S1 Fig. Flow chart of data analysis

**Training set**
- Moffitt cohort (early stage PDAC: n=63): Microarray platform
  - Sparse PCA (screen out noise genes)
  - 689 probesets
  - Univariate analysis of Cox model
  - Selection criteria: 25% FDR and associated with poor OS
  - 15 genes (18 probesets): PC1 scoring system

**Validation set**
- External cohort (Stratford et al cohort: early stage PDAC: n=102): Microarray platform
  - Test PC1 score
- Technical validation cohort (subset of Moffitt cohort: n=53): NanoString platform
  - Correlation analysis to screen out false positive genes
- Test for random noise gene signature

**Development of gene signature**