Time to Recurrence and Survival in Serous Ovarian Tumors Predicted from Integrated Genomic Profiles

Parminder K. Mankoo, Ronglai Shen, Nikolaus Schultz, Douglas A. Levine and Chris Sander

mRNA Gene Expression

Data from three different microarray expression platforms (Agilent 244K, Affymetrix Exon and Affymetrix U133A) were combined and normalized as reported in the TCGA paper [1]. All 11,864 genes from the unified TCGA dataset (derived from genes represented on all three platforms) were utilized for Cox Lasso regression analysis. All mRNA expression datasets were mean-centered. Mean-centered expression values were calculated by subtracting the mean expression value across patients from the gene estimate and dividing by the standard deviation across patients.

External data preparation Four external mRNA expression datasets were used to evaluate the performance of TCGA-data based signatures:

Tothill data: mRNA expression data for 243 tumor samples and 17255 genes [3]. We used only the 174 high-grade serous ovarian cancer samples to match the TCGA cohort.
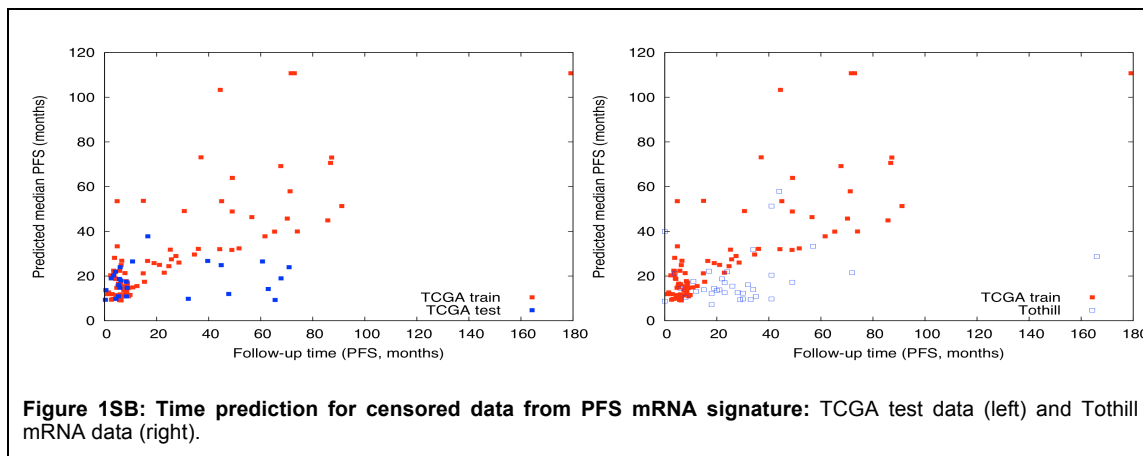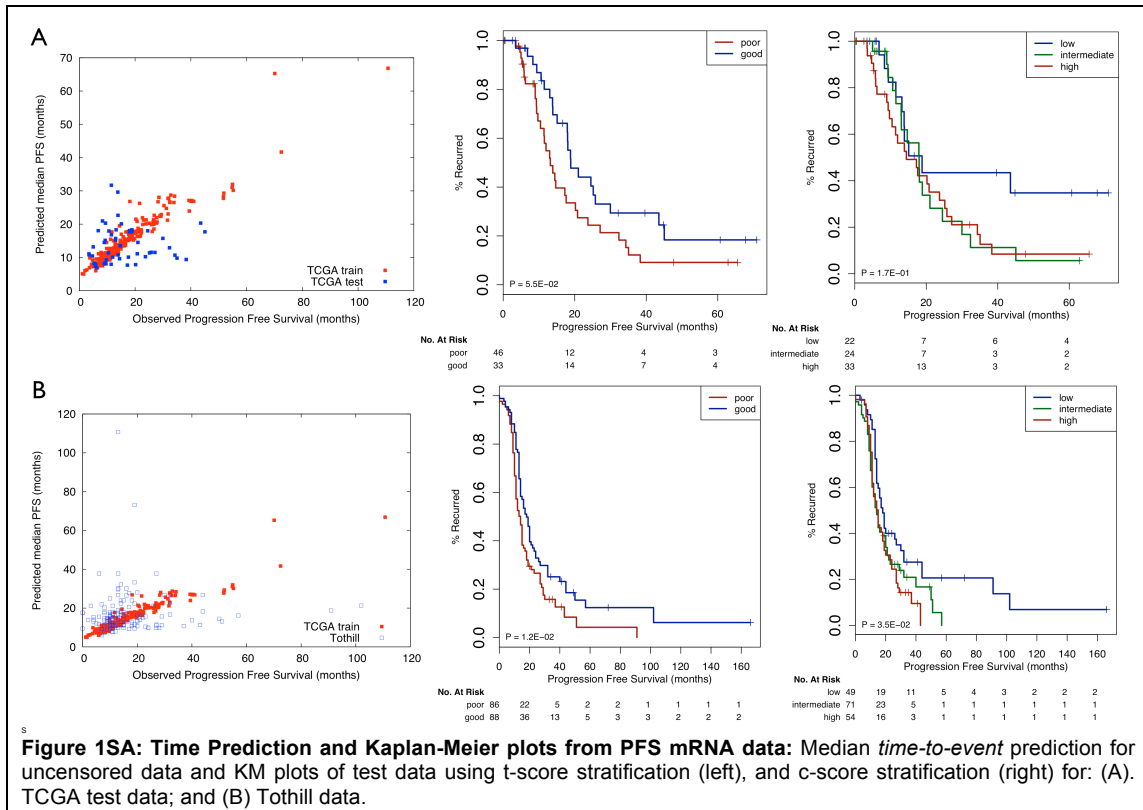*Dressman data:* mRNA expression data with overall survival information for 117 tumor samples and 12113 genes [4].
Bonome data: mRNA expression data for 166 tumor samples and 12114 genes [5]. Overall survival data was available for 163 patients and progression free survival data for 138 patients (clinical data as of February 4, 2008).
Yoshihara & Tanaka data: mRNA expression data for 110 patients and 18177 probes [6]. For each gene with multiple probes, we selected the probe with the largest sum of the squares of the individual expression values, resulting in 13154 unique genes.

Progression Free Survival: At λ=.830, we selected 181 features (Table ST4 [7]), cv.CPE = 0.79. Tertile stratification on the training data led to a p-value of 0.17 (c-score) and t-score stratification led to a p-value of 0.05, CPE.test = 0.77. Using the Tothill data set, tertile stratification based on c-score led to a p-value = 0.035 (and t-score p-value = 0.012) and CPE.tothill = 0.77. Our gene signature was not applicable to the Yoshihara & Tanaka dataset, as gene expression values were only available for 112 of the 181 genes in the progression-free-survival signature (Figure 1SA, 1SB).
Survival: At λ=.839, we selected 219 features (Table ST5 [8]), cv.CPE= 0.81. Tertile stratification based on c-score led to a p-value = 0.70 and t-score stratification led to a p-value of 0.09, CPE.test = 0.80. Using the Bonome dataset, tertile stratification led to a p-value = 0.05 (c-score) and p=0.18 (t-score) and CPE.bonome=0.75; using the Tothill dataset, tertile stratification led to a p-value = 0.048 (c-score) and p=0.014 (t-score), and CPE.tothill= 0.78; and using the Dressman dataset, tertile stratification led to a p-value = 0.008 (c-score) and p=0.033 (t-score). Our survival signature was not applicable to

Figure 1SA: Time Prediction and Kaplan-Meier plots from PFS mRNA data: Median *time-to-event* prediction for uncensored data and KM plots of test data using t-score stratification (left), and c-score stratification (right) for: (A). TCGA test data; and (B) Tothill data.

Figure 1SB: Time prediction for censored data from PFS mRNA signature: TCGA test data (left) and Tothill mRNA data (right).

Yoshihara & Tanaka dataset as gene expression values were only available for 143 of the 219 genes in the overall survival signature (Figures 1SC, 1SD).

2
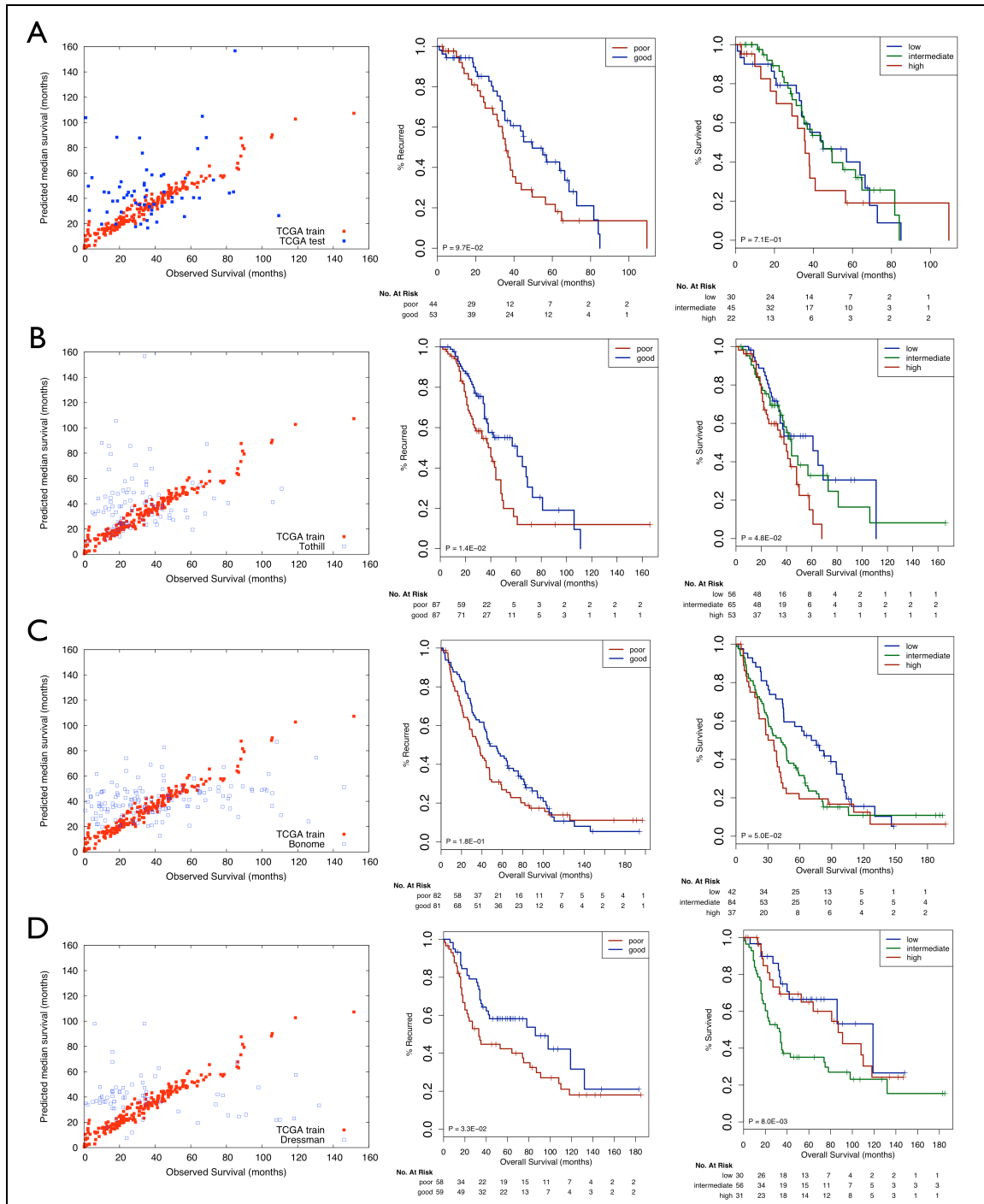
Figure 1SC: Time Prediction and Kaplan-Meier plots from OS mRNA signature: (A) TCGA test data; (B) Tothill dataset; (C) Bonome dataset and (D) Dressman dataset. KM plots using t-score (left) and c-score (right).

Figure 1SD: Prediction of median time-to-event for censored data from mRNA expression signatures: Median *time-to-event* prediction and comparison with follow-up times: A). TCGA test data; (B) Tothill dataset; (C) Bonome dataset; and, (D) Dressman dataset.
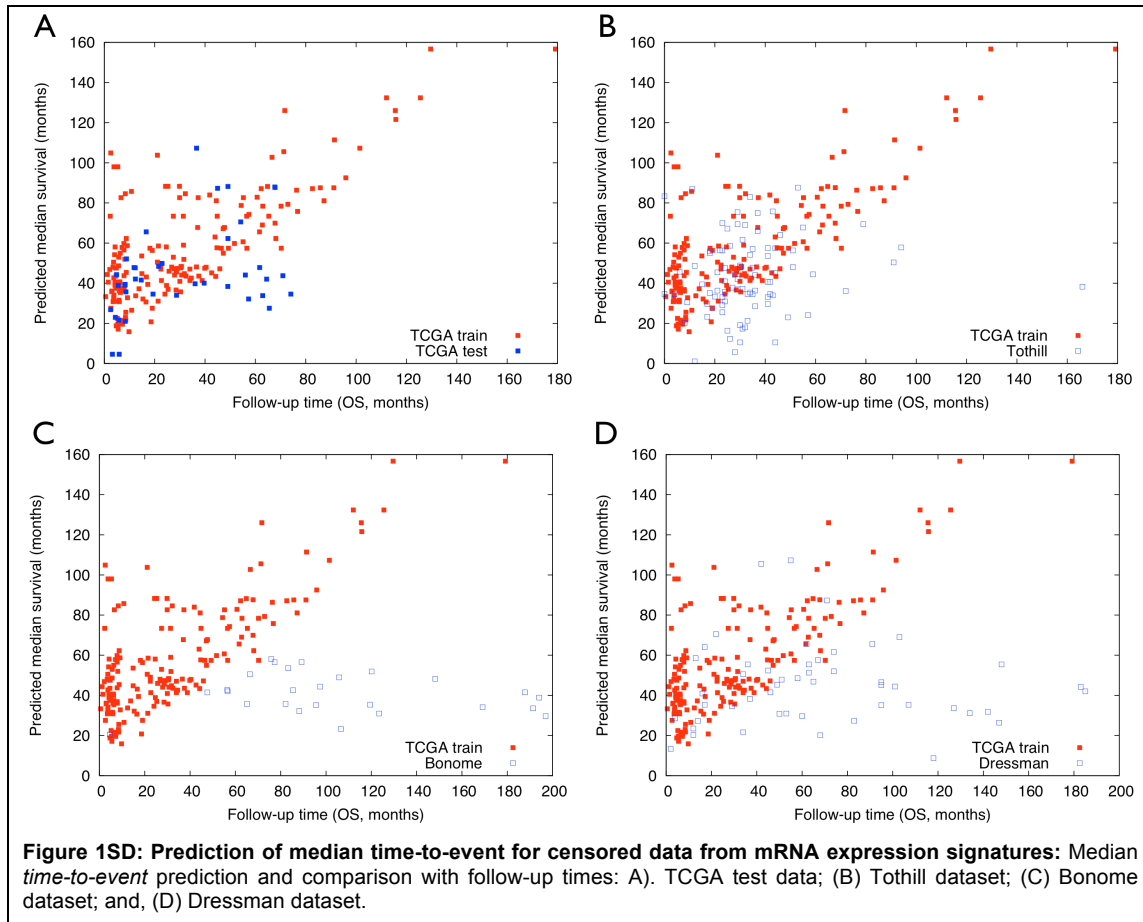
Functional and Disease association for genes in mRNA signatures

In order to prioritize functionally relevant genes and statistically and significantly associated canonical pathways within our mRNA signatures, we utilized Ingenuity pathway analysis (IPA) [7].

mRNA PFS signature: The top significant canonical pathways from mRNA TTP signature include: Granzyme A signaling; cardiomyocyte differentiation via BMP receptors; glycerolipid metabolism; Angrin interactions at neuromuscular junction; B cell development; HER-2 signaling in breast cancer and such (Figure 1SE). Other important categories (not statistically significant) were Androgen signaling and DNA double-strand break repair by homologous recombination. Statistically significant biological function categories include cell-to-cell signaling and interaction; molecular transport; lipid metabolism; cell death; cancer; and DNA replication, recombination and repair.

<u>mRNA OS signature</u>:

The top significant canonical pathways from the mRNA OS signature include: LXR/RXR activation; IL-6 and IL-10 signaling; PKA and p38MAPK signaling; TNFR1 signaling and PPAR signaling and such (Figure 1SF). General biological function categories (statistically significant) include: Inflammatory disease; lipid metabolism; drug metabolism, reproductive system disease; cell-to-cell signaling and interaction; and DNA replication, recombination and repair etc.

Figure 1SE: Canonical pathway analysis of 181 PFS mRNA features: Top functions with p-value cut-off < 0.1 are depicted.

Pathway	Genes
LXR/RXR Activation	*TNF, IL1F6, APOC4, IL1R1, APOC2, ABCG4, LBP, IL1R2*
LPS/IL-1 Mediated Inhibition of RXR Function	*TNF, ABCB9, APOC4, ALDH1A3, IL1R1, UST, CYP2A6 (includes others), APOC2, LIPC, LBP, IL1R2*
Hepatic Fibrosis / Hepatic Stellate Cell Activation	*VEGFA, TNF, IL1R1, FLT1, MYH4, LBP, CCR7, IL1R2*
Amyotrophic Lateral Sclerosis Signaling	*VEGFA, GRINA, CAPN6, GRID2, CACNA1C, PPP3CA*
IL-10 Signaling	*TNF, IL1F6, IL1R1, LBP, IL1R2*
Synaptic Long Term Potentiation	*GRINA, ATF4, PPP3R1, CACNA1C, PPP3CA, GRM3*
p38 MAPK Signaling	*PLA2G2D, TNF, IL1F6, ATF4, IL1R1, IL1R2*
Role of Cytokines in Mediating Communication between Immune Cells	*IL22, TNF, IL1F6, IFNA6*
Protein Kinase A Signaling	*PDE11A, ATF4, PYGB, PPP3R1, MYH4, CNGB3, RYR3, FLNC, GNAS, PPP3CA*
cAMP-mediated signaling	*PDE11A, ATF4, PPP3R1, CNGB3, DRD3, GNAS, PPP3CA, GRM3*
IL-6 Signaling	*TNF, IL1F6, IL1R1, LBP, IL1R2*
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	*VEGFA, TNF, IL1F6, ATF4, GPR50, IL1R1, PPP3R1, CEBPE, PPP3CA, IL1R2*
ILK Signaling	*VEGFA, TNF, ATF4, RHOT1, PPP2R2B, MYH4, FLNC*
Synaptic Long Term Depression	*PLA2G2D, PPP2R2B, GRID2, RYR3, GNAS, GRM3*
Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza	*TNF, IL1F6, IFNA6*
NF-κB Signaling	*TNF, IL1F6, IL1R1, FLT1, TANK, IL1R2*
Calcium Signaling	*GRINA, ATF4, PPP3R1, MYH4, RYR3, PPP3CA*
TNFR1 Signaling	*TNF, CASP2, TANK*
FXR/RXR Activation	*TNF, IL1F6, APOC2, LIPC*
Maturity Onset Diabetes of Young (MODY) Signaling	*CACNA1C, GAPDH*
Hepatic Cholestasis	*TNF, IL1F6, IL1R1, LBP, IL1R2*
PPAR Signaling	*TNF, IL1F6, IL1R1, IL1R2*
Glutamate Receptor Signaling	*GRINA, GRID2, GRM3*
Death Receptor Signaling	*TNF, CASP2, TANK*
Atherosclerosis Signaling	*PLA2G2D, TNF, IL1F6, F3*
Activation of IRF by Cytosolic Pattern Recognition Receptors	*TNF, IFNA6, TANK*
Tight Junction Signaling	*TNF, SYMPK, PPP2R2B, MYH4, CLDN6*
PXR/RXR Activation	*TNF, ABCB9, CYP2A6*
fMLP Signaling in Neutrophils	*PPP3R1, NOX3, GNAS, PPP3CA*
Hypoxia Signaling in the Cardiovascular System	*UBE2J1, VEGFA, ATF4*
Linoleic Acid Metabolism	*CYP4F12, PLA2G2D, CYP2A6*
FLT3 Signaling in Hematopoietic Progenitor Cells	*ATF4, GAB2, RPS6KA2*
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	*TNF, IL1F6, IL1R1, PPP3R1, PPP3CA, IL1R2*
IL-3 Signaling	*GAB2, PPP3R1, PPP3CA*
Nitric Oxide Signaling in the Cardiovascular System	*VEGFA, FLT1, CACNA1C*
IL-8 Signaling	*VEGFA, FLT1, RHOT1, NOX3, GNAS*
Ephrin Receptor Signaling	*EFNA5, VEGFA, GRINA, ATF4, GNAS*
Phospholipase C Signaling	*PLA2G2D, ATF4, RHOT1, PPP3R1, GNAS, PPP3CA*
Cellular Effects of Sildenafil (Viagra)	*KCNN2, MYH4, CACNA1C, GNAS*
Starch and Sucrose Metabolism	*PYGB, AGL, UXS1*
Cardiac β-adrenergic Signaling	*PDE11A, PPP2R2B, CACNA1C, GNAS*
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	*IL22, TNF, IL1F6*
Communication between Innate and Adaptive Immune Cells	*TNF, IL1F6, CCR7*
Melanocyte Development and Pigmentation Signaling	*ATF4, RPS6KA2, GNAS*

-log(p-value)

0   1.5   3.0   4.5   6.0

Figure 1SF: Canonical pathway analysis of 219 OS mRNA features: Top functions with p-value cut-off < 0.1 are depicted.

References:

1. TCGA (2011) Integrated Genomic Analyses of Ovarian Carcinoma. Nature 474: 609-615.
2. Park MY, Hastie T (2007) L1-regularization path algorithm for generalized linear models. J R Statist Soc B 69: 659-677.
3. Tothill RW, Tinker AV, George J, Brown R, Fox SB, et al. (2008) Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 14: 5198-5208.
4. Dressman HK, Berchuck A, Chan G, Zhai J, Bild A, et al. (2007) An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. J Clin Oncol 25: 517-525.
5. Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, et al. (2008) A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. Cancer Res 68: 5478-5486.
6. Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H, et al. Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. PLoS One 5: e9615.
7. http://cbio.mskcc.org/~mankoo/ST4.txt
8. http://cbio.mskcc.org/~mankoo/ST5.txt
9. IPA Ingenuity Pathway Analysis; http://www.ingenuity.com.