Unveiling Survival Signatures:
Assessing Contributions of Transcriptional Outliers on Breast Cancer Survival

JAMES R. EVANS[1*], AMAAN M. JOGIA-SATTAR[1*], Jee Y. Han[2, 3, 4, 5], Zhuyu Qiu[2, 3, 4, 5], Paul C. Boutros[2, 3, 4, 5, 6]

UCLA Health — Jonsson Comprehensive Cancer Center

Introduction | Abstract

Despite their potential as prognostic biomarkers and therapeutic targets, outlier genes with unique expression patterns in cancer have not been comprehensively characterized in the context of patient survival. The integration of outlier gene signatures with clinical measurements enables more nuanced profiling of tumor-specific heterogeneity. Leveraging TCGA RNA-Seq and METABRIC microarray data, we constructed univariate and multivariate Cox proportional hazards models to investigate the impact of outlier genes on patient survival. Our findings shed light on the role of outlier genes as proxies for patient survival, demonstrating their relevance in characterizing breast cancer survival patterns. Exploration of clinical variables in association with outlier genes and their molecular mechanisms can unveil avenues for therapeutic interventions targeting tumor-specific transcriptional variations. Future endeavors include an extension of this methodology across cancer types and advancing pattern identification in pan-cancer outlier contributions for clinical and molecular data.
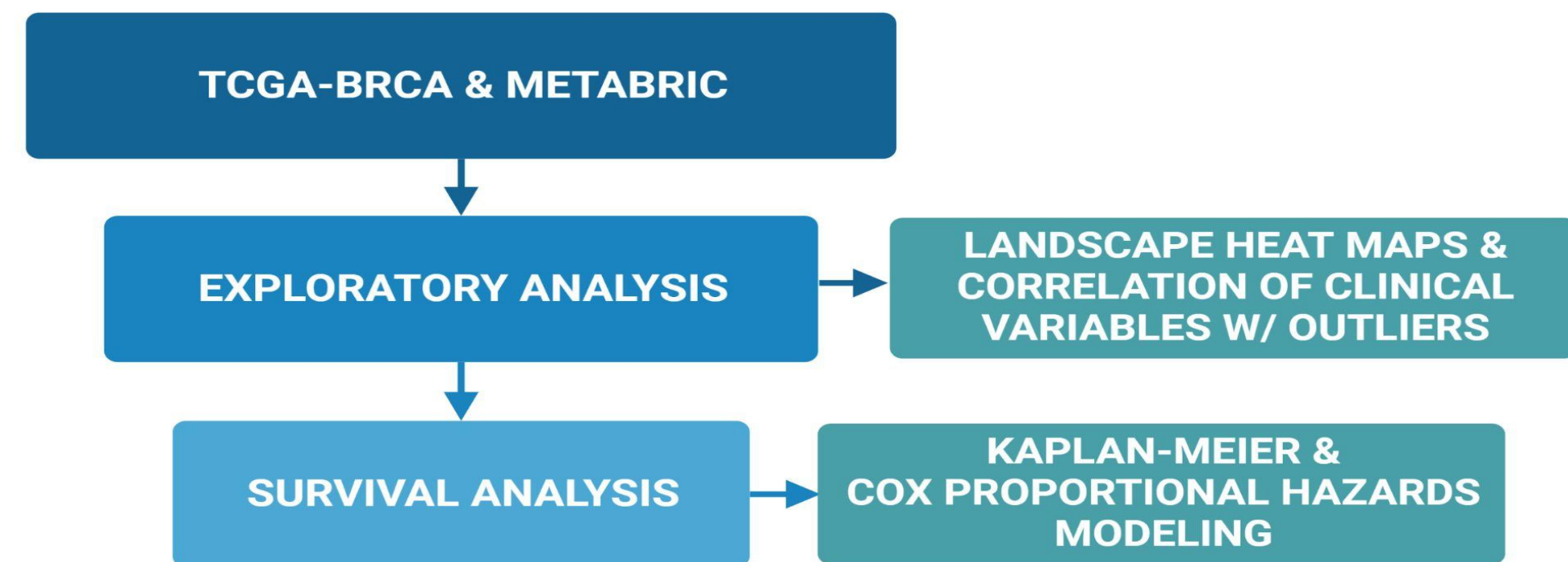
Methods | Schematic Workflow

Methods / Schematic Workflow

TCGA-BRCA Dataset:
- RNA-seq data from n = 1085 breast cancer patients, with 1187 identified outlier genes
- RNA abundance levels represented as Fragments Per Kilobase Million (FPKM)

METABRIC Dataset:
- Microarray data from n = 1991 breast cancer patients, with 4782 identified outlier genes
- RNA abundance levels represented as probe intensity values

Workflow: TCGA-BRCA & METABRIC → EXPLORATORY ANALYSIS → LANDSCAPE HEAT MAPS & CORRELATION OF CLINICAL VARIABLES W/ OUTLIERS; EXPLORATORY ANALYSIS → SURVIVAL ANALYSIS → KAPLAN-MEIER & COX PROPORTIONAL HAZARDS MODELING

References

1) Cancer Genome Atlas Research Network; Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113-20. doi: 10.1038/ng.2764. PMID: 24071849; PMCID: PMC3919969.
2) Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group; Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Apr 18;486(7403):346-52. doi: 10.1038/nature10983. PMID: 22522925; PMCID: PMC3440846.

BOUTROS LAB

Results | Discussion

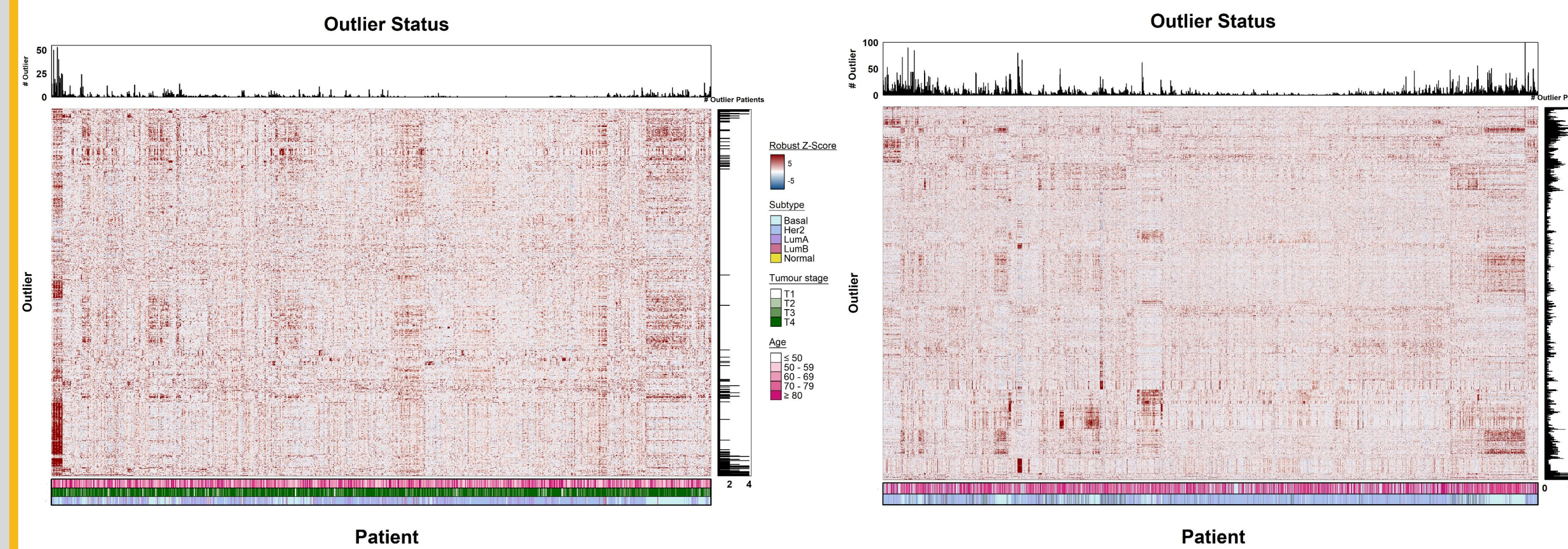### TCGA-BRCA & METABRIC Outlier Landscape

FIGURE 1. TCGA-BRCA Outlier Heat Map. Clustering of outliers by gene and patient

FIGURE 2. METABRIC Outlier Heat Map. Clustering of outliers by gene and patient

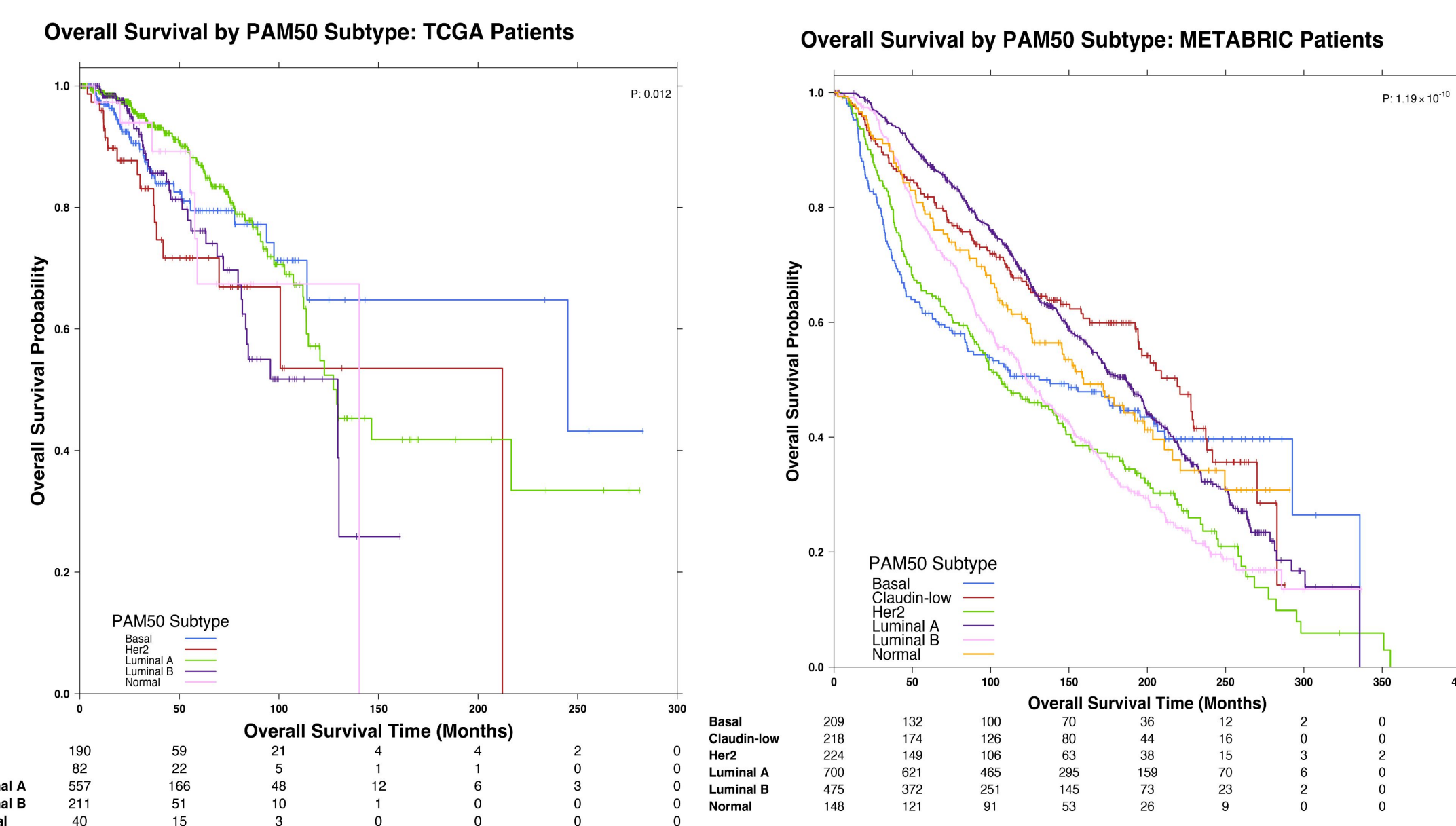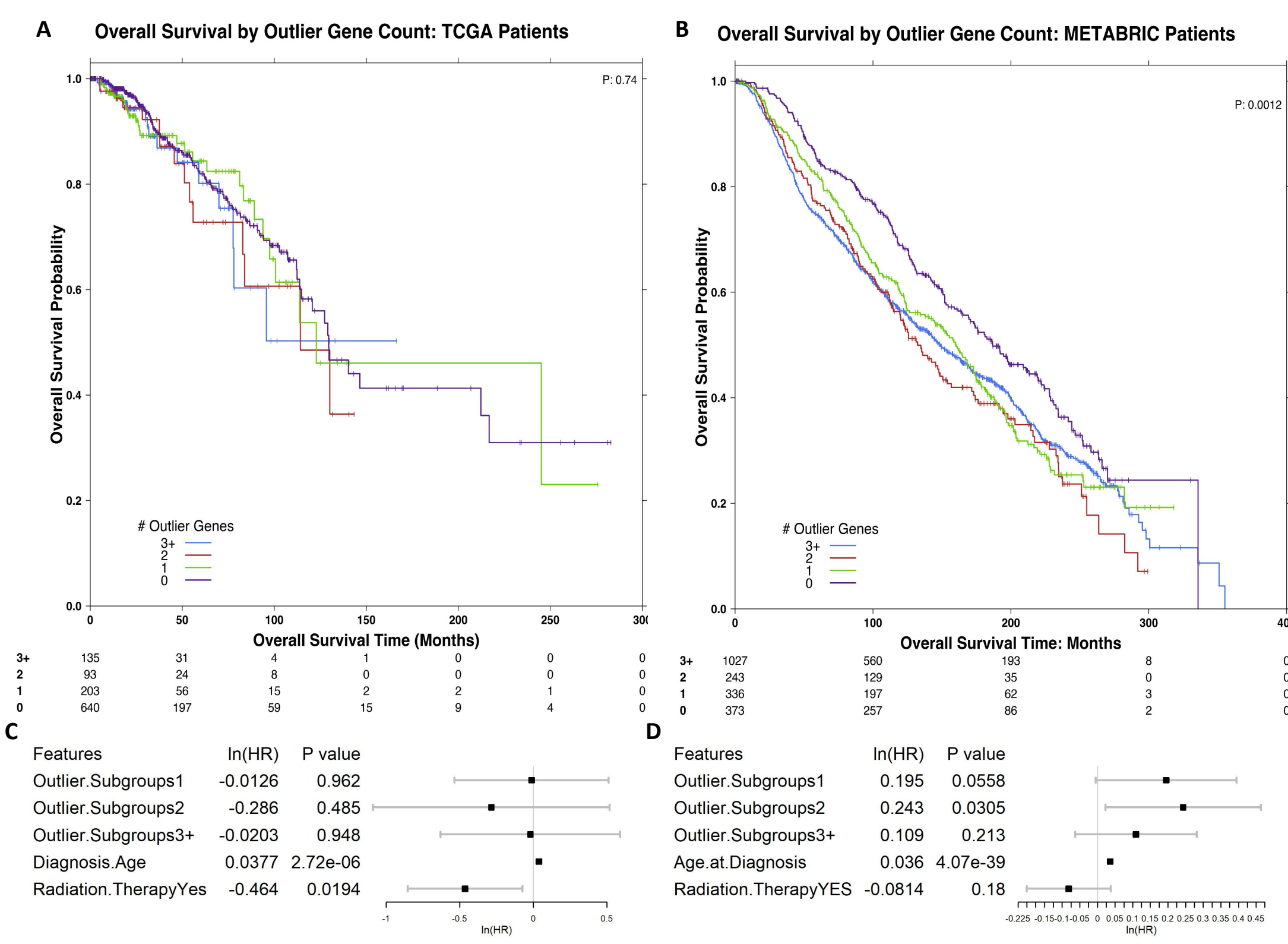### Survival Analysis: Outlier Gene Count and Molecular Subtype

FIGURE 5. Kaplan-Meier Survival Curves Stratified by PAM50 Molecular Cancer Subtype [Significance Level: α = 0.05]
A) TCGA-BRCA: Kaplan-Meier survival curves and log-rank test comparison for subtype strata
B) METABRIC: Kaplan-Meier survival curves and log-rank test comparison for subtype strata

FIGURE 6. Kaplan-Meier Survival Curves and Cox Proportional Hazards Models Stratified by Outlier Count [Significance Level: α = 0.05]
A) TCGA-BRCA: Kaplan-Meier survival curves and log-rank test comparison for outlier strata
B) METABRIC: Kaplan-Meier survival curves and log-rank test comparison for outlier strata
C) CoxPH Model Forest Plot for TCGA-BRCA Covariates: age, radiation therapy, prior diagnosis
D) CoxPH Model Forest Plot for METABRIC Covariates: age, radiation therapy

Features	ln(HR)	P value
Outlier.Subgroups1	-0.0126	0.962
Outlier.Subgroups2	-0.286	0.485
Outlier.Subgroups3+	-0.0203	0.948
Diagnosis.Age	0.0377	2.72e-06
Radiation.TherapyYes	-0.464	0.0194

Features	ln(HR)	P value
Outlier.Subgroups1	0.195	0.0558
Outlier.Subgroups2	0.230	0.0305
Outlier.Subgroups3+	0.109	0.213
Age.at.Diagnosis	0.036	4.07e-39
Radiation.TherapyYES	-0.0814	0.145

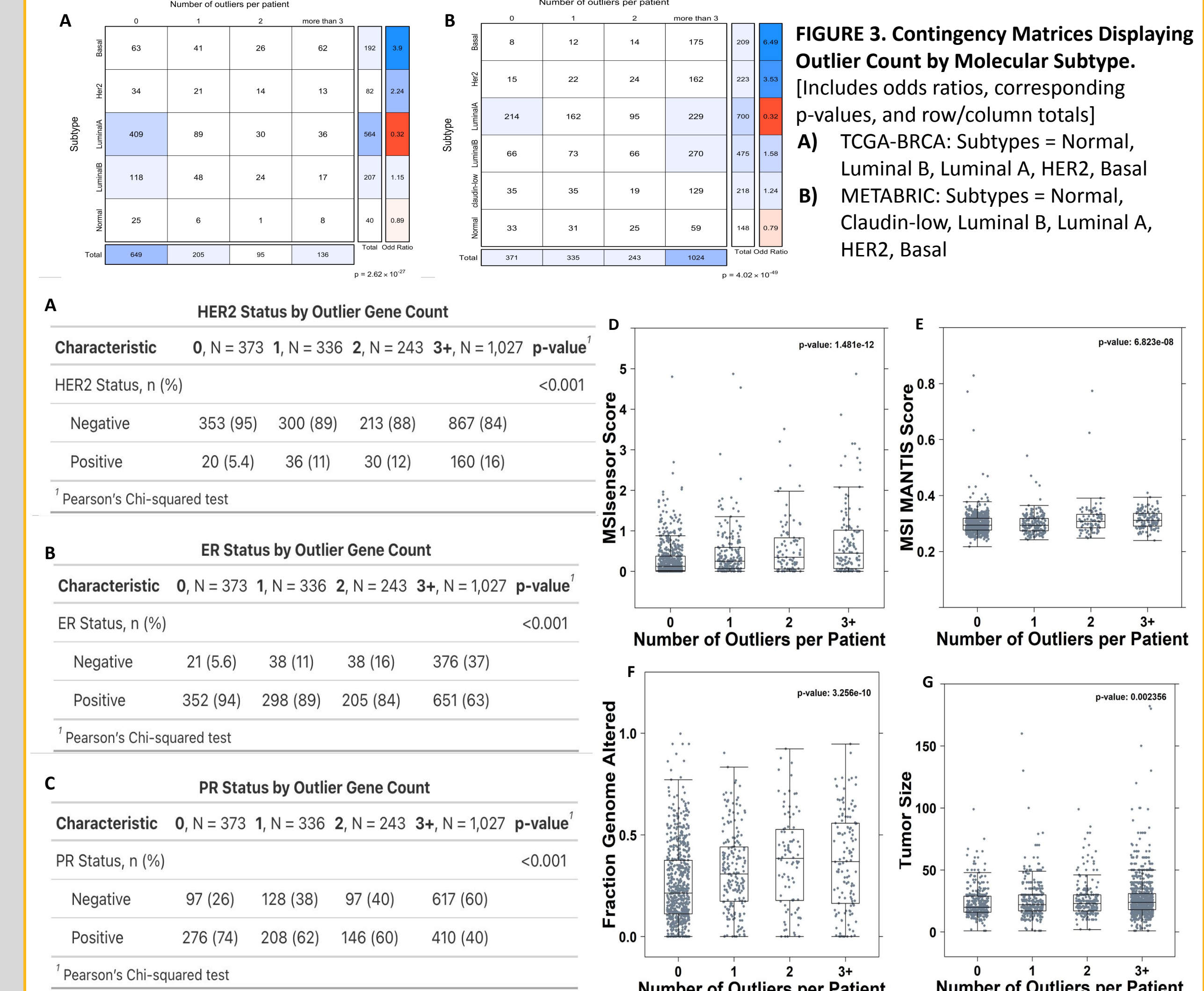### Clinical Variable Correlation

FIGURE 3. Contingency Matrices Displaying Outlier Count by Molecular Subtype. [Includes odds ratios, corresponding p-values, and row/column totals]
A) TCGA-BRCA: Subtypes = Normal, Luminal B, Luminal A, HER2, Basal
B) METABRIC: Subtypes = Normal, Claudin-low, Luminal B, Luminal A, HER2, Basal

Characteristic	0, N = 373	1, N = 336	2, N = 243	3+, N = 1,027	p-value
HER2 Status, n (%)					<0.001
Negative	353 (95)	300 (89)	213 (88)	867 (84)	
Positive	20 (5.4)	36 (11)	30 (12)	160 (16)	

Pearson's Chi-squared test

Characteristic	0, N = 373	1, N = 336	2, N = 243	3+, N = 1,027	p-value
ER Status, n (%)					<0.001
Negative	21 (5.6)	38 (11)	38 (16)	376 (37)	
Positive	352 (94)	298 (89)	205 (84)	651 (63)	

Pearson's Chi-squared test

Characteristic	0, N = 373	1, N = 336	2, N = 243	3+, N = 1,027	p-value
PR Status, n (%)					<0.001
Negative	97 (26)	128 (38)	97 (40)	617 (60)	
Positive	276 (74)	208 (62)	146 (60)	410 (40)	

Pearson's Chi-squared test

FIGURE 4. Relationship Between TCGA-BRCA & METABRIC Clinical Variables and Outlier Gene Count A-C) Relationship Between METABRIC Hormone Receptor Status and Outlier Gene Count. Contingency Tables displaying the relationship between: A) Human epidermal growth factor receptor 2 (HER2) and Outlier Gene Count B) Estrogen receptor (ER) and Outlier Gene Count C) Progesterone receptor (PR) and Outlier Gene Count Statistical significance was assessed using Pearson's Chi-Square Test, using significance level α = 0.05. D) TCGA-BRCA - MSI Sensor Score per Outlier Group (lower accuracy) E) TCGA-BRCA - MSI Mantis Score per Outlier Group (higher accuracy) F) TCGA-BRCA - Fraction of Genome Altered per Outlier Group G) METABRIC - Tumor Size per Outlier Group
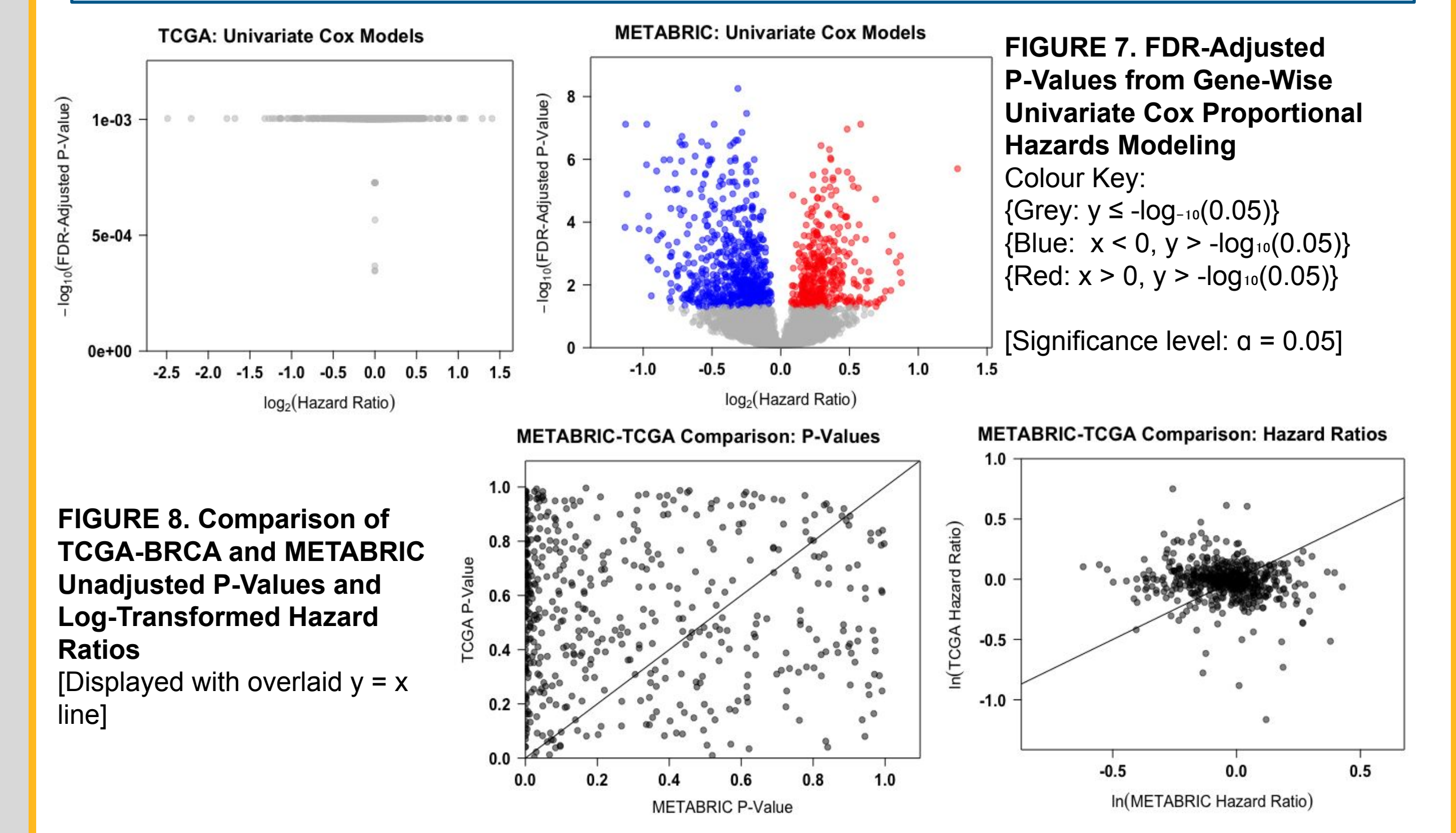
Survival Analysis: Outlier Gene Contributions

FIGURE 7. FDR-Adjusted P-Values from Gene-Wise Univariate Cox Proportional Hazards Modeling Colour Key:
{Grey: y ≤ -log₁₀(0.05)}
{Blue: x < 0, y > -log₁₀(0.05)}
{Red: x > 0, y > -log₁₀(0.05)}
[Significance level: α = 0.05]

FIGURE 8. Comparison of TCGA-BRCA and METABRIC Unadjusted P-Values and Log-Transformed Hazard Ratios [Displayed with overlaid y = x line]

Conclusion

- ER and PR status became increasingly negative with increasing outlier gene count.
- HER2 status became increasingly positive with increasing outlier gene count).
- METABRIC displayed reduced survival probability with increasing outlier gene count early post-diagnosis. TCGA had limited follow-up, requiring further investigation.
- Both diagnosis age and radiation therapy showed significant associations with survival in METABRIC covariate-adjusted Cox models, while diagnosis age was significantly associated in TCGA.
- METABRIC revealed significant outlier genes post-FDR correction. Limited significant outliers in TCGA; longer follow-up may clarify.
- Future endeavors include exploration of outlier genes' survival patterns in other cancer types for pan-cancer insight generation, as well as replication of survival analyses in cohorts with extended patient follow-up.

Acknowledgements | Funding

Author Affiliations

1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Department of Human Genetics, University of California, Los Angeles, USA
3 Department of Urology, University of California, Los Angeles, USA
4 Jonsson Comprehensive Cancer Center, University of California, Los Angeles, USA
5 Institute for Precision Health, University of California, Los Angeles, USA
6 Department of Medical Biophysics, University of Toronto, Toronto, Canada
* equal contributions