are evaluated over varying sample counts (SC) and repetitions (R). The Low TMB class consistently achieves high performance with F1 scores  $\geq 0.9$ , while the High TMB class maintains moderate performance ranging between 0.7 and 0.8. Notably, the N/A class exhibits poor performance with F1 scores  $< 0.4$  when using a sample count of 3 and repetition of 3. However, increasing the sample count and repetition progressively improves N/A class performance, reaching an F1 score of approximately 0.8.