

# What do we learn from Pan-Cancer Subtyping?

**TCGA Symposium**  
**May 12, 2014**

***Pan-Can Integrated Subtypes AWG***

***Leads: Josh Stuart, UCSC***

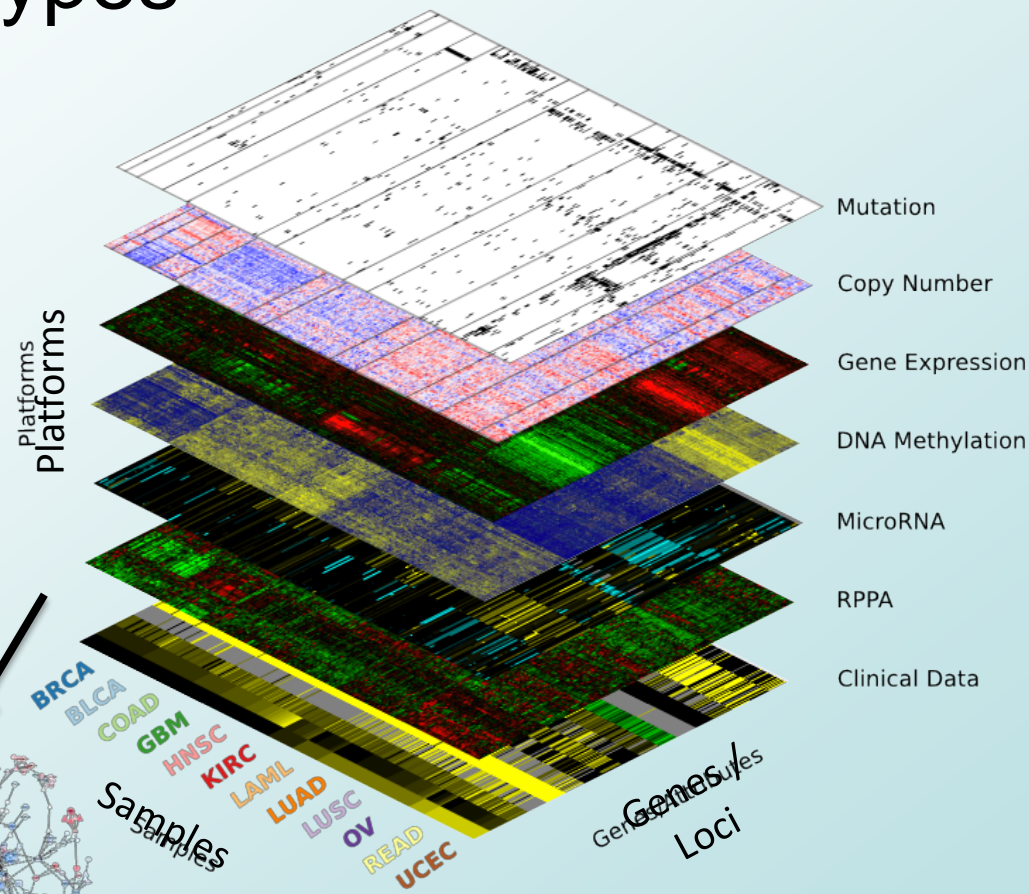
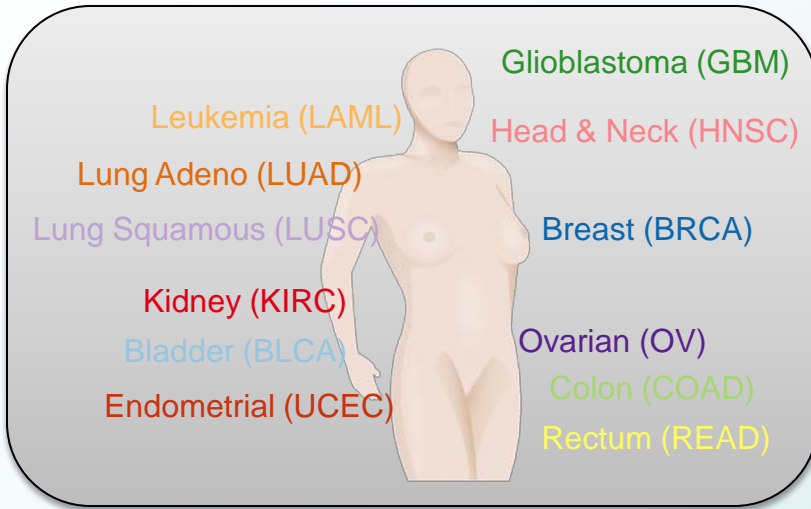
***Chris Benz, Buck***

***Chuck Perou, UNC***

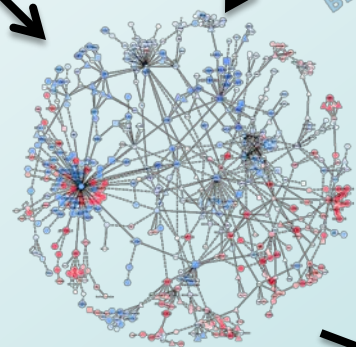
# Pan-Cancer: Integrative analysis across tumor types

12 Tumor Types

Omics Characterizations

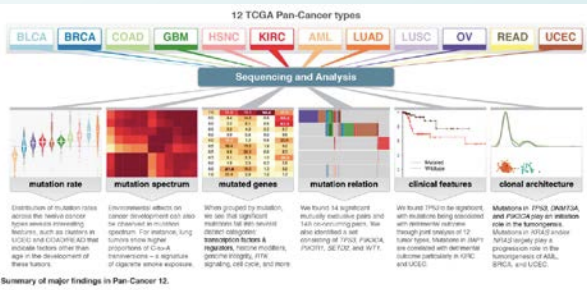


Thematic Pathways



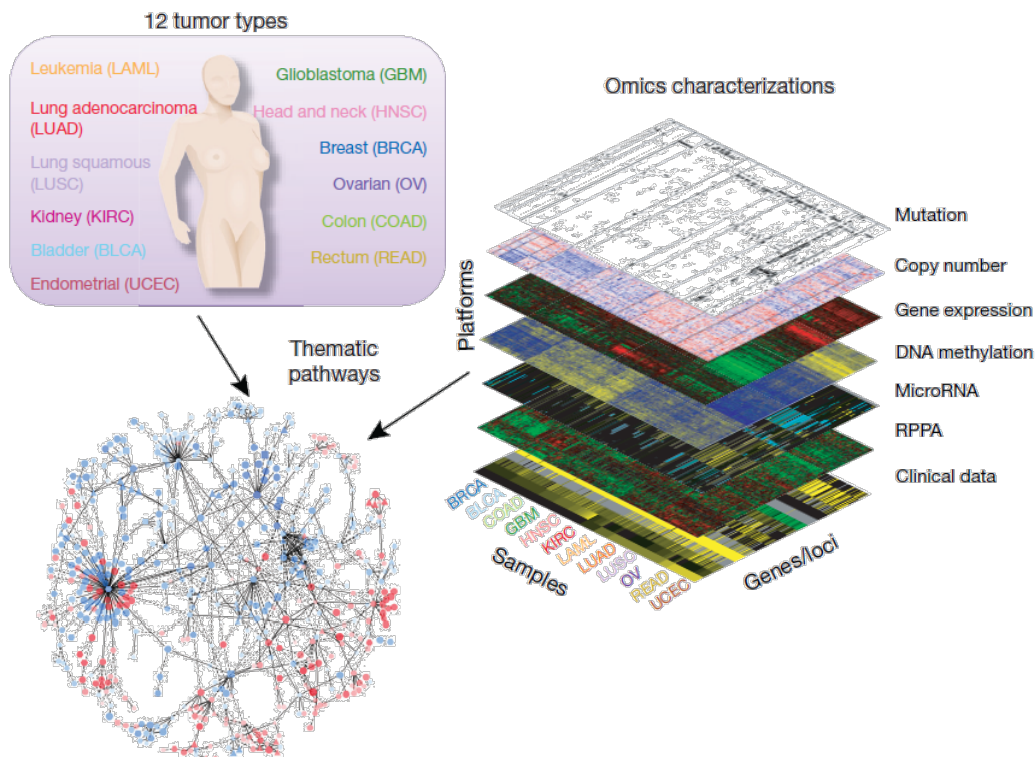
Data:

➤ Synapse - Sage Bionetworks



Findings: [nature.com/tcga](http://nature.com/tcga)

# Pan-Cancer-12 Dataset



Defined by Pan-Cancer AWG

~3500 Samples

**Tumor Type**      **# Samples**

AML	173
Bladder	122
Breast	845
Colon	190
Endometrial	370
GBM	168
Head & Neck	303
Kidney Clear Cell	480
Lung Adeno	355
Lung Squamous	259
Ovarian Serous	265
Rectum	72

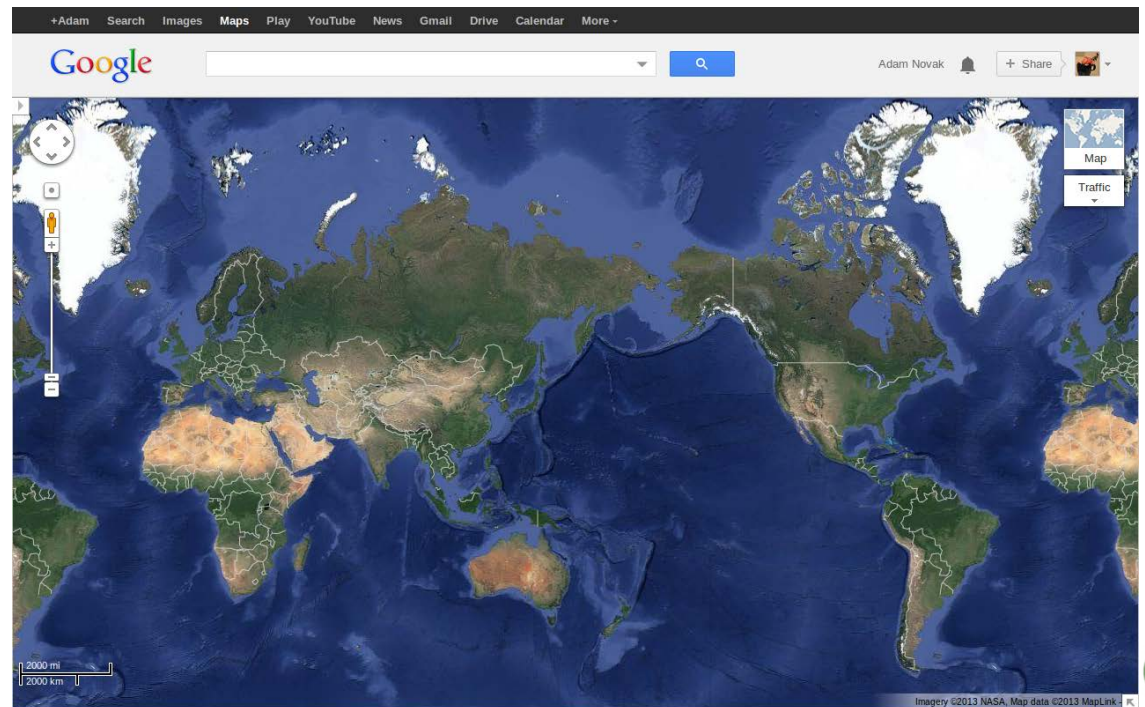


# Need a Map to navigate so much info



## Inspiration: Google Maps

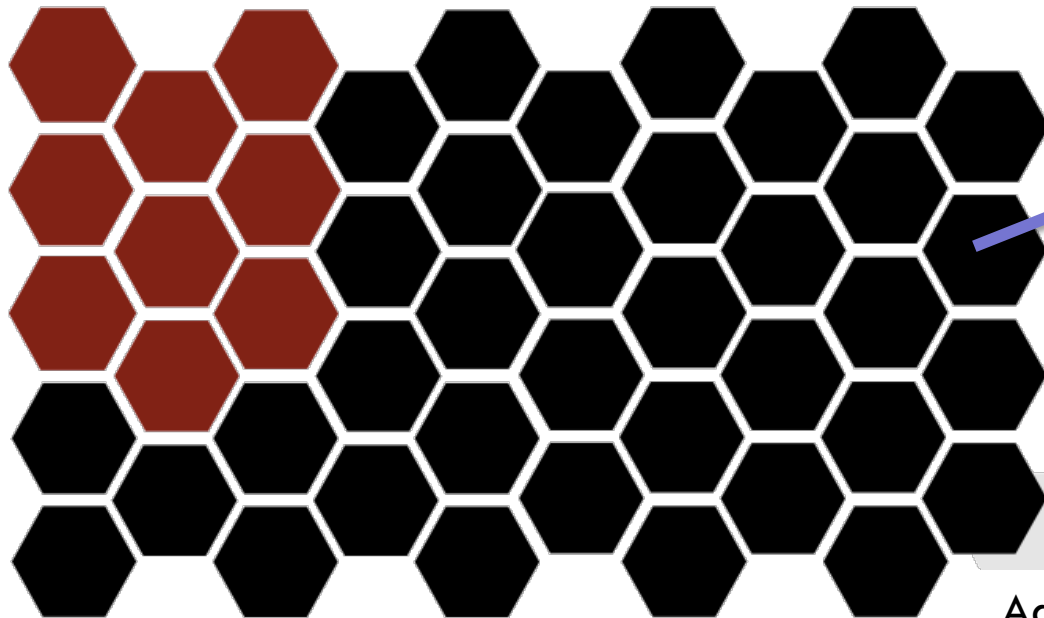
- Fixed, learnable coordinate system
- Natural human intuition
- Overlay stores, places, reviews, photos, video...
- And... Its not a heatmap!



# Each sample has its own Address

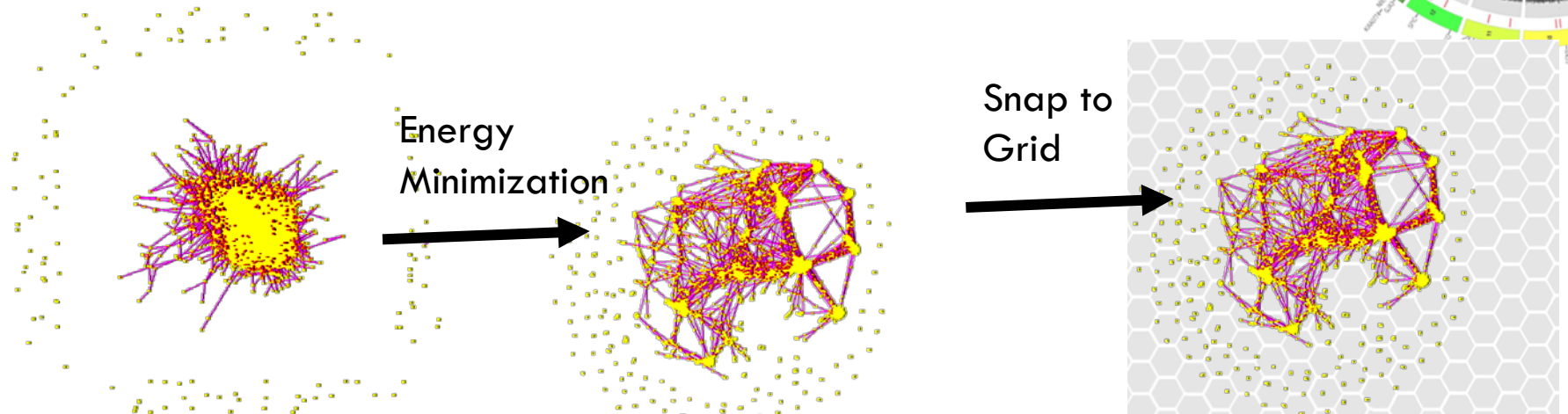


- Each sample = hexagon (the *address*)
- Hexagons good packing properties
- Colors display attributes: outcomes, mutations, etc



1 Sample  
(or signature)

# Connecting Signatures to Addresses

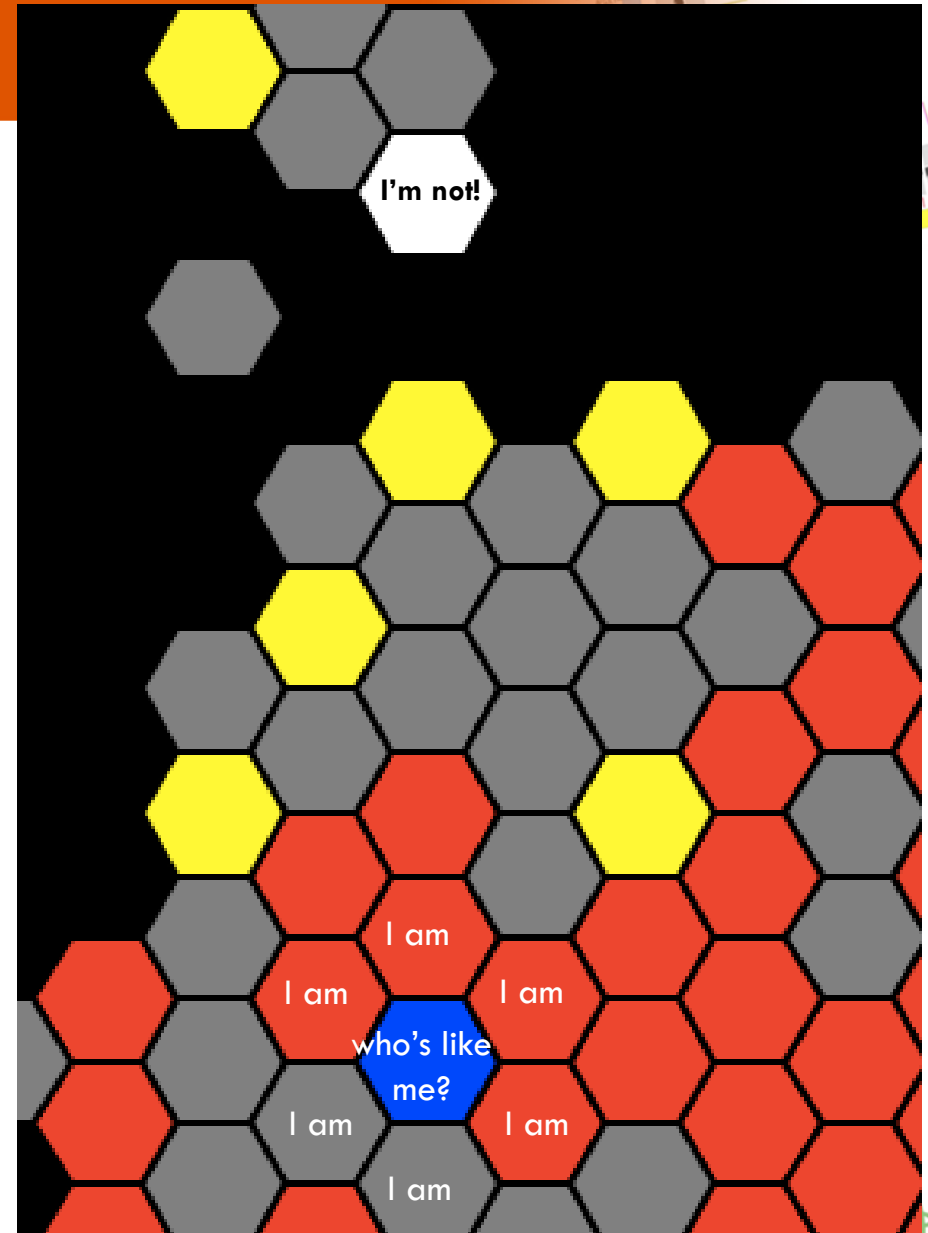


- Spring Layout: Low energy configuration of competing springs
- Snap to grid: Associate one point per hexagon.
- Samples w/ similar signatures  $\rightarrow$  same zip code.

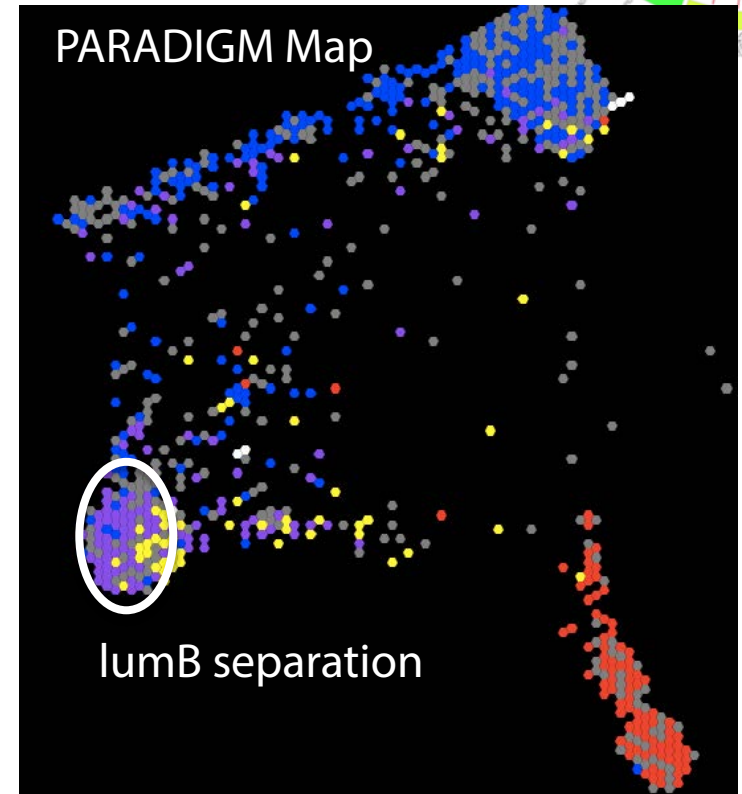
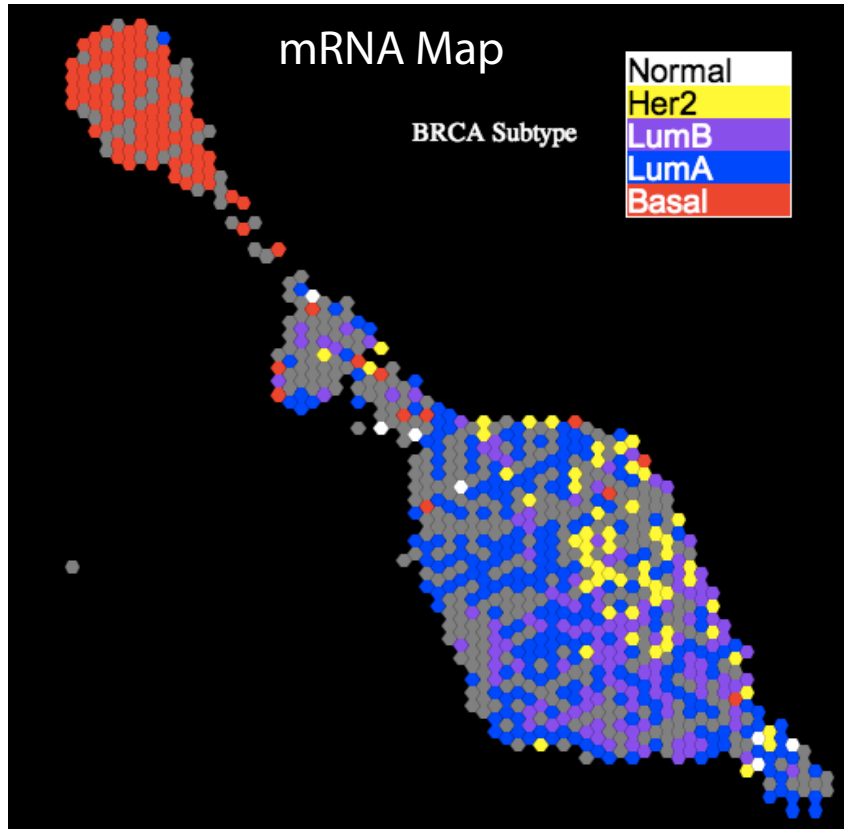
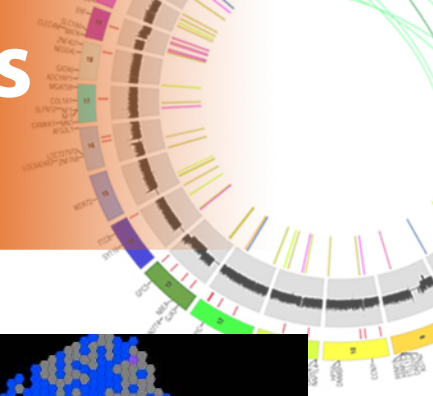
Layout Engine: DrL. Sandia National Labs

# Identify similar samples in the same zip code

- Map address reflects *molecular similarity*
- “Zip codes” carry information
  - like: Russian Hill, Berkeley, Silicon Valley, Bethesda, ...



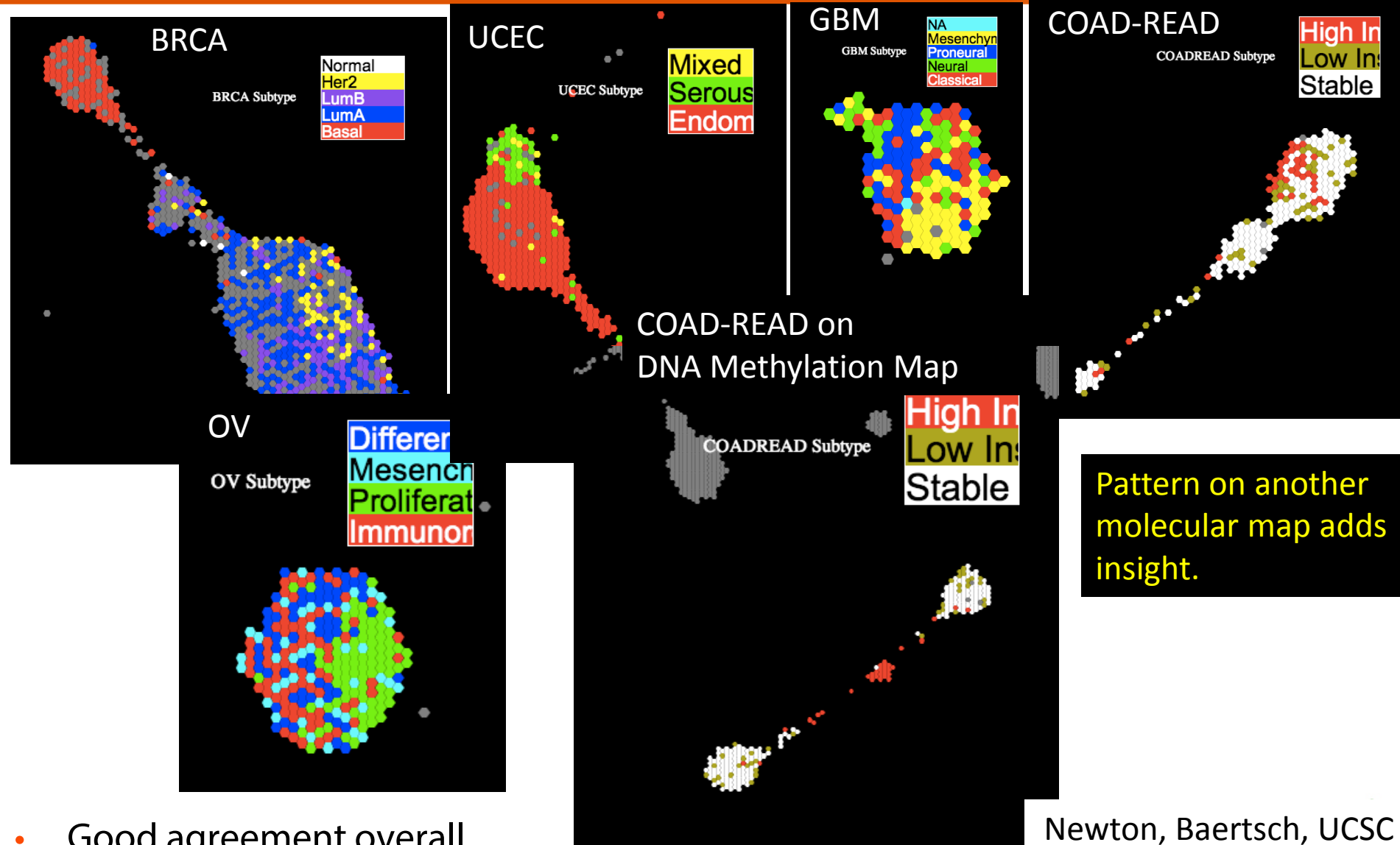
# Are disease-specific AWG subtypes recappeded in TumorMap?



- BRCA subtypes resolve clearly on mRNA and PARADIGM maps.



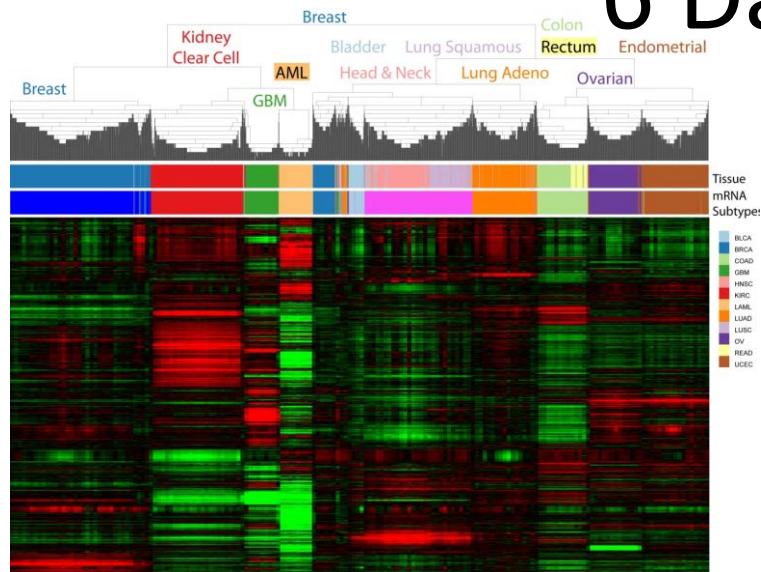
# Are disease-specific AWG subtypes recappeded in TumorMap?



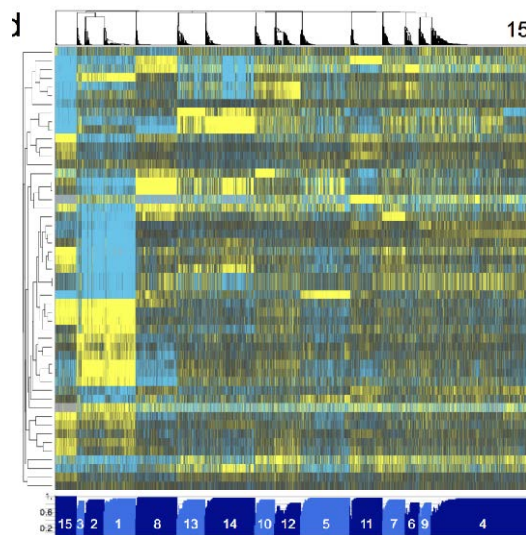
- Good agreement overall.

Newton, Baertsch, UCSC

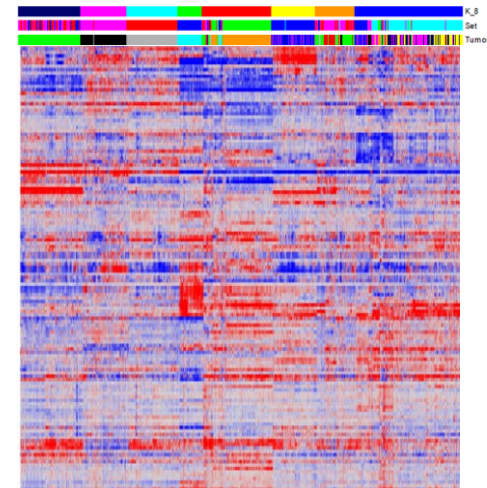
# 6 Data Platforms



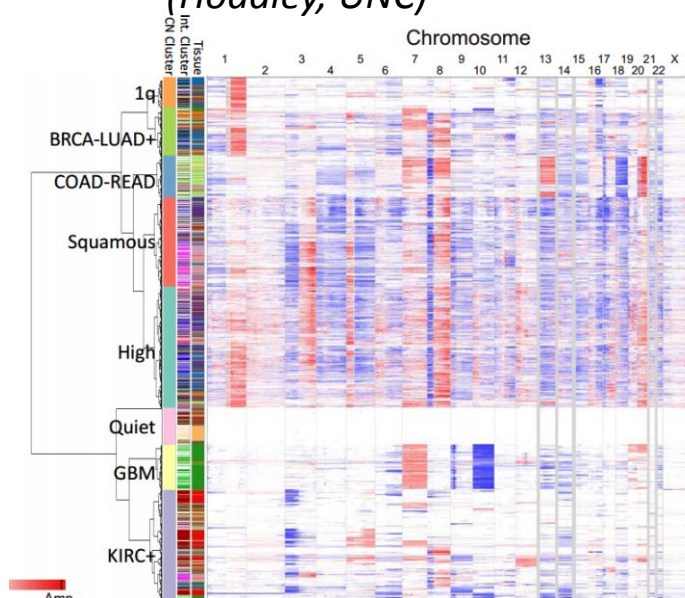
mRNA  
(Hoadley, UNC)



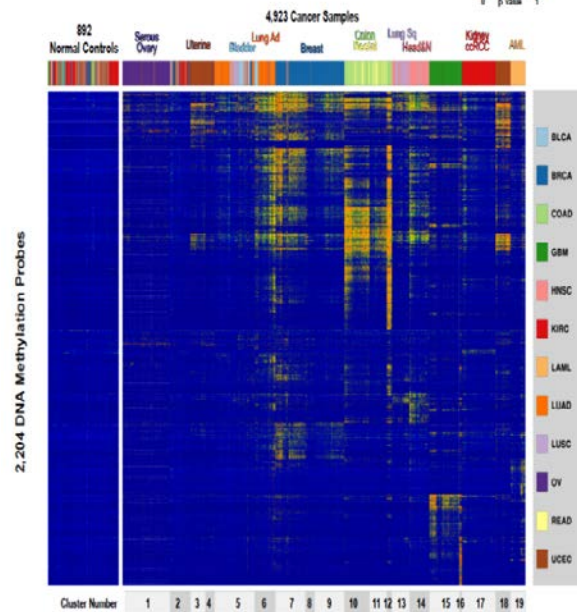
microRNA  
(Hoadley, UNC)



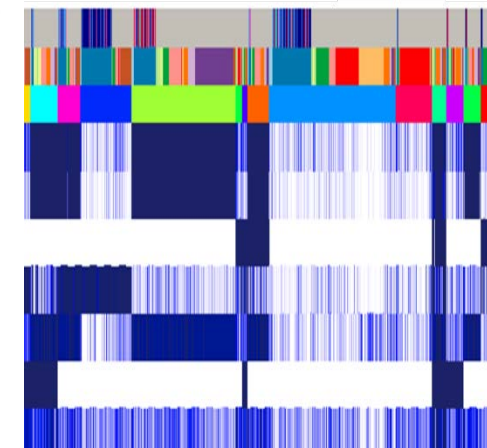
Protein  
(Akbari, MDACC)



DNA Copy Number  
(Cherniack, Broad)



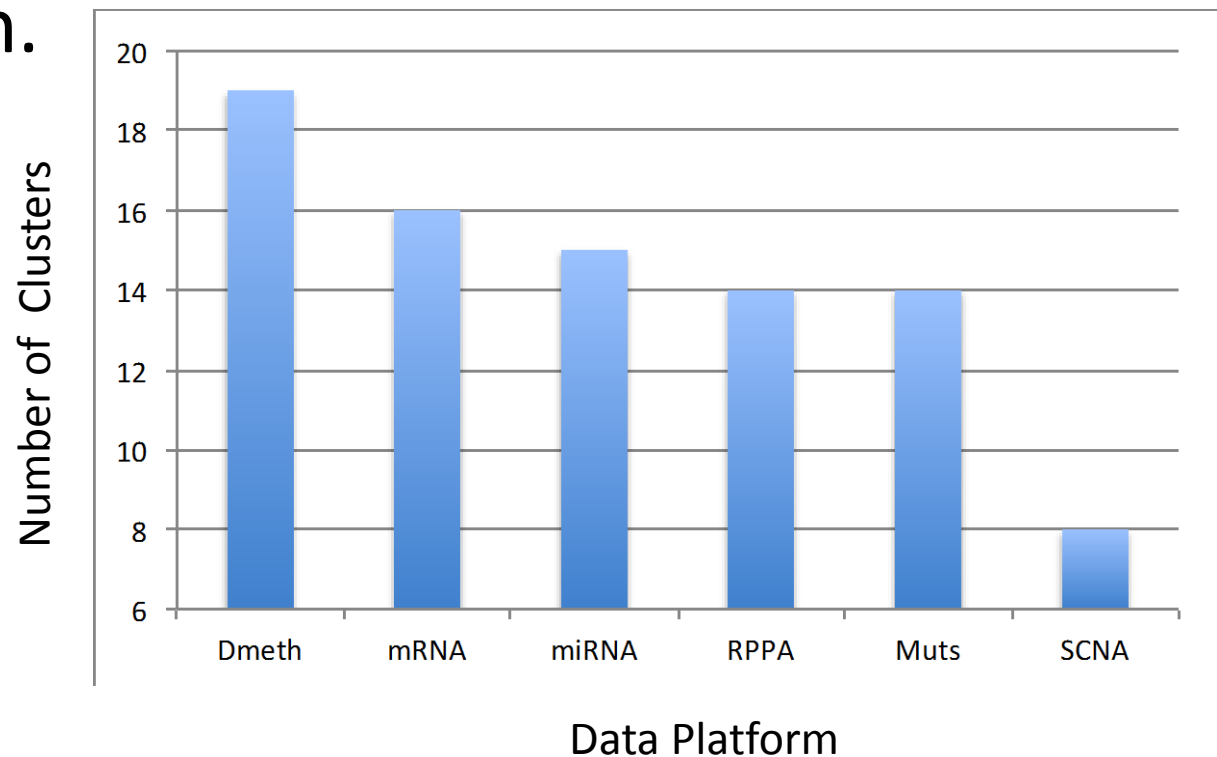
DNA Methylation  
(Shen, USC)



Exome-Mutations  
(not used)  
(Uzunangelov, UCSC)

# Single Platform Subtypes

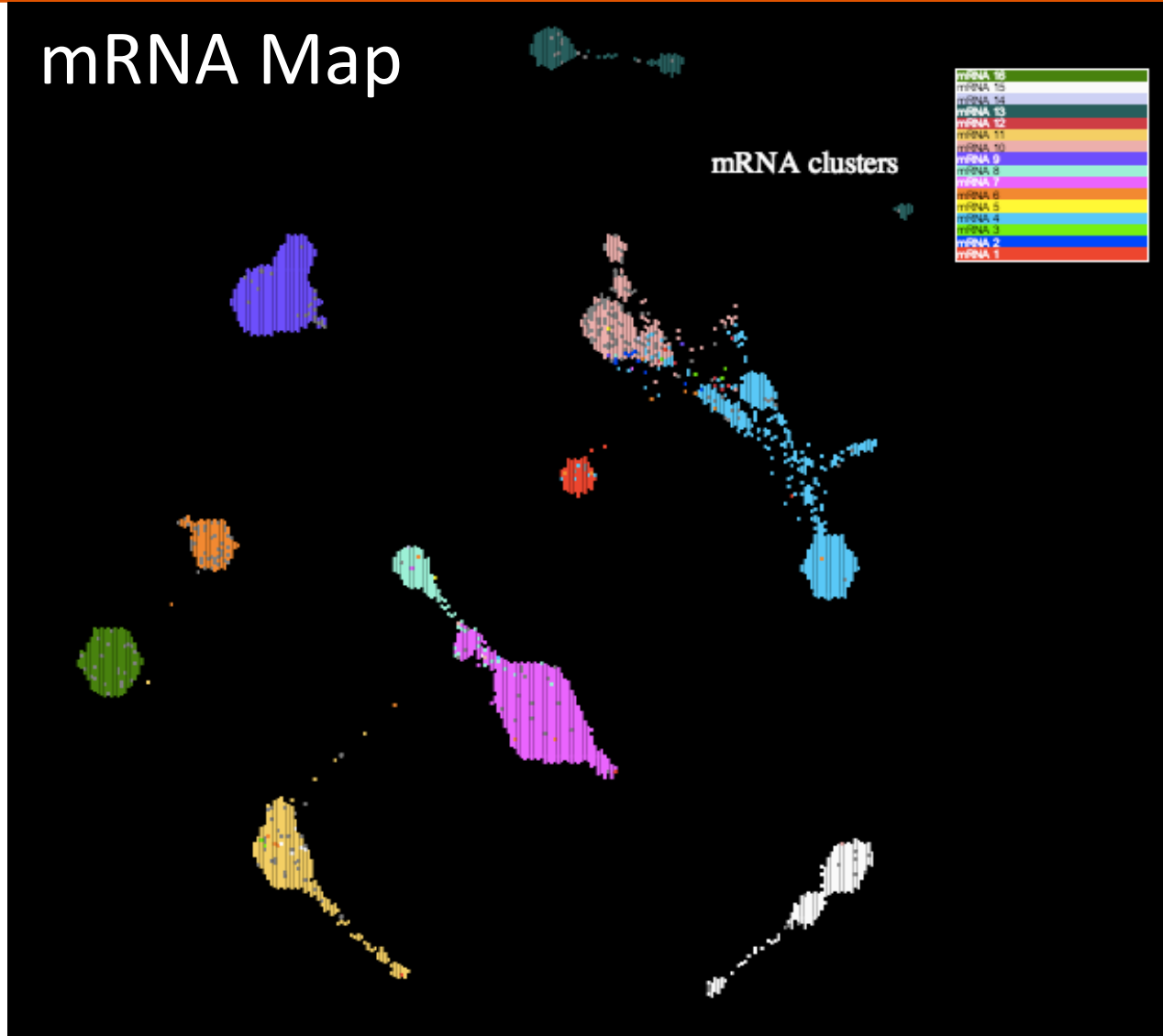
- 6 platforms, each produced 8-19 different clusters.
  - DNA Methylation had the most.
- All subtypes show a strong correlation with tissue of origin.



# Single Platform Subtypes Recap on TumorMap



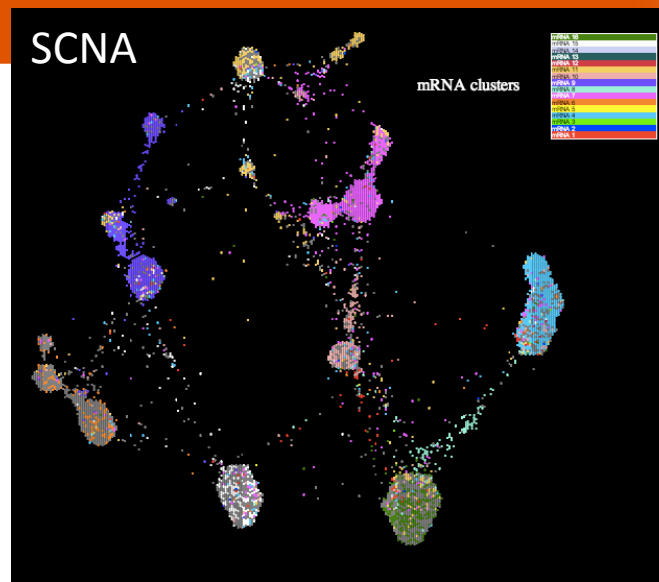
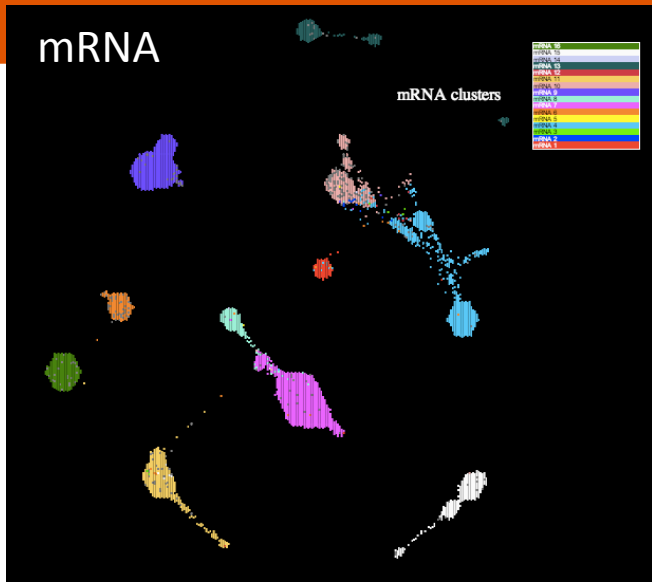
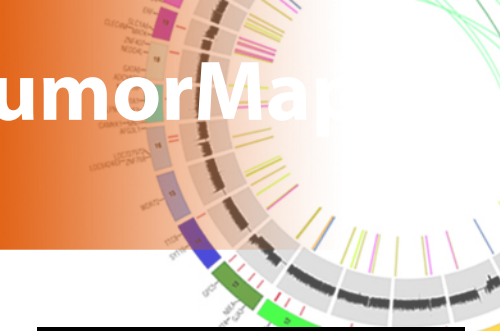
## mRNA Map



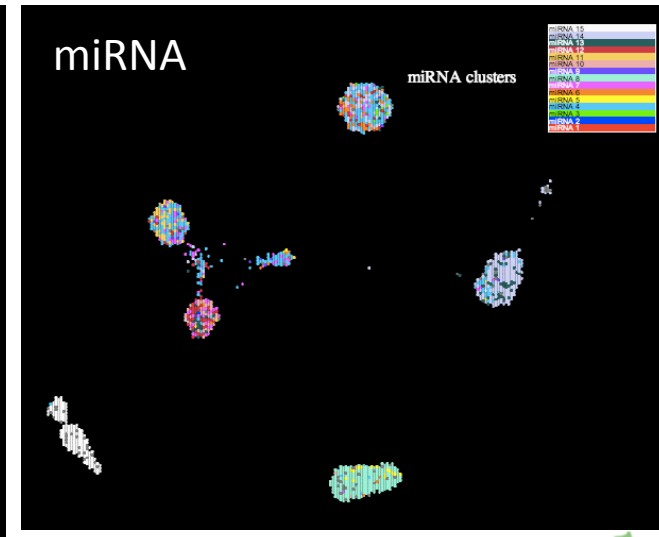
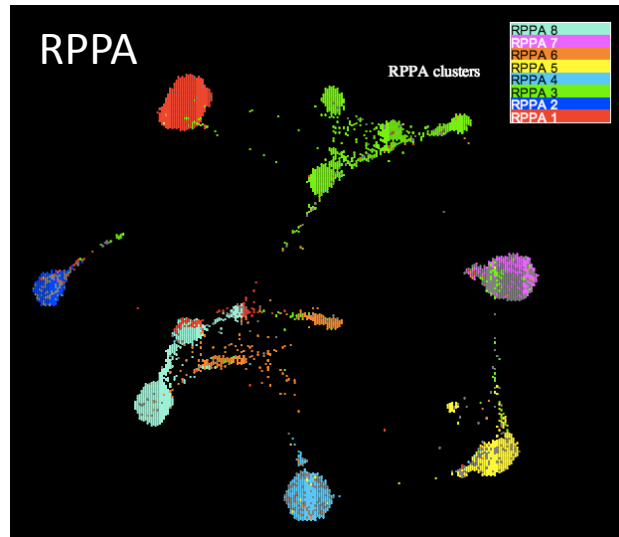
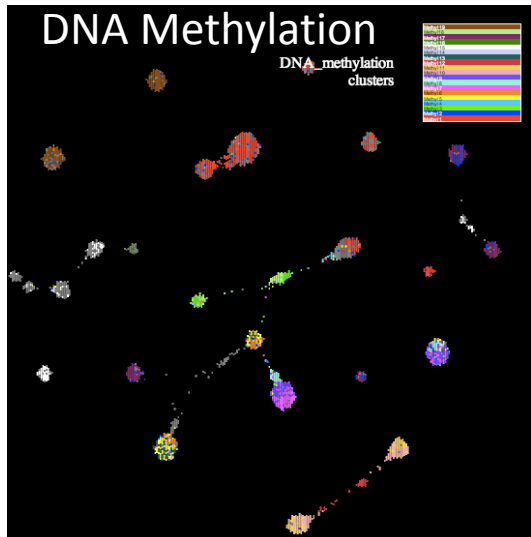
Colors reflect the subtypes obtained using mRNA platform.



# Single Platform Subtypes Recap on TumorMap

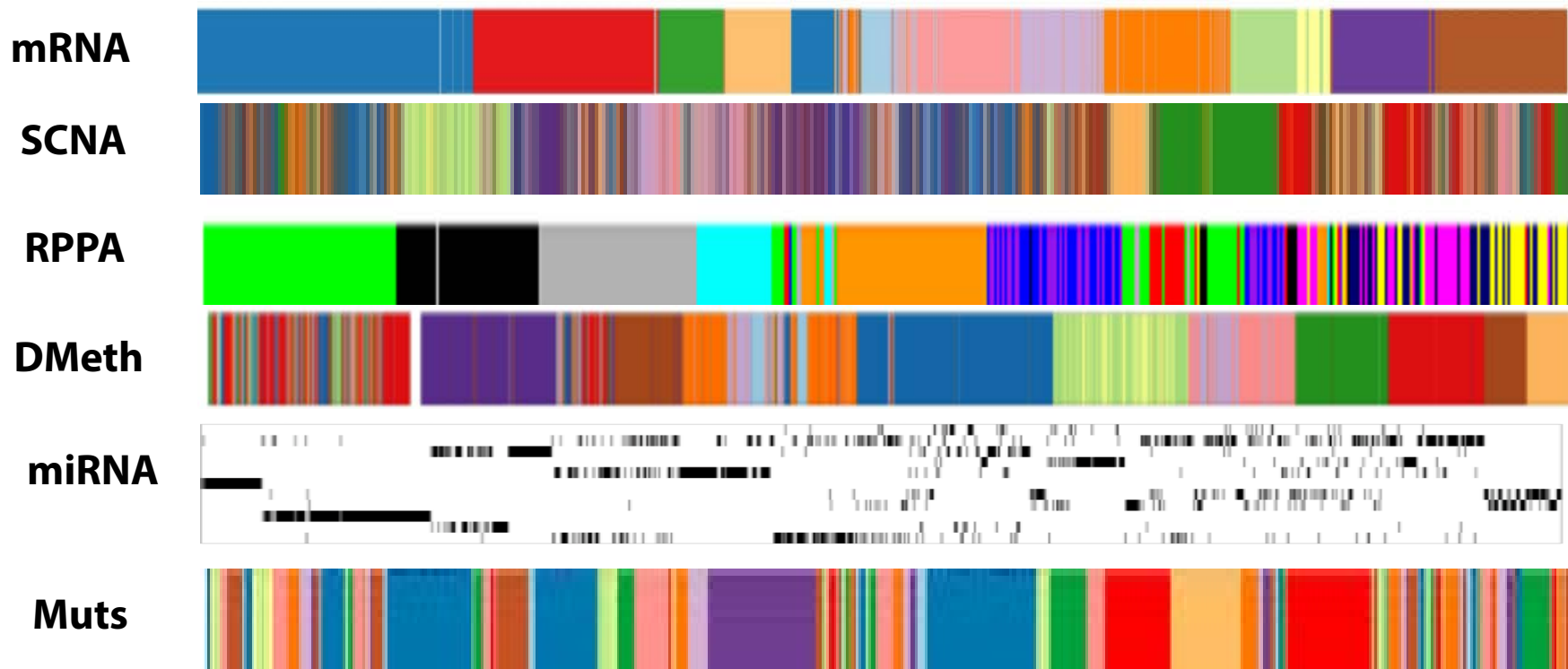
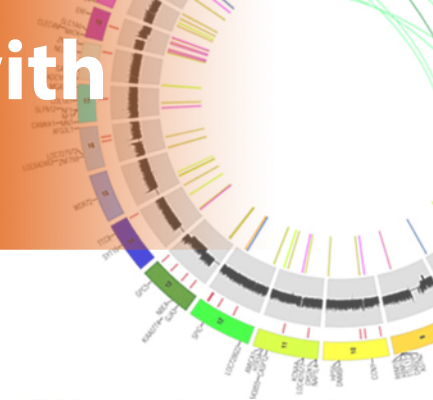


Colors reflect the subtypes obtained using each different platform.



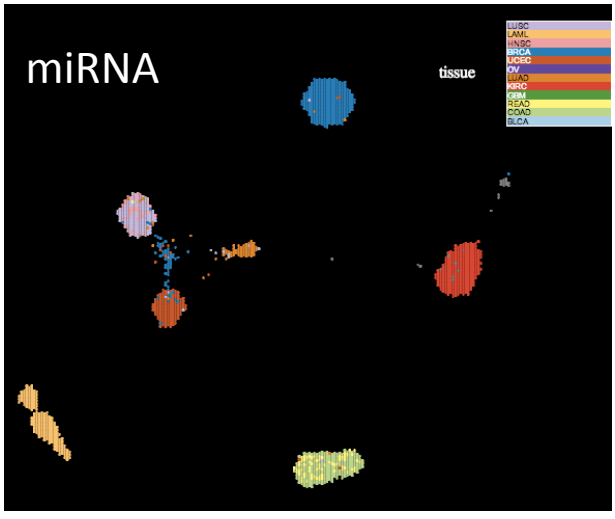
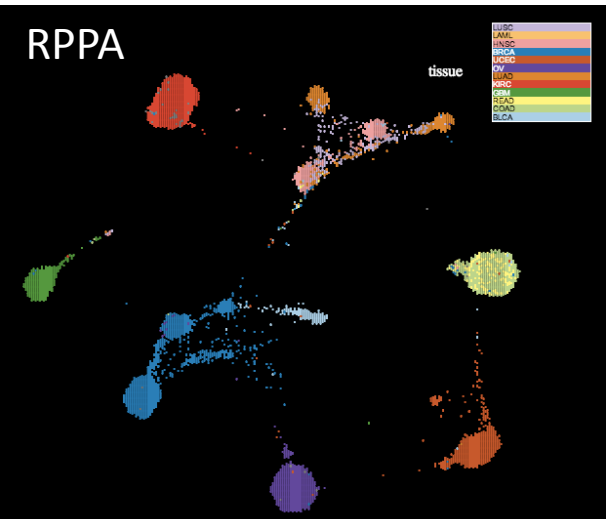
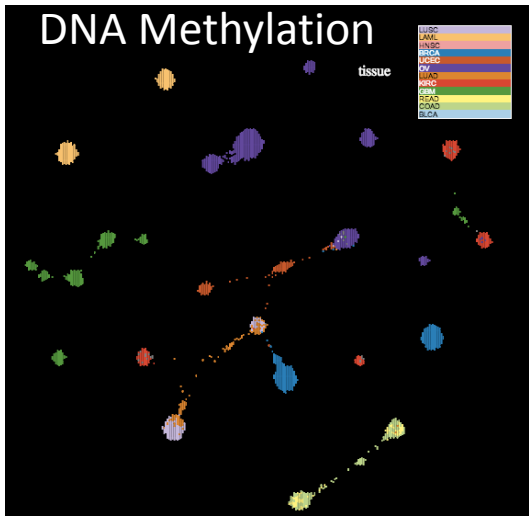
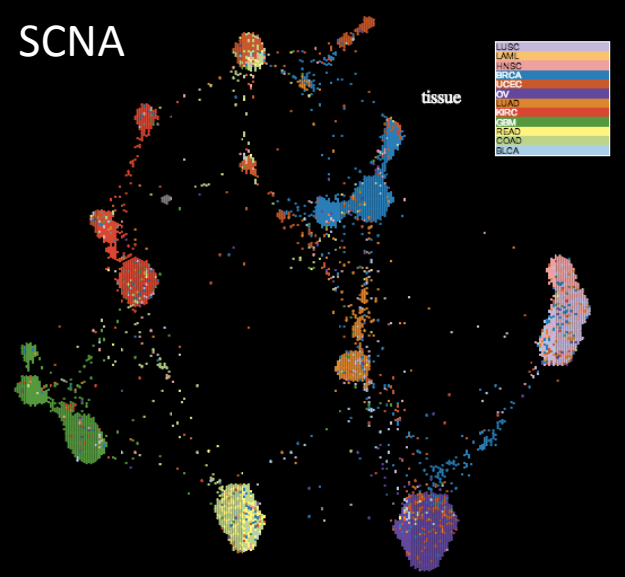
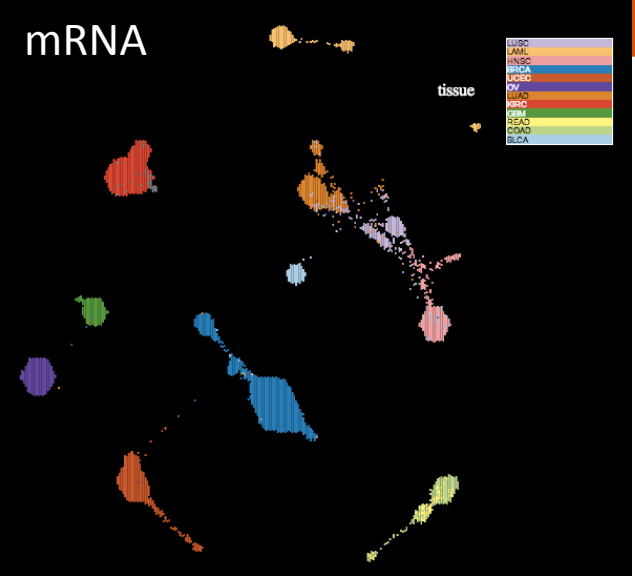
- Good agreement w/ most. miRNA still needs work...

# Single Platform Subtypes Correlated with Tissue of Origin



- Exome mutation clusters show least amount of tissue correlation, but still appreciable (~70%, Kandoth et al. *Nature* 2013).

# Single platform maps are tissue driven

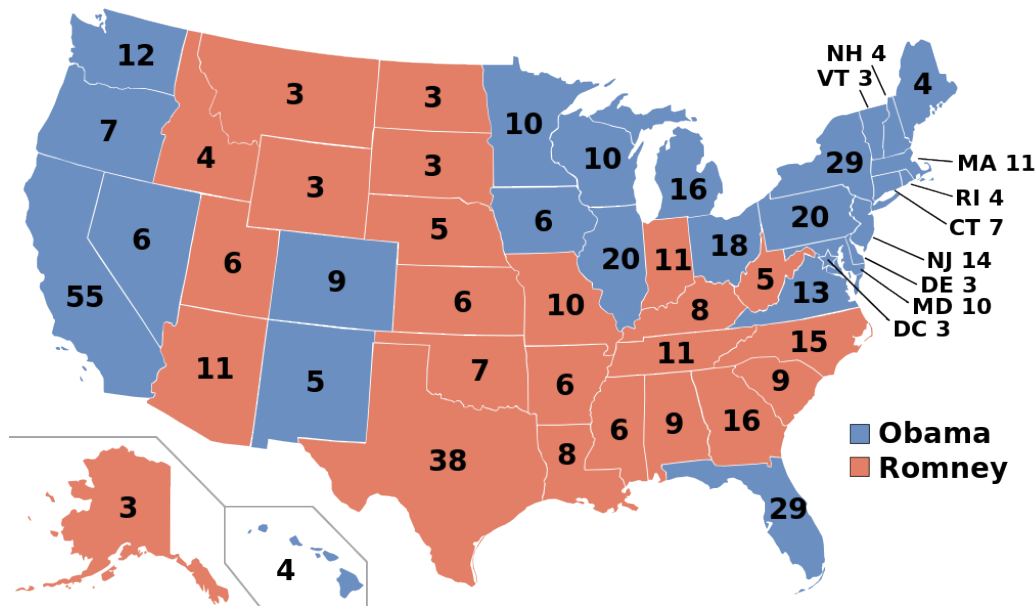


- Each layout driven by a different data platform. † Newton, Baertsch, UCSC

# Cluster of Cluster Assignments - COCA



- Question: How do get one cluster solution from many?
- Answer: Democracy!
  - But like the electoral college system:  
Every Platform Gets a Vote for Each of its Clusters



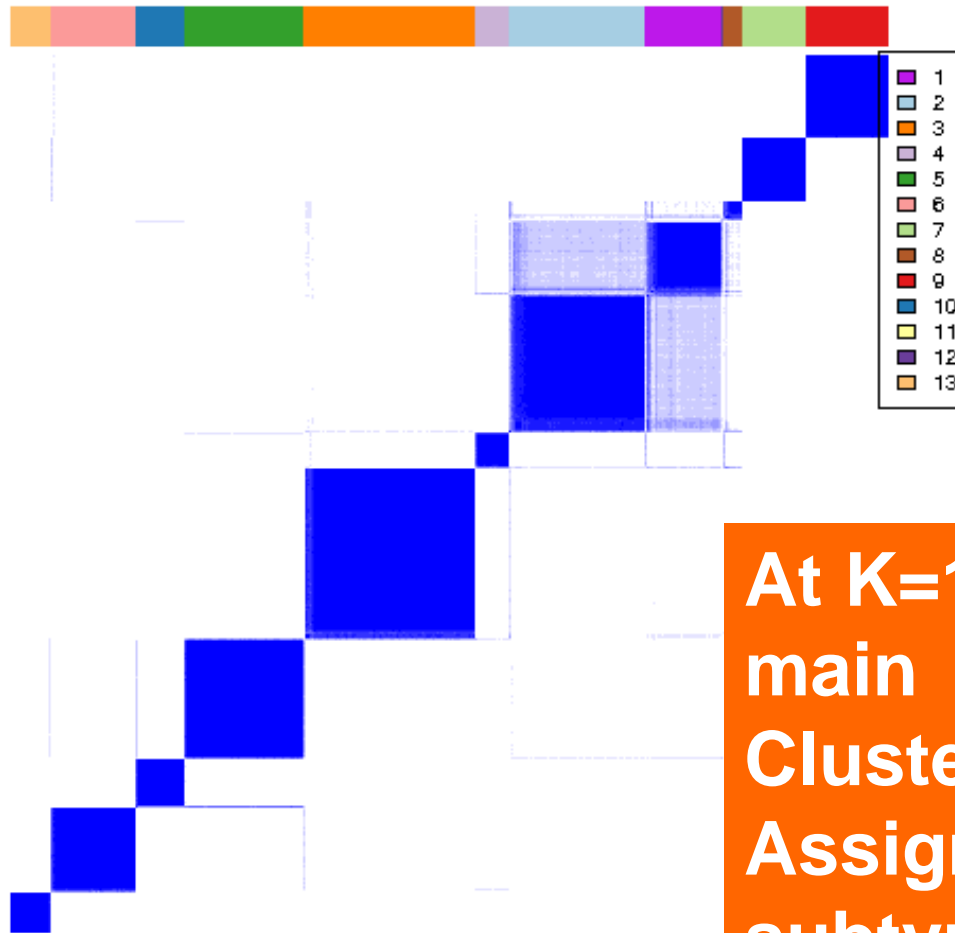
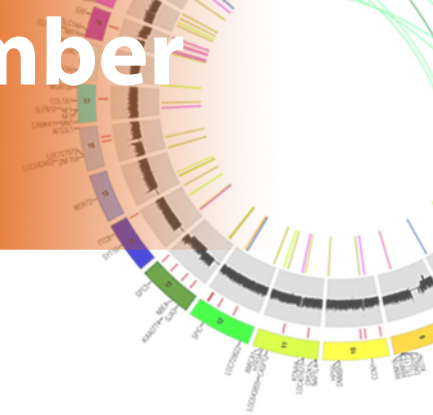
Katherine Hoadley, UNC



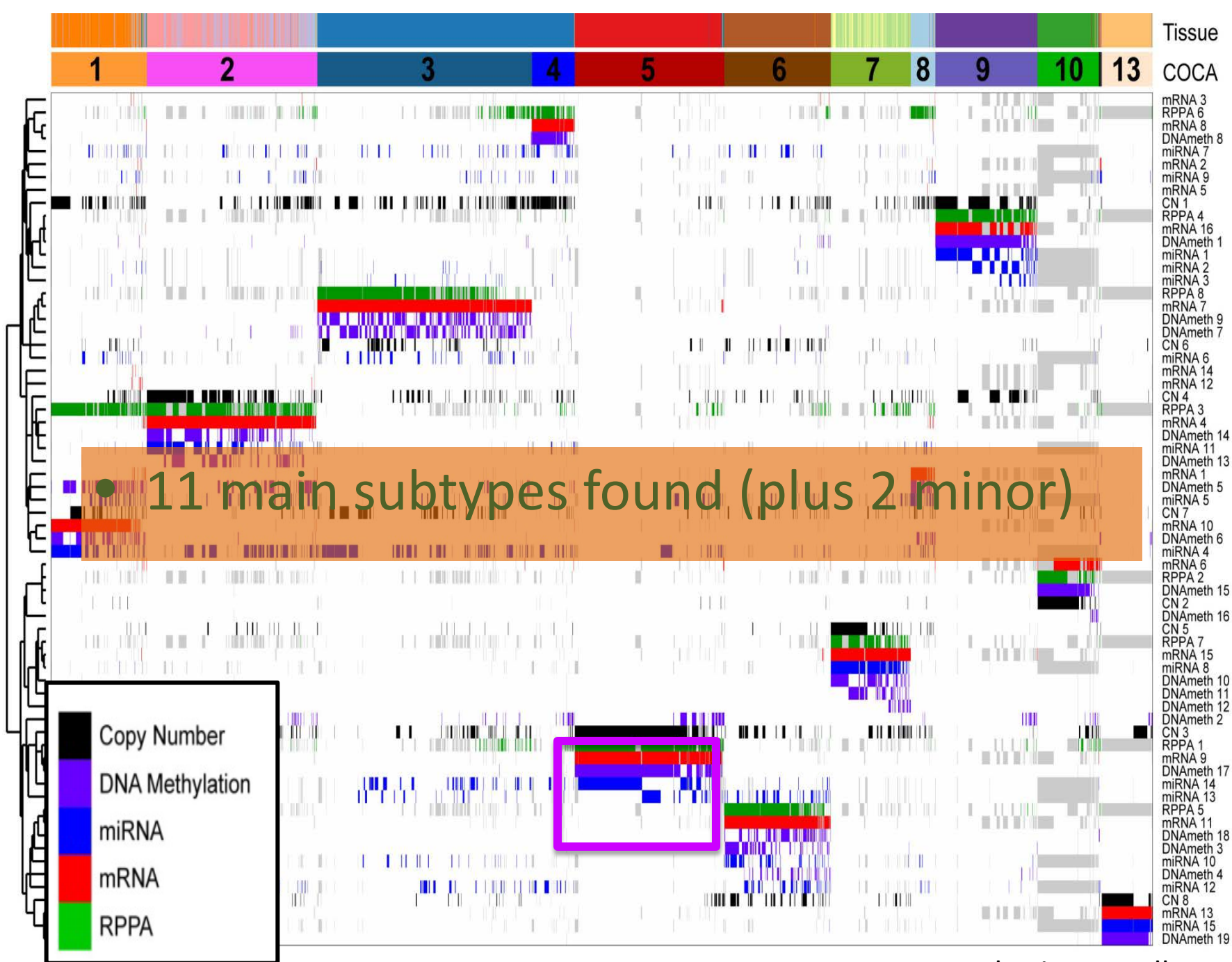
# Cluster Of Cluster Assignments (COCA subtypes)

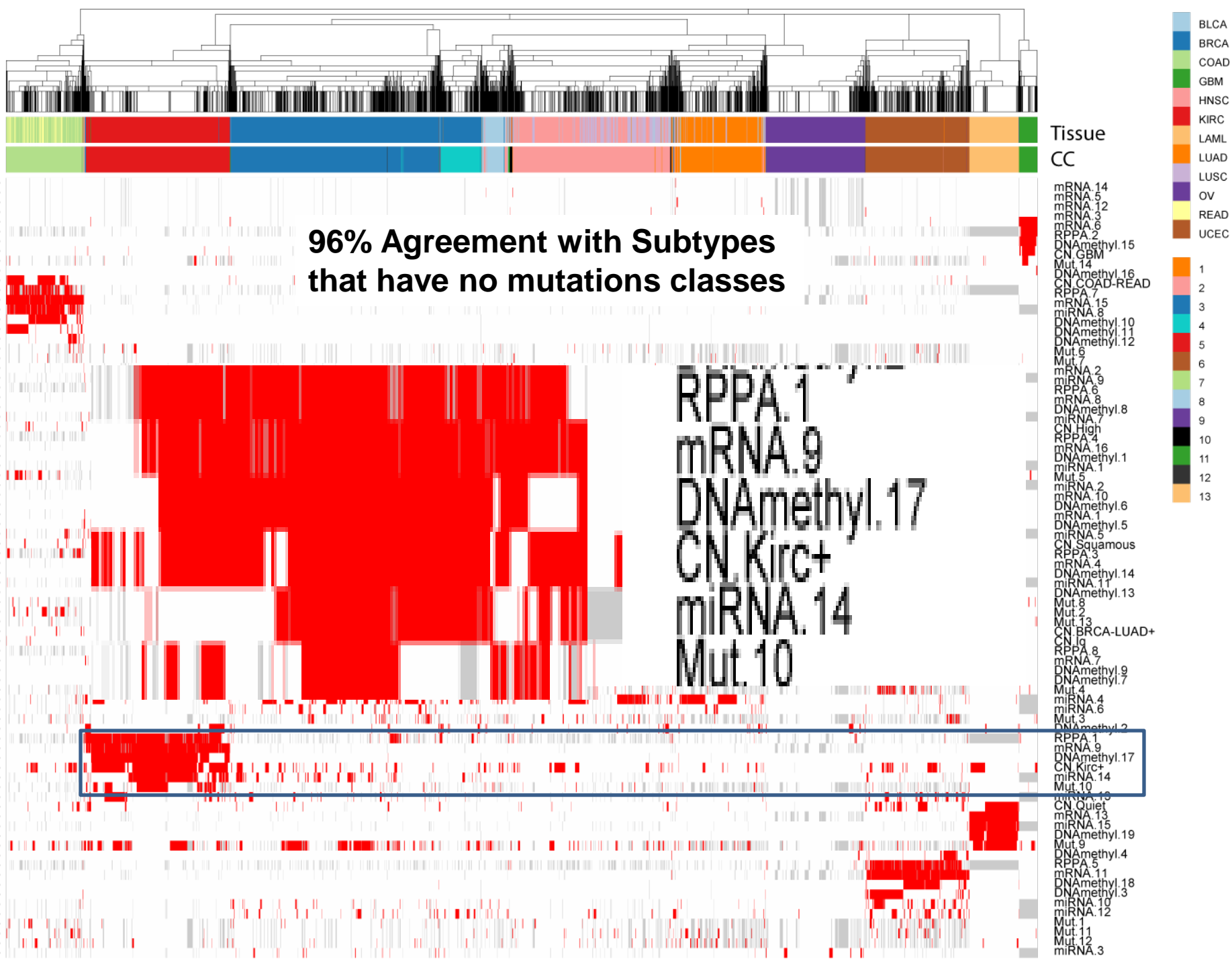
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	NAME	TCGA-BH-	TCGA-B6-	TCGA-E2-	TCGA-BH-	TCGA-A2-	TCGA-E2-	TCGA-AN-	TCGA-AN-	TCGA-B6-	TCGA-B6-	TCGA-BH-	TCGA-E2-	TCGA-BH-	TCGA-B6-	TCGA-
2	CN.1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	
3	CN.2	0	0	0	0	0	0	0	0	1	1	0	0	0	0	
4	CN.3	1	0	1	0	1	0	0	0	0	0	0	0	0	0	
5	CN.4	0	0	0	0	0	0	1	1	0	0	0	1	1	0	
6	CN.5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
7	Methy.1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	
8	Methy.2	0	0	0	0	0	0	0	1	0	1	0	0	0	0	
9	Methy.3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
10	Methy.4	1	1	0	0	0	0	0	0	1	0	0	0	0	0	
11	Methy.5	0	0	1	1	0	1	0	0	0	0	0	1	1	1	
12	miRNA.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	miRNA.2	0	0	0	1	0	0	0	0	1	0	0	0	0	0	
14	miRNA.3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
15	miRNA.4	1	1	1	0	1	0	1	1	0	1	1	0	1	1	
16	miRNA.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	miRNA.6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
18	miRNA.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	PAM50.Basal	0	0	0	0	1	0	0	0	0	0	1	1	1	1	
20	PAM50.HER2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	
21	PAM50.LumA	0	0	0	1	0	1	1	1	1	1	0	0	0	0	
22	PAM50.LumB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	PAM50.Normal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	RPPA.Basal	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
25	RPPA.Her2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
26	RPPA.X	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
27	RPPA.LumA	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
28	RPPA.LumB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
29	RPPA.ReactiveII	0	0	0	1	1	0	0	0	0	1	0	1	1	0	
30	RPPA.ReactiveI	0	0	1	0	0	0	1	0	1	0	1	0	0	0	
31																

# Consensus Clustering defines number of groups



At  $K=13$ , we have 11 main Cluster of Cluster Assignment (COCA) subtypes

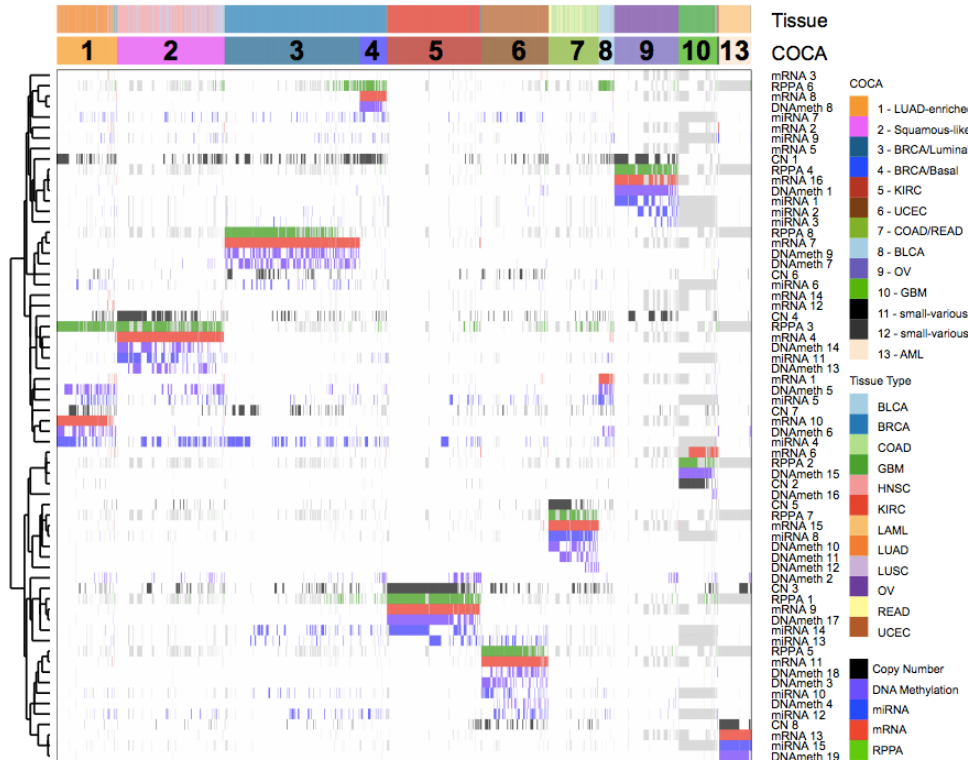




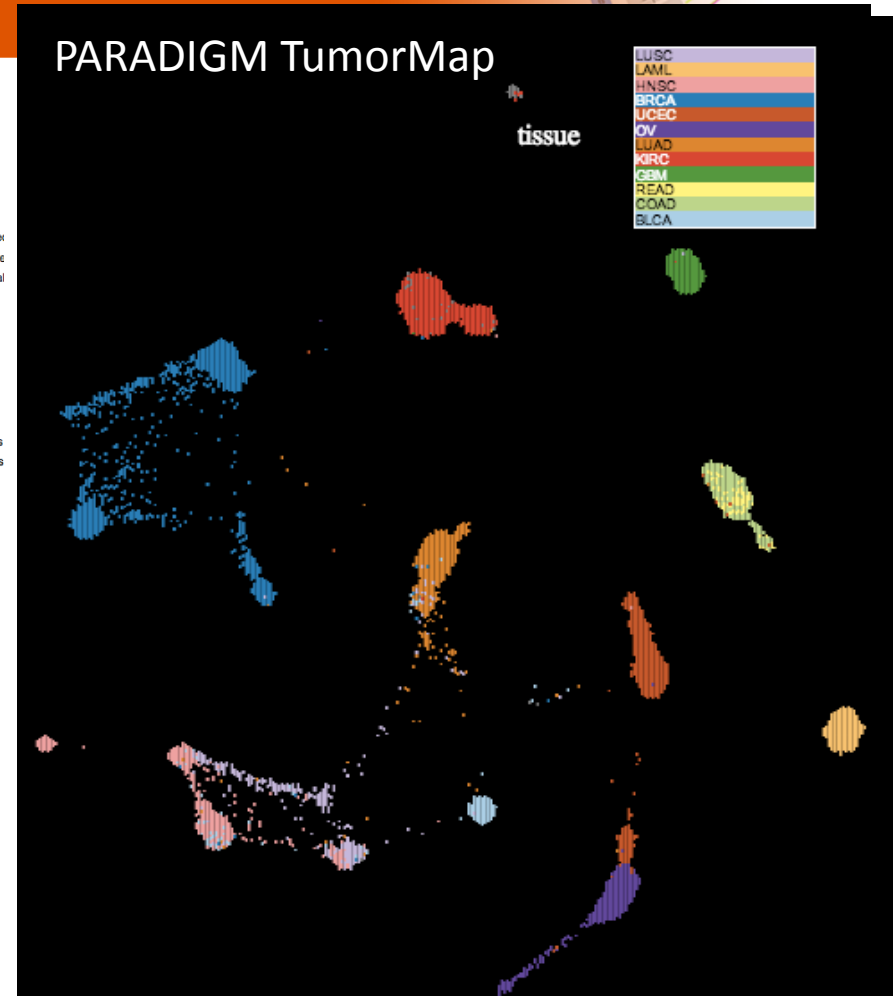


# 11 main COCA Subtypes

A

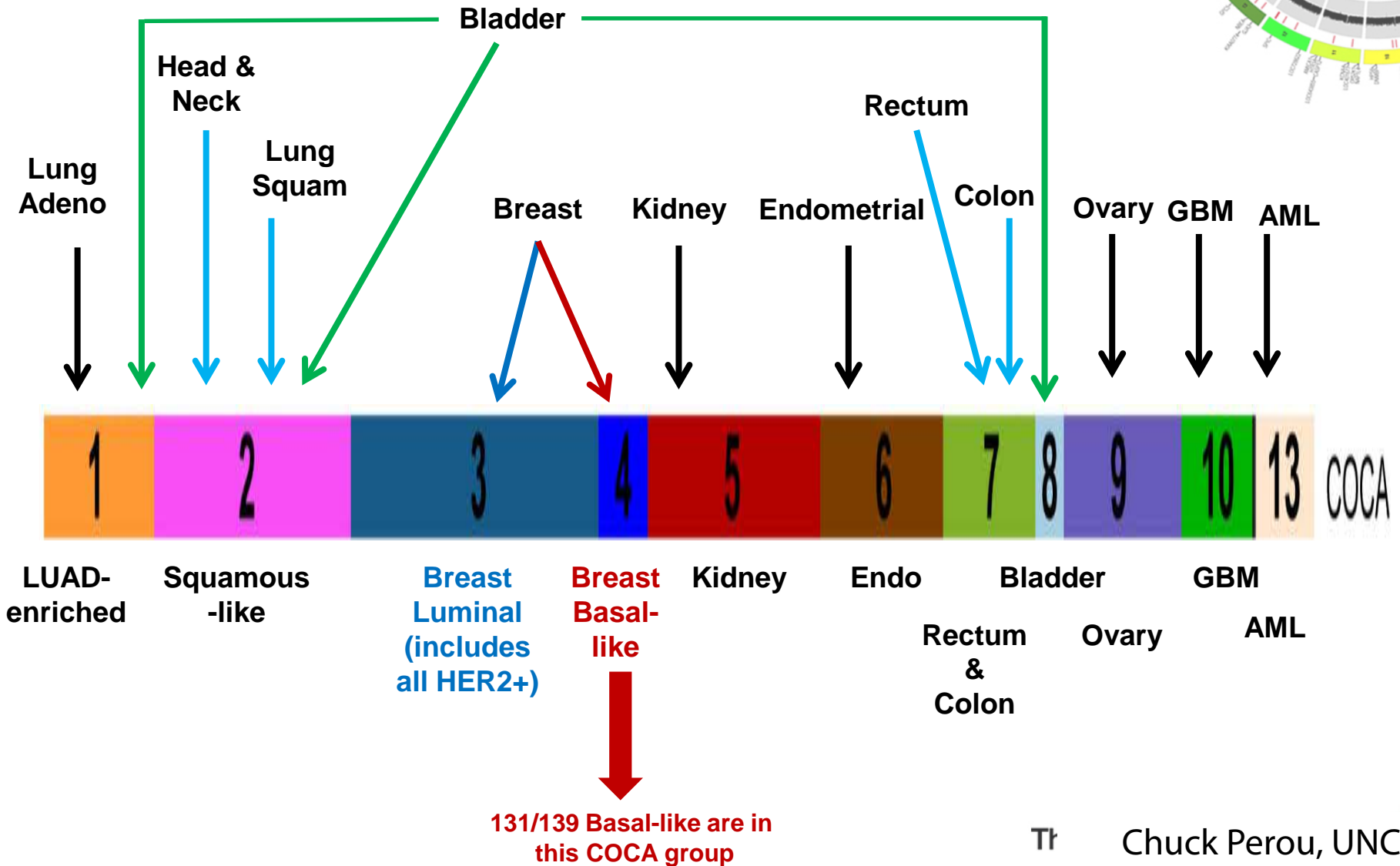


## PARADIGM TumorMap



- 11 main subtypes found (plus 2 minor)
- ~90% of samples cluster with their tissue
- PARADIGM TumorMap corresponds well to COCA

# 12 Tissue of Origin Sites Translate into 11 COCA Subtypes



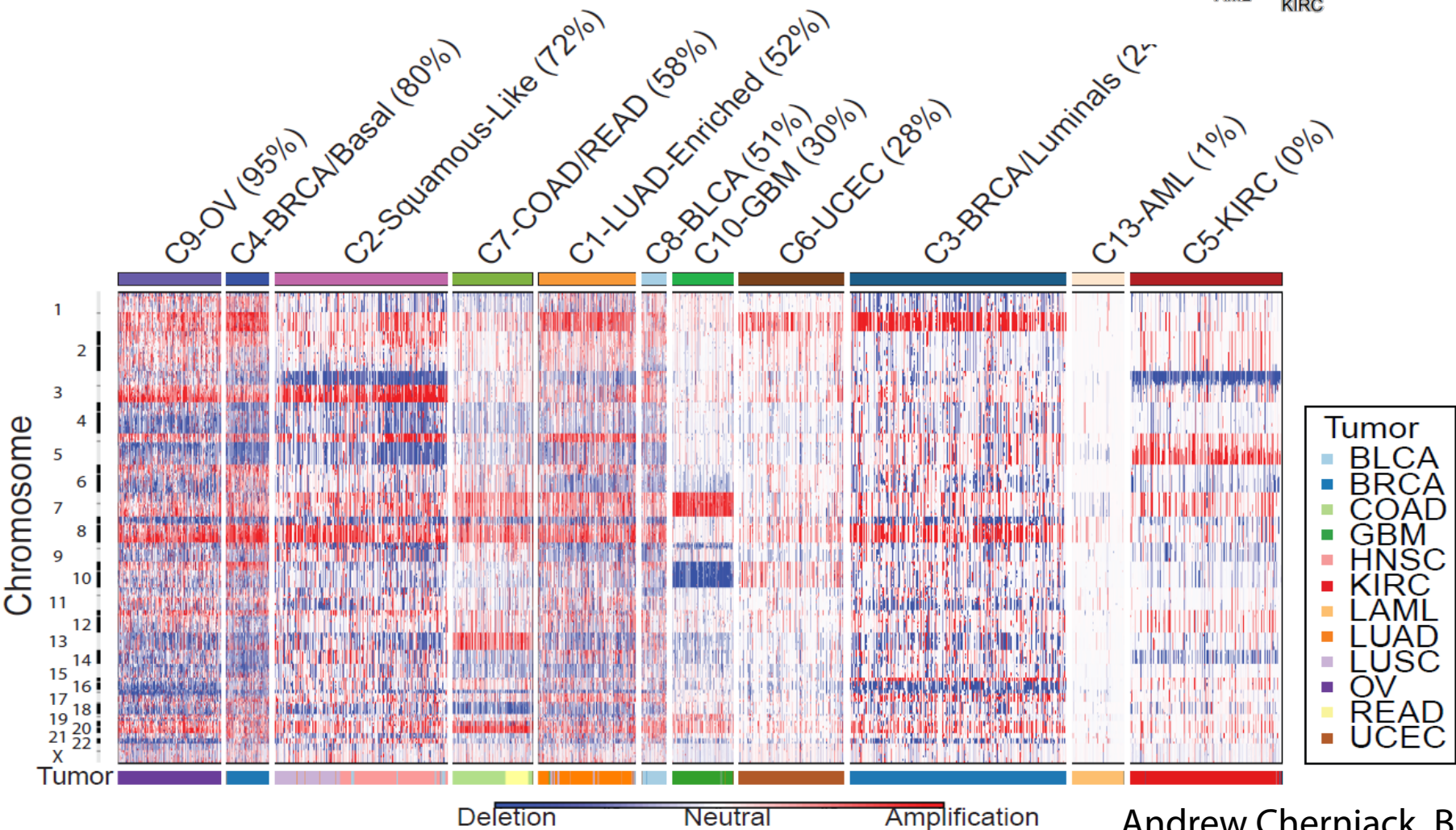
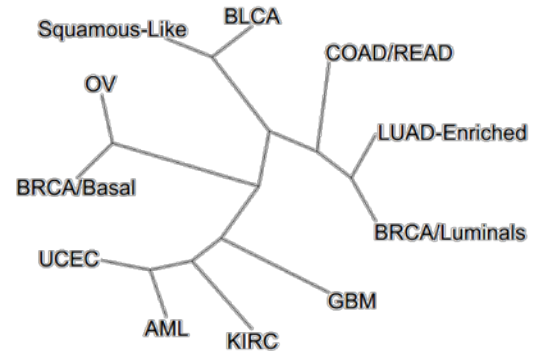
# Mutations according to COCA subtypes

Gene	1- Adeno-enriched	2- Squamous	3- BRCA Luminal	4- BRCA Basal	5- LIRC	6- UCEC	7- COAD	8- Bladder	9- OV	10- GBM	13- AMI	Total
<b>TP53</b>	52%	<b>72%</b>	24%	<b>80%</b>	1%	2%	58%	1%	1%	20%	9%	<b>41%</b>
<b>PIK3CA</b>	7%	19%	<b>40%</b>	4%	3%	<b>51%</b>	18%	17%	1%	9%	0%	<b>20%</b>
<b>PTEN</b>	3%	4%	4%	3%	4%	1%	1%	3%	9%	1%	0%	<b>10%</b>
<b>APC</b>	6%	4%	0%	2%	2%	5%	<b>82%</b>	5%	2%	1%	0%	6%
MLL3	18%	11%	7%	5%	4%	5%	3%	25%	2%	4%	1%	8%
<b>VHL</b>	0%	0%	0%	0%	<b>52%</b>	1%	0%	0%	0%	0%	0%	7%
<b>KRAS</b>	<b>24%</b>	<b>0%</b>	1%	0%	0%	20%	<b>46%</b>	2%	1%	1%	4%	7%
MLL2	10%	20%	2%	0%	3%	1%	2%	1%	1%	3%	1%	7%
ARID1A	8%	5%	2%	0%	2%	3%	1%	20%	2%	2%	1%	7%
PBRM1	2%	3%	0%	2%	32%	2%	0%	5%	0%	1%	0%	6%
NAV3	20%	11%	1%	2%	1%	5%	2%	5%	2%	1%	0%	5%
PIK3R1	2%	2%	3%	1%	0%	<b>31%</b>	2%	0%	0%	15%	0%	5%
NF1	12%	5%	2%	3%	2%	4%	1%	11%	3%	8%	1%	5%
SETD2	7%	3%	1%	1%	12%	3%	3%	8%	2%	2%	1%	5%
ATM	7%	4%	2%	2%	3%	6%	6%	8%	1%	2%	0%	4%
<b>EGFR</b>	<b>11%</b>	<b>4%</b>	1%	0%	2%	1%	2%	0%	1%	<b>25%</b>	1%	4%
FBXW7	1%	6%	0%	2%	0%	12%	12%	6%	1%	0%	0%	1%

Only 3 Genes > 10% frequency

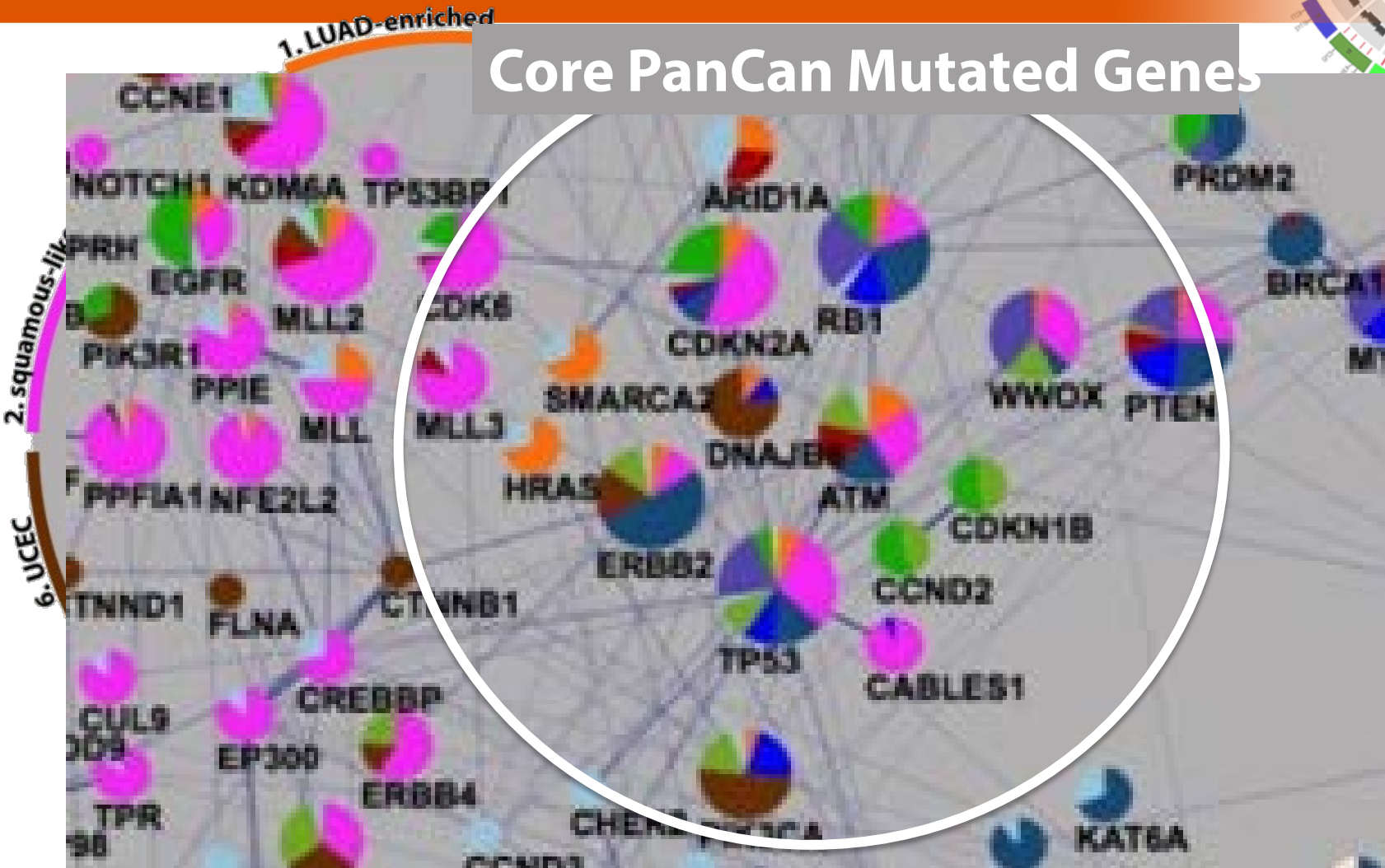
chromatin remodelers, as a class, account for many

# DNA Copy # according to COCA subtypes



# Interconnected mutated networks reveal subtype and tissue preferential

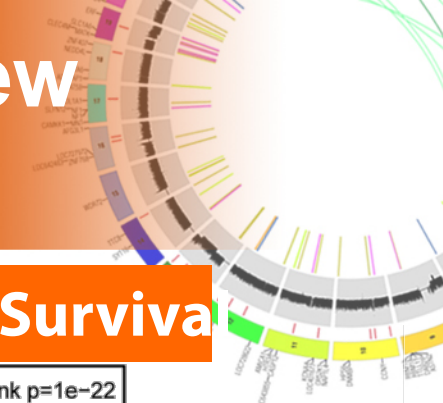
## Core PanCan Mutated Genes



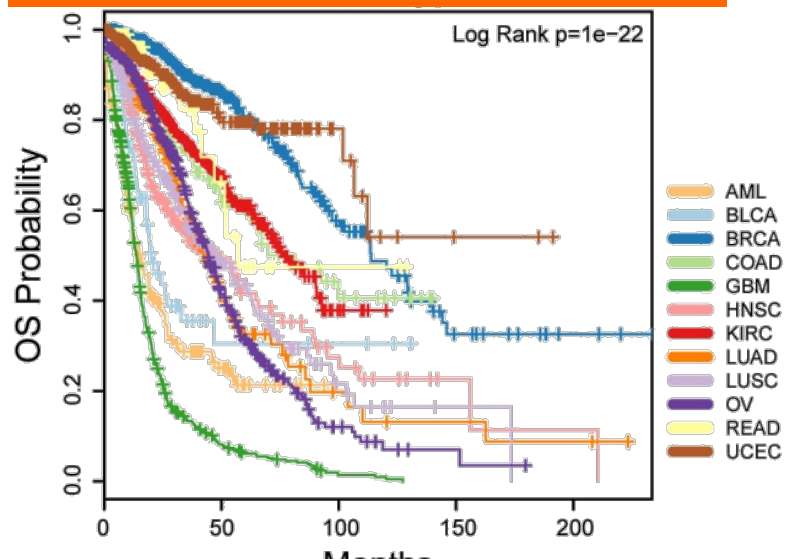
- HotNet2 mutated subnetworks spanning all tumor types.



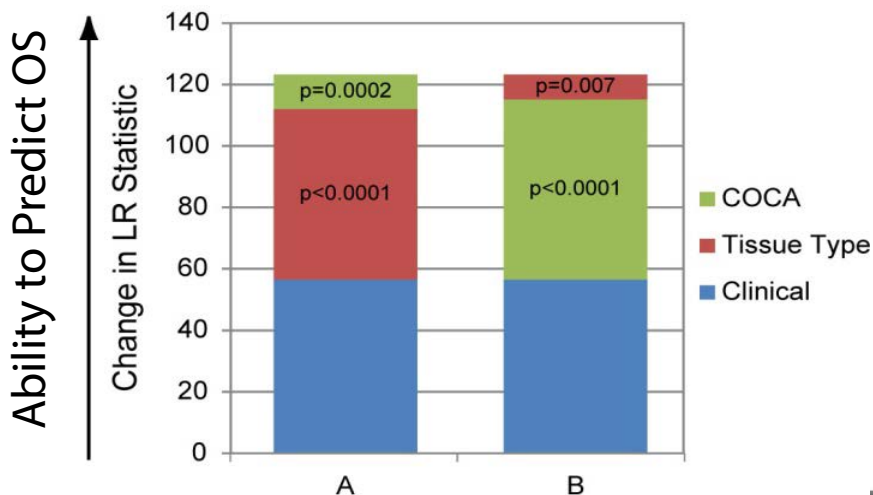
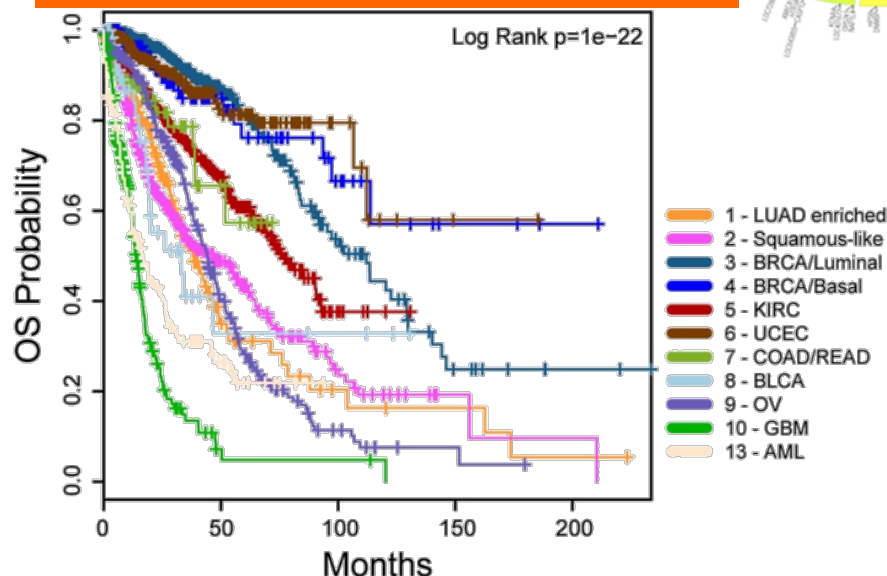
# Do Integrated Subtypes Provide New Prognostic Information?



## Tissue → Overall Survival

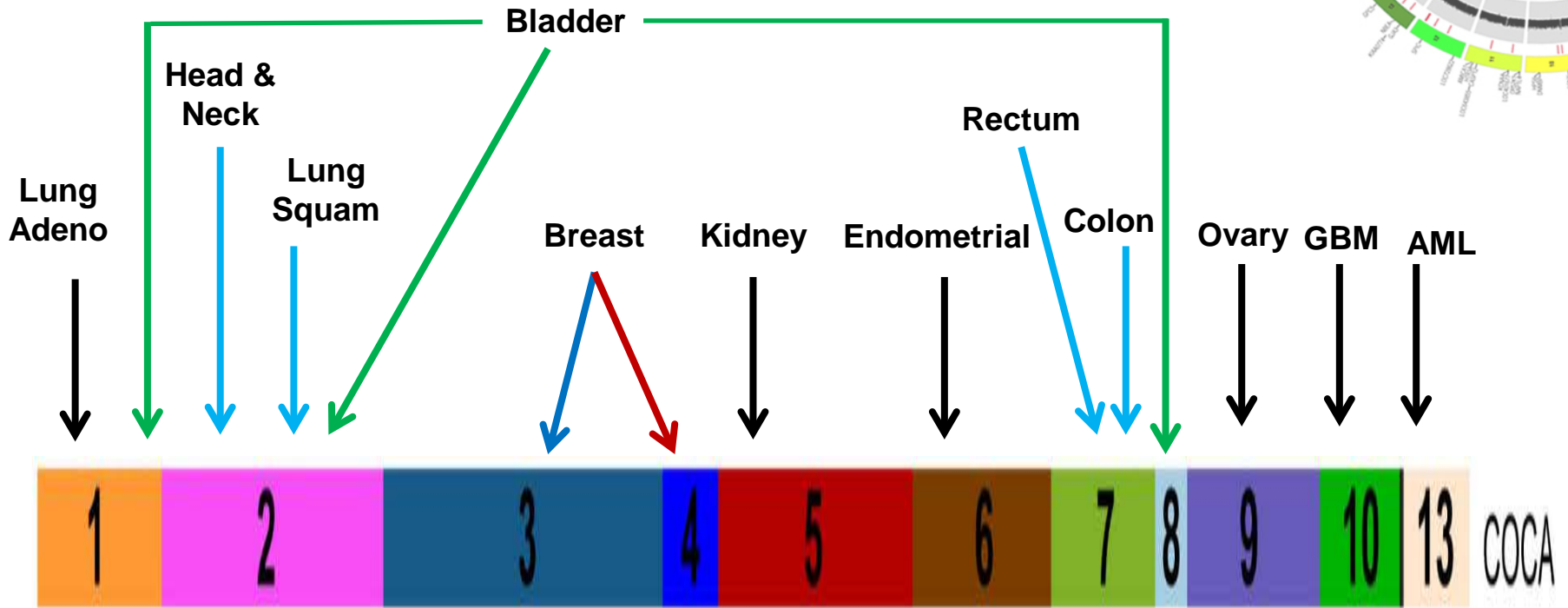
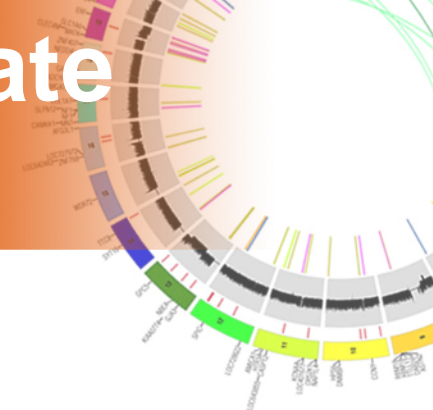


## COCA → Overall Survival



Improvement w/  
integrated subtypes  
over clinical *and* tissue

# 12 Tissue of Origin Sites Translate into 11 COCA Subtypes



Breast Luminal (includes all HER2+)

Breast Basal-like



131/139 Basal-like are in this COCA group

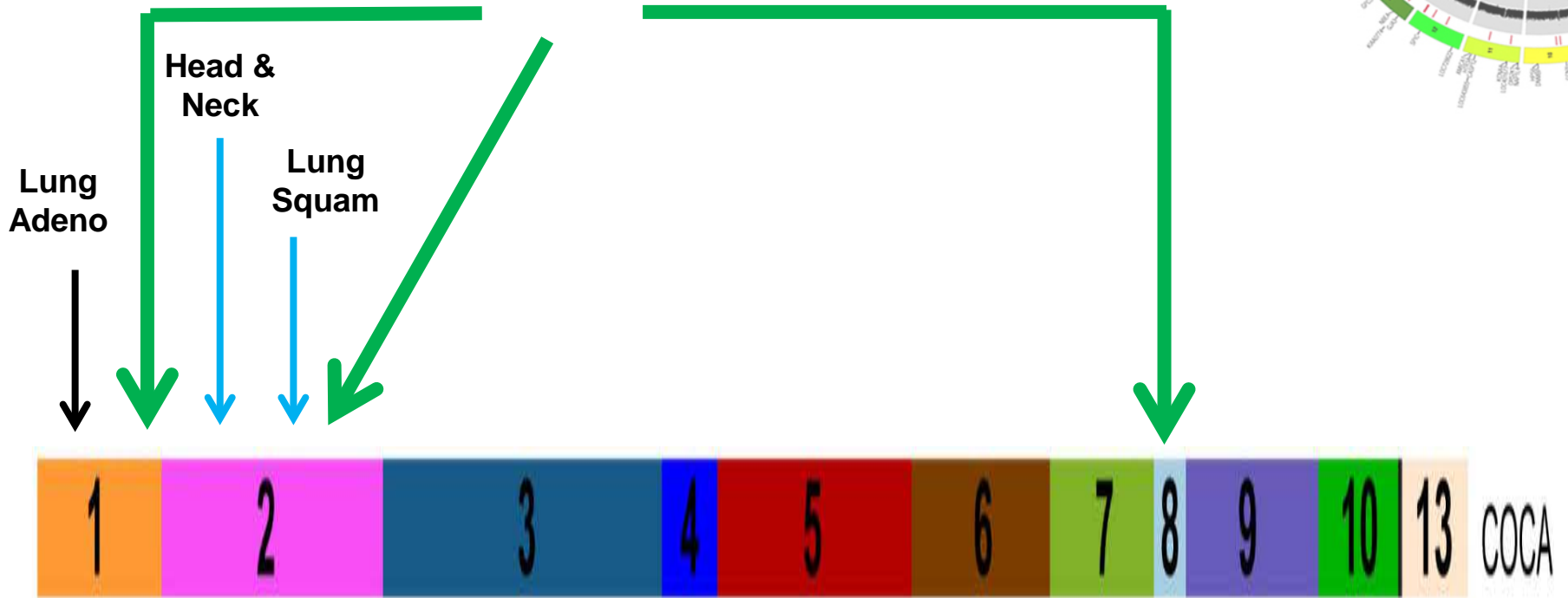
Th

Chuck Perou, UNC

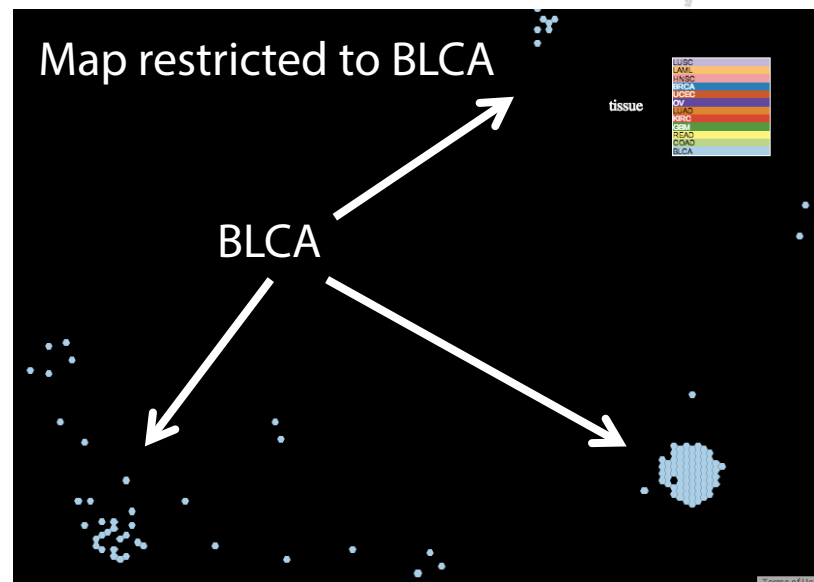
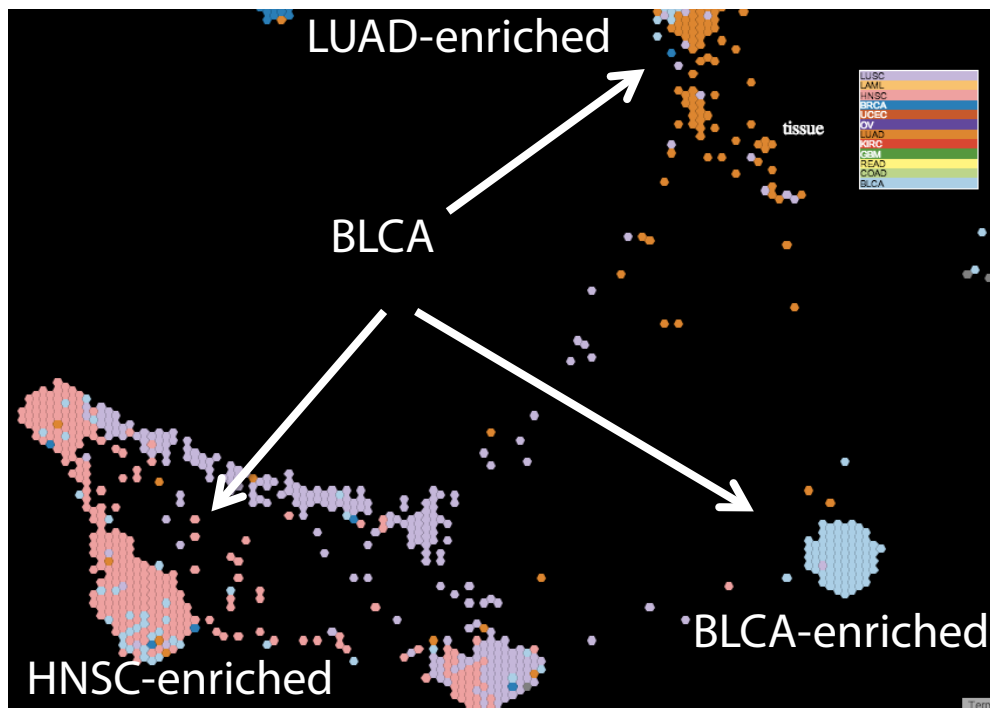
# BLCA samples diverge into 3 integrated subtypes



**Bladder**

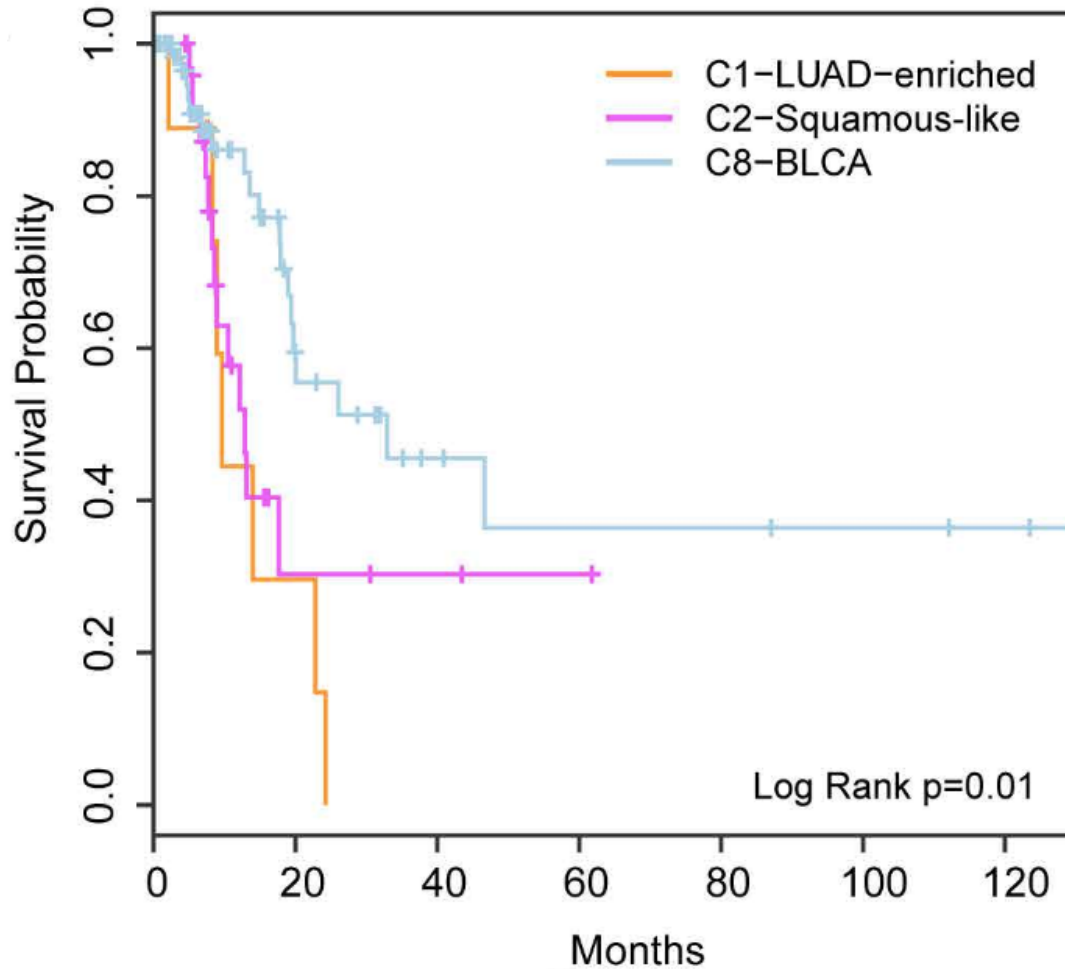


# BLCA divergence in Pan-Can-12



- BLCA diverge into bladder-enriched, squamous, and LUAD-enriched islands

# Integrated subtyping of BLCA distinguishes patient outcomes



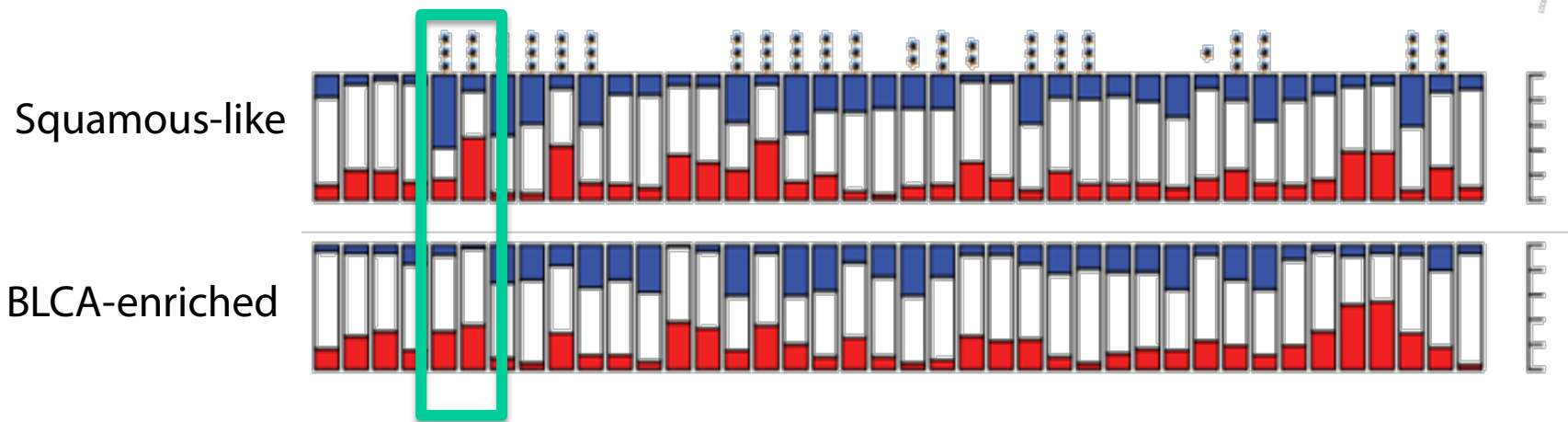
- COCA clusters distinguish different survival classes for BLCA



# Expression determinants of BLCA divergence



- Squamous-like BLCA show significant genomic differences



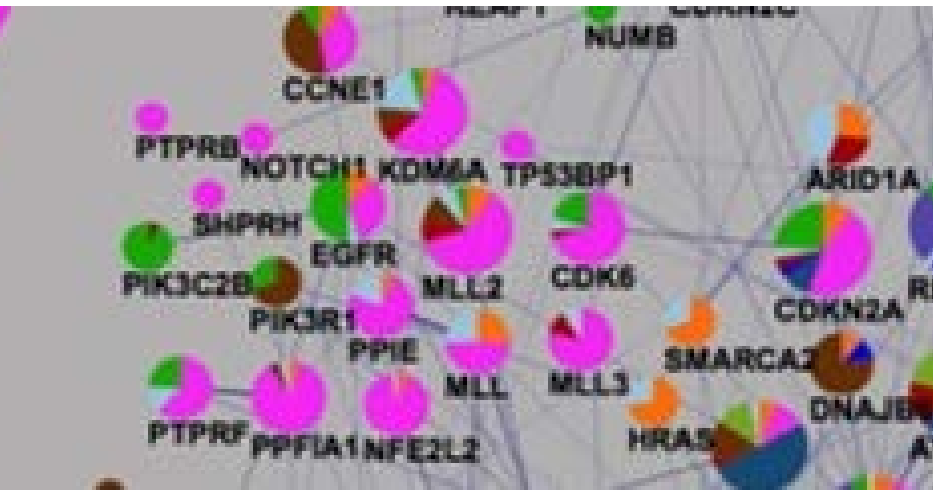
3p Loss  
in Squamous

BLCA-squamous  
mutated chromatin  
remodelers MLLs  
KDM5A, EP300



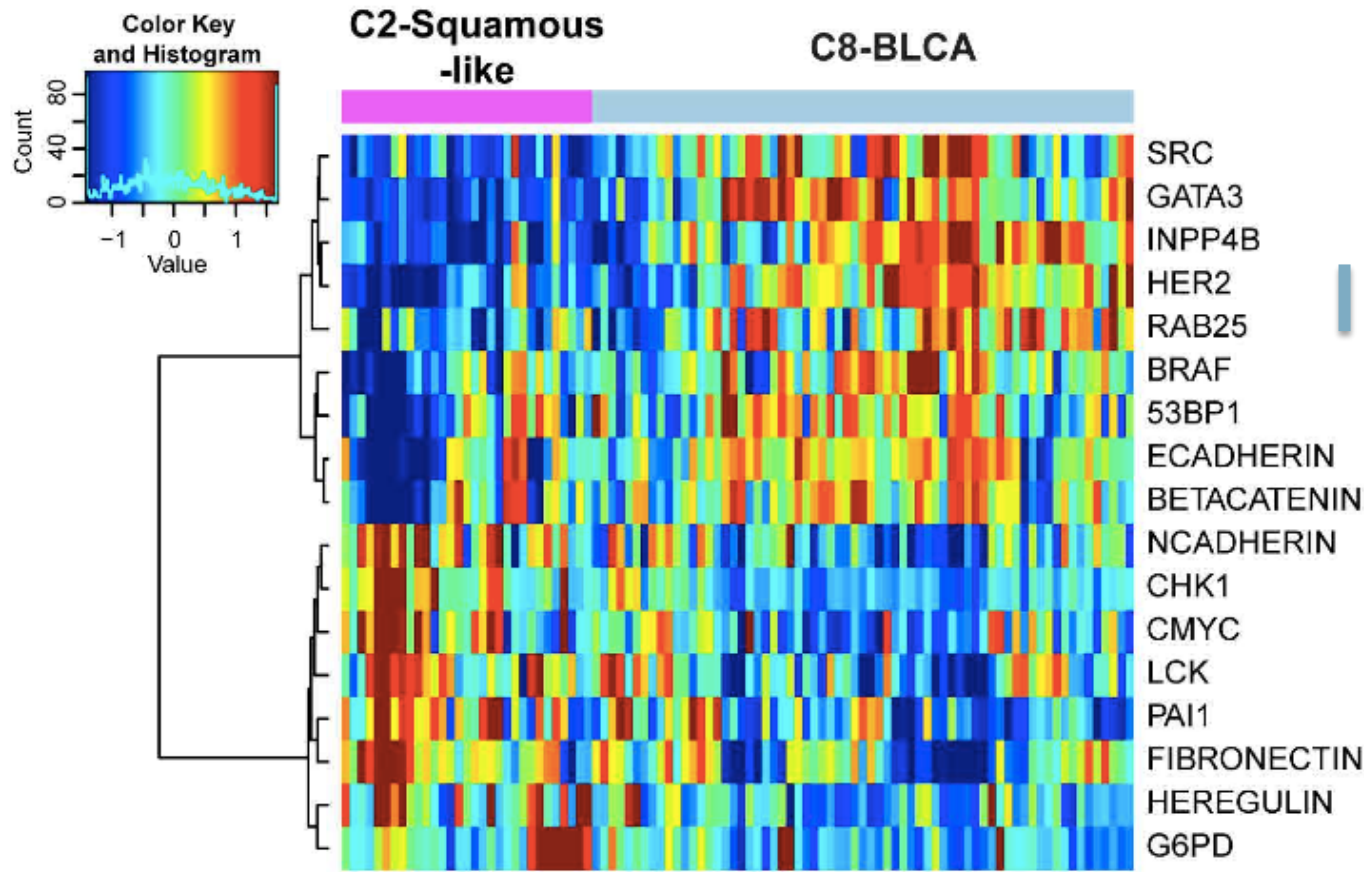
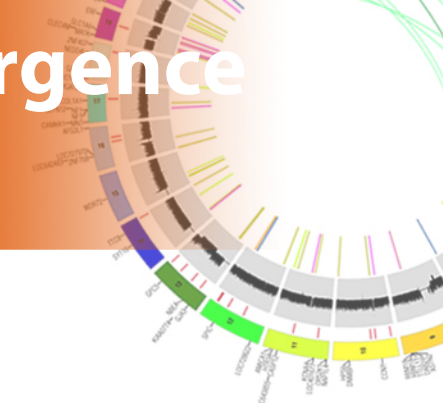
Test for differential mutation frequency (FDR)

4. squamous-like



under-represented non-significance enriched

# Expression determinants of BLCA divergence

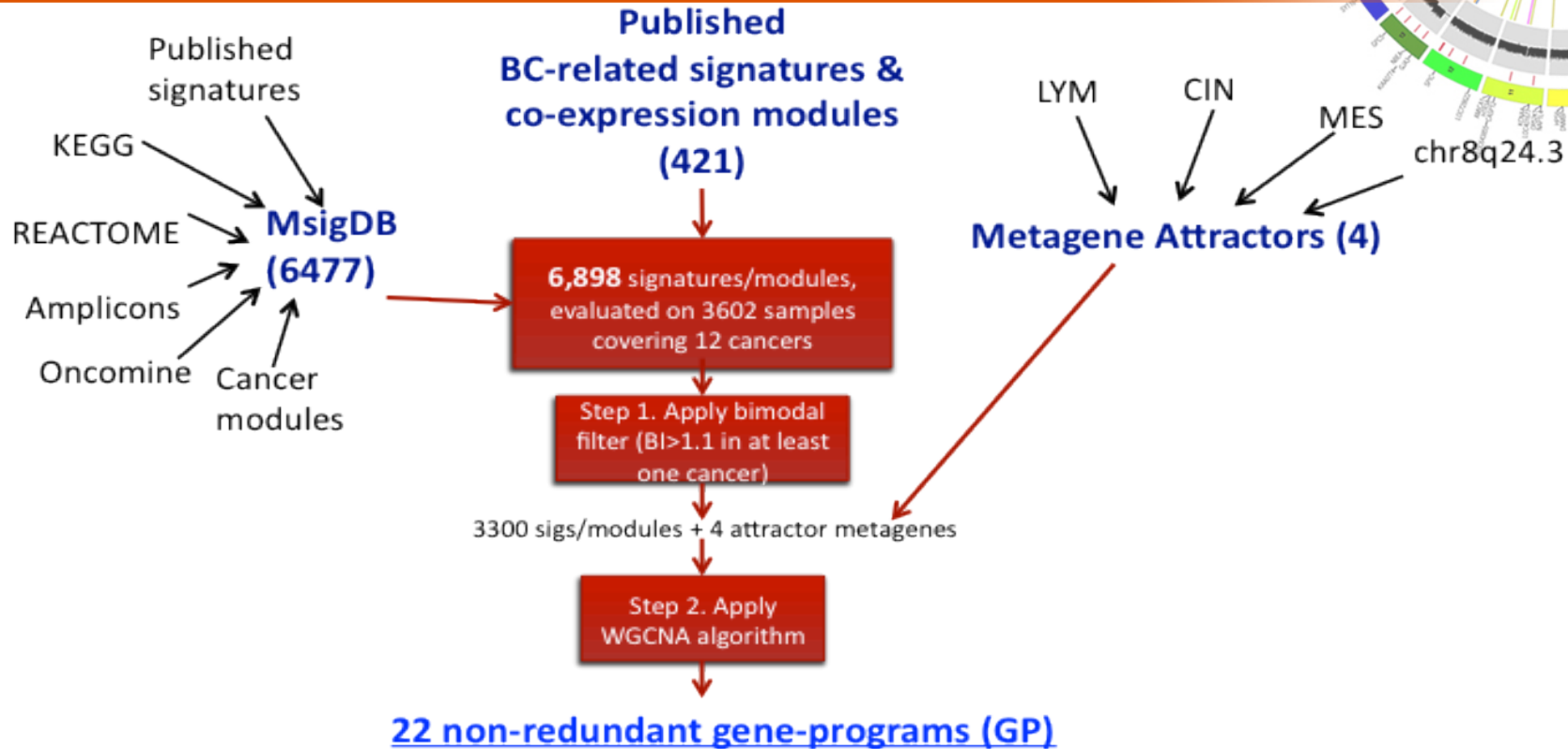


See BLCA AWG

EMT and proliferation

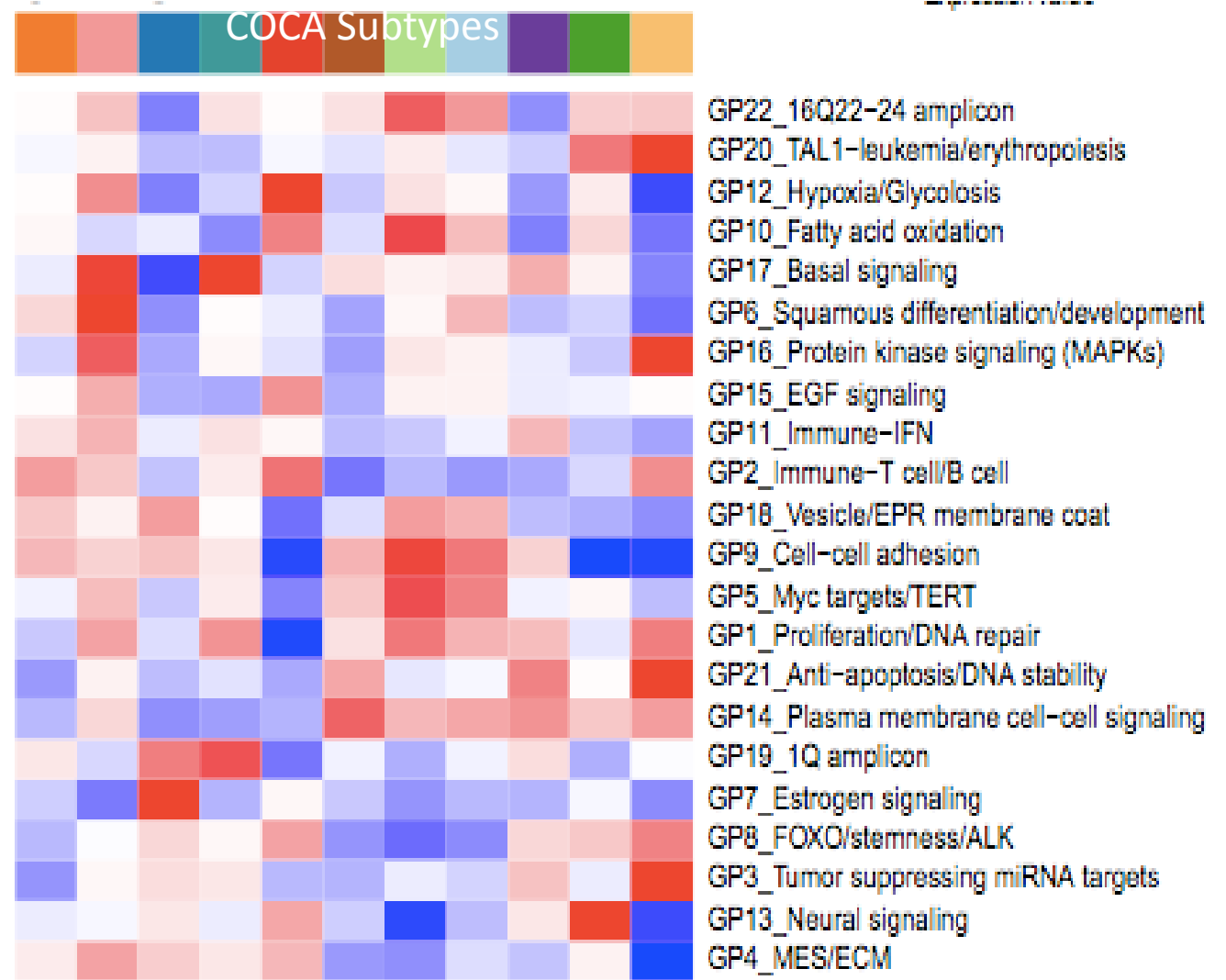
- Higher HER2 and Rab25 in non-squamous BLCA – consistent w/ BLCA AWG
- Markers of EMT expressed in squamous BLCA cases

# Gene Programs – functionally coupled genes coregulated across PanCan-12



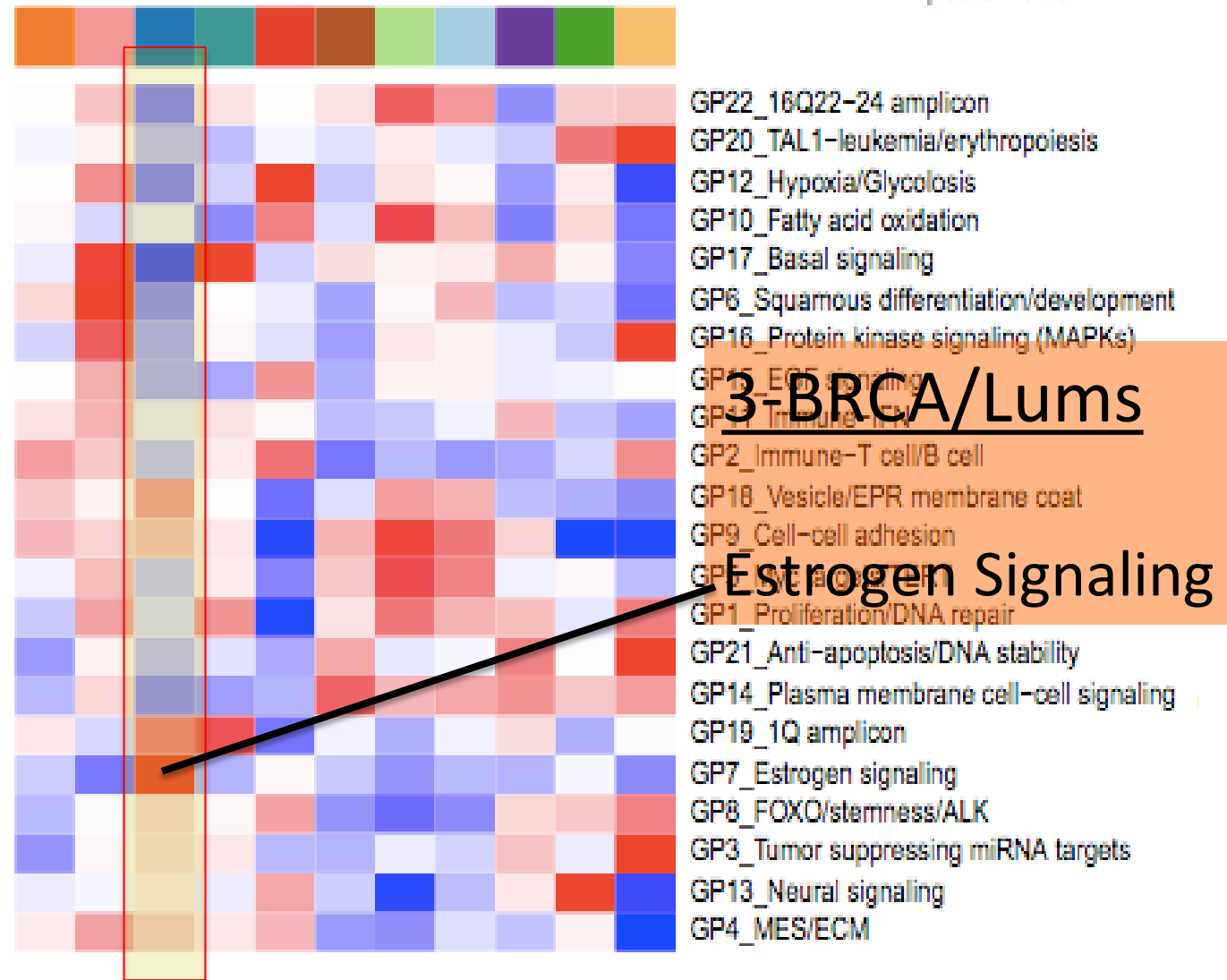
- Identified 22 sets of functionally-related genes coregulated in PanCan-12.
- Gene Programs can recapitulate the integrated subtypes

# Gene Programs: Surrogates of Integrated Subtypes



- 90% classification accuracy (LDA)

# Gene Programs: Surrogates of Integrated Subtypes



- 90% classification accuracy (LDA)



# Viewing Gene Programs on the TumorMap

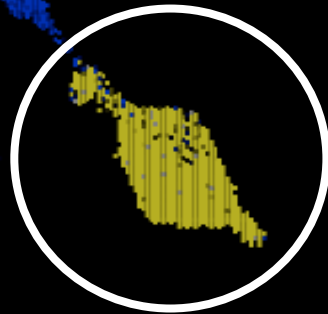
## ER Signaling "Weather Map"

GP7\_Estrogen  
signaling program

High



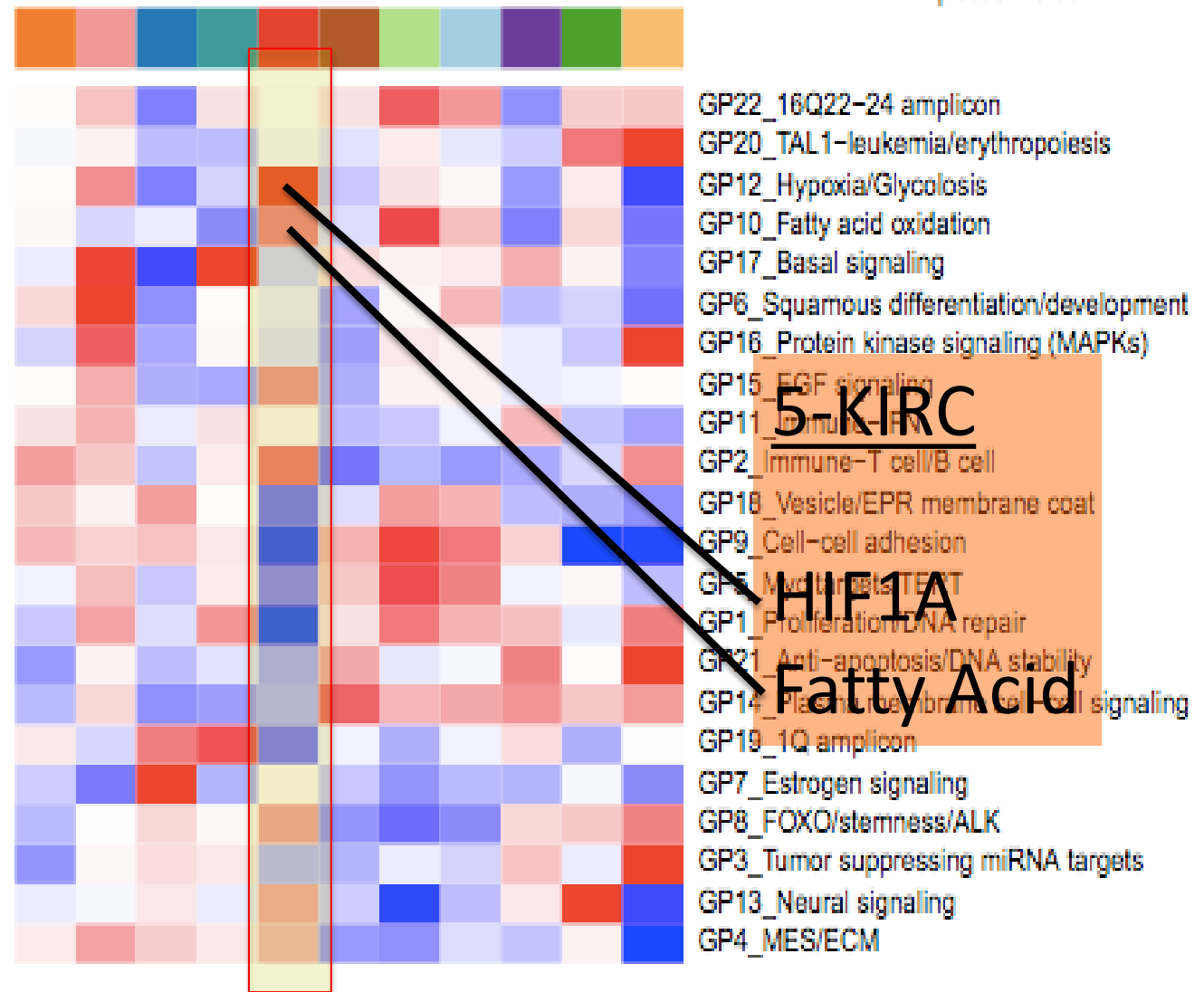
Low



BRCA Luminals Show High ER signaling



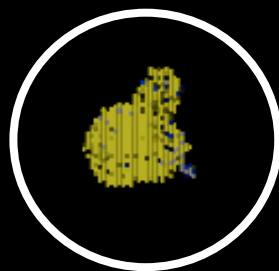
# Gene Programs: Surrogates of Integrated Subtypes



- 90% classification accuracy (LDA)

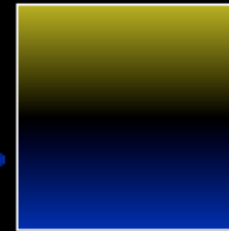
# Viewing Gene Programs on the TumorMap

HIF1A "Weather Map"



GP12\_Hypoxia/Glycolysis program

High

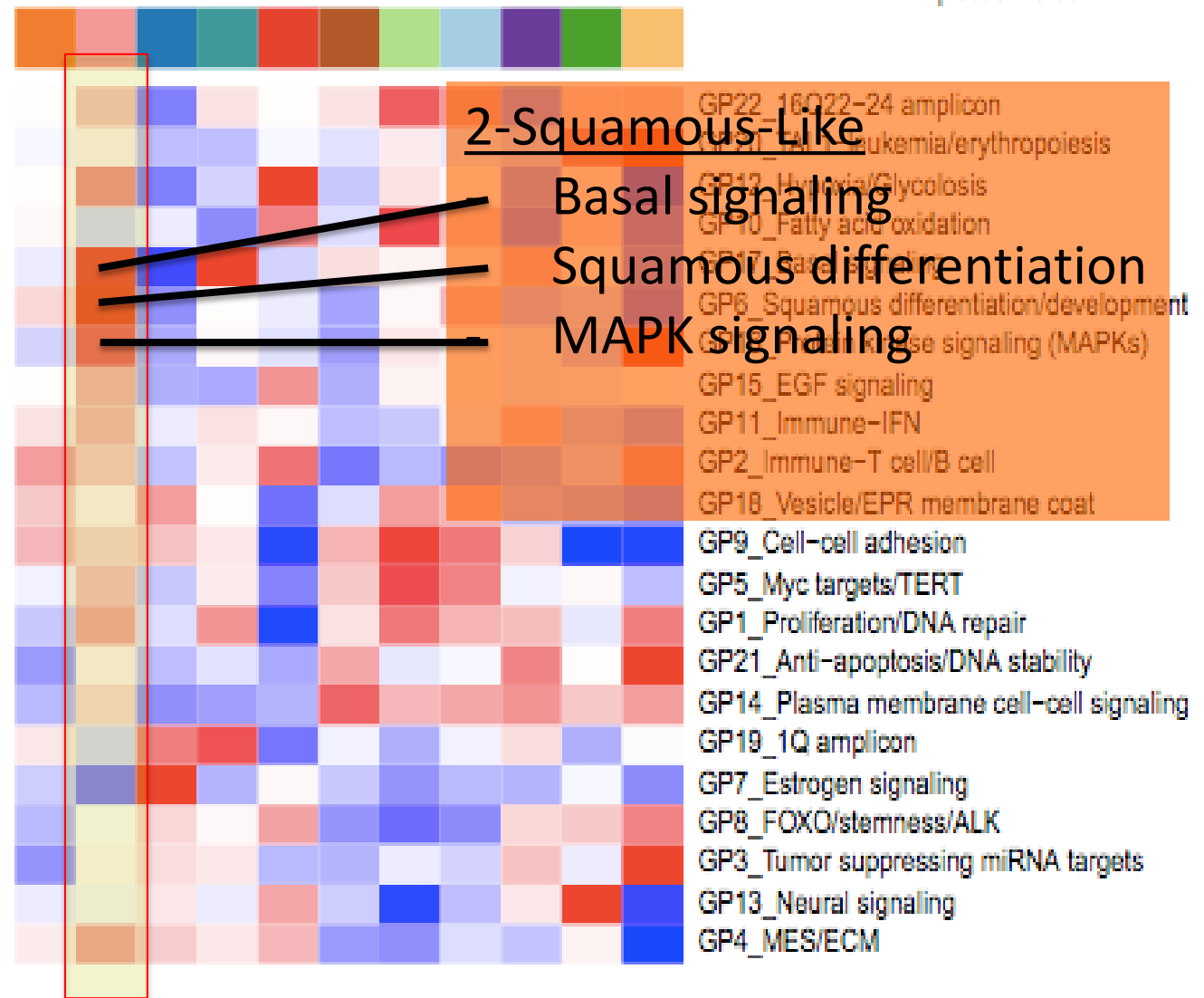


Low

KIRC w/ high hypoxia due to VHL mutations

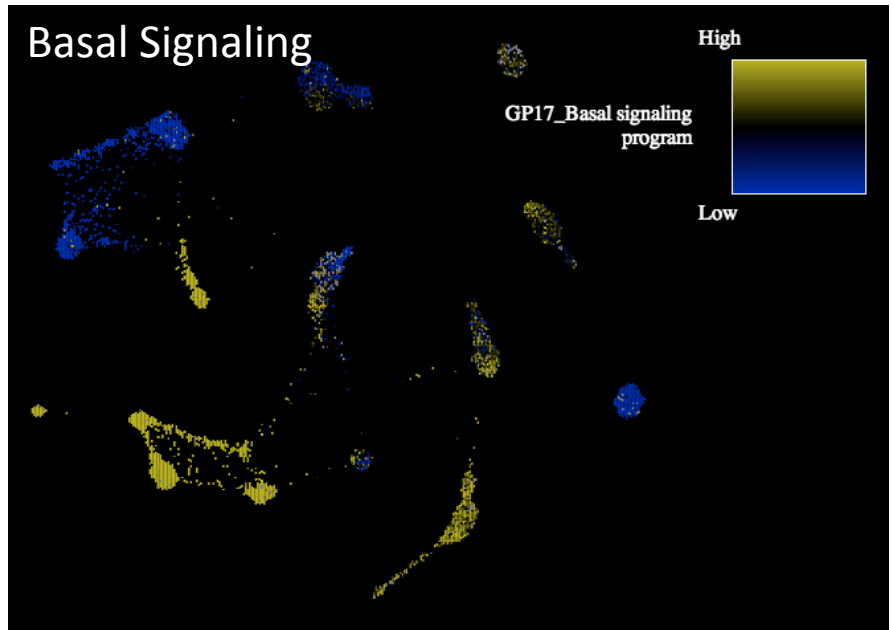
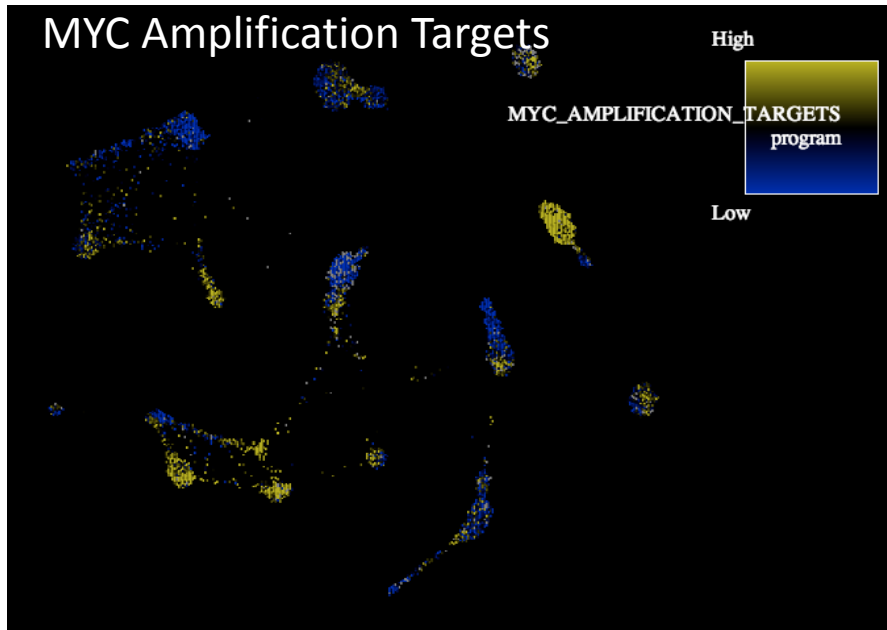
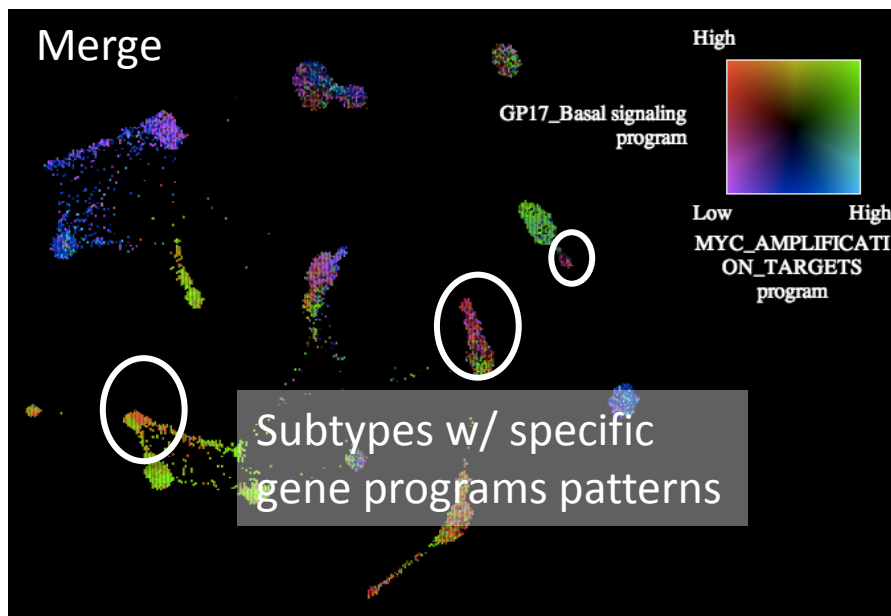
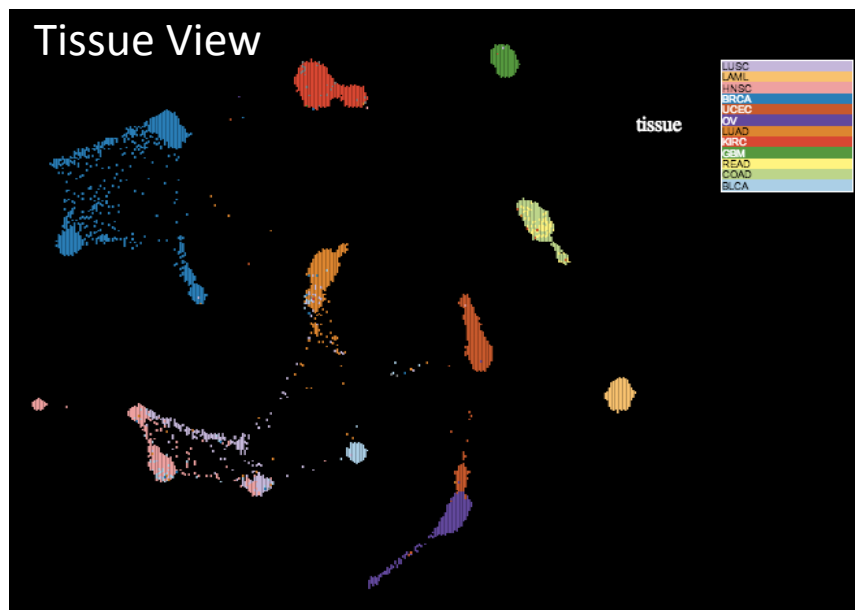


# Gene Programs: Surrogates of Integrated Subtypes



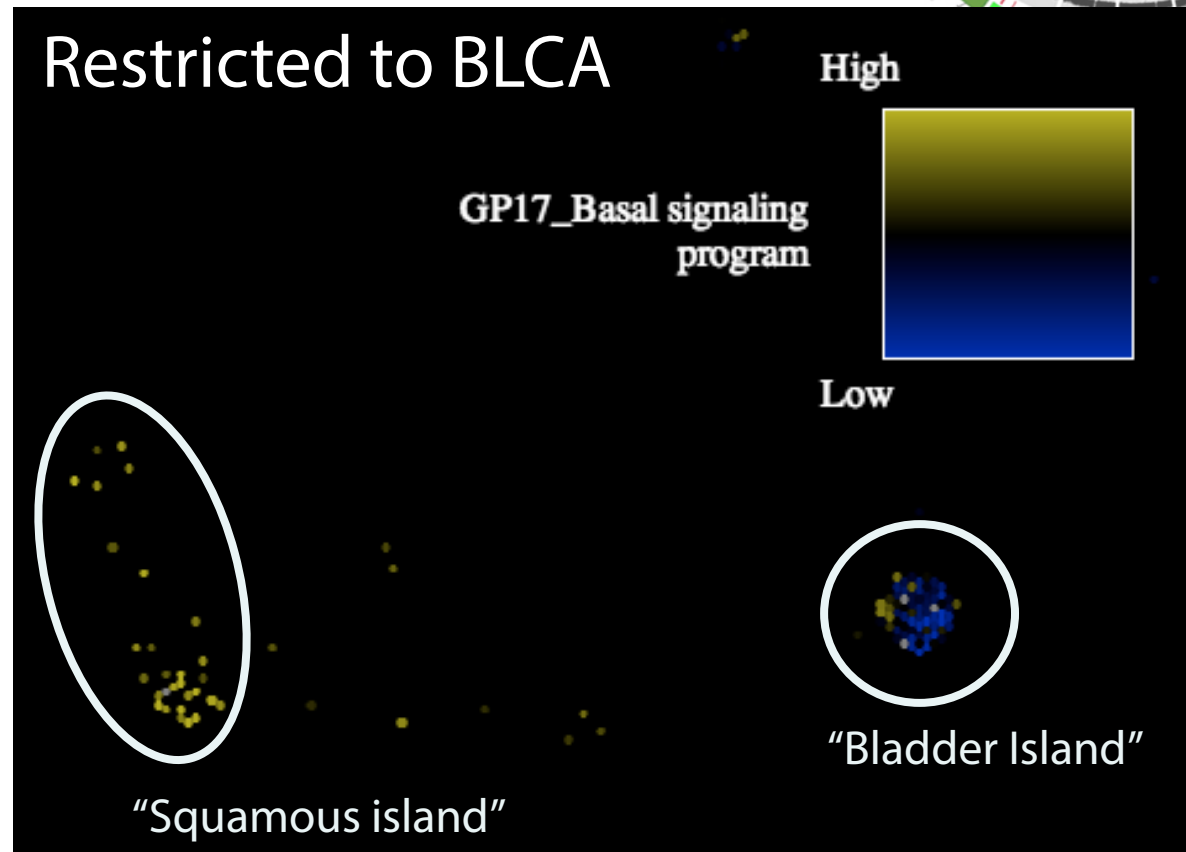
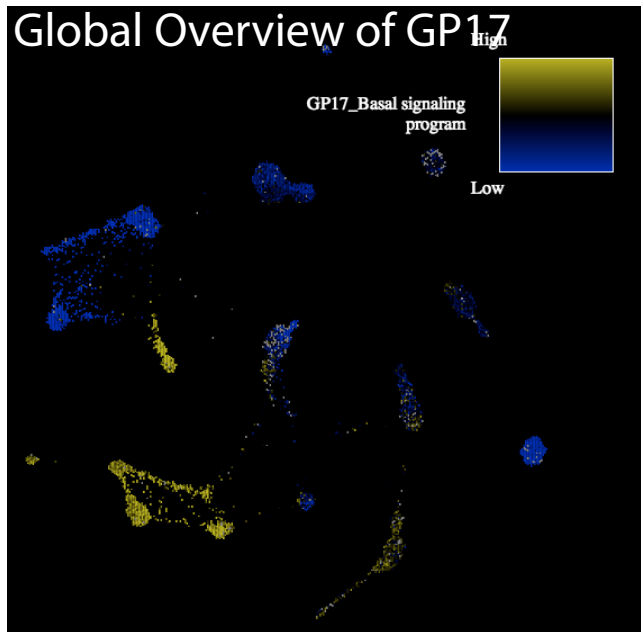
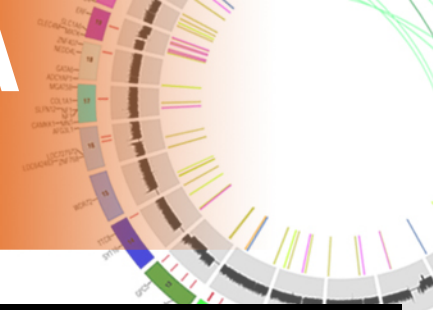
- 90% classification accuracy (LDA)

# Viewing Gene Programs on the TumorMap





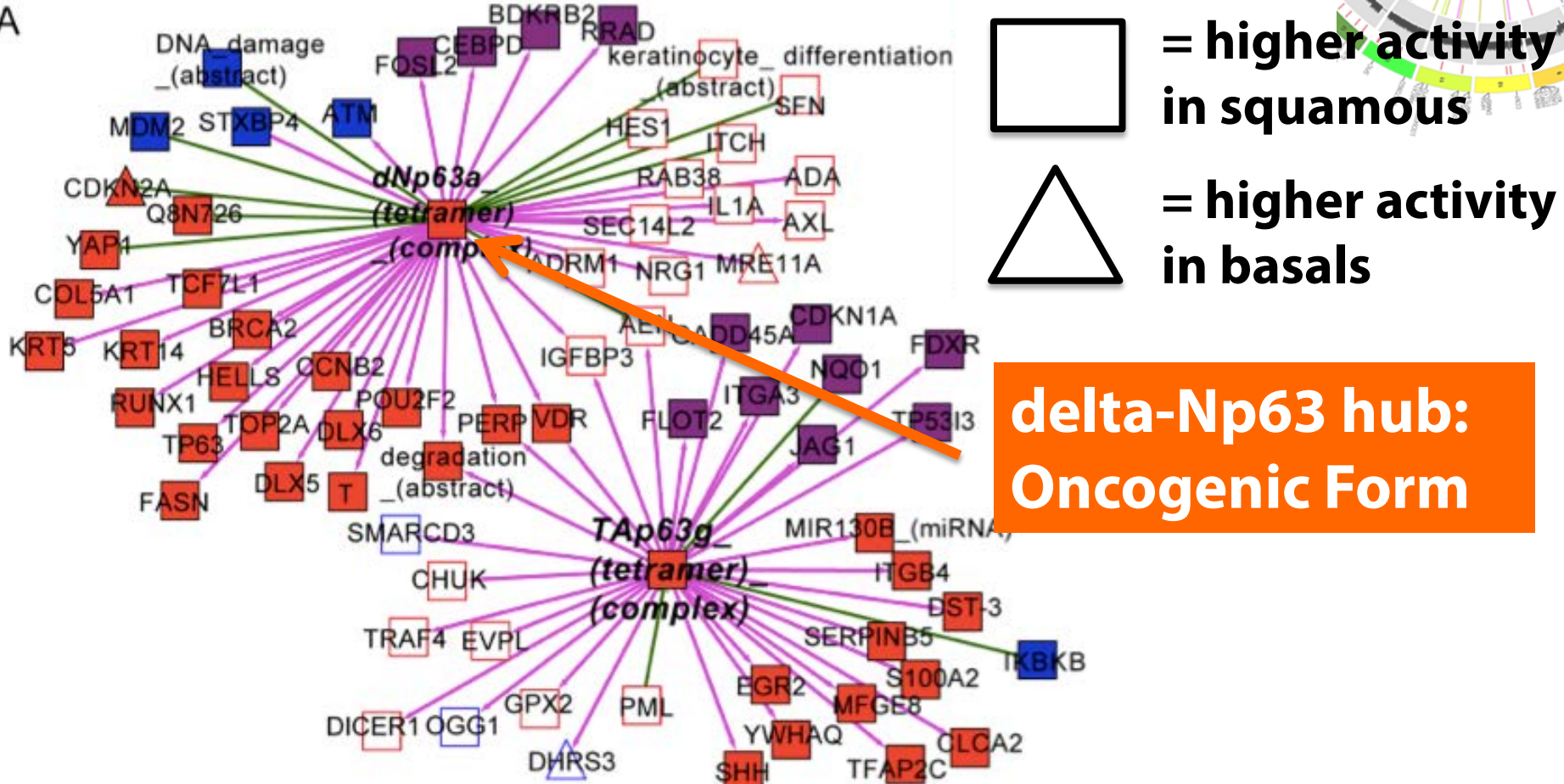
# Gene Program markers of BLCA divergence



- Squamous BLCA cases show higher GP17

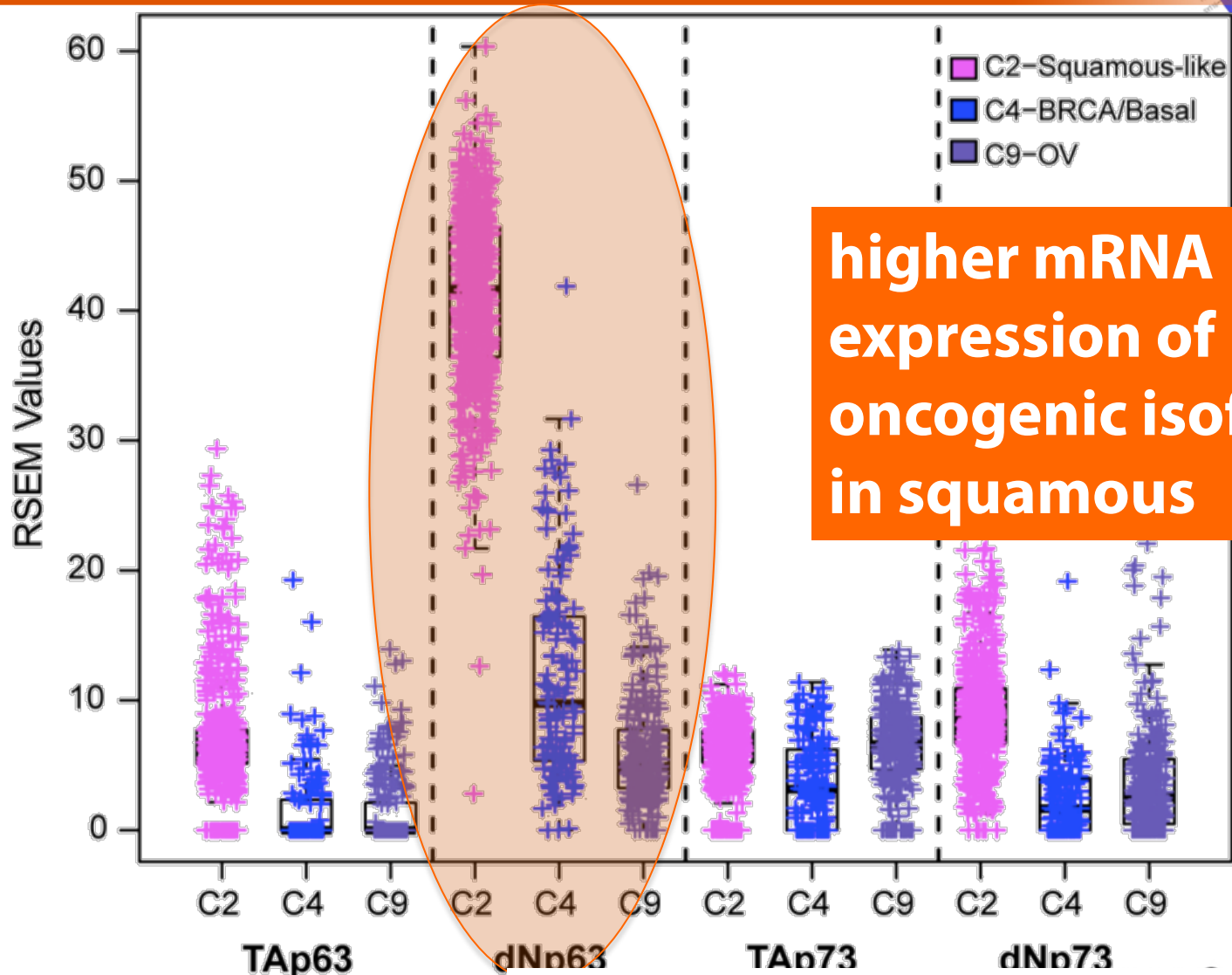
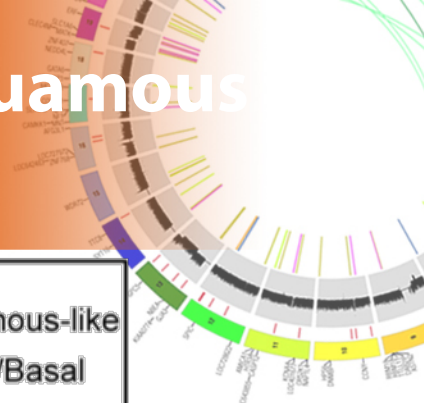
# Oncogenic Tp63 forms are more active in Squamous vs BRCA/Basal (or OV) TP53 mutants

A



- Most targets in network higher activity in Squamous (more “squares in diagram”)

# Oncogenic Tp63 forms are more active in Squamous vs BRCA/Basal (or OV) TP53 mutants



**higher mRNA expression of oncogenic isoform in squamous**

# Published as a Resource



 Synapse CONTRIBUTE to the CURE





Multi-platform integration of 12 cancer types reveals cell-of-origin classes with distinct molecular signatures ★

Synapse ID: syn2468297

Wiki

Files

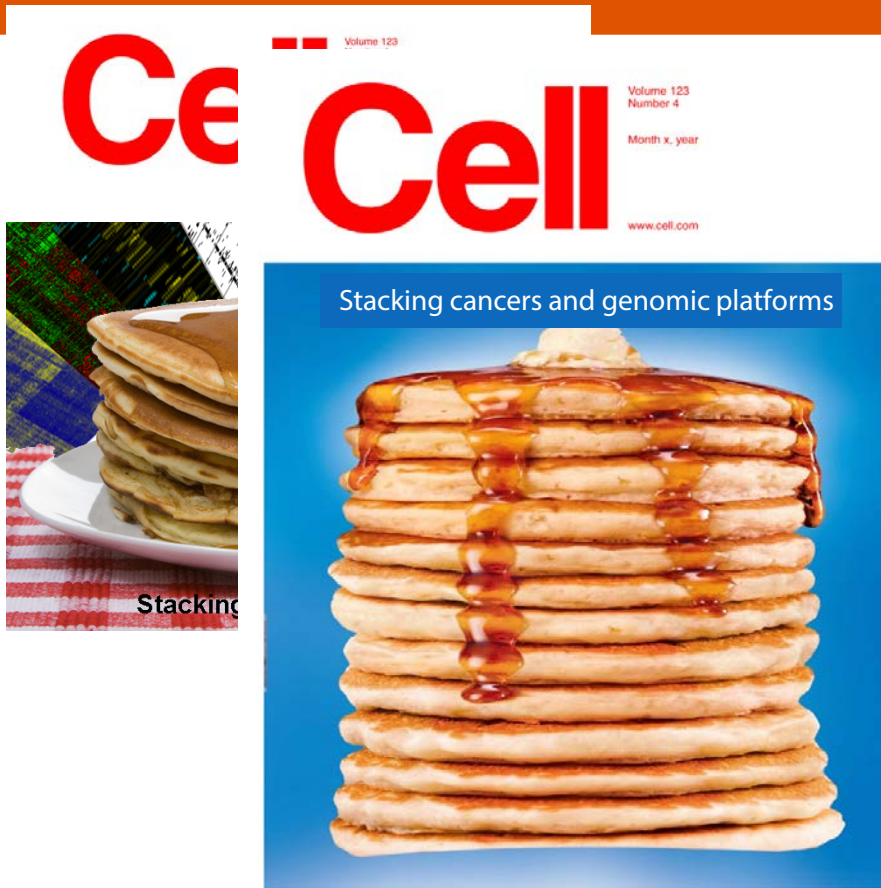
## Single Platform derived Clustering

platform	Source	Description
RNA		mRNA expression clustering of 12 Pan-Cancer tumor types.
miRNA		Unsupervised clustering of miRNA-seq data. (1) Schematic of a miRNA primary transcript (pri), the trimmed pre-miRNA (pre), reference miRBase 5p representation used. (2) From the NMF rank survey the profile of average silhouette width suggests a 15-group solution (gray triangle). (3) Consensus clusters consistently over 200 iterations. (4) NMF consensus clustering. Top to bottom: normalized abundance heatmap for 51 discriminatory miRNAs, a table of group number (c), number of samples (n) and average silhouette width (w) for each group, and for the overall set of 4229 tumor samples.
SCNA		Clustering of SCNAs across all tumors. SCNAs in tumors (vertical axis) are plotted by chromosomal location (horizontal axis). Tumors are clustered based on a 2.0 analysis of the entire set. The heatmap shows the presence of amplifications (red) and deletions (blue) throughout the genome. Colorstrips on the side indicate chromosome location.
Methylation		DNA methylation subtypes. Shown in the heatmap are DNA methylation beta values for 4,923 tumors (columns) of twelve tumor types at 2,204 CpG loci (rows).

- Datasets and subtypes provide pivot for further analyses
- All datasets hosted on a Synapse project page
  - Links to all relevant PanCan-12 datasets

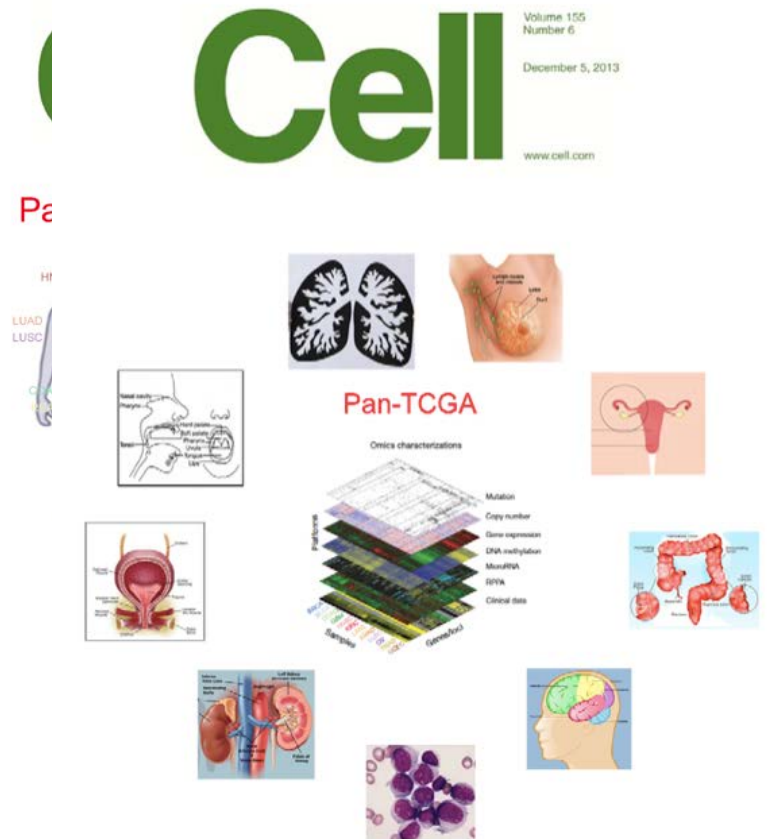


# To appear in *Cell*



Hoadley and Zhu

- Vote for your favorite cover image
- Pancakes, Part II – Original Pancake House – 7703 Woodmont Ave.

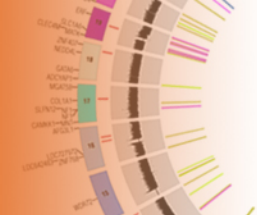


Zhong Chen



# Summary

- Analysis of 12 tumor types w/ 6 platforms display tissue-of-origin as dominant
- Integrated analysis reveals 11 major groups, with some tumor types merging together (HNSCC, Lung Squamous, some Bladder) and others separating (breast luminal vs. Basal-like)
- 1:10 re-classified cases based on the map.
  - Rate similar to EGFR mutations in NSCLC cancers
  - Convergences and Divergences of tissues
- Intriguing subtype-specific differences in TP53 pathway activity between OV, BRCA-Basals, and the Squamous tumors
- Classification adds prognostic information independent of tissue and stage.
  - E.g. COCA clusters define clear prognostic groups for BLCA
- Clearly more investigation will be beneficial; especially those that subtract away tissue-of-origin signals (see Verhaak paper)





# TCGA Pan-Can AWG

**MDACC**

**John Weinstein**  
**Rehan Akbani**

Lauren Byers  
Han Liang  
Roel Verhaak  
Gordon Mills



**MSKCC**

Chris Sander  
Giovanni Ciriello  
Anders Jacobsen

**Baylor**

Mark Hamilton  
David Wheeler

**USC**

Peter Laird  
Hui Shen



**Broad**

Andrew Cherniak  
Matt Meyerson  
Gaddy Getz  
Rameen Beroukhim  
Scott Carter  
Travis Zack  
Mike Lawrence  
Angela Brooks



**Sage Bionetworks**

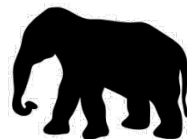
Adam Margolin  
Larsson Ohmberg

**Harvard**

Raju Kucherlapati

**ISB**

Ilya Shmulevich  
Sheila Reynolds



**UCSC**

Vlado Uzunangelov  
Sam Ng  
Evan O. Paull  
Kyle Ellrott  
David Haussler  
Jing Zhu



**UPF Barcelona**

Nuria Lopez-Bigas  
**David Tamborero**  
Abel Gonzalez-Perez

**UNC**

**Katherine Hoadley**  
**Chuck Perou**



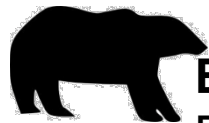
**NCI / NHGRI**

**Julia Zhang**  
Zhong Chen  
Carter Van Waes



**Brown**

Ben Raphael  
Max Leiserson



**UBC**

Gordon Robertson  
Andy Chu



**UCSF/Buck Inst**

**Chris Benz**  
**Eric Collisson**  
**Christina Yau**  
**Denise Wolf**

**WashU**

Li Ding  
Cyriac Kandoth  
Beifang Liu  
Mike McLellan

# UCSC Integrative Genomics Group

Kyle Ellrott



Sam Ng



Robert Baertsch



Evan Paull



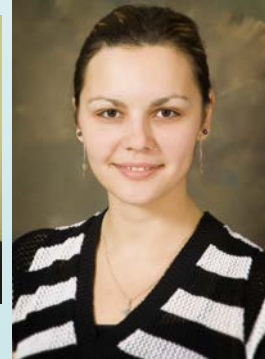
Yulia Newton



Adrian Bivol



Olena Morozova



Dan Carlin



Marcos Woehrmann



Artem Sokolov



Ted Goldstein



Chris Wong



Vlado Uzunangelov



Kiley Graim



CENTER FOR BIOMOLECULAR SCIENCE & ENGINEERING  
promoting discovery and invention for human health and well-being





**David Haussler**

# Acknowledgments

**Chris Benz,**



**Buck Institute for Aging**

- Christina Yau

**Collaborators**

- Li Ding, WashU
- Matthew Eills, WashU
- Elaine Mardis, WashU
- Rick Wilson, WashU
- Cyriac Kandoth, WashU
  
- Joe Gray, OHSU
- Laura Heiser, OHSU
- Nuria Lopez-Bigas, UPF
- Abel Gonzalez, UPF
- Adam Margolin, Sage
- Larsson Omberg, Sage

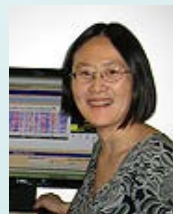
**UCSC Cancer Genomics**

- Adam Ewing
- Chris Wilks
- Sofie Salama
- Steve Benz

**UCSC Genome Browser Staff**

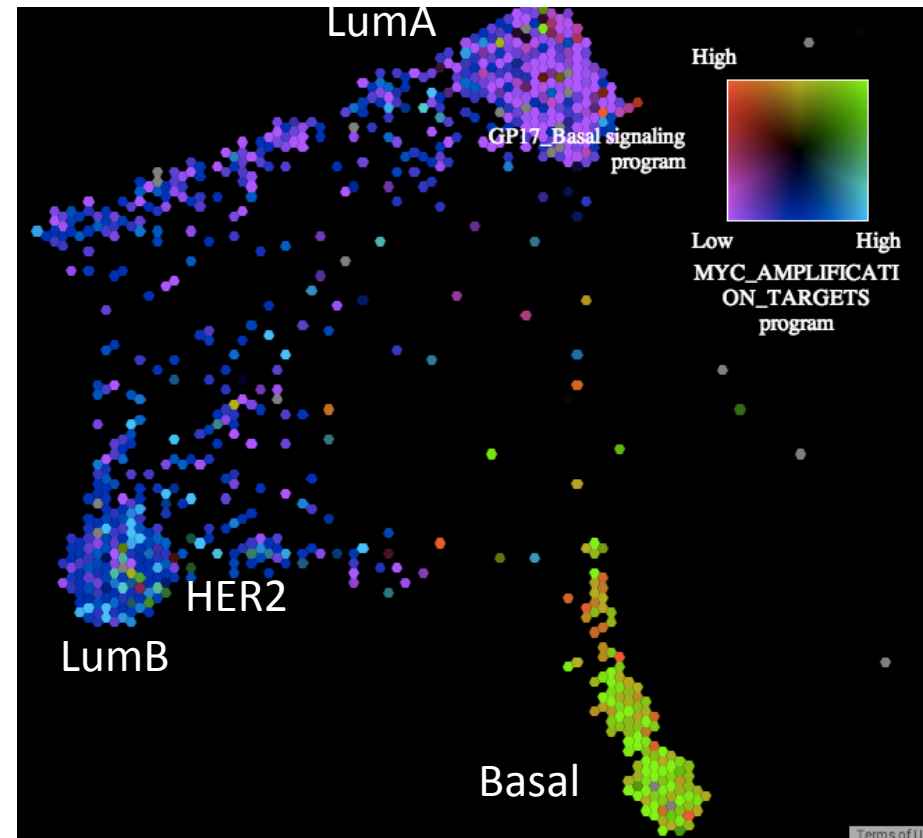
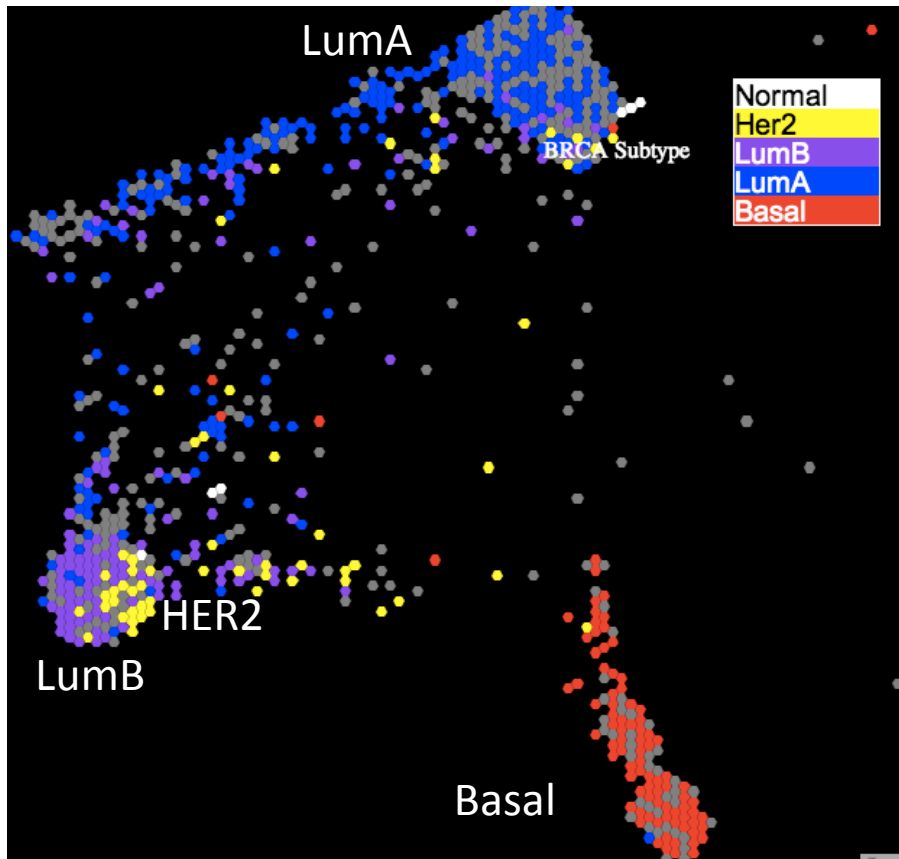
- Mark Diekins
- Melissa Cline
- Jorge Garcia
- Erich Weiler

**Jing Zhu**



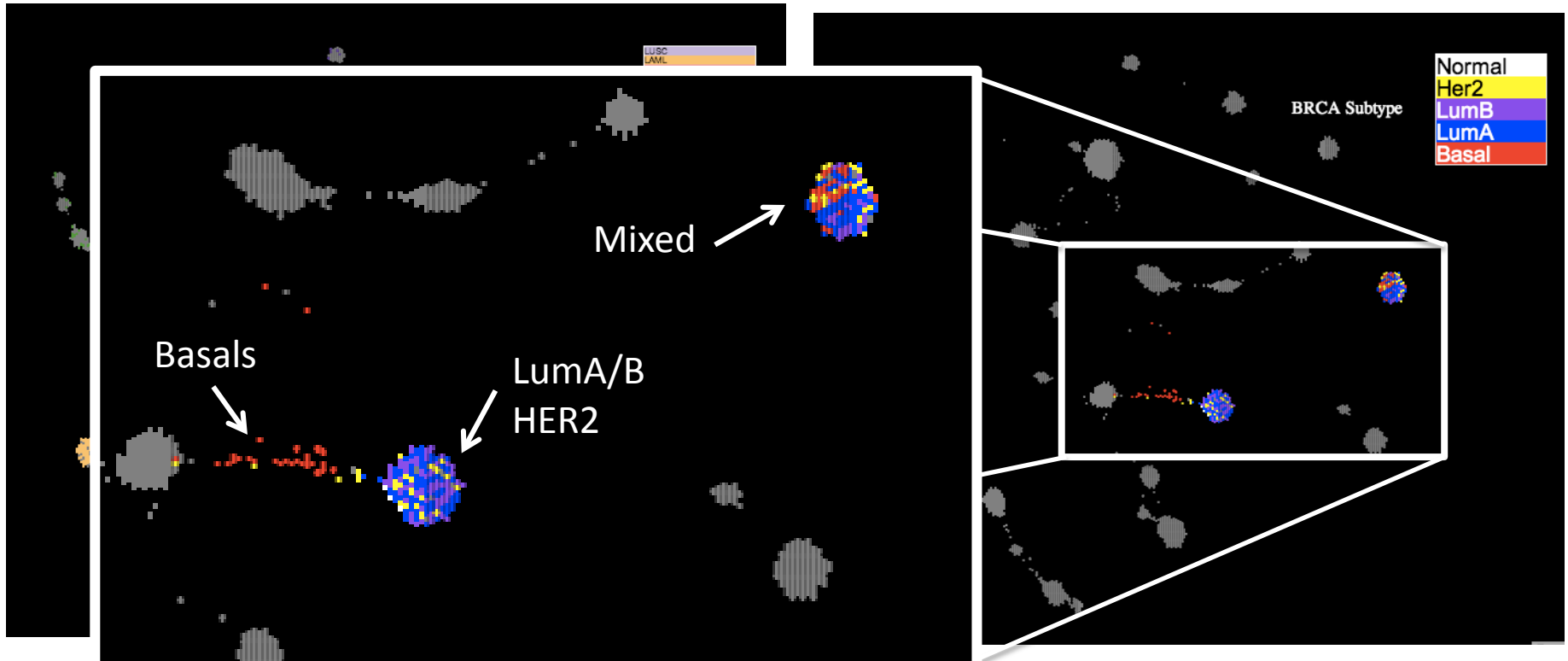
# Supplemental

# LumB's w/ Different Gene Program Expression



- LumB BRCA (and HER2's) have high MYC but low basal signaling.

# DNA Methylation View of BRCA Subtypes



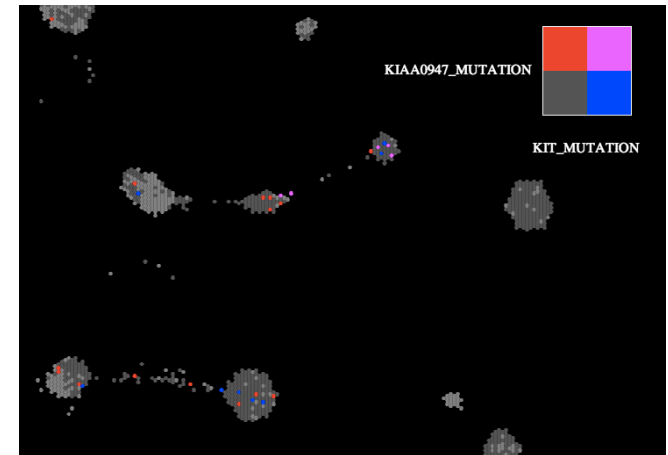
- 3 distinct DNA methylation subtypes revealed:
  - One all BRCA-basals, one all luminals and HER2, and one mixed



# BRCA DNA Methylation Subtypes



Slightly more TP53 in BM3 (P<0.04)



- Created a contrast between BRCA methylation subtypes {1,2} vs {3}.
- SCNA in {1,2}: deletions in CNTN5 (11q22.1) RB1 (13q14.2), ITM2B (13q14.3)
- SCNA in {3}: chr3-12484849-12485147 amplification
- Slightly more TP53 in BRCA-Methylation subtype 3 (BM3) (P<0.04).
- More mutations in ORF KIAA0947 in BM1 and BM2.
  - ORF interacts with transcriptional elongation proteins ELL, ELL2, EAF1, MED26 (mediator). Cell 2011 – Takahashi et al.

# BRCA Methylation Subtypes (indep of transcriptional subtypes)



- Correcting for Subtype.
- Restrict only to the luminals. Enriched in BRCA-Meth-2 vs 3: RB1 deletions ( $P < 0.0076$ )
- Restrict to basals. Only found one amplicon on 6p21 (POLR1C, POLH, KLK2, CUL9, ...)
- POLH could be *very* interesting as it copies past thymidine dimers and causes hypermutation rates. Do the subset of patients have higher mutation rates?

# BLCA Subtypes

- Bladder cancers are split into two subtypes on the map.
- One set tightly associated with squamous cancers of HNSC and LUSC
- The other with the rest of the BLCA

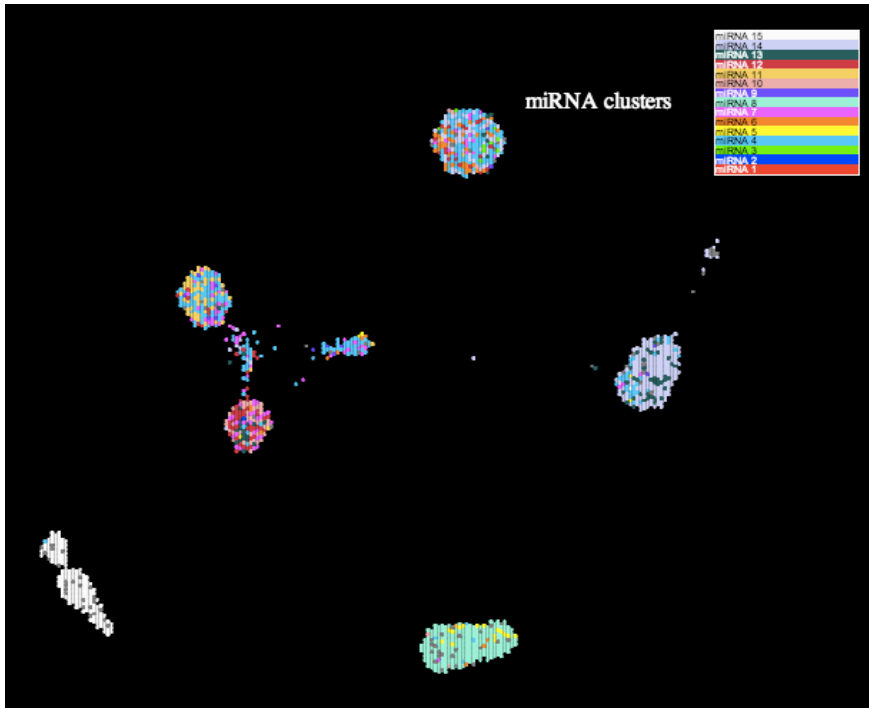
# BLCA non-squamous genomic determinants



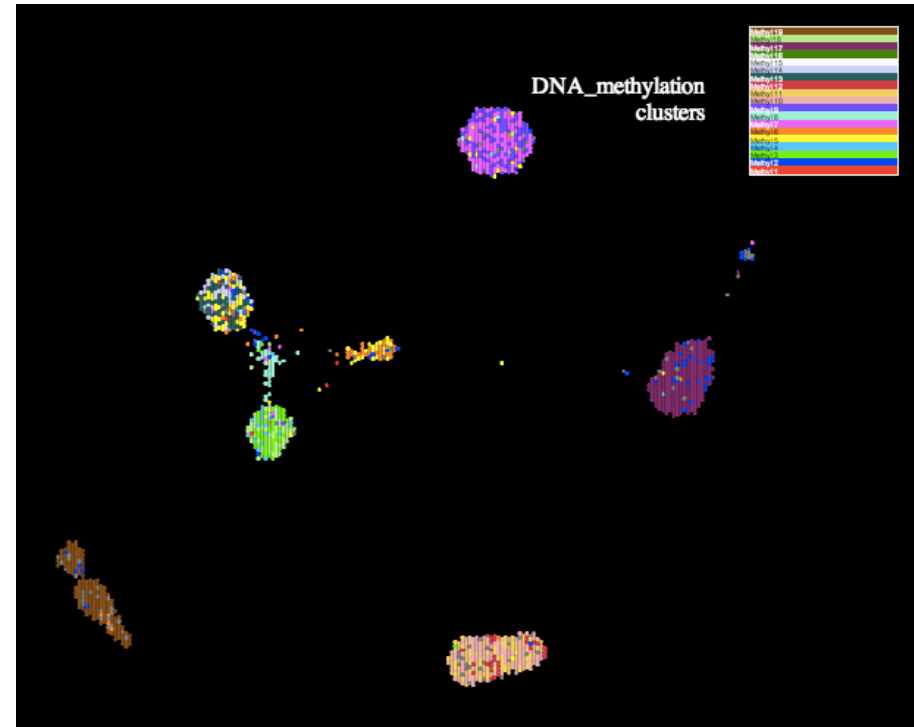
- The non-squamous BLCA tumors are characterized by mutations in several genes including ERBB2, BAP1, STAG2, PDGFRA, and the ORF KIAA0947.

# Closer Look at miRNA subtypes

miRNA subtypes  
on miRNA map

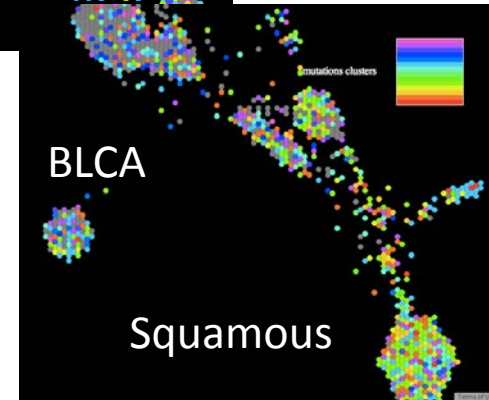
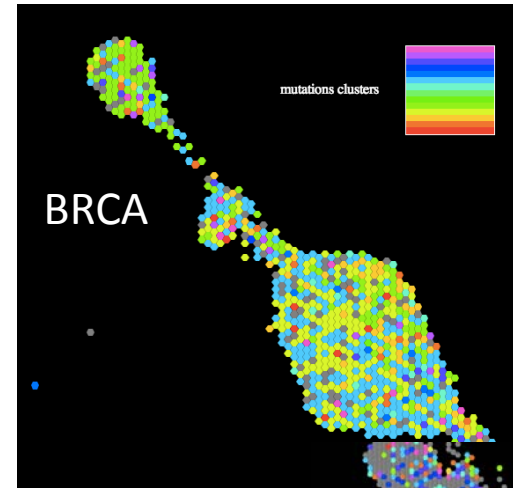
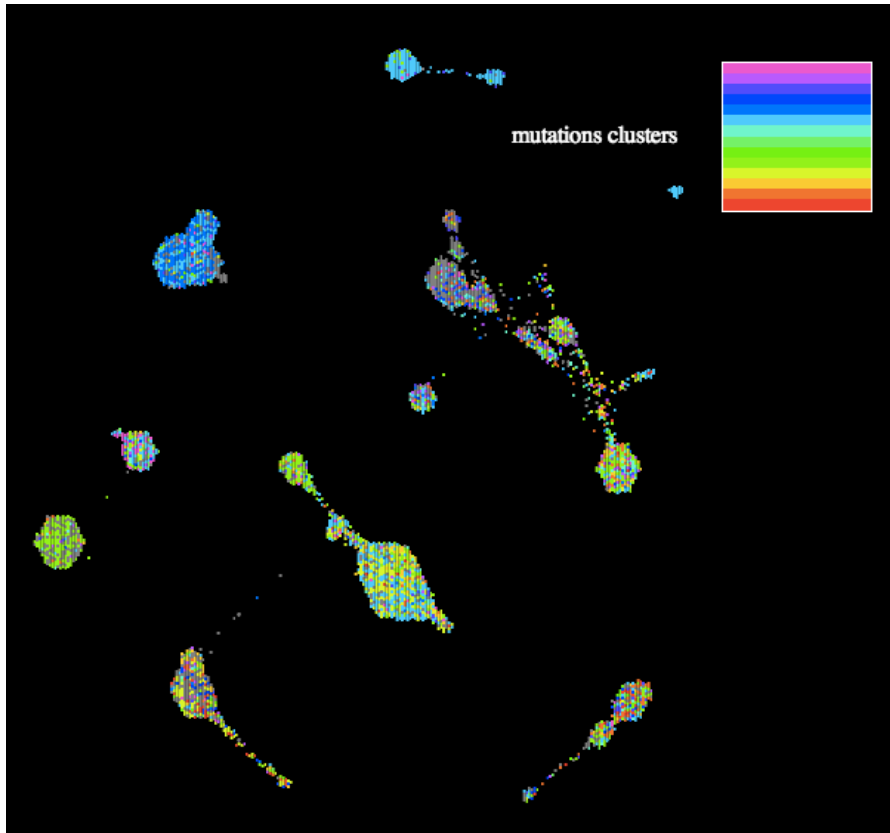


DNA methylation subtypes  
on miRNA map



- The DNA methylation pancan subtypes correspond better to the TumorMap miRNA clusters than the miRNA pancan subtypes!

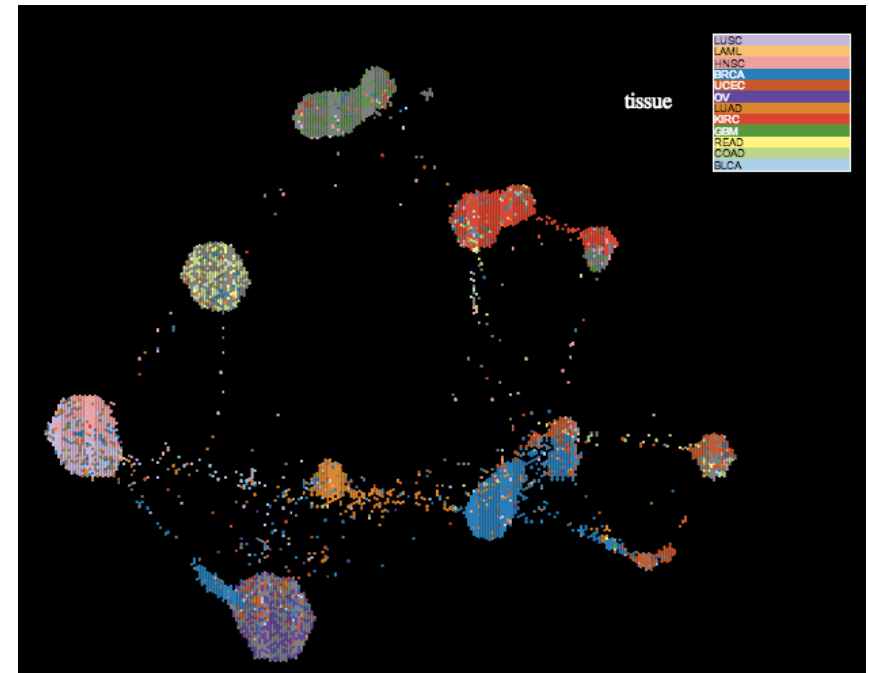
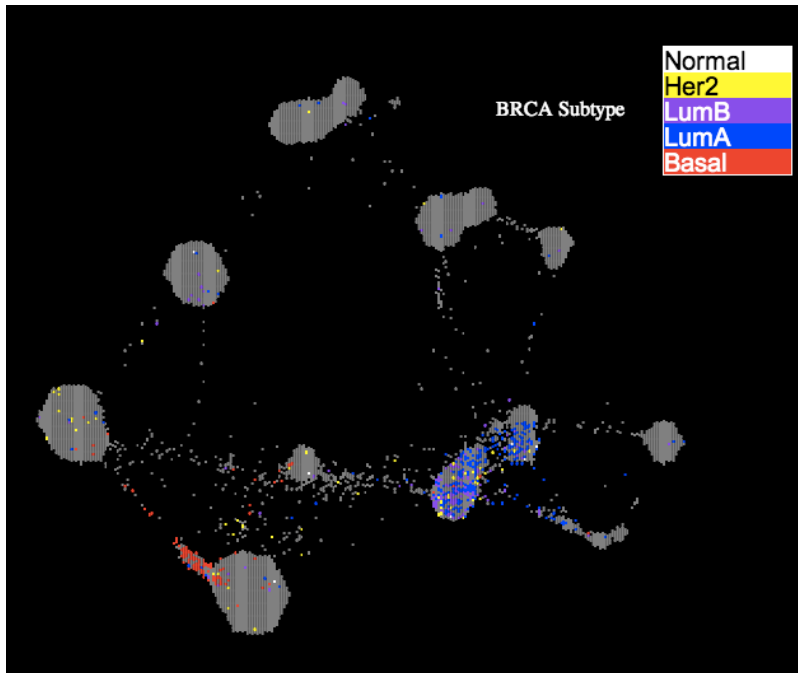
# Mutation Clusters: To Be Continued



- Clustering by mutations gives more tissue-orthogonal clustering.
- But mutation clustering is difficult
- This solution used prior pathway knowledge to help unify evidence.

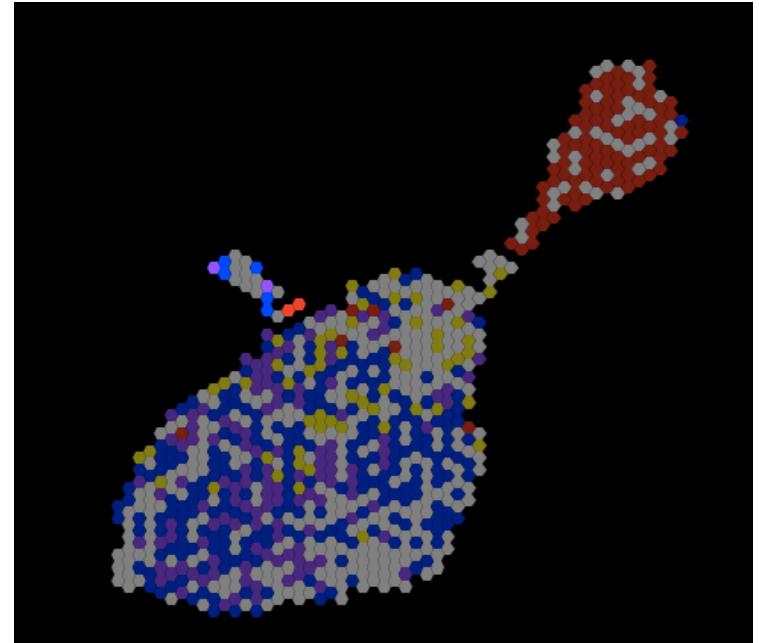
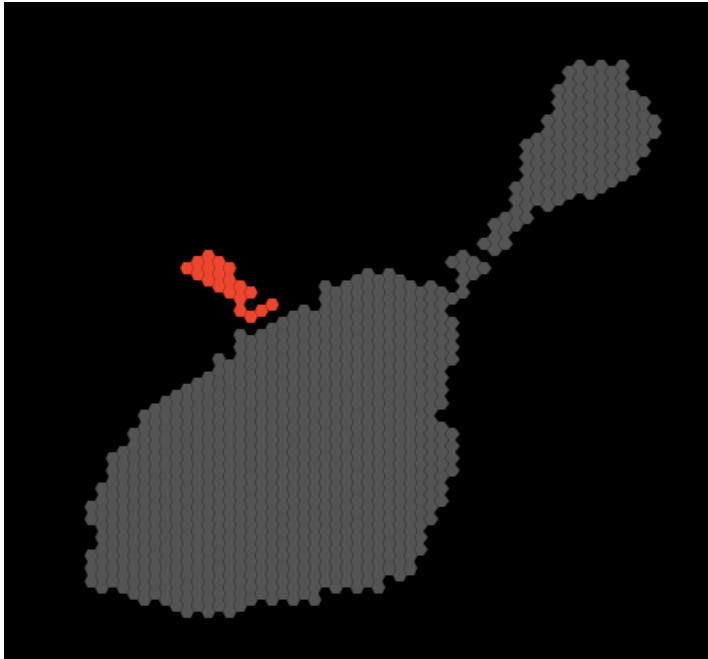


# Copy Number View of BRCA Subtypes



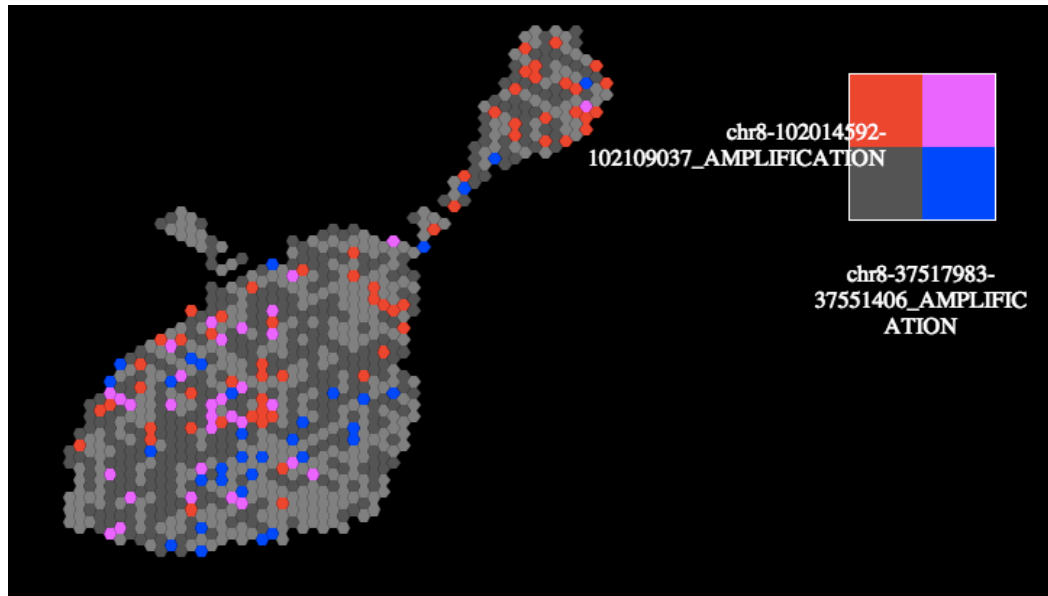
- BRCA transcriptional subtypes also reflected in the somatic copy number data.
- BRCA basals similar to OV.
  - And to a lesser extent, the Squamous group (HNSC/LUSC)
- BRCA luminals similar to UCEC and LUAD

# Interesting minor BRCA subtype



- A luminal area distinct from the major luminal BRCA area.
- What distinguishes this subtype?

# Minor BRCA subtype



- Absence of amplifications
  - chr8p11 amplifications absent in the subtype
    - Genes: LETM2 (8p11.23), WHSC1L1 and POLB (8p11.2)
  - chr11q14 amps absent
    - Genes: ALG8 (11q14.1), KCTD14 (11q14.1)
- Absence of mutations
  - GATA3, MLL3, MAP2K4, PTEN, NCOR1, SYNE1, DMD, PIK3R1, NF1, SPEN, BRCA2, CTCF, TBX3,