Title: Molecular subtypes of high grade serous ovarian cancer across racial groups and gene expression platforms

Authors: Natalie R. Davidson*[1], Mollie E. Barnard*[2,3], Ariel A. Hippen[4], Amy Campbell[4], Courtney Johnson[5], Gregory P. Way[1,4], Brian K. Dalley[6], Andrew Berchuck[7], Lucas A. Salas[8], Lauren C. Peres[9], Jeffrey R. Marks[10], Joellen M. Schildkraut[†5], Casey S. Greene[†1,4], Jennifer A. Doherty[†2,8]

*Denotes equal contributions
[†]Denotes equal contributions

Author affiliations:
[1] Department of Biomedical Informatics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[2] Huntsman Cancer Institute and the Department of Population Health Sciences at the Spencer Fox Eccles School of Medicine, University of Utah, Salt Lake City, UT, USA
[3] Slone Epidemiology Center, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA
[4] Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[5] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[6] Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
[7] Department of Obstetrics and Gynecology, Duke University, Durham, NC
[8] Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA
[9] Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[10] Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA

Keywords: epithelial ovarian cancer, racial disparities, subtype assignment, gene expression clustering, cancer survival

Funding statement: This work was supported by the National Cancer Institute (NCI) of the National Institutes of Health (R01 CA200854 to JAD and JMS, R01 CA142081 to JMS, R01 CA076016 to JMS, R01 CA237170 to CSG and JAD, K99/R00CA218681 to LCP). Research reported in this publication utilized the High-Throughput Genomics and Cancer Bioinformatics Shared Resource at Huntsman Cancer Institute at the University of Utah and was supported by the National Cancer Institute of the National Institutes of Health under Award Number P30CA042014. Dr. Barnard received support from the National Cancer Institute (K00 CA212222) and the Karin Grunebaum Cancer Research Foundation. The content of the manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Conflict of interest declaration: LCP received grant funding from Bristol Myers Squibb for work external to this manuscript. The remaining authors declare no potential conflicts of interest.

Corresponding authors:

1. Mollie E. Barnard, ScD
   72 E. Concord St.
   Boston, MA, 02118
   T: (617) 206-6142

F: (617) 738-5119
barnard@bu.edu

2. Natalie R. Davidson, PhD
Department of Biomedical Informatics
University of Colorado
Anschutz Medical Campus
1890 N Revere Ct, Aurora, CO
natalie.davidson@cuanschutz.edu

Abstract

Introduction: Black individuals with ovarian cancer experience poorer survival compared to non-Hispanic White individuals. High-grade serous carcinoma (HGSC) is the most common ovarian cancer histotype, with gene expression subtypes that are associated with differential survival. We characterized HGSC gene expression in Black individuals and considered whether gene expression differences by race may contribute to disparities in survival.

Methods: We performed gene expression analyses using RNA-Seq data from Black and White individuals with HGSC, as well as array-based genotyping data from four existing studies of HGSC. Our main analysis assigned subtypes by identifying dataset-specific clusters using K-means clustering for K=2-4. The cluster- and dataset-specific gene expression patterns were summarized by moderated t-scores that differentiate an individual cluster from all other clusters within each dataset. We compared the calculated gene expression patterns for each cluster across datasets by calculating the Pearson correlation between the two summarized vectors of moderated t-scores. Following K=4 subtype assignment and mapping to The Cancer Genome Atlas (TCGA)-derived HGSC subtypes, we used multivariable-adjusted Cox proportional hazards models to estimate subtype-specific survival separately for each dataset.

Results: When we compared HGSC gene expression between RNA-Seq data and array-based gene expression data, cluster-specific gene expression was similar across data types. Comparing the Black study population to White and Japanese study populations, the immunoreactive subtype was more common (39% versus 23%-28%) and the differentiated subtype less common (7% versus 22%-31%). Patterns of subtype-specific survival were similar between the Black and White populations with RNAseq data; when compared to mesenchymal HGSC, the risk of death was similar for proliferative and differentiated cases and lower, but not statistically significantly lower, for immunoreactive cases (HR=0.79 [0.55, 1.13] in the Black population and HR=0.86 [0.62, 1.19] in the White population). Patterns of subtype-specific survival were also similar in a Japanese population and in two of three predominantly White populations with array-based gene expression data.

Conclusions: A single pipeline can be used to assign gene expression subtypes to HGSC with array-based or RNA-Seq gene expression data. We observed that while the prevalence of the immunoreactive subtype was higher, and the prevalence of the differentiated subtype was lower among Black individuals compared to White individuals, subtype-specific survival was similar.

Introduction

Ovarian cancer is a highly fatal malignancy comprised of multiple histologically-defined subtypes (i.e., “histotypes”). High-grade serous carcinoma (HGSC) is the most common histotype (Phung et al., 2023), and

an important contributor to ovarian cancer mortality (Fortner et al., 2023; Gaitskell et al., 2022; Peres et al., 2019). HGSC is also molecularly heterogeneous, which is relevant to both prognosis and treatment. Prior studies have described between three and five molecularly distinct subtypes (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014; Tan et al., 2013; Tothill et al., 2008; Verhaak et al., 2013; Way et al., 2016), and, while no gold standard exists for defining these subtypes, they are commonly mapped to the four TCGA-derived subtypes, which are similar to those reported in Tothill et al., 2008: mesenchymal (Tothill C1.MES), proliferative (Tothill C5.PRO), immunoreactive (Tothill C2.IMM), and differentiated (Tothill C4.DIF) (Cancer Genome Atlas Research Network, 2011; Talhouk et al., 2020). Key characteristics of the mesenchymal subtype include high expression of HOX genes, increased stromal components, and poor survival (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014; Talhouk et al., 2020). The proliferative subtype has been characterized by low expression of ovarian tumor markers (e.g., MUC1, MUC16), high expression of transcription factors, and intermediate survival (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014; Talhouk et al., 2020). Defining characteristics of the immunoreactive subtype include the enrichment of genes and pathways associated with an immune response, including CD3+/CD8+ T-cell markers and genes in the CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation, and more favorable survival (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014; Talhouk et al., 2020). The differentiated subtype has been the most difficult to characterize and reproduce (Wang et al., 2017; Way et al., 2016), yet multiple studies have described it as having high expression of ovarian tumor markers (e.g., MUC1, MUC16, SLPI) and intermediate to good survival (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014).

Since HGSC gene expression-based subtypes have potential clinical utility, it is of interest to develop a clinical-grade, gene expression-based subtype classifier that is applicable across diverse populations. One recently developed gene-set assay, the Predictor of high-grade-serous Ovarian carcinoma molecular subTYPE (PrOTYPE) assay, was derived using array-based gene expression data from mostly non-Hispanic White individuals and migrated to a NanoString platform (Talhouk et al., 2020). The PrOTYPE assay has been reported to classify HGSC into the four TCGA-derived subtypes with >95% accuracy (Talhouk et al., 2020). Here, we leverage RNA-sequencing (RNA-seq) data to characterize HGSC gene expression subtypes among a more racially diverse study population, including a population of >300 self-identified Black individuals. Prior studies have noted that Black individuals have poorer overall ovarian cancer survival (Siegel et al., 2022), and poorer three-year and six-year HGSC survival when compared to White individuals (Peres et al., 2020; Schildkraut et al., 2023), so inclusion of Black individuals in the derivation of HGSC molecular subtypes is especially important.

In the present study, we used K-means clustering to assign HGSC tumors from Black, White, and Japanese individuals to molecular subtypes (Way et al., 2016). Following subtype assignment, we compared subtype-specific gene expression, subtype frequency, and subtype-specific survival across all possible pairs of studies with different racial distributions (e.g., Black, White, Japanese), and different data types (array-based gene expression data, and RNA-seq).

Methods

Primary study population

We included epithelial ovarian cancer cases enrolled in one of two population-based case-control studies, the North Carolina Ovarian Cancer Study (NCOCS, diagnosis dates 1999-2005) (Moorman et al., 2002), and the African American Cancer Epidemiology Study (AACES, diagnosis dates 2010-2015) (Schildkraut et al., 2013, 2023). Both of these studies enrolled epithelial ovarian cancer cases covering a range of histotypes, grades, and stages, though some of the most aggressive epithelial ovarian cancer cases were missed because they

were not feeling well enough or were already deceased by the time they were invited to participate in research (Schildkraut et al., 2023). Written informed consent was obtained for NCOCS participants, while AACES participants provided verbal consent and signed medical record and pathology release forms to allow for access to tumor tissue. All cases were confirmed via centralized pathology review. Both the NCOCS and AACES studies were approved by the Duke Medical Center Institutional Review Board (IRB) and the IRBs of participating enrollment sites.

Together, the AACES and NCOCS studies included 747 Black ovarian cancer cases, 464 of which were HGSC. Of the Black HGSC cases, 325 provided consent to participate in biospecimen-based research and had adequate tissue available to pursue RNA extraction (**Supplemental Figure 1A**). Fifty-three of these cases were excluded from gene expression analyses due to a history of neoadjuvant chemotherapy, which can influence observed gene expression (**Supplemental Figure 1B**). Following these exclusions, there were 272 self-identified Black or African American cases (3 Hispanic; 269 non-Hispanic) who we subsequently refer to as “SchildkrautB”. The NCOCS study included 1,014 White cases, 484 of which were HGSC. Of these HGSC cases, 316 provided consent to participate in biospecimen-based research and had sufficient tissue available to pursue RNA extraction (**Supplemental Figure 2A**). None of these cases had neoadjuvant chemotherapy prior to tissue collection, so all were considered eligible for gene expression analyses. We subsequently refer to this set of 316 non-Hispanic White cases as “SchildkrautW”.

Demographic characteristics, disease characteristics, and vital status were available for all HGSC cases included in SchildkrautB and SchildkrautW. Information on age at diagnosis was obtained from questionnaires and pathology reports. Tumor stage, debulking status, and use of neoadjuvant chemotherapy were abstracted from medical records and pathology reports. The proportions of intratumoral CD3+ T cells and CD3+/CD8+ suppressor T cells were obtained from multiplex immunofluorescence staining of formalin-fixed paraffin-embedded (FFPE) tissue (Peres et al., 2022). Vital status was assessed using data from population-based cancer registries, obituaries, LexisNexis, and the National Death Index.

Acquisition of gene expression data

RNA was extracted from FFPE tumor tissue and stored at -80°C. An initial quality control (QC) evaluation revealed substantial RNA degradation, so a repurification step consisting of DNAase treatment and purification on a Zymo research spin column was completed before library preparation to reduce the bulk of degraded RNA product (i.e., RNA product <200 nucleotides in length). Following repurification, RNA libraries were prepared from total RNA samples (5-100 ng) using reagents from the Illumina Stranded mRNA Prep (cat# 20020189) and the Illumina RNA UD Indexes Set (20091657) for reverse transcription, adapter ligation, and PCR amplification. Amplified libraries were hybridized to biotin-labeled probes from the Illumina Exome Panel (cat# 20020183) using the Illumina RNA Fast Hyb Enrichment kit (20040540) to generate strand-specific libraries enriched for coding regions of the transcriptome. The quality of exon-enriched libraries was assessed on an Agilent Technologies 2200 TapeStation using a D1000 ScreenTape assay (cat# 5067-5582 and 5067-5583). The molarity of adapter-modified molecules was defined by quantitative PCR using the Kapa Biosystems Kapa Library Quant Kit (cat#KK4824). Individual libraries were normalized to 0.95 nM in preparation for Illumina sequence analysis. Sequencing libraries were chemically denatured and applied to an Illumina NovaSeq flow cell using the NovaSeq XP workflow (20043131). Following the transfer of the flow cell to an Illumina NovaSeq 6000 instrument, a 150 x 150 cycle paired-end sequence run was performed using a NovaSeq 6000 S4 reagent Kit v1.5 (20028312).

Quantification of gene expression data

We trimmed adapters and filtered read quality using fastp (S. Chen et al., 2018). We filtered to reads with a PHRED score of at least 15 and a length of at least 20 base pairs. While a minimum PHRED score of 15 may include some reads with low quality, we found that most bases across all samples in SchildkrautB and SchildkrautW had a quality score greater than 30 (**Supplemental Figure 3**). We quantified paired-end reads with Salmon (version 1.4.0) (Patro et al., 2017) using GRCh38 release 95. We used the seqBias and gcBias flags to correct for sequence-specific biases. We also used the recommended rangeFactorizationBins parameter value 4, which improves quantification accuracy on difficult-to-quantify transcripts. We then filtered out low-expression genes by excluding genes with a median expression of 0 within a dataset (SchildkrautB: 10,620 genes removed, SchildkrautW: 10,410 genes removed). We library-size normalized samples using upper quantile normalization. This normalization matches the 85th percentile across samples to correct for library size differences across samples.

Overview of comparator study populations

Consistent with the data processing pipeline used by Way *et al.* (Way et al., 2016), we included data from the TCGA (platform: AffymetrixHT_HG-U133A), Yoshihara (platform: AgilentG4112F), and Tothill (platform: AffymetrixHG-U133Plus2) datasets included in the R package curatedOvarianData (Ganzfried et al., 2013), and additional data from the dataset “Mayo” (GEO Accession: GSE74357, platform: AgilentG4112F). All gene expression data generated from these studies was derived from fresh frozen tumor tissue. Using the R package doppelgangR (Waldron et al., 2016), we identified four more duplicate pairs across studies than found previously (Way et al., 2016) and one pair of duplicate samples in SchildkrautW.

Clustering Methods

We performed clustering analyses by modifying the pipeline from Way *et al.* (Way et al., 2016) for inclusion of RNA-Seq data. Briefly, for K-means clustering, we used the R package cluster (version 2.0.6) (Maechler, 2018), and for non-negative matrix factorization (NMF), we used the R package NMF (version 0.20.6) (Gaujoux & Seoighe, 2010). We identified the top 1,500 genes with the largest Median Absolute Deviance (MAD) from each of the six datasets (SchildkrautB, SchildkrautW, TCGA, Mayo, Yoshihara, and Tothill), then used their union set of 4,355 genes for clustering (**Supplemental Table 1**). Consensus clusters for both K-means and NMF were identified using multiple random starts for each value of K=2-4 (K-means starts: 100; NMF starts: 10). Each dataset was clustered independently using the same set of genes.

We compared the similarity of clusters across datasets and clustering methods using the same approach as Way *et al.* (Way et al., 2016). We used significance analysis of microarray (SAM) (Schwender et al., 2006; Tusher et al., 2001) to get cluster-specific vectors of moderated t-scores; this provided us with gene-specific expression patterns of each cluster within each dataset. We restricted to genes that were assayed and expressed across all datasets, which we denote as “common genes" (**Supplemental Table 2**). Using these 8,360 “common genes", we compared clusters across: (1) datasets, and (2) clustering methodologies by calculating the Pearson correlation between two vectors of moderated t-scores. To use this approach on the RNA-Seq datasets, we log10(x+1) transformed the normalized counts to more closely match the data distribution of the microarray datasets. We performed a simulation study to examine SAM’s performance on log10 transformed counts relative to DESeq2’s performance on raw counts and found comparable performance (**Supplemental Note 1**).

Additionally, we compared the similarity of clusters in a global and aligned principal components analysis (PCA) projection. We filtered the datasets to only include the MAD genes used in our clustering pipeline, then

within-sample scaled the expression values to ensure comparative ranges of expression across all datasets. We log10(x+1) transformed the RNA-Seq expression values before scaling to better match the RNA-Seq and microarray data distributions. After normalization, we concatenated the datasets to calculate a global PCA projection. We then subtracted each dataset's total centroid from each dataset's cluster centroid to align each dataset such that it was centered at the origin of the PCA projection. This allowed us to compare the relative differences between each cluster across all datasets.

Survival analyses by cluster

We generated Kaplan-Meier survival curves to visualize overall survival and subtype-specific survival within each dataset. We also used Cox proportional hazards (PH) models to calculate hazard ratios (HR) and 95% confidence intervals (CIs) quantifying the differences in survival across HGSC subtypes. For five of the six datasets (SchildkrautB, SchildkrautW, TCGA, Mayo, Tothill), Cox PH models were adjusted for three factors that independently influence survival: age (in 5-year age groups), stage (I, II, III and IV), and debulking status (optimal debulking, suboptimal or missing debulking status). Yoshihara *et al.* did not provide information on age at HGSC diagnosis, so models for this dataset were adjusted for stage and debulking status only.

Cluster comparison between pipeline runs and consensusOV

To ensure that our run of the updated Way *et al.* (Way et al., 2016) pipeline was producing similar results to the results from previously published subtype predictors, we compared the assigned cluster labels for each case in the TCGA, Mayo, Yoshihara, and Tothill datasets. To quantify the similarity between pipeline results, we rephrased the cluster comparison problem as a prediction problem. The ground-truth labels were the previously published cluster IDs (Way et al., 2016) or consensusOV (G. M. Chen et al., 2018) predictions for each case, and the predicted labels were our newly assigned cluster labels. consensusOV version 1.16.0 was run using default parameters and no significance filtering. We used balanced accuracy as the metric for comparing current and published labels due to variations in cluster sizes. Balanced accuracy is defined in the following equations.

$$Balanced\ Accuracy = (Sensitivity + Specificity) / 2$$

$$Sensitivity = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Negatives}$$

$$Specificity = \frac{Number\ of\ True\ Negatives}{Number\ of\ True\ Negatives + Number\ of\ False\ Positives}$$

Sensitivity analysis of K-means cluster assignments

K-means is sensitive to outliers and local minima; therefore, we quantified the stability of the cluster labels when using a subsample within each dataset. First, we removed any samples that were more than 1.5x the interquartile range from the first or third quartiles in the first five PCs. Using the remaining samples, we subsampled 80% of the samples within each dataset and re-ran the K-means clustering. We then matched the new cluster labels to the original cluster labels using a greedy approach. We matched labels by finding the cluster IDs with the highest sample overlap. We did this iteratively, removing each cluster from consideration once its matched cluster ID was assigned.

Code and data availability

We provide all software under the BSD 3-Clause License. Our code contains scripts for the analyses completed for this manuscript, including RNA-Seq quantification, quality control, clustering, and figure

generation. To support reproducibility, we provide code to recreate the environments and re-run both the RNA-Seq data analysis (https://github.com/greenelab/hgsc_rnaseq_cluster/) and our updated clustering pipeline (https://github.com/greenelab/hgsc_rnaseq_clustering_pipeline/). All derived expression data are publicly available and accessible under the GEO accession code (will be added when paper is published). Raw gene expression data and epidemiologic variables used in this manuscript are available through dbGaP, under accession number (will be added when paper is published).

Results

Data quality

For both SchildkrautB and SchildkrautW, we further excluded samples due to possible technical artifacts (**Supplemental Tables 3 and 4)**. In SchildkrautB and SchildkrautW, 14 and 4 cases were resequenced, respectively, due to insufficient read depth, and the transcriptomic profiles from the first sequencing attempt were excluded. After resequencing, these 18 cases attained a read depth comparable to all other sequenced samples and progressed to the next step of the quality control pipeline (**Supplemental Figures 1B and 2B**). All other samples in both datasets were assessed to be of high quality, with over 83.3% of reads having a base quality of at least 30 (**Supplemental Figure 3**). After normalization, we found that read count distributions were similar across sequencing batches, including samples previously identified by the sequencing core as low-quality (**Supplemental Figure 4C-H)**. However, a small subset of samples had lower-than-expected read counts. To account for this, we removed all samples where the median normalized read count was below 925 or where the bottom 25th quantile read count for a sample was below 30. This additional read count filter removed 10 SchildkrautB cases and 5 SchildkrautW cases from further consideration (**Supplemental Figure 4A-B**). The final technical exclusion for SchildkrautB and SchildkrautW removed samples that were flagged by doppelgangR (Waldron et al., 2016) as having overly similar expression patterns, suggesting they originated from the same tumor. Only SchildkrautW had a pair of samples identified as too similar (NCO0557 and NCO0625).

Following the SchildkrautB and SchildkrautW technical exclusions, we applied the Way pipeline exclusions to the TCGA, Mayo, Yoshihara, and Tothill datasets, and we applied doppelgangR exclusions to all datasets. The sizes of our final analytic data sets were as follows: SchildkrautB (n=262); SchildkrautW (n=309); TCGA (n = 499; phs000178) (Cancer Genome Atlas Research Network, 2011); Mayo (n = 377; GSE74357) (Konecny et al., 2014); Yoshihara (n = 255; GSE32062.GPL6480) (Yoshihara et al., 2012), and Tothill (n = 241; GSE9891) (Tothill et al., 2008).

Validation of updated clustering pipeline

We updated a previously published bioinformatic subtyping pipeline designed specifically for HGSC (Way et al., 2016). The original pipeline identified clusters across four microarray studies that span the United States, Japan, and Australia. While most of our pipeline was the same, we used balanced accuracy to quantify the similarity in output between pipelines. We found that balanced accuracy for our K-means clusters was consistently ≥0.91 across all datasets for K=2-4 (**Supplemental Figure 5**). The limited differences that arose were primarily due to changes in stochastic elements of the clustering methods and software packages, such as differences in computing environments, random seeds, and package versions. Another source of variation between the output of the pipelines was the addition of the Schildkraut datasets since the genes used for clustering must be shared across all datasets and have a high MAD in at least one dataset (**Methods)**. After adding the Schildkraut datasets, 893 genes that were previously denoted as common were reclassified as

MAD genes, 609 MAD genes were re-classified as common genes, and 911 genes were removed from consideration. A Venn diagram of the removed genes is provided in **Supplemental Figure 6**.

Similar to Way *et al.* (Way et al., 2016), we wanted to ensure that clustering outcomes were robust across methods. We compared the clusters from two different clustering methods: K-means and NMF. We varied the number of clusters from 2 to 4 and compared the clusters between methods in each dataset. We derived a gene expression pattern for each cluster using the significance analysis of microarray (SAM) moderated t-score (**Methods**). Using this cluster-specific gene expression pattern, we calculated the Pearson correlation between moderated t-scores for clusters identified by each method. We found extremely high cluster concordance in all datasets between the clusters identified by NMF and K-means when we selected 2 and 3 clusters. We saw diminished concordance between methods when we used four clusters. Clusters in all datasets (except the Mayo dataset for K=4) had the highest concordance with the matched cluster along the diagonal (**Supplemental Figure 7**). Furthermore, since K-means clustering is sensitive to outliers, we additionally re-ran K-means clustering in subsets of 80% of the data and found that, in all datasets, the majority of samples maintained their same cluster label in more than 50% of the subsampled reclusterings (**Supplemental Figure 8)**.

Our final clustering validation was to compare it to another external HGSC subtype classifier, consensusOV (G. M. Chen et al., 2018). consensusOV combines four methods (Cancer Genome Atlas Research Network, 2011; Helland et al., 2011; Konecny et al., 2014; Tothill et al., 2008; Verhaak et al., 2013) into a consensus classifier. Since consensusOV can only assign each sample to one of the four Verhaak-defined clusters, we only compare our K-means results for K=4. We found that our subtype calls were very similar to the consensusOV calls across all microarray datasets, with a minimum per-class balanced accuracy of 0.72 (**Supplemental Table 5**). We saw more discordance between our calls and consensusOV when we compared RNA-Seq samples, with per-class balanced accuracy ranging from 0.589-0.818 (**Supplemental Table 5**).

Gene expression patterns in the self-identified Black study population compared with other study populations

Expression patterns for each cluster were summarized in two ways, first by plotting the principal components of normalized and aligned centroids in each dataset, and second, using SAM-moderated t-scores (Way et al., 2016). In both approaches, we sought to determine if the relative cluster distances for each individual dataset were consistent across datasets. In our principal components (PC) based approach, we normalized and projected all datasets into a shared PC space, then centered them at the origin of the space. For K=3, PC2 and PC3 could separate each cluster (**Figure 1a**). For K=4, PC2 and PC3 could not independently separate each cluster, but, together, could still differentiate them (**Figure 1b**). Furthermore, in the first three PCs, the SchildkrautB dataset did not cluster away from any other datasets, and clustered closely with the SchildkrautW, TCGA, and Tothill datasets.

In our second, complementary approach, we compared differential gene expression between clusters by calculating the Pearson correlation for each cluster's SAM-moderated t-score across pairs of datasets. **Figure 2** compares the cluster-specific gene expression pattern between the SchildkrautB cases and cases from the other five datasets. For K=3, SchildkrautB cluster-specific gene expression was highly correlated with cluster-specific gene expression across all datasets, and showed the strongest correlations with SchildkrautW cluster-specific gene expression. This provides strong evidence that the derived clusters from Black cases are the same as the derived clusters from White and Japanese cases. A high correlation between clusters was also found when performing pairwise comparisons among all six datasets (**Supplemental Figure 9**). Similar to the findings of Way *et al.* (Way et al., 2016), we found that the correlation of cluster-specific gene expression patterns was diminished when using four, as opposed to three, clusters to describe the data. We found this

diminished correlation for four, as opposed to three, clusters when comparing SchildkrautB to each of the five other populations (**Figure 2B vs. Figure 2A**) and when comparing gene expression patterns across all possible pairs of datasets (**Supplemental Figure 10 vs. Supplemental Figure 9**).

Subtype distributions and characteristics by study population

We mapped the K-means clusters with K=4 to the four TCGA-derived HGSC subtypes to compare the frequency of subtypes across datasets and evaluate how cancer characteristics vary by subtype. The immunoreactive subtype was more common (39% vs 23-28%), and the differentiated subtype less common (7% vs 22-31%) in the self-identified Black study population when compared to the White and Japanese study populations (**Figure 3**). In analyses restricted to SchildkrautB and SchildkrautW, FIGO stage differed by HGSC subtype among both Black (p=0.040) and non-Hispanic White individuals (p=0.009); those with differentiated HGSC were more likely to have a lower FIGO stage at diagnosis (**Table 1**). Immune infiltration also differed by HGSC subtype, particularly among Black individuals (p=0.001 for CD3+ T cells; p=0.045 for CD3+/CD8+ suppressor T cells; **Table 1**). In both Black and White individuals, immune infiltration was higher for tumors categorized as mesenchymal or immunoreactive and lower for tumors categorized as proliferative.

Survival patterns for clusters

Overall HGSC survival varied by study (**Supplemental Figure 11,** p-value for test of heterogeneity in a stage-adjusted model <0.001), though patterns of subtype-specific survival were generally similar across studies (**Figure 4**). Multivariable-adjusted hazard ratios and 95% CIs indicated that, when compared to individuals with mesenchymal tumors, those with immunoreactive tumors had better survival in most (SchildkrautB HR=0.79 [0.55, 1.13]; SchildkrautW HR=0.86 [0.62, 1.19]; Mayo HR=0.54 [0.39, 0.75]; Yoshihara HR=0.65 [0.39, 1.09]; Tothill HR=0.54 [0.31, 0.95]), but not all (TCGA HR=1.01 [0.69, 1.49]) study populations (**Table 2**). Meanwhile, the risk of death among those with proliferative HGSC was not statistically significantly different from the risk of death among those with mesenchymal tumors in any study population (**Table 2**), and it was close to the null value of 1.00 in both SchildkrautB (HR=0.98 [0.66, 1.46]) and SchildkrautW (HR=1.05 [0.70, 1.57]). The risk of death among those with differentiated HGSC was not statistically significantly different from the risk of death among those with mesenchymal tumors in five of the six study populations, and was lower in the Mayo study population (HR=0.61 [0.44, 0.84]; **Table 2**).

Discussion

We performed a cross-platform and cross-population analysis of HGSC, including four existing datasets, plus newly generated RNA-Seq gene expression data from 262 Black and 309 White HGSC cases. When comparing RNA-Seq data from the 309 White cases to microarray gene expression data from the predominantly White populations that comprise the Tothill and TCGA study populations, we observed high cluster stability within each dataset when advancing from K=2 to K=4 (**Supplemental Figure 8**) and consistent cluster-specific gene expression profiles across datasets (**Supplemental Figures 9 and 10**). This indicates that similar HGSC gene expression clusters can be defined using combined data from array-based technologies in fresh frozen tissue and RNA-seq technologies in FFPE tissue. We also observed consistent cluster composition (**Figure 1**) and consistent gene expression profiles when comparing Black HGSC cases to White and Japanese HGSC cases (**Figure 2**). This indicates that HGSC gene expression clusters are consistent across Black, White, and Japanese individuals, so it is unlikely that racial differences in HGSC characteristics are a key driver behind poorer HGSC survival in Black populations.

Our interest in determining how an existing HGSC subtype clustering pipeline performs with RNA-Seq data was motivated by the increasing use of RNA-Seq technology to interrogate cancer gene expression (Peng et al., 2015). Previously published HGSC subtype clustering approaches were designed around array-based gene expression data (Bonome et al., 2008; Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014; Tothill et al., 2008; Way et al., 2016; Yoshihara et al., 2012), and while array-based and RNA-Seq technologies observe the same underlying biological processes, the data they produce follow different data distributions (Marioni et al., 2008; Robinson & Smyth, 2007). This is most clear in the methodological differences between array-based and RNA-Seq differential gene expression methods (Anders & Huber, 2010; Lönnstedt & Speed, 2002; Robinson et al., 2010; Smyth, 2004). Here, we demonstrated that, despite the difference in data distributions generated by array-based and RNA-Seq technologies, the Way *et al.* (Way et al., 2016) subtype pipeline built for array-based gene expression data can be applied to normalized and log-transformed RNA-Seq data. The high cluster correlations across all four array-based datasets using fresh frozen tissue and both RNA-Seq datasets using FFPE tissue indicate that the delineation of HGSC subtypes is agnostic to sequencing technology and fresh frozen versus FFPE tissue.

Consistent with previous studies, we were able to assign all cases to one of the four TCGA-derived HGSC subtypes; however, as was previously seen in Way *et al.* (Way et al., 2016), we observed that cluster-specific gene expression was more concordant across datasets when clusters were assigned using K=2 or K=3 compared to K=4. This was most evident when comparing each dataset's cluster-specific expression profiles to the cluster-specific expression profiles across all of the other datasets (**Figure 2, Supplemental Figures 9 and 10**). When considering K=3, the largest off-diagonal correlation we observed over all datasets was 0.07 (**Supplemental Figure 9**). In contrast, when comparing gene expression across clusters for K=4, we found much larger positive off-diagonal correlations for the clusters observed in the array-based TCGA, Mayo, Yoshihara, and Tothill datasets, with the largest per-dataset correlations ranging from 0.19 to 0.33 (**Supplemental Figure 10**). We also observed large positive off-diagonal correlations for the RNA-Seq-based SchildkrautB and SchildkrautW datasets, with the largest per-dataset correlations ranging from 0.17 to 0.34 (**Supplemental Figure 10**). Since cluster-specific gene expression was more concordant when K=2 or K=3 versus K=4 for all datasets, we encourage future studies to test whether a number other than four best represents HGSC subtypes. Our results are consistent with a model where either two HGSC gene expression axes (e.g., mesenchymal-like and immune) or a set of three HGSC subtypes (e.g., as derived using K=3) may more effectively describe the biological variation in HGSC gene expression than the four HGSC subtypes most commonly seen in the literature.

Beyond methodological advances, our work also provides strong evidence that HGSC subtypes can be reproduced among Black HGSC cases and that Black, White, and Japanese HGSC cases share similar subtype-specific gene expression profiles. At increasing values of K, we observed patterns in cluster composition among Black HGSC cases that were consistent with those observed among White and Japanese HGSC cases (**Figure 1, Supplemental Figure 8**). Further, we observed strong correlations between cluster-specific gene expression in Black cases and cluster-specific gene expression in White and Japanese cases, especially when comparing clusters defined using K=2 and K=3 (**Figure 2**). Patterns of subtype-specific survival were also generally consistent across populations. When compared to mesenchymal HGSC, the risk of death was similar for proliferative and differentiated cases, and lower, but not statistically significantly lower, for immunoreactive cases both in SchildkrautB (HR=0.79 [0.55, 1.13]) and SchildkrautW (HR=0.86 [0.62, 1.19], **Table 2, Figure 4**).

The primary difference we observed when comparing HGSC subtypes in Black individuals (i.e., cases in SchildkrautB) to HGSC subtypes in all other study populations was that more Black HGSC cases had gene expression profiles consistent with the TCGA immunoreactive subtype (39% compared to 23%-28%) and fewer Black HGSC cases had gene expression profiles consistent with the TCGA differentiated subtype (7%

compared to 22%-31%; **Figure 3**). Differences in the Tothill, C1-C6, gene expression subtype signatures for Black (n=29) versus White (n=156) ovarian cancers were observed previously, though in different proportions (Schildkraut et al., 2013), so it is possible that there exists true variation in the proportion of HGSC subtypes for Black versus non-Black HGSC cases. However, it is also possible that variations in study design across the six study populations contributed to the different subtype distributions that we observed. For example, case-control studies like the AACES and NCOCS are unable to enroll cases with rapidly fatal HGSC (Schildkraut et al., 2023). This could have skewed the observed subtype distribution for Black cases toward a greater proportion of less aggressive, immunoreactive tumors, and artificially inflated estimates of overall survival in the SchildkrautB study population (**Supplemental Figure 11**), as has been posited previously (Schildkraut et al., 2023).

A key strength of this study was that we were able to update a previously published subtype clustering pipeline to accept either array-based gene expression data or RNA-Seq data and validate our modifications. Another important strength was our creation of the first large RNA-Seq dataset of HGSC in self-identified Black cases, consisting of 262 high-quality expression profiles. This dataset allowed us to compare the expression profiles of HGSC subtypes in Black cases against other study populations, and it provided an opportunity to evaluate differences in subtype frequency and survival in Black HGSC cases compared to non-Black HGSC cases. An important limitation of this study was that we lacked adequate data to explore whether the observed racial variation in the proportions of gene expression subtypes and survival outcomes was due to technical, biological, sociodemographic, or access-to-care differences.

In summary, we have updated an existing HGSC gene expression subtype classifier so that it is compatible with both array-based gene expression data and RNA-Seq data. This advancement will facilitate reproducible HGSC subtyping for research purposes and is available for use in future studies of HGSC. We have also demonstrated that the HGSC subtypes generated by the updated Way *et al.* (Way et al., 2016) gene expression subtype classifier generalize to racially diverse populations, and we have indicated that HGSC subtype-specific gene expression and subtype-specific survival are consistent across Black, White and Asian study populations. Given our findings, we expect that a clinical HGSC gene expression assay would benefit prognostication and treatment strategies similarly for women from multiple racial and ethnic backgrounds.

Acknowledgements

We would like to acknowledge the AACES interviewers, Christine Bard, LaTonda Briggs, Whitney Franz (North Carolina) and Robin Gold (Detroit). We also acknowledge the individuals responsible for facilitating case ascertainment across the ten sites including: Jennifer Burczyk-Brown (Alabama); Rana Bayakly and Vicki Bennett (Georgia); the Louisiana Tumor Registry; Lisa Paddock and Manisha Narang (New Jersey); Diana Slone, Yingli Wolinsky, Steven Waggoner, Anne Heugel, Nancy Fusco, Kelly Ferguson, Peter Rose, Deb Strater, Taryn Ferber, Donna White, Lynn Borzi, Eric Jenison, Nairmeen Haller, Debbie Thomas, Vivian von Gruenigen, Michele McCarroll, Joyce Neading, John Geisler, Stephanie Smiddy, David Cohn, Michele Vaughan, Luis Vaccarello, Elayna Freese, James Pavelka, Pam Plummer, William Nahhas, Ellen Cato, John Moroney, Mark Wysong, Tonia Combs, Marci Bowling, Brandon Fletcher (Ohio); Martin Whiteside (Tennessee) and Georgina Armstrong and the Texas Registry, Cancer Epidemiology and Surveillance Branch, Department of State Health Services. We would also like to acknowledge the AACES investigators, Anthony J Alberg, Elisa V Bandera, Jill Barnholtz-Sloan, Melissa Bondy, Michele L Cote, Ellen Funkhouser, Edward Peters, Ann G Schwartz, Paul Terry, and Patricia G Moorman. This study would not have been possible without the efforts of the North Carolina Central Tumor Registry and all of the staff of the NCOCS. We also thank Christine Lankevich for her management of the data collection for the North Carolina Study and Rex C. Bentley for

review of the pathology in the NCOCS.We would like to thank Rex C. Bentley and Ann M. Mills for the review of the pathology in the AACES.

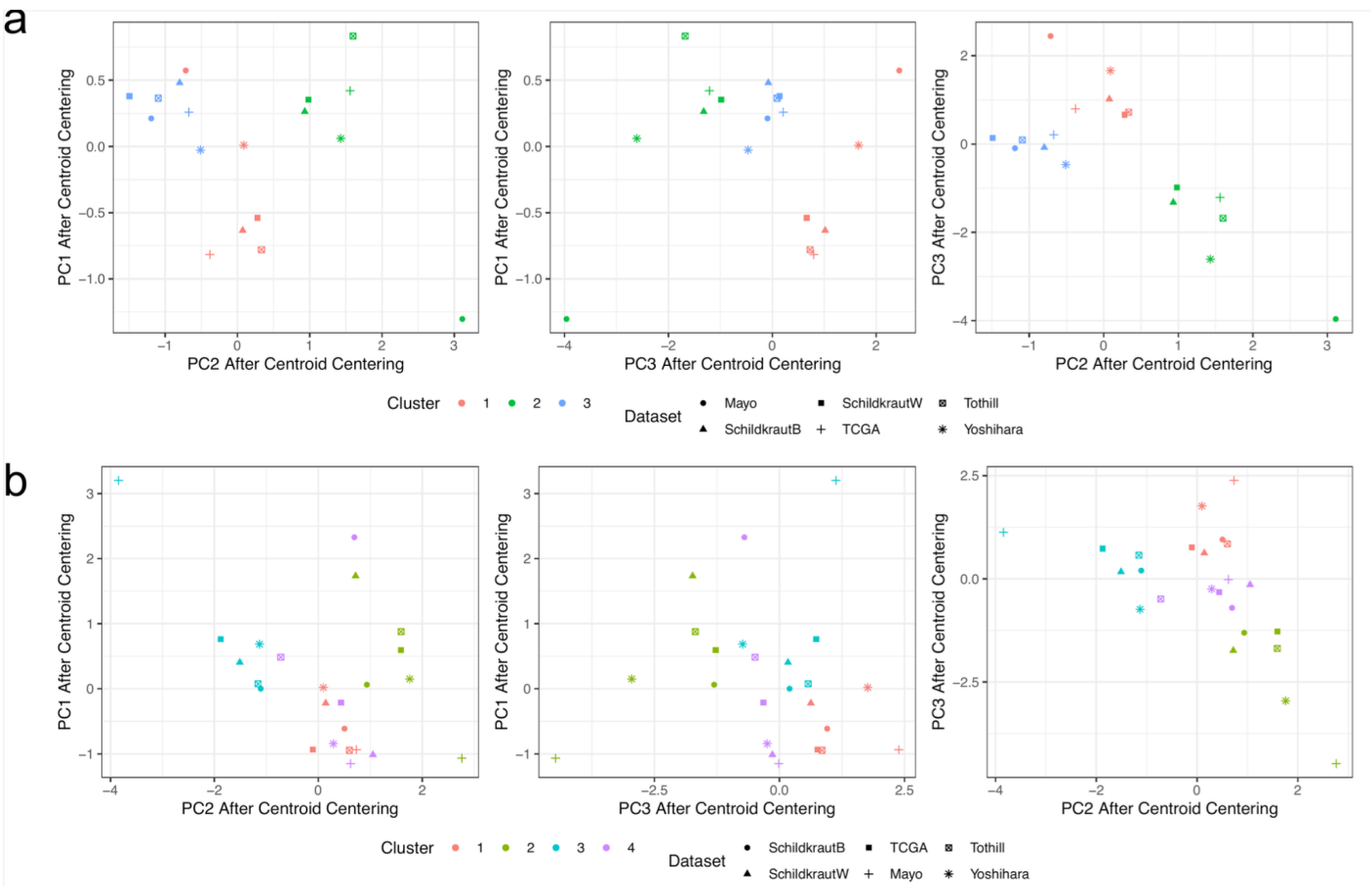


Figure 1. Principal Components Analysis (PCA) plots of cluster centroids for each dataset, after dataset alignment. **Panel a** compares K-means cluster centroids (K=3), and **Panel b** compares K-means cluster centroids (K=4) for each dataset considered in this study. We find that the principal components separate each cluster centroid in a consistent way across almost all datasets. For K=3, PC2 and PC3 are both able to separate each cluster independently, but for K=4 the combination of PC2 and PC3 are needed to separate each cluster. Furthermore, for K=3, we see that the Yoshihara and Mayo datasets have centroids that are much higher in PC1 than the other datasets. This trend continues for the Mayo dataset when K=4, in all PCs.

Figure 2. Significance analysis of microarray (SAM) moderated t-score Pearson correlation heatmaps of clusters across datasets. **Panel a** compares K-means clusters (K=3) between SchildkrautB and every other dataset considered in this study. Across each dataset we find a strong positive correlation with the clusters in SchildkrautB, with matched cluster correlations ranging from 0.37-0.86, and mismatched cluster correlations ranging from -0.65-0.03. **Panel b** performs the same comparison, but for K=4. In this comparison, we see much more inconsistency between matched clusters, with some mismatched clusters having a higher correlation than some matched clusters.

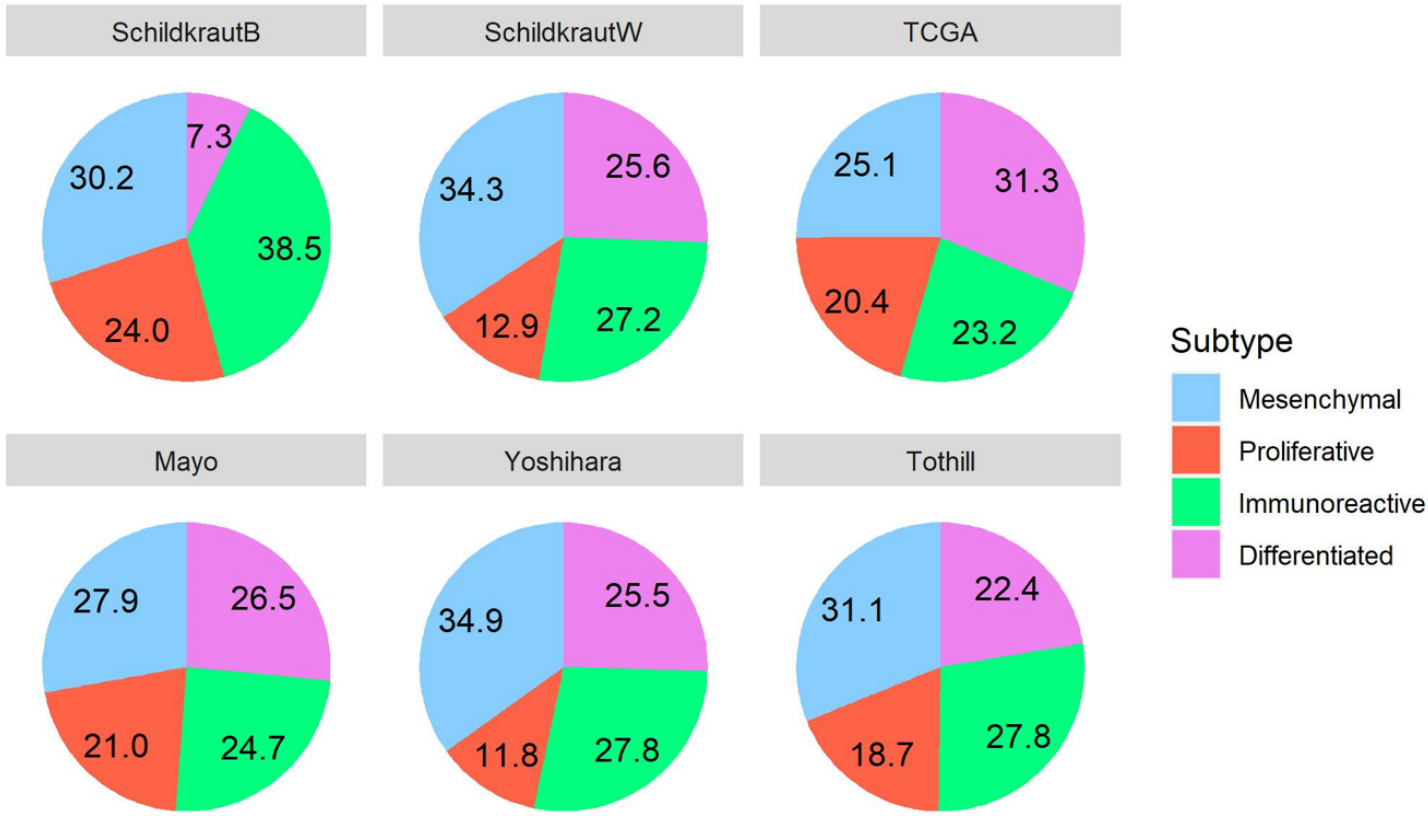


Figure 3. Distribution of subtypes across datasets.

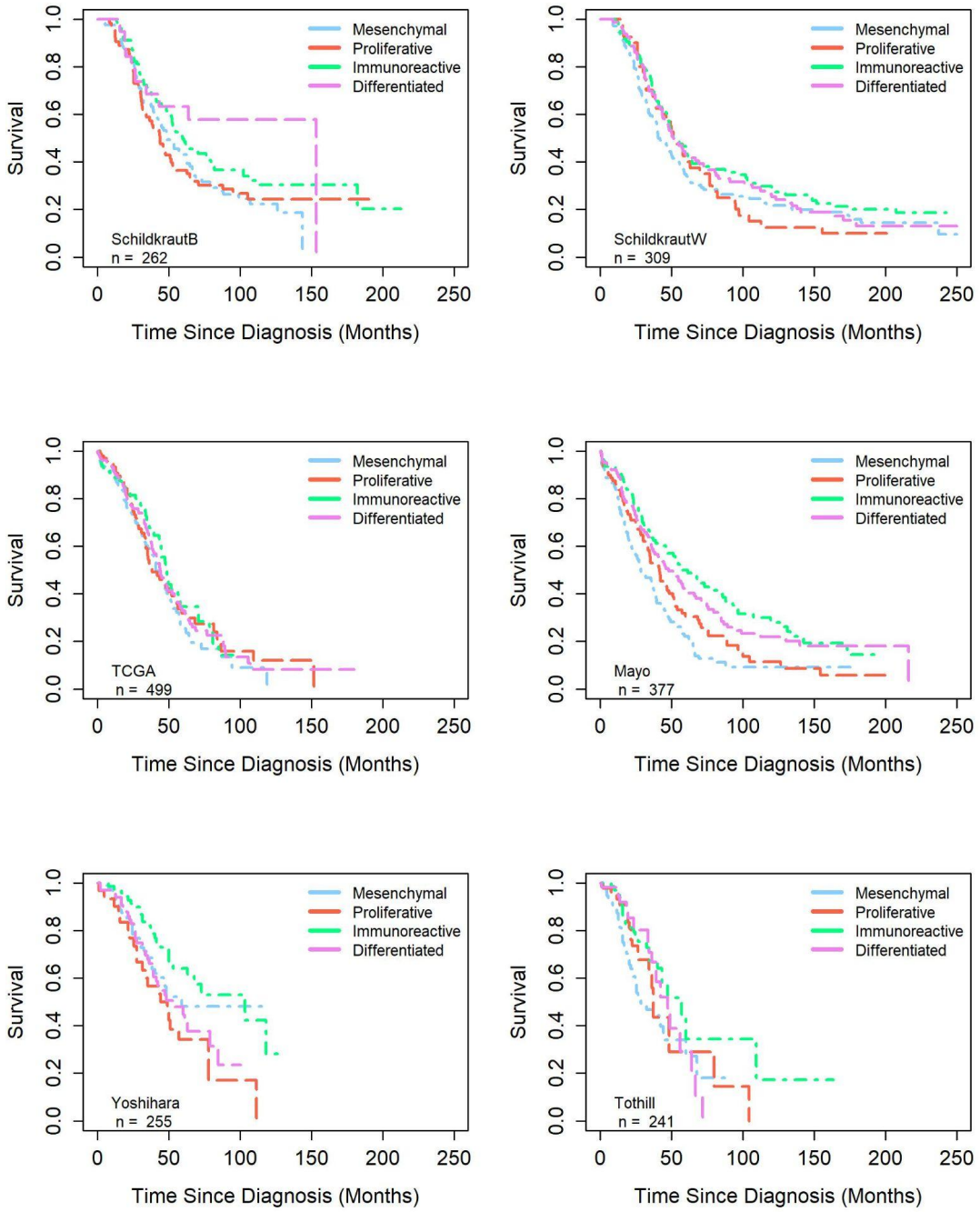


Figure 4. Kaplan Meier survival curves comparing subtype-specific survival by dataset.

Table 1. Characteristics of the Schildkraut study populations, by subtype

	Mesenchymal	Proliferative	Immunoreactive	Differentiated	p-value	Mesenchymal	Proliferative	Immunoreactive	Differentiated	p-value
	Schildkraut Black					Schildkraut White				
Case count	79	63	101	19		106	40	84	79	
Age (%)										
Under 40	1 (1.3)	2 (3.2)	2 (2.0)	1 (5.3)		4 (3.8)	0 (0.0)	1 (1.2)	2 (2.5)	
40-44	8 (10.1)	2 (3.2)	6 (5.9)	3 (15.8)		2 (1.9)	1 (2.5)	7 (8.3)	6 (7.6)	
45-49	9 (11.4)	5 (7.9)	14 (13.9)	2 (10.5)		6 (5.7)	5 (12.5)	10 (11.9)	9 (11.4)	
50-54	17 (21.5)	8 (12.7)	13 (12.9)	3 (15.8)		17 (16.0)	2 (5.0)	17 (20.2)	13 (16.5)	
55-59	8 (10.1)	16 (25.4)	25 (24.8)	4 (21.1)		16 (15.1)	9 (22.5)	17 (20.2)	16 (20.3)	
60-64	16 (20.3)	9 (14.3)	16 (15.8)	1 (5.3)		20 (18.9)	13 (32.5)	17 (20.2)	13 (16.5)	
65-69	6 (7.6)	9 (14.3)	15 (14.9)	2 (10.5)		24 (22.6)	1 (2.5)	9 (10.7)	8 (10.1)	
70-74	12 (15.2)	8 (12.7)	7 (6.9)	1 (5.3)		17 (16.0)	9 (22.5)	6 (7.1)	12 (15.2)	
75+	2 (2.5)	4 (6.3)	3 (3.0)	2 (10.5)	0.301	0 (0)	0 (0)	0 (0)	0 (0)	0.024
FIGO Stage (%)										
I	5 (6.3)	4 (6.3)	13 (12.9)	7 (36.8)		1 (0.9)	1 (2.5)	6 (7.1)	10 (12.7)	
II	8 (10.1)	6 (9.5)	7 (6.9)	3 (15.8)		5 (4.7)	1 (2.5)	2 (2.4)	4 (5.1)	
III	60 (75.9)	49 (77.8)	74 (73.3)	8 (42.1)		97 (91.5)	36 (90.0)	73 (86.9)	59 (74.7)	
IV	4 (5.1)	3 (4.8)	6 (5.9)	0 (0.0)		3 (2.8)	2 (5.0)	3 (3.6)	2 (2.5)	
Missing	2 (2.5)	1 (1.6)	1 (1.0)	1 (5.3)	0.040	0 (0.0)	0 (0.0)	0 (0.0)	4 (5.1)	0.009
Debulking (%)										
Optimal	33 (41.8)	24 (38.1)	45 (44.6)	8 (42.1)		28 (26.4)	14 (35.0)	23 (27.4)	26 (32.9)	
Suboptimal	25 (31.6)	15 (23.8)	21 (20.8)	2 (10.5)		8 (7.5)	3 (7.5)	2 (2.4)	4 (5.1)	
Missing	21 (26.6)	24 (38.1)	35 (34.7)	9 (47.4)	0.34	70 (66.0)	23 (57.5)	59 (70.2)	49 (62.0)	0.59
Neoadjuvant chemotherapy (%)										
No	68 (86.1)	54 (85.7)	85 (84.2)	17 (89.5)		106 (100.0)	40 (100.0)	84 (100.0)	79 (100.0)	NA
Missing	11 (13.9)	9 (14.3)	16 (15.8)	2 (10.5)	0.94					
Proportion CD3+ T cells (%)	4.24 (4.68)	1.51 (3.18)	3.88 (4.54)	1.04 (1.24)	0.001	2.46 (2.36)	0.72 (1.25)	4.35 (6.55)	2.24 (2.20)	0.053
Proportion CD3+/CD8+ Suppressor T cells (%)	1.38 (2.30)	0.52 (1.80)	1.42 (2.33)	0.32 (0.53)	0.045	1.15 (1.18)	0.22 (0.30)	2.39 (4.48)	1.16 (1.21)	0.084

Table 2. Subtype distribution and risk of death by study population

Study and molecular subtype	Case count (%)	Model 1 HR[a] (95% CI)	Model 2 HR[b] (95% CI)
Schildkraut Black			
Mesenchymal	79 (30.2)	(ref)	(ref)
Proliferative	63 (24.0)	1.13 (0.76, 1.67)	0.98 (0.66, 1.46)
Immunoreactive	101 (38.5)	0.83 (0.59, 1.19)	0.79 (0.55, 1.13)
Differentiated	19 (7.3)	0.87 (0.42, 1.83)	0.83 (0.40, 1.72)
Schildkraut White			
Mesenchymal	106 (34.3)	(ref)	(ref)
Proliferative	40 (12.9)	1.00 (0.67, 1.49)	1.05 (0.70, 1.57)
Immunoreactive	84 (27.2)	0.85 (0.61, 1.18)	0.86 (0.62, 1.19)
Differentiated	79 (25.6)	1.01 (0.73, 1.40)	0.98 (0.70, 1.36)
TCGA			
Mesenchymal	125 (25.1)	(ref)	(ref)
Proliferative	102 (20.4)	0.82 (0.56, 1.20)	0.84 (0.57, 1.23)
Immunoreactive	116 (23.2)	0.92 (0.63, 1.35)	1.01 (0.69, 1.49)
Differentiated	156 (31.3)	0.93 (0.67, 1.28)	1.03 (0.74, 1.44)
Mayo			
Mesenchymal	105 (27.9)	(ref)	(ref)
Proliferative	79 (21.0)	0.81 (0.59, 1.12)	0.92 (0.66, 1.28)
Immunoreactive	93 (24.7)	0.54 (0.39, 0.75)	0.54 (0.39, 0.75)
Differentiated	100 (26.5)	0.63 (0.46, 0.87)	0.61 (0.44, 0.84)
Yoshihara[c]			
Mesenchymal	89 (34.9)	(ref)	(ref)
Proliferative	30 (11.8)	1.35 (0.77, 2.35)	1.48 (0.86, 2.57)
Immunoreactive	71 (27.8)	0.62 (0.37, 1.04)	0.65 (0.39, 1.09)
Differentiated	65 (25.5)	1.16 (0.73, 1.85)	1.36 (0.85, 2.19)
Tothill			
Mesenchymal	75 (31.1)	(ref)	(ref)
Proliferative	45 (18.7)	0.77 (0.43, 1.37)	0.79 (0.44, 1.42)
Immunoreactive	67 (27.8)	0.55 (0.32, 0.95)	0.54 (0.31, 0.95)
Differentiated	54 (22.4)	0.64 (0.37, 1.12)	0.65 (0.37, 1.14)

[a] Multivariable model adjusted for age (in 5-year age groups) and stage (I, II, III, IV)

[b] Multivariable model adjusted for age (in 5-year age groups), stage (I, II, III, IV), and debulking status (optimal, not optimal)

[c] Yoshihara dataset does not include information on age, so multivariable models are not adjusted for age in this dataset

Works Cited

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Bonome, T., Levine, D. A., Shih, J., Randonovich, M., Pise-Masison, C. A., Bogomolniy, F., Ozbun, L., Brady, J., Barrett, J. C., Boyd, J., & Birrer, M. J. (2008). A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Research*, *68*(13), 5478–5486. https://doi.org/10.1158/0008-5472.CAN-07-6595

Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, *474*(7353), 609–615. https://doi.org/10.1038/nature10166

Chen, G. M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D. M. A., Parmigiani, G., Birrer, M., Haibe-Kains, B., & Waldron, L. (2018). Consensus on Molecular Subtypes of High-Grade Serous Ovarian Carcinoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *24*(20), 5037–5047. https://doi.org/10.1158/1078-0432.CCR-18-0784

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Fortner, R. T., Trewin-Nybråten, C. B., Paulsen, T., & Langseth, H. (2023). Characterization of ovarian cancer survival by histotype and stage: A nationwide study in Norway. *International Journal of Cancer. Journal International Du Cancer*, *153*(5), 969–978. https://doi.org/10.1002/ijc.34576

Gaitskell, K., Hermon, C., Barnes, I., Pirie, K., Floud, S., Green, J., Beral, V., Reeves, G. K., & Million Women Study Collaborators. (2022). Ovarian cancer survival by stage, histotype, and pre-diagnostic lifestyle factors, in the prospective UK Million Women Study. *Cancer Epidemiology*, *76*, 102074. https://doi.org/10.1016/j.canep.2021.102074

Ganzfried, B. F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., Jazic, I., Wang, X. V., Ahmadifar, M., Birrer, M. J., Parmigiani, G., Huttenhower, C., & Waldron, L. (2013). curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database: The Journal of Biological Databases and Curation*, *2013*, bat013. https://doi.org/10.1093/database/bat013

Gaujoux, R., & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, *11*, 367. https://doi.org/10.1186/1471-2105-11-367

Helland, Å., Anglesio, M. S., George, J., Cowin, P. A., Johnstone, C. N., House, C. M., Sheppard, K. E., Etemadmoghadam, D., Melnyk, N., Rustgi, A. K., Phillips, W. A., Johnsen, H., Holm, R., Kristensen, G. B., Birrer, M. J., Australian Ovarian Cancer Study Group, Pearson, R. B., Børresen-Dale, A.-L., Huntsman, D. G., … Bowtell, D. D. L. (2011). Deregulation of MYCN, LIN28B and LET7 in a molecular subtype of aggressive high-grade serous ovarian cancers. *PloS One*, *6*(4), e18064. https://doi.org/10.1371/journal.pone.0018064

Konecny, G. E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K. R., Dering, J., Ginther, C., Chen, H.-W., Dowdy, S., Cliby, W., Gostout, B., Podratz, K. C., Keeney, G., Wang, H.-J., Hartmann, L. C., Slamon, D. J., & Goode, E. L. (2014). Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *Journal of the National Cancer Institute*, *106*(10). https://doi.org/10.1093/jnci/dju249

Lönnstedt, I., & Speed, T. (2002). REPLICATED MICROARRAY DATA. *Statistica Sinica*, *12*(1), 31–46. http://www.jstor.org/stable/24307034

Maechler, M. (2018). cluster: Cluster Analysis Basics and Extensions. *R Package Version 2.0.7–1*. https://cir.nii.ac.jp/crid/1370290617548437512

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*(9), 1509–1517. https://doi.org/10.1101/gr.079558.108

Moorman, P. G., Schildkraut, J. M., Calingaert, B., Halabi, S., Vine, M. F., & Berchuck, A. (2002). Ovulation and ovarian cancer: a comparison of two methods for calculating lifetime ovulatory cycles (United States). *Cancer Causes & Control: CCC*, *13*(9), 807–811. https://doi.org/10.1023/a:1020678100977

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Peng, L., Bian, X. W., Li, D. K., Xu, C., Wang, G. M., Xia, Q. Y., & Xiong, Q. (2015). Large-scale RNA-Seq Transcriptome Analysis of 4043 Cancers and 548 Normal Tissue Controls across 12 TCGA Cancer Types. *Scientific Reports*, *5*, 13413. https://doi.org/10.1038/srep13413

Peres, L. C., Colin-Leitzinger, C., Sinha, S., Marks, J. R., Conejo-Garcia, J. R., Alberg, A. J., Bandera, E. V., Berchuck, A., Bondy, M. L., Christensen, B. C., Cote, M. L., Doherty, J. A., Moorman, P. G., Peters, E. S.,

Moran Segura, C., Nguyen, J. V., Schwartz, A. G., Terry, P. D., Wilson, C. M., … Schildkraut, J. M. (2022). Racial Differences in the Tumor Immune Landscape and Survival of Women with High-Grade Serous Ovarian Carcinoma. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *31*(5), 1006–1016. https://doi.org/10.1158/1055-9965.EPI-21-1334

Peres, L. C., Cushing-Haugen, K. L., Köbel, M., Harris, H. R., Berchuck, A., Rossing, M. A., Schildkraut, J. M., & Doherty, J. A. (2019). Invasive Epithelial Ovarian Cancer Survival by Histotype and Disease Stage. *Journal of the National Cancer Institute*, *111*(1), 60–68. https://doi.org/10.1093/jnci/djy071

Peres, L. C., Sinha, S., Townsend, M. K., Fridley, B. L., Karlan, B. Y., Lutgendorf, S. K., Shinn, E., Sood, A. K., & Tworoger, S. S. (2020). Predictors of survival trajectories among women with epithelial ovarian cancer. *Gynecologic Oncology*, *156*(2), 459–466. https://doi.org/10.1016/j.ygyno.2019.12.011

Phung, M. T., Pearce, C. L., Meza, R., & Jeon, J. (2023). Trends of Ovarian Cancer Incidence by Histotype and Race/Ethnicity in the United States 1992-2019. *Cancer Research Communications*, *3*(1), 1–8. https://doi.org/10.1158/2767-9764.CRC-22-0410

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* , *23*(21), 2881–2887. https://doi.org/10.1093/bioinformatics/btm453

Schildkraut, J. M., Iversen, E. S., Akushevich, L., Whitaker, R., Bentley, R. C., Berchuck, A., & Marks, J. R. (2013). Molecular signatures of epithelial ovarian cancer: analysis of associations with tumor characteristics and epidemiologic risk factors. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *22*(10), 1709–1721. https://doi.org/10.1158/1055-9965.EPI-13-0192

Schildkraut, J. M., Johnson, C., Dempsey, L. F., Qin, B., Terry, P., Akonde, M., Peters, E. S., Mandle, H., Cote, M. L., Peres, L., Moorman, P., Schwartz, A. G., Epstein, M., Marks, J., Bondy, M., Lawson, A. B., Alberg, A. J., & Bandera, E. V. (2023). Survival of epithelial ovarian cancer in Black women: a society to cell approach in the African American cancer epidemiology study (AACES). *Cancer Causes & Control: CCC*,

34(3), 251–265. https://doi.org/10.1007/s10552-022-01660-0

Schwender, H., Krause, A., & Ickstadt, K. (2006). Identifying interesting genes with siggenes. *The Newsletter of the R Project Volume 6/5, December 2006*, *6*, 45. http://mirror.psu.ac.th/pub/cran/doc/Rnews/Rnews_2006-5.pdf#page=45

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(1), 7–33. https://doi.org/10.3322/caac.21708

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*, Article3. https://doi.org/10.2202/1544-6115.1027

Talhouk, A., George, J., Wang, C., Budden, T., Tan, T. Z., Chiu, D. S., Kommoss, S., Leong, H. S., Chen, S., Intermaggio, M. P., Gilks, B., Nazeran, T. M., Volchek, M., Elatre, W., Bentley, R. C., Senz, J., Lum, A., Chow, V., Sudderuddin, H., … Anglesio, M. S. (2020). Development and Validation of the Gene Expression Predictor of High-grade Serous Ovarian Carcinoma Molecular SubTYPE (PrOTYPE). *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *26*(20), 5411–5423. https://doi.org/10.1158/1078-0432.CCR-20-0103

Tan, T. Z., Miow, Q. H., Huang, R. Y.-J., Wong, M. K., Ye, J., Lau, J. A., Wu, M. C., Bin Abdul Hadi, L. H., Soong, R., Choolani, M., Davidson, B., Nesland, J. M., Wang, L.-Z., Matsumura, N., Mandai, M., Konishi, I., Goh, B.-C., Chang, J. T., Thiery, J. P., & Mori, S. (2013). Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Molecular Medicine*, *5*(7), 1051–1066. https://doi.org/10.1002/emmm.201201823

Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., Johnson, D. S., Trivett, M. K., Etemadmoghadam, D., Locandro, B., Traficante, N., Fereday, S., Hung, J. A., Chiew, Y.-E., Haviv, I., Australian Ovarian Cancer Study Group, Gertig, D., DeFazio, A., & Bowtell, D. D. L. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *14*(16), 5198–5208. https://doi.org/10.1158/1078-0432.CCR-08-0196

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. In *Proceedings of the National Academy of Sciences* (Vol. 98, Issue 9, pp.

5116–5121). https://doi.org/10.1073/pnas.091062498

Verhaak, R. G. W., Tamayo, P., Yang, J.-Y., Hubbard, D., Zhang, H., Creighton, C. J., Fereday, S., Lawrence, M., Carter, S. L., Mermel, C. H., Kostic, A. D., Etemadmoghadam, D., Saksena, G., Cibulskis, K., Duraisamy, S., Levanon, K., Sougnez, C., Tsherniak, A., Gomez, S., … Cancer Genome Atlas Research Network. (2013). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *The Journal of Clinical Investigation*, *123*(1), 517–525. https://doi.org/10.1172/JCI65833

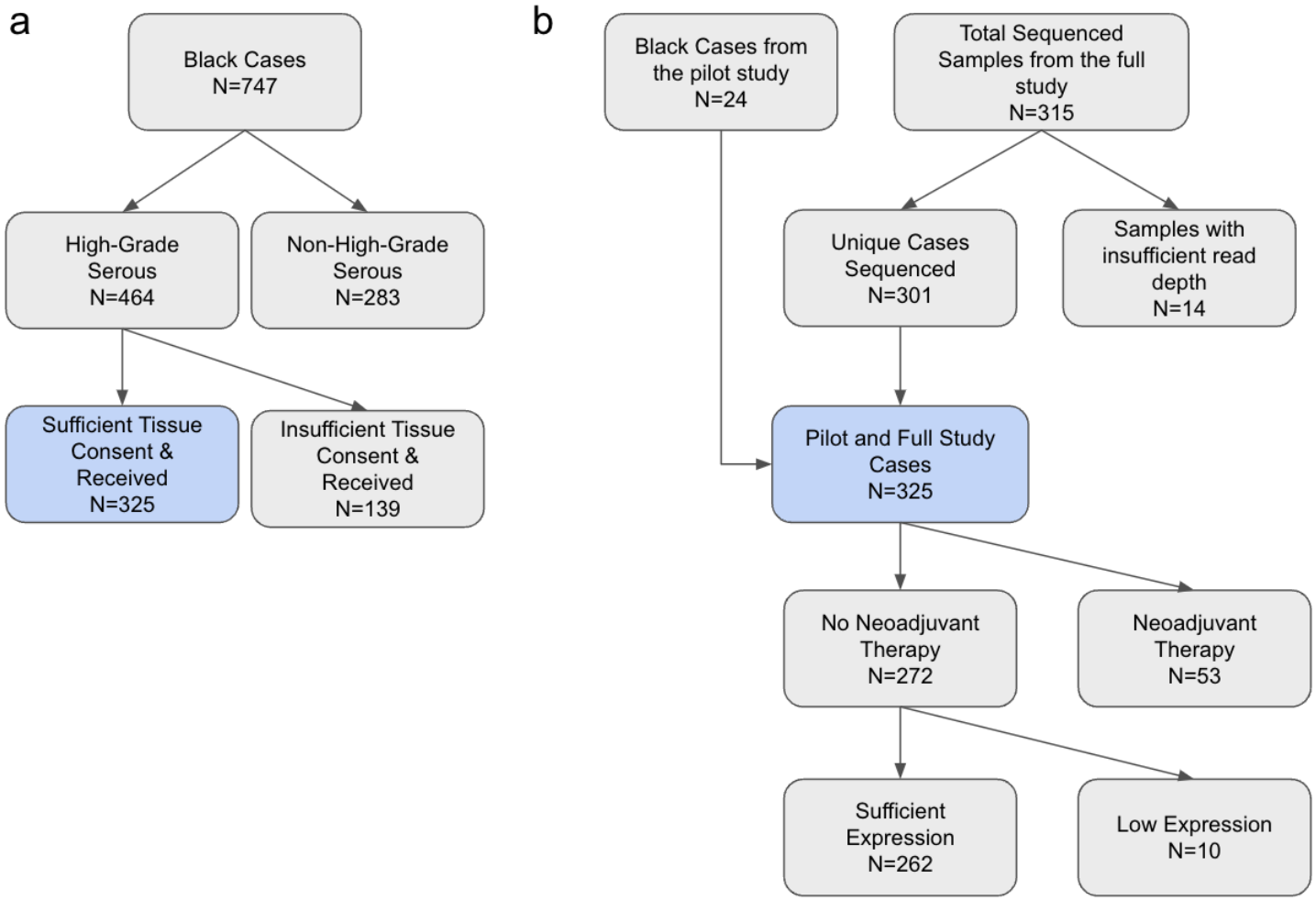
Waldron, L., Riester, M., Ramos, M., Parmigiani, G., & Birrer, M. (2016). The Doppelgänger Effect: Hidden Duplicates in Databases of Transcriptome Profiles. *Journal of the National Cancer Institute*, *108*(11). https://doi.org/10.1093/jnci/djw146

Wang, C., Armasu, S. M., Kalli, K. R., Maurer, M. J., Heinzen, E. P., Keeney, G. L., Cliby, W. A., Oberg, A. L., Kaufmann, S. H., & Goode, E. L. (2017). Pooled Clustering of High-Grade Serous Ovarian Cancer Gene Expression Leads to Novel Consensus Subtypes Associated with Survival and Surgical Outcomes. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *23*(15), 4077–4085. https://doi.org/10.1158/1078-0432.CCR-17-0246

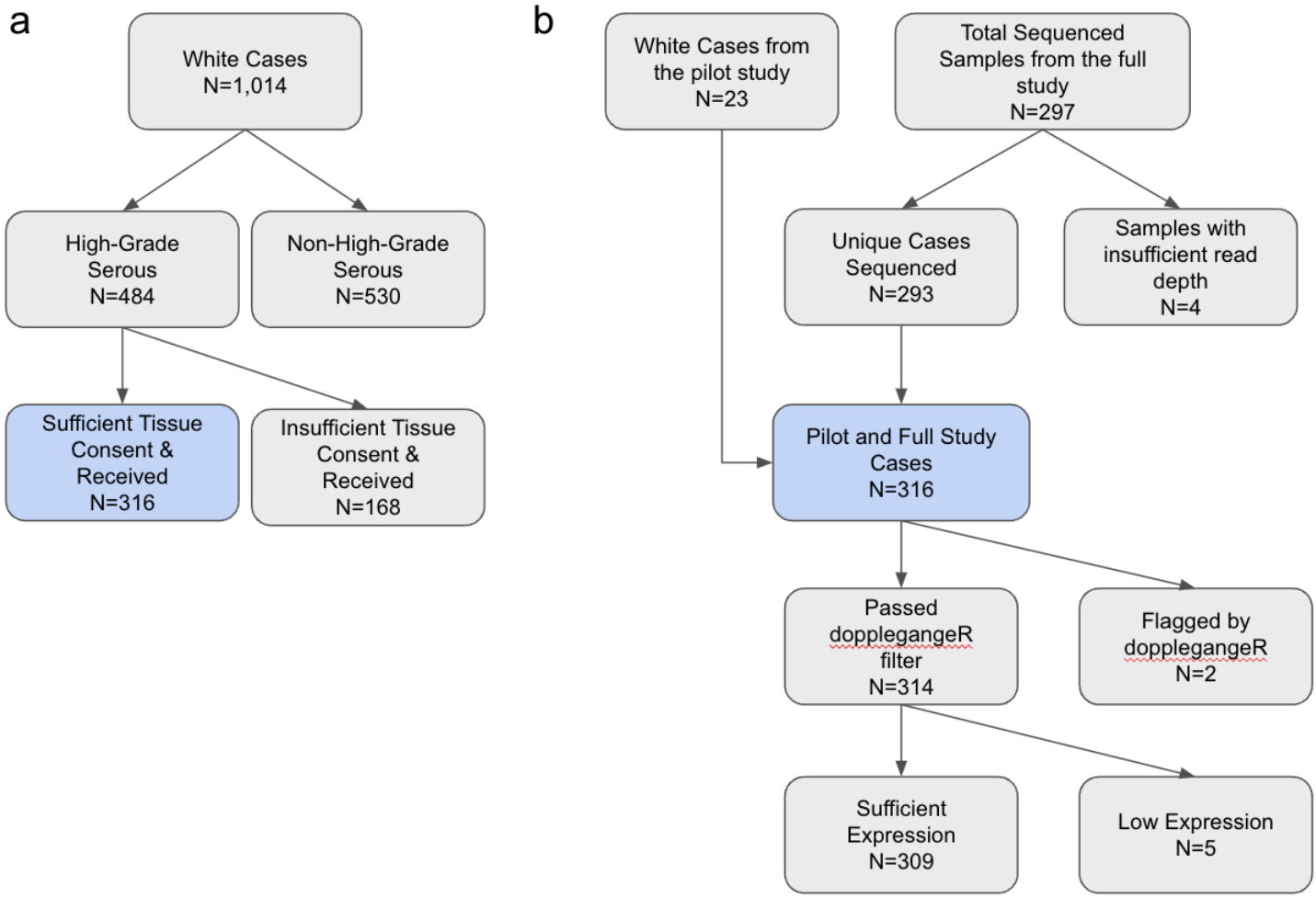
Way, G. P., Rudd, J., Wang, C., Hamidi, H., Fridley, B. L., Konecny, G. E., Goode, E. L., Greene, C. S., & Doherty, J. A. (2016). Comprehensive Cross-Population Analysis of High-Grade Serous Ovarian Cancer Supports No More Than Three Subtypes. In *G3 Genes|Genomes|Genetics* (Vol. 6, Issue 12, pp. 4097–4103). https://doi.org/10.1534/g3.116.033514

Yoshihara, K., Tsunoda, T., Shigemizu, D., Fujiwara, H., Hatae, M., Fujiwara, H., Masuzaki, H., Katabuchi, H., Kawakami, Y., Okamoto, A., Nogawa, T., Matsumura, N., Udagawa, Y., Saito, T., Itamochi, H., Takano, M., Miyagi, E., Sudo, T., Ushijima, K., … Japanese Serous Ovarian Cancer Study Group. (2012). High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *18*(5), 1374–1385. https://doi.org/10.1158/1078-0432.CCR-11-2725

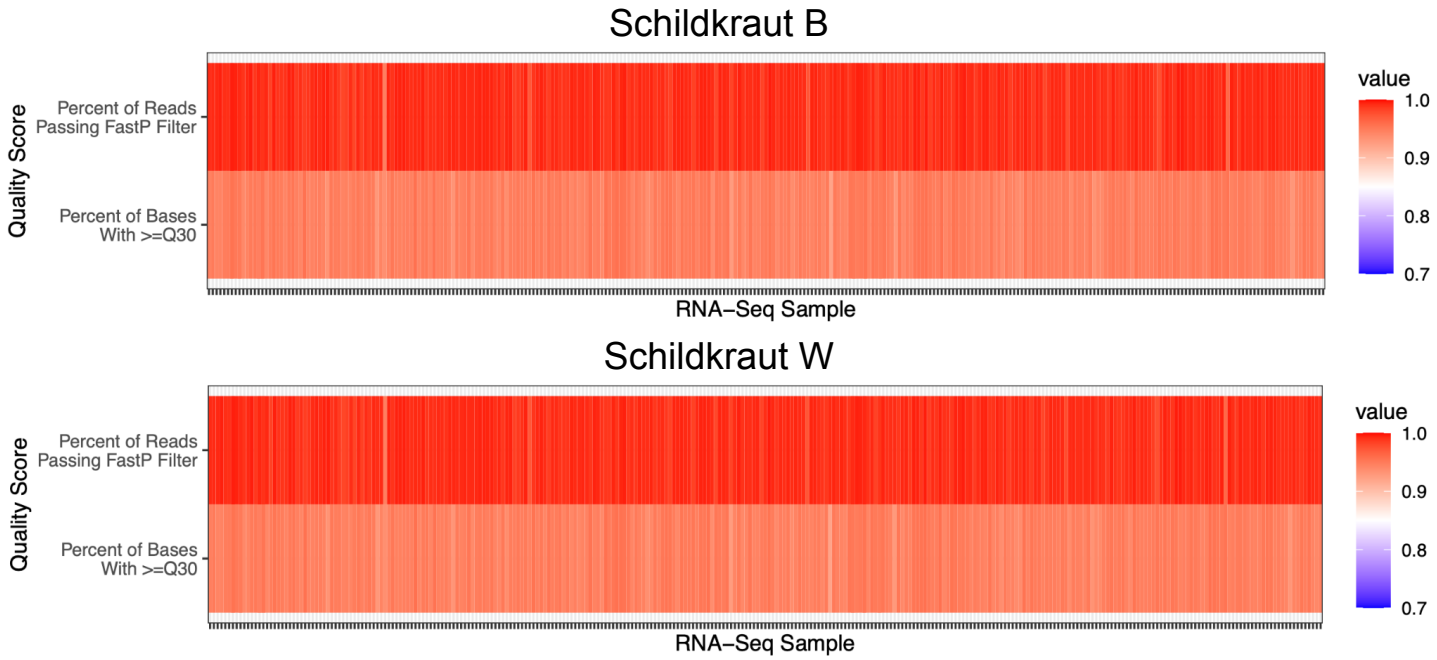
Supplemental Figures



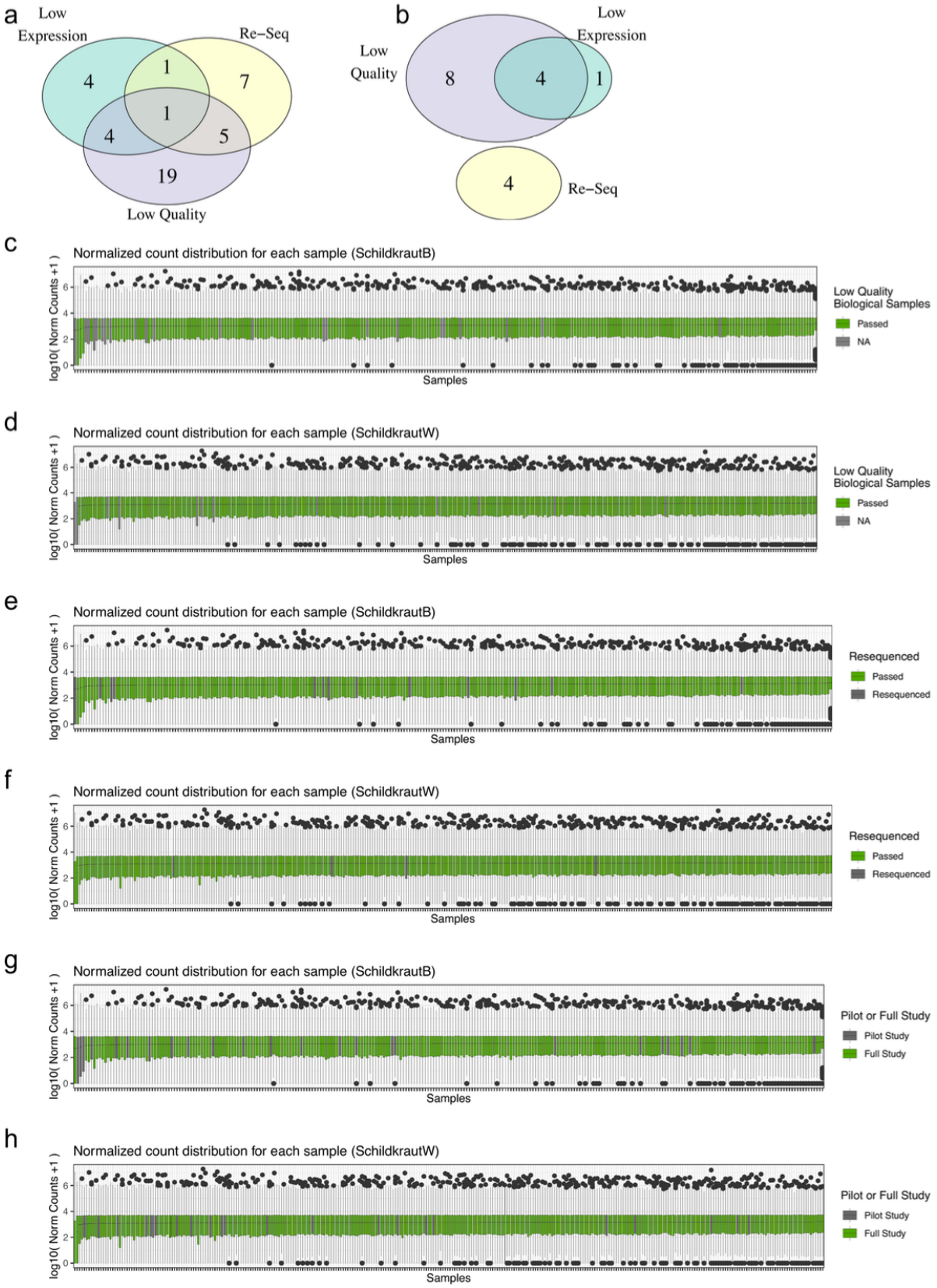
Supplemental Figure 1. Panels a and b depict the case selection process for Black study participants before (**a**) and after (**b**) sequencing. Blue rectangles indicate the cases that are the same in panels a and b, and are included in **Supplemental Table 3**. All cases listed as “samples with insufficient read depth” were successfully resequenced with sufficient read depth so, while 14 samples were excluded, no participants were excluded between “Total Sequenced Samples from the full study” and “Unique Cases Sequenced.”



Supplemental Figure 2. Panels a and b depict the case selection process for White study participants before (**a**) and after (**b**) sequencing. Blue rectangles indicate the cases that are the same in panels a and b, and are included in **Supplemental Table 4**. All cases listed as “samples with insufficient read depth” were successfully resequenced with sufficient read depth so, while 4 samples were excluded, no participants were excluded between “Total Sequenced Samples from the full study” and “Unique Cases Sequenced.”

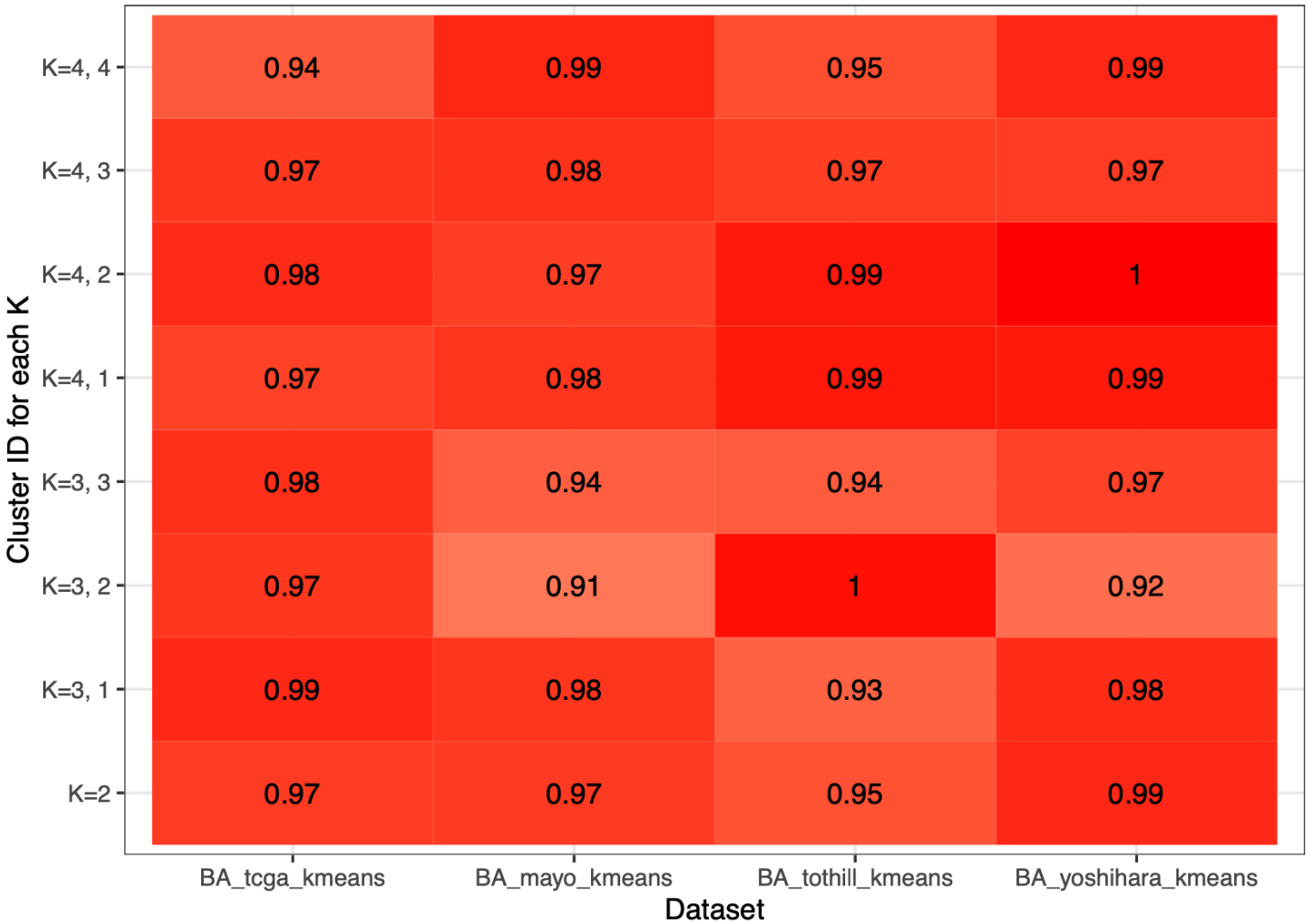


Supplemental Figure 3. This figure depicts the percentage of reads that passed the FastP filtering (top), and the percentage of bases that had a quality score ≥30 in the SchildkrautB and SchildkrautW datasets. Each column is an RNA-seq sample and its coloring is based upon its quality score for each metric. For SchildkrautB the sample with the minimum percent of reads passing the FastP filter was 89.8%, and the sample with the minimum percentage of bases with a quality score ≥30 was 83.3%. For SchildkrautW the sample with the minimum percent of reads passing the FastP filter was 92.4% and the sample with the minimum percentage of bases with a quality score ≥30 was 88.2%.

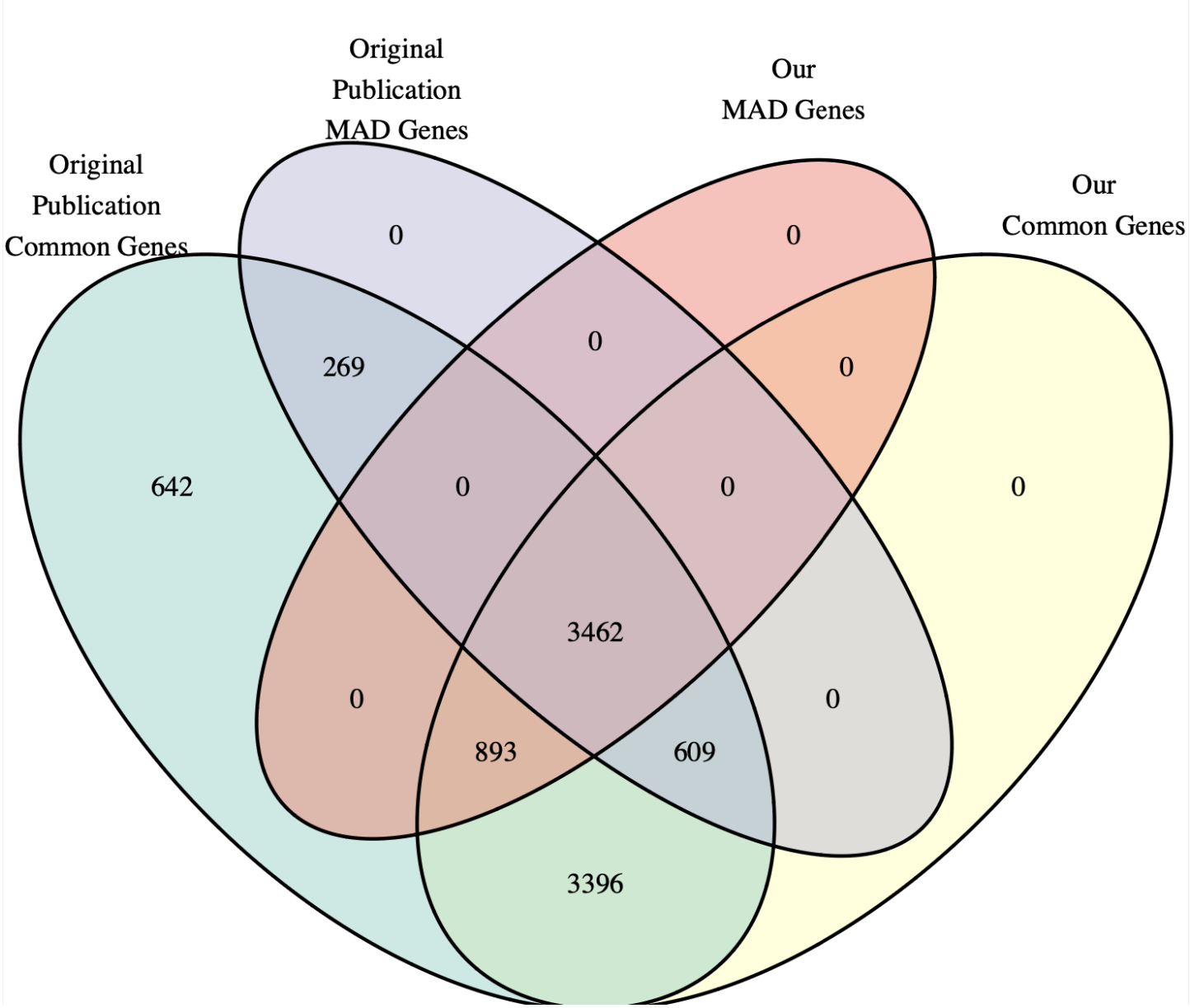


Supplemental

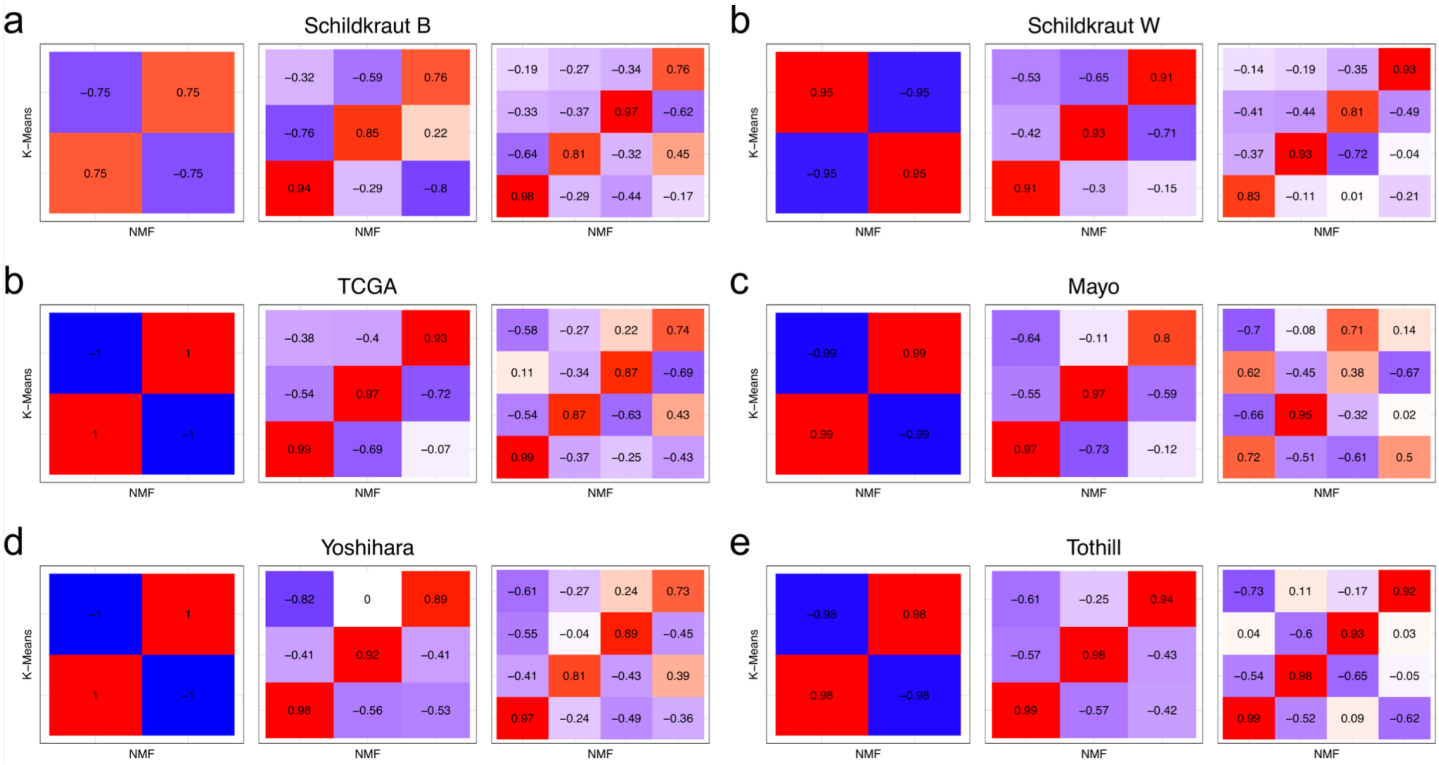
Figure 4. SchildkrautB and SchildkrautW sample quality evaluation. **Panels A and B** depict venn diagrams of samples that needed to be resequenced, had lower-than-expected read counts (i.e., low expression), or were flagged by the sequencing core as potentially low-quality. Ultimately, 10 cases were excluded in SchildkrautB (Panel A) and 5 were excluded in SchildkrautW (Panel B), all due to low expression. **Panels C-H** are boxplots showing the range of gene expression for each sample and colored by a sample feature related to sequencing quality or batch. Each boxplot is a sample, and the boxplots are ordered by median normalized expression.



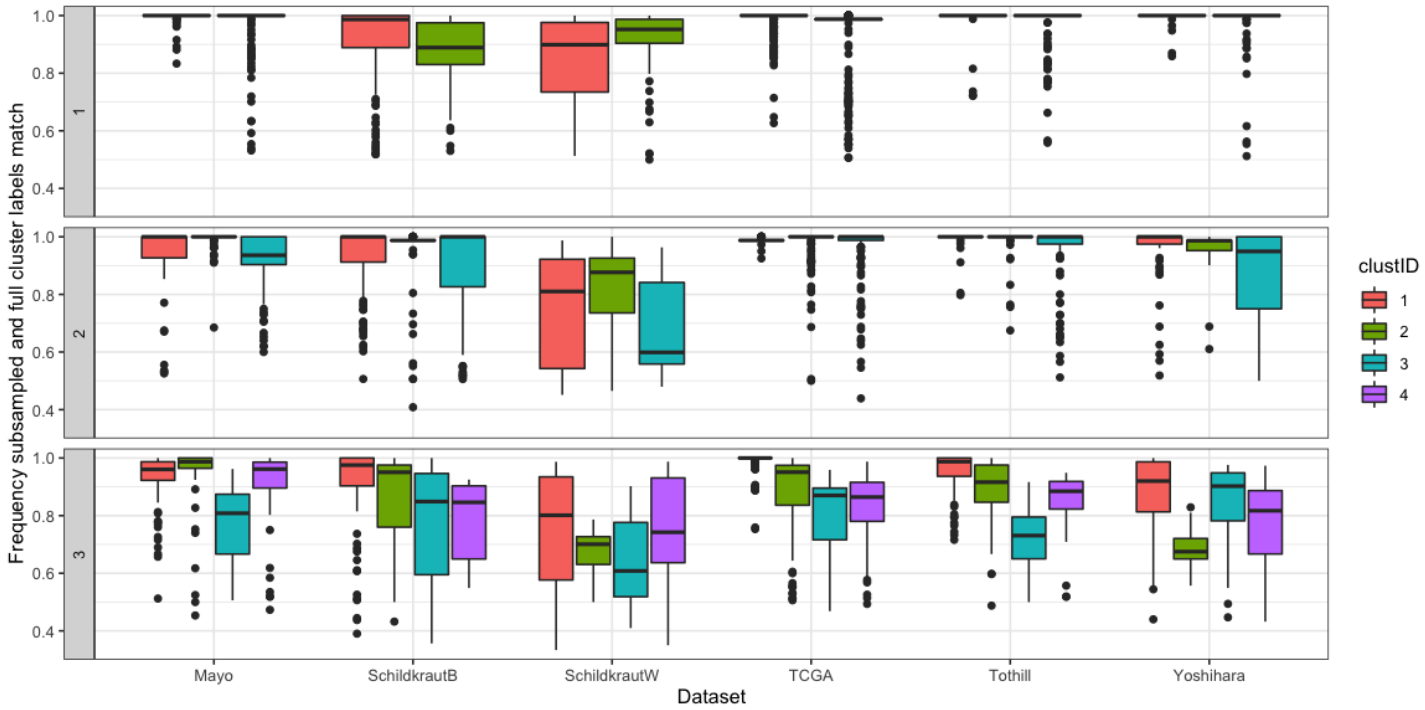
Supplemental Figure 5. This heatmap shows the balanced accuracy when comparing each sample's previously assigned cluster-ID and newly assigned cluster-ID. For all considered values of K, each dataset achieved a balanced accuracy of at least 0.91.



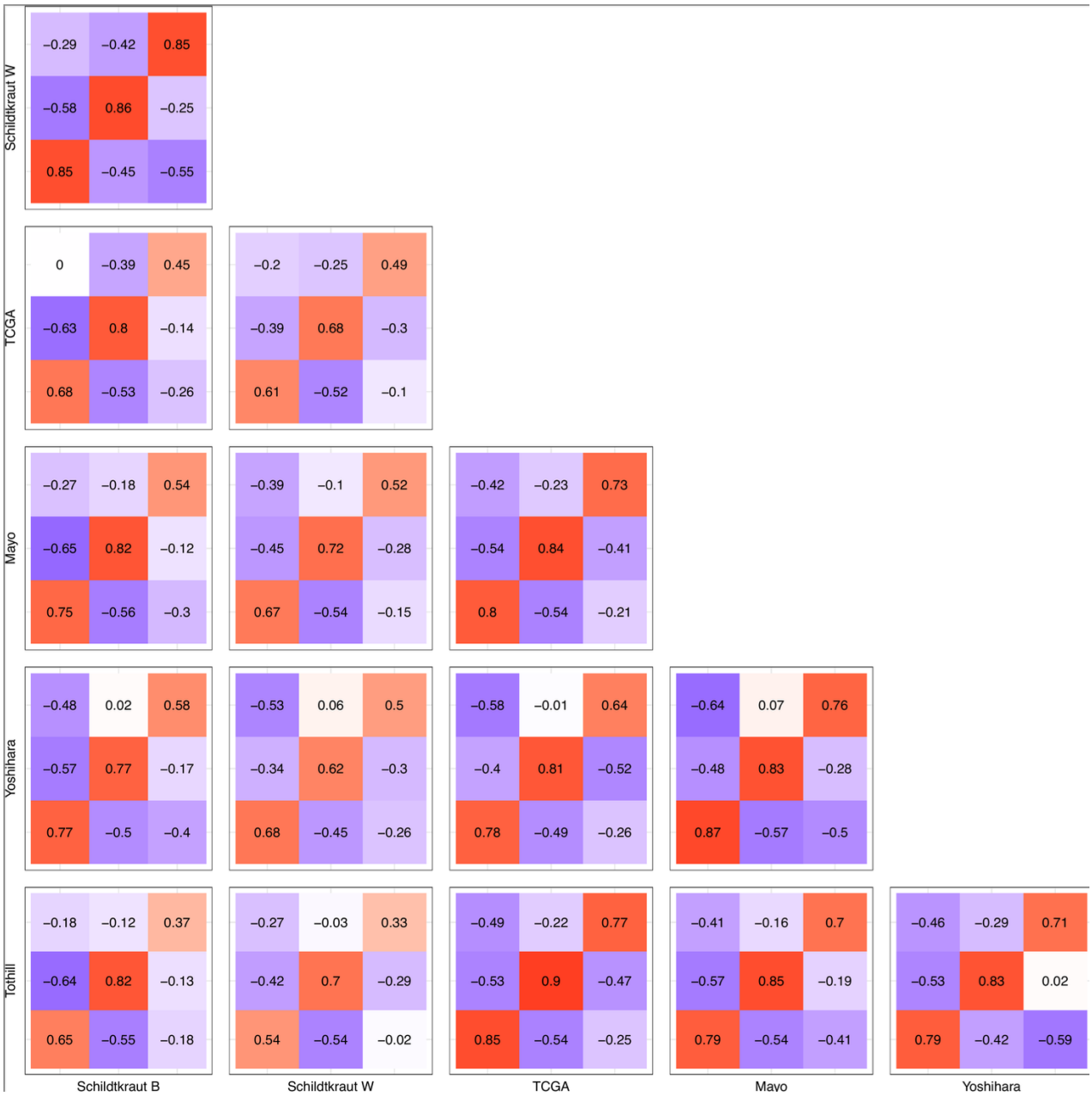
Supplemental Figure 6. Genes considered for clustering and cluster comparison. This venn diagram compares the genesets used in this work to those used in the previously published analysis (Way et al., 2016). The MAD genes (the genes used for clustering), are a subset of the common genes (the genes used to compare clusters across datasets). No new genes were added in our work, and 911 were removed from the common genes. This is expected since the common genes must be common to all analytic datasets, and our analysis included two additional datasets.



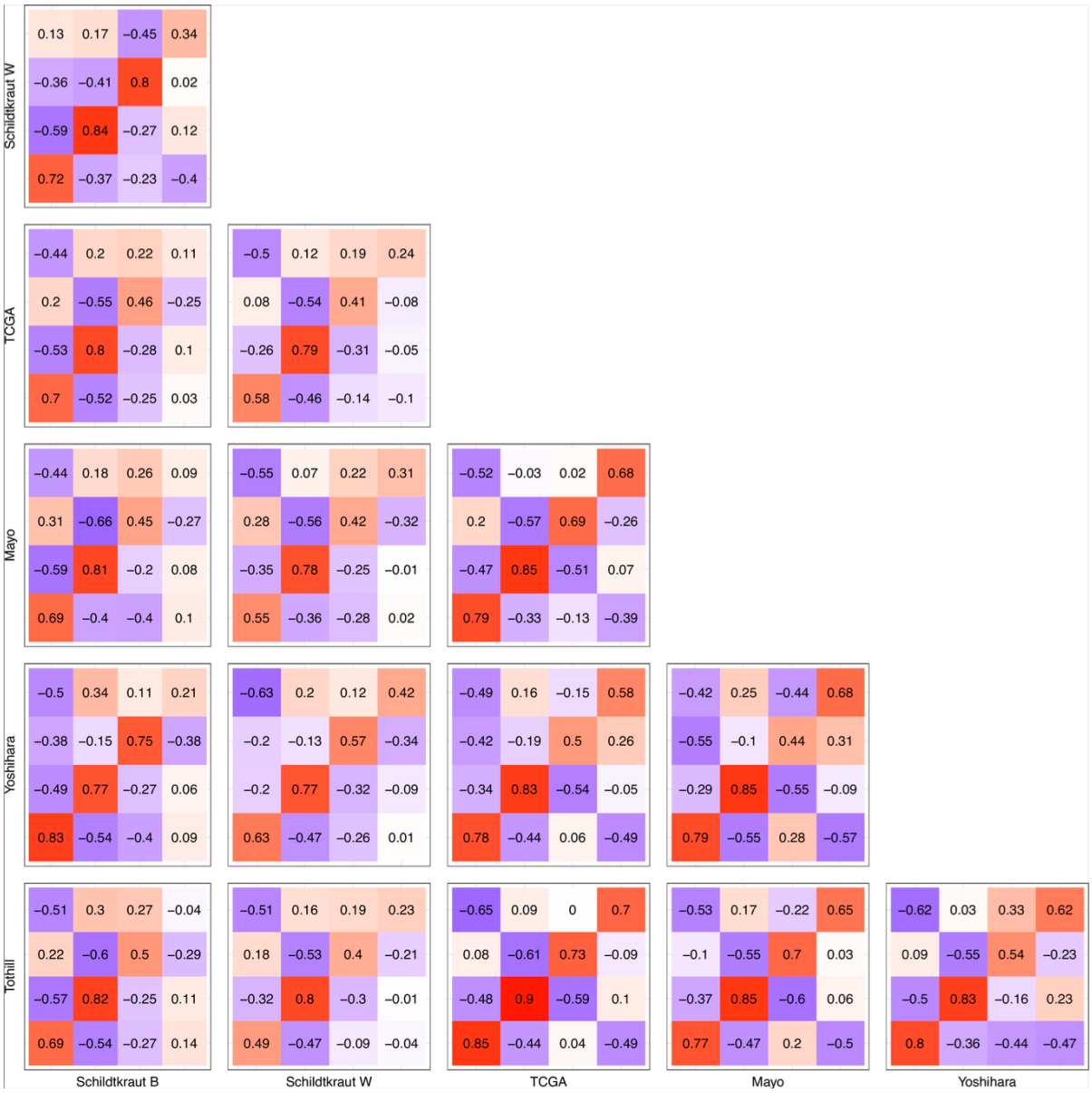
Supplemental Figure 7. Significance analysis of microarray (SAM) moderated t-score Pearson correlation heatmaps of clusters across two clustering methods. **Panels A-E** depict heatmaps of cluster agreement between K-means and NMF for each dataset fo K=2, 3, and 4. Blue indicates a negative correlation and red indicates a positive correlation, with the moderated t-score Pearson correlation coefficient written in each cell. We found that the clusters identified by K-means and NMF were highly consistent in all datasets for K=2 and 3. When we used K=4, we found higher off-diagonal correlation coefficients indicating mismatched clusters. The highest, off-diagonal correlation coefficients in each dataset were as follows: SchildkrautB = 0.45, SchildkrautW = 0.01, TCGA = 0.43, Mayo = 0.71, Yoshihara=0.39, and Tothill 0.11. In K=3 we found the highest off-diagonal correlation coefficient in each dataset to be: SchildkrautB = 0.22, SchildkrautW = -0.15, TCGA = -0.07, Mayo = -0.12, Yoshihara=0, and Tothill -0.25.



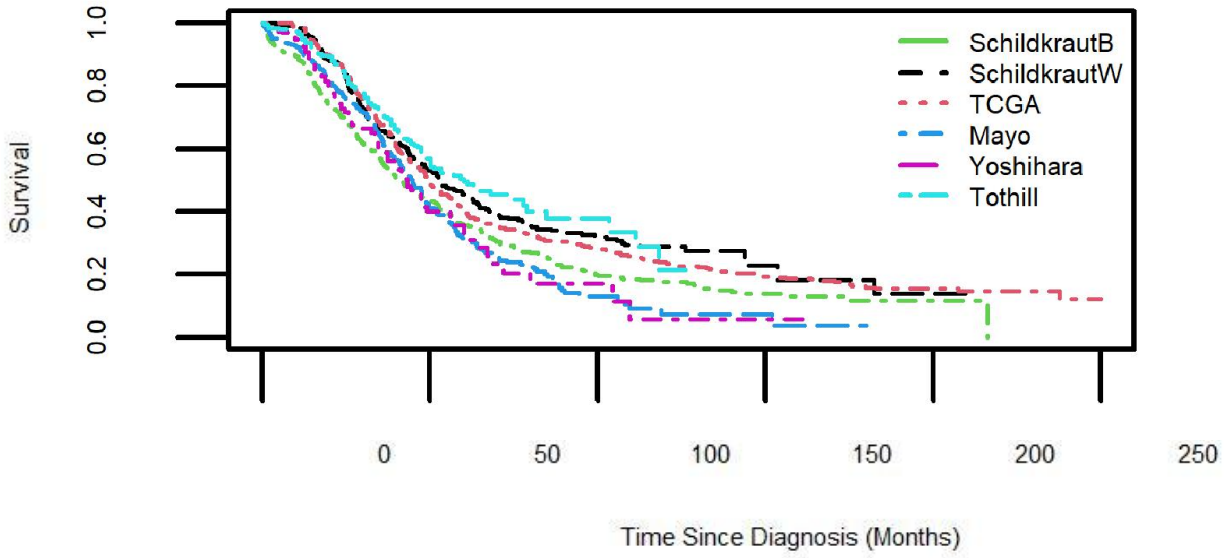
Supplemental Figure 8. Analysis of cluster stability for K=2-4 across all datasets. For each dataset, we selected 80% of samples and re-ran K-means clustering 100 times. After matching the subsampled and original cluster labels, we calculated the frequency with which a sample retained its original cluster label. For all datasets, we found that the majority of samples kept their original cluster labels in more than 50% of the subsampled re-clusterings.



Supplemental Figure 9. Significance analysis of microarray (SAM) moderated t-score Pearson correlation heatmaps of K-means clusters across each dataset for K=3. Blue indicates a negative correlation and red indicates a positive correlation, with the moderated t-score Pearson correlation coefficient written in each cell. Our comparison of each dataset against every other dataset shows that the clusters are highly correlated between all pairs of datasets.



Supplemental Figure 10. Significance analysis of microarray (SAM) moderated t-score Pearson correlation heatmaps of K-means clusters across each dataset for K=4. Blue indicates a negative correlation and red indicates a positive correlation, with the moderated t-score Pearson correlation coefficient written in each cell. In our comparison of each dataset against every other dataset, we see that the clusters have reduced correlation between each dataset in comparison to our K=3 analysis.



Supplemental Figure 11. Kaplan Meier survival curves comparing survival across datasets.