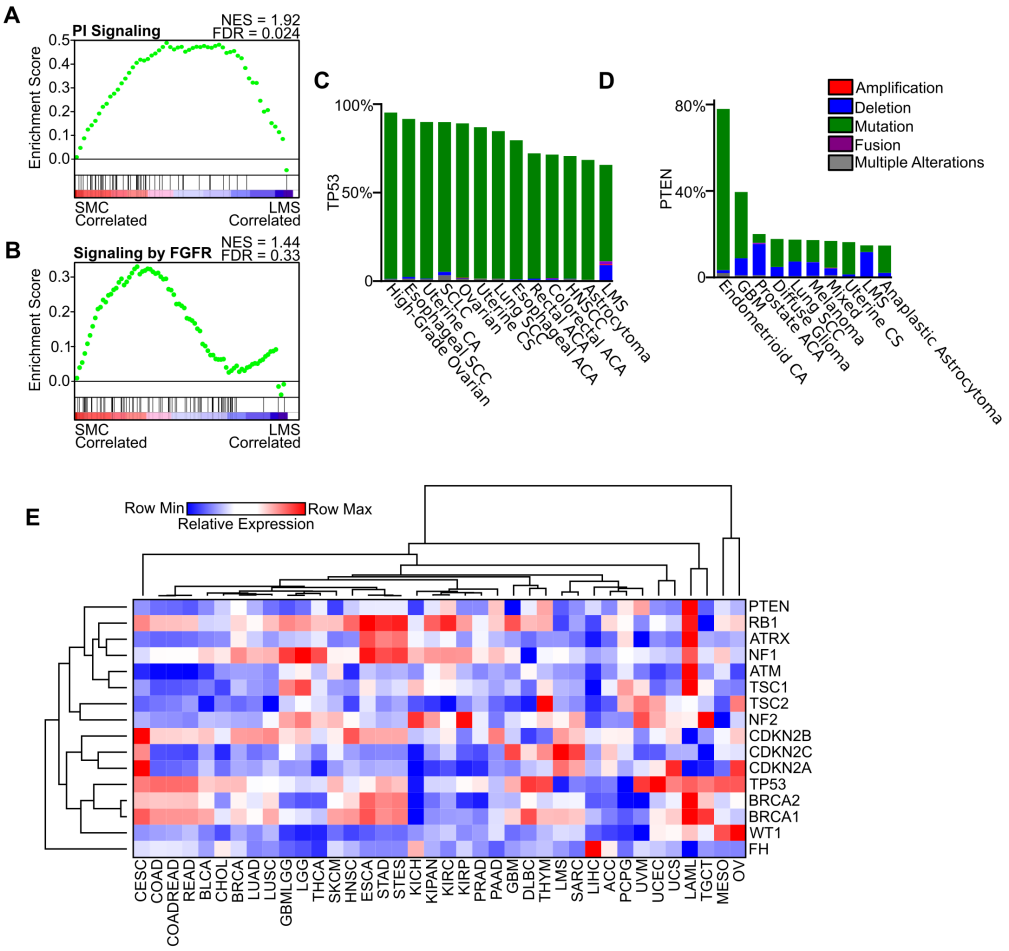

Figure S1. Smooth muscle enriched gene sets and tumor suppressors in LMS. **A-B,** KEGG Phosphatidylinositol Signaling System and Reactome Signaling by FGFR gene sets comparing normal smooth muscle and LMS. **C-D,** Frequency of *TP53* and *PTEN* alterations across cancer subtypes, with the top most altered subtypes shown. **E,** Heatmap showing hierarchical clustering of expression for select tumor suppressor genes in TCGA data sets. See Tables S1-S2 for cancer abbreviations.

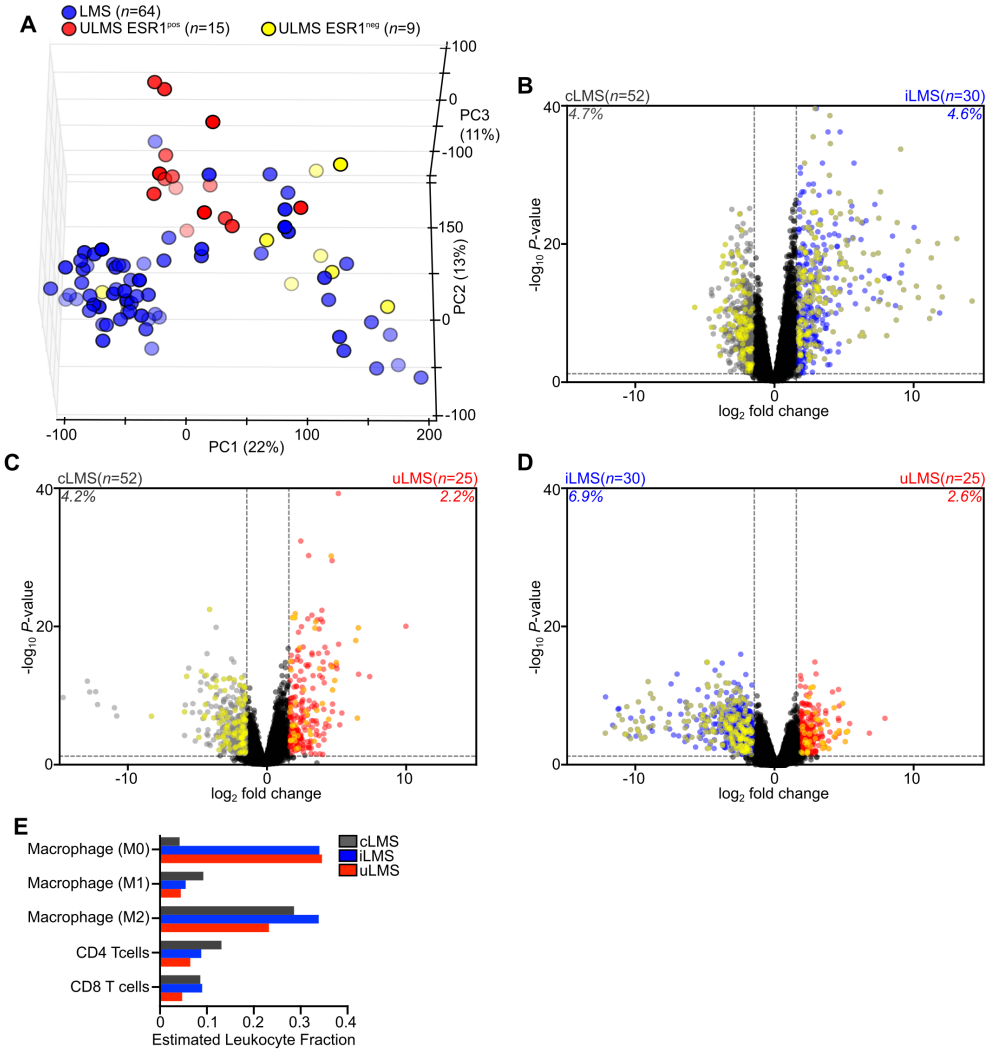

Figure S2. A unique subset of genes differentiates molecular subtypes of LMS. **A,** PCA of 88 TCGA LMS samples, with colors stratifying ELMS (blue) and ULMS with positive or negative ESR1 expression (red and yellow, respectively). **B-D,** Differential expression of RNA-seq data from the top 10,000 expressed genes comparing cLMS and iLMS (*B*), cLMS and uLMS (*C*) and iLMS and uLMS (*D*). Genes highlighted in yellow are uniquely expressed in each subtype. The percent of differentially expressed genes is indicated for each subtype comparison. **E,** CIBERSORT analysis of macrophage, CD4 and CD8 T cell subsets across LMS subtypes.

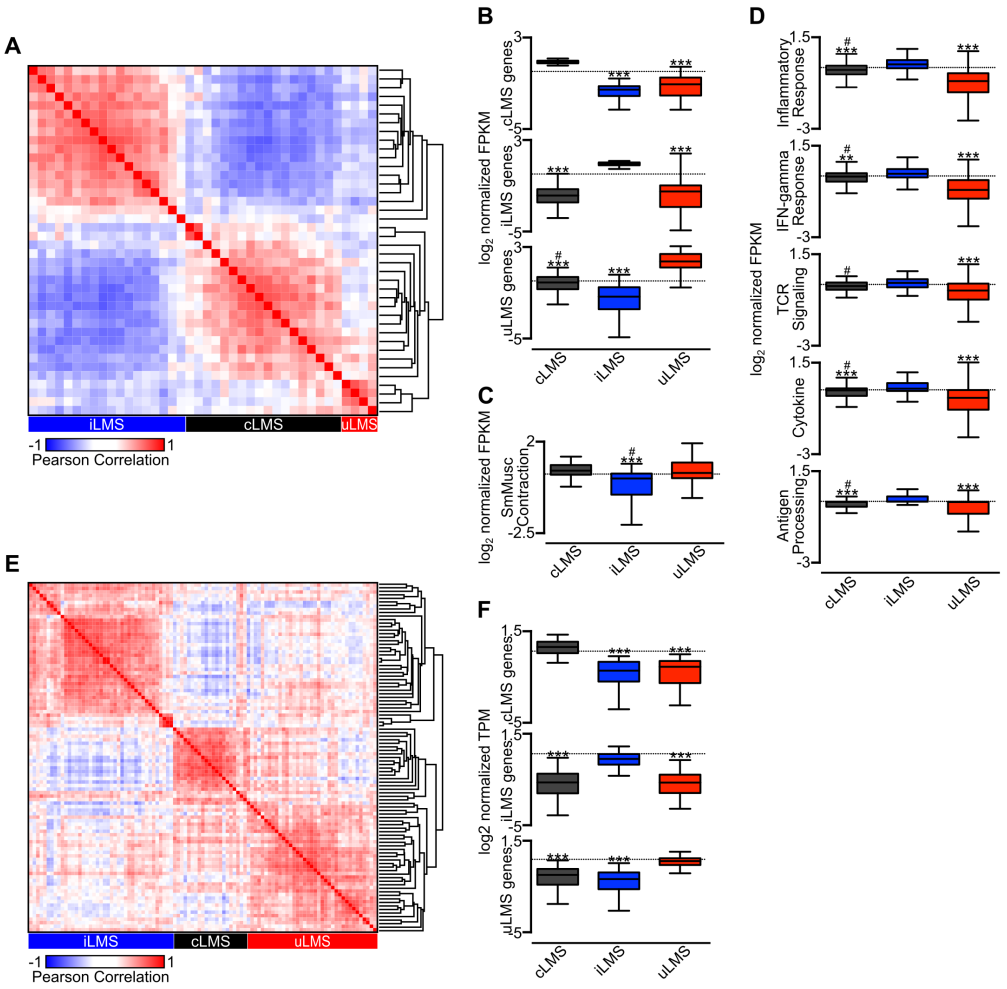

Figure S3. Validation of LMS subtypes in separate cohorts. **A,** Pearson correlation matrix of the top 100 most differentially expressed genes identified in the primary cohort in an independent RNA-seq dataset of 40 LMS samples. The LMS subtype represented by each cluster is indicated. **B,** Log_2 FPKM values normalized to the mean of all LMS samples in the independent RNA-seq data set. For each box plot, the indicated subtype-specific gene list was used to compare expression values across subtypes. **C,** Log_2 normalized FPKM values across LMS subtypes for the KEGG Vascular Smooth Muscle Contraction gene list in the independent RNA-seq data set. **D,** Log_2 normalized FPKM values across LMS subtypes for the indicated gene sets including Hallmark Inflammatory Response, Hallmark IFN-gamma Response, KEGG TCR Signaling Pathway, KEGG Cytokine-Cytokine Receptor Interaction and KEGG Antigen Processing and Presentation in the independent RNA-seq data set. **E,** Pearson correlation matrix of the top 100 most differentially expressed genes identified in the primary cohort in an independent 3SEQ dataset of 99 LMS samples. The LMS subtype represented by each cluster is indicated. **F,** Log_2 TPM values normalized to the mean of all LMS samples in the 3SEQ data set. For each box plot, the indicated subtype-specific gene list was used to compare expression values across subtypes. Data were analyzed by one-way ANOVA with Tukey's *post-hoc* test (compared to the indicated LMS subtype; **, $P<0.01$; ***, $P<0.001$; compared to the non-reference LMS subtype; #,$P<0.05$).

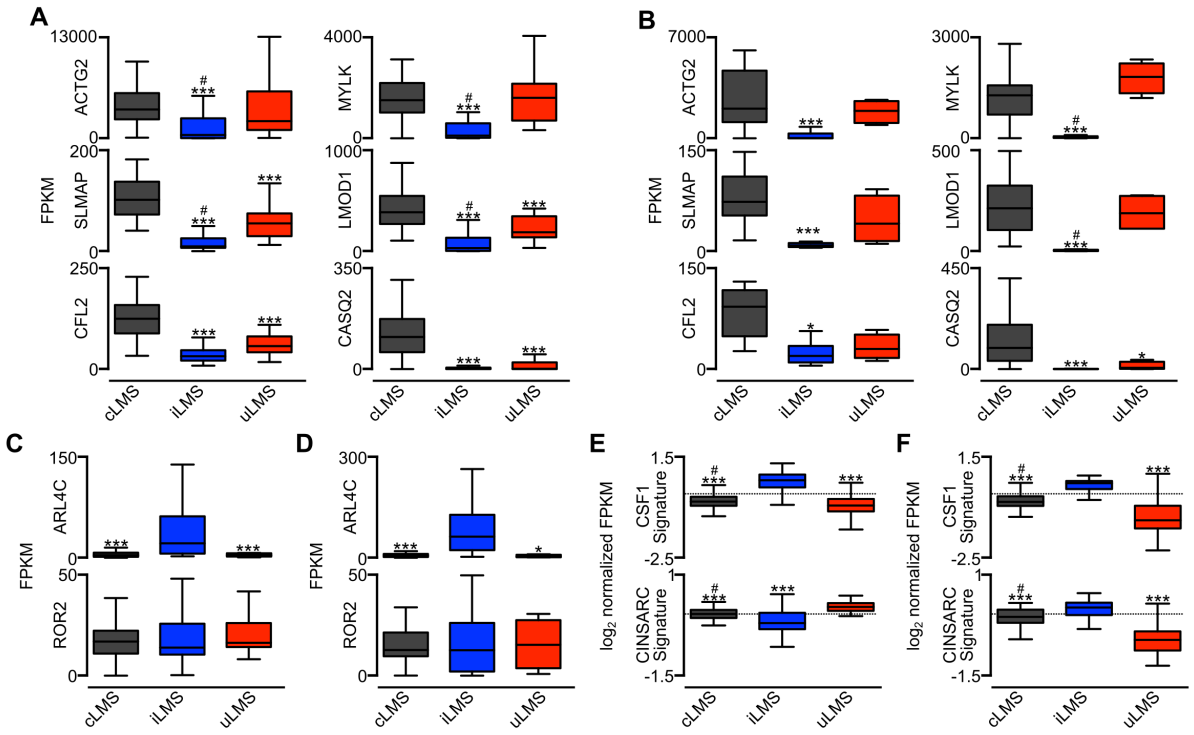

Figure S4. Expression of previously described differentiating transcripts in LMS subtypes. **A,** Box plots showing FPKM expression values of previously described differentiating transcripts in LMS subtypes in the primary RNA-seq cohort. **B,** Box plots showing FPKM expression values of previously described differentiating transcripts in LMS subtypes in the RNA-seq validation cohort. **C-D,** Box plots showing FPKM expression values of ARL4C and ROR2 in LMS subtypes in the primary (*C*) and validation (*D*) RNA-seq cohorts. **E-F**, Gene expression signatures for CSF1 and CINSARC in the primary (*E*) and validation (*F*) RNA-seq cohorts. Data were analyzed by one-way ANOVA with Tukey's *post-hoc* test (compared to the indicated LMS subtype; *,*P*<0.05; ***, *P*<0.001; compared to the non-reference LMS subtype; #,*P*<0.05).

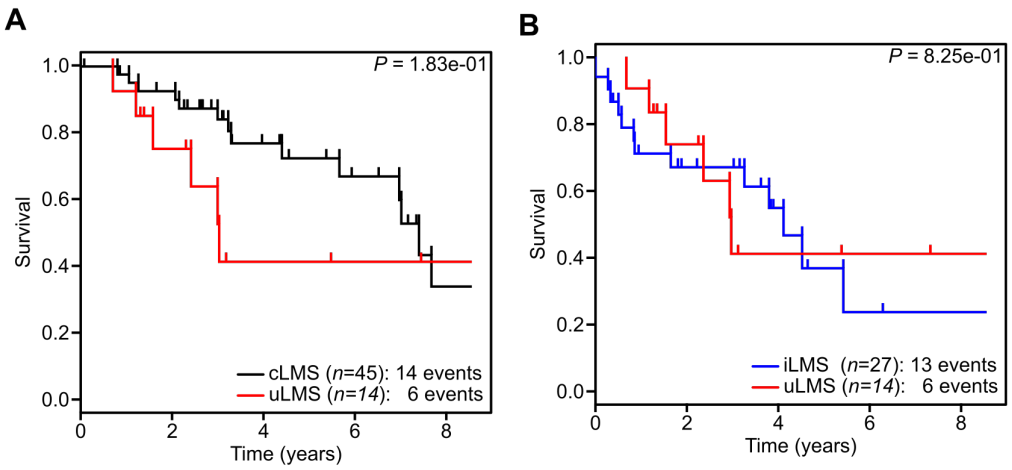

Figure S5. Kaplan-Meier analysis of LMS subtypes. **A,** Kaplan-Meier analysis of disease-specific survival comparing cLMS and uLMS subtypes. **B,** Kaplan-Meier analysis of disease-specific survival comparing iLMS and uLMS subtypes. Survival data were analyzed by log-rank test.

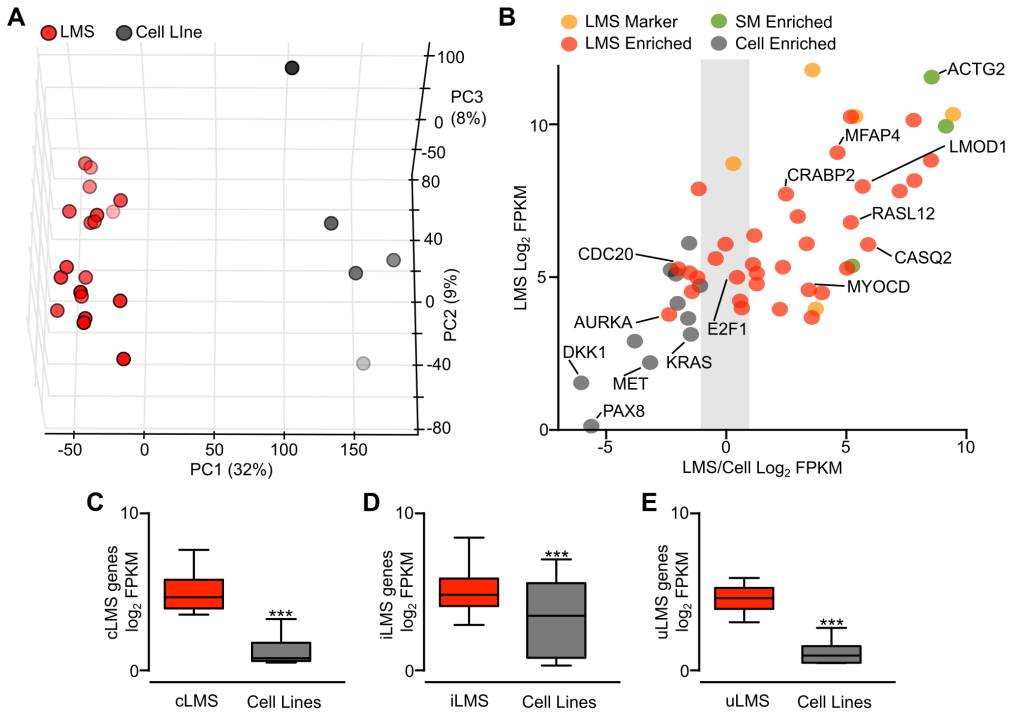

Figure S6. LMS cell lines lack an LMS-related gene expression program. **A,** PCA of LMS tumors (red, *n*=19) and cell lines (gray, *n*=5), all sequenced using identical methods at the study's primary institution. **B,** Plot of LMS $\log_2$ FPKM RNA-seq expression versus the $\log_2$ ratio of LMS (*n*=147) to cell line (*n*=5) FPKM for select genes enriched in LMS, smooth muscle or cell lines. The gray box includes values below a 2-fold ratio change. **C-E,** For each box plot, the indicated subtype-specific gene list was used to compare $\log_2$ FPKM RNA-seq expression values between the indicated LMS subtype and cell lines. Data were analyzed by unpaired t-test (compared to the indicated LMS subtype; ***, *P*<0.001).

Table S1. TCGA cancer subtypes.

Table S2. cBioPortal cancer subtypes.

Table S3. LMS genes within recurrently amplified regions.

Table S4. LMS RNA-seq samples.

Table S5. LMS subtype-specific transcript expression levels.

Table S6. LMS and cell line transcript expression levels.