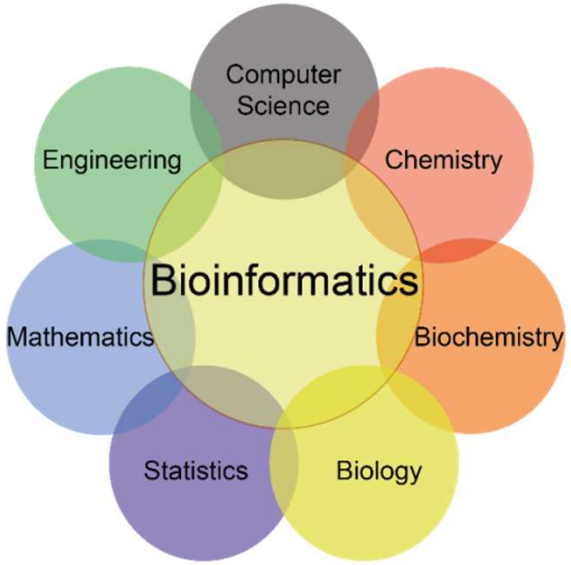
Translational Bioinformatics: Go Deep and Go Broad

- "Working Examples in Deciphering Molecular Heterogeneity of Ovarian Cancer"

Chen Wang wang.chen@mayo.edu
Associate Professor of Biomedical Informatics
Department of Quantitative Health Sciences
Department of Obstetrics and Gynecology
Mayo Clinic

Sep. 20, 2021 - Texas A&M Institute of Data Science Seminar Series

Bioinformatics



1. What's bioinformatics?

"an interdisciplinary* field that develops methods and software tools for understanding biological data" – wiki

*: biology, computer, informatics, machine learning

2. What bioinformatics do?

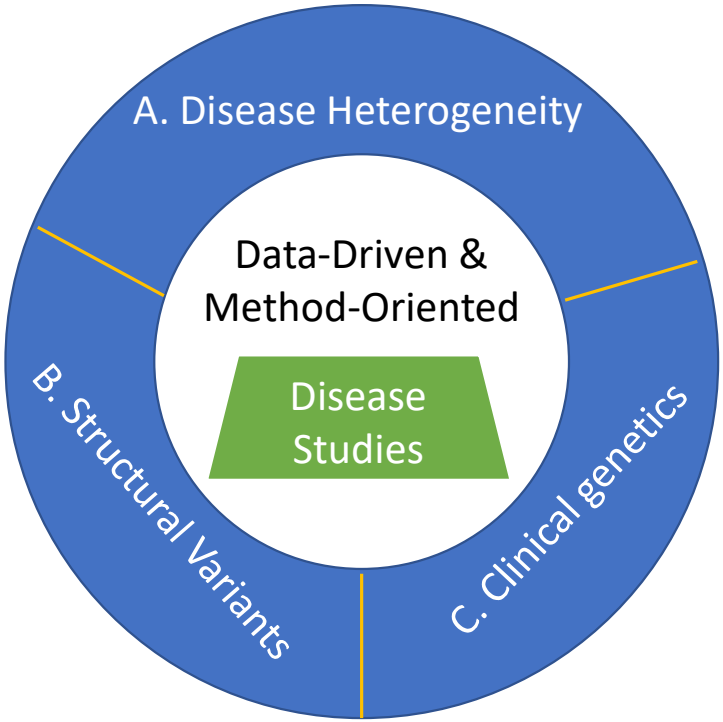
Omics** data-driven ways of …

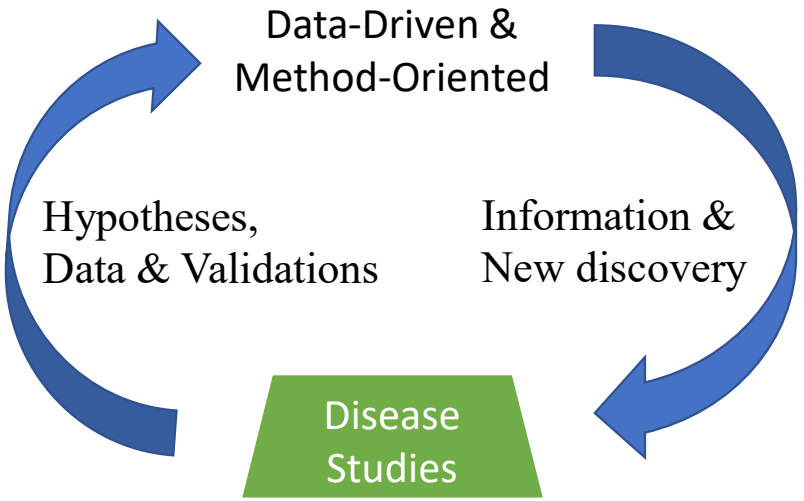
- Discovery: hypothesis formulation
- Validation: hypothesis testing
- Translation: from research discoveries to clinical applications

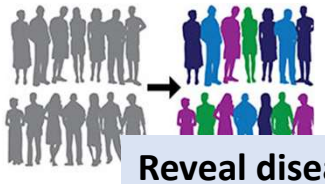
**: genomics, transcriptomics, epigenomics, proteomics, Interactome…

My Research and Collaborative Interests

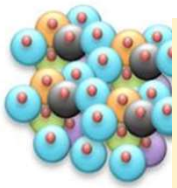
- *I*. Omics data-driven and bioinformatics methodology-oriented
- *II*. iterating between disease studies and method advancements





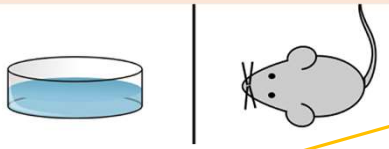


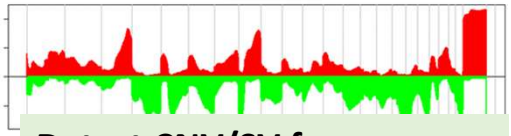
Reveal disease subtypes
unsupervised
semi-supervised
subtypes w. outcomes



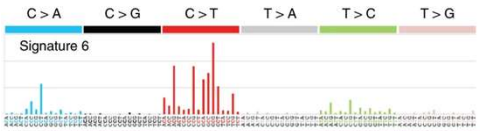
Decipher cellular heterogeneity
cellular sub-population (scRNAseq)
tumor-stroma interactions
cell-cell cross-talk

Build links to model systems
cell-line
organoid
PDX

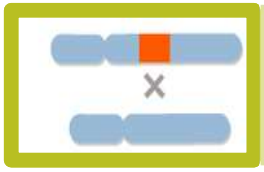




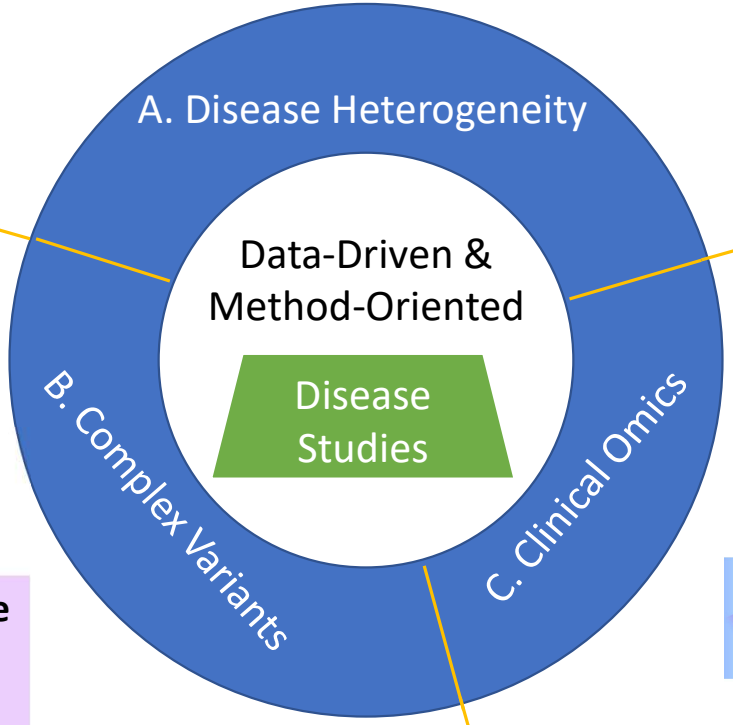
Detect CNV/SV from seq
for exome, whole-genome



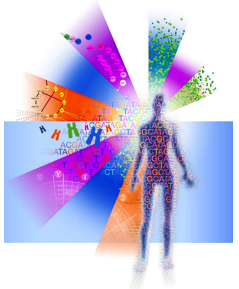
Characterize variant signature
genome instability
mutational signature



Interpret variant's consequence
functional impacts
disease etiology



Research -> Clinic
rigorous performance evaluation
streamlined result interpretation

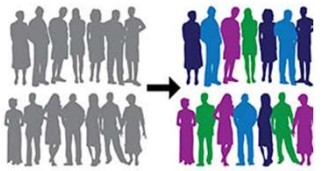


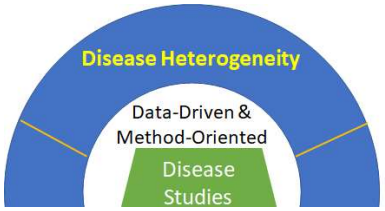
Clinic -> Research
clinical data reuse for research discovery
broad data integrations (EHRs, imaging)

Today's focused topic : Disease Heterogeneity

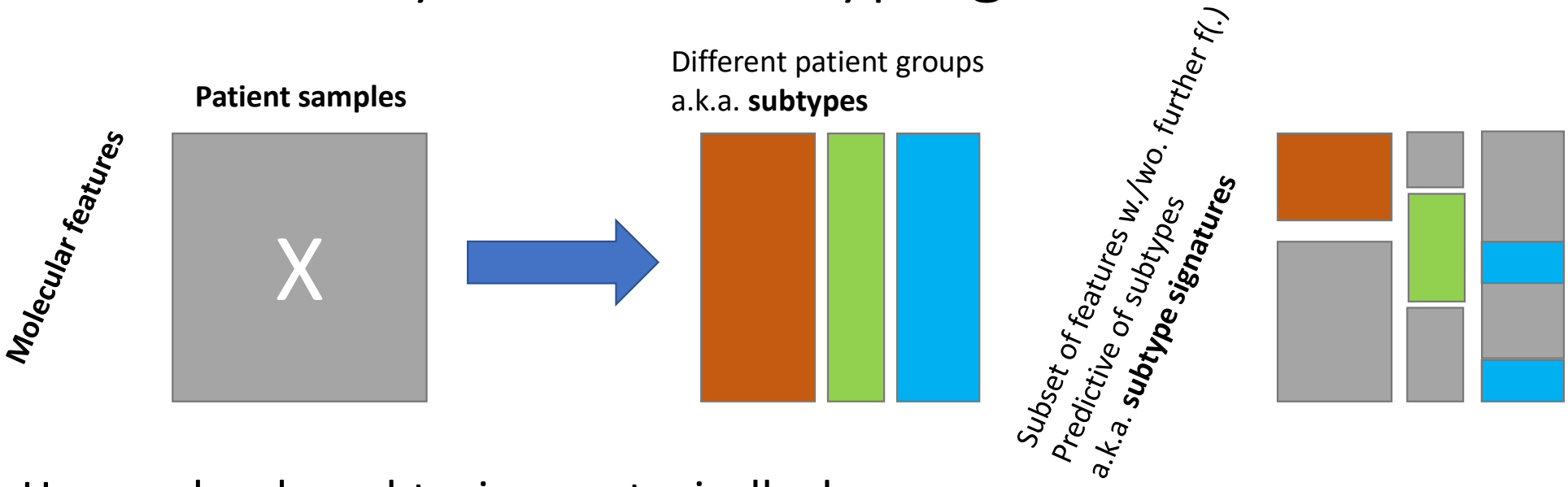
Why to study disease heterogeneity?

- **Clinical goals for disease studies**
  - Etiology: what caused it? How to prevent? Early detect?
  - Diagnosis & Prognosis: which treatment?
  - Therapeutic development: what are targetable to cure disease?
- **Bioinformatics goals for advancing omics analytics**
  - DNA as blueprints: genetics & genomics heterogeneity
  - mRNA & Proteins as dynamic profiles: molecular heterogeneity
  - Cell -> tissue -> disease: microenvironment
  - Varieties of discovery methods: subtyping





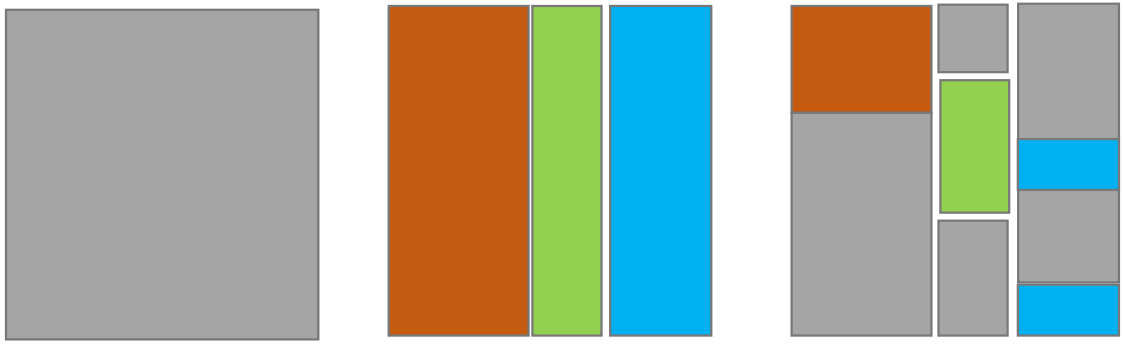
Technically, what's subtyping



How molecular subtyping are typically done

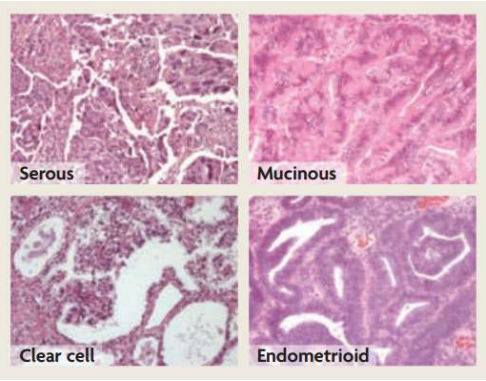
- Clustering of observed data **X** directly
- Decompose observed data **X** to latent space
  - *Matrix decomposition* : PCA, ICA, NMF x linear/kernel, X = WH
  - *Probabilistic representation*: subtype membership => data X distribution
  - *Graph-based mining*: patient as graph-node w. different features; graph-cut, topology

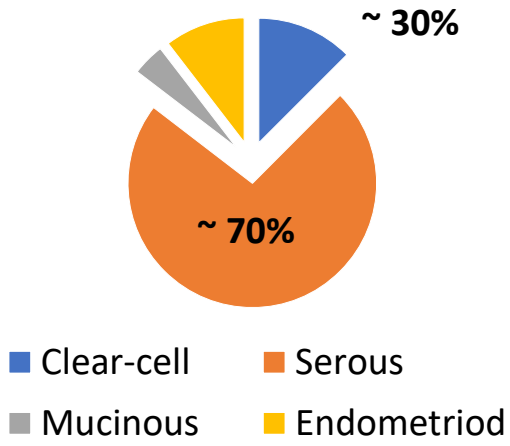
Technical considerations of subtyping



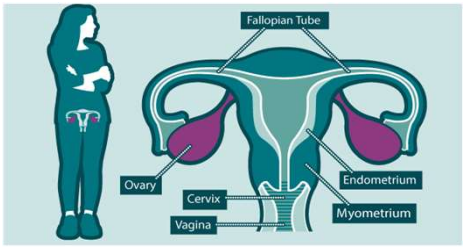
- **General Challenge**: curse of dimensionality
  - E.g. 20,000 features x 50 samples; 450,000 features x 200 samples
- Considerations for **unsupervised** solution
  - What's "overfitting" for un-supervised approach?
  - How many subtypes?
  - Are different subtypes exclusive or transitional to each other?
  - How to generalize from discovery dataset to validation dataset?
- **Semi-supervised** subtyping
  - How to utilize partially labeled data aspects (features and/or samples)?
  - How to leverage known sample-sample, feature-feature relationships?

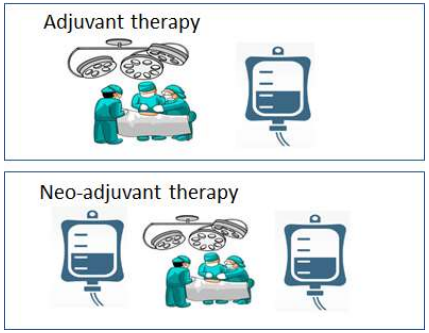
1. go deep: clinical aspects of *ovarian cancer* heterogeneity



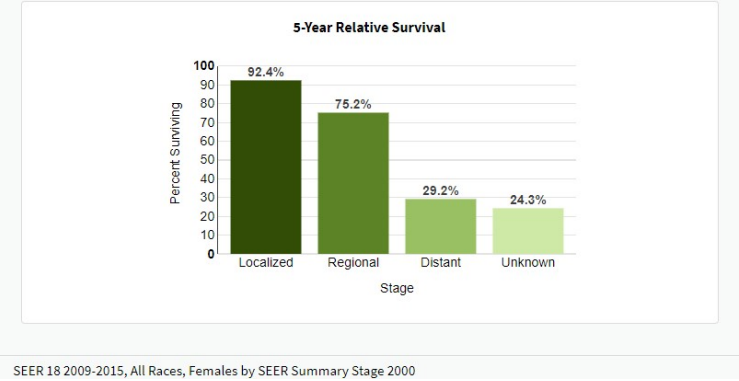


- Despite its heterogeneity, traditional treatments are relatively homogenous

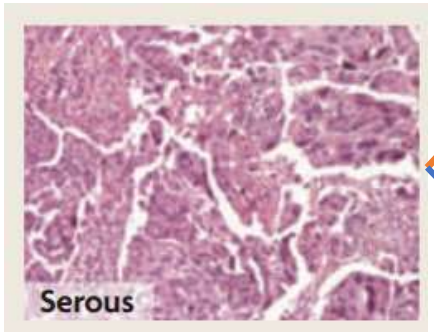




- Stage is the leading prognosis factor

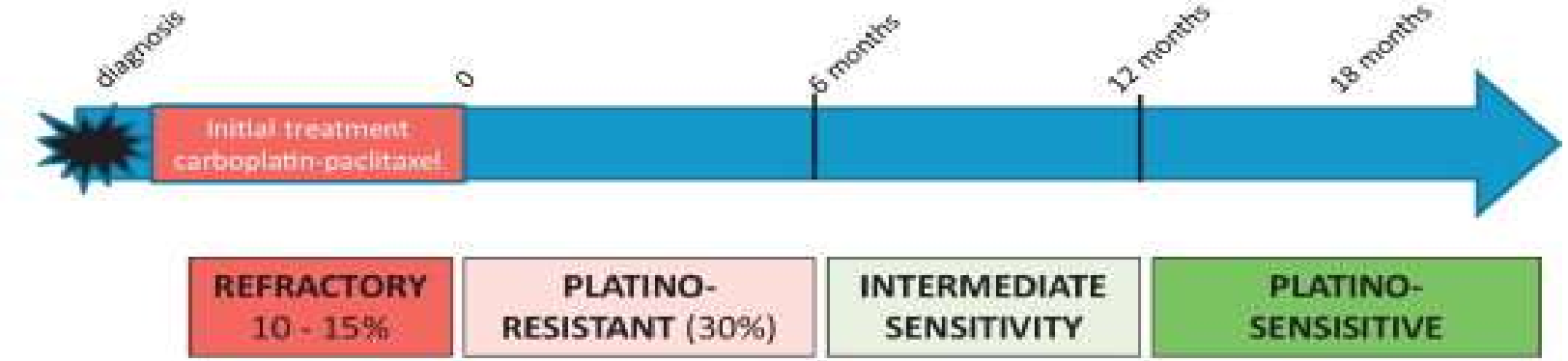


Heterogeneity of HGSOC



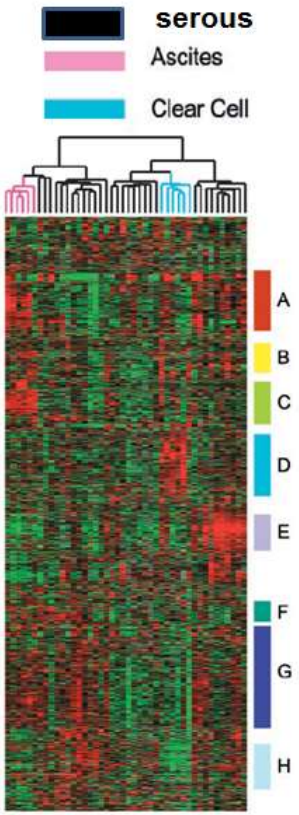
High-grade serous ovarian cancer(HGSOC)
~95%, most diagnosed in advanced stage
Median onset age: 63 years

Low-grade serous ovarian cancer(LGSOC)
~5% of all serous cases,
Median onset age: 47 years



Diverse responses to platinum-based chemo
Any clues in molecular & genomics?

2. Molecular Heterogeneity of HGSOC



2003: Marci E. Schaner, "Gene Expression Patterns in Ovarian Carcinomas", Mol Biol Cell. Nov 2003

- 36 tumors
- cDNA microarray

2 subtypes

N = 36

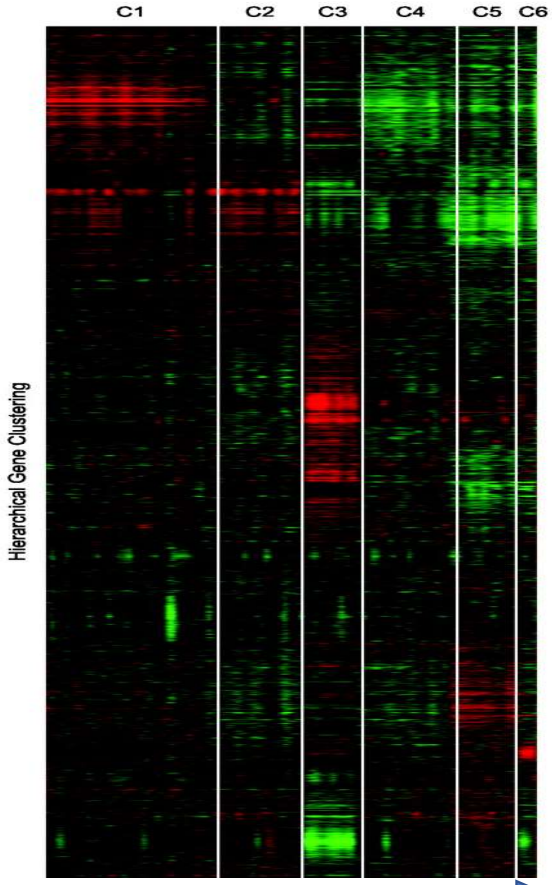
2003

2. Molecular Heterogeneity of HGSOC

2008: Tothill RW et al.“Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome“: Clin Cancer Res 14:5198-5208, 2008

~180 serous tumors

Affymetrix U133A

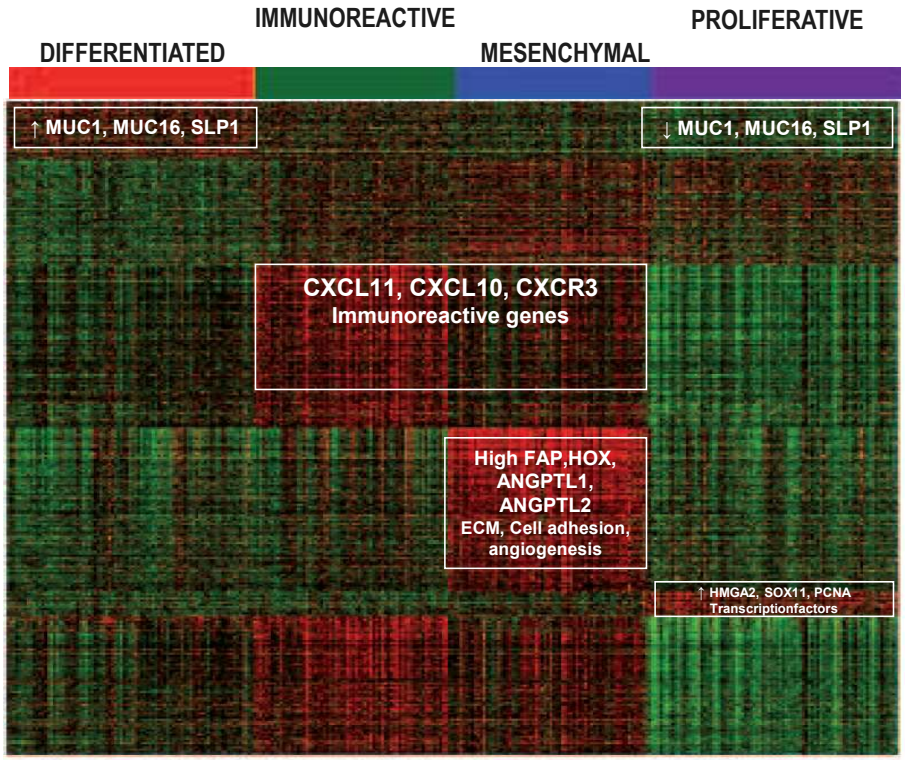


2. Molecular Heterogeneity of HGSOC

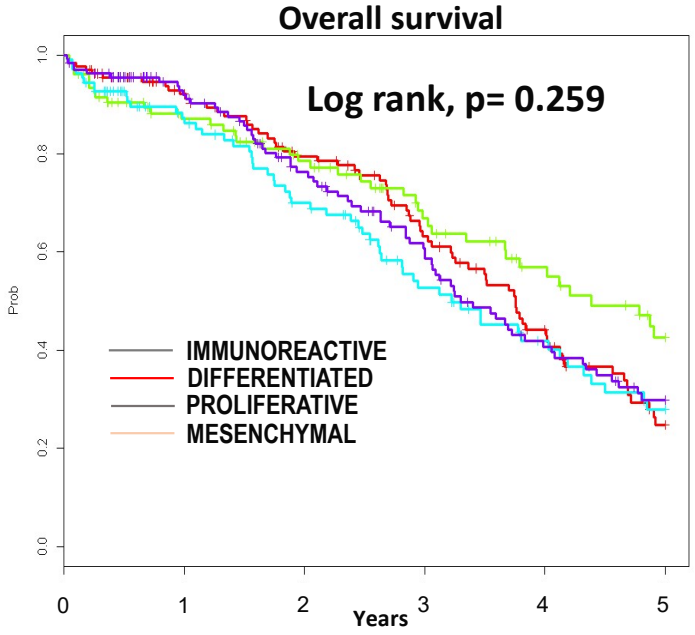
2011: The Cancer Genome Atlas Research Network.
"Integrated genomic analyses of ovarian carcinoma".
Nature 474:609-15, 2011

~500 serous tumors

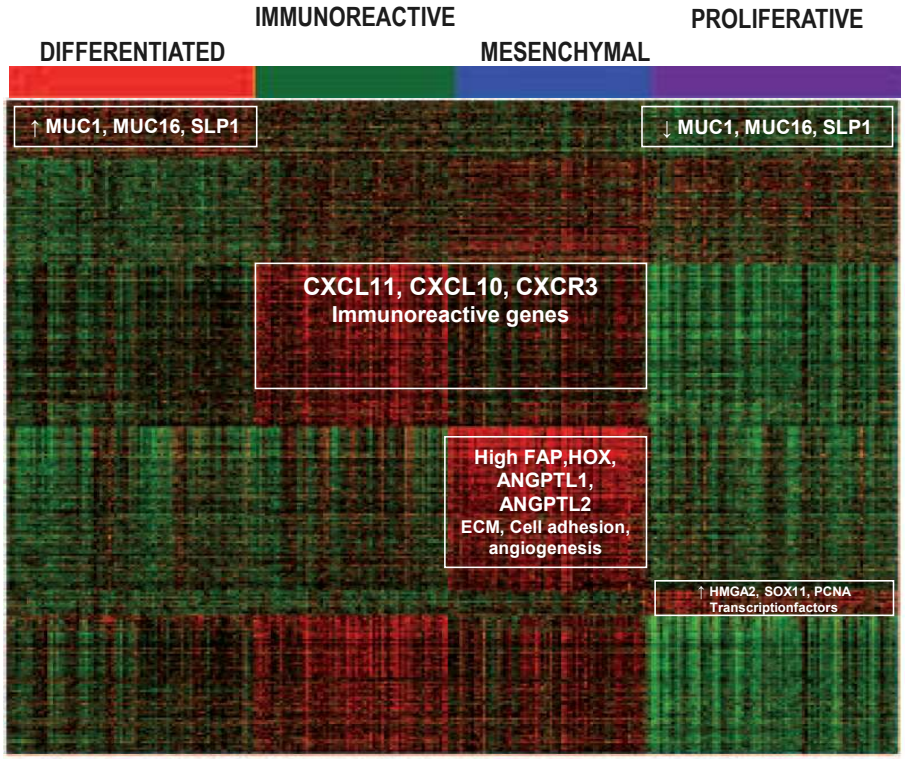
3 microarray platforms measured on all of the samples



2. Molecular Heterogeneity of HGSOC



No prognosis associations



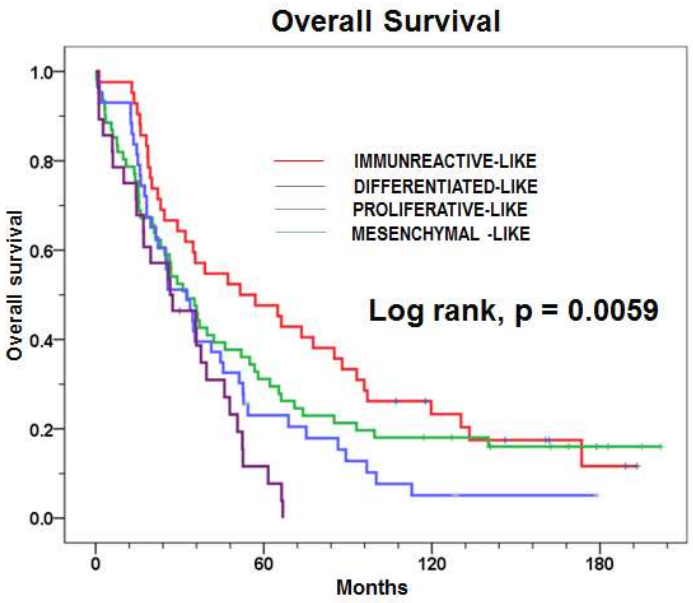
2 subtypes
N = 36

6 subtypes
N = 180

4 subtypes
N = 500

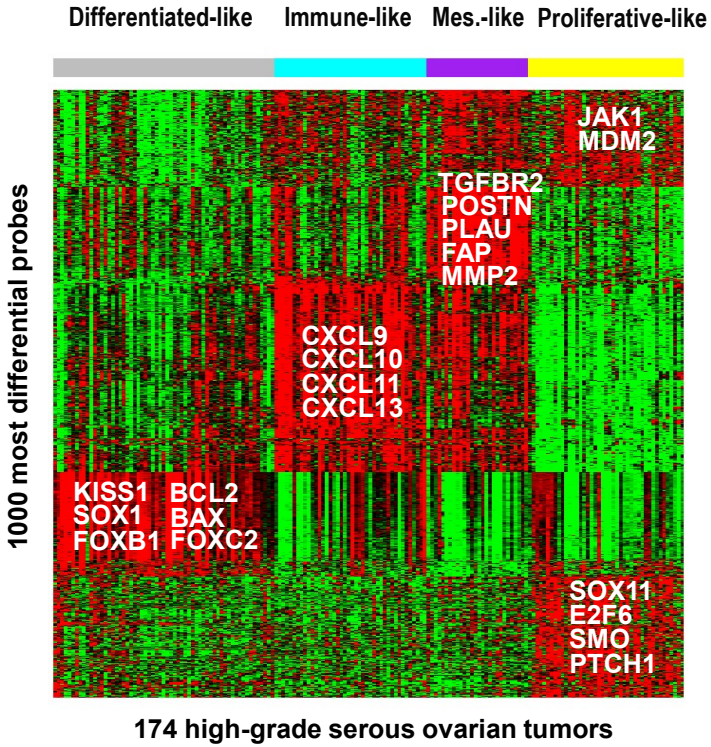
2003 2008 2011

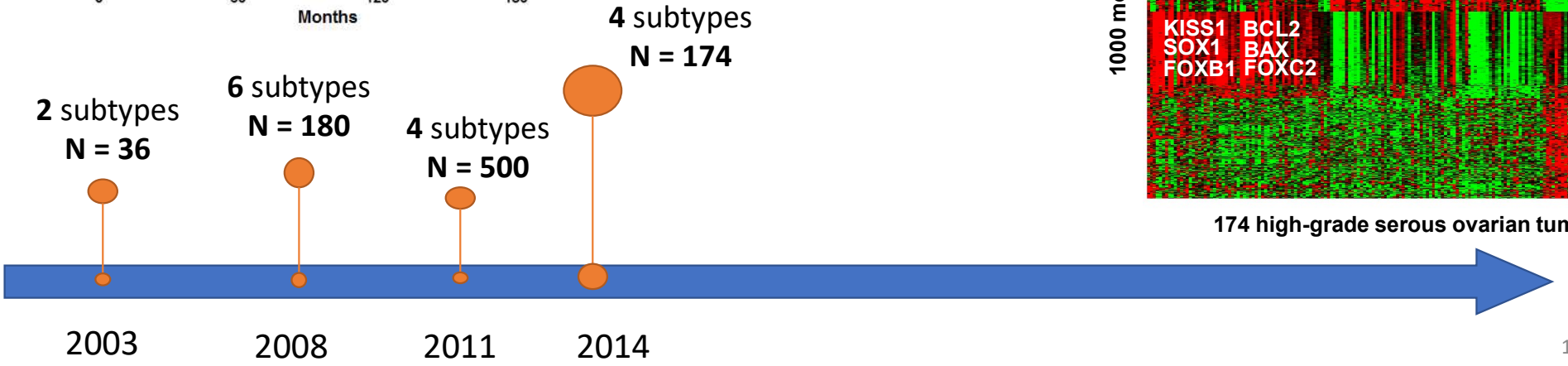
2. Molecular Heterogeneity of HGSOC

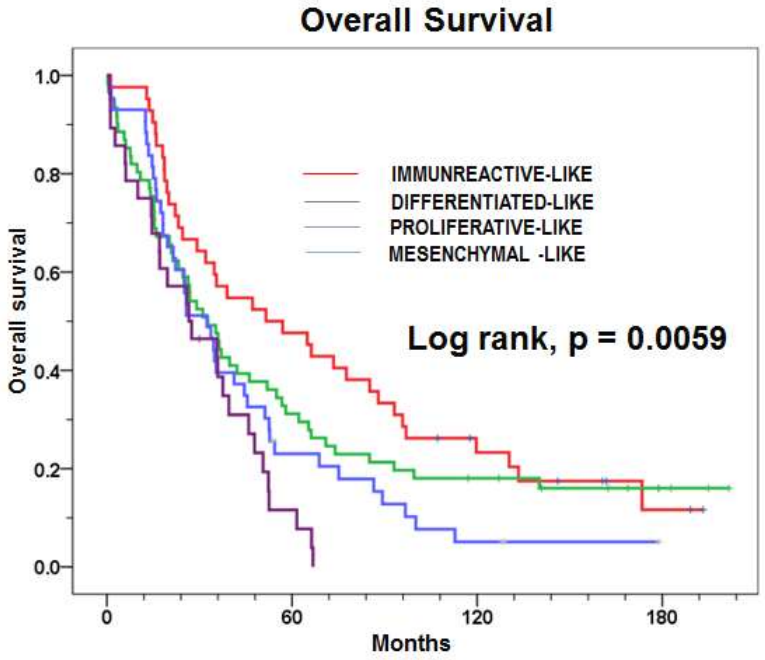


Konecny G**, Wang C**, Hamidi H, et al. "Prognostic and therapeutic relevance of molecular subtypes in high grade serous ovarian cancer". Journal of the National Cancer Institute, 2014.

Significant survival association in Mayo Clinic cohort





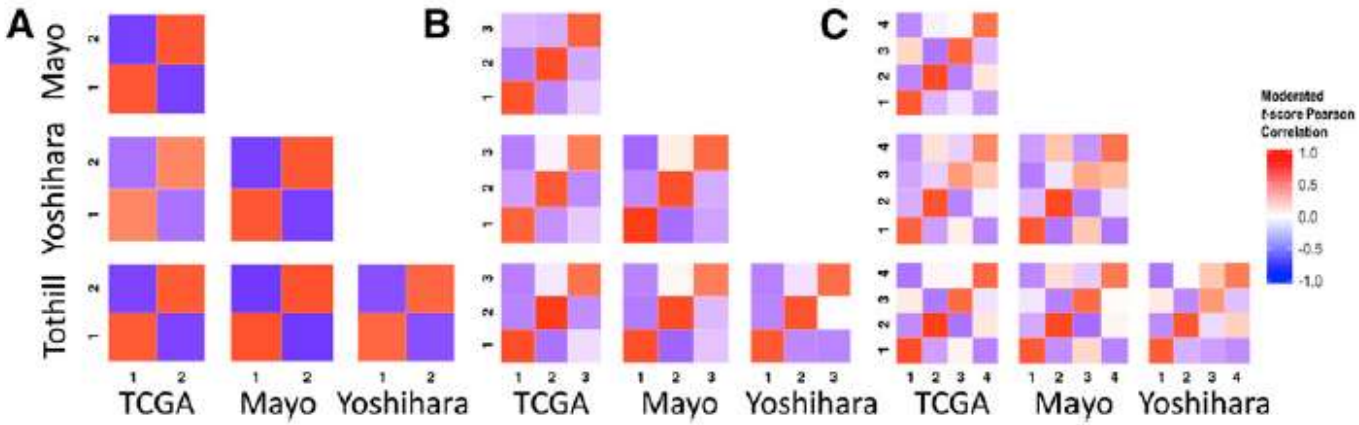


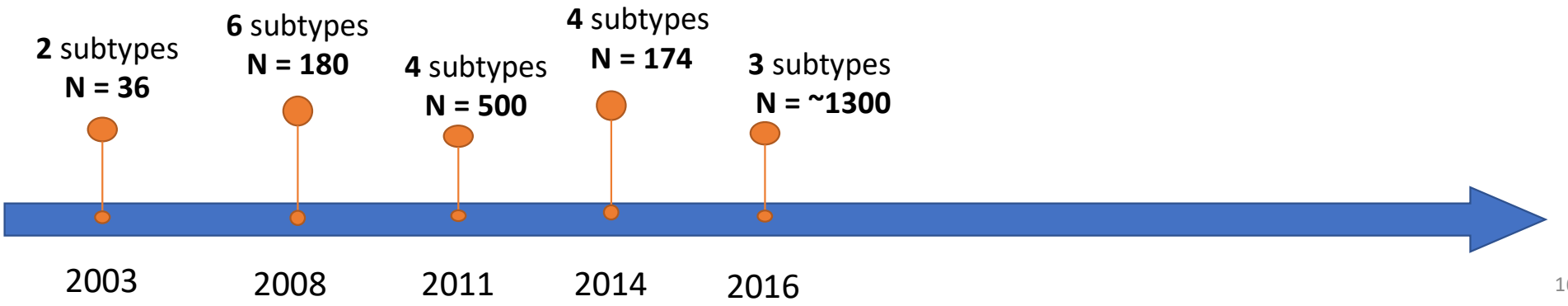
Factors		Estimate HR (95% CI)	Wald test
Age		1.02 (1.00, 1.03)	P=0.0154
Stage	IV vs. II-III	1.74 (1.19, 2.54)	P=0.0045
Grade	4 vs. 2-3	1.23 (0.88, 1.71)	P=0.2217
Debulking	optimal vs. others	0.42 (0.30, 0.61)	P<0.0001
Cluster	C1: Immune-like	1.0 (reference)	
	C2 : Diff.-like	1.25 (0.80, 1.95)	P=0.3293
	C3: Prolif.-like	1.89 (1.18, 3.02)	P=0.0079
	C4: Mes.-like	2.45 (1.43, 4.18)	P=0.0011

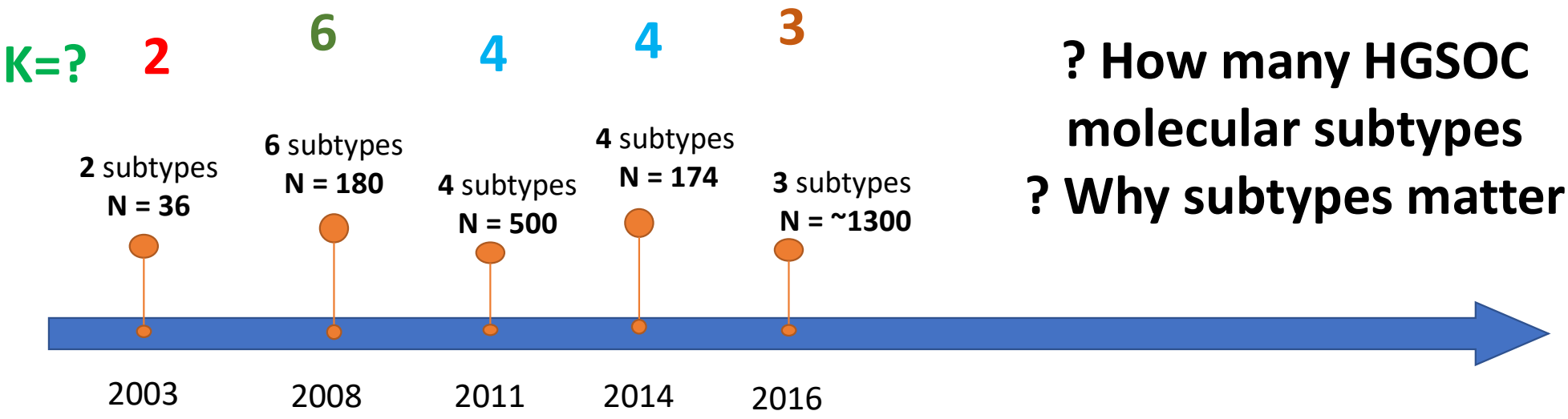
Multivariate analysis

2. Molecular Heterogeneity of HGSOC

2016 "Comprehensive Cross-Population Analysis of High-Grade Serous Ovarian Cancer Supports No More Than Three Subtypes", G3: Genes, Genomes, Genetics







? How many HGSOC molecular subtypes

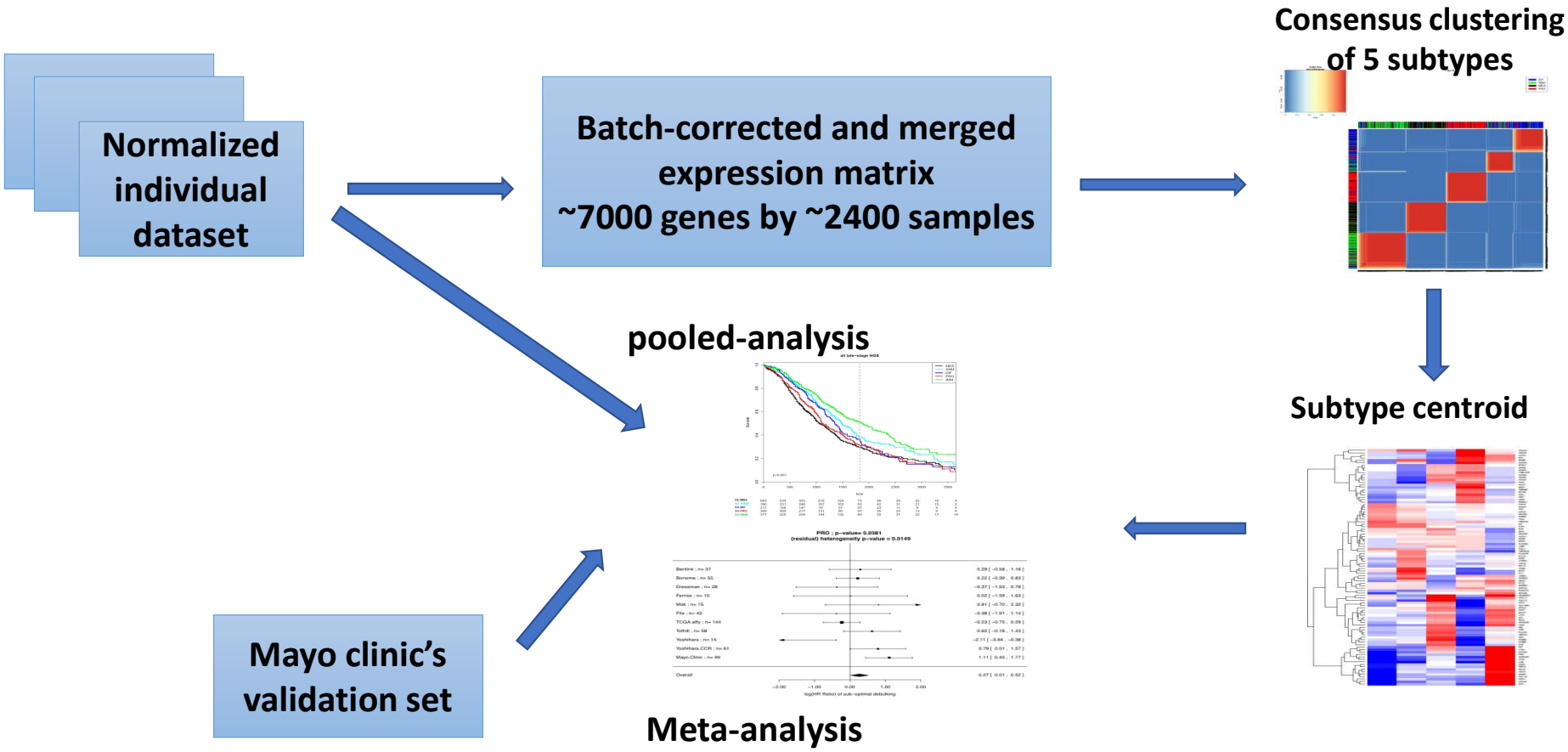
? Why subtypes matter

Motivations:

- **Can we fully utilize all the available public ovarian tumor expression samples to identify tumor subtypes?**
  (independent of individual studies and microarray platform)
- **Whether knowledge of tumor subtypes benefit ovarian cancer treatment decisions?** (e.g. adjuvant vs. neoadjuvant)

CCR 2017, **C Wang**, SM Armasu, et al.

2. A *de-novo* subtyping study for HGSOC



	Stage		Histology				Vital		Recurrence		Debulking	
Study Name	early	late	clearcell	endo	mucinous	ser	deceased	living	norecurrence	recurrence	optimal	suboptimal
Bentink, 1	1	128	0	0	0	129	73	56	0	0	98	28
Bonome, 2	0	185	0	0	0	185	129	56	42	153	90	95
Crijns, 3	0	157	0	0	0	157	113	44	0	0	0	0
Denkert, 4	9	71	2	6	0	68	21	59	50	26	0	0
Dressman, 5	1	115	0	0	0	117	67	50	0	0	63	54
Ferriss, 6	0	58	5	1	1	47	36	22	6	48	26	30
Mateescu, 7	31	76	6	8	8	79	76	31	27	80	0	0
Mok, 8	0	53	0	0	0	53	41	12	0	0	28	11
Pils, 9	9	185	0	0	0	171	57	137	70	124	137	57
TCGA.affy, 10	43	520	0	0	0	568	290	270	279	299	367	140
Tothill, 11	42	240	0	20	0	264	113	169	94	188	160	88
Wu, 12	42	53	8	37	13	41	0	0	0	0	0	0
Yoshihara, 13	0	110	0	0	0	110	46	64	34	76	57	53
Yoshihara.CCR, 14	0	260	0	0	0	260	121	139	0	0	103	157
Total	**178**	**2211**	**21**	**72**	**22**	**2249**	**1183**	**1109**	**602**	**994**	**1129**	**713**

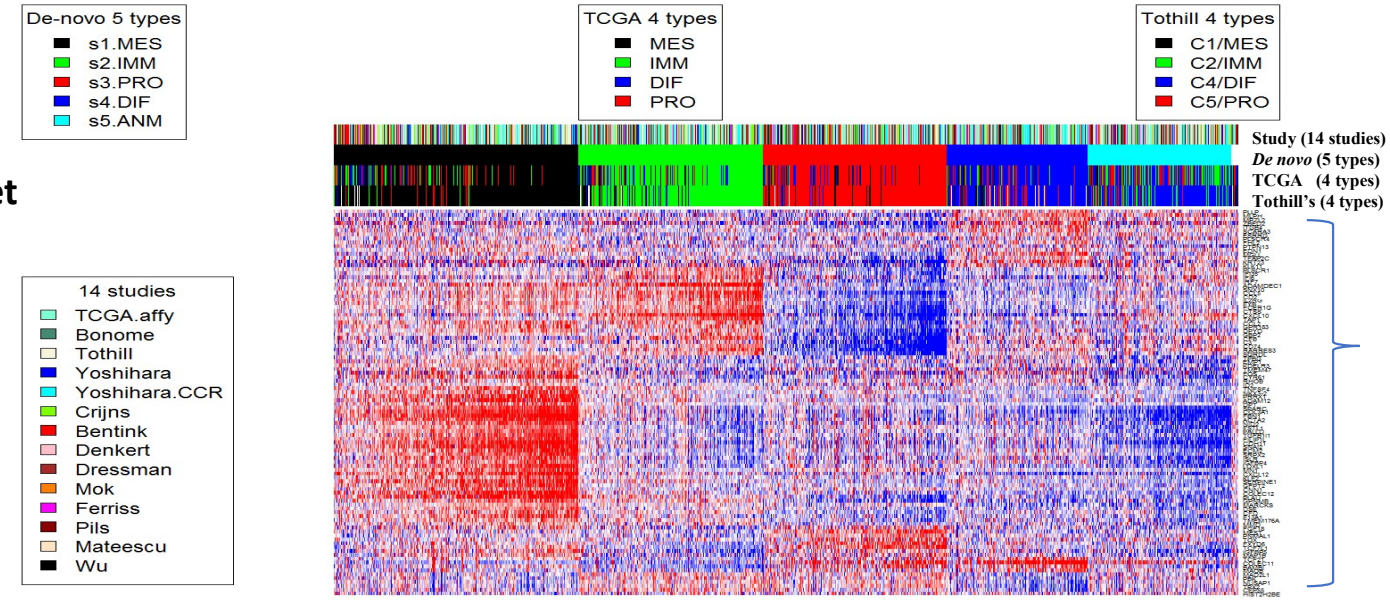
references

[1] Bentink S, Haibe-Kains B, Risch T, et al. Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. PLoS One 2012;7:e30269.
[2] Bonome, T., et al., A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. Cancer Res, 2008. 68(13): p. 5478-86.
[3] Crijns, A.P., et al., Survival-related profile, pathways, and transcription factors in ovarian cancer. PLoS Med, 2009. 6(2): p. e24.
[4] Denkert, C., et al., A prognostic gene expression index in ovarian cancer - validation across different independent data sets. J Pathol, 2009. 218(2): p. 273-80
[5] Dressman, H.K., et al., An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. J Clin Oncol, 2007. 25(5): p. 517-25.
[6] Ferriss, J.S., et al., Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance. PLoS One, 2012. 7(2):
[7] Mateescu, B., et al., miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. Nat Med, 2011. 17(12): p. 1627-35.
[8] Mok, S.C., et al., A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. Cancer Cell, 2009. 16(6): p. 521-32.
[9] Pils, D., et al., Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. Cancer Sci, 2012. 103(7): p. 1334-41.
[10] TCGA, Integrated genomic analyses of ovarian carcinoma. Nature, 2011. 474(7353): p. 609-15
[11] Tothill, R.W., et al., Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res, 2008. 14(16): p. 5198-208
[12] Wu, R., et al., Mouse model of human ovarian endometrioid adenocarcinoma based on somatic defects in the Wnt/beta-catenin and PI3K/Pten signaling pathways. Cancer Cell, 2007.
[13] Yoshihara, K., et al., Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. PLoS One, 2010. 5(3): p. e9615.
[14] Yoshihara, K., et al., High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. Clin Cancer Res, 2012.

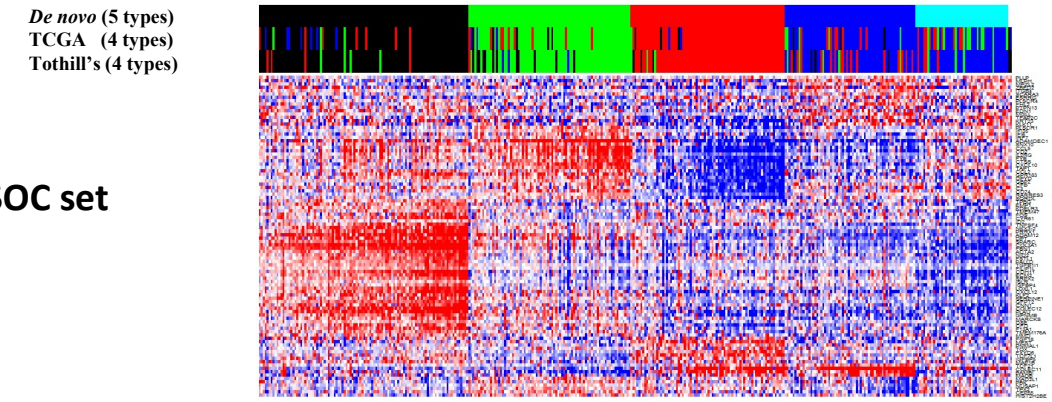
14 HGSOC mRNA studies

2. *de-novo* subtyping system

Public HGSOC set
(n=2,103)



Mayo Clinic HGSOC set
(n=381)



S1.MES: mesenchymal
S2.IMM: immunoreactive
S3.PRO: proliferative
S4.DIF: differentiated
S5.ANM: anti-mesenchymal

Transcriptome to Proteomics
- molecular subtypes confirmed again

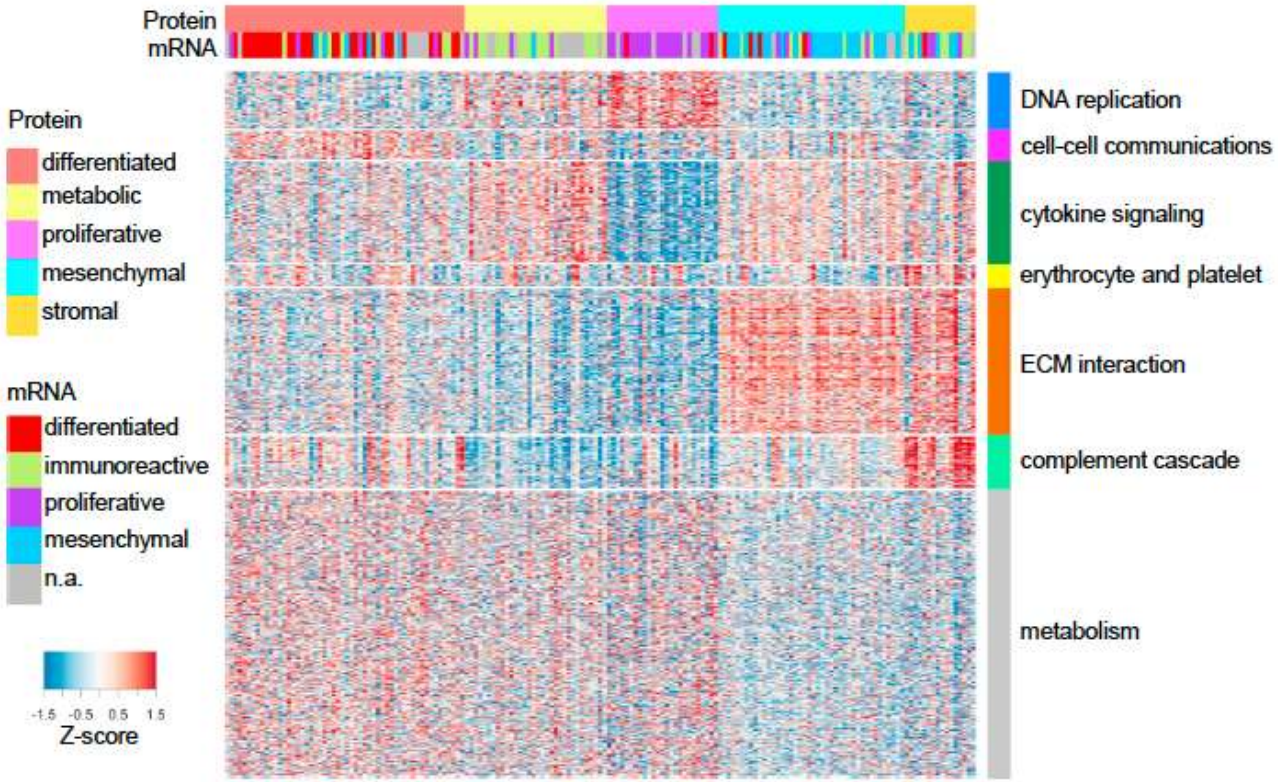
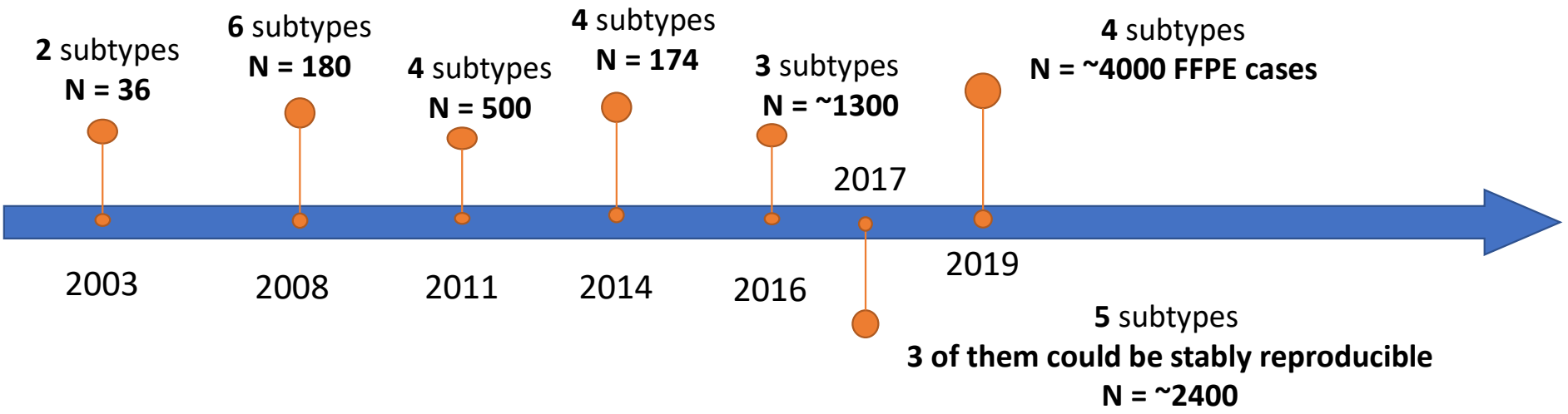


Figure 2. Proteomic Subtypes and Corresponding Driving Protein Modules

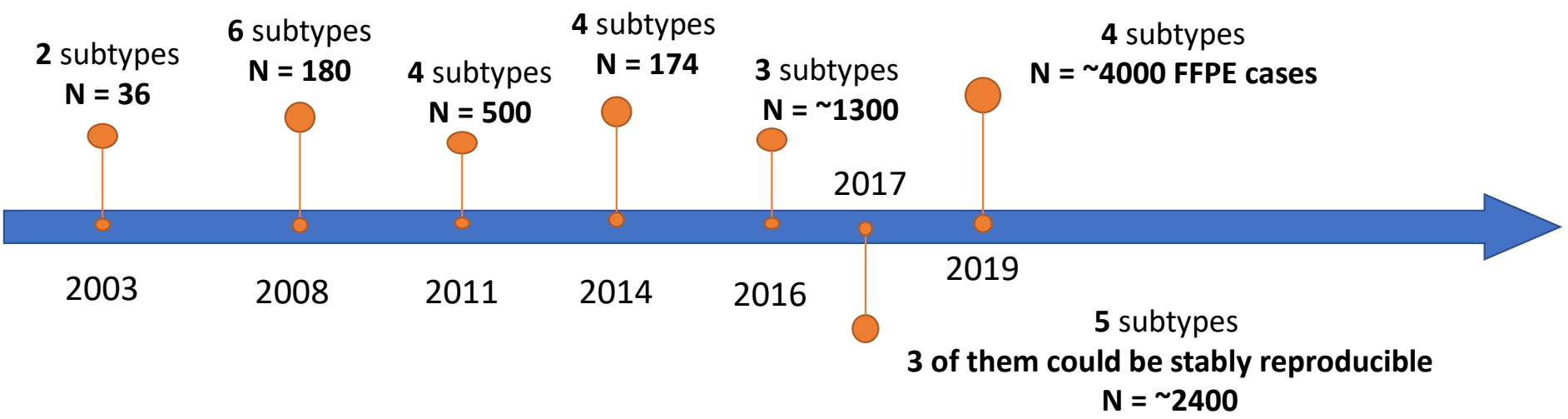
"Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer", *Cell* 2016



Molecular Heterogeneity of HGSOC, and why it matters?

Clinic

Research

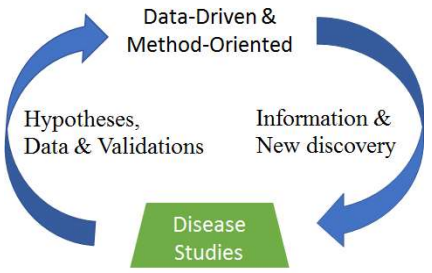


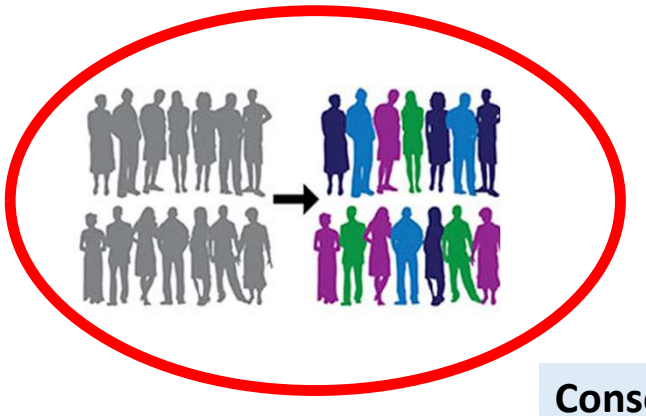
Molecular Heterogeneity of HGSOC, and why it matters?

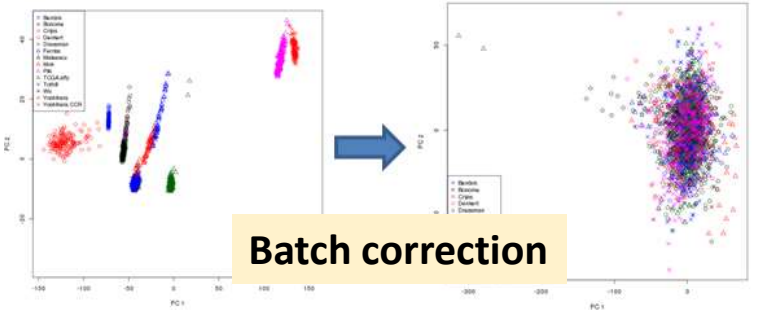
Clinic ⟷ Research

Translational Bioinformatics Methodologies

Informatics methods behind the scenes

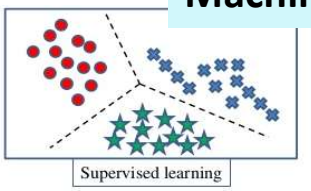


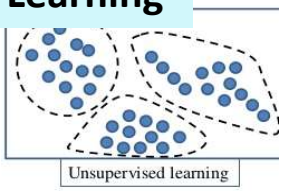


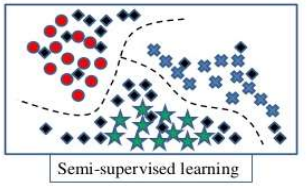


Batch correction

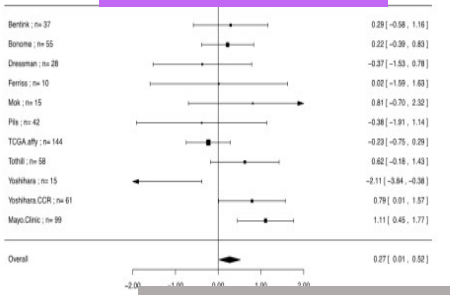
Machine Learning



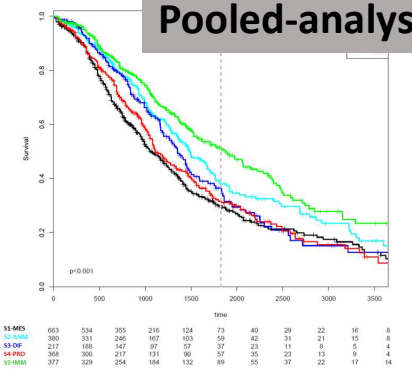




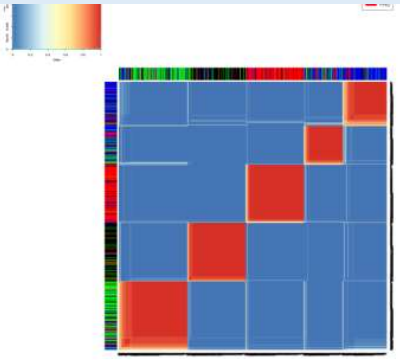
Meta-analysis



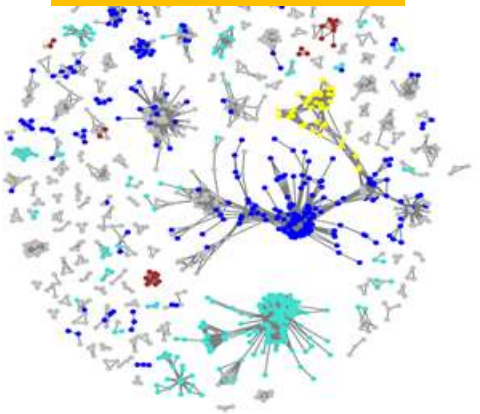
Pooled-analysis



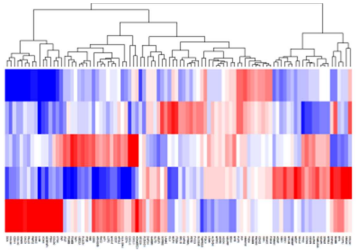
Consensus Clustering

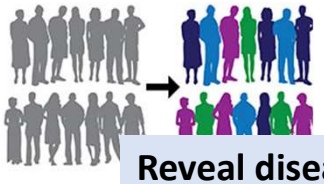


Network Analysis

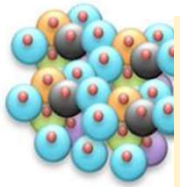


Predictive Signatures



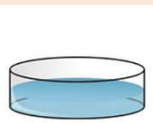


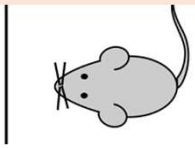
Reveal disease subtypes
- unsupervised
- semi-supervised
- subtypes w. outcomes



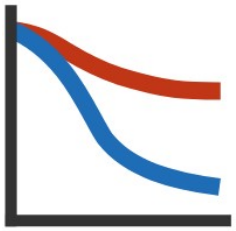
Decipher cellular heterogeneity
- cellular sub-population (scRNAseq)
- tumor-stroma interactions
- cell-cell cross-talk

Build links to model systems
- cell-line
- organoid
- PDX

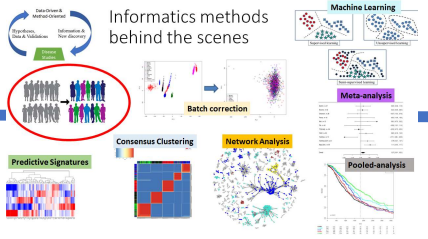




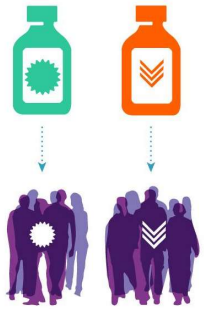
3. Go Deep & Broad: bioinformatics findings -> clinical relevance



Clinic

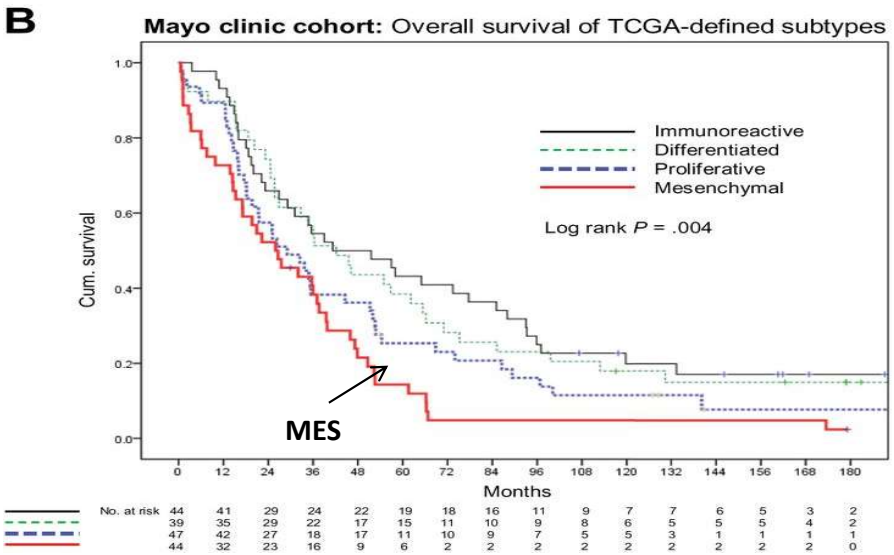


Research

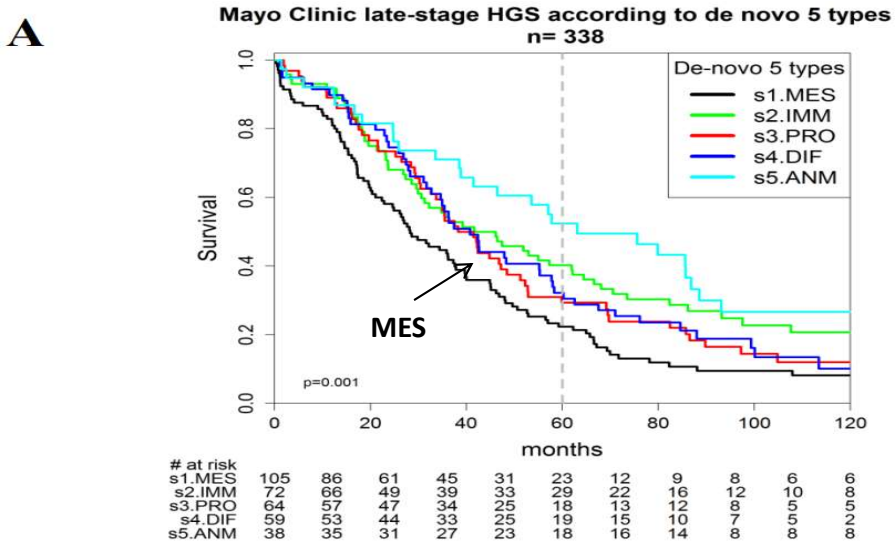


Survival associations of HGSOC subtypes

- Significant associations with survival



JNCI 2014, G Konecny, **C Wang,** et al.
Mayo Clinic HGSOC, n = 174

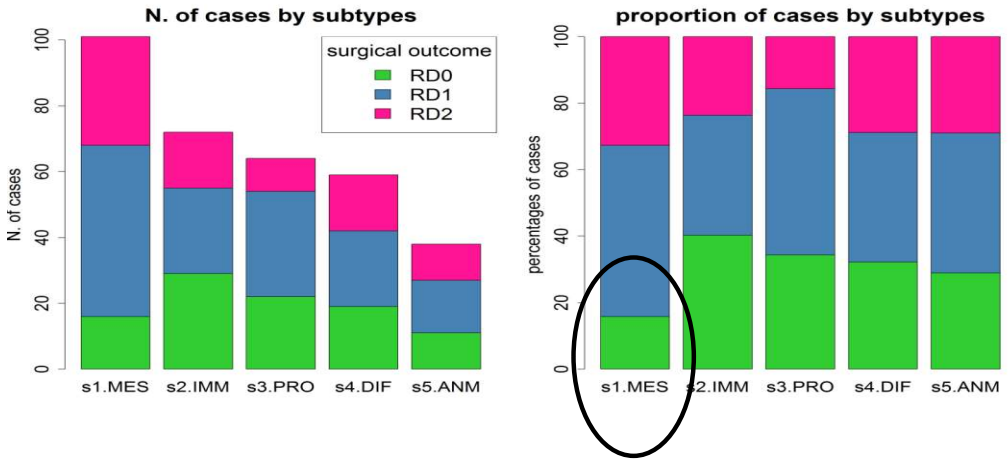


CCR 2017, **C Wang**, SM Armasu, et al.
Mayo Clinic HGSOC, n = 338

Surgical associations of HGSOC subtypes

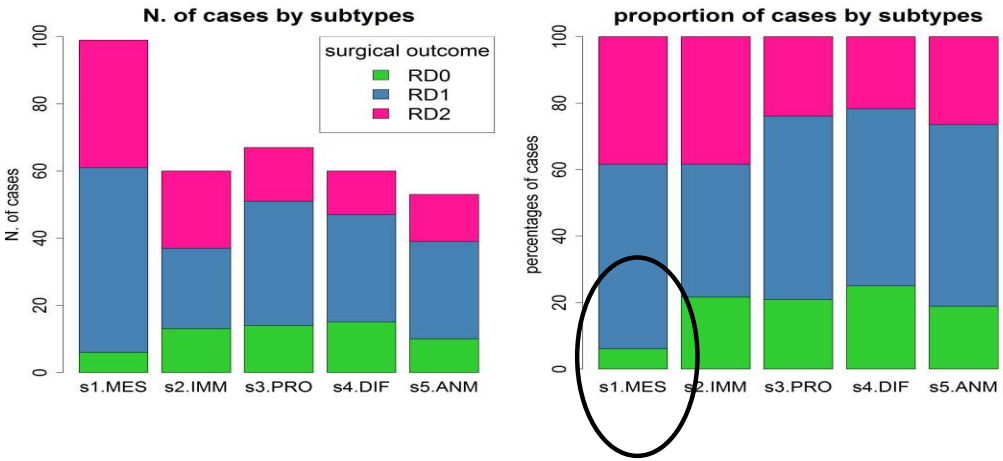
s1.MES subtype RD0 rate ~= 15%
Other subtypes' avg RD0 rate ~= 30%

Mayo Clinic stage IIIC/IV HGSOC



s1.MES subtype RD0 rate ~= 12%
Other subtypes' avg RD0 rate > 20%

TCGA stage IIIC/IV HGSOC

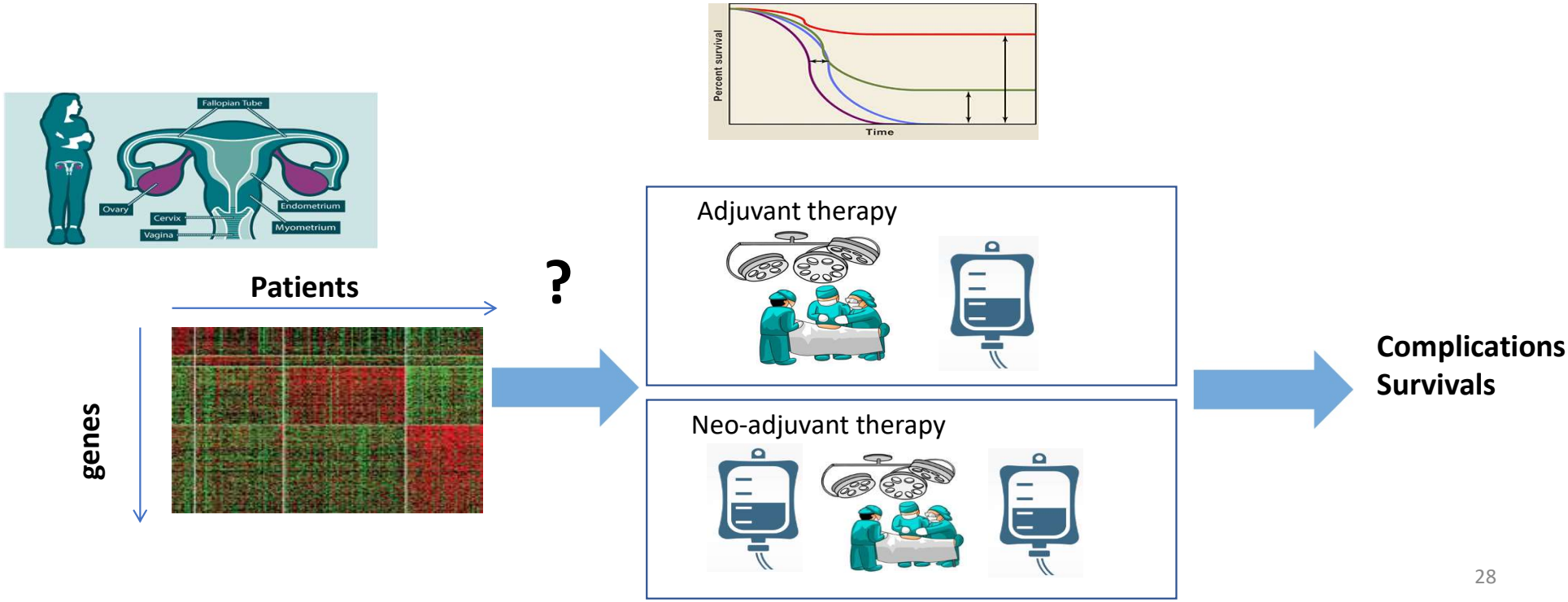


Mesenchymal subtype → √ → Increased residual disease → √ → Shorter survival

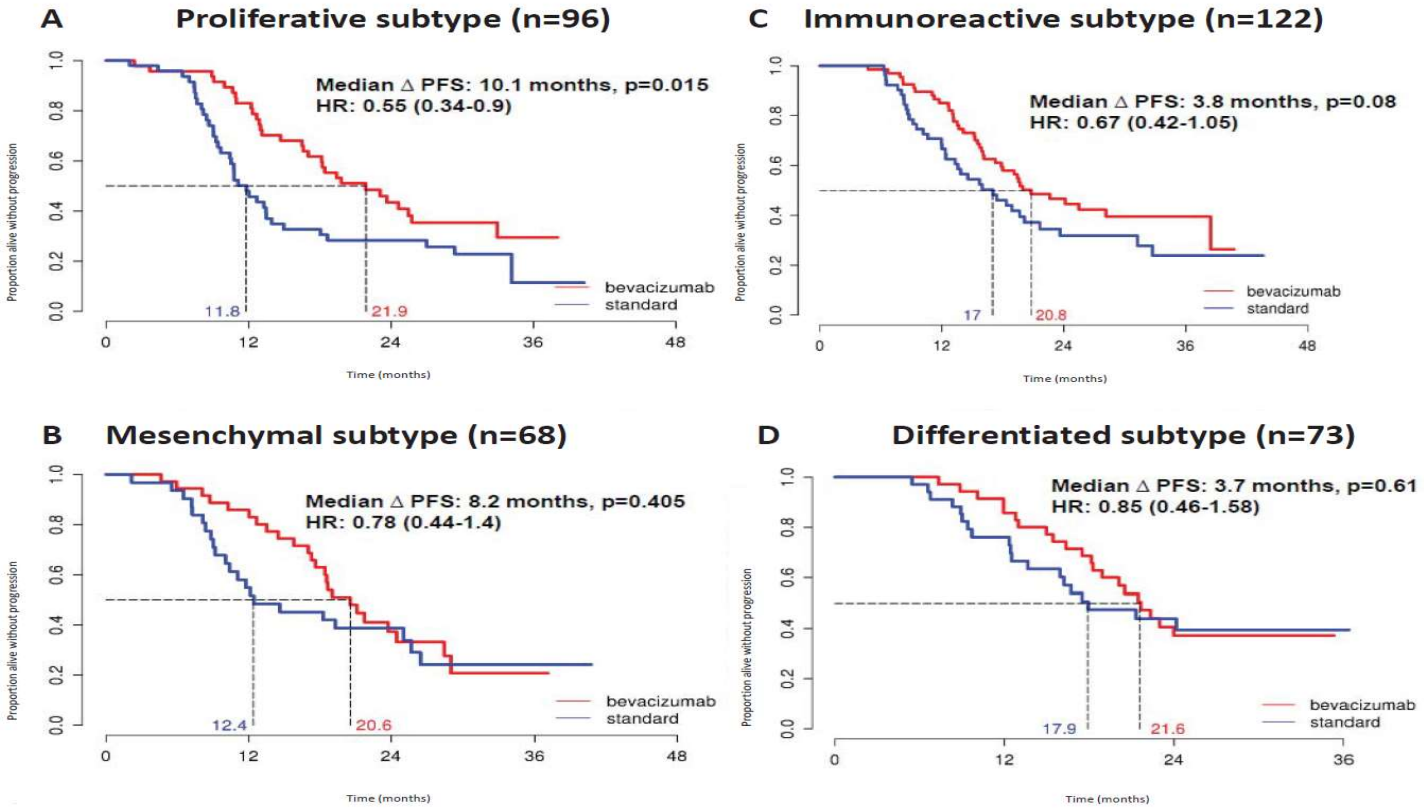
√

The clinical **associations** we learnt so far (MES):

- MES subtype is associated with worst survival
- MES subtype is associated with lowest RD0%
- MES subtype is associated with higher disease burden (upper abdominal or miliary diseases)
- MES subtype is also associated with significantly increased grade-3/-4 complications



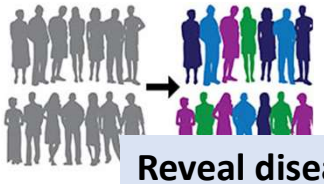
HGSOC Molecular subtypes are also associated with PFS in a retrospective analysis of phase-III trial



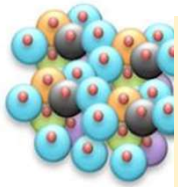
Bevacizumab may differentially improve ovarian cancer outcome in patients with proliferative and mesenchymal molecular subtypes

Stefan Kommoss, Boris Winterhoff, Ann Oberg, Gottfried E. Konecny, Chen Wang, Shaun M Riska, Jian-Bing Fan, Matthew J. Maurer, Craig April, Viji Shridhar, Friedrich Kommoss, Andreas du Bois, Felix Hilpert, Sven Mahner, Klaus Baumann, Willibald Schroeder, Alexander Burges, Ulrich Canzler, Jeremy Chien, Andrew C Embleton, Mahesh Parmar, Richard Kaplan, Timothy Perren, Lynn C. Hartmann Ellen L. Goode, Sean C. Dowdy, and Jacobus Pfisterer

CCR 2017

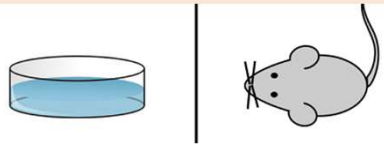


Reveal disease subtypes
- unsupervised
- semi-supervised
- subtypes w. outcomes



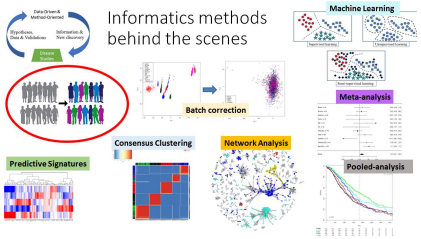
Decipher cellular heterogeneity
- cellular sub-population (scRNAseq)
- tumor-stroma interactions
- cell-cell cross-talk

Build links to model systems
- cell-line
- organoid
- PDX

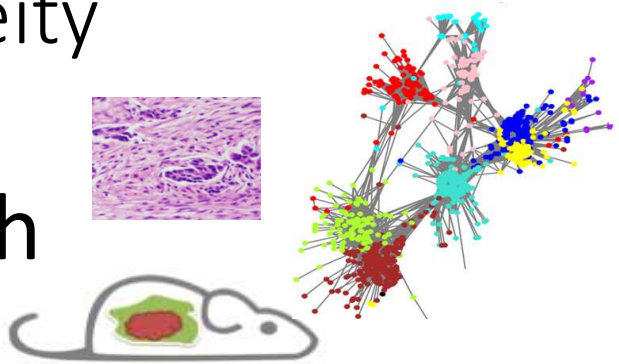


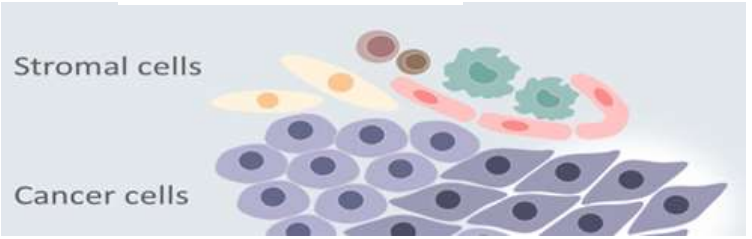
4. Go Deep again: tissue and microenvironment heterogeneity

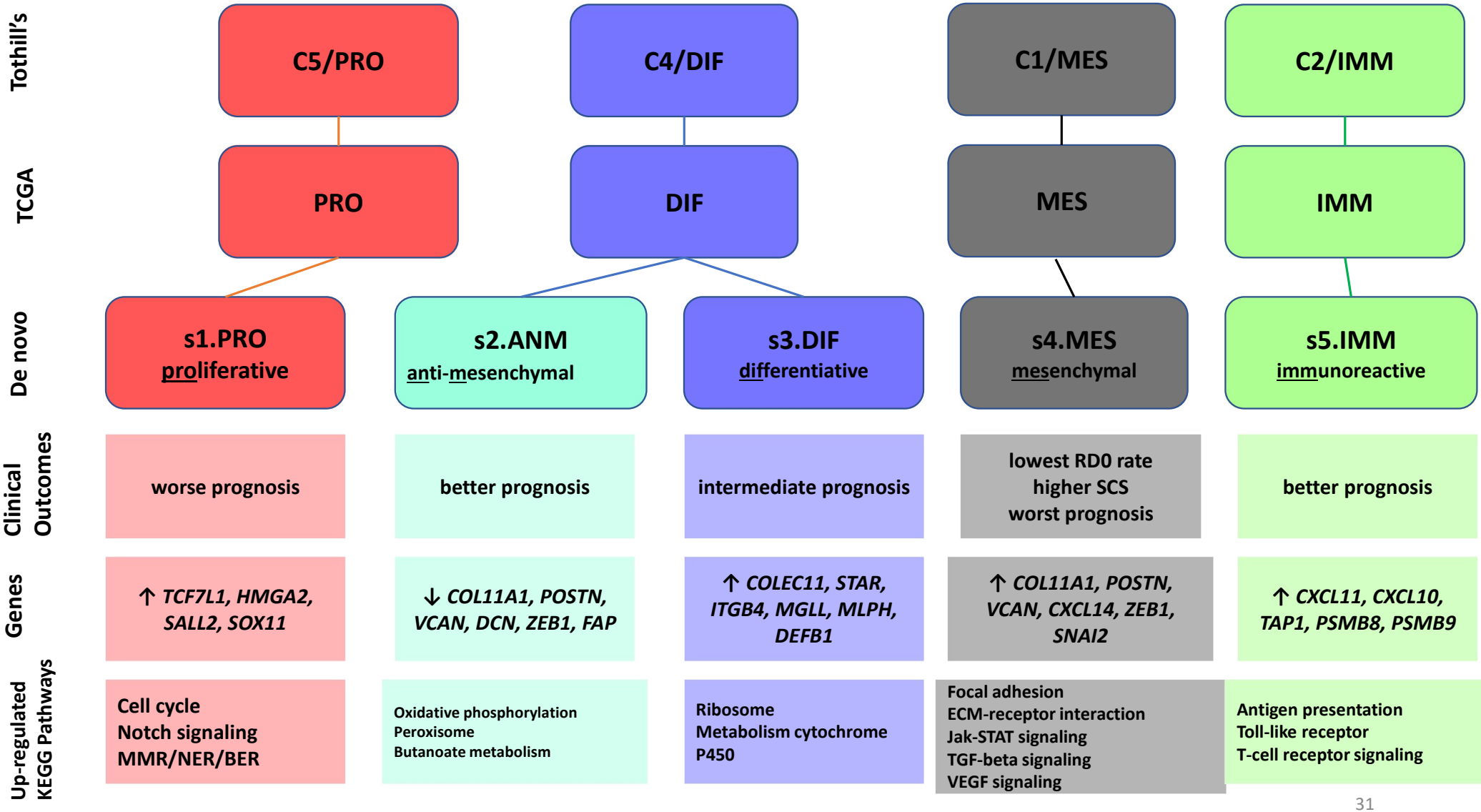
Clinic



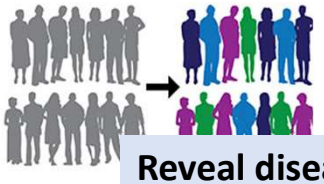
Research



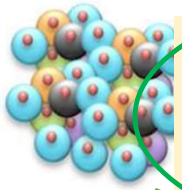




Schematic summaries of de novo subtypes w.r.t. previous subtype systems and associated changes.



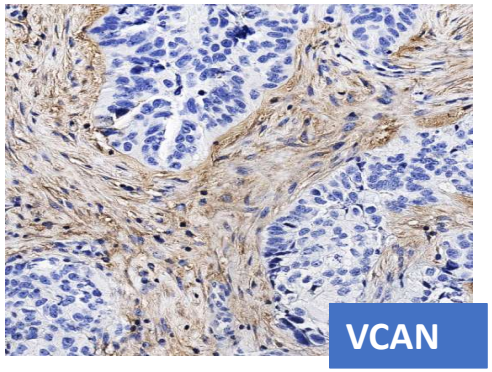
Reveal disease subtypes
unsupervised
semi-supervised
subtypes w. outcomes

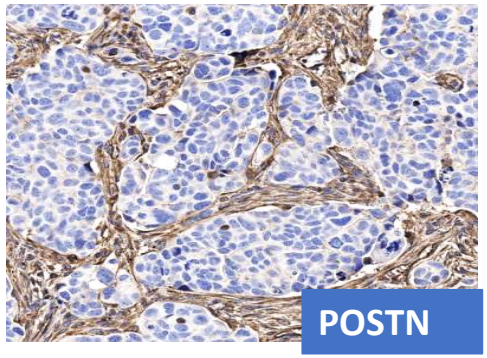


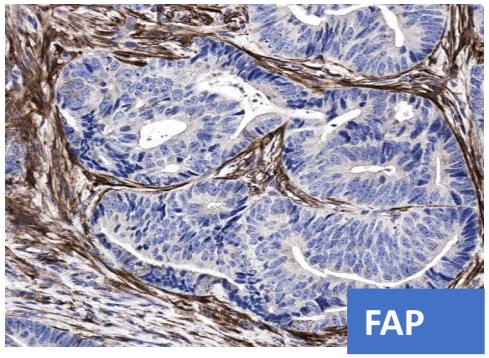
Decipher cellular heterogeneity
cellular sub-population (scRNAseq)
tumor-stroma interactions
cell-cell cross-talk

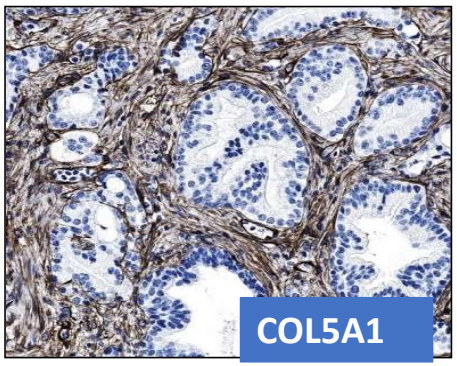
Q Zhang, **C Wang**, WA Cliby
"Cancer-associated stroma significantly contributes to the mesenchymal subtype signature of serous ovarian cancer"
Gynecologic Oncology 2019

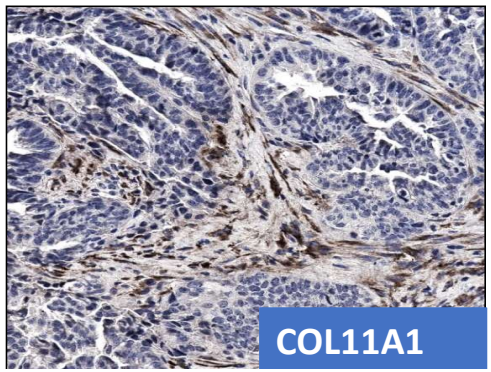
The leading MES-subtype genes are almost exclusively expressed in stroma

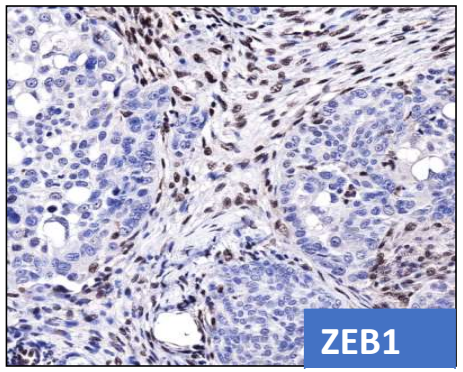








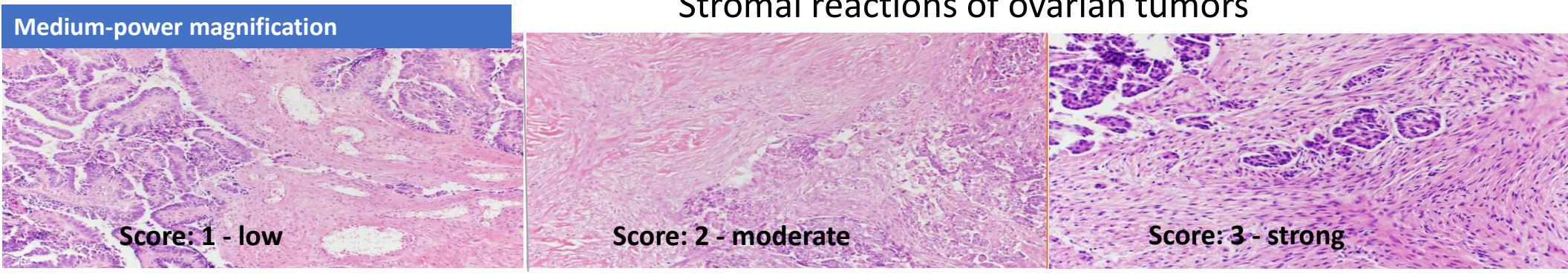


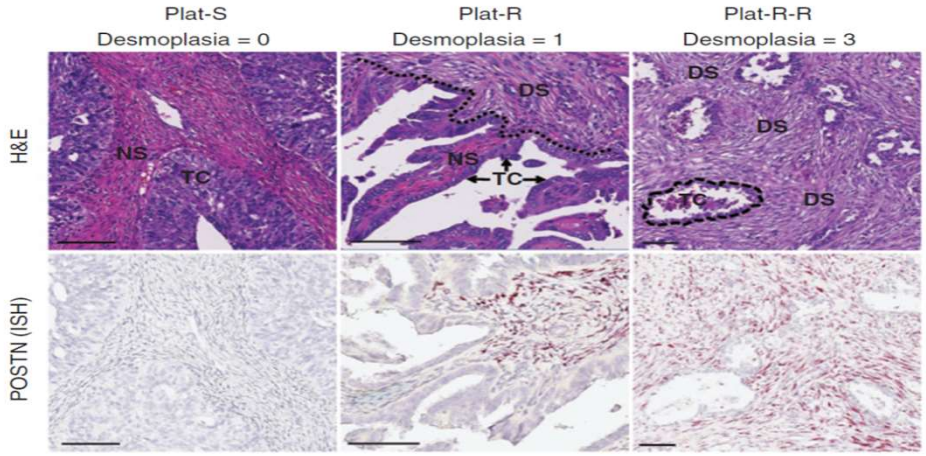


Ongoing works for characterizing tumor-stroma interactions for mesenchymal (MES) subtype tumors

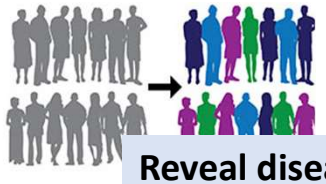
- MES subtype tumor is known of lower purity, infiltrating stroma, and desmoplasia features.

Stromal reactions of ovarian tumors

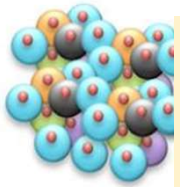




Upregulation of Periostin and Reactive Stroma Is Associated with Primary Chemoresistance and Predicts Clinical Outcomes in Epithelial Ovarian Cancer, CCR 2015

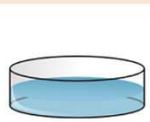


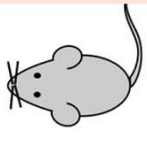
Reveal disease subtypes
- unsupervised
- semi-supervised
- subtypes w. outcomes

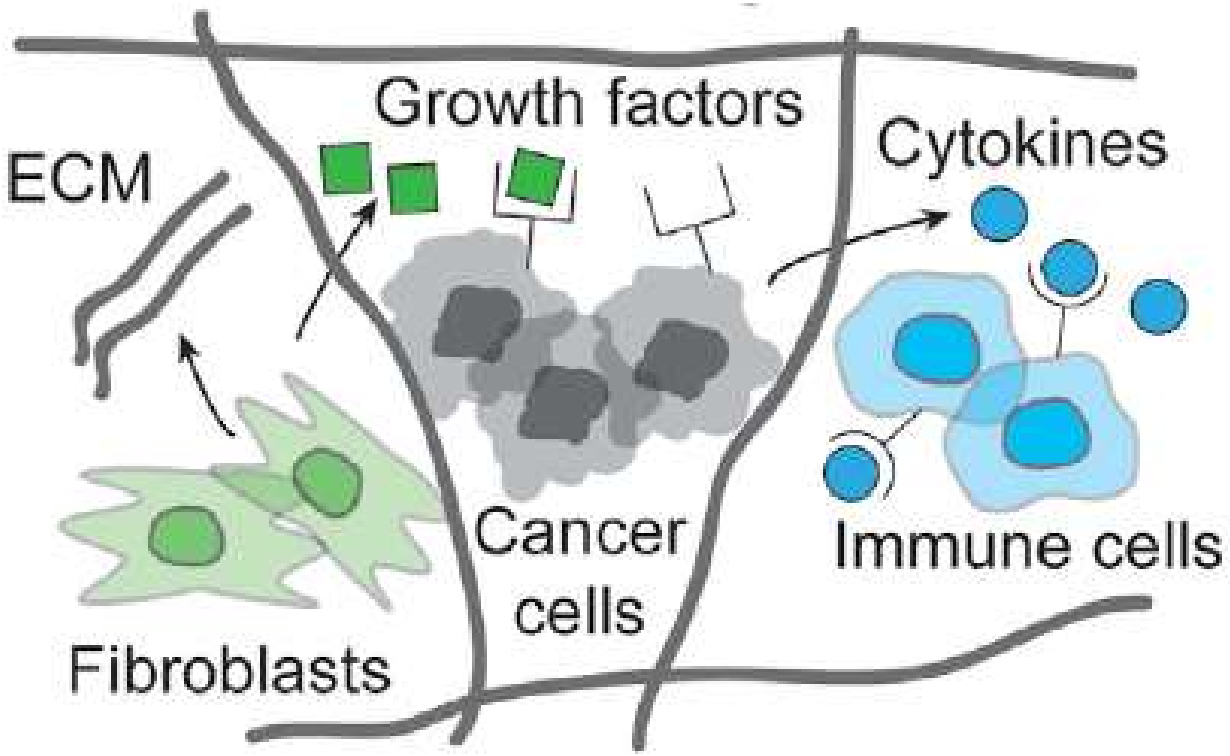


Decipher cellular heterogeneity
- cellular sub-population (scRNAseq)
- tumor-stroma interactions
- cell-cell cross-talk

Build links to model systems
- cell-line
- organoid
- PDX

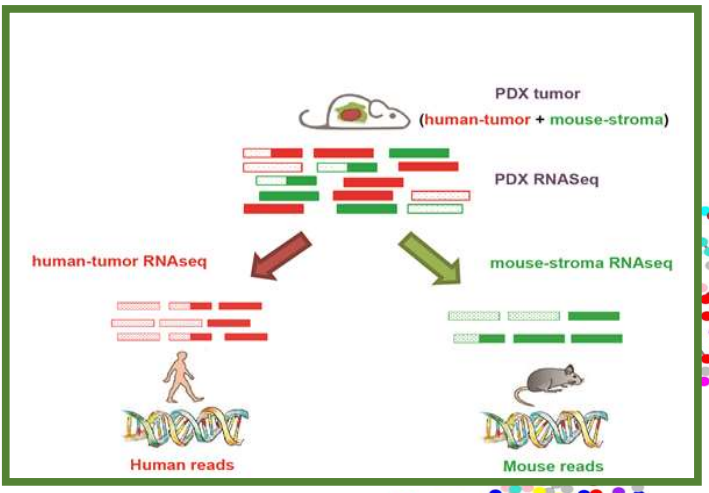






part of graphic abstract from https://www.cell.com/cell-reports/pdf/S2211-1247(18)31636-X.pdf

Gene co-expression networks (PDX stroma)

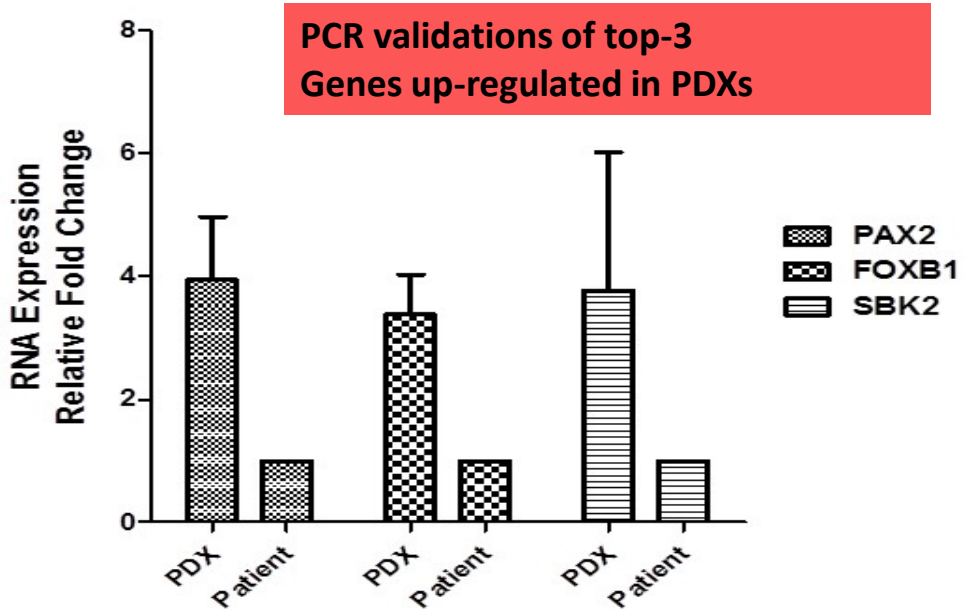


MES-subtype genes are largely stroma-specific

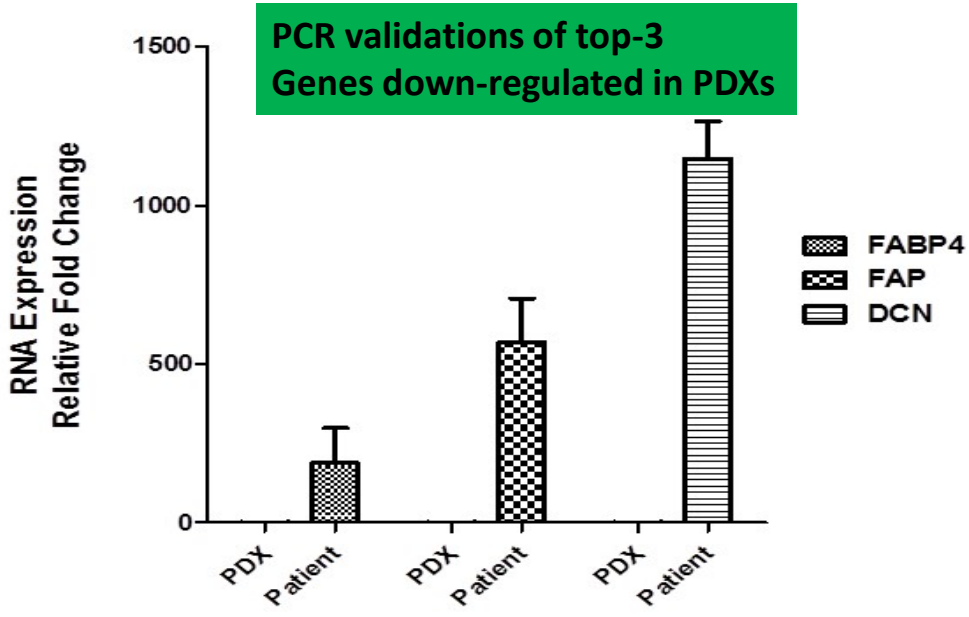
~200 genes
Up-regulated in PDX versus donor tumors

~1700 genes
down-regulated in PDX versus donor tumors

PCR validations of top-3
Genes up-regulated in PDXs

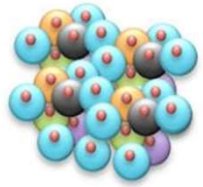


PCR validations of top-3
Genes down-regulated in PDXs

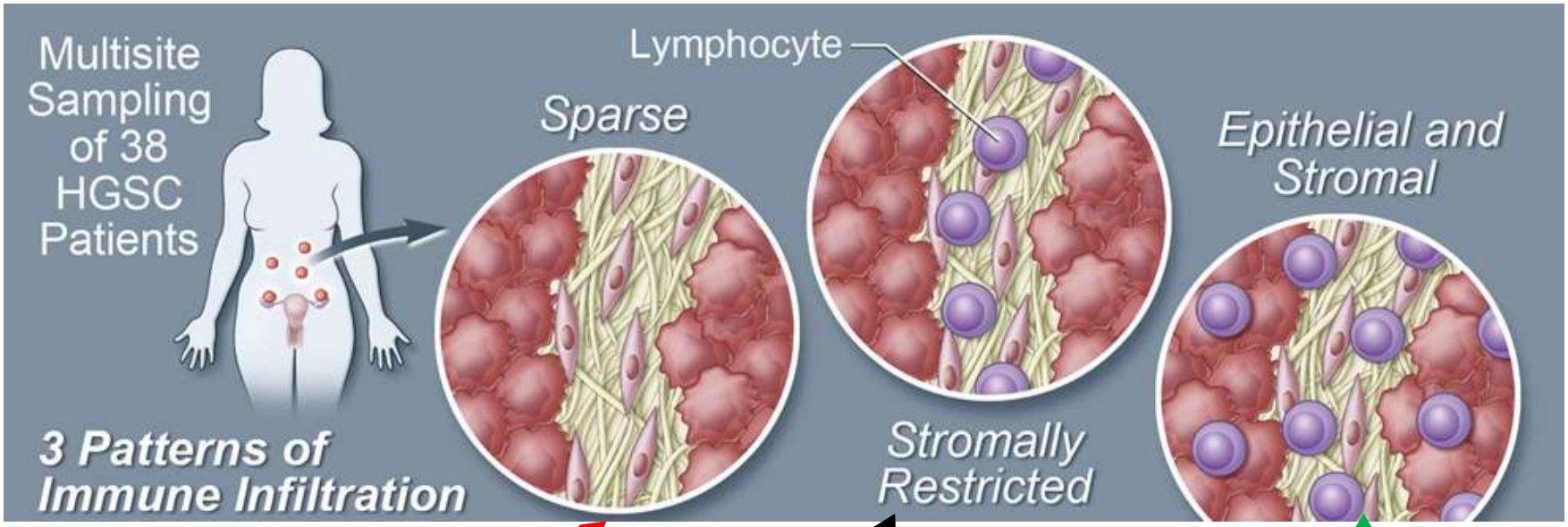


"Gene expression differences between matched pairs of ovarian cancer patient tumors and patient-derived xenografts". Liu Y, Chanana P, Davila JI, Hou X, Zanfagnin V, McGehee CD, Goode EL, Polley EC, Haluska P, Weroha, SJ, **Wang C**. Sci Rep. 2019

Microenvironment and immunology



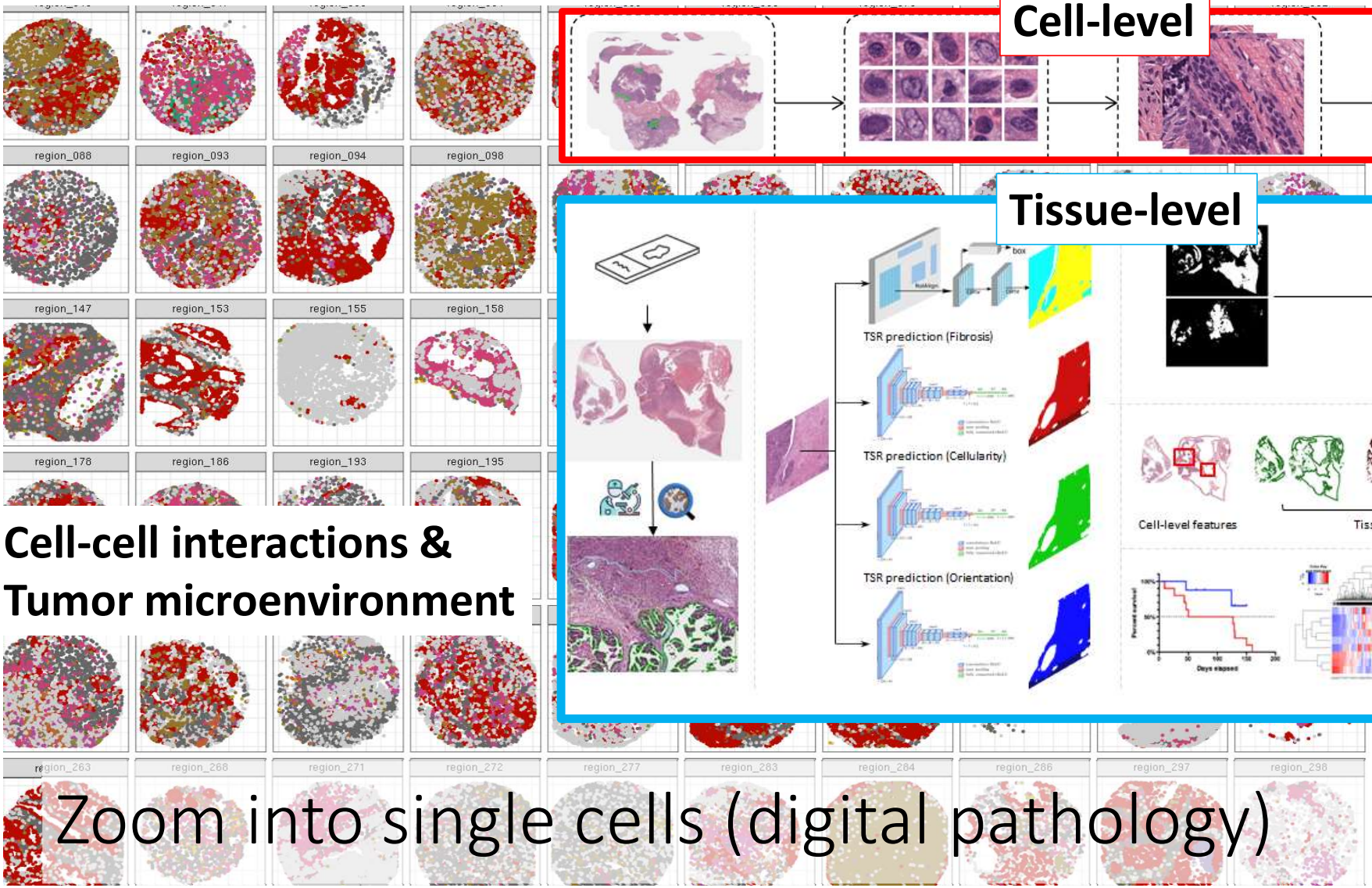
"Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer", Cell 2018



Enriched for Proliferative-subtype

Enriched for Immunoreactive subtype

Enriched for Mesenchymal subtype



Cell-cell interactions &
Tumor microenvironment

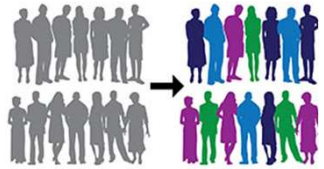
Zoom into single cells (digital pathology)

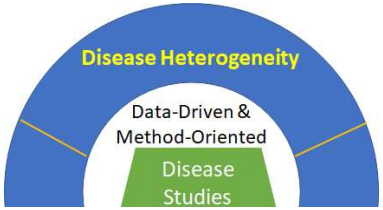
5. Data Sciences Reflections

through working examples for ovarian cancer

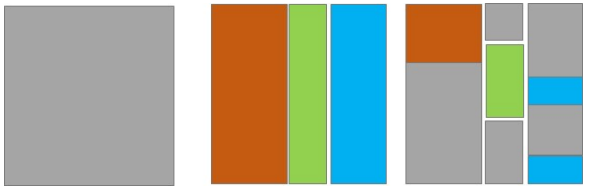
- Subtyping lead to many discoveries, and these evidences begin to converge (OC studies as working examples)
  - DNA - epigenomics - mRNA/protein - microenvironment - single-cell

- Opportunities & Challenges for Data Sciences
  - Tremendous opportunities for methodology developments
    - Heterogenous Data integration
    - Data-knowledge integration





Data-science opportunities & challenges

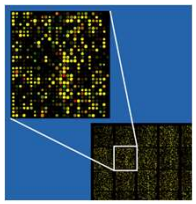


- Heterogeneity: different levels
  - **Data source & generation differences**

  Ignore or accommodate?

Sample source heterogeneity
- Collection biases and differences
- Systematic missingness and/or bias of covariates
- Privacy concerns (not completely sharing)

Feature heterogeneity
- Measurement Platform
  - mRNA
    - Microarray – 8K
    - RNAseq – 20K
    - Nanostring – 300
  - DNA: panel, WES, WGS
- Distribution difference
  - Microarray: Log-norm
  - RNAseq: Poisson



Features

...

Sample

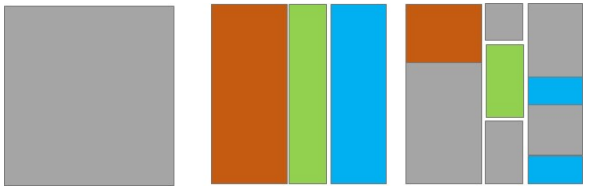
X1

X2

...

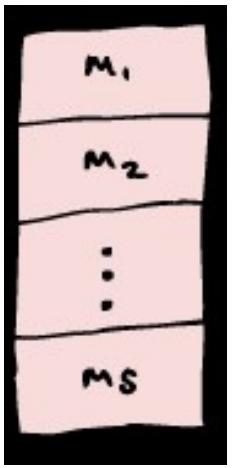
Xn

Data-science opportunities & challenges

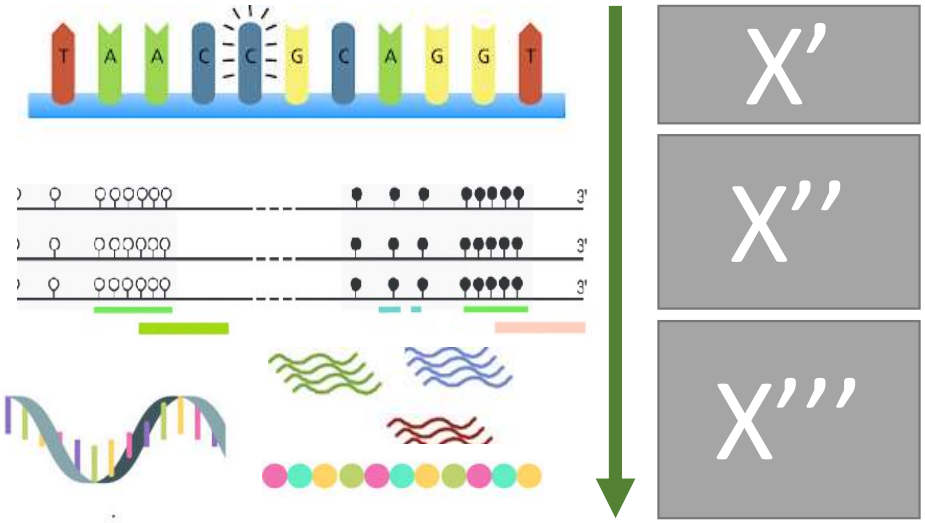


- Heterogeneity: different levels
  - Data source and/or site differences
  - **Data modalities and representations**

  Concatenate or integrate?



Features



X’

X’’

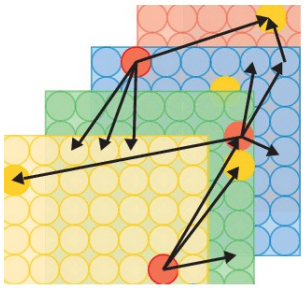
X’’’

$$K = \sum w_p K^p$$

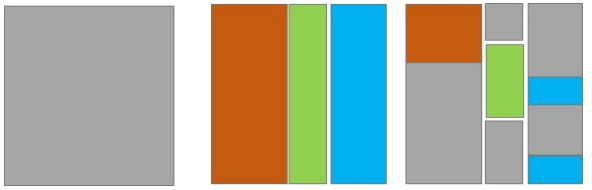
...

?

Causality Inference

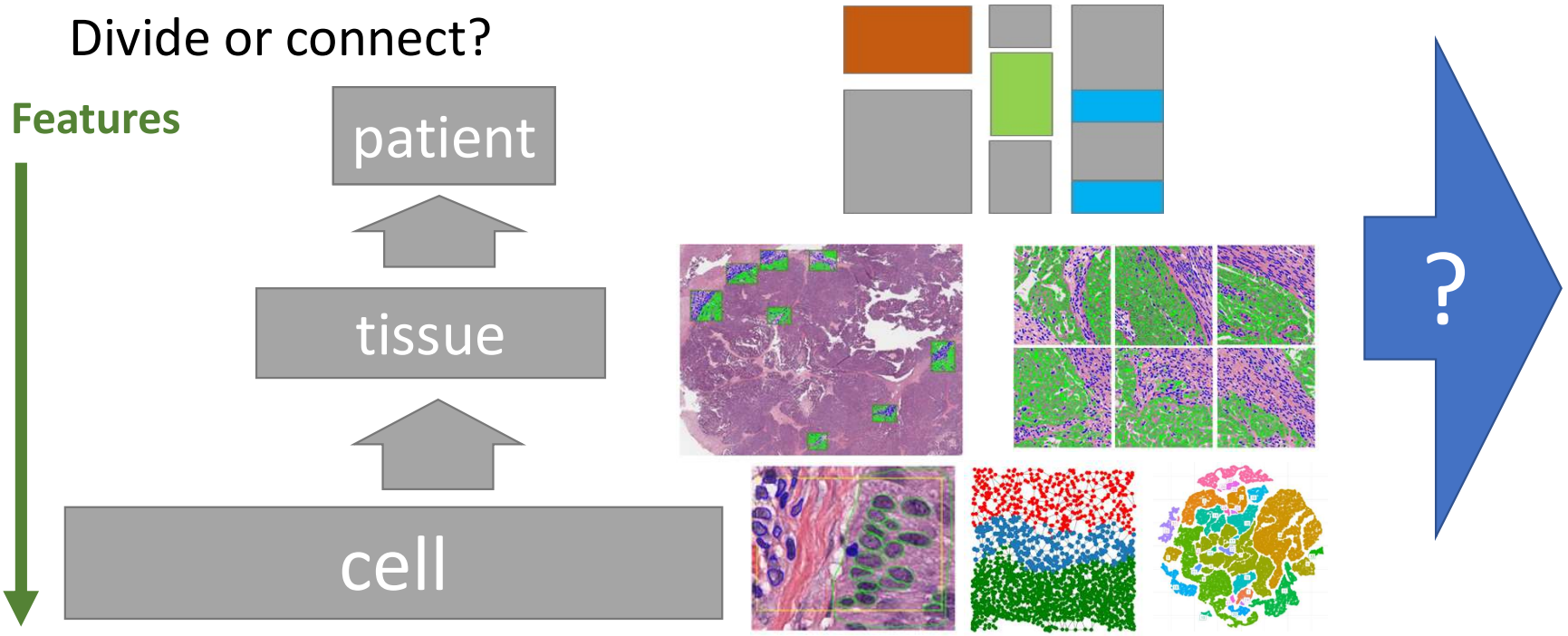


Data-science opportunities & challenges

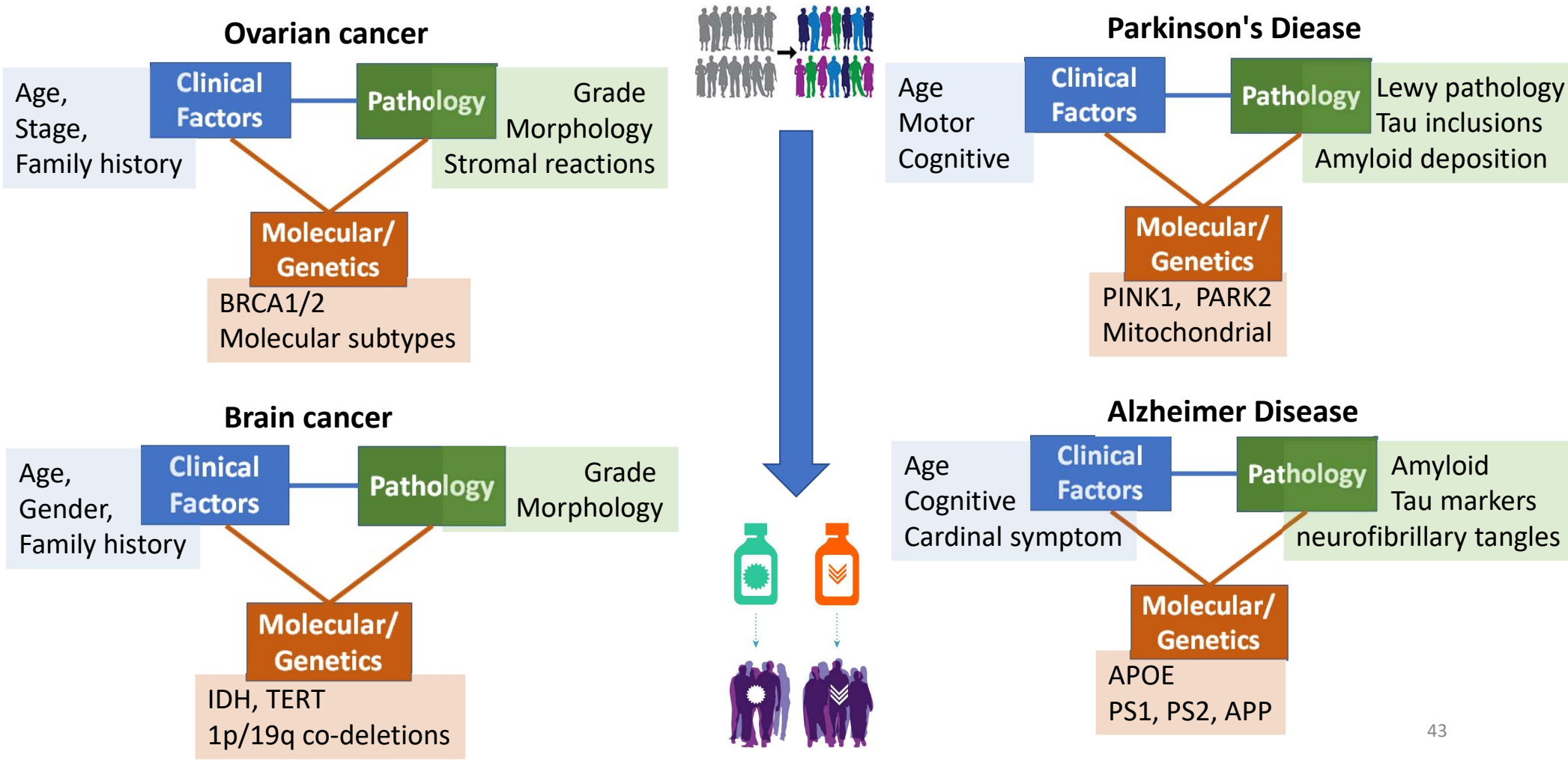


- Heterogeneity: different levels
  - Data source and/or site differences
  - Data modalities and representations
  - **Multi-resolution data**

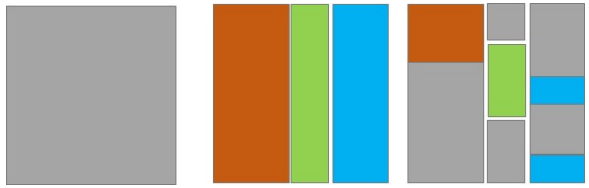
Divide or connect?

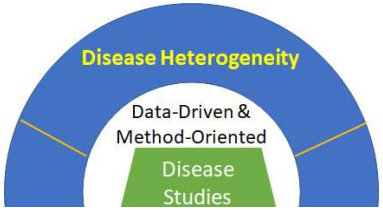


Parallels between Cancer and Neurological Disease



Questions?

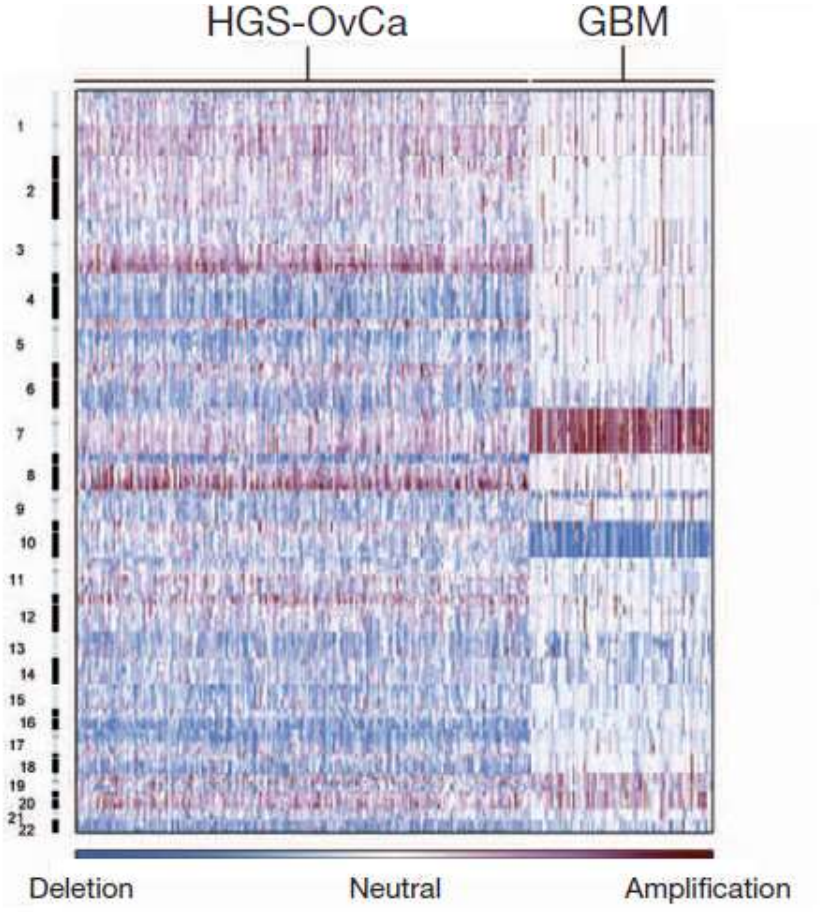



Disease Heterogeneity
Data-Driven &
Method-Oriented
Disease
Studies

Supplementary

- Genetics and Genomics heterogeneity of ovarian cancer – HRD in OC and pan-Caner studies

Genomics heterogeneity of ovarian cancer



- GBM has recurrent chr7-gain (MYC) and chr10-loss (PTEN)
- Ovarian CNV profiles are messy, yet different with DNA repair deficiency: **HRD**
  - Homologue recombination deficiency

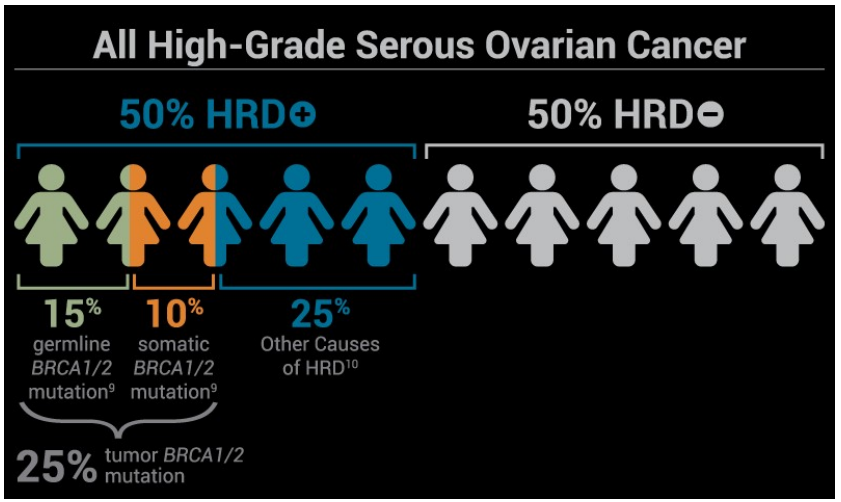
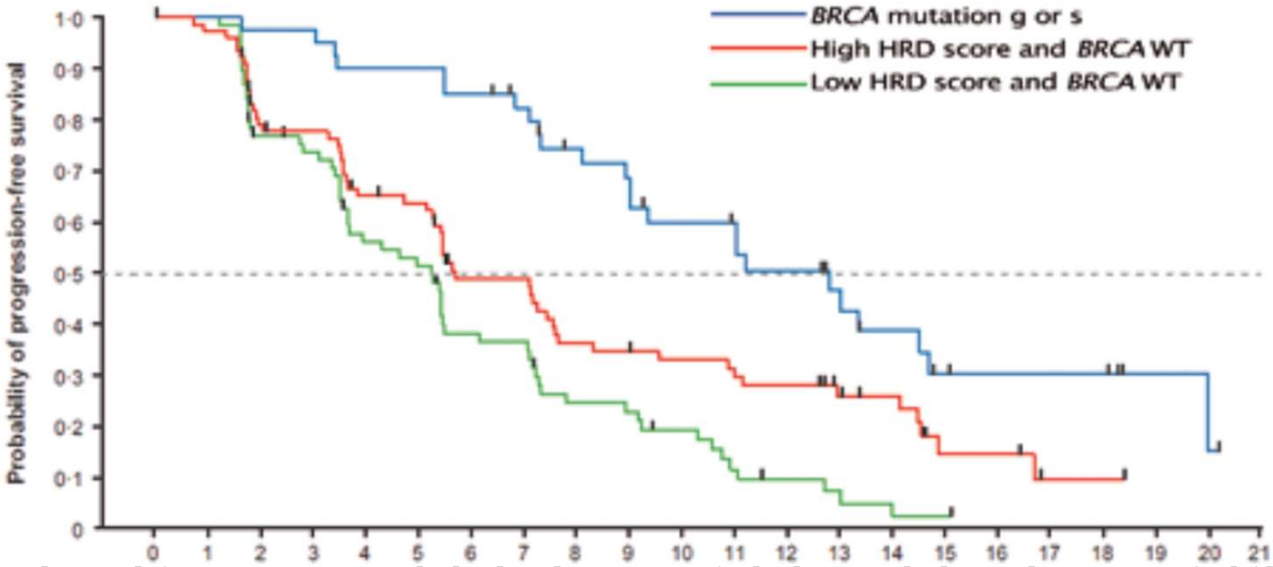


Fig. 1a, from TCGA-ovarian, Nature 2011

Clinical significances of HRD score/status

g – germline, s – somatic; HRD – homologous recombination deficiency; WT – wild type

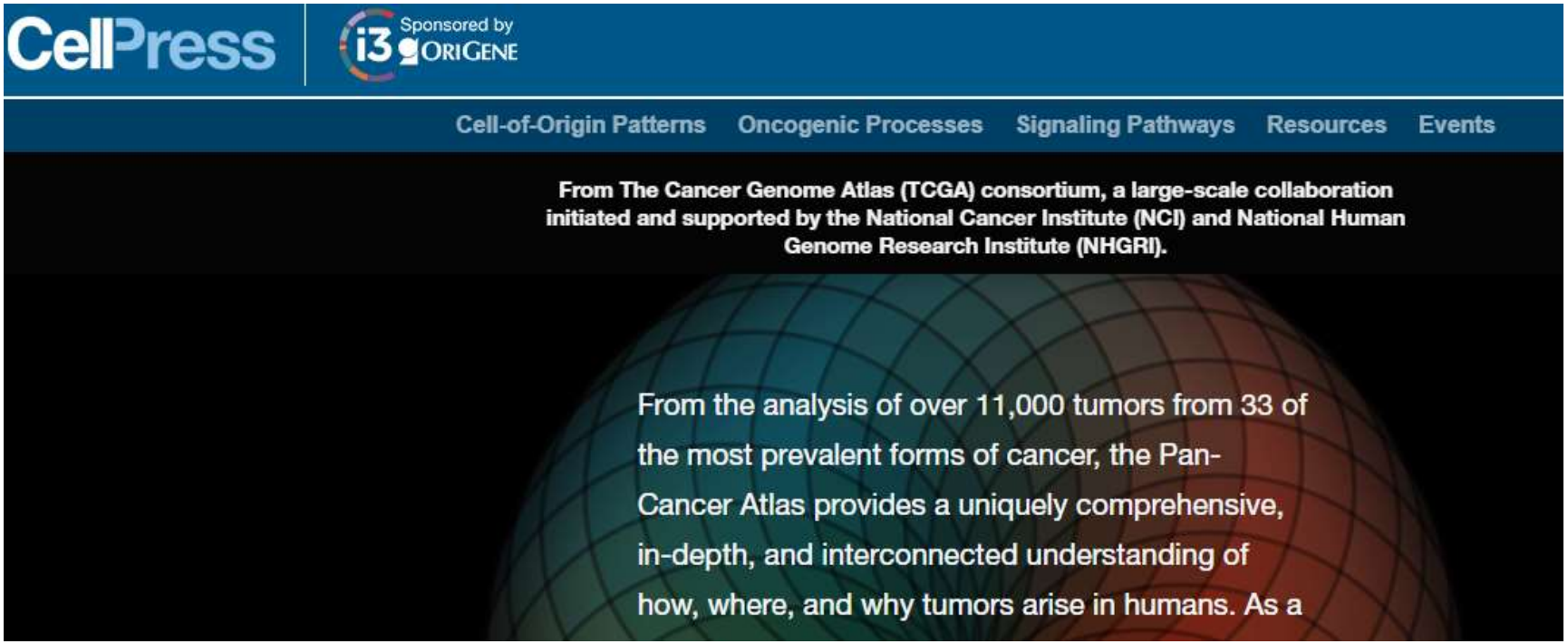
- PARPi



The ARIEL 2 international, multicentre, open-label, phase II trial showed that the PARP inhibitor rucaparib extended progression-free survival in patients with relapsed, platinum-sensitive high-grade serous ovarian cancers.

EM Swisher et al. (2017) Lancet Oncol 18:75–87.

4. Go even Broader – across all the TCGA cancers



I co-authored 6 of TCGA panCan papers published in 2018; and serve as last-author of DNA damage study

Pathogenic germline variants, Cell, 2018

Oncogenic signaling pathways, Cell, 2018

PanCan Aneuploidy, Cancer Cell 2018

PanCan Gyn Cancers, Cancer Cell 2018

Squamous carcinomas, Cell Reports 2018

DNA damage repair, Cell Reports 2018

Go even Broader – across all the TCGA cancers

(researchers from 20+ U.S. institutions)

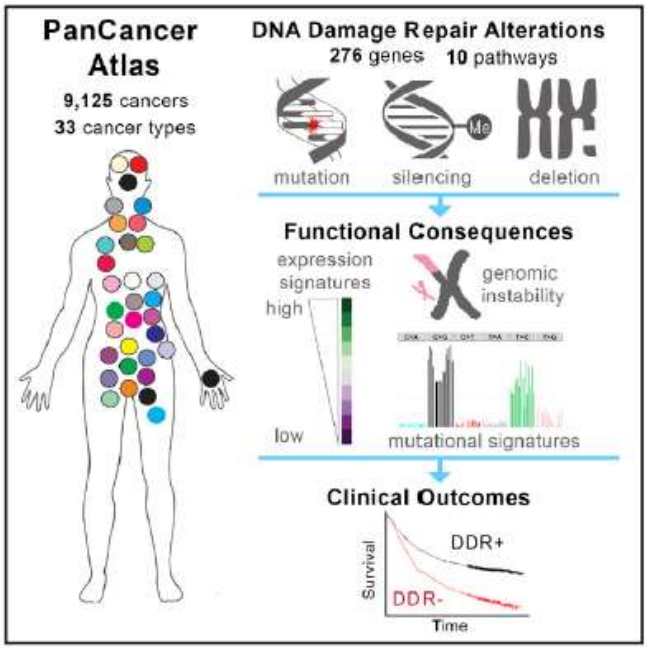
Resource

Cell Reports

Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas

Knijnenburg et al. (2018) *Cell Reports* 23:239-254

Chen Wang Yonghong Xiao Ray Monnat



Theo Knijnenburg

Nyasha Chambwe

Linghua Wang

Mike Zimmermann

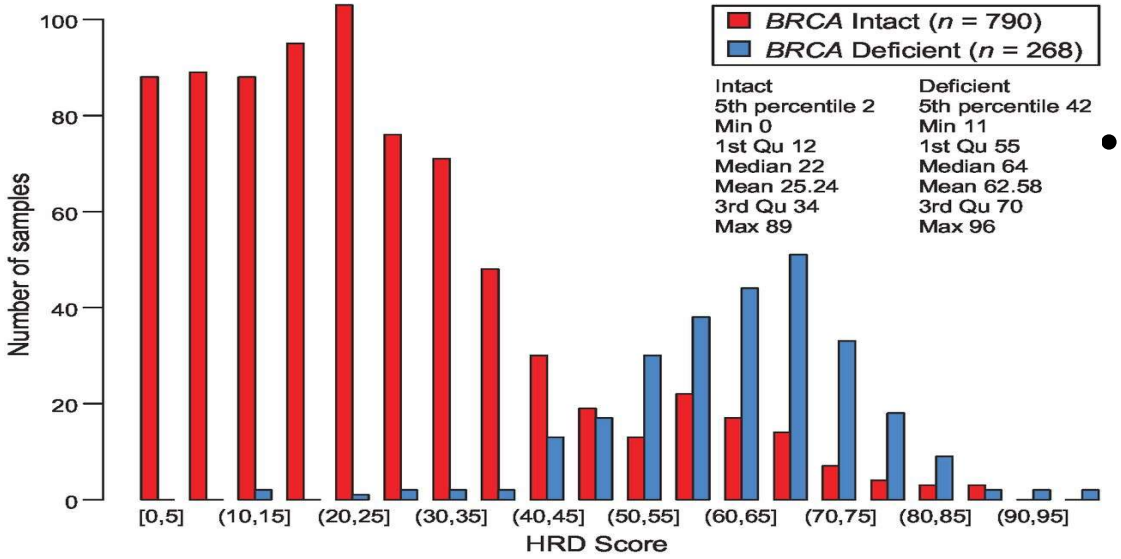
Andy Cherniak

Galen Gao

Ilya Shmulevich

CNV-burden scores defining HRD status

HR-deficiency (HRD)
HRD-score >= 42



HR-proficiency (HRP)
HRD-score < 42

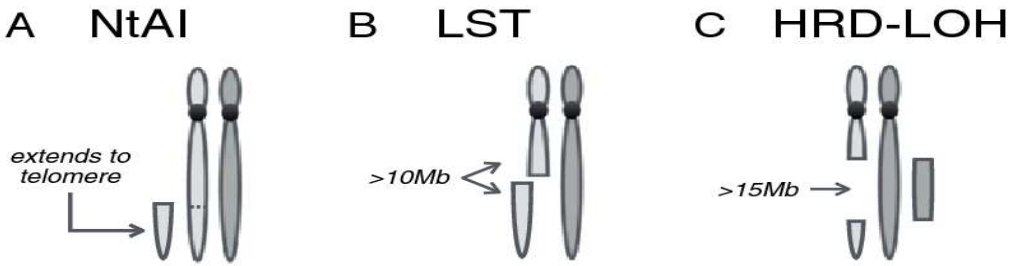
- Copy number profiling as footprints to infer BRCAness status
- We published HRD scores for all the TCGA samples in panCan DDR study:

Resource

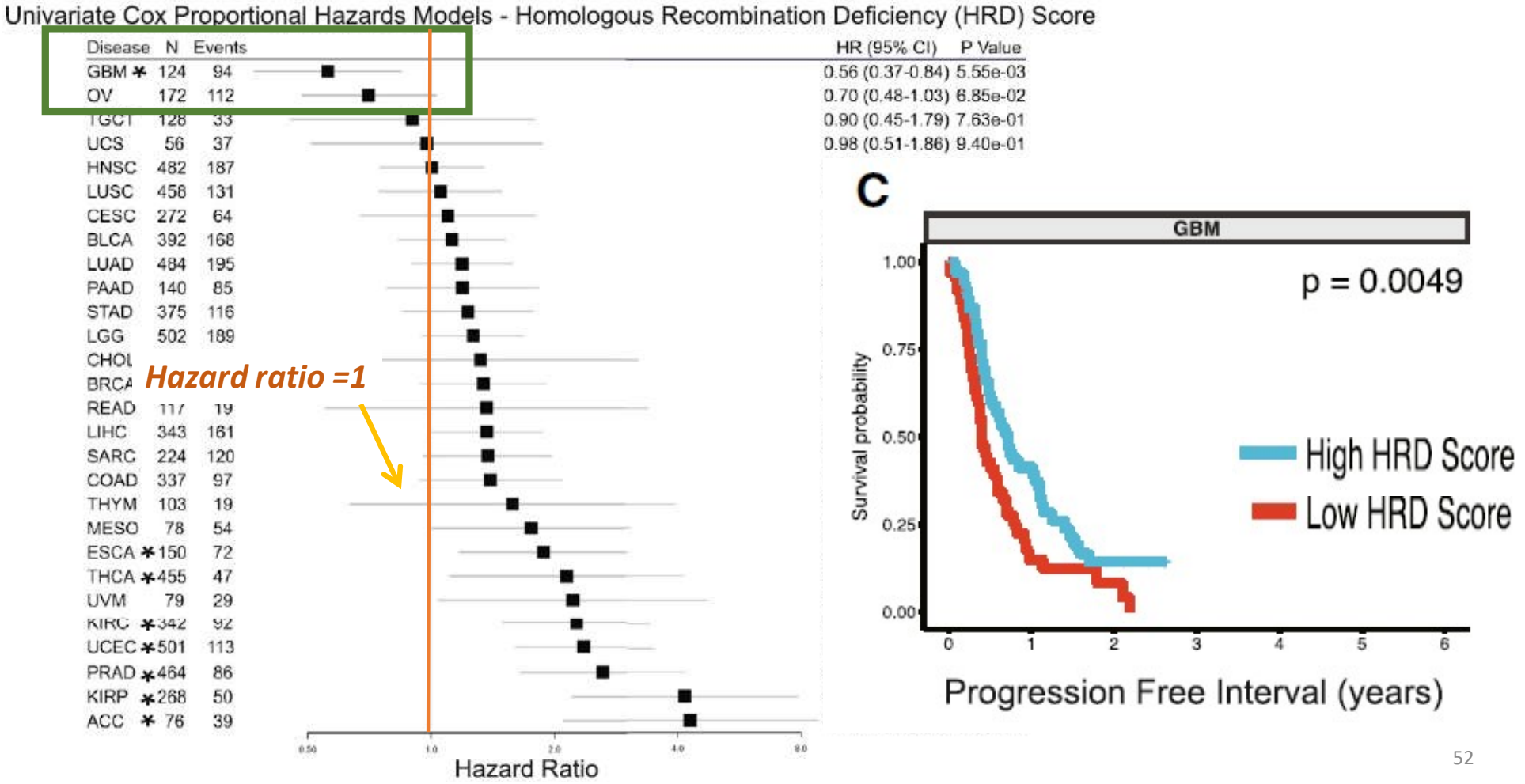
Cell Reports

Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas

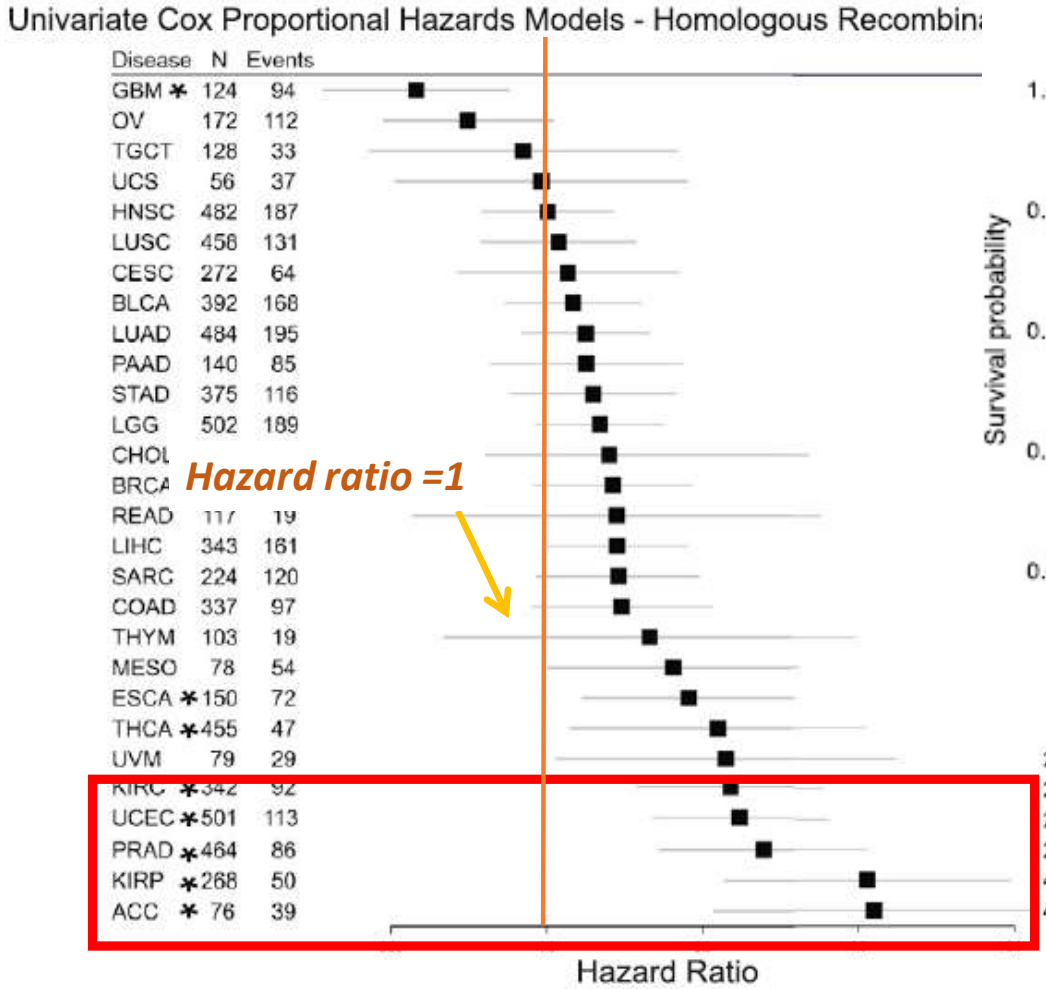
Knijnenburg et al. (2018) *Cell Reports* 23:239-254

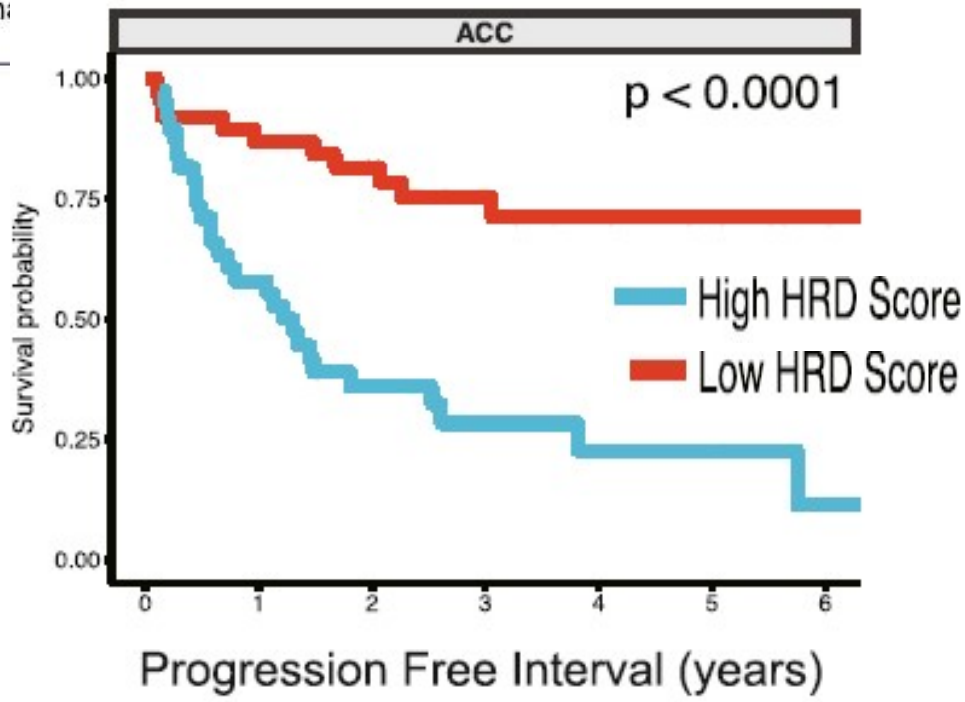


For GBM and ovarian, the higher HRD, the longer PFI



For some cancers, the higher HRD, the shorter PFI





ACC: Adrenocortical carcinoma