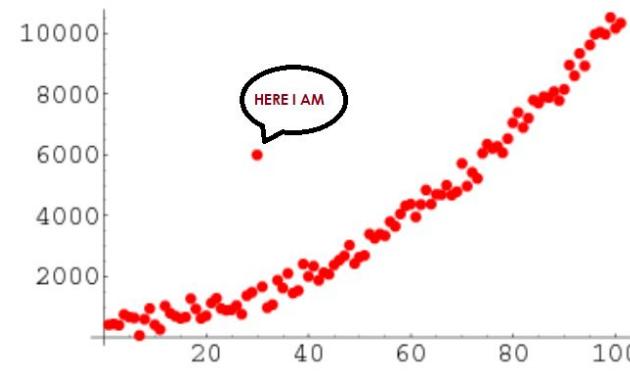
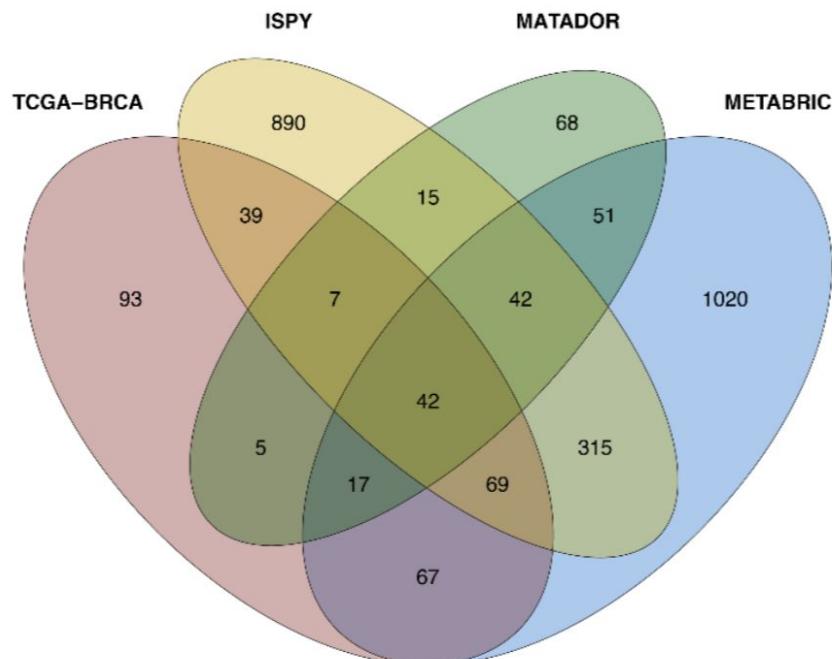


Unveiling Survival Signatures: Assessing Contributions of Transcriptional Outliers on Breast Cancer Survival

James R. Evans, Amaan M. Jogia-Sattar, Jee Y. Han, Zhuyu Qiu, Paul C. Boutros

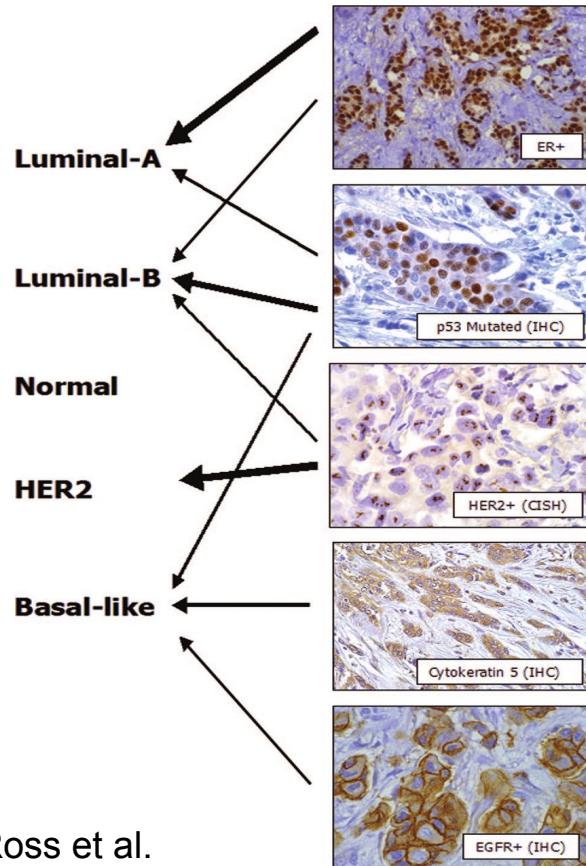
Context: Pan-Cancer Outlier Gene Analysis



TCGA RNA-seq (33 cancer types)

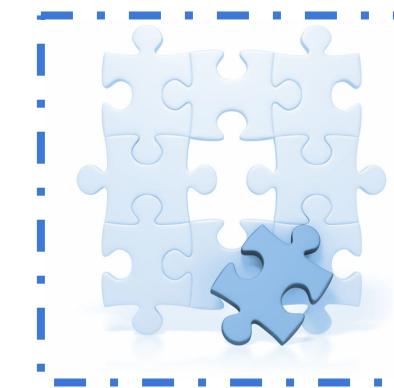
LAML	Acute Myeloid Leukemia	n = 151	LUAD	Lung Adenocarcinoma	n = 526
ACC	Adrenocortical Cancer	n = 79	LUSC	Lung Squamous Cell Carcinoma	n = 501
CHOL	Bile Duct Cancer	n = 36	SKCM	Melanoma	n = 103
BLCA	Bladder Cancer	n = 411	MESO	Mesothelioma	n = 86
BRCA	Breast Cancer	n = 1097	UVM	Ocular melanomas	n = 80
CESC	Cervical Cancer	n = 304	OV	Ovarian Cancer	n = 379
COAD	Colon Cancer	n = 470	PAAD	Pancreatic Cancer	n = 177
UCEC	Endometrioid Cancer	n = 548	PCPG	Pheochromocytoma & Paraganglioma	n = 178
ESCA	Esophageal Cancer	n = 161	PRAD	Prostate Cancer	n = 498
GBM	Glioblastoma	n = 168	READ	Rectal Cancer	n = 167
HNSC	Head and Neck Cancer	n = 500	SARC	Sarcoma	n = 262
KICH	Kidney Chromophobe	n = 65	STAD	Stomach Cancer	n = 375
KIRC	Kidney Clear Cell Carcinoma	n = 534	TGCT	Testicular Cancer	n = 150
KIRP	Kidney Papillary Cell Carcinoma	n = 288	THYM	Thymoma	n = 119
DLBC	Large B-cell Lymphoma	n = 48	THCA	Thyroid Cancer	n = 502
LIHC	Liver Cancer	n = 374	UCS	Uterine Carcinosarcoma	n = 56
LGG	Lower Grade Glioma	n = 529			

Background: Histopathological Subtyping



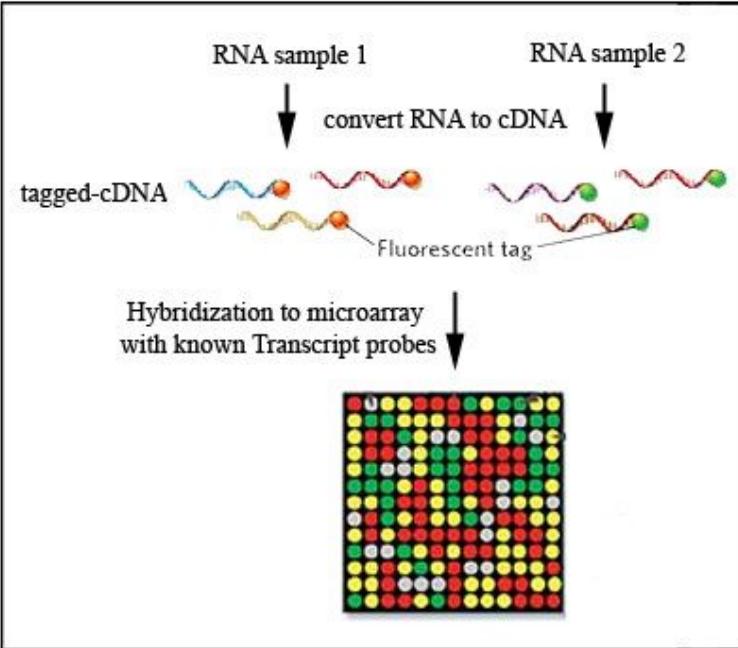
Subtypes	Molecular Signatures	Characteristics	Treatment options ^a
Luminal A	ER+, PR±, HER2-, Low Ki67	~70%, Most common Best prognosis	Hormonal Therapy Targeted Therapy
Luminal B	ER+, PR±, HER2±, High Ki67	10%–20% Lower survival than Luminal A	Hormonal Therapy Targeted Therapy
HER2	ER-, PR-, HER2+	5%–15%	Targeted Therapy
Triple Negative	ER-, PR-, HER2-	15%–20% More common in black women Diagnosed at younger age Worst prognosis	Limited Targeted Therapy
Normal-like	ER+, PR±, HER2-, Low Ki67	Rare Low proliferation gene cluster expression	Hormonal Therapy Targeted Therapy

^a Besides conventional surgical and non-surgical treatment.



High-Throughput Technological Advances

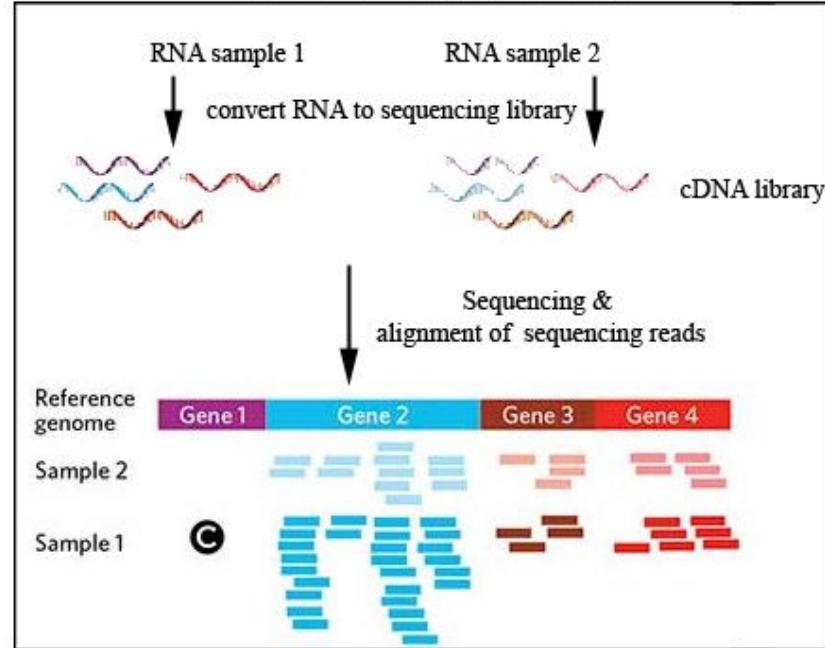
Microarray



relative intensity
=
expression levels

Low sensitivity
Low dynamic range
known transcript only
No alternative splicing information
lower cost

RNA Sequencing (RNA-Seq)

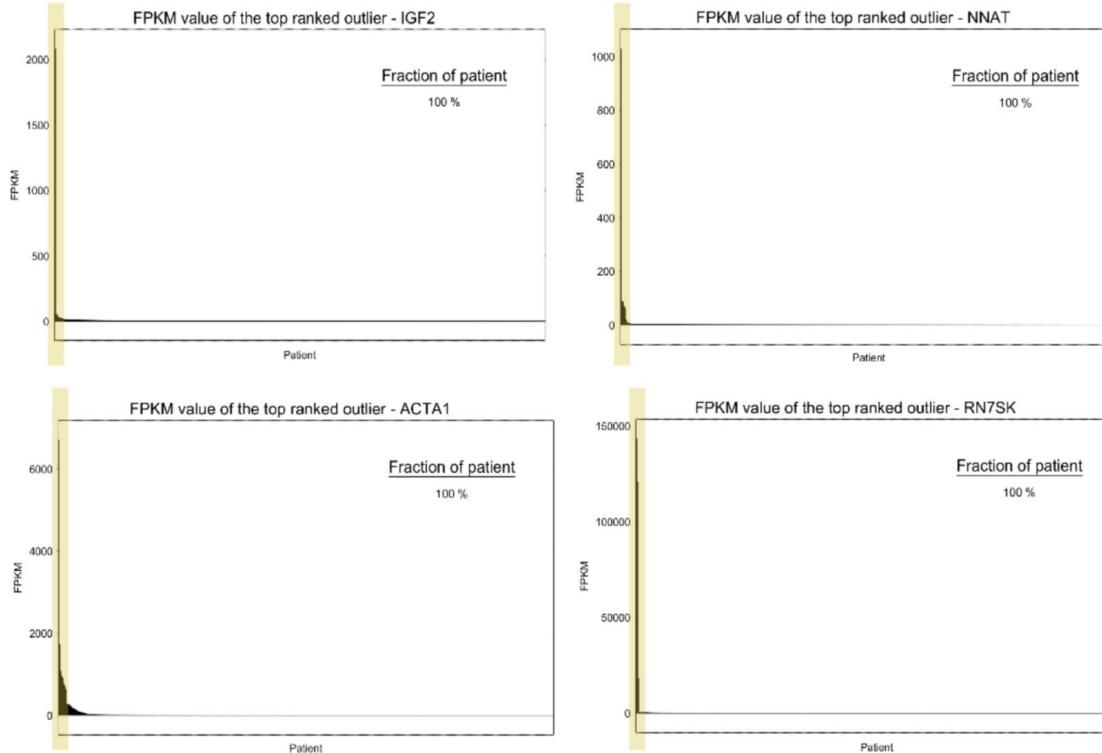


High sensitivity
High dynamic range
Novel transcripts sequences identified
structural variation & alternative splicing revealed
unlimited sample comparisons

Sequencing Reads
=
expression levels

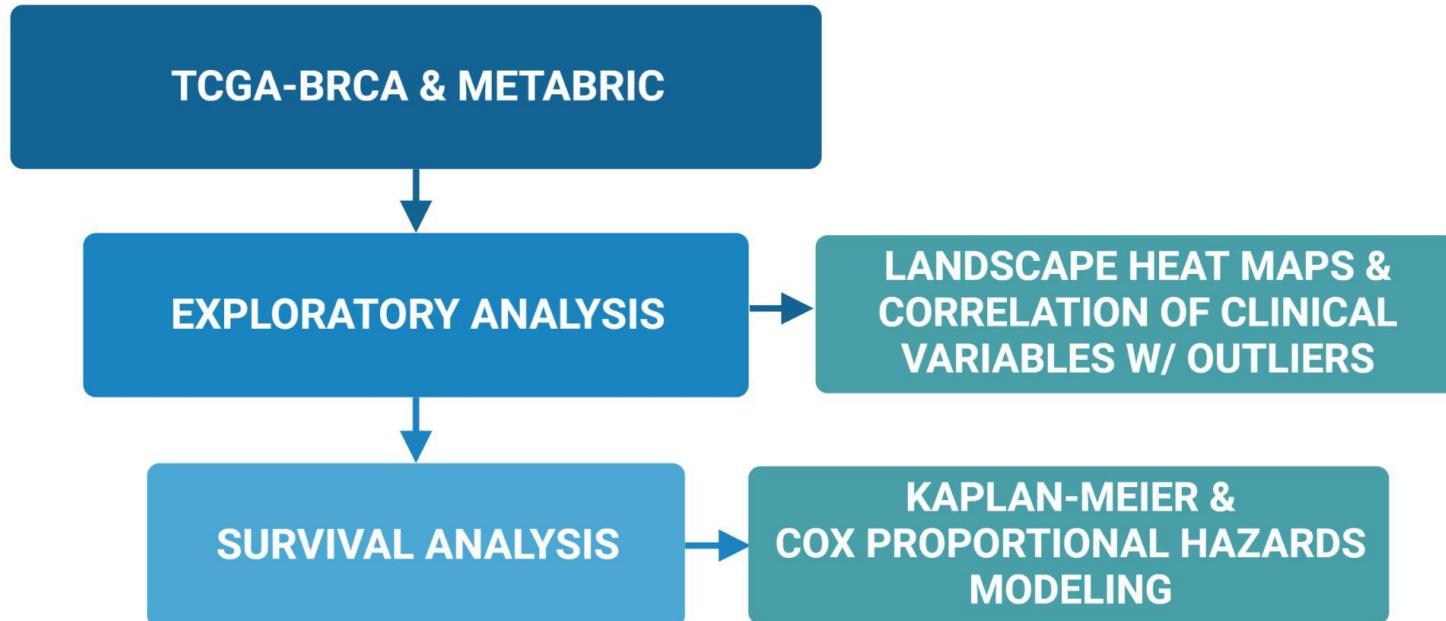
What are gene expression outliers?

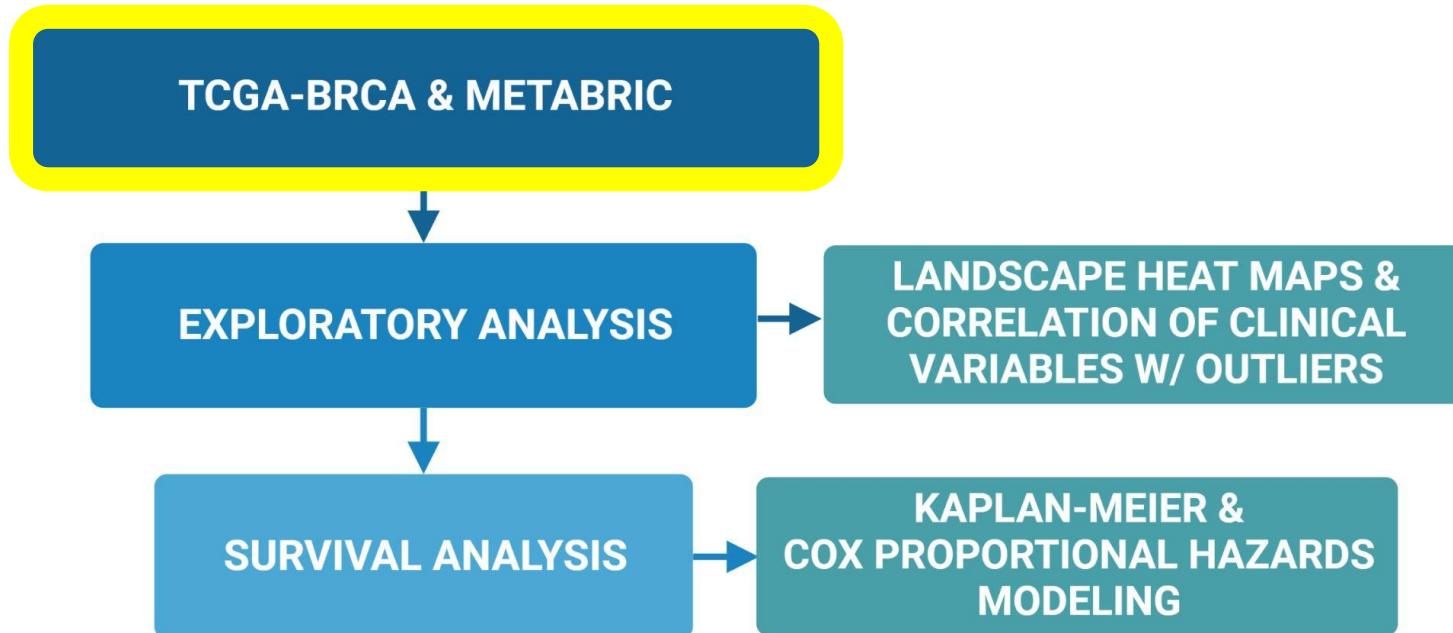
TCGA Prostate Cancer (498 samples)



How do transcriptional outliers contribute to breast cancer survival outcomes, and are they viable prognostic biomarkers for routine clinical practice?

Methods





Dataset Descriptions / Dimensions

Dataset	Project name	Patient set	Study goal	Study design
TCGA-BRCA	The Cancer Genome Atlas breast invasive carcinoma	Various subtypes/treatment/surgery/stage status	The genomic and molecular diversity of breast cancer	Multiple platforms: whole-exome sequencing, RNA sequencing, microarray, and methylation array
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium	Various subtypes/treatment/surgery/stage status	An integrated genomic/transcriptomic analysis of breast cancers with long-term clinical outcomes	Targeted sequencing and gene expression microarray

Dataset	Data type	Sample size	# of transcripts	
TCGA-BRCA	RNA-seq	1085	17820	TCGA-BRCA Outliers: 1187 METABRIC Outliers: 4782
METABRIC	Microarray	1991	19820	METABRIC: Microarray Probe Intensity Values

TCGA-BRCA: Normalized RNA-Seq (FPKM)

Clinical Variables Present

TCGA-BRCA vs. METABRIC

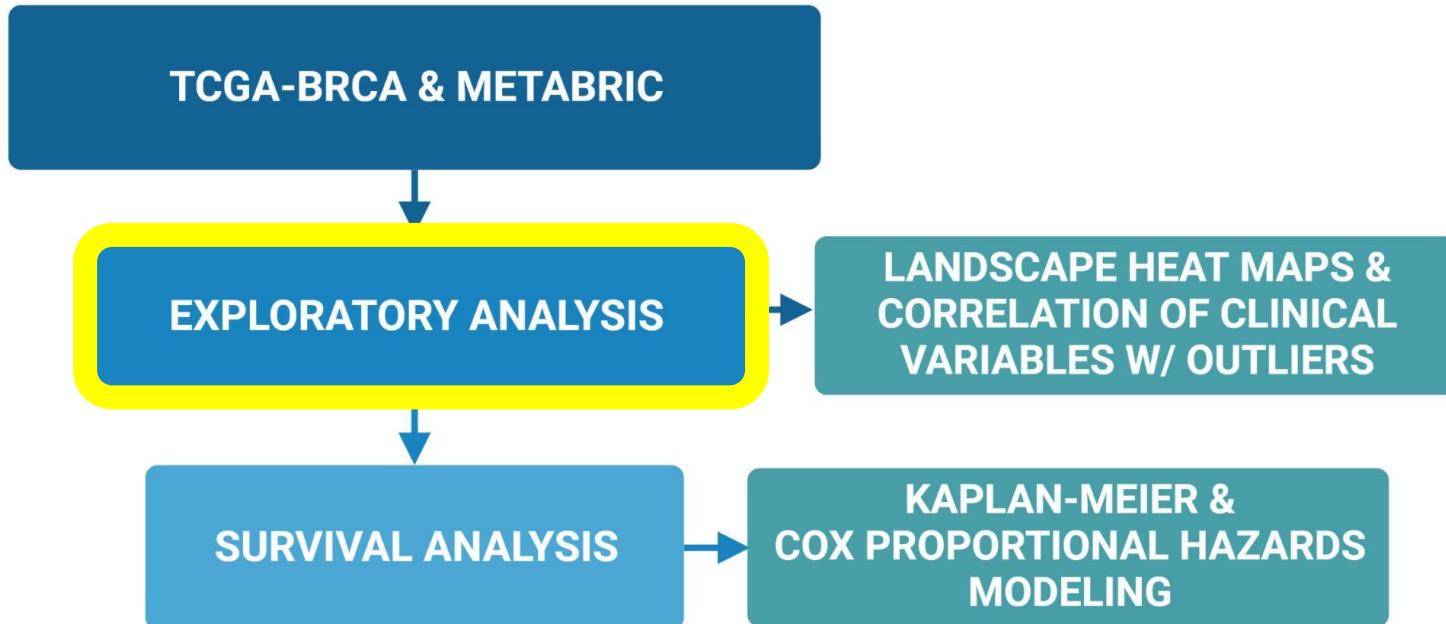
- Few overlapping clinical variables are measured in each dataset
- One similar treatment was recorded for both datasets, Radio Therapy
- TCGA-BRCA has numerous clinical variables that are utilizing multiple methods for measuring.

TCGA-BRCA

```
> colnames(brca_clinic.order)
[1] "Patient.ID"
[2] "Sample.ID"
[3] "Diagnosis.Age"
[4] "Neoplasm.Stage.American.Joint.Committee.on.Cancer.Code"
[5] "American.Joint.Committee.on.Cancer.Publication.Version.Type"
[6] "Aneuploidy.Score"
[7] "Buffa.Hypoxia.Score"
[8] "Cancer.type"
[9] "TCGA.PanCanAtlas.Cancer.Type.Acronym"
[10] "Cancer.Type.Detailed"
[11] "Last.Communication.Contact.from.Initial.Pathologic.Diagnosis.Date"
[12] "Birth.from.Initial.Pathologic.Diagnosis.Date"
[13] "Last.Alive.Less.Initial.Pathologic.Diagnosis.Date.Calculated.Day.Value"
[14] "Disease.Free.Months."
[15] "Disease.Free.Status"
[16] "Months.of.disease.specific.survival"
[17] "Disease.specific.Survival.status"
[18] "Ethnicity.Category"
[19] "Form.completion.date"
[20] "Fraction.Genome.Altered"
[21] "Neoplasm.Histologic.Grade"
[22] "Neoadjuvant.Therapy.Type.Administered.Prior.To.Resection.Text"
[23] "ICD.10.Classification"
[24] "International.Classification.of.Diseases.for.Oncology..Third.Edition.ICD.O.3.Histology.Code"
[25] "International.Classification.of.Diseases.for.Oncology..Third.Edition.ICD.O.3.Site.Code"
[26] "Informed.consent.verified"
[27] "In_PanCan_Pathway_Analysis"
[28] "MSI.MANTIS.Score"
[29] "MSIsensor.Score"
[30] "mutation.count"
[31] "New.Neoplasm.Event.Post.Initial.Therapy.Indicator"
[32] "Oncotype.Code"
[33] "Overall.Survival..Months."
[34] "Overall.Survival.Status"
[35] "Other.Patient.ID"
[36] "American.Joint.Committee.on.Cancer.Metastasis.Stage.Code"
[37] "Neoplasm.Disease.Lymph.Node.Stage.American.Joint.Committee.on.Cancer.Code"
[38] "American.Joint.Committee.on.Cancer.Tumor.Stage.Code"
[39] "Person.Neoplasm.Cancer.Status"
[40] "Progress.Free.Survival..Months."
[41] "Progression.Free.Status"
[42] "Primary.Lymph.Node.Presentation.Assessment"
[43] "Prior.Diagnosis"
[44] "Race.Category"
[45] "Radiation.Therapy"
[46] "Ragnum.hypoxia.Score"
[47] "Number.or.Samples.per.Patient"
[48] "Sample.Type"
[49] "Sex"
[50] "Somatic.Status"
[51] "Subtype"
[52] "Tissue.Retrospective.Collection.Indicator"
[53] "Tissue.Retrospective.Collection.Indicator"
[54] "Tissue.Source.Site"
[55] "Tissue.Source.Site.Code"
[56] "TMB..nonsynonymous.."
[57] "Tumor.Disease.Anatomic.Site"
[58] "Tumor.Type"
[59] "Patient.Weight"
[60] "Winter.Hypoxia.Score"
```

METABRIC

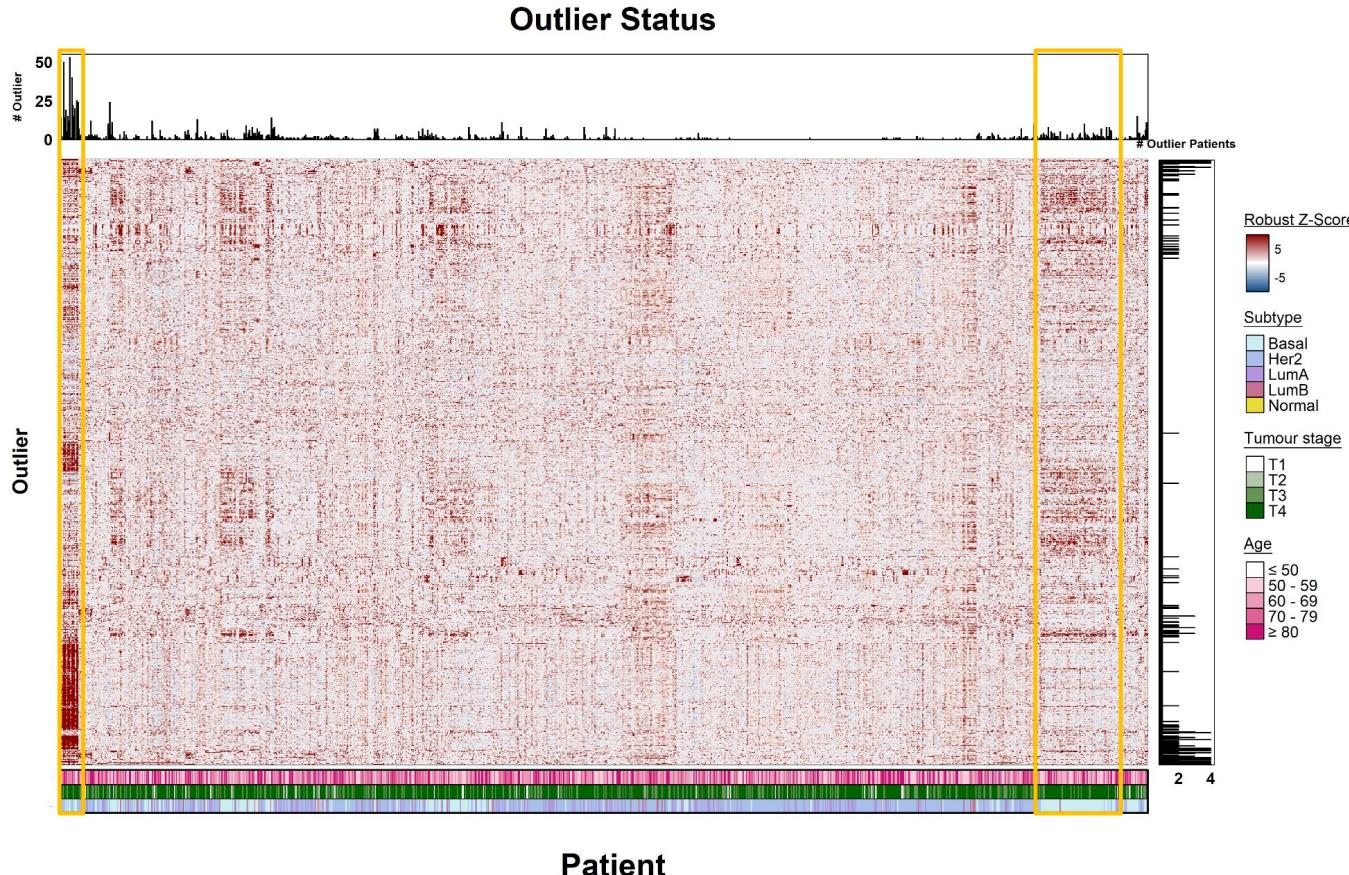
```
> colnames(outlier.meta.clinic)
[1] "Patient.ID"
[2] "outlier.Subgroups"
[3] "Lymph.nodes.examined.positive"
[4] "Nottingham.prognostic.index"
[5] "cellularity"
[6] "Chemotherapy"
[7] "Cohort"
[8] "ER.status.measured.by.IHC"
[9] "HER2.status.measured.by.SNP6"
[10] "Hormone.Therapy"
[11] "Inferred.Menopausal.State"
[12] "Sex"
[13] "Integrative.Cluster"
[14] "Age.at.Diagnosis"
[15] "overall.survival..Months."
[16] "overall.survival.status"
[17] "Pam50...Claudin_low.Subtype"
[18] "x3.Gene.Classifier.subtype"
[19] "Patient.s.vital.Status"
[20] "Primary.Tumor.Laterality"
[21] "Radiation.Therapy"
[22] "Tumor.Other.Histologic.subtype"
[23] "Type.of.Breast.Surgery"
[24] "Relapse.Free.Status"
[25] "Relapse.Free.Status..Months."
[26] "outlier.class.x"
[27] "outlier.totals.x"
[28] "Cancer.Type"
[29] "Cancer.Type.Detailed"
[30] "ER.Status"
[31] "HER2.Status"
[32] "Neoplasm.Histologic.Grade"
[33] "Oncotree.Code"
[34] "PR.Status"
[35] "Sample.Type"
[36] "Tumor.Size"
[37] "Tumor.Stage"
[38] "TMB..nonsynonymous.."
[39] "outlier.class.y"
[40] "outlier.totals.y"
```



TCGA-BRCA RNA Abundance Landscapes

UCLA Health

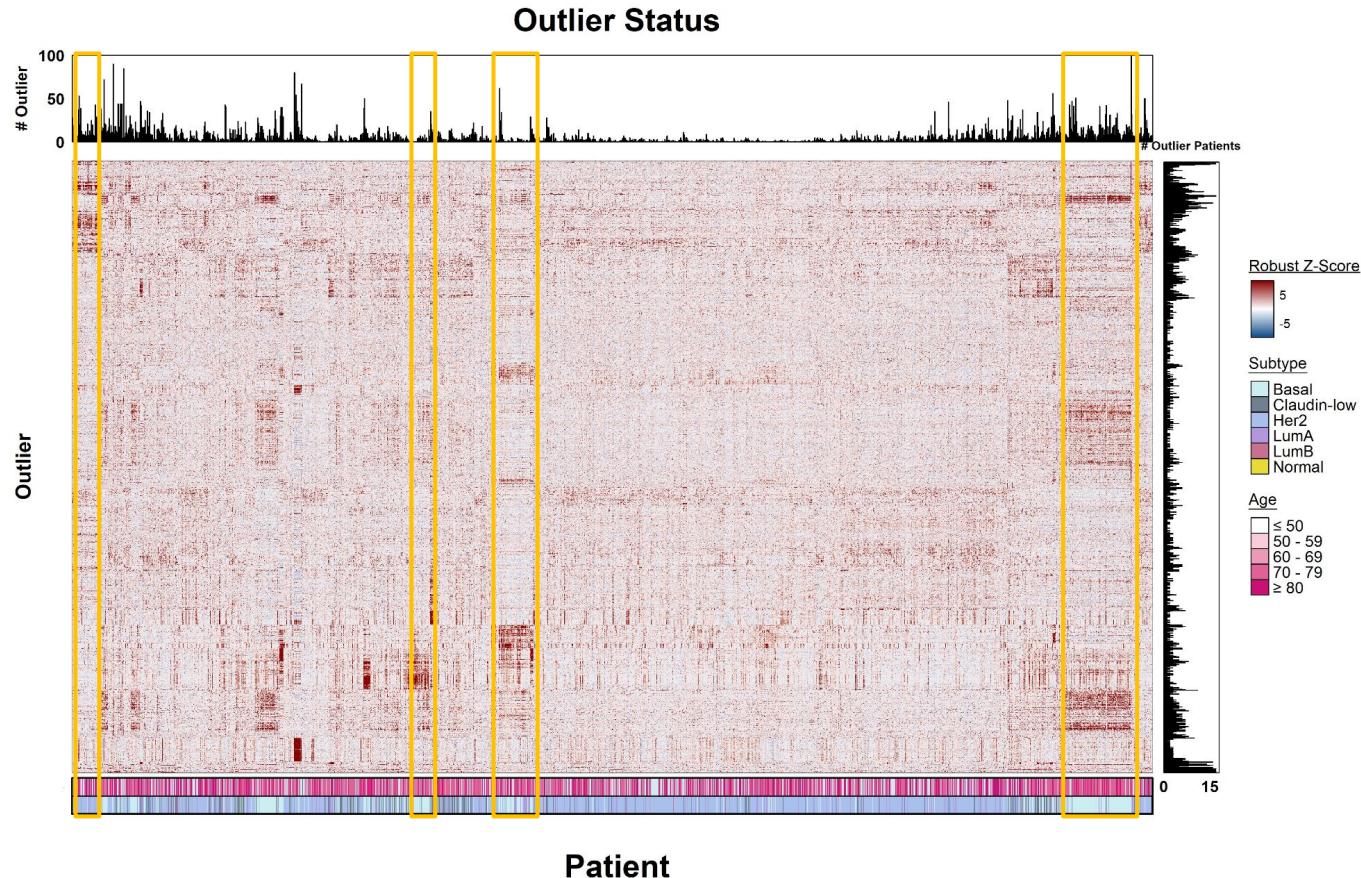
Jonsson Comprehensive
Cancer Center



METABRIC RNA Abundance Landscapes

UCLA Health

Jonsson Comprehensive
Cancer Center



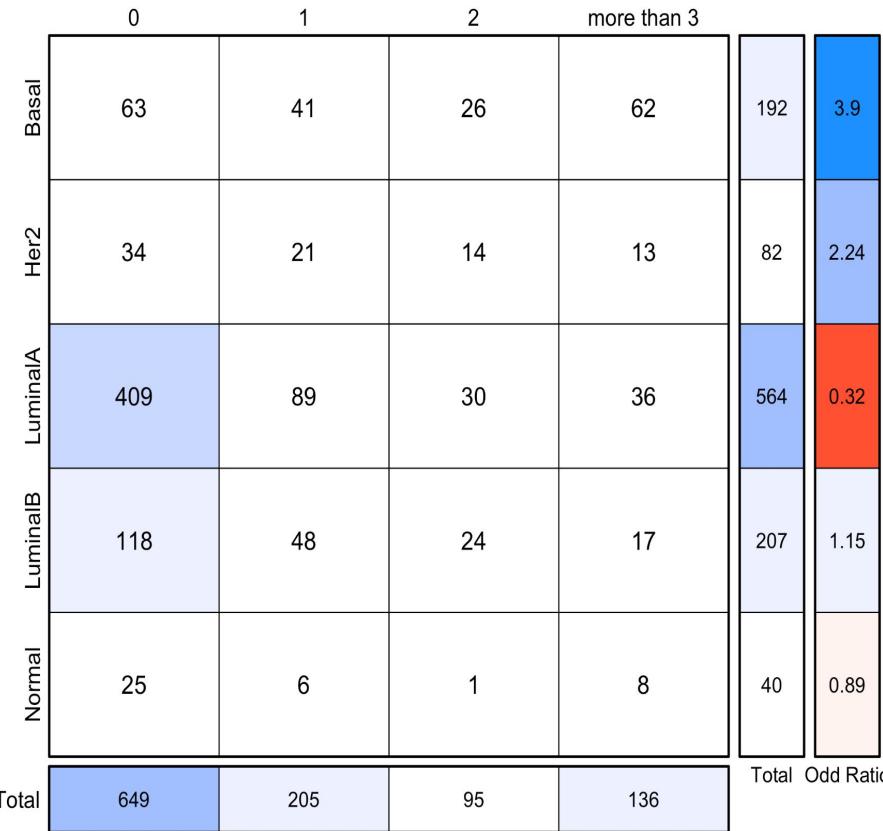
Subtype Classification by Outlier Gene Count



Jonsson Comprehensive
Cancer Center

TCGA-BRCA

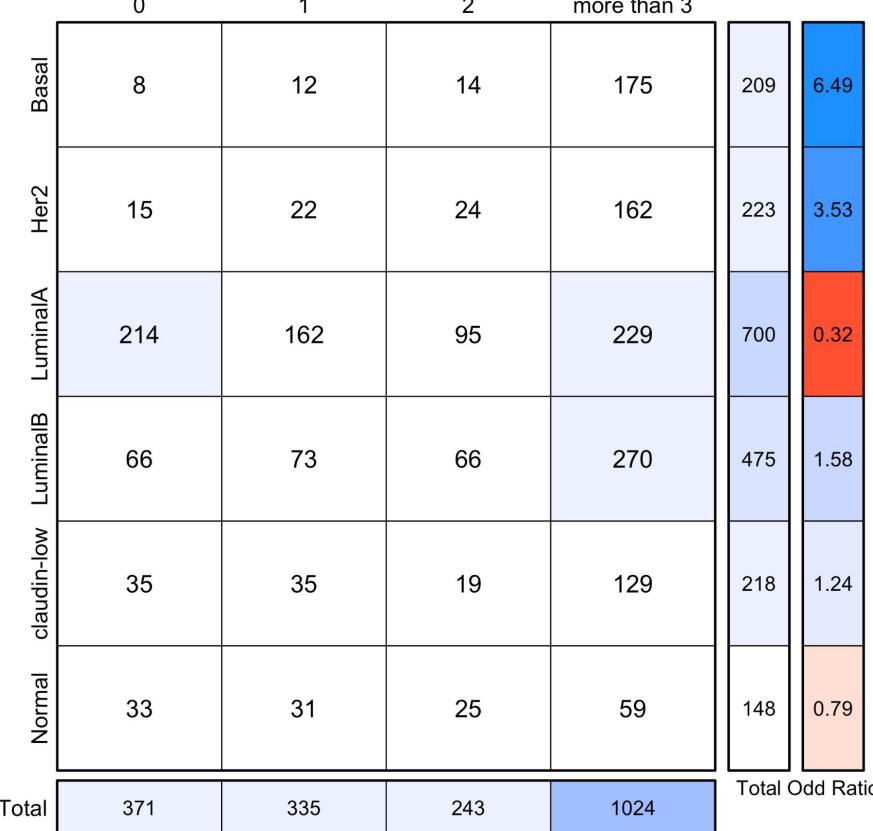
Number of outliers per patient



$p = 2.62 \times 10^{-27}$

METABRIC

Number of outliers per patient



$p = 4.02 \times 10^{-49}$

Outlier Genes and Clinical Variables

HER2 Status by Outlier Gene Count

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

¹ Pearson's Chi-squared test

ER Status by Outlier Gene Count

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

¹ Pearson's Chi-squared test

PR Status by Outlier Gene Count

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

¹ Pearson's Chi-squared test

(Frequency of Associated
Status)

HER2+

ER-

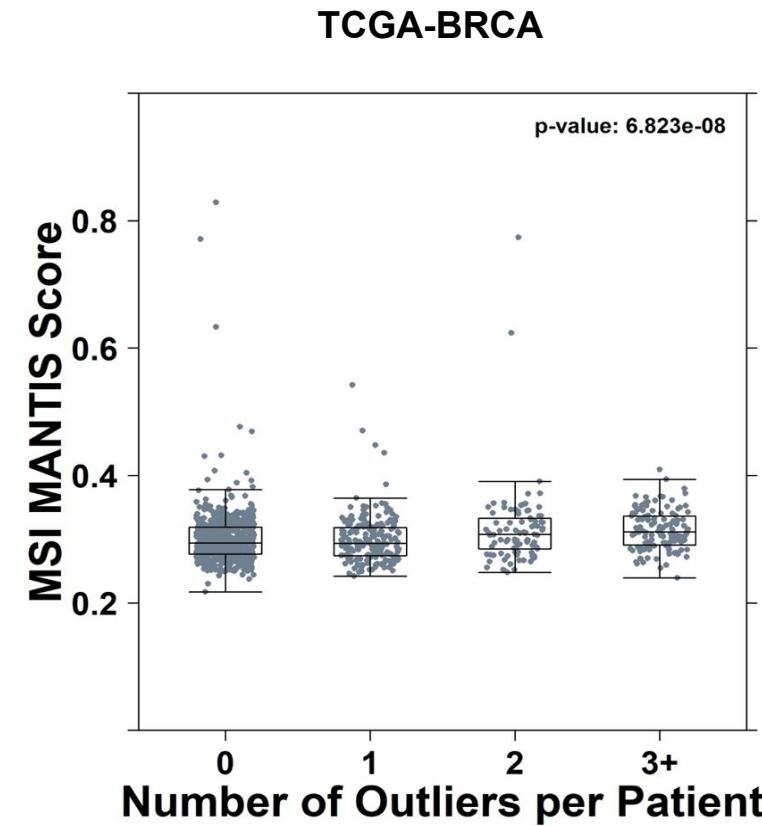
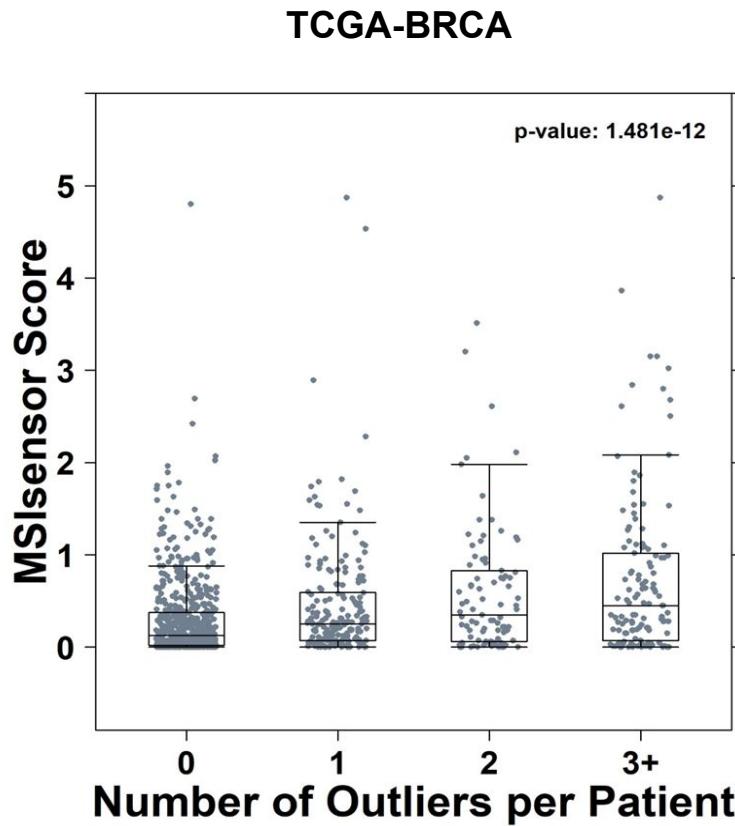
PR-

Outlier Genes

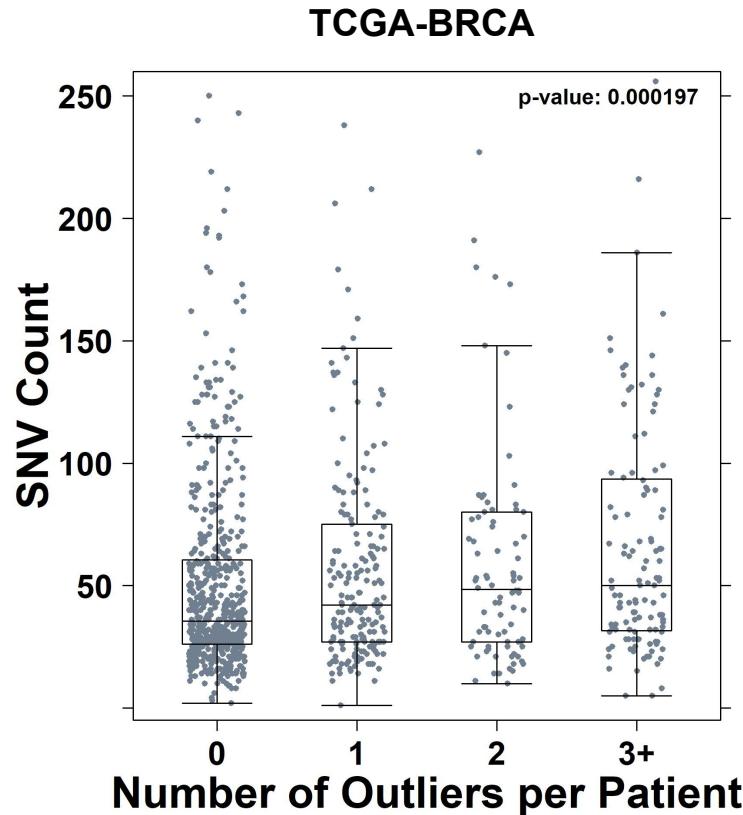
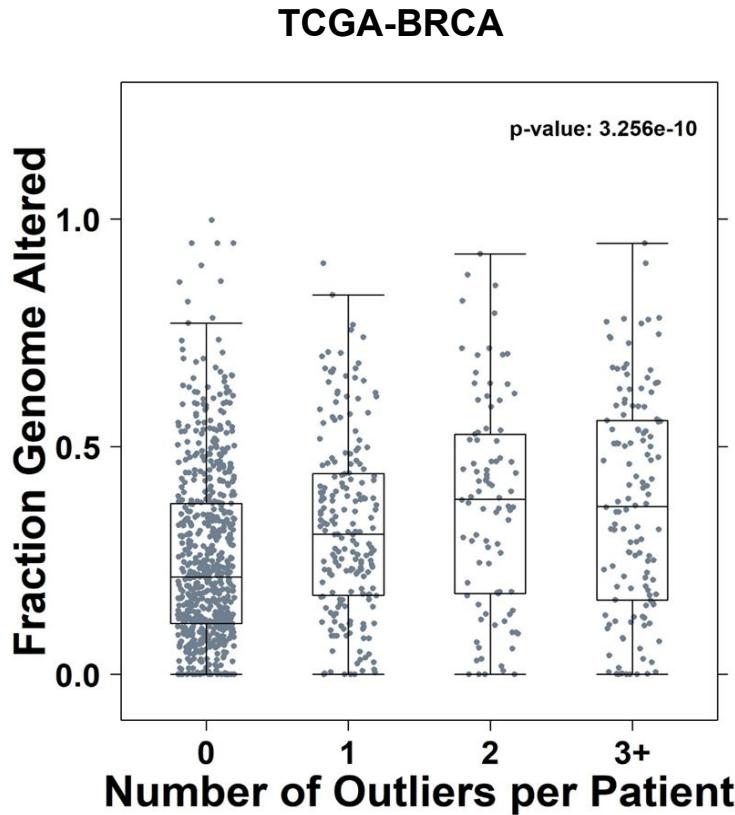
Outlier Genes and Clinical Variables

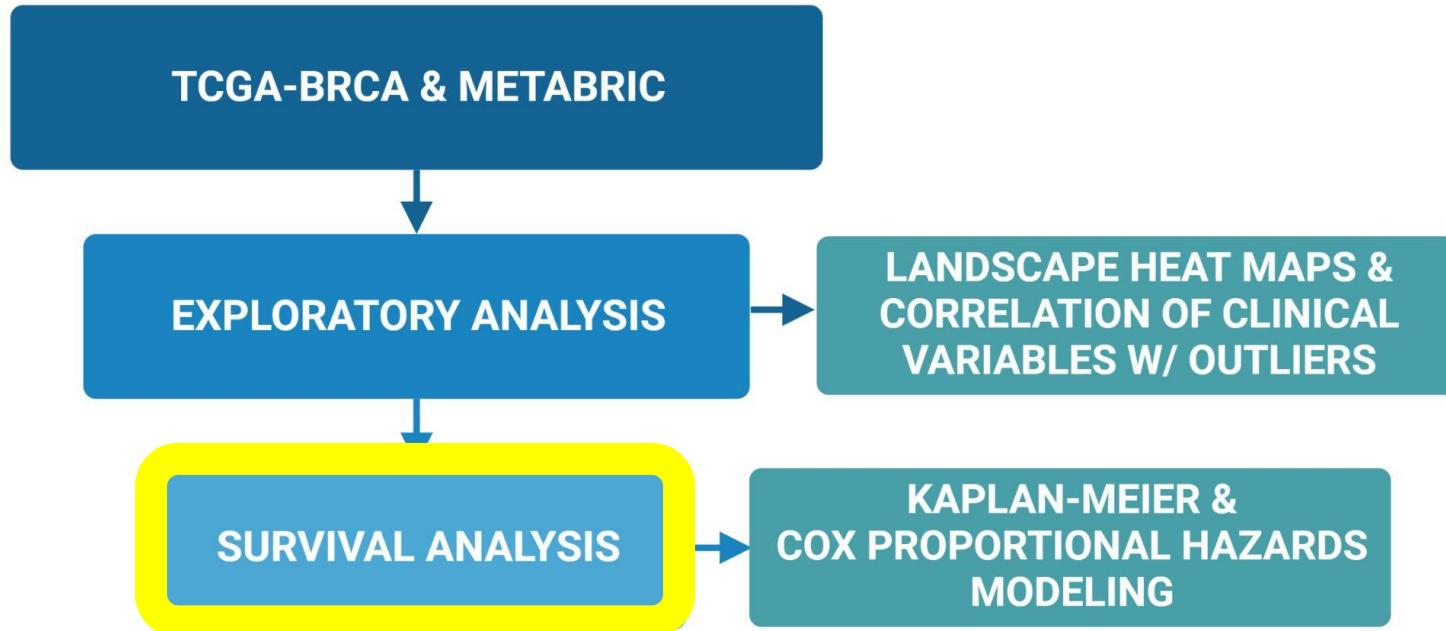
UCLA Health

Jonsson Comprehensive
Cancer Center



Outlier Genes and Clinical Variables





TCGA-BRCA

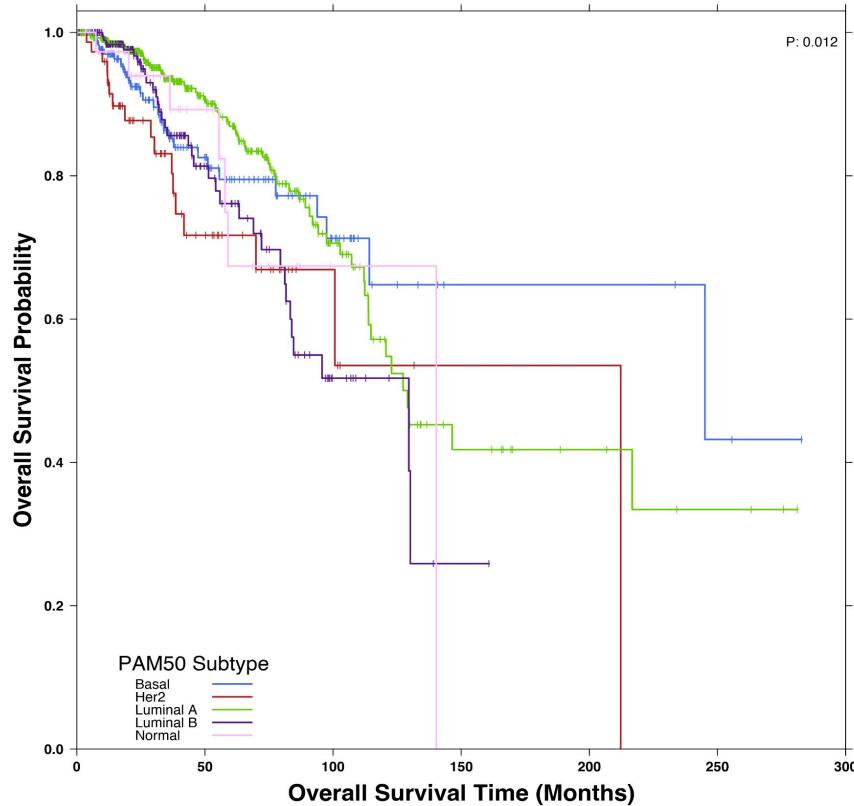
Survival Statistic	Group ID	Death/All Events	Median Survival Time (95% CI)
Overall Survival		151/1071(8.88%)	27.008(24.953,30.082)
Outlier Groups	0	92/640 (14.38%)	29.917 (26.038,32.909)
	1	28/203 (13.79%)	25.052 (22.816,31.726)
	2	15/93 (16.13%)	21.271 (17.687,33.304)
	3+	16/135 (11.85%)	23.835(20.383,29.819)
Subtype Groups	Basal	28/191 (14.66%)	26.926(21.172,33.041)
	Her2	17/82 (20.73%)	21.353(16.504,33.205)
LumA	66/557 (11.85%)	28.865(25.742,32.449)	
LumB	33/201 (16.42%)	23.934(20.876,29.950)	
Normal	7/40 (17.5%)	35.424(21.304,53.128)	

METABRIC

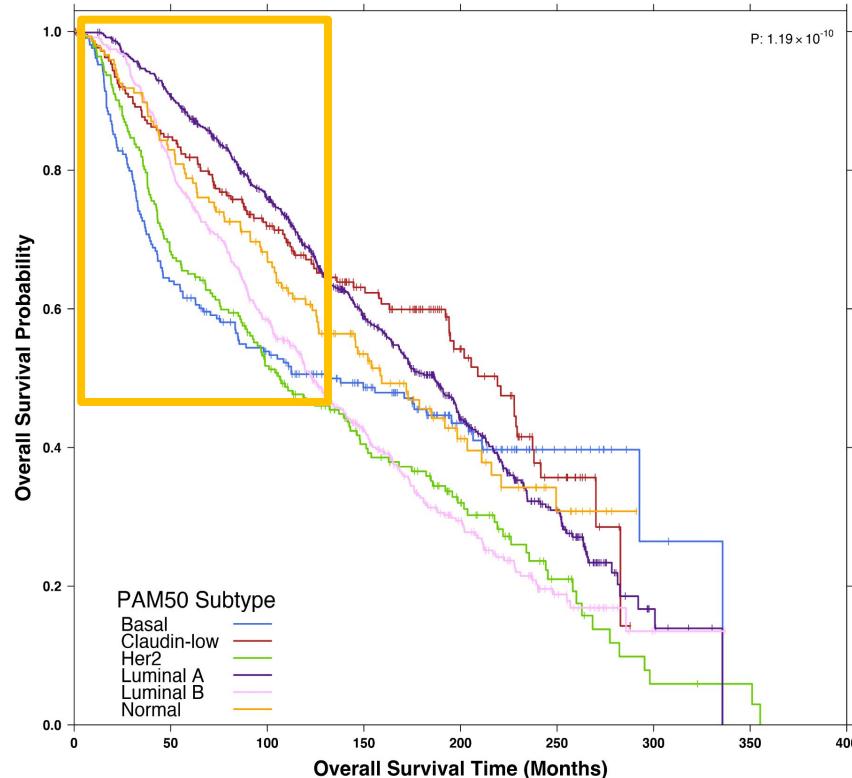
Survival Stat	Group ID	Death/All Events	Median Survival Time (CI 95%)
Overall Survival		1144/1980 (57.78%)	116.5(111.833,121.967)
Outlier Groups	0	192/373 (51.47%)	132.067(125.833,144.767)
	1	202/336 (60.12%)	121.383(107.267,134.267)
	2	142/243 (58.44%)	105(90.133,114.233)
	3+	608/1028 (59.14%)	110.8(103.833,117.6)
Subtype Groups	Basal	115/209 (55.02%)	85.5(68.7,112.933)
	Her2	157/223 (70.4%)	96.833(79.967,107.1)
LumA	378/700 (54%)	130.083(124.267,140.067)	
LumB	315/475 (66.18%)	104.1(95.867,116.633)	
Normal	79/148 (53.38%)	121.067(105.967,134.5)	

Kaplan-Meier Curves by PAM50 Subtype

Overall Survival by PAM50 Subtype: TCGA Patients



Overall Survival by PAM50 Subtype: METABRIC Patients



Basal 190 59 21 4 4 2 0

Her2 82 22 5 1 1 0 0

Luminal A 557 166 48 12 6 3 0

Luminal B 211 51 10 1 0 0 0

Normal 40 15 3 0 0 0 0

Basal 209 132 100 70 36 12 2 0 0 0

Claudin-low 218 174 126 80 44 16 0 0 0 0

Her2 224 149 106 63 38 15 3 2 0 0

Luminal A 700 621 465 295 159 70 6 0 0 0

Luminal B 475 372 251 145 73 23 2 0 0 0

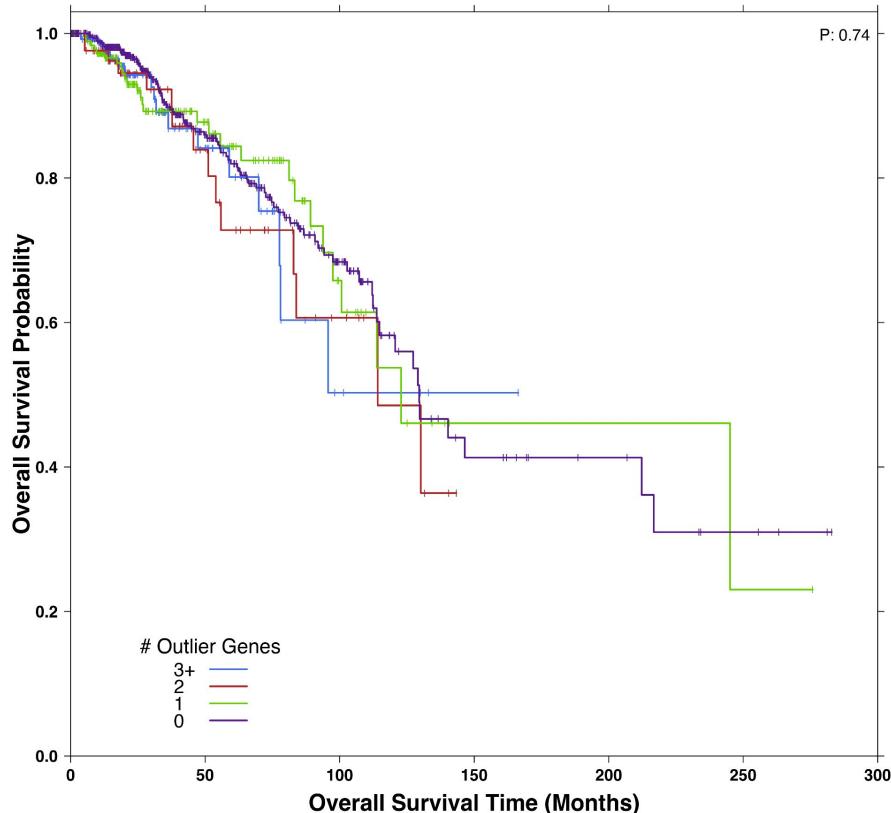
Normal 148 121 91 53 26 9 0 0 0 0

Kaplan-Meier Curves by Outlier Gene Count

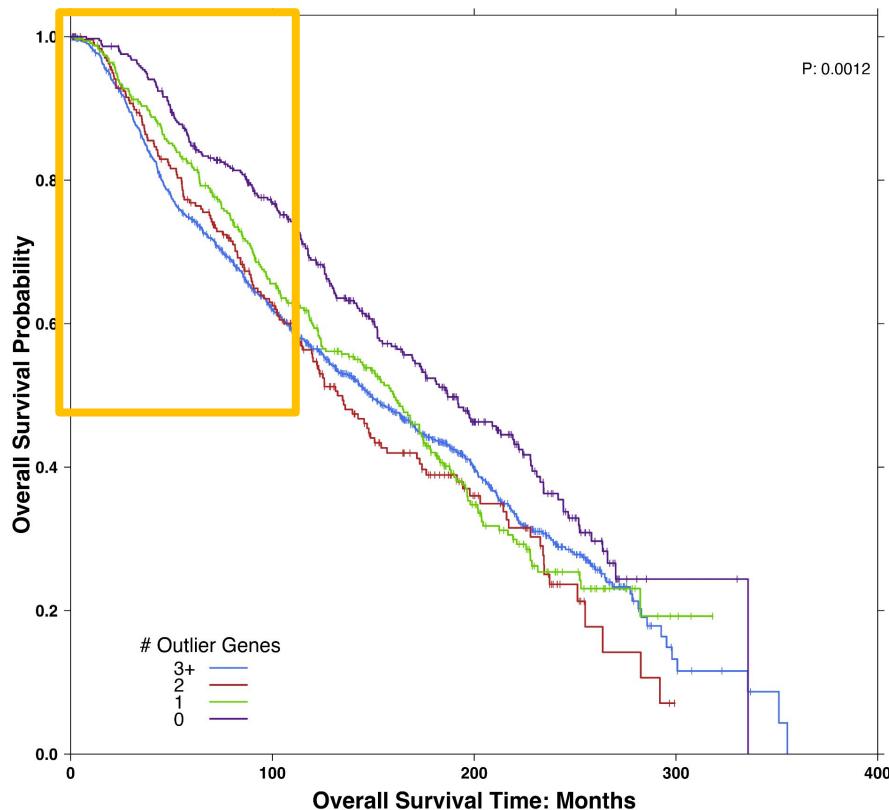


Jonsson Comprehensive
Cancer Center

Overall Survival by Outlier Gene Count: TCGA Patients



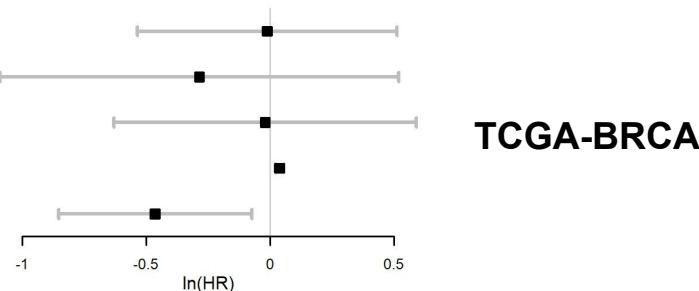
Overall Survival by Outlier Gene Count: METABRIC Patients



Multivariate Survival Modeling

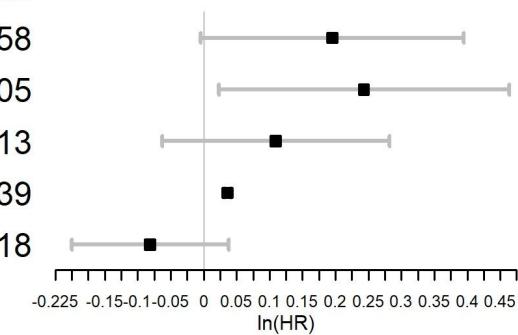
MODEL: `coxph(Surv(time,event) ~ Outlier.groups + strata(subtype)+ age + Radiation.Therapy)`

Features	In(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



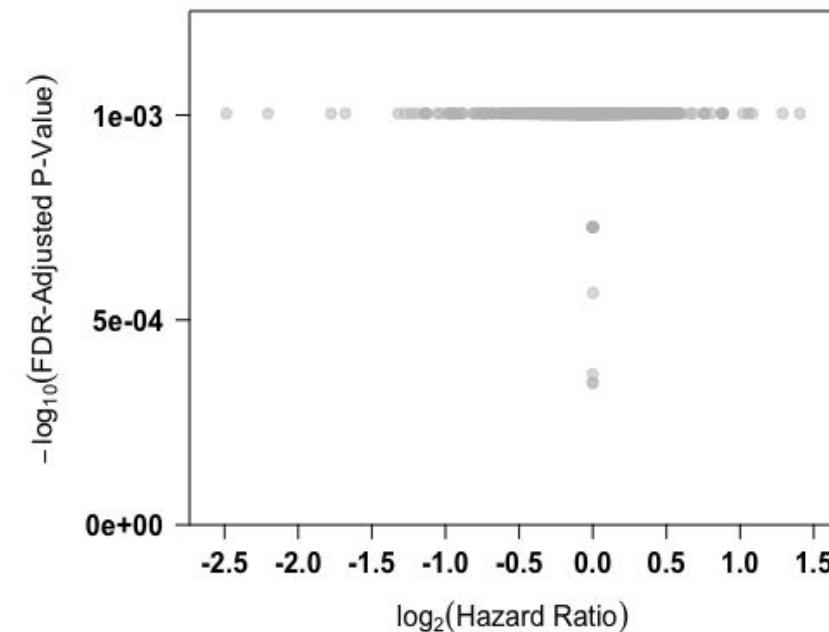
Features	In(HR)	P value
Outlier.Subgroups1	0.195	0.0558
Outlier.Subgroups2	0.243	0.0305
Outlier.Subgroups3+	0.109	0.213
Age.at.Diagnosis	0.036	4.07e-39
Radiation.TherapyYES	-0.0814	0.18

METABRIC

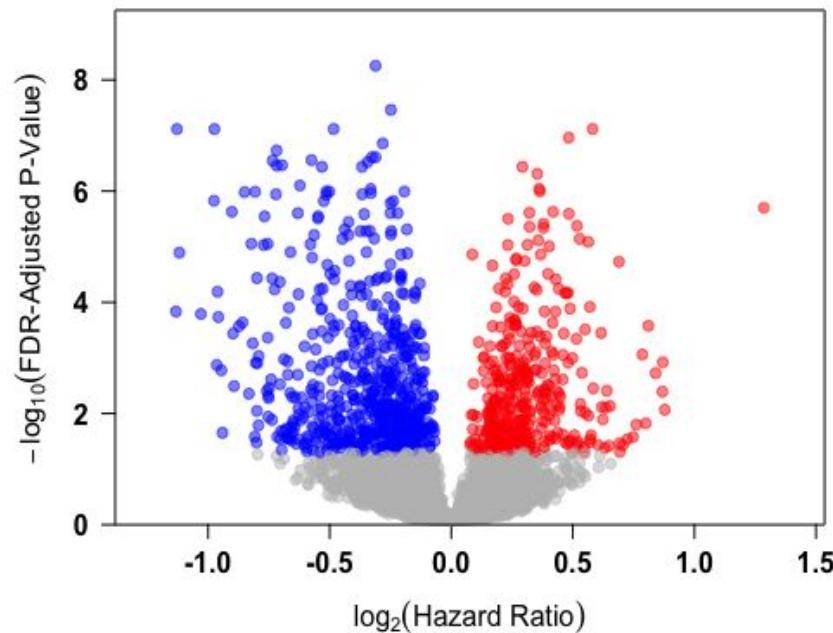


Gene-Wise Outlier Contributions

TCGA: Univariate Cox Models



METABRIC: Univariate Cox Models

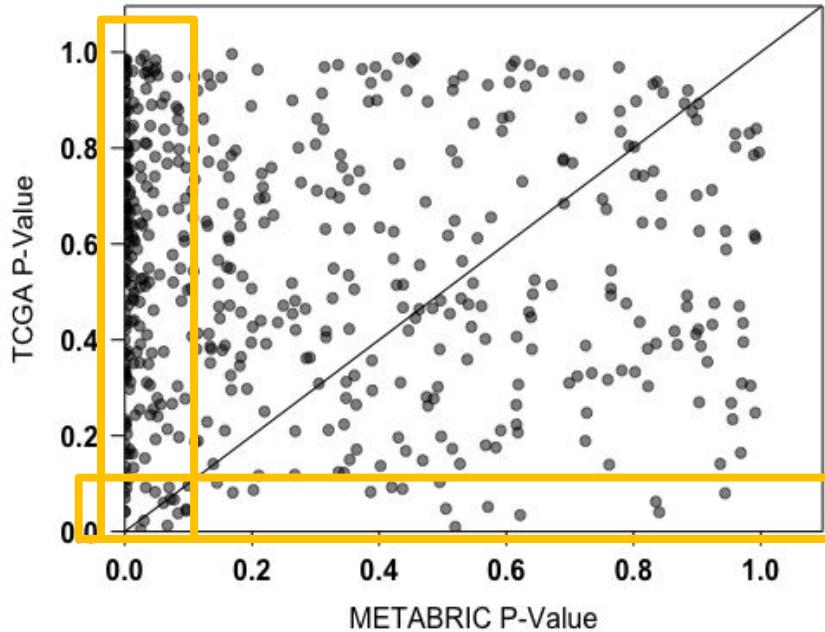


Outlier Contributions: Dataset Comparison

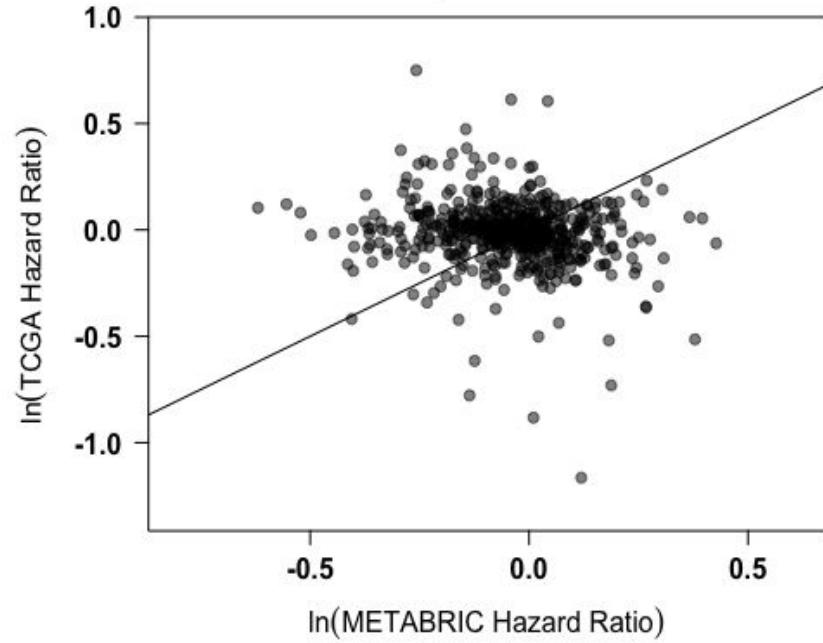
UCLA Health

Jonsson Comprehensive
Cancer Center

METABRIC-TCGA Comparison: P-Values



METABRIC-TCGA Comparison: Hazard Ratios



Conclusions / Future Directions



UCLA Health

Jonsson Comprehensive
Cancer Center

- ER and PR status became increasingly negative with increasing outlier gene count.
- Genomic instability seems to introduce the capability for outlier gene expression in cancers.
- 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 TCGA-BRCA covariate-adjusted Cox models, while diagnosis age was significantly associated in METABRIC.
- 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.

References

- 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.
- 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.
- Feng, Yixiao & Spezia, Mia & Huang, Shifeng & Yuan, Chengfu & Zeng, Zongyue & Zhang, Linghuan & Ji, Xiaojuan & Liu, Wei & Huang, Bo & Luo, Wenping & Liu, Bo & Lei, Yan & Du, Scott & Vuppalapati, Akhila & Luu, Hue & Haydon, Rex & He, Tong-Chuan & Ren, Guosheng. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & Diseases.* 5. 10.1016/j.gendis.2018.05.001.
- Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget.* 2017 Jan 31;8(5):7452-7463. doi: 10.18632/oncotarget.13918. PMID: 27980218; PMCID: PMC5352334.
- Ross, Jeffrey & Hatzis, Christos & Symmans, W & Pusztai, Lajos & Hortobágyi, Gabriel. (2008). Commercialized Multigene Predictors of Clinical Outcome for Breast Cancer. *The oncologist.* 13. 477-93. 10.1634/theoncologist.2007-0248.
- Segaert P, Lopes MB, Casimiro S, Vinga S, Rousseeuw PJ. Robust identification of target genes and outliers in triple-negative breast cancer data. *Stat Methods Med Res.* 2019;28(10-11):3042-3056. doi:10.1177/0962280218794722
- Shin G, Kang T-W, Yang S, Baek S-J, Jeong Y-S, Kim S-Y. GENT: Gene Expression Database of Normal and Tumor Tissues. *Cancer Informatics.* 2011;10. doi:10.4137/CIN.S7226

Acknowledgements

UCLA Health

Jonsson Comprehensive
Cancer Center

Special Thanks:

- Paul Boutros
- Zhuyu Qiu
- Jee Yun Han

Boutros Lab Members

Aaron Holmes	Jee Yun Han	Raag Agrawal
Alfredo Gonzalez	Jenna Giafaglione	Rupert Hugh-White
Amaan Jogia-Sattar	Jieun (Grace) Oh	
Amanda Lozano	John Sahrmann	Ruilin Wang
Brandon Tsai	Joseph Salmingo	Sarah Hiyari
Caroline Chen	Julie Ramirez	Selina Wu
Chenghao (Trevor)	Karthik	Shupeng Luxu
Zhu	Guruvayurappan	Stefan Eng
Dan Knight	Lydia Liu	Sorel Fitz-Gibbon
Edward Hwang	Mao Tian	Takafumi Yamaguchi
Faizal Eeman	Muriele Varuzza	Tsumugi Gebo
Helena Winata	Nicholas Wang	Wenshu Zhang
Jack Dodson	Nicole Zeltser	Yash Patel
James Evans	Noe Reyna	
Jaron Arbet	Paul Boutros	Zhuyu Qiu