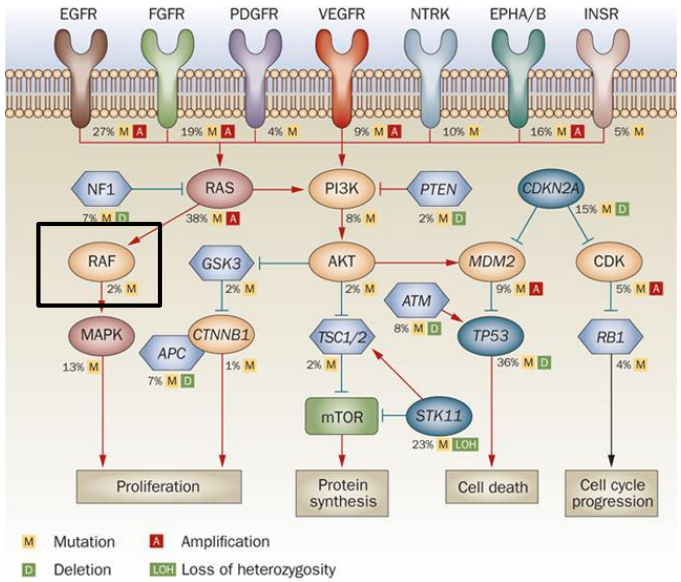
Most Random Gene Expression Signatures are Significantly Associated with Breast Cancer Outcome

Venet, et al.
PLoS Computational Biology, 2011

Molly Carroll

Biomedical Research Methods

1. Characterize mechanism in the model
2. Derive a marker that changes when the mechanism is altered
3. Show correlation of marker with disease outcome



Ding, L. et al. Nature (2008).

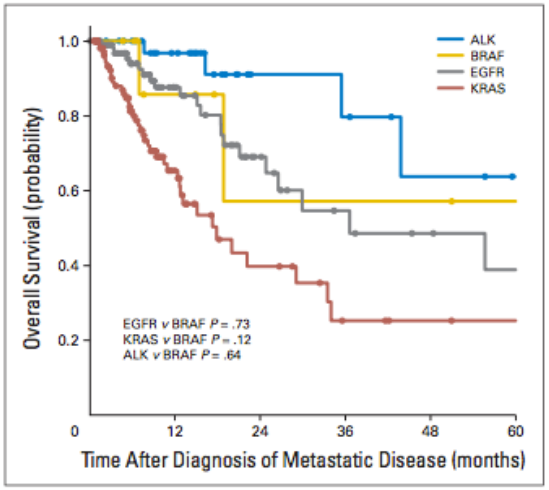


Fig 2. Kaplan-Meier curve for overall survival in patients with advanced stage (IIIB/IV) disease.

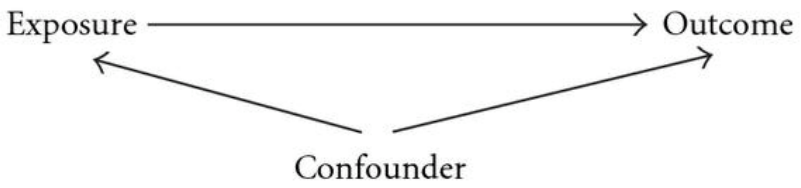
Paik, PK. et al. Journal of Clinical Oncology (2011)

Hazard Ratio:

$$\frac{\text{chances of an event (death) occurring in variable condition}}{\text{chances of an event (death) occurring in control condition}}$$

Confounding Variable Problem

- Some signatures are markers of mechanisms- ie. Epithelial mesenchymal transition
- Several signatures have equivalent prognostic outcome
- Are all mechanisms independent drivers or is there a confounding factor?



(Proliferation?)

Advances made in Methods

Step 2: Increase in genome-wide expression profiling leading to automated screen for markers and increased signatures

Step 3: Rise of cohorts with genome-wide expression profiles and patient follow-ups

Need to test negative controls to check relation of signature to outcome

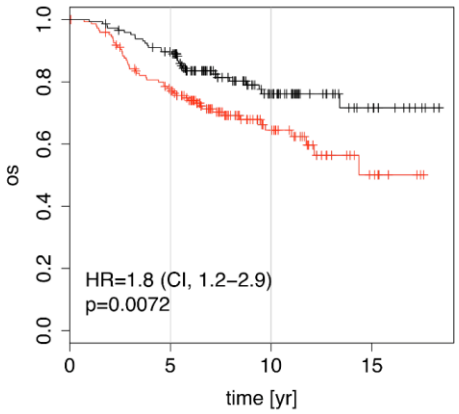
Typical: Signature of interest more strongly related to outcome than signature of no oncological rationale

Proposed: Random signature is more likely to be correlated with cancer outcome than not

Results- Fig 1

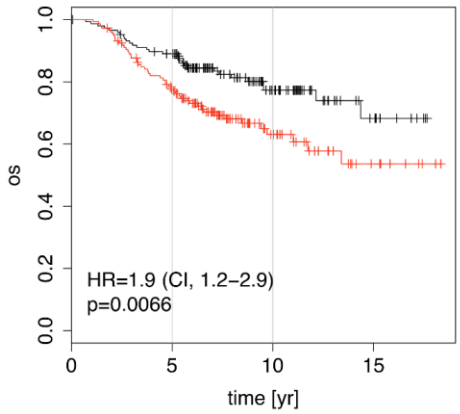
A

Post-prandial laughter sig.



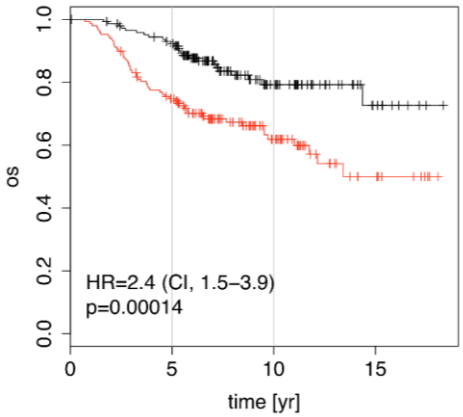
B

Localization of skin fibroblasts sig.

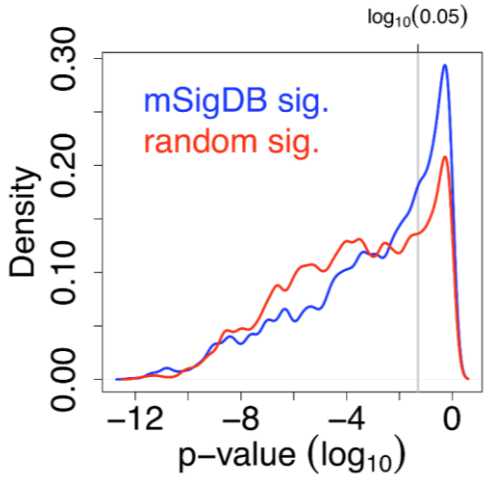


C

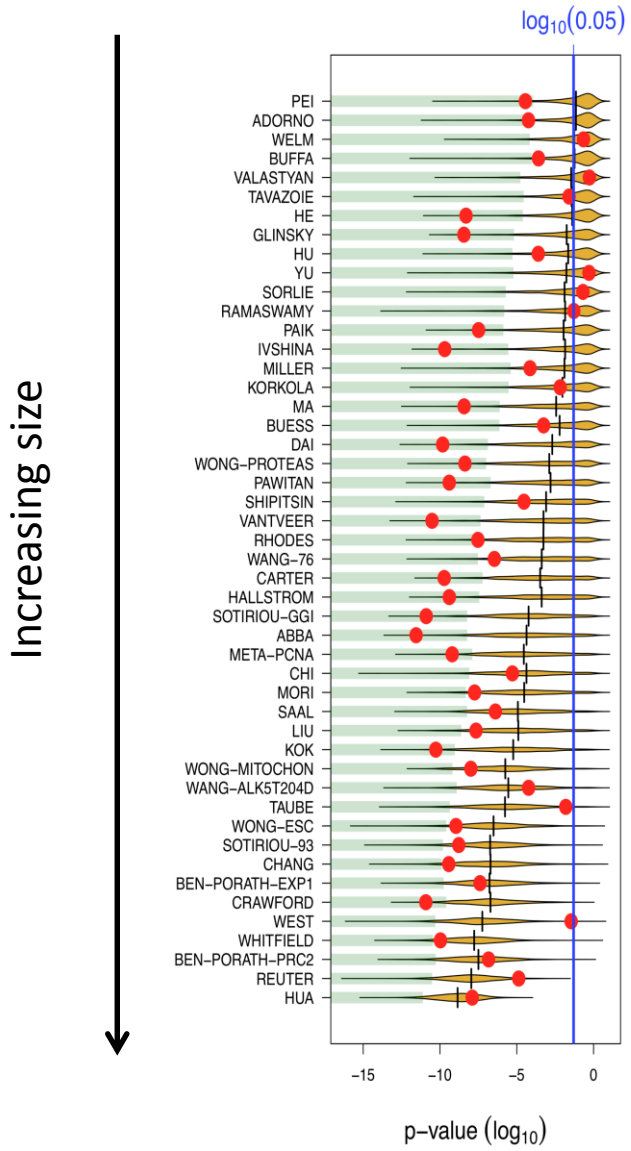
Social defeat in mice sig.



D

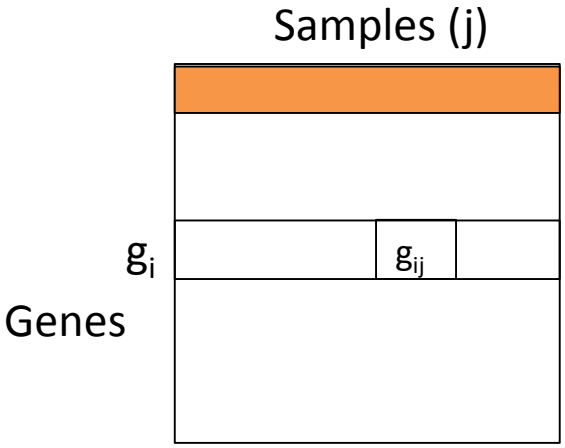


Results- Fig 2



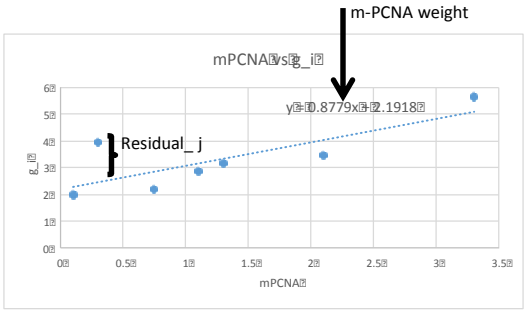
- Compared published breast cancer signature p-value of association with random signatures of equal size
- Used NKI cohort of patients

Methods: Meta-PCNA and Data Adjustment



m-PCNA index

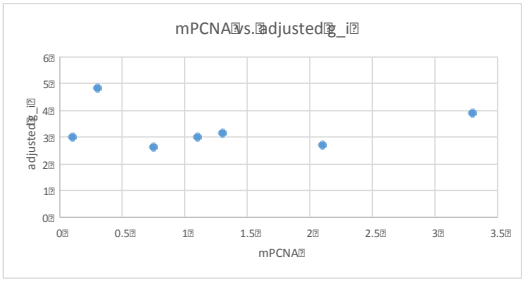
j	mPCNA_j	g_ij
1	0.1	1.957143
2	0.3	3.957143
3	0.75	2.157143
4	1.1	2.857143
5	1.3	3.157143
6	2.1	3.457143
7	3.3	5.657143



j	g_ij	linear fit	residual_j
1	1.957143	2.279579	-0.32244
2	3.957143	2.455137	1.502006
3	2.157143	2.850143	-0.693
4	2.857143	3.157369	-0.30023
5	3.157143	3.332927	-0.17578
6	3.457143	4.035159	-0.57802
7	5.657143	5.088507	0.568636

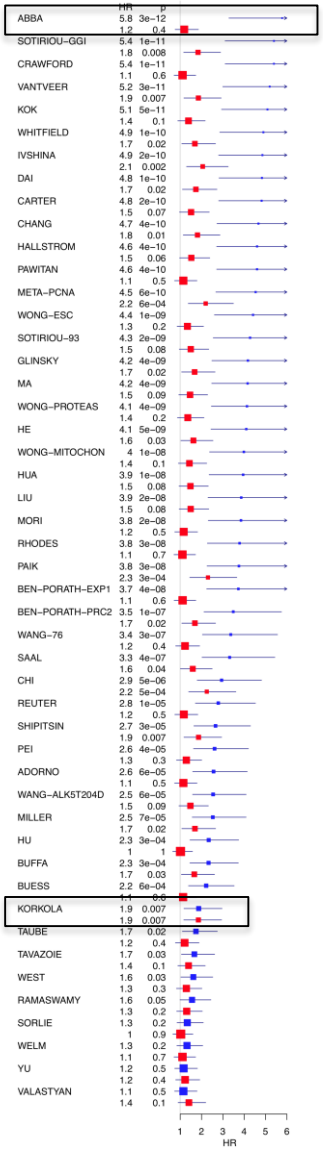
g_ij=weight*(mPCNAj) + intercept +error_ij

adj_g_ij = avg(g_i) + error_ij

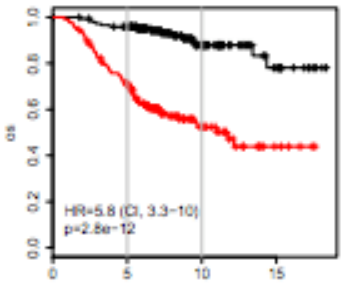


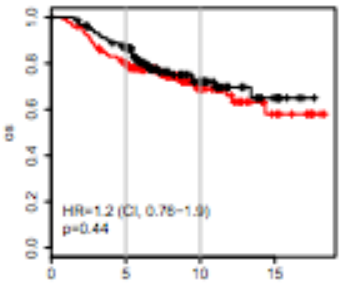
- Pearson correlation between PCNA and all genes in by Ge et al. via genome-wide expression profiling of healthy tissues
- 131 genes were top 1% that correlated with PCNA=> meta-PCNA sig.
- m-PCNA index of tissue: median expression of the genes
- Used linear regression (R's 'lm' function) to fit a sample's individual gene expression to m-PCNA gene

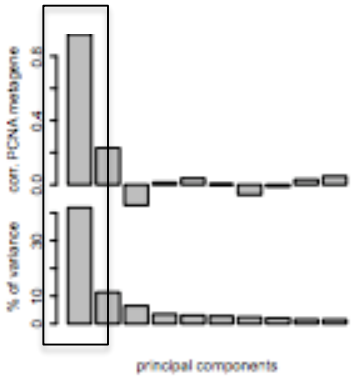
Results: Figure 3 and Supplmental



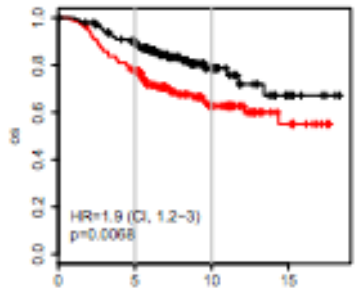
Abba Signature

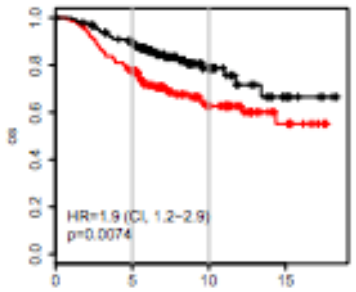


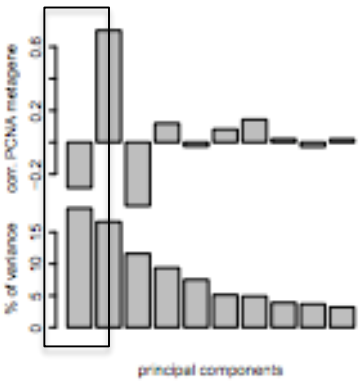




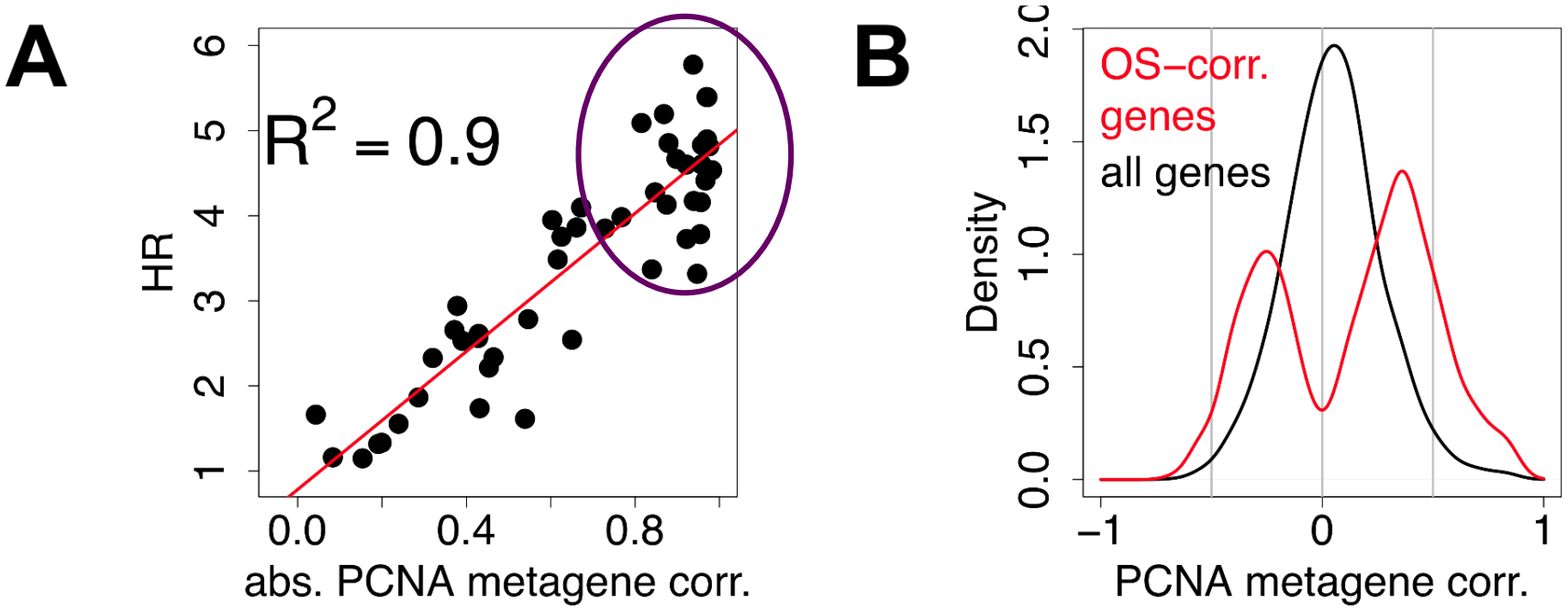
Korkola Signature



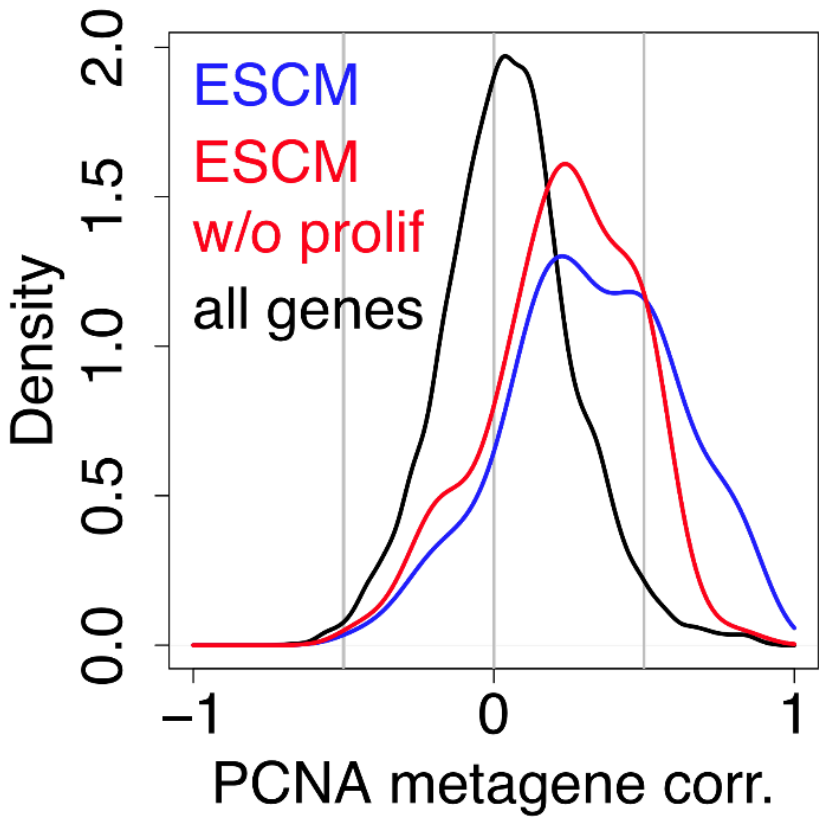




Results: Figure 4



Results: Figure 5



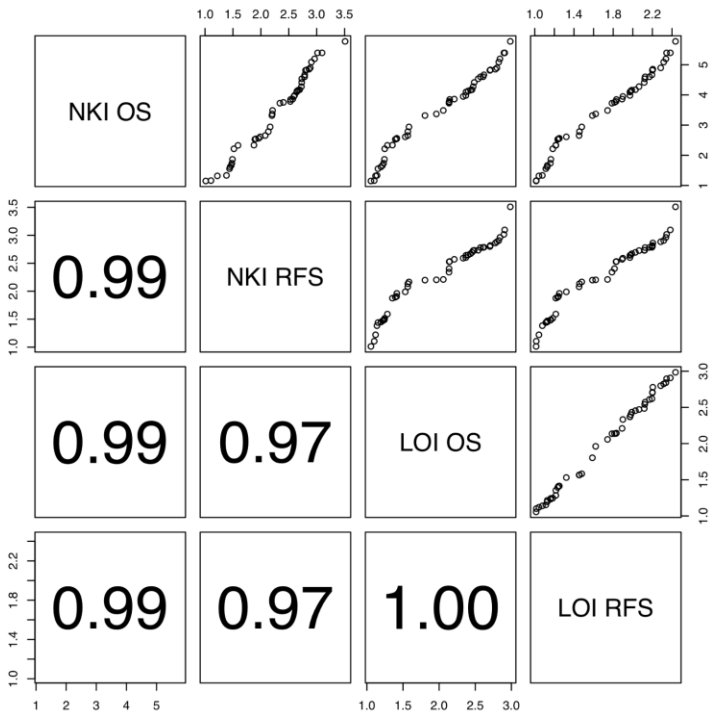
- ESCM: signature of gene sets associated with embryonic stem cell identity from Wong et al.
- Purging of cell cycle genes did not eliminate high correlation of ESCM with PCNA metagene

Correlations with meta-PCNA extend far beyond cell-cycle genes

Results: Figure 6

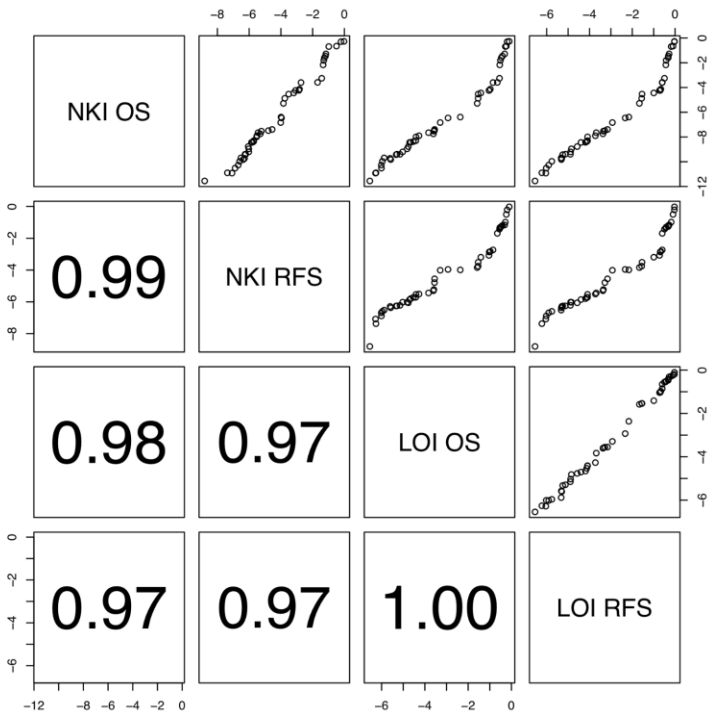
A

Hazard Ratio



B

Log rank p-values



Conclusions and Moving Forward

- Random single and multiple genes expression markers have high probability to be associated with BC outcome
- Most published signatures are not significantly more associated with outcome than random signatures
- Meta-PCNA metagene integrates most of the outcome-related information in BC transcriptome
- This information is present in 50% of the transcriptome and can't be removed by purging cell cycle genes from a signature

- Development of larger cohorts with various sub-types of a cancer included may help find better prognostic signatures
  - The NKI cohort represented bulk tumors from a wide spectrum of patients
  - Couldn't use NKI cohort to detect transcriptional signatures in specific cells (stromal, epithelial, etc) or patient groups (ER+, HER2 amplification)