

Why single cell studies are needed to solve reproducibility issues in clinical gene expression studies of bulk tumors

Aedín Culhane, PhD

Twitter @AedinCulhane

aedin@jimmy.harvard.edu

MAQC Meeting 2019



Harvard T.H. Chan
School of Public Health

Develop a classifier that predicts Boston from Seattle?



Hypothesis?

CITY *of* **BOSTON**



Hypothesis?

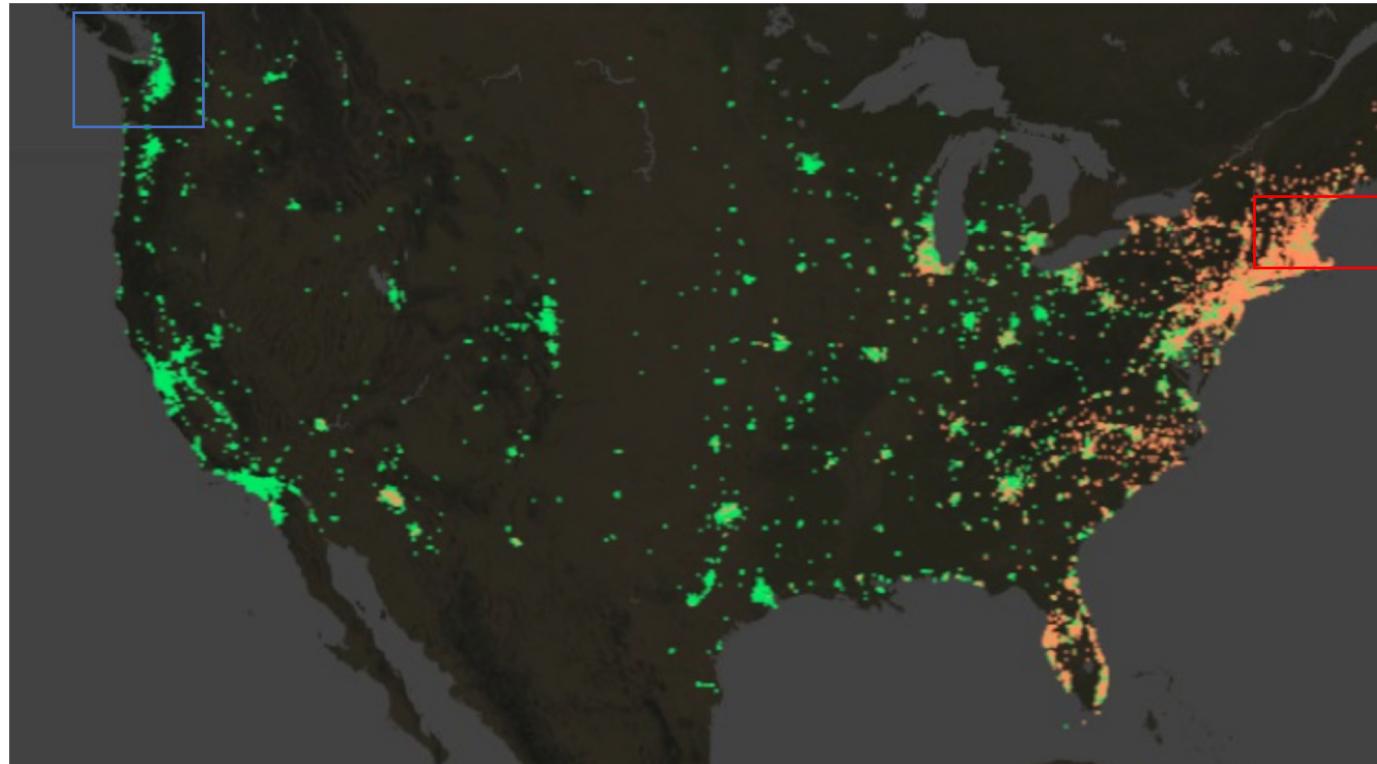


Collect prospective data

- Collected data from a Boston Globe survey 11,100 Starbucks and 7,200 Dunkins drinkers.
- At the time of the survey;
- Dunkins has 10,000 stores in 32 countries and sales of nearly \$9 billion.
- Starbucks has 20,000 stores on six continents with sales > \$13
- Examine those in Seattle and Boston

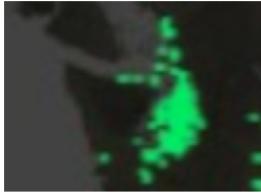
http://archive.boston.com/yourtown/specials/starbucks_vs_dunkin_donuts/

Extract average results for Boston and Seattle



http://archive.boston.com/yourtown/specials/starbucks_vs_dunkin_donuts/

Averages of coffee survey results agree



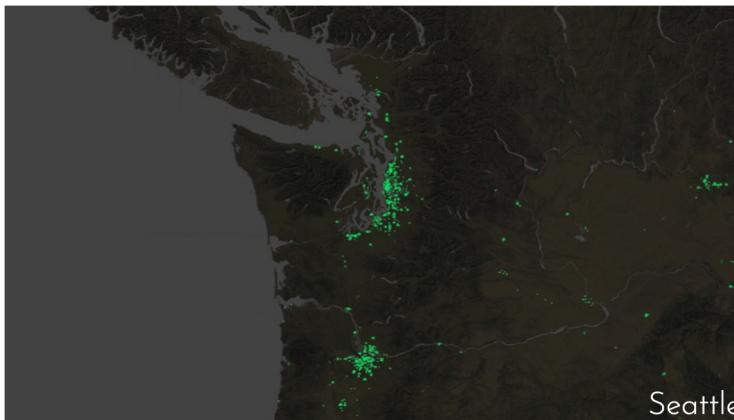
CITY of BOSTON

MY CLASSIFIER
Seattle drinks Starbucks
Boston drinks Dunkin'

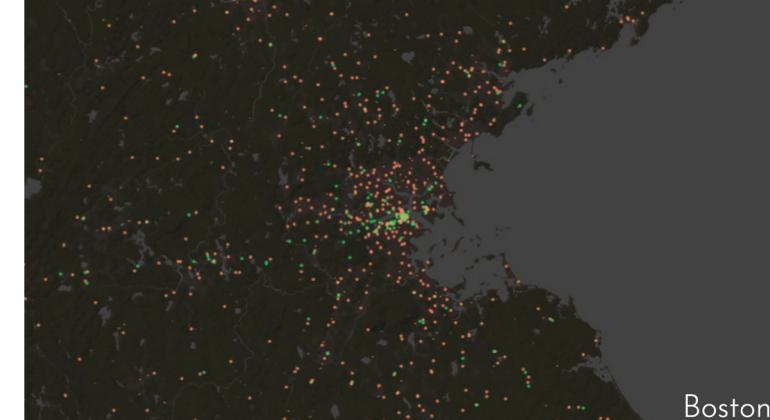
Paper “Coffee is a robust, reliable biomarker of Boston v Seattle”



Paper “Coffee is a robust, reliable biomarker of Boston v Seattle”



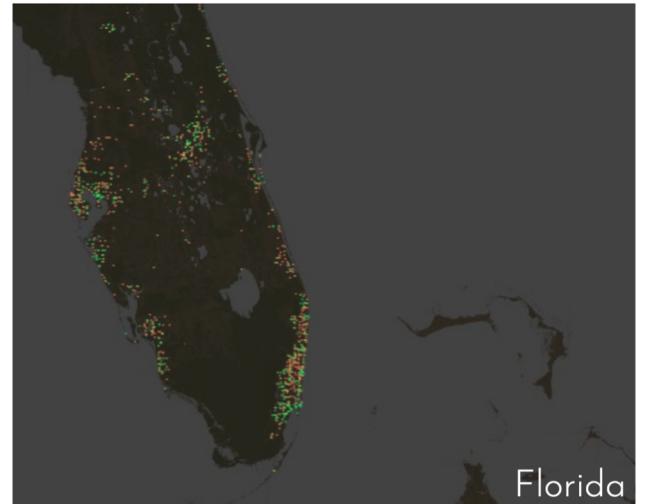
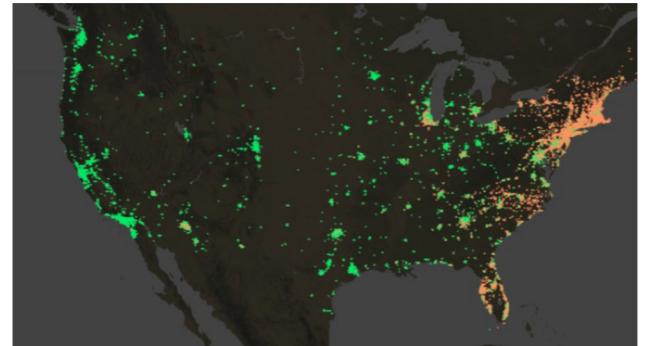
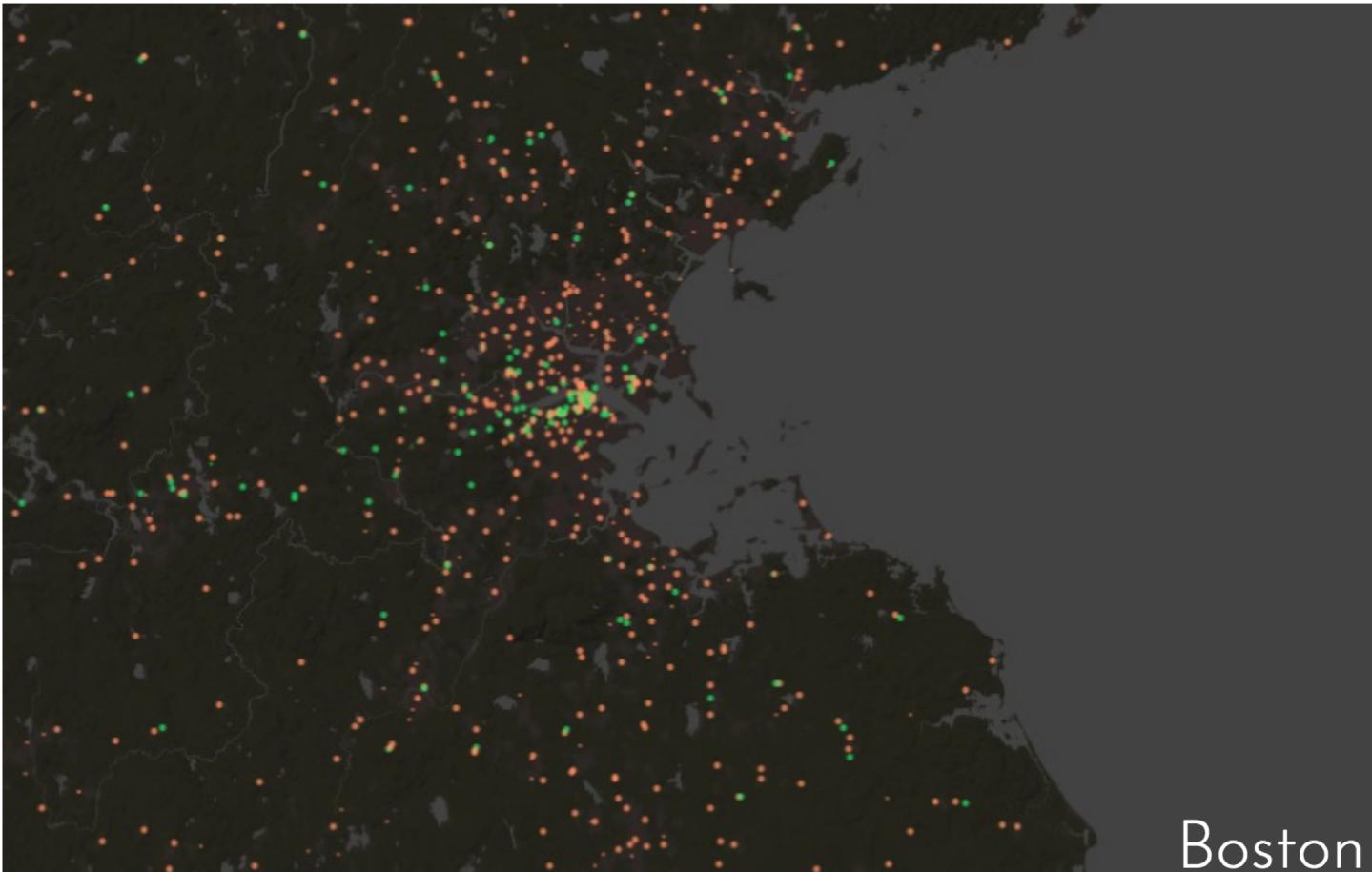
Seattle



Boston

http://archive.boston.com/yourtown/specials/starbucks_vs_dunkin_donuts/

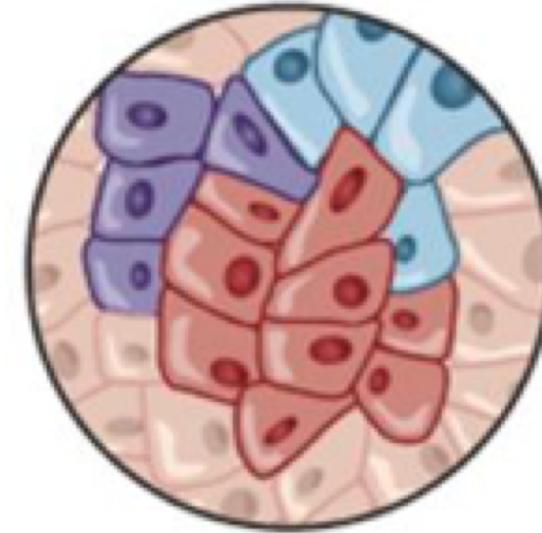
Issues with using an average (or bulk value)



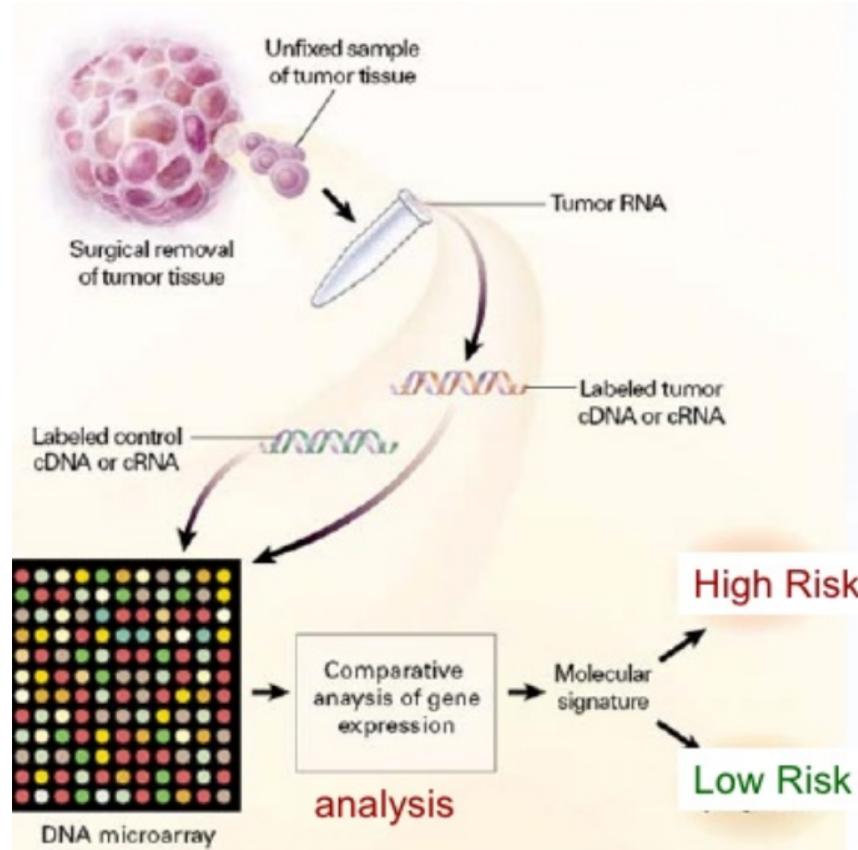
Most Tumor 'Omics Data is a Cell Admixture

- Bulk tumor tissue is an admixture of epithelial tumor cells and infiltrating immune, stromal endothelial, other cells.

Bulk analysis



Dream of Genomics “Bench to Bedside”



Most RNAseq and microarray gene expression of tumors were obtained on **bulk tissue**

Gene signatures
Molecular subtype classifications

And the basis of much of the literature and databases of genomic annotation

Biomarkers, Gene Signatures

- Diagnostic biomarkers that distinguish phenotypically similar medical conditions (often categorical)
 - mild, moderate or severe
 - Low v high grade
 - benign v malignant
 - **Molecular subtypes**
 - Histological subtypes
- Prognostic biomarkers that predict outcome or course of a disease (developed on naïve tissue, without regard to treatment)
- Predictive biomarkers that predict response to a specific treatment (pharmacogenomics)

Gene Set Analysis, Gene Annotation, Pathways

- Many cell type biomarkers, pathways annotation has been extrapolated from non-pure population of cells (tissue)
- These are widely used in gene set analysis



Lack of bench-> bedside translation of gene signatures.

Though many gene-expression-based prognostic signatures have been reported in the literature, very few are used in clinical practice

Many Prognostic gene signatures of ovarian cancer have been described

OPEN ACCESS Freely available online



Angiogenic mRNA and microRNA Gene Expression Signature Predicts a Novel Subtype of Serous Ovarian Cancer

Cancer Research

A Gene Signature Predicting for Survival in Suboptimal Debulked Patients with Ovarian Cancer

Tomas Bonome, Douglas A. Levine, Joanna Shih, et al.

Cancer Cell Article

A Gene Signature Predictive for Outcome in Advanced Ovarian Cancer Identifies a Survival Factor: Microfibril-Associated Glycoprotein 2



The Journal of Molecular Diagnostics

Volume 14, Issue 3, May–June 2012, Pages 214–222



BJC
British Journal of Cancer

Journal home > Archive > Genetics and Genomics > Full text

BJC OPEN



Genetics and Genomics
British Journal of Cancer (2011) 105, 304–311. doi:10.1038/bjc.2011.219
www.bjcancer.com
Published online 7 June 2011

A seven-gene prognostic model for platinum-treated ovarian carcinomas

R Sabater^{1,2}, P Finetti¹, J Bonnesea¹, J Jacquemier^{1,3}, J Adelaidet¹, E Lambaudie⁴, P Viens^{2,5}, D Birbaum¹ and F Bertuccia^{1,2,3}

THE JOURNAL OF Pathology

Original Paper

A prognostic gene expression index in ovarian cancer—validation across different independent data sets[†]



VOLUME 28 • NUMBER 22 • AUGUST 1 2010
JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

OPEN ACCESS Freely available online

PLOS MEDICINE

Regular article

Genes with Bimodal Expression Are Robust Diagnostic Targets that Define Distinct Subtypes of Epithelial Ovarian Cancer with Different Overall Survival

Dawn N. Kermagis, Allison H.S. Hall, Michael B. Datto, et al.

Survival-Related Profile, Pathways, and Transcription Factors in Ovarian Cancer

Anne P. G. Crijns¹, Rudolf S. N. Fehrmann^{1,2,3*}, Steven de Jong², Frans Gerbens², Gert Jan Meersma¹, Harry G. Klip¹, Harry Hollema², Robert M. W. Hofstra³, Gerard J. te Meerman², Elisabeth G. E. de Vries², Ate G. J. van der Zee^{1,2}

OPEN ACCESS Freely available online

PLOS one

Gene Expression Profile for Predicting Survival in Advanced-Stage Serous Ovarian Cancer Across Two Independent Datasets

Molecular and Cellular Pathobiology

Cancer Research

Activation of NF-κB Signaling by Inhibitor of NF-κB Kinase β Increases Aggressiveness of Ovarian Cancer

Lidia Hernandez¹, Sarah C. Hsu¹, Ben Davidson², Michael J. Birrer³, Elise C. Kohn¹, and Christina M. Annunziata¹

Clinical Cancer Research

Home | OnlineFirst | Current Issue | Past Issues | Subscriptions | Alerts | Feedback | AACR Publications | AACR Home

High-Risk Ovarian Cancer Based on 126-Gene Expression Signature Is Uniquely Characterized by Downregulation of Antigen Presentation Pathway

JNCI JOURNAL OF THE NATIONAL CANCER INSTITUTE

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals | Medicine | JNCI J Natl Cancer Inst | Volume 104, Issue 9, pp. 670-681.



Nominations for the eighth biennial Carcinogenesis awards now open. Click for more information.

A DNA Repair Pathway–Focused Score for Prediction of Outcomes in Ovarian Cancer Treated With Platinum-Based Chemotherapy

Josephine Kang, Alan D. D'Andrea and David Kozono

« Previous | Next Article » Table of Contents

This Article

JNCI J Natl Cancer Inst (2012) 104, 670–681.
doi:10.1093/jncnjc177

Search this journal:
Advanced >

Current Issue

January 16, 2013 105 (2)

nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive

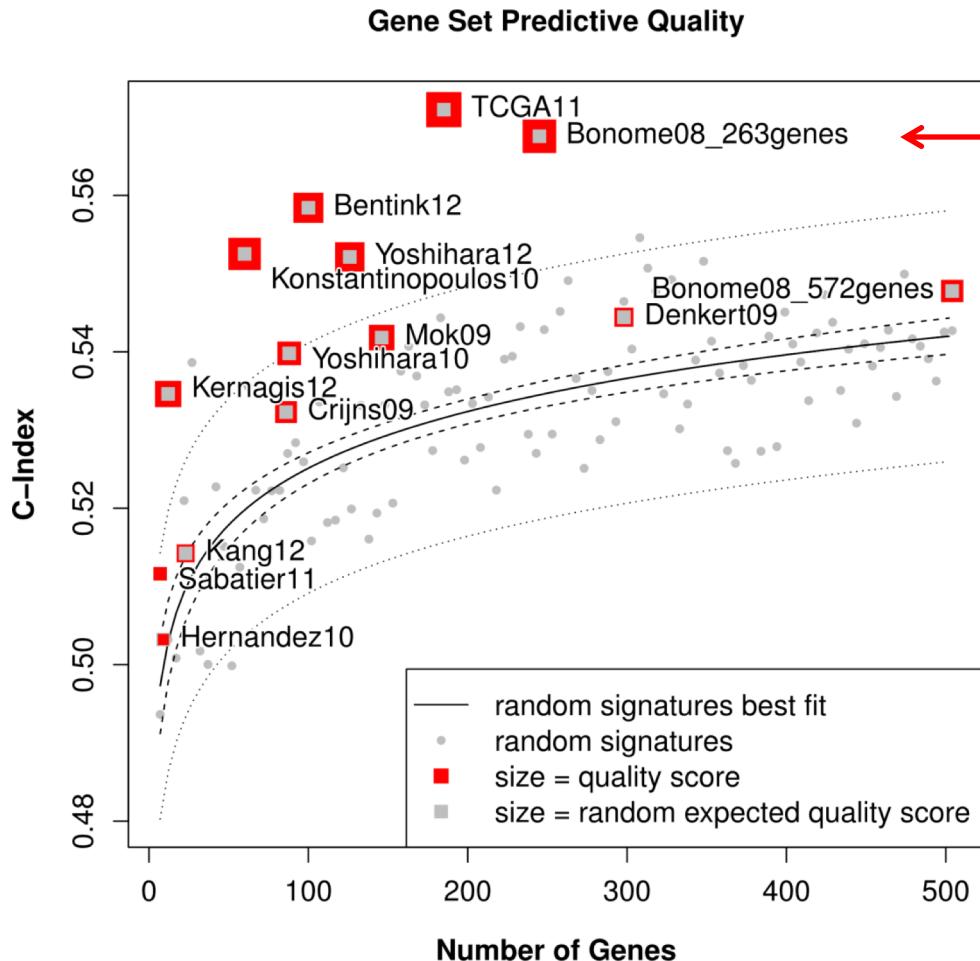
Archive > Volume 474 > Issue 7353 > Articles > Article

NATURE | ARTICLE

日本語要約

Integrated genomic analyses of ovarian carcinoma

Many published prognostic signatures no better than random



About half of gene signatures provide prognostic “value added” over 97.5% of gene random signatures

Lack of robustness in molecular studies of Ovarian Cancer

“Current subtypes do not meet statistical criteria for robustness to re-clustering across multiple datasets (Prediction Strength < 0.6). “

	IMMUNOREACTIVE	DIFFERENTIATED	PROLIFERATIVE	MESENCHYMAL
Patient Ages	~ 61 years	~ 55 years	~ 64 years	~ 59 years
Risk (5-year survival %)	Low (50%)	High (34%)	High (34%)	Very High (20%)
Purity of TCGA samples (ABSOLUTE)	71%	87%	91%	62%
Lymphocyte Infiltration (TCGA samples)	~ 24%	~ 41%	< 5%	< 5%
Neutrophil Infiltration (TCGA samples)	~8%	~10%	< 5%	< 5%

Chen et al., Clin Cancer Res. 2018 Oct 15;24(20):5037-5047.

Subtype also called C1, C2, C4,C5

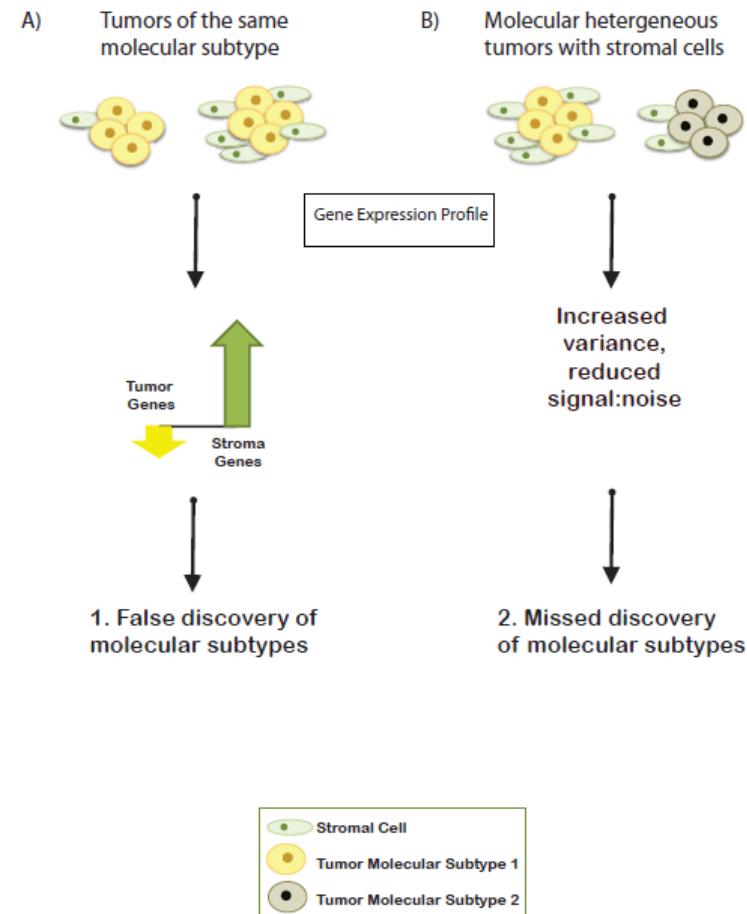
Impact of Admixture on molecular studies of Ovarian Cancer

- Hypothesis;

Tissue admixture will impacts discovery of molecular signature or subtypes

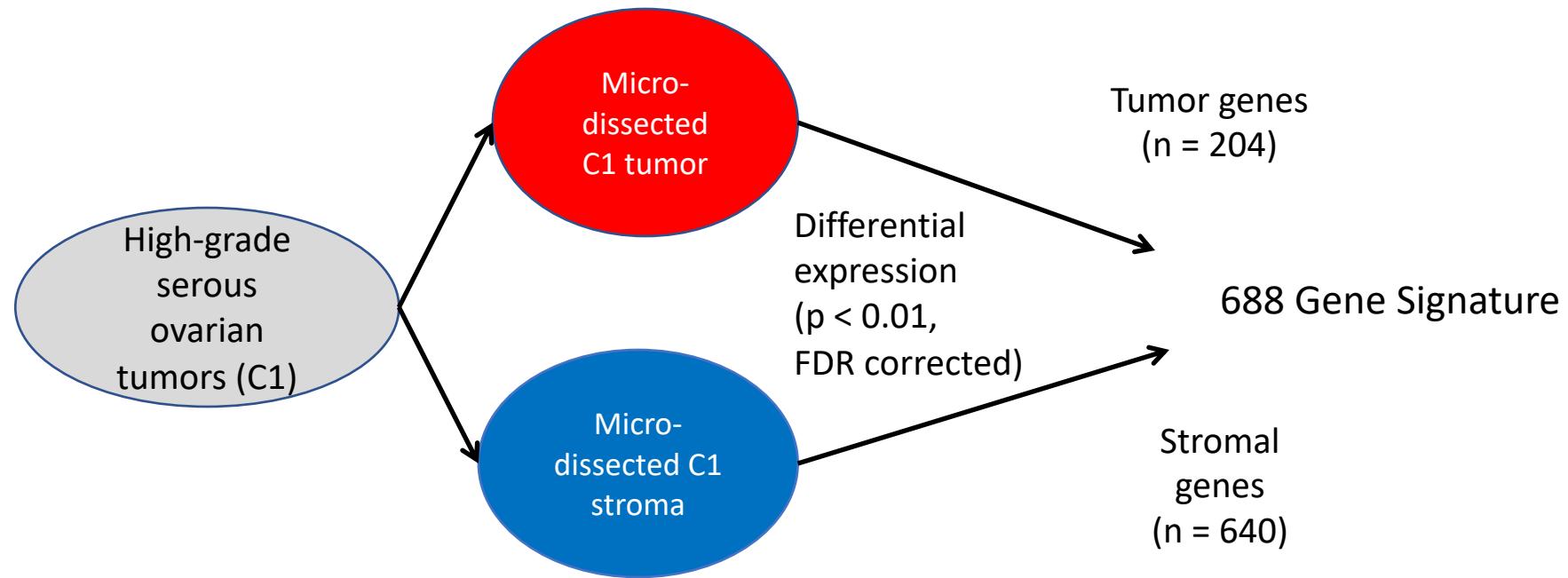
- The % tumor in tissue is variable between studies, subtypes discovered in one study maybe difficult to reproduce
- Molecular subtypes and gene signature will capture cell composition or cell lineage of cells in the tissue

<https://www.biorxiv.org/content/10.1101/496406v1.full>



Matt Schwede

Compared tumors to microdissected tissue to their matched bulk tumors



AOCS C1

Micro-dissected stroma clustered bulk tumor. Micro-dissected tumor did not cluster with its bulk

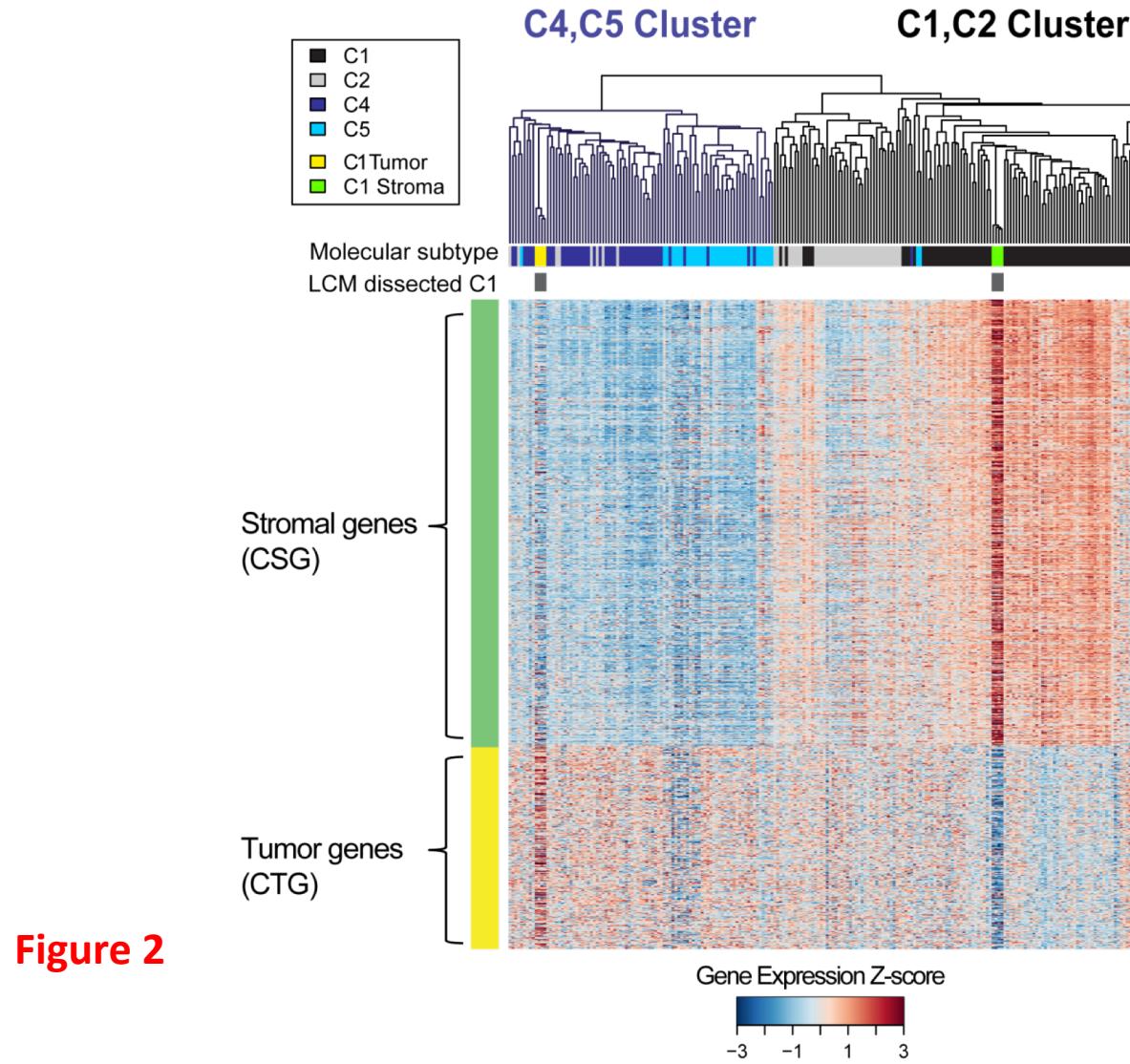


Figure 2

688 Tumor-Stroma
signature distinguished
AOCS subtypes

Splits C1, C2, C4,C5

C1 stroma clusters with C1 tumor

C1 tumor clusters with C4

Global differential gene expression analysis between each pair of AOCS tumor molecular subtypes

Affymetrix_ID	Gene_Symbol	
C1_vs_C2	533	343
C1_vs_C4	1280	821
C1_vs_C5	2011	1267
C2_vs_C4	591	384
C2_vs_C5	1816	1157
C4_vs_C5	928	619

How many of these genes are stromal genes?

Global differential gene expression analysis between each pair of molecular subtypes

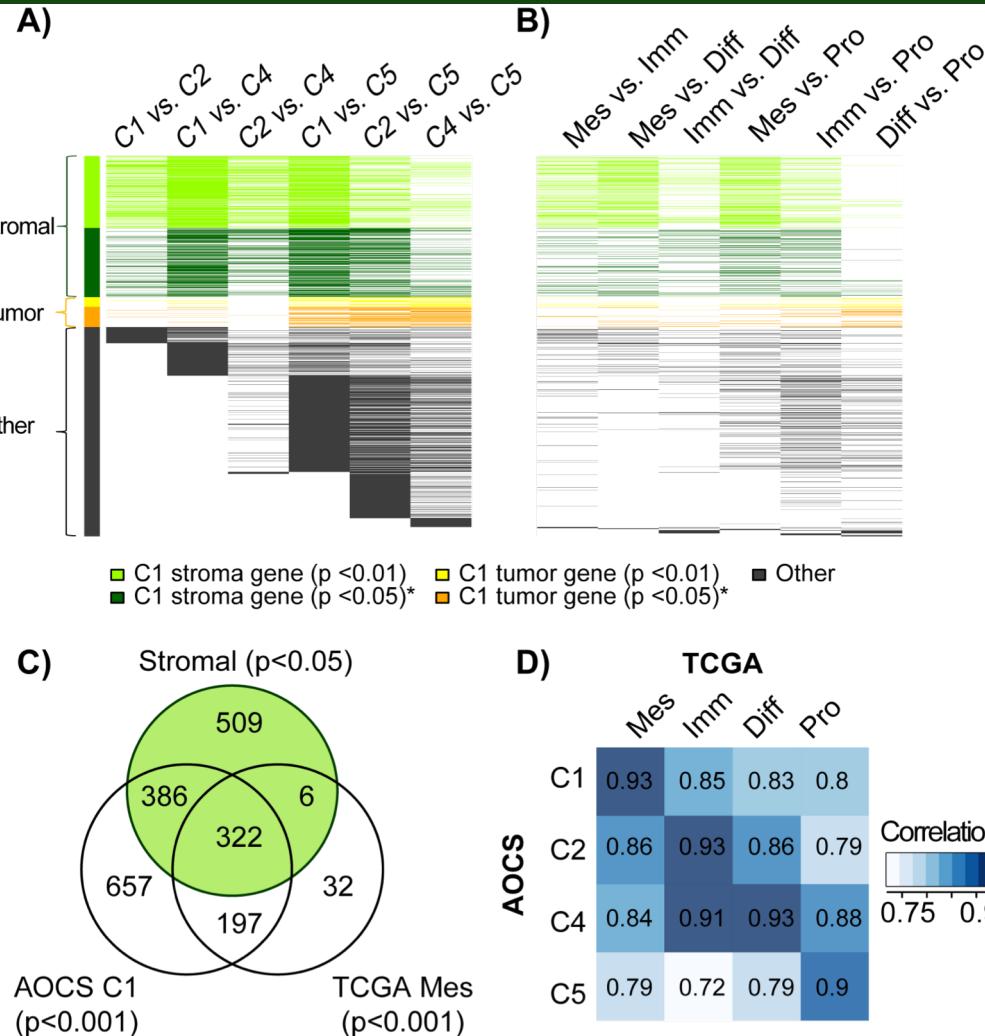


Figure 3

Testing robustness of the AOCS TCGA Subtype Classifiers

- From AOCS supplementary data (Tothill et al.,)
 - Extracted molecular subtype gene list
 - Extracted predicted subtype of each tumor
- For each molecular subtypes (C1,C2,C4,C5)
 - Calculated mean expression of all subtype genes. Did not try to build classifier of C3,C6 or NC (not classified samples)
- To predict classification of new sample
 - Max correlation of new samples to mean profile of C1,C2,C4,C5 (nearest centroid approach)

Stability of subtype with increasing % stroma

100% tumor

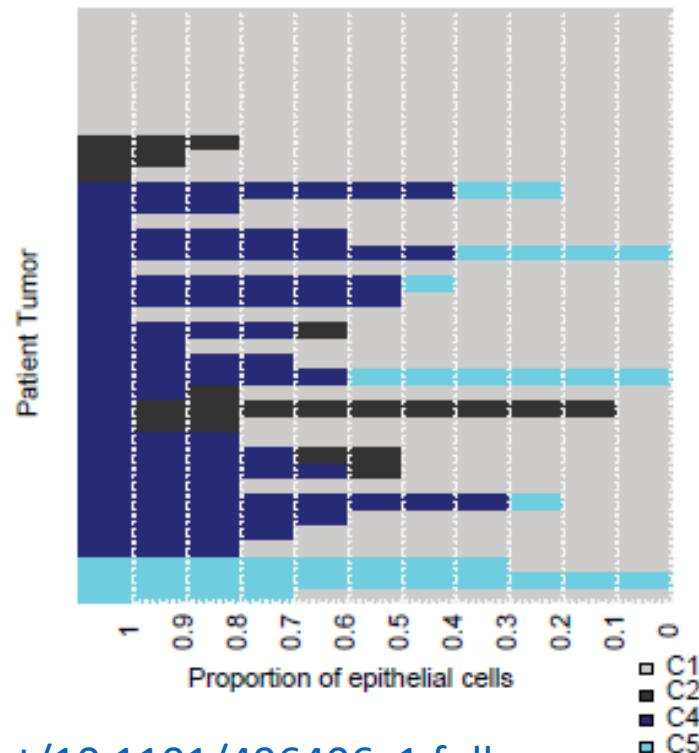
24/68 tumors are C4

8/38 are C1 , 3/8 are C2 or C5.

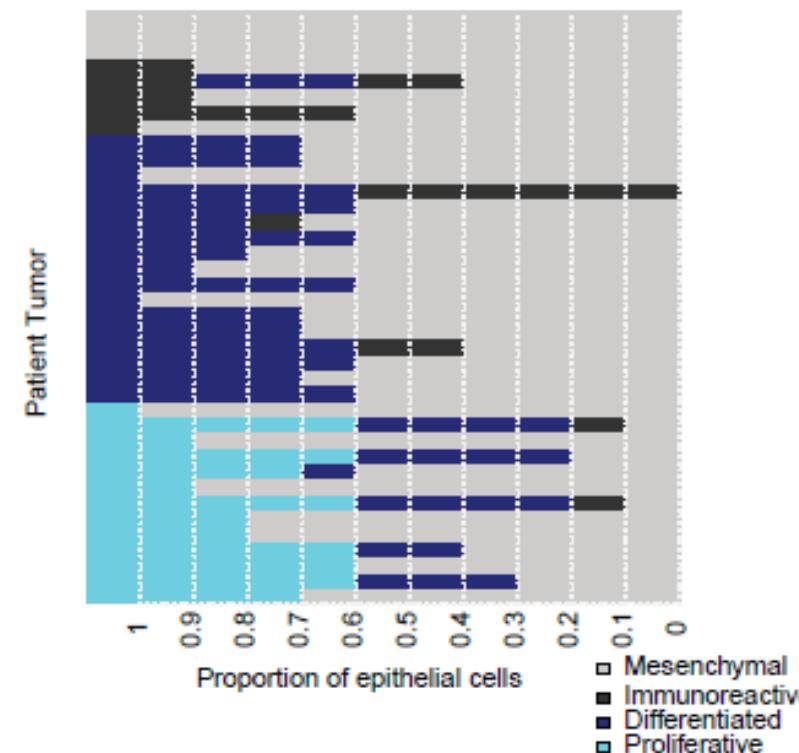
10% stroma, 4/38 tumors are assigned to a different subtype

30% stroma, 15/38 samples are Assigned to a different subtype

AOCS



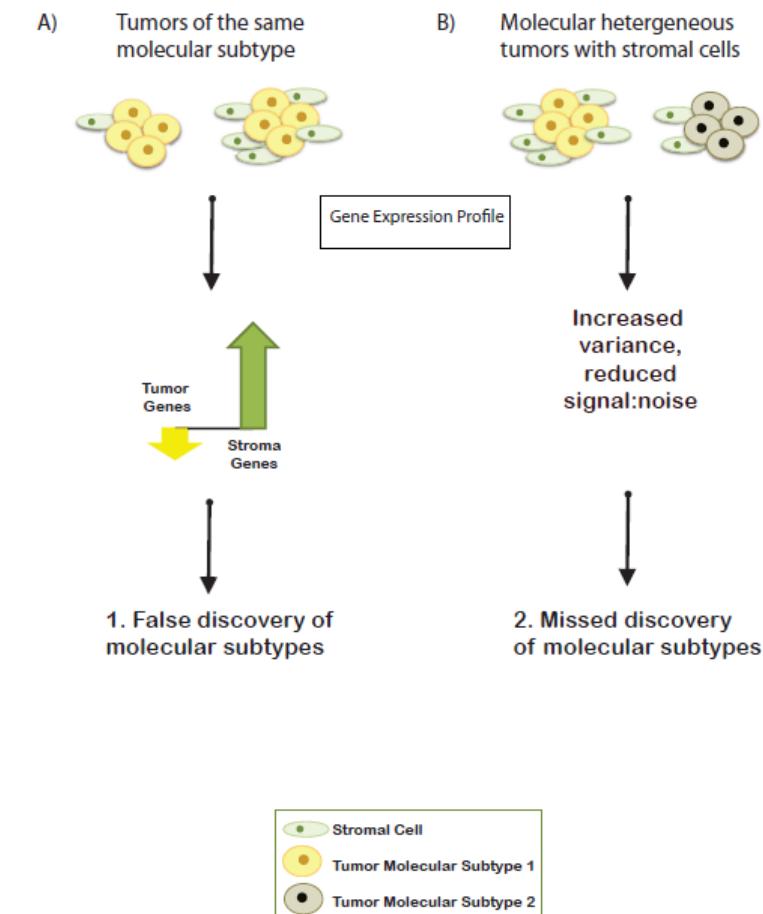
TCGA



<https://www.biorxiv.org/content/10.1101/496406v1.full>

Robustness of Molecular Subtypes

- Cell composition (tumor-stroma ratio) is an important variable in published ovarian cancer gene signatures of molecular subtype.
- Reproducible clinical translation of molecular subtype gene signatures would require rigorous tissue sampling standards



<https://www.biorxiv.org/content/10.1101/496406v1.full>

Is stroma gene expression a component in prognostic Ovarian Gene Signature?

- Examined 61 ovarian gene signatures. Approx 20% enriched in stromal genes.
- These predicted with % stromal content (pathologist estimate) in ovarian, breast and prostate cancer
- Prognostic genes could be identified in both microdissected stromal and epithelial of tumor.

Published gene signatures		Overlap with tumor and stromal genes			TCGA clinical covariates
Study	Description	# genes	# tumor genes	# stromal genes	% Stroma p-value
Tothill et al.	Down in C1	180	2	1	6.76×10^{-11} (+)
	Up in C1	478	0	281	
	Good PFS	300	25	0	8.59×10^{-13} (+)
	Poor PFS	300	0	138	
	Good OS	300	36	0	5.26×10^{-11} (+)
	Poor OS	300	0	138	
Bentink et al.	Non-angiogenic	19	1	1	1.63×10^{-9} (+)
	Angiogenic	74	0	39	
Bignotti et al.	Up in primary	36	1	0	1.16×10^{-9} (+)
	Up in metastasis	89	0	57	
Spentzos et al.	Favorable prognosis	43	5	0	1.66×10^{-10} (+)
	Unfavorable prognosis	73	0	21	
Bonomo et al.	Good prognosis	272	6	2	9.44×10^{-11} (+)
	Poor prognosis	288	0	46	
Biade et al.	Benign cluster	21	0	8	1.08×10^{-12} (-)
	Malignant cluster	15	4	0	
Konstantinopoulos et al.	BRCA-like	32	0	2	0.73 (-)
	Non-BRCA-like	27	0	2	



Schwede et al.,

Published gene signatures		Overlap with tumor and stromal genes			TCGA clinical covariates			
Study	Description	# genes	# tumor genes	# stromal genes	% Stroma p-value	Stage p-value	RFS p-value	OS p-value
Tothill et al.	Down in C1	180	2	1	6.76×10 ⁻¹¹ (+)	0.00116 (+)	0.1420 (-)	0.0198 (-)
	Up in C1	478	0	281				
	Good PFS	300	25	0				
	Poor PFS	300	0	138	8.59×10 ⁻¹³ (+)	0.000236 (+)	0.0936 (-)	0.00447 (-)
	Good OS	300	36	0				
	Poor OS	300	0	138				
Bentink et al.	Non-angiogenic	19	1	1	1.63×10 ⁻⁹ (+)	0.00124 (+)	0.119 (-)	0.0313 (-)
	Angiogenic	74	0	39				
Bignotti et al.	Up in primary	36	1	0	1.16×10 ⁻⁹ (+)	7.54×10 ⁻⁵ (+)	0.0983 (-)	0.0165 (-)
	Up in metastasis	89	0	57				
Spentzos et al.	Favorable prognosis	43	5	0	1.66×10 ⁻¹⁰ (+)	0.00529 (+)	0.0487 (-)	0.0118 (-)
	Unfavorable prognosis	73	0	21				
Bonomo et al.	Good prognosis	272	6	2	9.44×10 ⁻¹¹ (+)	0.554 (+)	0.437 (-)	0.0426 (-)
	Poor prognosis	288	0	46				
Biade et al.	Benign cluster	21	0	8	1.08×10 ⁻¹² (-)	0.739 (+)	0.220 (+)	0.310 (+)
	Malignant cluster	15	4	0				
Konstantinopoulos et al.	BRCA-like	32	0	2	0.73 (-)	0.148 (+)	0.0126 (+)	0.0167 (+)
	Non-BRCA-like	27	0	2				

Are tumor:stroma genes specific to an epithelial carcinoma cell type?

Published gene signature Study	Signature	Ovarian	Breast		Prostate
		(% stroma) TCGA (n = 518)	Boersma (n = 34)	Casey (n = 28)	% non-tumor content ³ Wang (n = 109)
Schwede— TG, SG	C1 stromal genes	5.6×10^{-13} (+)	4.3×10^{-8} (+)	7.1×10^{-8} (+)	1.5×10^{-9} (+)
	C1 tumor genes	3.8×10^{-11} (+)	2.8×10^{-8} (+)	5.1×10^{-7} (+)	4.9×10^{-8} (+)
AOCS (Tothill et al.)	Up in C1				1.1×10^{-14} (+)
	Down in C1				
Bentink et al.	Poor PFS	8.6×10^{-13} (+)	4.0×10^{-9} (+)	4.1×10^{-8} (+)	1.2×10^{-9} (+)
	Good PFS				2.0×10^{-17} (+)
Bignotti et al.	Poor OS	5.3×10^{-11} (+)	4.6×10^{-9} (+)	9.7×10^{-8} (+)	1.7×10^{-5} (+)
	Good OS				6.5×10^{-13} (+)
Spentzos et al.	Angiogenic	1.6×10^{-9} (+)	1.5×10^{-7} (+)	0.93 (+)	1.1×10^{-6} (+)
	Non-angiogenic				6.8×10^{-14} (+)
Bonomo et al.	Up in metastasis	1.2×10^{-9} (+)	4.5×10^{-7} (+)	5.2×10^{-5} (+)	7.2×10^{-4} (+)
	Up primary tumor				7.2×10^{-8} (+)
Biade et al.	Unfav prognosis	1.7×10^{-10} (+)	1.8×10^{-5} (+)	0.0074 (-)	4.1×10^{-14} (+)
	Fav prognosis				1.0×10^{-7} (+)
Konstantinopoulos et al.	Poor prognosis	9.4×10^{-11} (+)	3.5×10^{-6} (+)	1.5×10^{-6} (+)	7.7×10^{-5} (+)
	Good prognosis				3.0×10^{-11} (+)
Konstantinopoulos et al.	Malignant cluster	1.1×10^{-12} (-)	9.8×10^{-8} (-)	5.2×10^{-6} (-)	6.5×10^{-10} (-)
	Benign cluster				7.1×10^{-13} (-)
Schwede et al.,	BRCA-like	0.729 (-)	0.0043 (-)	7.6×10^{-6} (-)	0.34 (+)
	Non-BRCA-like				0.19 (+)



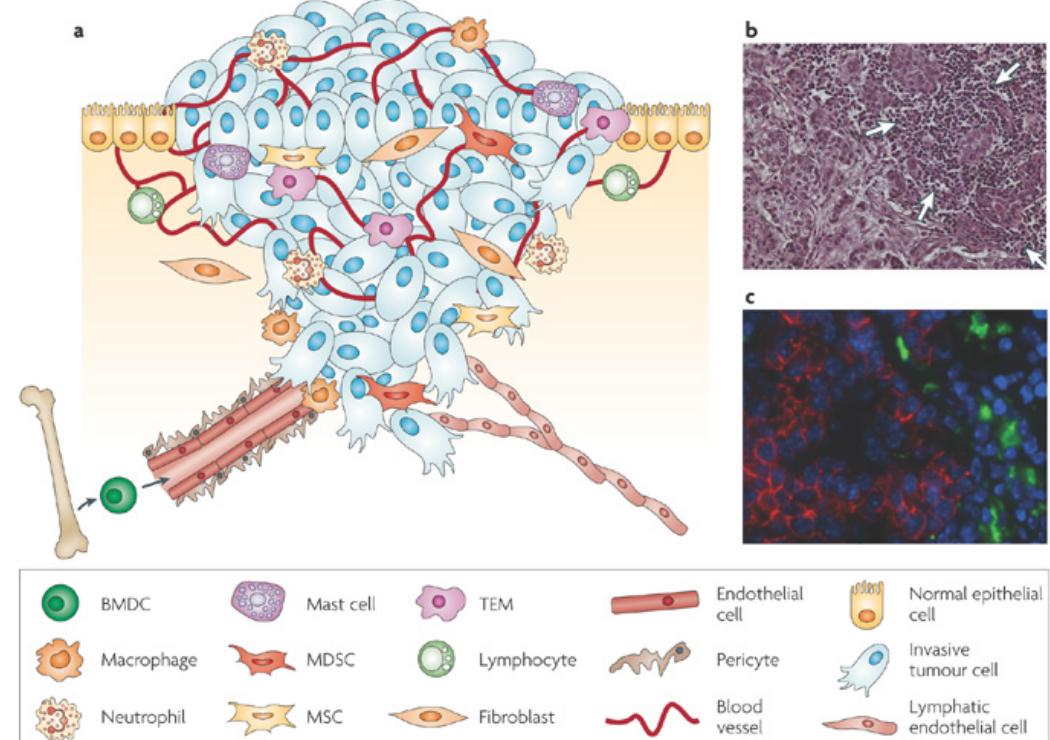
Summary: Robust Molecular Subtypes & Gene Signature

- Cell composition (tumor-stroma ratio) is an important variable in published ovarian cancer gene signatures.
- Reproducible clinical translation of molecular subtype gene signatures would require rigorous tissue sampling standards
- Ratio of tumor:stroma is prognostics in HGSO
- Is this due to tumor sampling or an intrinsic to the tumor biology?

<https://www.biorxiv.org/content/10.1101/496406v1.full>

Tissue Gene Expression: “Observed Signal” is an average over many sources..

- Tumor cell
 - Cell lineage
 - mutations (driver pathway)
 - Copy number effects
- Tumor microenvironment
 - Cell lineages of many cells
 - Immune infiltrate
 - Activated stroma cells



“Hypothetical future workflow”

Annual Review of Biomedical Data Science

From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture

Xi Chen,¹ Sarah A. Teichmann,^{1,2,3}
and Kerstin B. Meyer¹

Annu. Rev. Biomed. Data Sci. 2018. 1:29–51

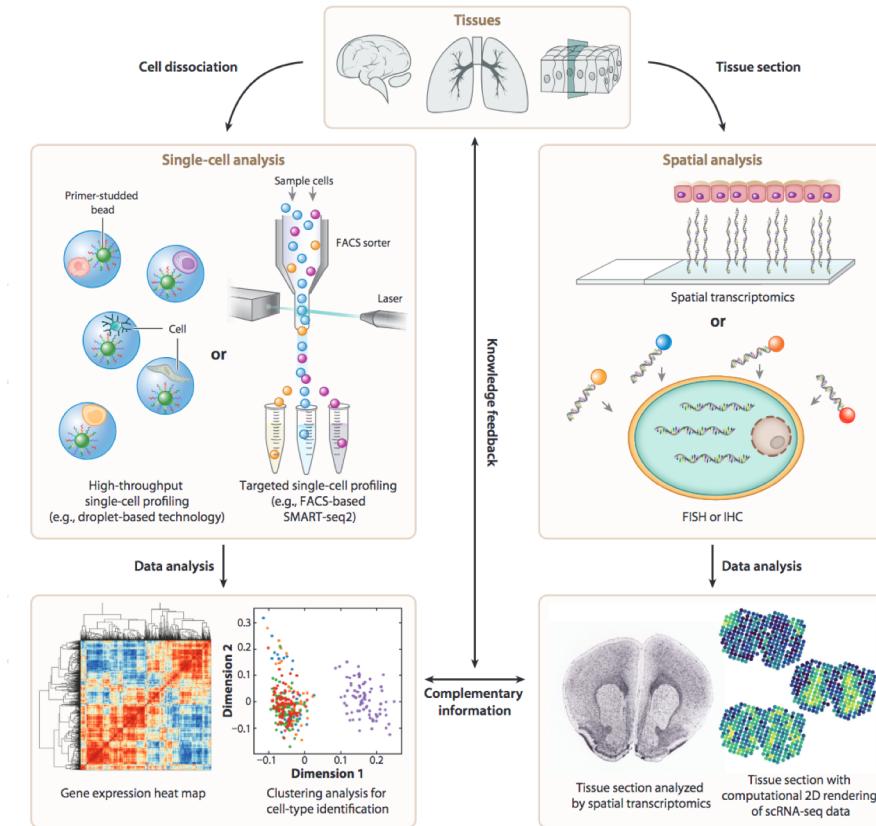
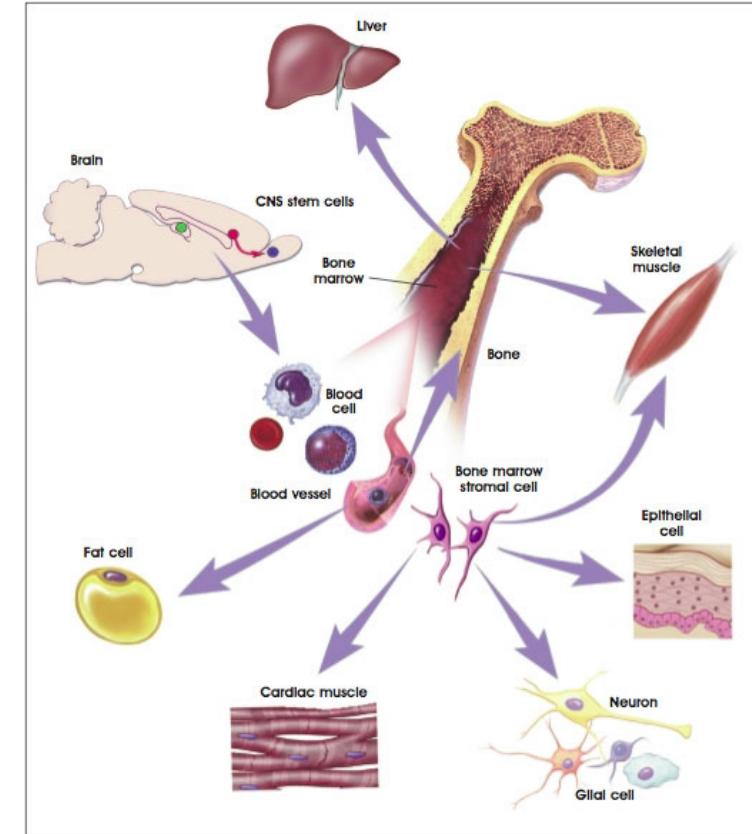


Figure 4

A hypothetical future workflow which combines single-cell genomics and spatial methods to understand tissue architecture.

Summary

- Our organs are complex and many human disease including cancer are multicellular within a complex system
- Deconvolution is a challenge on real data
- Microdissected data, co-culture cell models and leveraging other data can help
- We have developed latent variable approaches and biclustering methods for data integration and multi-omics Gene Set analysis to extract gene patterns that reflect cell states



Acknowledgements

Azfar Basunia

Daniel Gusenleitner

Chen Meng * (with Amin, Bernard)

Matthew Schwede

Oana A. Zeleznik* (with Gerhard)

Technische Universitaet Muenchen, Germany

Amin Moghaddas Gholami, Bernard Kuster

Graz University of Technology, Graz, Austria

Gerhard G. Thallinger

TCGA PanCanAtlas Immune Response Working Group

Vésteinn Thorsson

Ilya Shmulevich

Benjamin Vincent

Thanks also to collaborators

Constanine Mitsiades (DFCI)

Levi Waldron (CUNY)

Vince Carey (Channing)

Toni Choueiri (DFCI)

Kathleen Mahoney (BIDMC)

Elana Fertig (John Hopins)

Rafa Irizarry (DFCI)

Benjamin Haibe Kains (Pharmacodb, Univ Toronto)

Mike Birrer (MGH)

David Livingston (DFCI)

David Harrington (DFCI)

John Quackenbush (DFCI)



@AedinCulhane aedin@jimmy.harvard.edu

