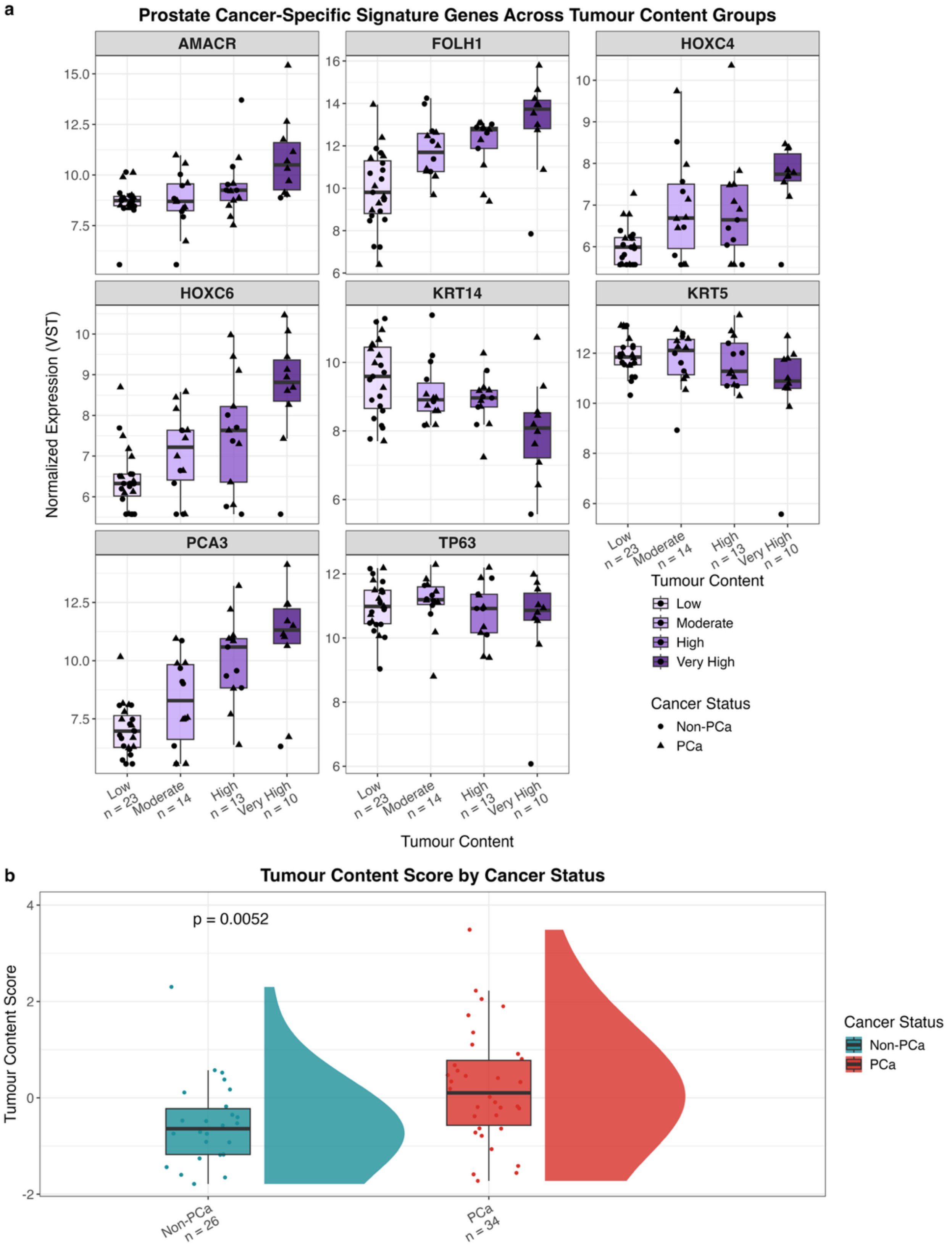
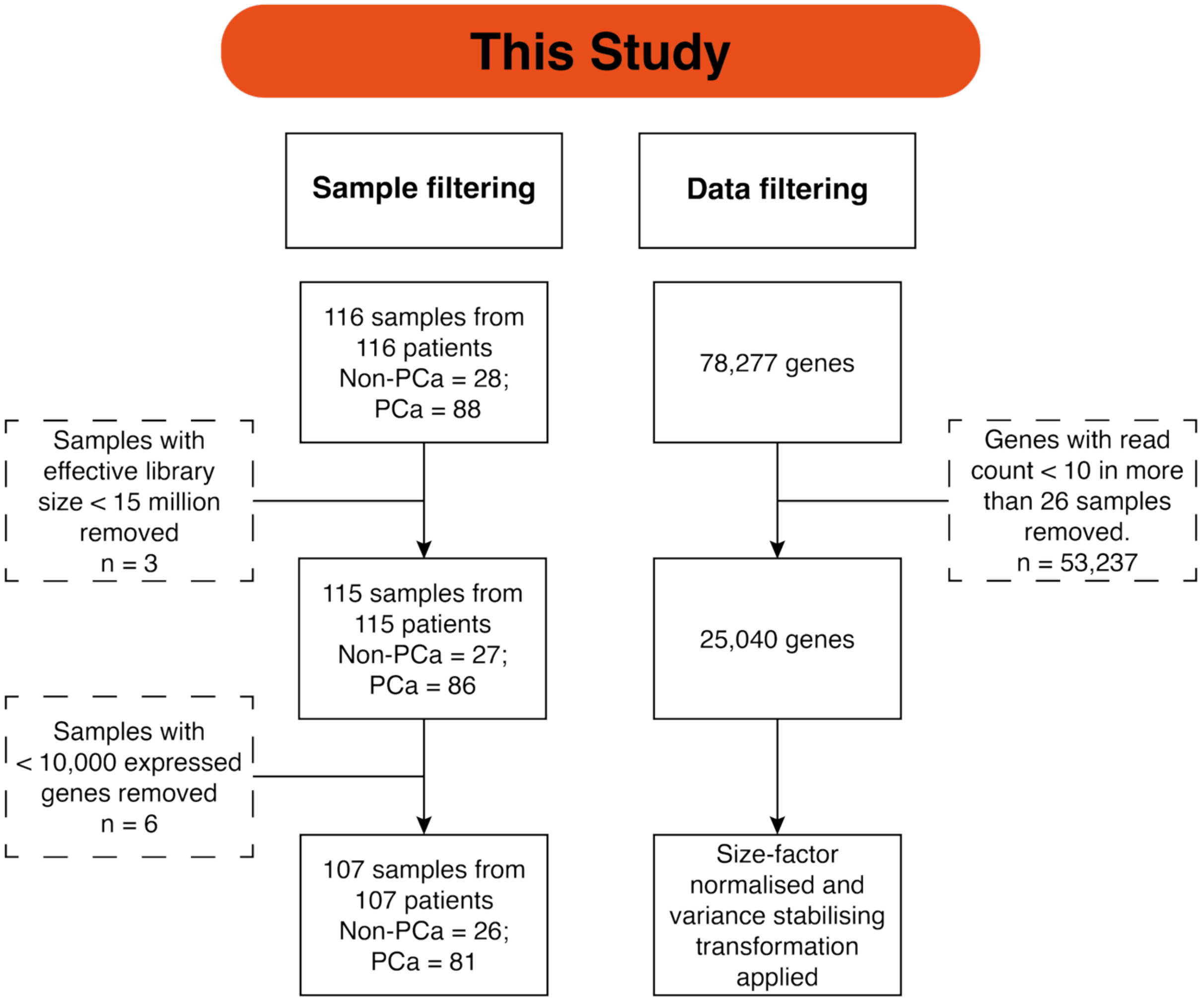
Additional File 2: Supplementary Figures

A unique transcriptomic landscape defines African-specific grade group 1 prostate cancer

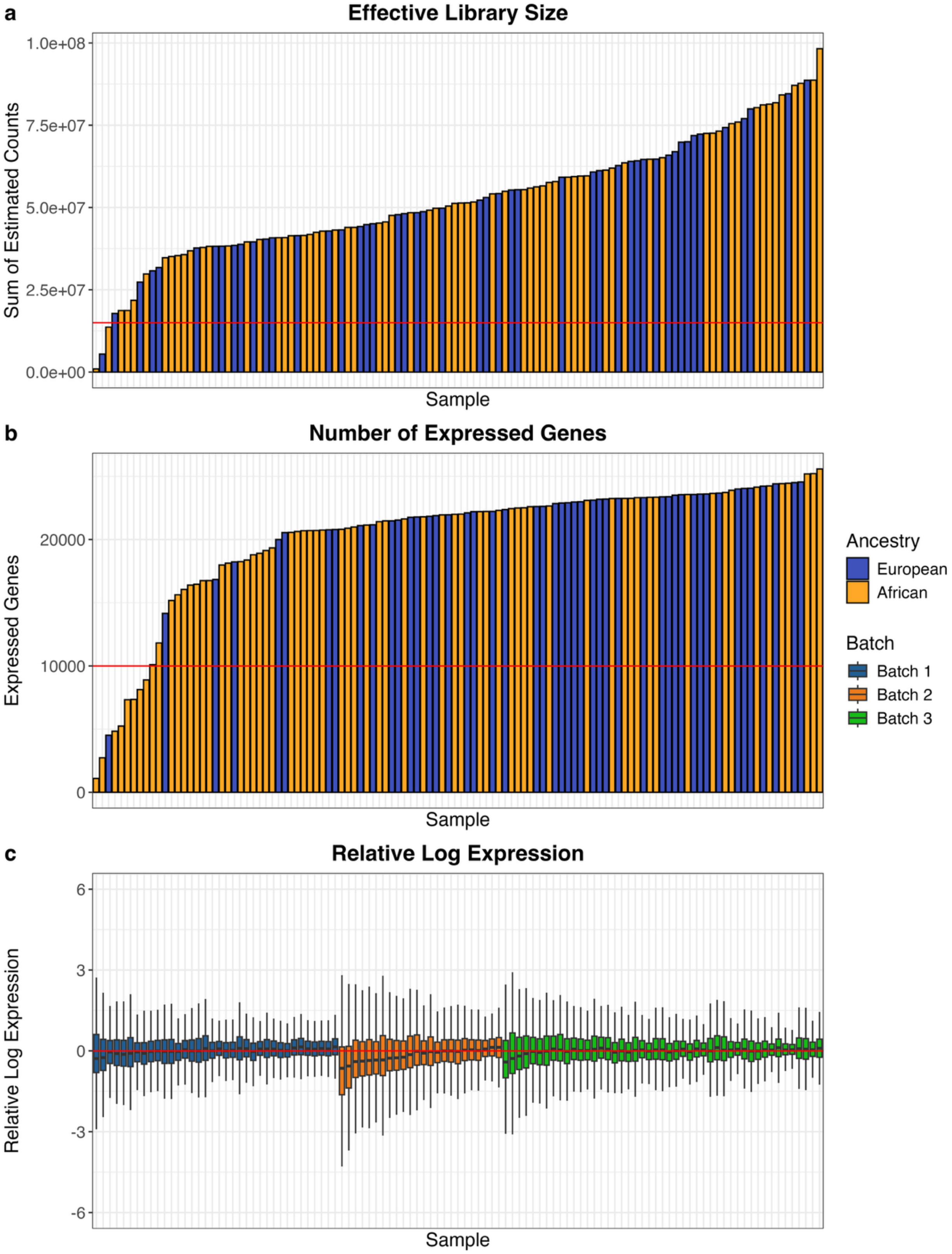
Eva Ferlev Jensby, Korawich Uthayopas, Md Mehedi Hasan, Melanie Louw, Shingai B.A. Mutambirwa, Philip D. Stricker, Raymond A. Campbell, Philippe Lamy, Luigi Marchionni, Pan Prostate Cancer Group, Southern African Prostate Cancer Study, M.S. Riana Bornman, Anthony T. Papenfuss, Karina D. Sørensen & Vanessa M. Hayes



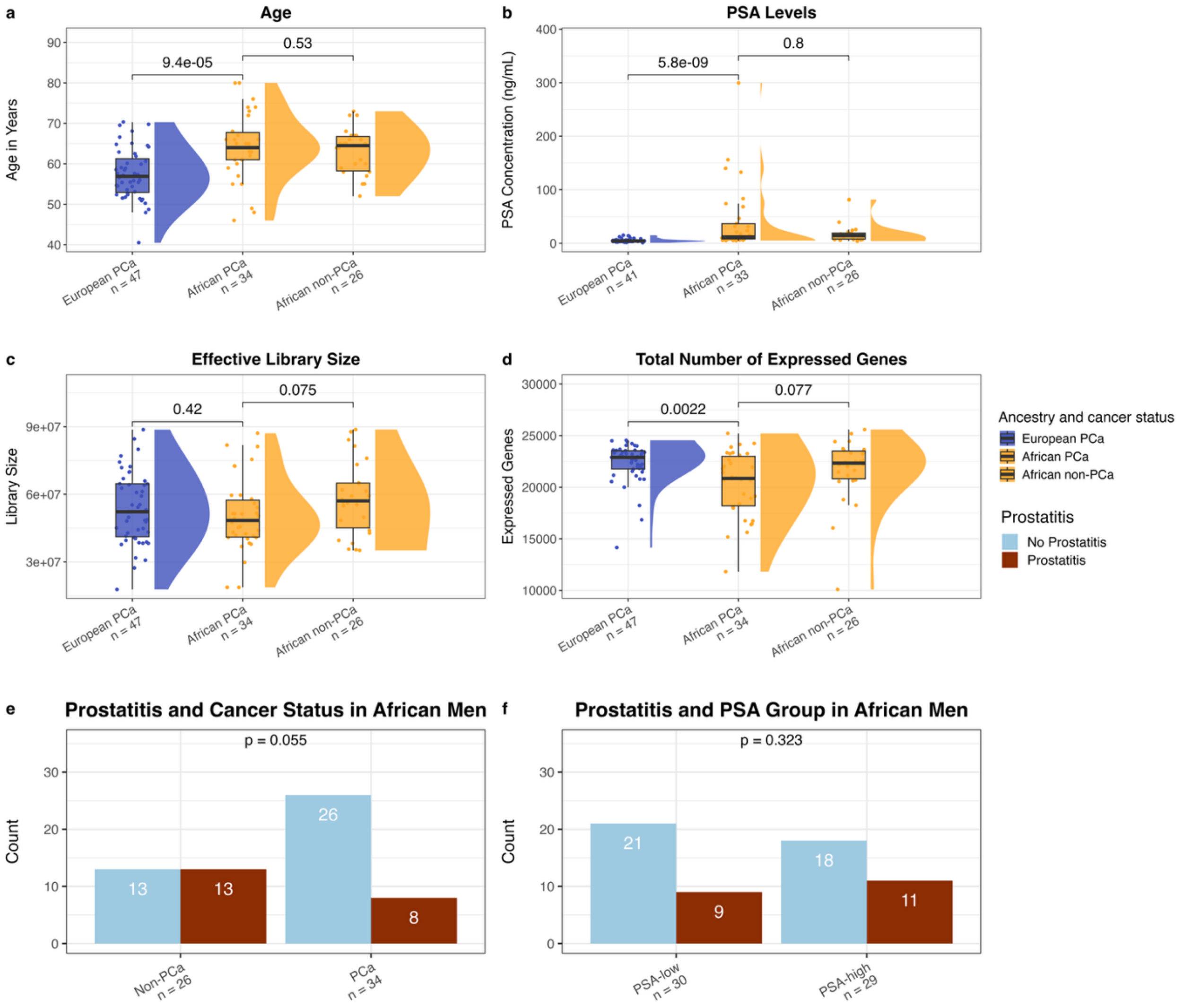
Supplementary Fig. S1: Tumour content estimation in African ancestry samples (PCa = 34, non-PCa = 26). a, Expression of 8 tumour content signature genes stratified by tumour content group (low, moderate, high, very high). Boxplots coloured by tumour content category. **b,** Tumour content scores stratified by cancer status (34 PCa, 26 non-PCa). Colours indicate cancer status. *P*-value calculated using Wilcoxon rank-sum test. PCa = prostate cancer.



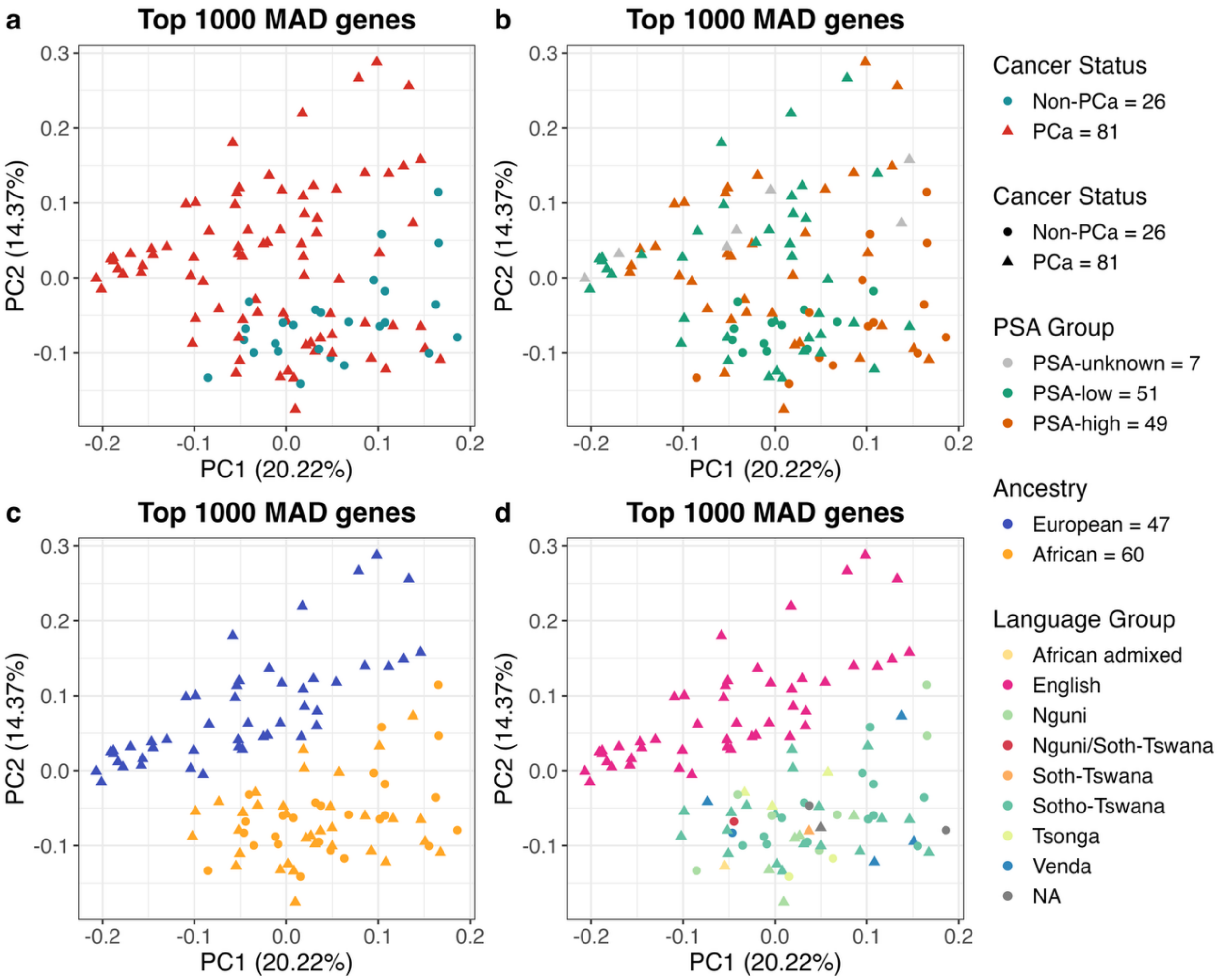
Supplementary Fig. S2: Sample and data filtering workflow from initial cohort (n = 116 total, 88 GG1-PCa, 28 non-PCa) to final analysed dataset (n = 107 total, 81 GG1-PCa, 26 non-PCa). Left panel shows sample-level filtering with samples excluded if the effective library size was <15 million or the number of expressed genes was < 10,000. Right panel shows gene-level filtering and normalisation with genes removed if counts were < 10 in more than 26 samples, followed by variance stabilising transformation (includes size-factor normalisation). PCa = prostate cancer.



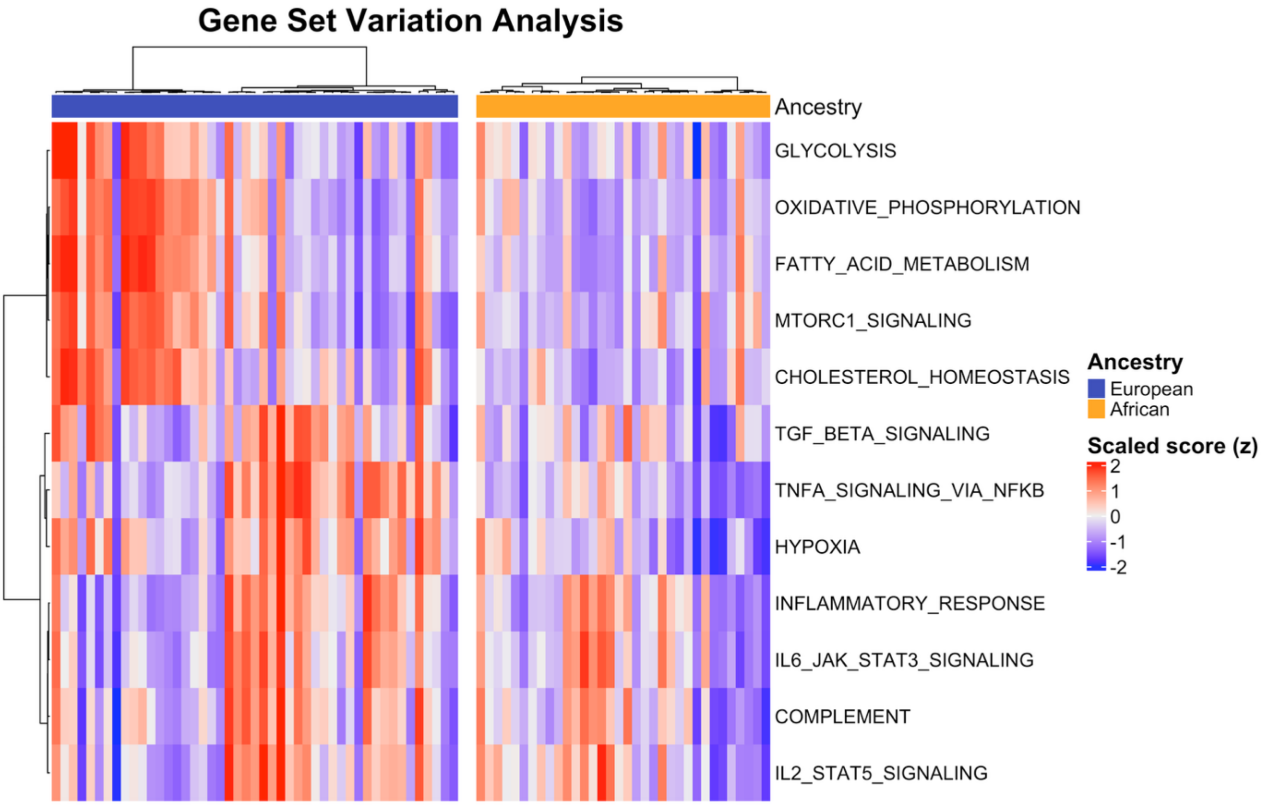
Supplementary Fig. S3: Sample quality assessment of all samples before filtering (n = 116, panels a-b) and after sample and gene filtering (n = 107, panel c). a, Effective library sizes. Red line indicates exclusion threshold of 15 million counts. Samples ordered by library size. Colours indicate ancestry. **b,** Number of expressed genes. Red line indicates exclusion threshold of 10,000 genes. Samples ordered by gene count. Colours indicate ancestry. **c,** Relative log expression of variance-stabilised data across samples after filtering. Samples ordered by sequencing batch. Each boxplot represents gene-level deviations from median expression for one sample. Red line indicates zero deviation. Colours indicate sequencing batch.



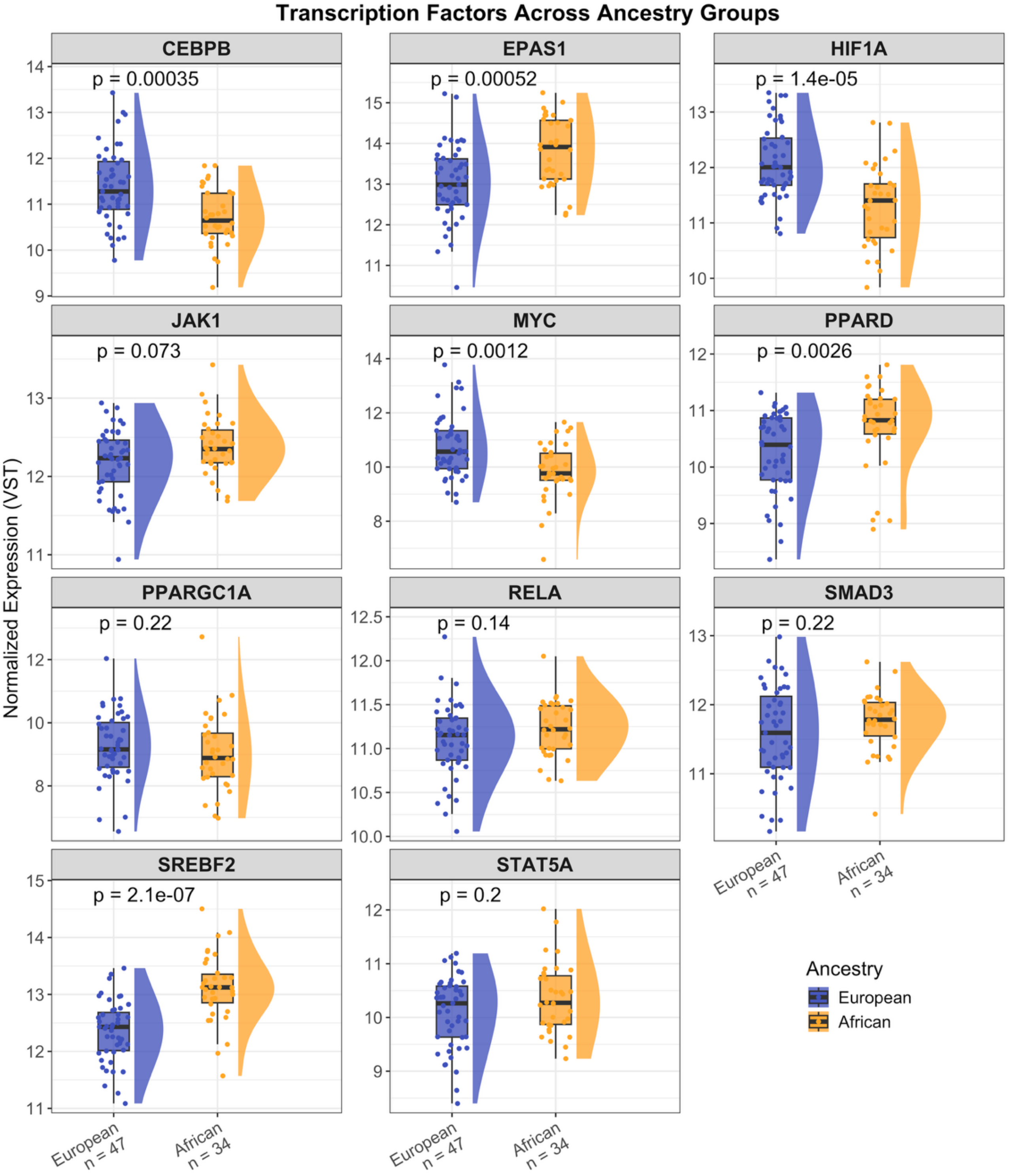
Supplementary Fig. S4: Clinical characteristics of 107 analysed samples comprising 60 African ancestry samples (34 GG1-PCa, 26 non-PCa) and 47 European ancestry GG1-PCa samples. Boxplots show the distribution of **a,** age, **b,** PSA levels, **c,** effective library size (total mapped reads), and **d,** number of expressed genes across European ancestry PCa, African ancestry PCa, and African ancestry non-PCa groups. Colours indicate ancestry group. *P*-values calculated using Wilcoxon rank-sum test. Bar plots show prostatitis prevalence in African ancestry samples stratified by **e,** cancer status and **f,** PSA groups. Colours indicate prostatitis status. *P*-values calculated using Fisher's exact test. PCa = prostate cancer; PSA = prostate-specific antigen.



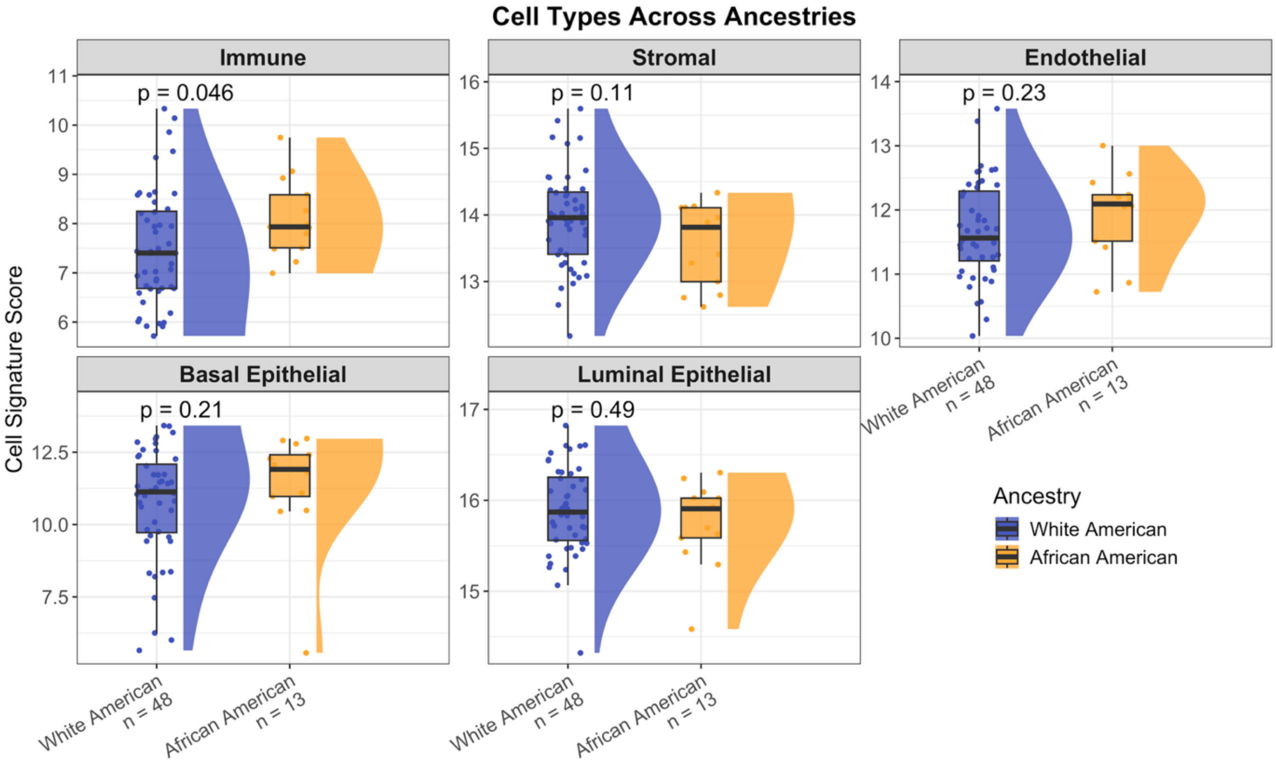
Supplementary Fig. S5: Principal component analysis of the top 1,000 genes ranked by median absolute deviation (MAD) from variance-stabilised data (n = 107). Each panel shows the first two principal components coloured according to **a,** cancer status, **b,** PSA group, **c,** ancestry, and **d,** language group (ethnic/linguistic groups within African ancestry samples; European ancestry samples shown as English). MAD = median absolute deviation; NA = not available; PC = principal component”; PCa = prostate cancer, PSA = prostate-specific antigen.



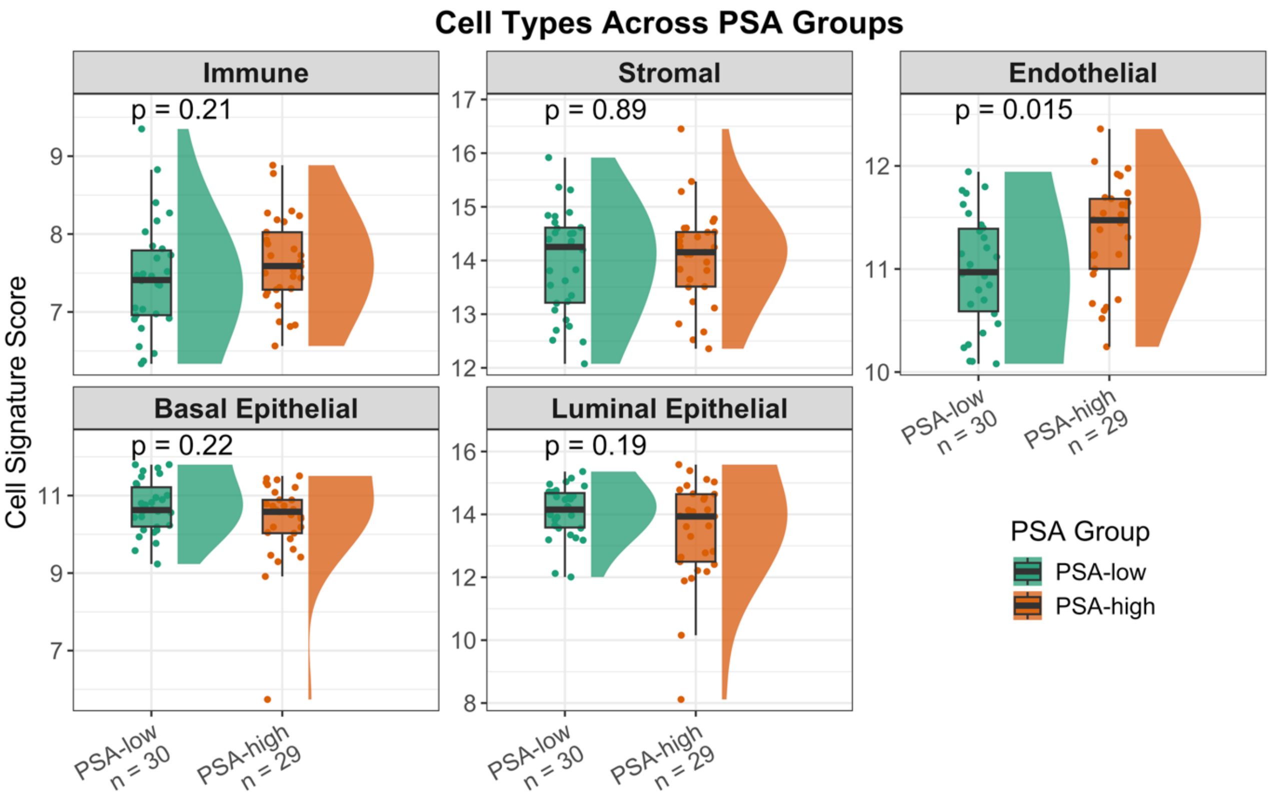
Supplementary Fig. S6: Gene set variation analysis of GG1-PCa samples (34 African, 47 European). Heatmap of GSVA enrichment scores for 12 selected metabolic and immune pathways showing significant negative enrichment between ancestry groups in GSEA analysis. Enrichment scores are z-score transformed across samples for visualisation. Samples are clustered within ancestry groups (hierarchical clustering). GSEA = gene set enrichment analysis; GSVA = gene set variation analysis.



Supplementary Fig. S7: Transcription factor analysis in GG1-PCa samples (34 African, 47 European). Boxplots showing expression of 11 master regulator transcription factors for key pathways identified through GSEA. Colours indicate ancestry group. *P*-values calculated using Wilcoxon rank-sum test.



Supplementary Fig. S8: Cell type signature scores across ancestries in TCGA Grade Group 1 tissues (African American = 13, White American = 48). Boxplots showing basal epithelial, luminal epithelial, stromal, endothelial, and immune scores by ancestry. Colours indicate ancestry group. *P*-values calculated using Wilcoxon rank-sum test.



Supplementary Fig. S9: Cell type signature scores across PSA groups in African ancestral prostate tissues (n = 59: PSA-low = 30, PSA-high = 29). Boxplots showing basal epithelial, luminal epithelial, stromal, endothelial, and immune cell type scores stratified by PSA group. Samples with unknown PSA status excluded (n = 1). Colours indicate PSA group. *P*-values calculated using Wilcoxon rank-sum test. PSA = prostate-specific antigen.