NATIONAL CANCER INSTITUTE

CSH Cold Spring Harbor Laboratory

Computational Framework

for Single-Cell Genomics of Tumors

U01CA188590 to A. Krasnitz and M. Wigler

- A template for integrating bench genomics, computation and pathology: an illustration from prostate cancer
  - Detection
  - Inference
  - Visualization
  - Clinical utility
- Early detection of cancer from blood
  - Survey of potential
  - A roadmap to validation

Cyto-pathological assessment of prostate cancer

Current practice

High PSA (>4ng/ml), suspicious DRE

↓

TRUS/CT/MRI

↓

Core or FNA biopsy

↙ ↘

Localized, Gleason<6
Or
Localized, Gleason≤6, >60 yrs
Surveillance

Else aggressive treatment, RP if possible

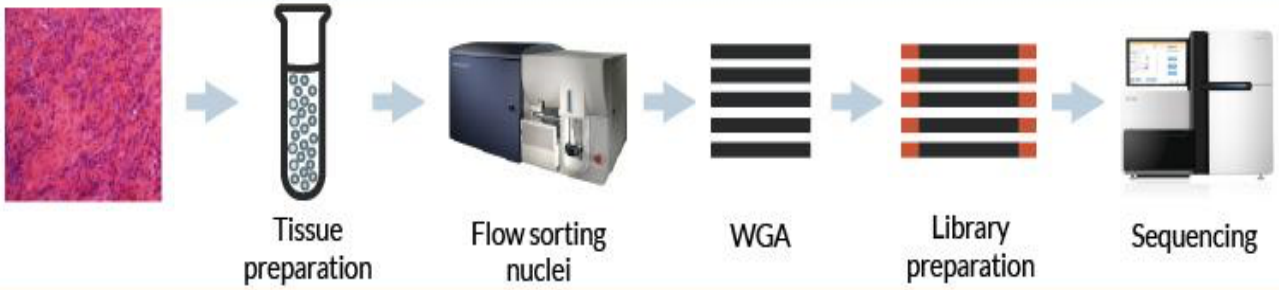
Shortcomings of conventional pathology:

- 65% probability that any 2 pathologists disagree by ≥ 1 unit of Gleason score.
- Differing scores on core vs. post-RP biopsies.
- Some of low-scoring cases may be aggressive due to subclonal cell populations that go undetected.

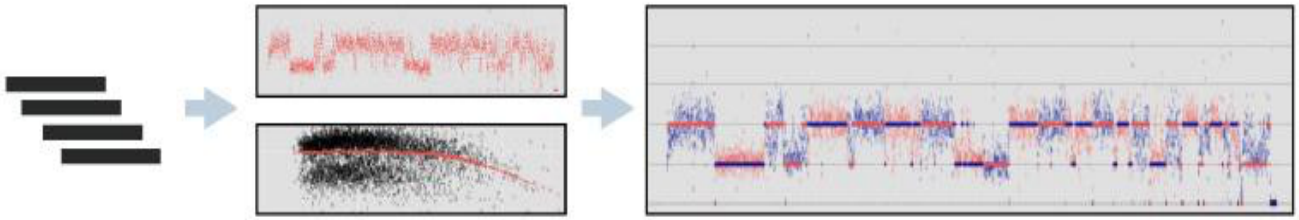
Case	Age	Sample[$]	Sectors	Gleason Score Biopsy	Gleason Score Final*	Proportion of sectors with pathology[%]	Highest Involvement of Cancer[+]	Mean Involvement Of Cancer[>]
NYU003.Benign.1	47	PBXW	13	Benign	NA	0/13	0	0
NYU002.Pin.1	72	PBXW	13	HGPIN	NA	0/13	0	0
COR002.GS6.1	62	TCRP	5	6 (3+3)	6 (3+3)	2/5		
NYU005.GS6.2	64	PBXW	14	7 (3+4)	6 (3+3)[#]	4/14	30	5
NYU001.GS7.1	63	PBXW	14	7 (4+3)	7 (3+4)	8/14	100	40
NYU007.GS7.2	65	PBXW	13	6 (3+3)	7 (3+4)[^]	1/13	30	2
NYU010.GS7.3	79	PBXW	15	7 (3+4)	NA	6/15	90	11
NYU004.GS7.4	75	PBXW	14	8 (4+4)	7 (4+3)[#]	6/14	100	23
NYU011.GS7.5	63	PBXW	10	7 (4+3)	7 (4+3)	5/10	60	14
COR001.GS9.1	77	TCRP	6	9 (5+4)	9 (5+4)	4/6		
COR003.GS9.2	80	TCRP	5	8 (4+4)	9 (4+5)[^]	3/5		
Median Age	65	Total	122	-----	-----	39/122		

Can single-cell genomic profiling help desambiguate pathology?
In particular, clones of cells with major genomic alterations ➔likely aggressive malignancy.
Can we detect them?

Single Nucleus Sequencing

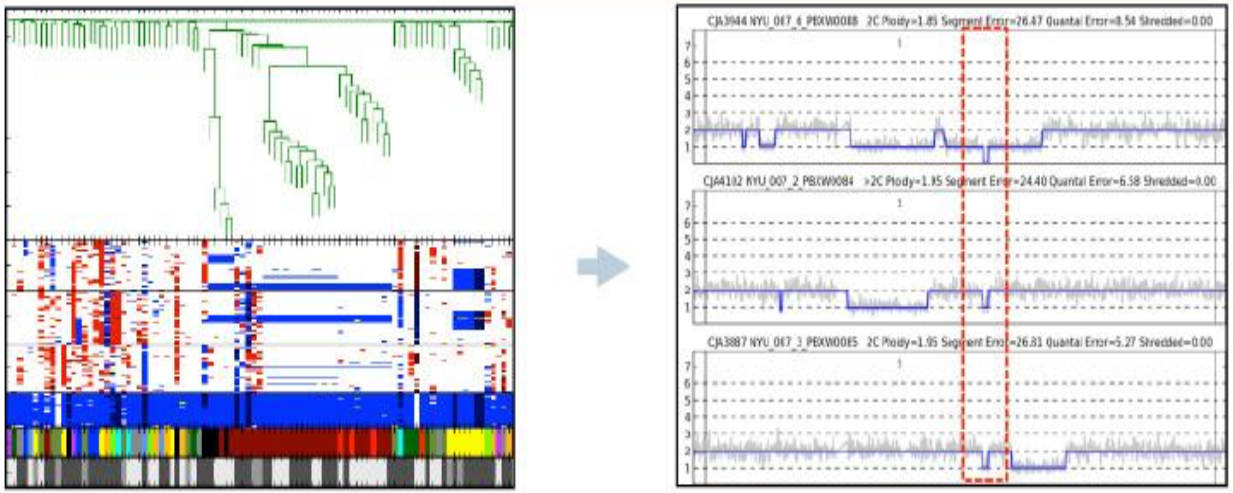


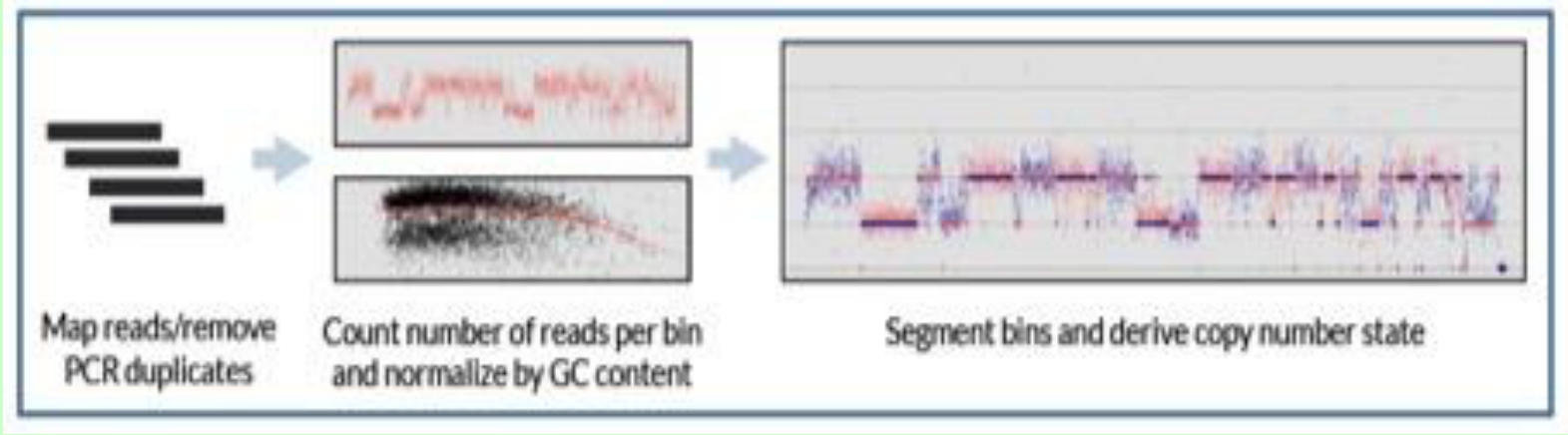
Data Processing



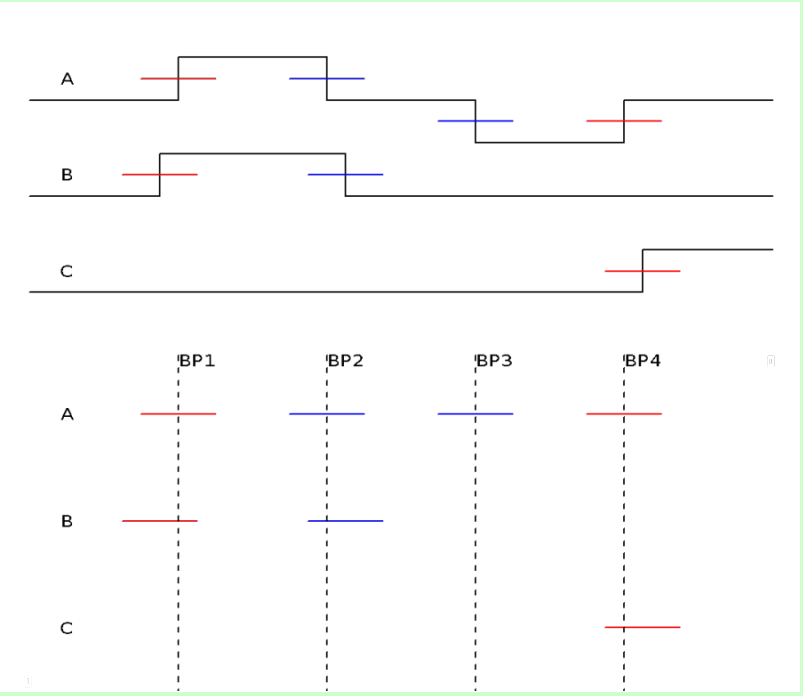
Data Integration and Visualization

Single-Cell Genomics Viewer (SCV)





Profiles as collections of “smeared” break-points



→

Break-point incidence table

	A	B	C
BP1	+	+	-
BP2	+	+	-
BP3	+	-	-
BP4	+	-	+

↓

B

A		+	-
	+	2	2
	-	0	0

↓

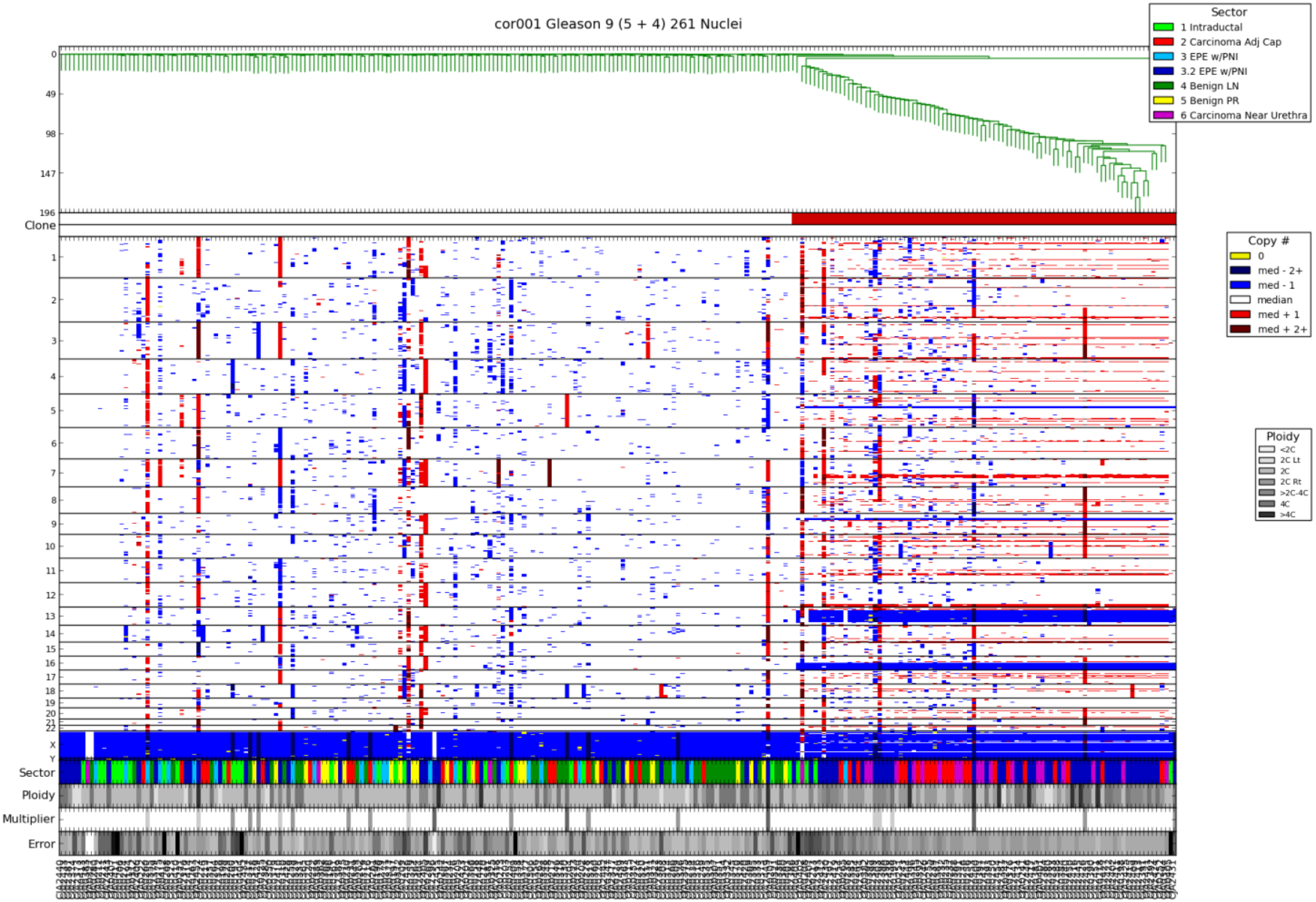
Fisher’s exact test: is this table surprisingly diagonal?

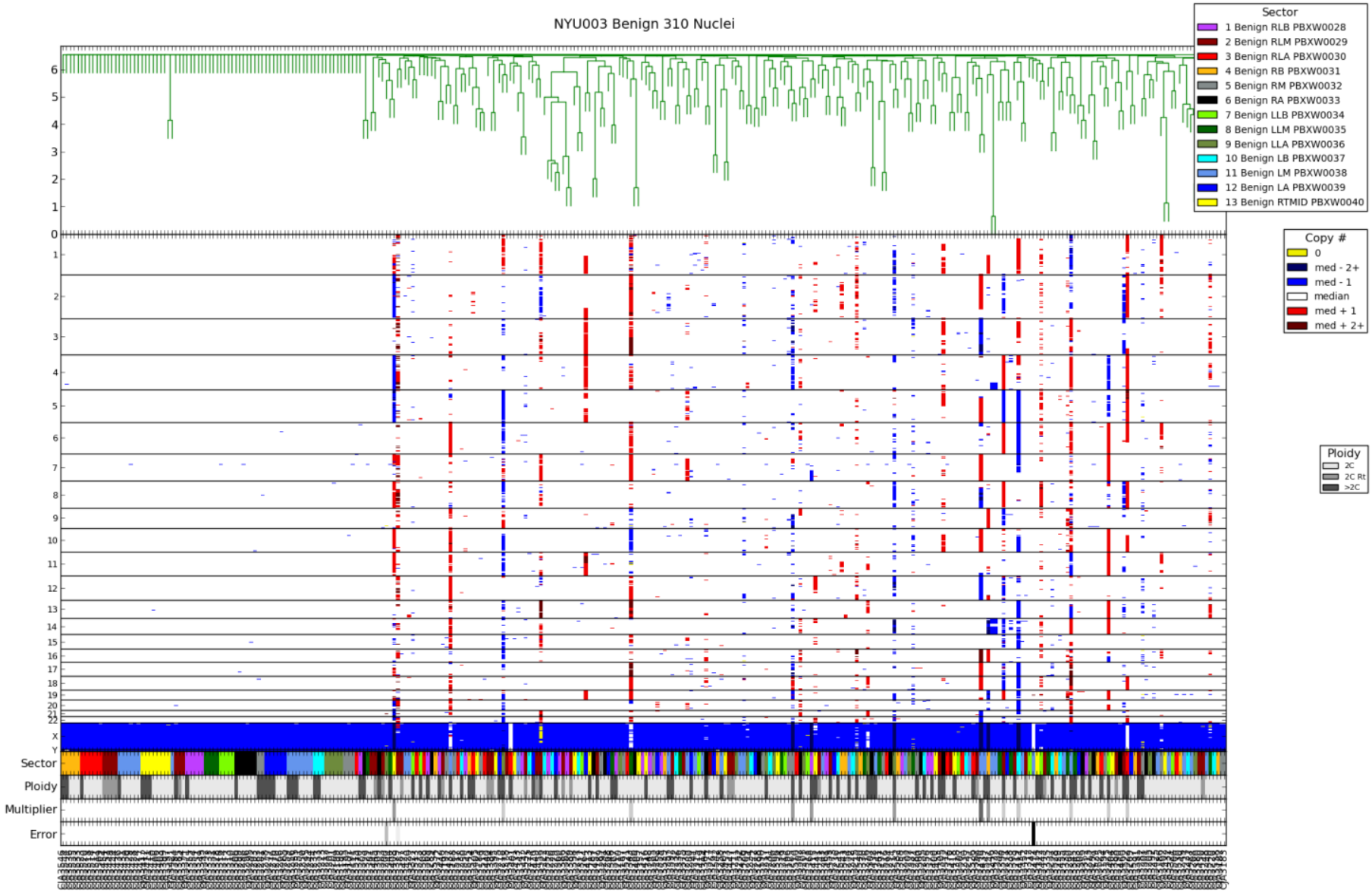
←

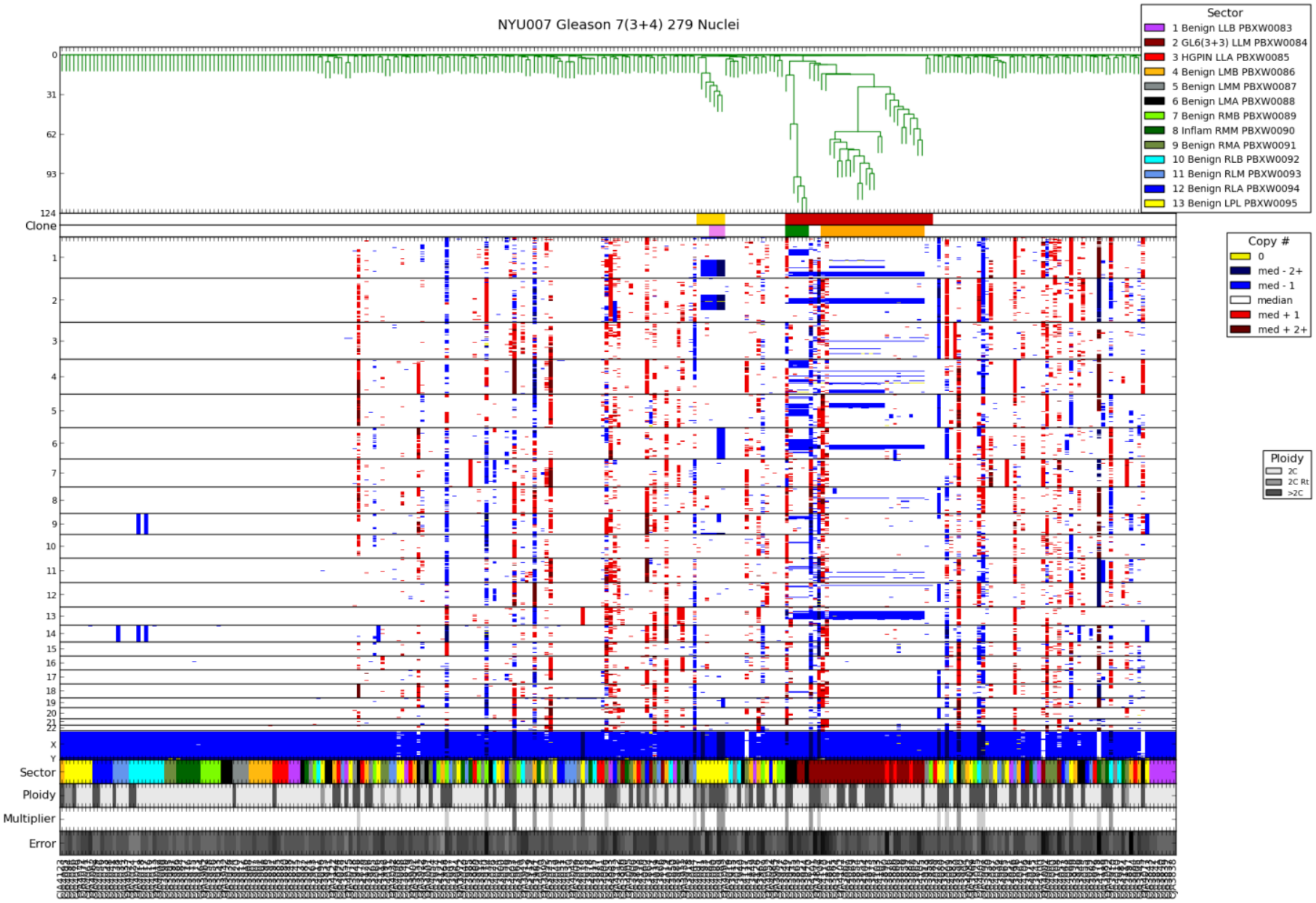
Use Fisher’s p-values as pairwise dissimilarities

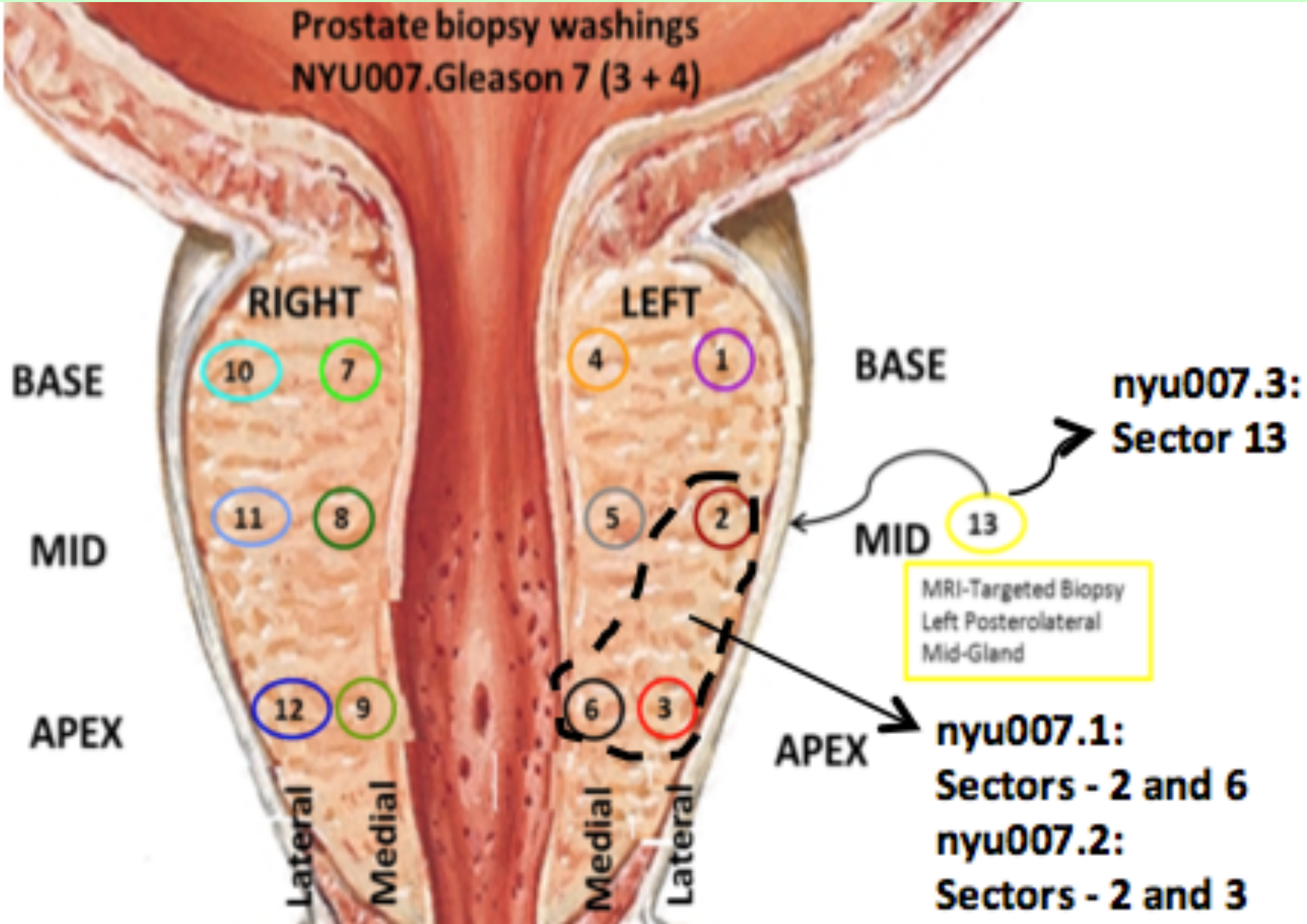
What is Single-Cell Genome Viewer?

https://github.com/KrasnitzLab/SCGV


cor001 Gleason 9 (5 + 4) 261 Nuclei
Sector
1 Intraductal
2 Carcinoma Adj Cap
3 EPE w/PNI
3.2 EPE w/PNI
4 Benign LN
5 Benign PR
6 Carcinoma Near Urethra
Copy #
0
med - 2+
med - 1
median
med + 1
med + 2+
Ploidy
<2C
2C Lt
2C
2C Rt
>2C-4C
4C
>4C
Clone
X
Y
Sector
Ploidy
Multiplier
Error


NYU003 Benign 310 Nuclei
Sector
1 Benign RLB PBXW0028
2 Benign RLM PBXW0029
3 Benign RLA PBXW0030
4 Benign RB PBXW0031
5 Benign RM PBXW0032
6 Benign RA PBXW0033
7 Benign LLB PBXW0034
8 Benign LLM PBXW0035
9 Benign LLA PBXW0036
10 Benign LB PBXW0037
11 Benign LM PBXW0038
12 Benign LA PBXW0039
13 Benign RTMID PBXW0040
Copy #
0
med - 2+
med - 1
median
med + 1
med + 2+
Ploidy
2C
2C Rt
>2C
Sector
Ploidy
Multiplier
Error


NYU007 Gleason 7(3+4) 279 Nuclei
Sector
1 Benign LLB PBXW0083
2 GL6(3+3) LLM PBXW0084
3 HGPIN LLA PBXW0085
4 Benign LMB PBXW0086
5 Benign LMM PBXW0087
6 Benign LMA PBXW0088
7 Benign RMB PBXW0089
8 Inflam RMM PBXW0090
9 Benign RMA PBXW0091
10 Benign RLB PBXW0092
11 Benign RLM PBXW0093
12 Benign RLA PBXW0094
13 Benign LPL PBXW0095
Copy #
0
med - 2+
med - 1
median
med + 1
med + 2+
Ploidy
2C
2C Rt
>2C
Clone
Sector
Ploidy
Multiplier
Error


Prostate biopsy washings
NYU007.Gleason 7 (3 + 4)
RIGHT
LEFT
BASE
MID
APEX
10
7
11
8
12
9
4
1
5
2
6
3
13
BASE
MID
APEX
nyu007.3:
Sector 13
MRI-Targeted Biopsy
Left Posterolateral
Mid-Gland
nyu007.1:
Sectors - 2 and 6
nyu007.2:
Sectors - 2 and 3
Lateral
Medial
Medial
Lateral

Genomic/clonal measures complement conventional pathology

Case	Age	Sample[$]	Sectors	Gleason Score Biopsy	Gleason Score Final*	Proportion of sectors with pathology[%]	Proportion of sectors with clonality[@]	Highest Involvement of Cancer[+]	Mean Involvement Of Cancer	Multiple Clones and/or Subclones	Clonal Heterogeneity[&]	Number of Clonal Features[~]	Proportion of Clonal Cells (Clonal/Total)
NYU003.Benign.1	47	PBXW	13	Benign	NA	0/13	0/13	0	0	no	0	0	0/310
NYU002.Pin.1	72	PBXW	13	HGPIN	NA	0/13	0/13	0	0	no	0	1	0/660
COR002.GS6.1	62	TCRP	5	6 (3+3)	6 (3+3)	2/5	2/5			no	1	34	4/451
NYU005.GS6.2	64	PBXW	14	7 (3+4)	6 (3+3)[#]	4/14	1/14	30	5	no	1	0	8/309
NYU001.GS7.1	63	PBXW	14	7 (4+3)	7 (3+4)	8/14	8/14	100	40	yes	2	54	147/712
NYU007.GS7.2	65	PBXW	13	6 (3+3)	7 (3+4)[^]	1/13	4/13	30	2	yes	3	31	42/279
NYU010.GS7.3	79	PBXW	15	7 (3+4)	NA	6/15	2/15	90	11	yes	3	25	20/341
NYU004.GS7.4	75	PBXW	14	8 (4+4)	7 (4+3)[#]	6/14	5/14	100	23	yes	2	41	51/314
NYU011.GS7.5	63	PBXW	10	7 (4+3)	7 (4+3)	5/10	4/10	60	14	yes	2	50	21/221
COR001.GS9.1	77	TCRP	6	9 (5+4)	9 (5+4)	4/6	3/6			no	1	285	85/261
COR003.GS9.2	80	TCRP	5	8 (4+4)	9 (4+5)[^]	3/5	3/5			yes	2	69	117/389
Median Age	65	Total	122	-----	-----	39/122	32/122			-----	-----	-----	495/4247

Genomic/clonal measures complement conventional pathology

Evaluation Criteria	Correlation with the Gleason Score (Diagnostic Biopsy)	Correlation with the Gleason Score (Diagnostic Biopsy) p-value*	Correlation with the Gleason Score (Revised)^	Correlation with the Gleason Score (Revised) p-value*
Clonal Heterogeneity	0.36	0.26	0.86	0.01
Proportion of clonal cells	0.46	0.14	0.79	0.01
Number of clonal features	0.55	0.08	0.79	0.01
Proportion of sectors with clonality	0.55	0.08	0.79	0.01
Proportion of sectors with patholgy	0.71	0.02	0.7	0.03
Highest Involvement of Cancer	0.83	0.01	0.78	0.02
Mean Involvement Of Cancer	0.80	0.01	0.70	0.03
Gleason Score Biopsy	1.00	0.002	0.64	0.06

Status: bench

- CN alterations common across all grades
- Clones rare below Gleason=6
- Massive clones in all Gleason≥7 cases
- Clones in 2 out of 3 Gleason 6 cases
- Clones are predominantly located in high-Gleason areas
- However, there are exceptions. Evidence for migration?
- Potential to meaningfully supplement conventional pathology
- Near future: pooling DNA from clonal cells for deeper analysis

Status: computing

- SCGV released
- Upstream pipeline: Docker image coming soon
- Near future: modules to handle a variety of DNA prep protocols
- Farther down the road: can pathology image analysis reveal more when supplemented by genomics? Is machine learning from images possible, with genomics as ground truth?

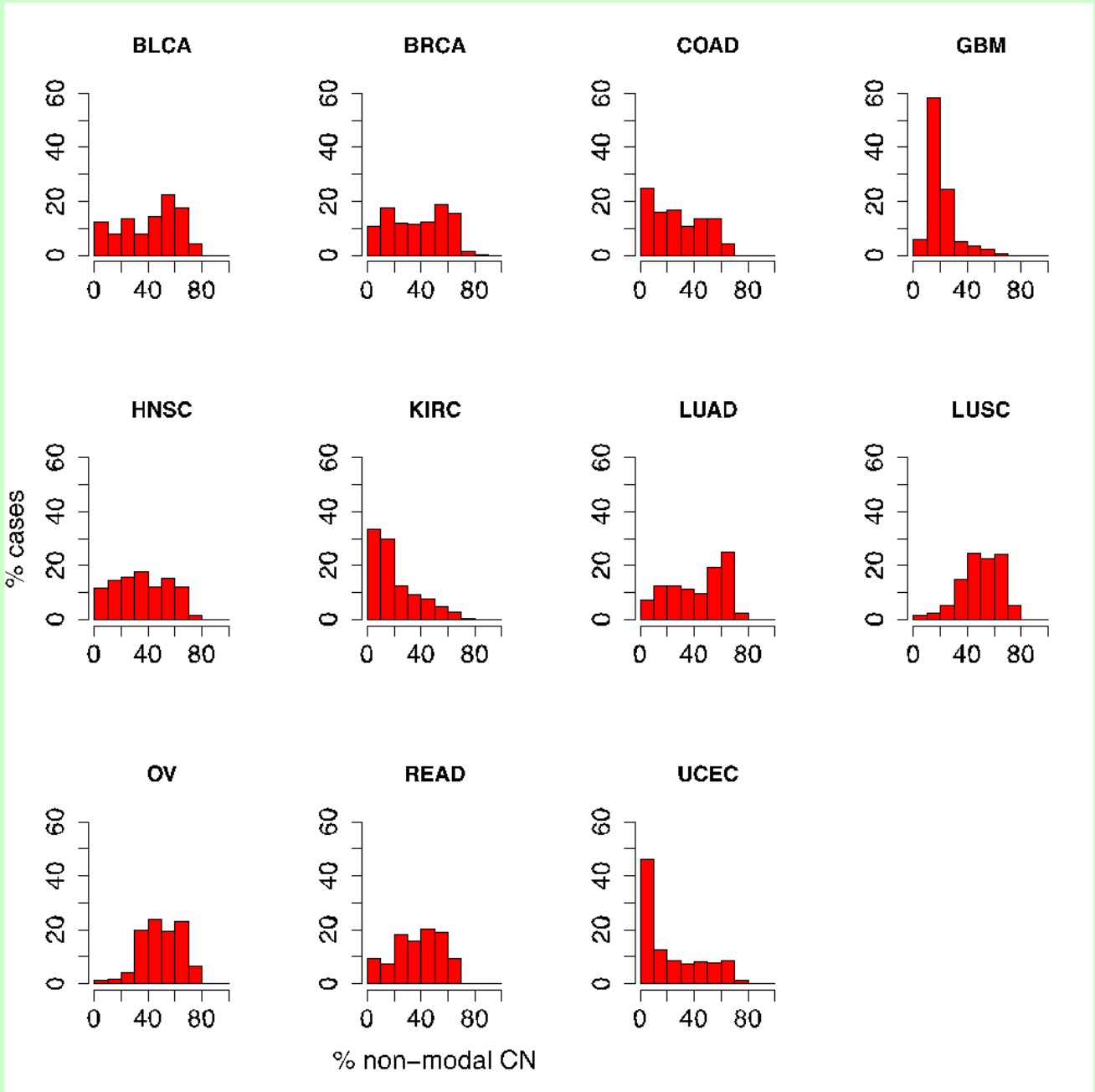
Early detection of cancer

Setting

- Existing blood-based molecular screening methods (e.g., PSA) lack sensitivity, specificity and universality
- Genome-wide DNA copy number (CN) variation is ubiquitous in multiple tumor types
- Tumor cells bear recurrent, clone-wide CN signature
- Use single-cell computational pipeline to peek into the future

A. Krasnitz et al, Trends in Molecular Medicine 23 (2017) 4

Genome involvement in CN variation by cancer type



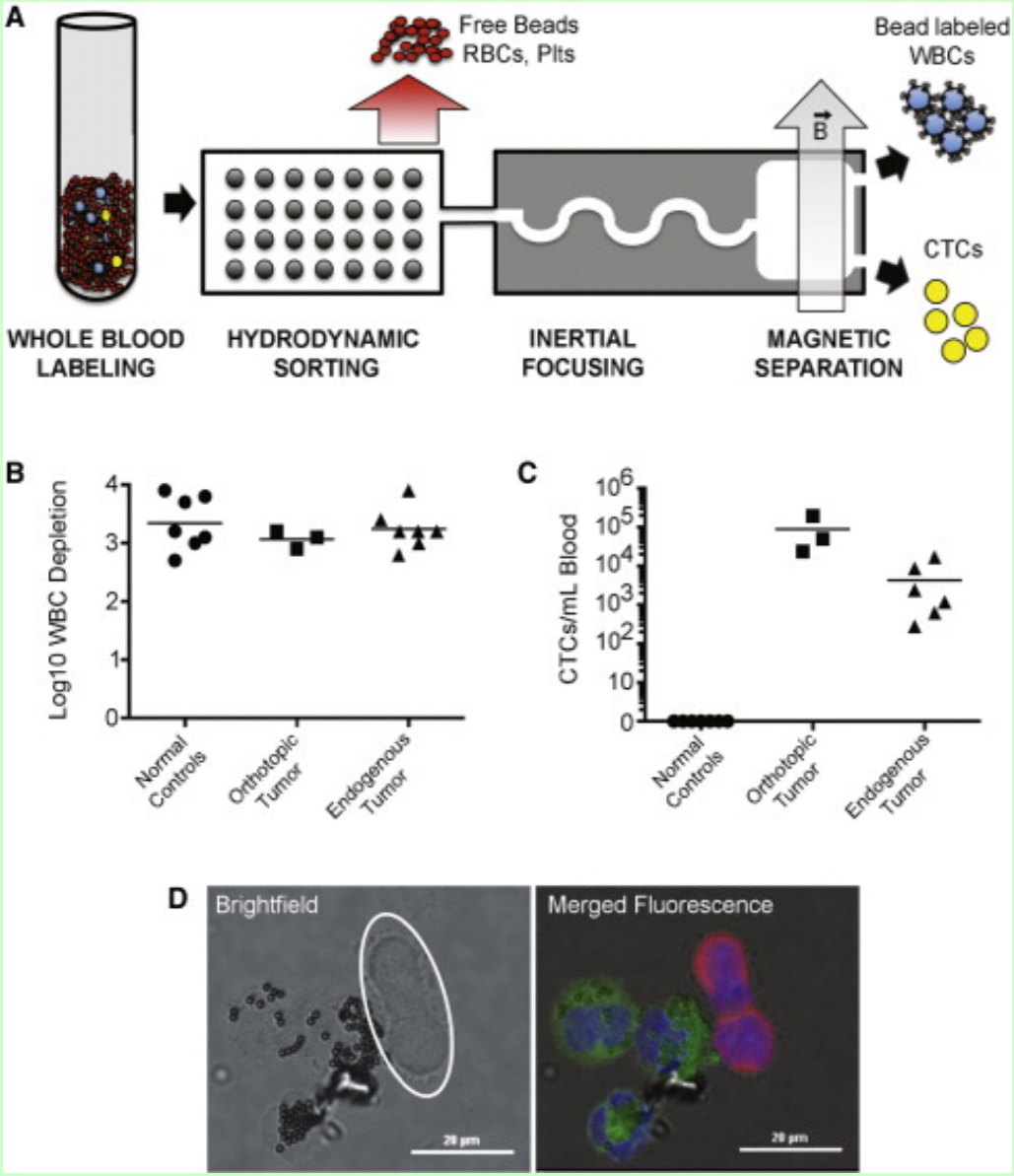
Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable
- Highly cancer-specific
- Highly sensitive
- Affordable ($1K per test)

Scenario: draw 10mL of blood ~ 1B nucleated cells.

10 of these are circulating tumor cells.

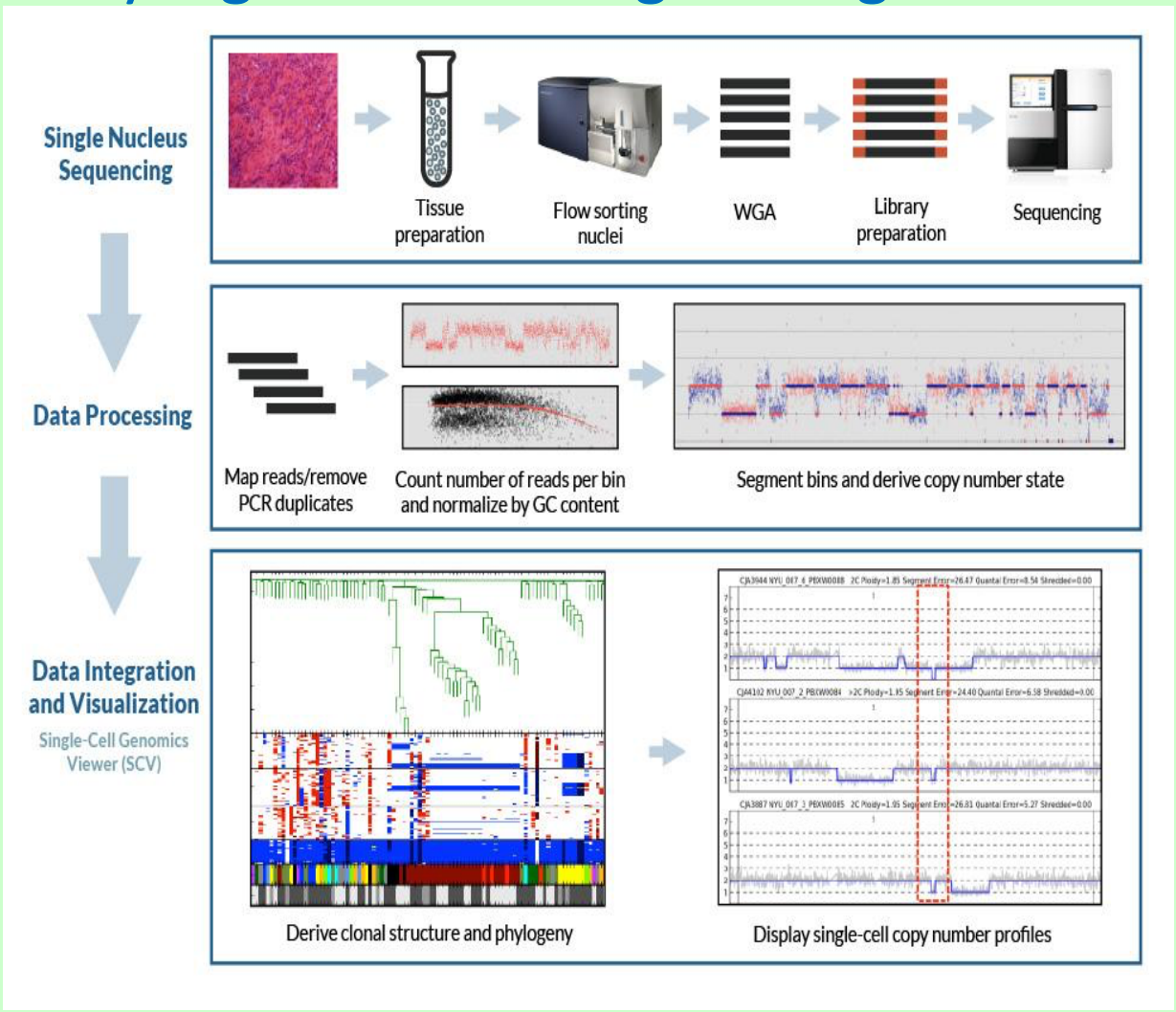
Key ingredient 1: deplete leukocytes



1K residual cells
per 10 mL of blood

(D. Ting et al,
Cell Rep. 2014)

Key ingredient 2: single-cell genomics



Very sparse (0.003X) sequencing of individual cells ($1/cell)
Assuming 10 cells from a cancer clone,
Knowing what we know about cancer types (TCGA),
How successful would we be in detecting them?

Assessment of Feasibility

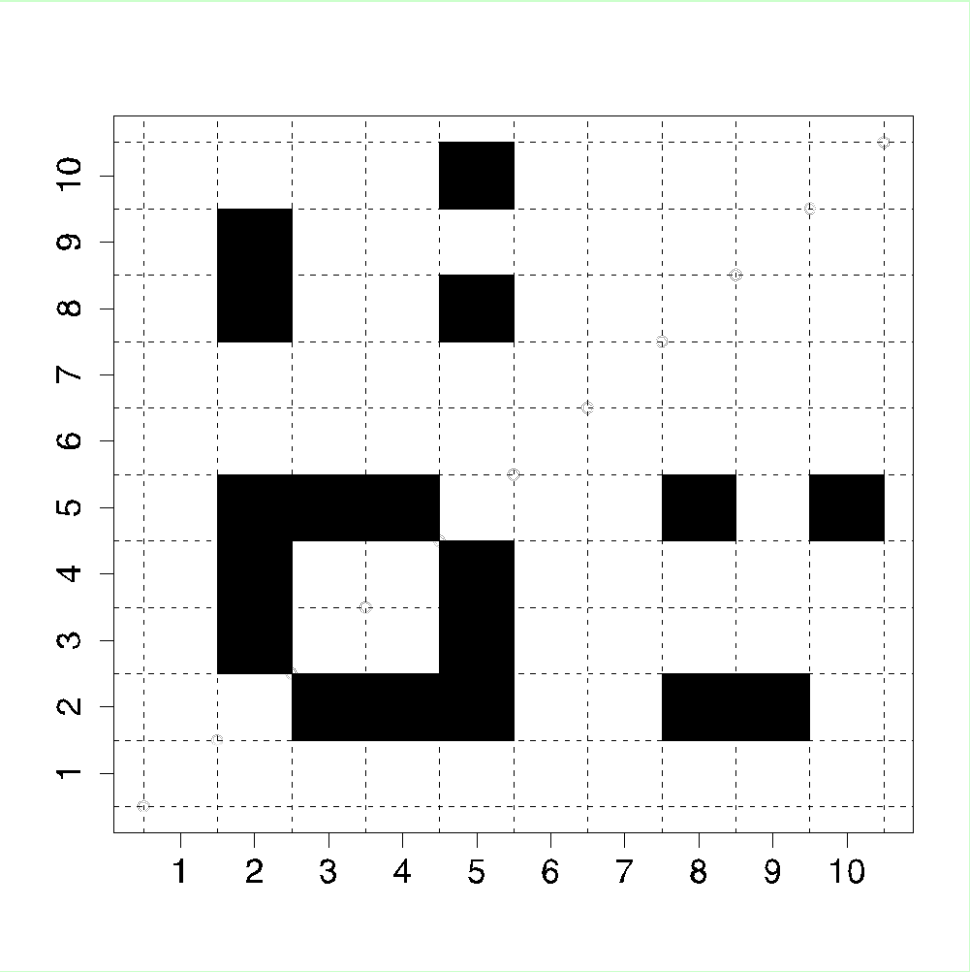
Input

- 1306 sequencing read sets from diploid cells
- 3852 published integer-valued copy number profiles of cancer genomes (TCGA)
- 11 tumor types represented

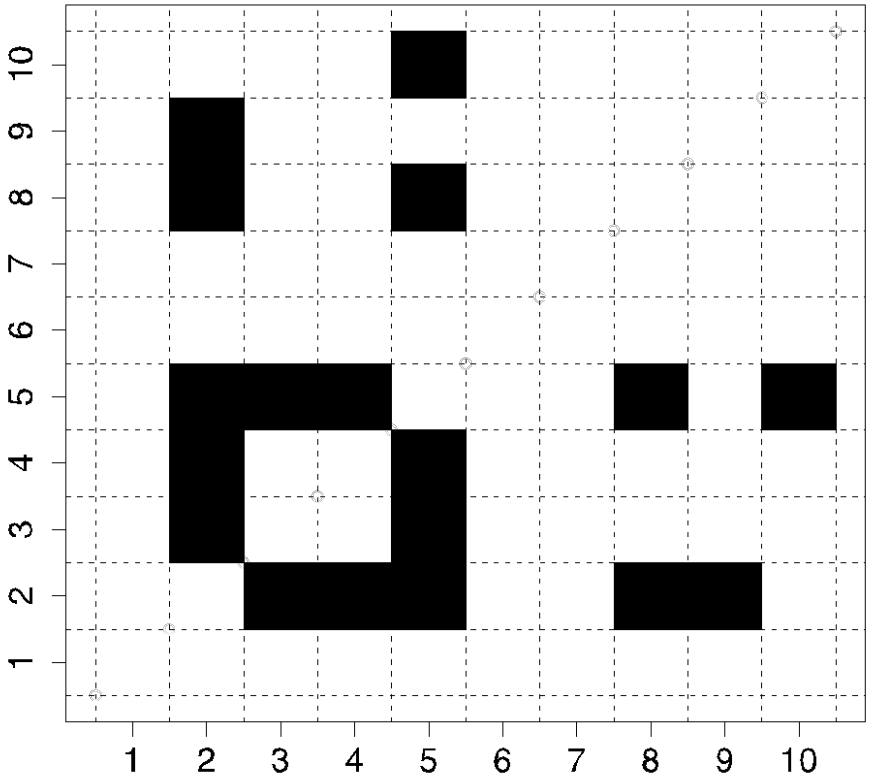
Simulation

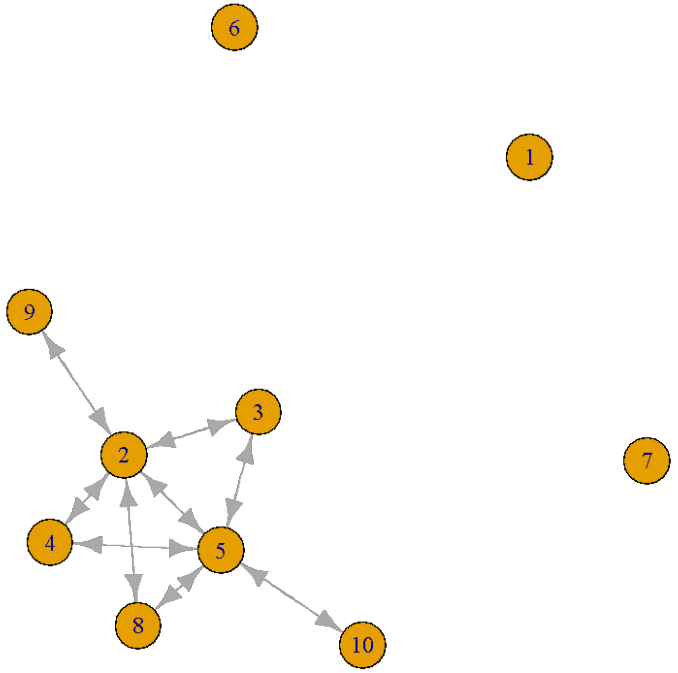
- Resample read sets from diploid cells to reflect the desired copy-number profile and coverage
- 3 kinds of cell populations
  - Clonal cells with CN profiles of TCGA tumors
  - Diploid cells
  - Cells with unstable genomes: a mixture of chromosomes from all TCGA tumors
- Compute pairwise correlations among CN profiles
- Assume 10 clonal cells per blood sample

From correlations to connectivities

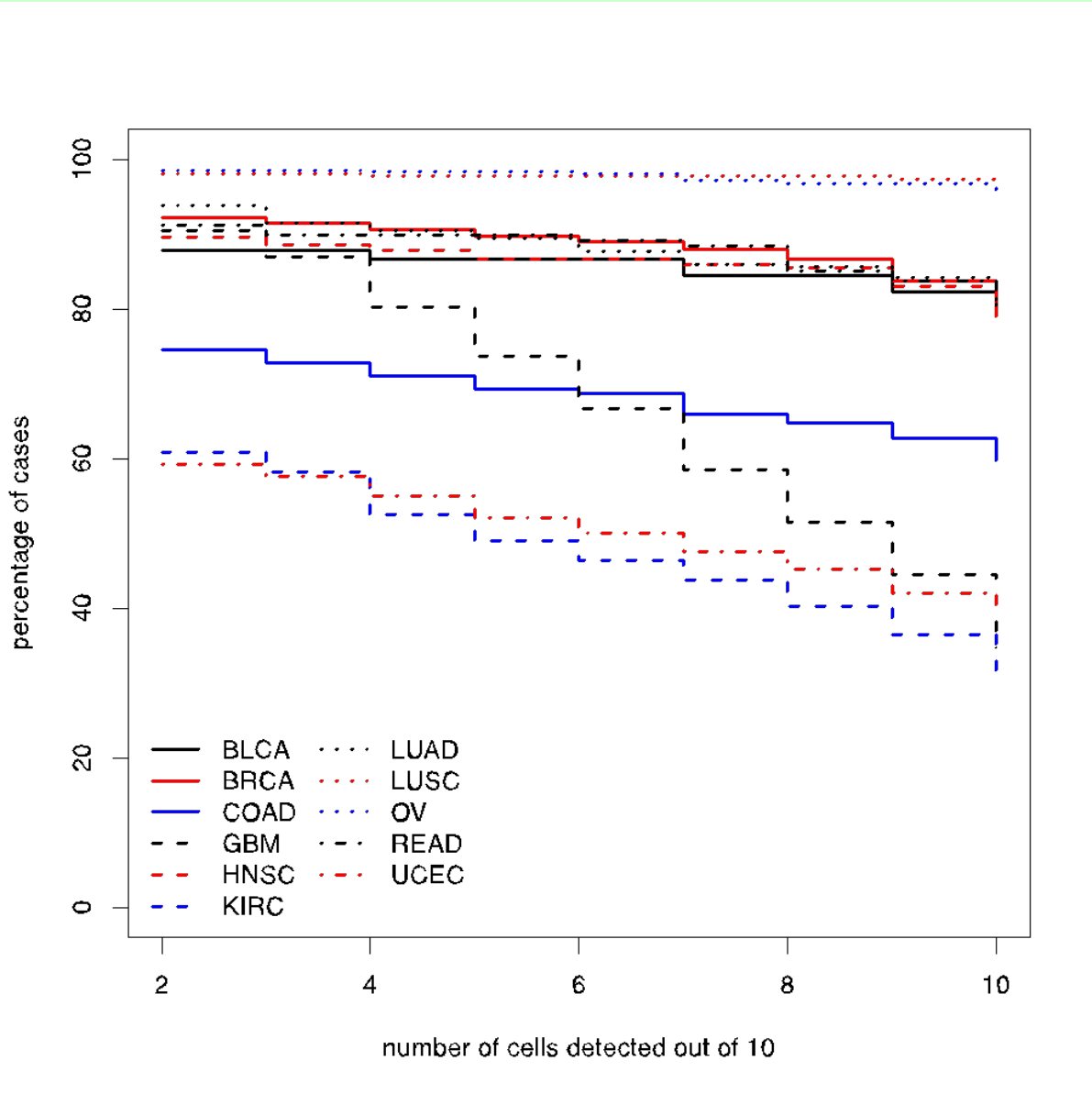


From correlations to connectivities

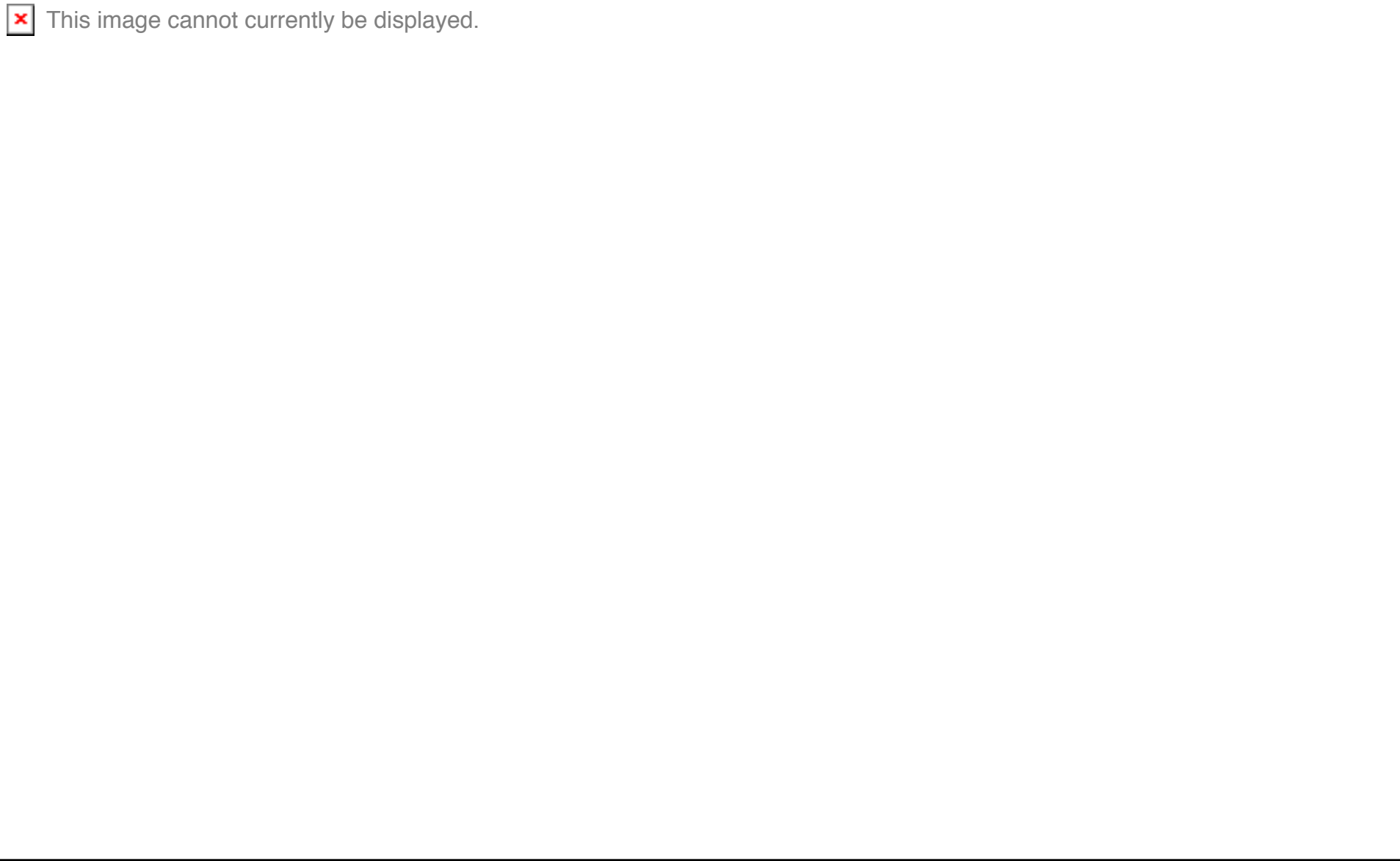




High sensitivity to major tumor types

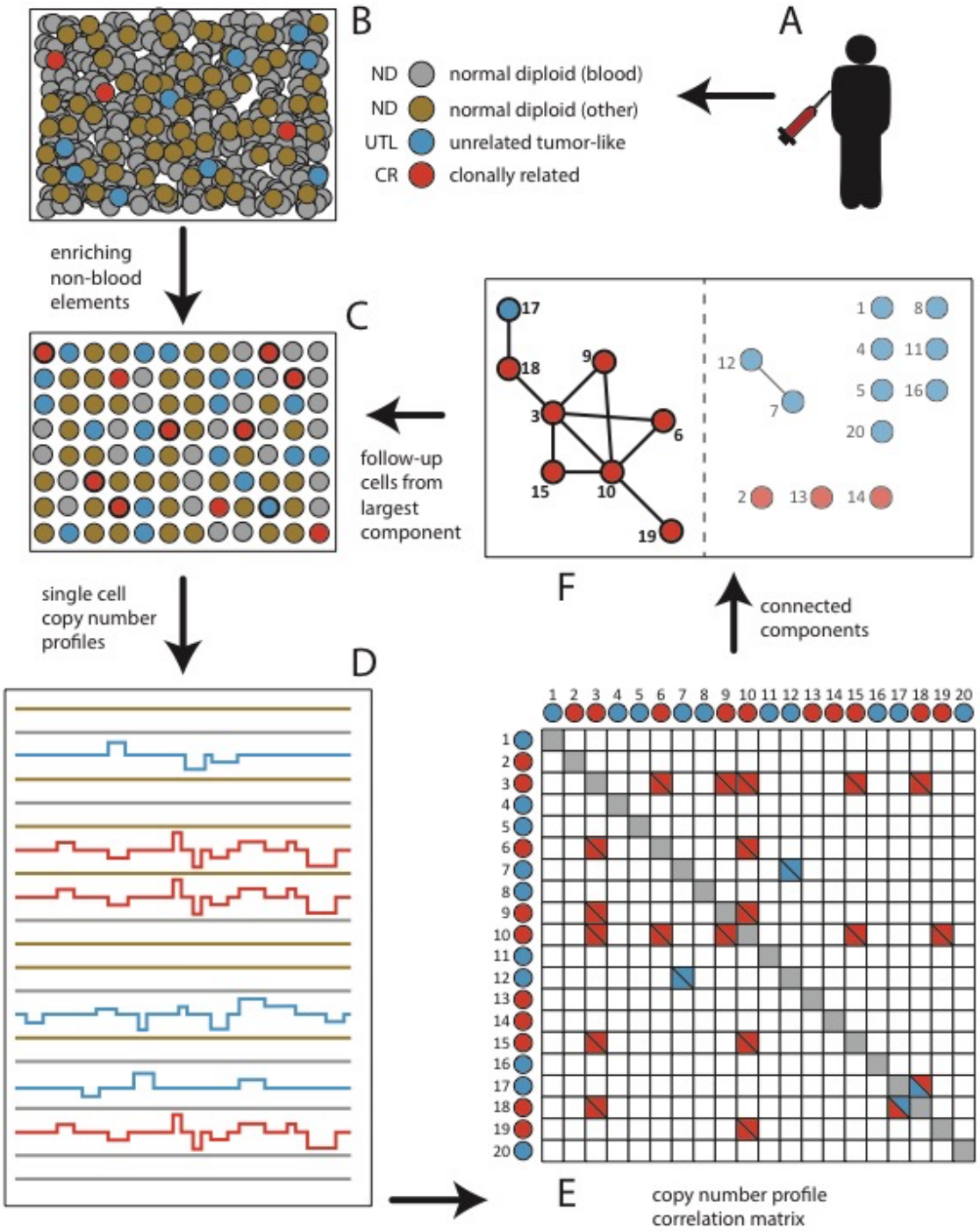


Specificity: non-clonal tumor-like cells may occasionally correlate…



…but they are unlikely to form large clusters

Cluster size	10 cells	20 cells	50 cells	100 cells	200 cells
2	0.0086	0.028	0.15	0.43	0.84
3	3.00×10^{-4}	0.0018	0.028	0.13	0.51
4	1.00×10^{-4}	1.00×10^{-4}	0.0036	0.044	0.3
5	0	1.00×10^{-4}	4.00×10^{-4}	0.012	0.16
6	0	0	2.00×10^{-4}	0.0031	0.08
7	0	0	0	6.00×10^{-4}	0.034
8	0	0	0	3.00×10^{-4}	0.013
9	0	0	0	0	0.0049
10	0	0	0	0	0.0012
12	0	0	0	0	1.00×10^{-4}


A
B
ND normal diploid (blood)
ND normal diploid (other)
UTL unrelated tumor-like
CR clonally related
enriching non-blood elements
C
follow-up cells from largest component
F
connected components
single cell copy number profiles
D
E
copy number profile correlation matrix

Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable
- Highly cancer-specific
- Highly sensitive
- Affordable ($1K per test)

Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable ✔
- Highly cancer-specific
- Highly sensitive
- Affordable ($1K per test)

Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable ✔
- Highly cancer-specific ✔
- Highly sensitive
- Affordable ($1K per test)

Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable ✔
- Highly cancer-specific ✔
- Highly sensitive ✔
- Affordable ($1K per test)

Goal: design a molecular blood-based early-detection method for solid tumors that is

- Widely applicable ✔
- Highly cancer-specific ✔
- Highly sensitive ✔
- Affordable ($1K per test) ✔ (assuming 1K cells @$1/cell)

Further steps

- Test feasibility in newly diagnosed patients: sensitive detection of patient-specific genomic tumor signature in blood
- Other analytes: urine; bronchoalveolar lavage
- Retain and pool libraries from clonal populations for deeper analysis / determination of origin (surface markers, methylation profiling)

Research Team

CSHL

Mike Wigler
Joan Alexander
Jude Kendall
Alex Krasnitz
Viacheslav Zhygulin
Guoli Sun
Junyan Song

SCGV Developer

Lubomir Chobardjiev

Clinical Collaborators (GU Oncology/Pathology)

Mt. Sinai

Dr. Ashutosh Tewari

NYU

Dr. Herbert Lepor

Adopters @

MSKCC

Breast tumor biopsies

CSHL

Prostate cancer (mouse models)
PDAC organoids

Supported by

NCI-ITCR (U01CA188590)

Simons Foundation