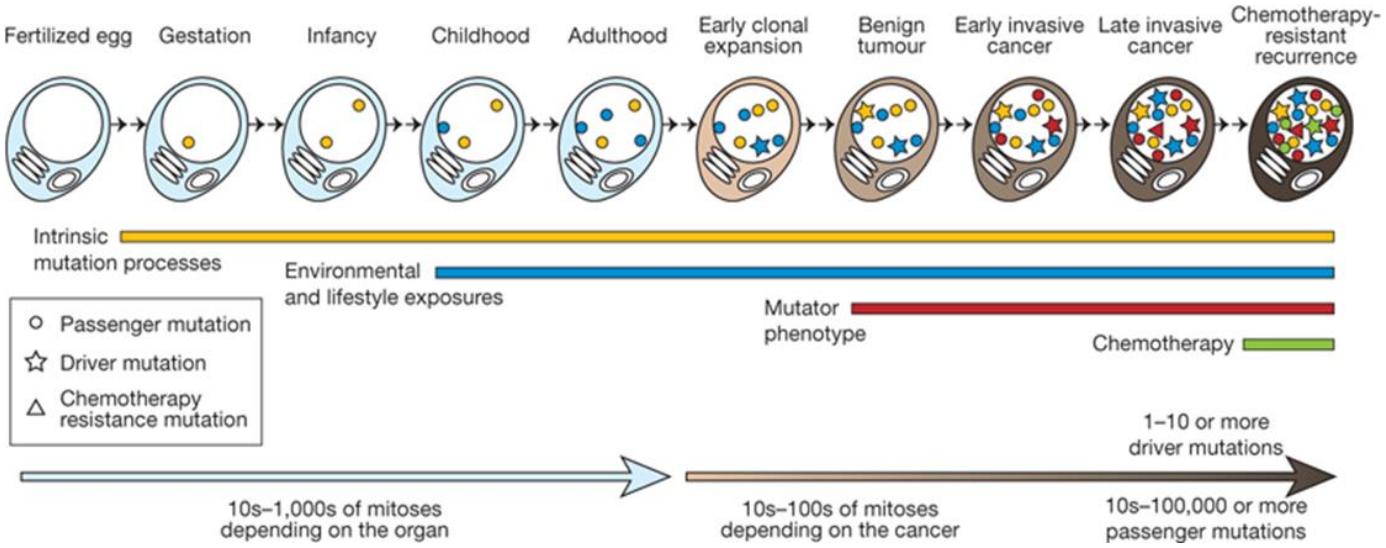
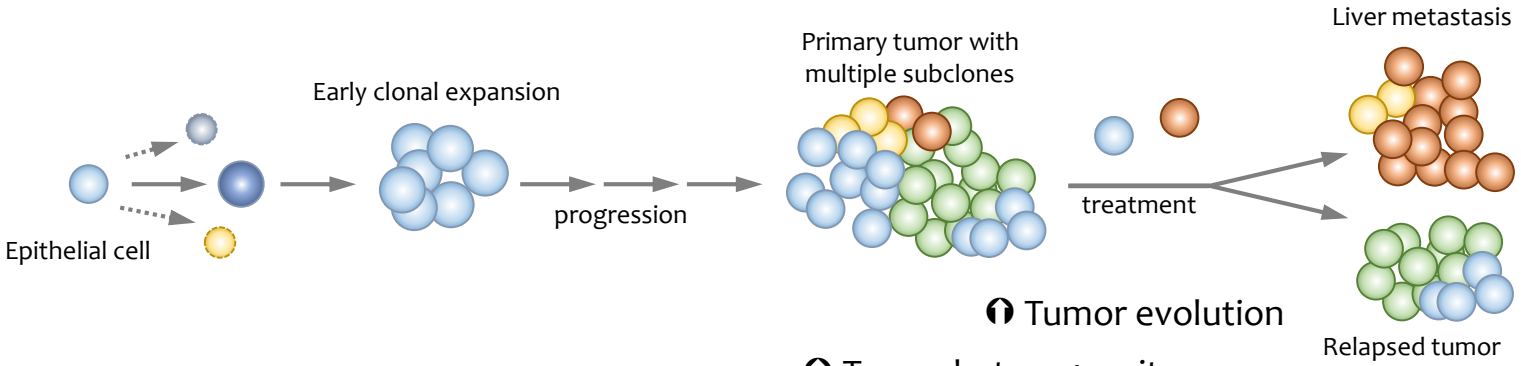
서울대학교 대학원 유전공학특론
2018.5.23 (Wed)

단일세포 유전체분석 방법

박 웅 양
삼성서울병원 유전체연구소
성균관대학교 의과대학 분자세포생물학교실

Tumor evolution and heterogeneity

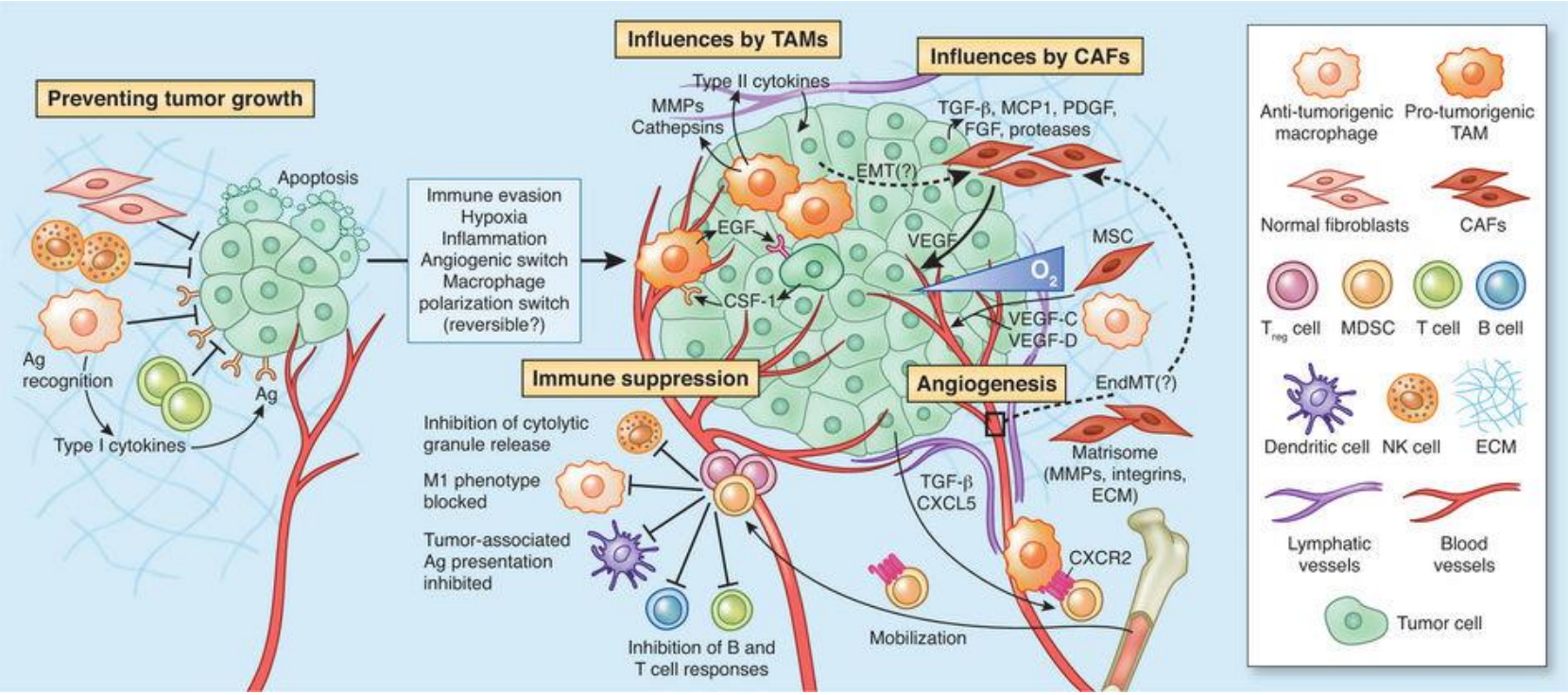




- Tumor evolution
- Tumor heterogeneity
- Circulating tumor cells
- Cancer stem cells (rare cell population)
- Subclonal differentiation and drug response

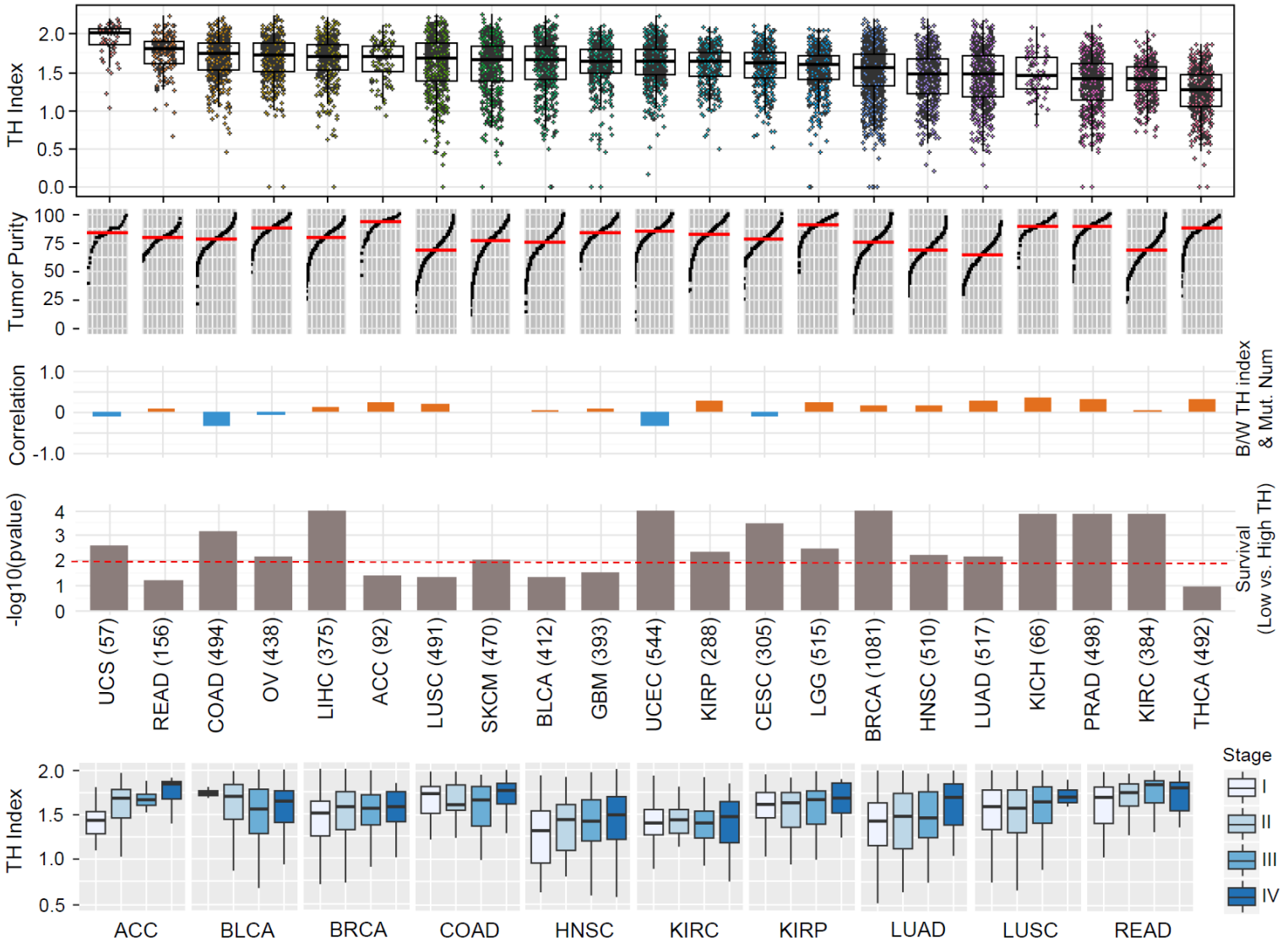
Tumor microenvironment as a therapeutic target

- Cancers should survive within a complex tissue and depends on sustained growth, invasion and metastasis.
- Stromal cells in the tumor microenvironment can be an attractive therapeutic target with reduced risk of resistance and tumor recurrence.



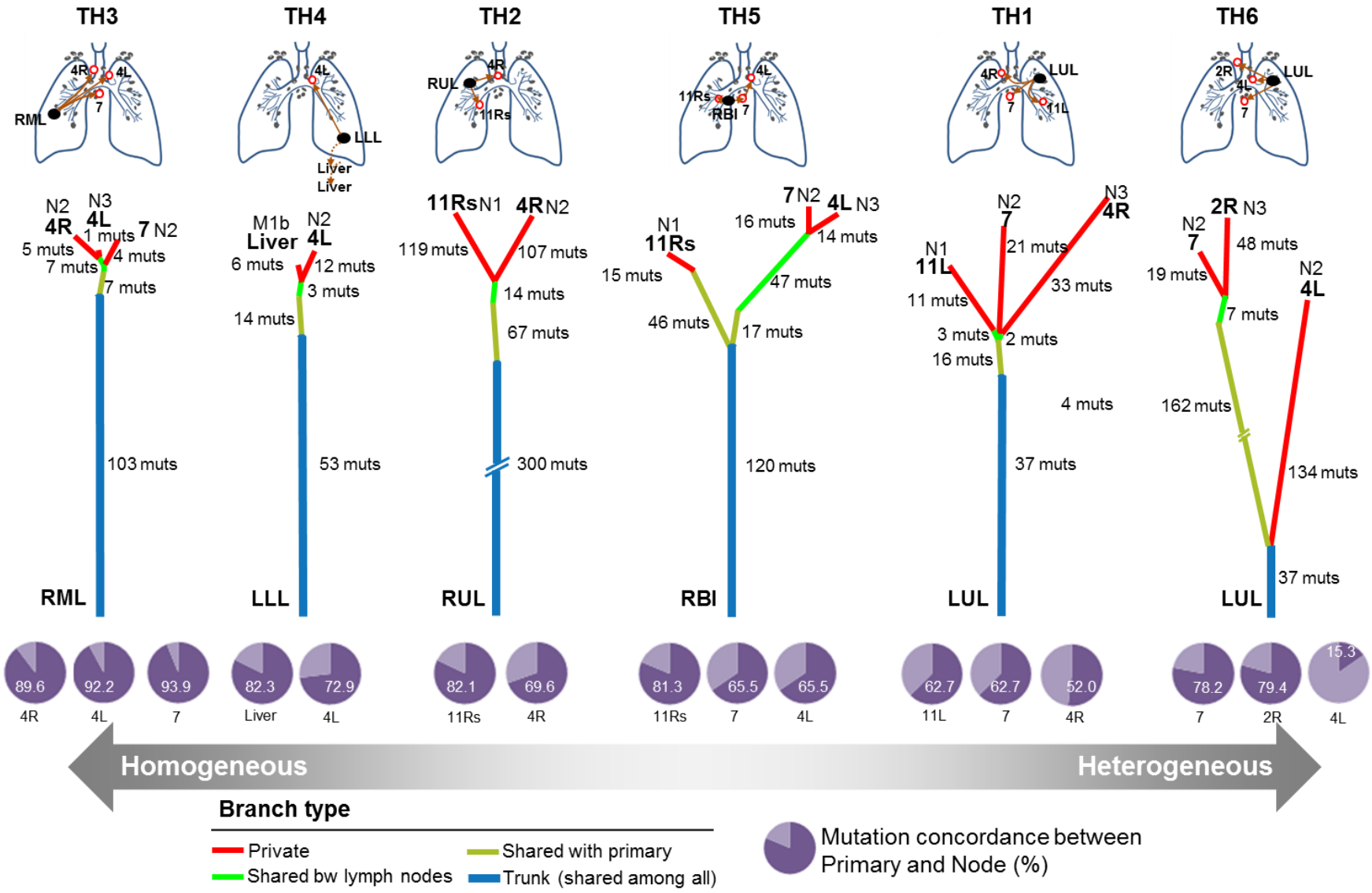
Quail DF and Joyce JA, Nature Medicine, 2013

Tumor heterogeneity analysis using whole exome sequencing data



Joung JG et al. Clinical Cancer Research, 2017

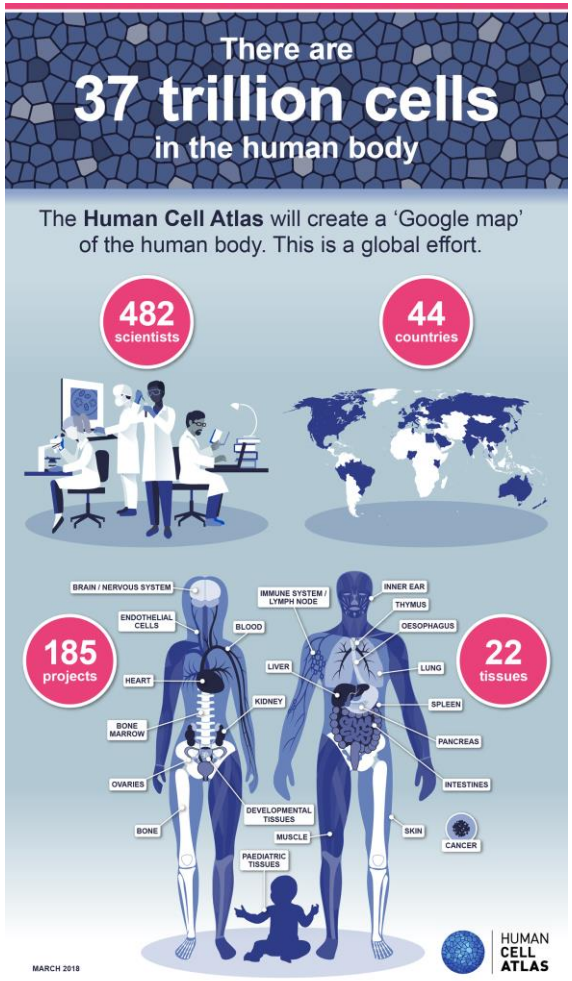
Tumor evolution analysis using longitudinal or multi-regional samples

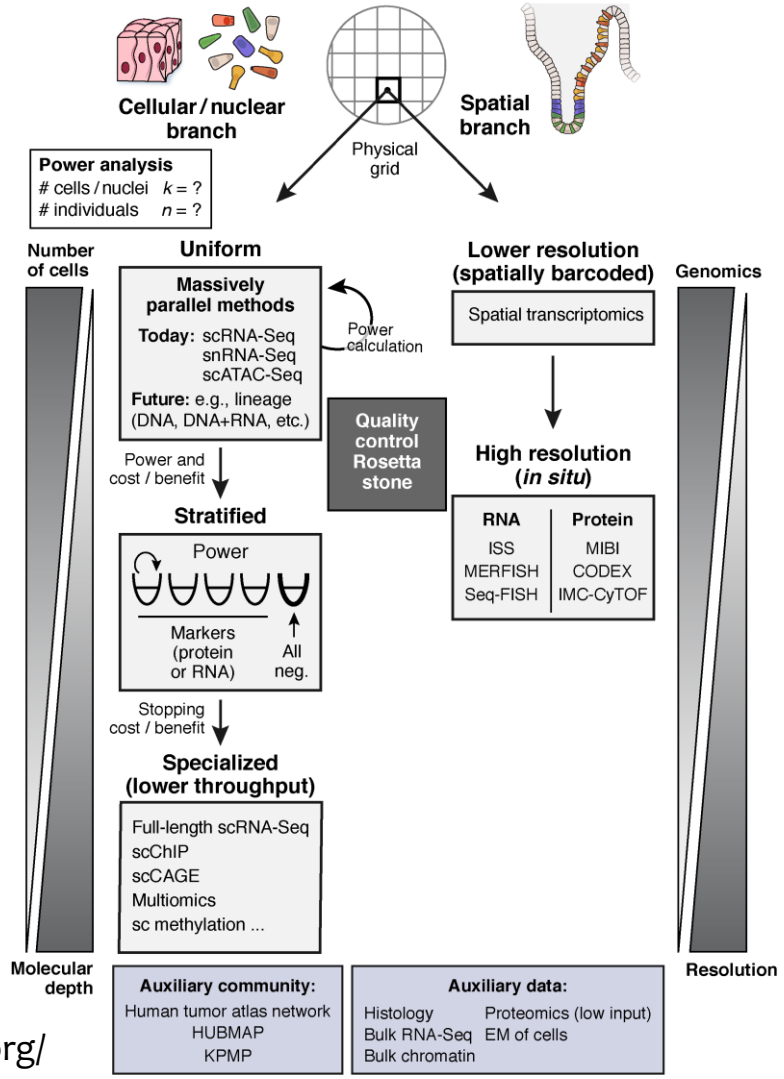


Um SW, Joung JG et al. Cancer Research, 2016

How many cell types in our body?

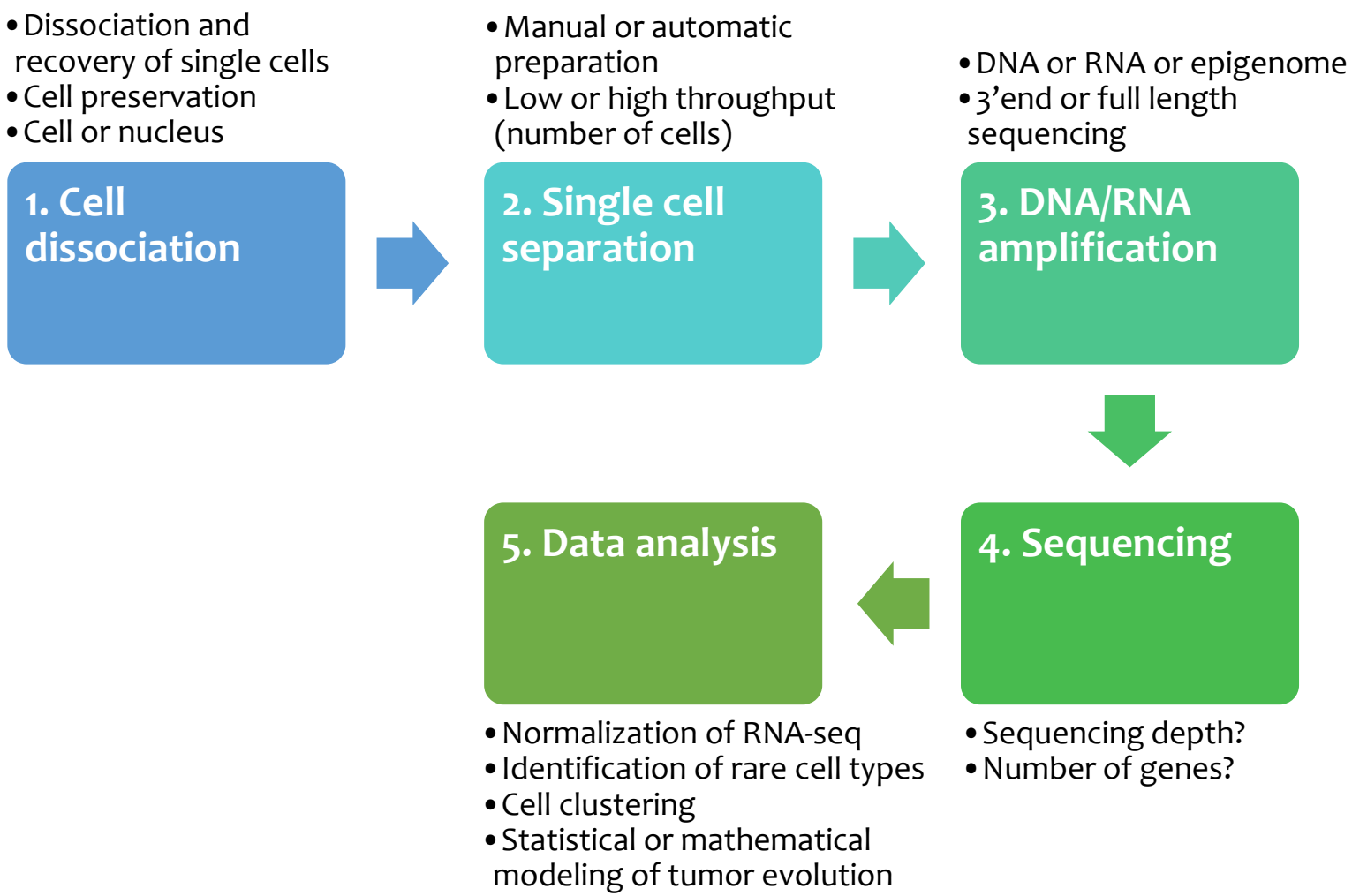
- The Human Cell Atlas (HCA) will be made up of comprehensive reference maps of all human cells — the fundamental units of life — as a basis for understanding fundamental human biological processes and diagnosing, monitoring, and treating disease.





https://www.humancellatlas.org/

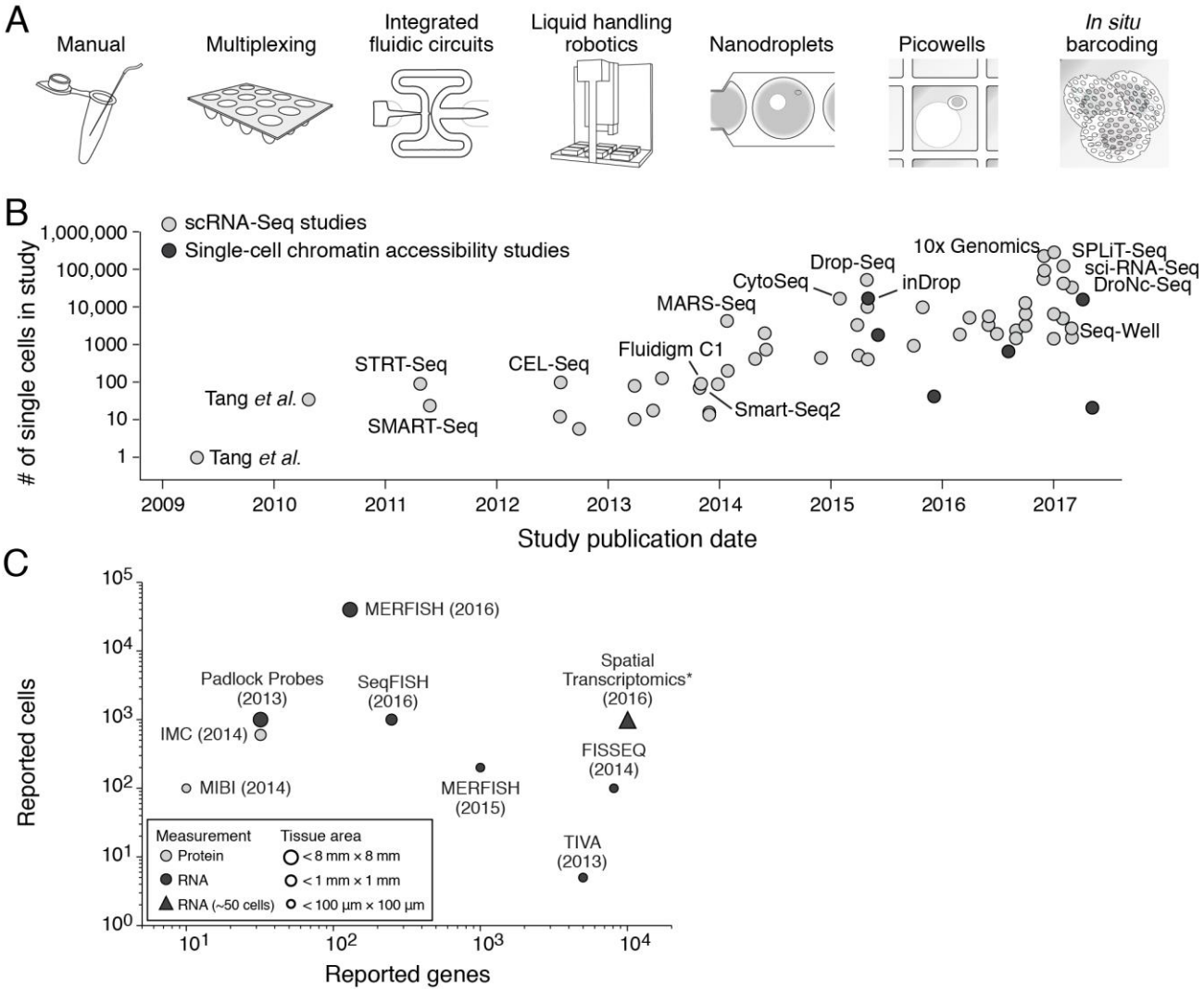
Workflow of single cell genome analysis



Cell dissociation

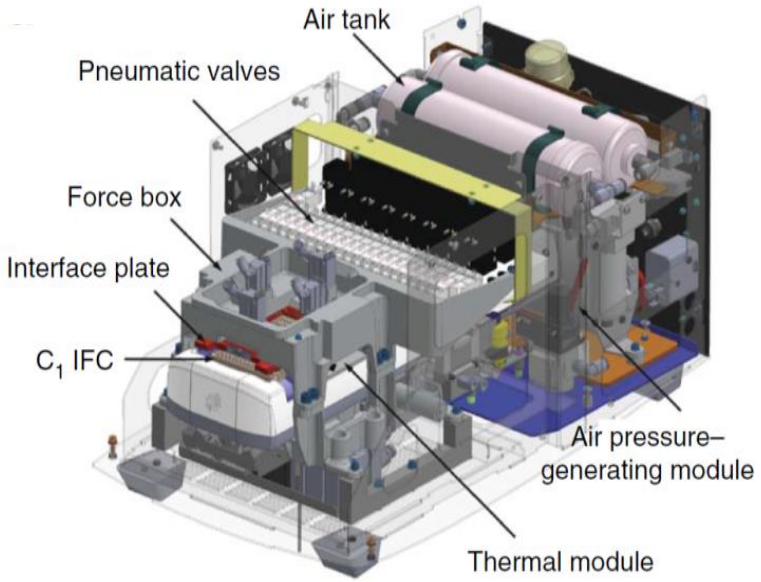
- In vitro cultured cells
- Dissociated single cells from tissue
  - Freshly dissociated cells
  - Cell freezing
- Nuclear seq

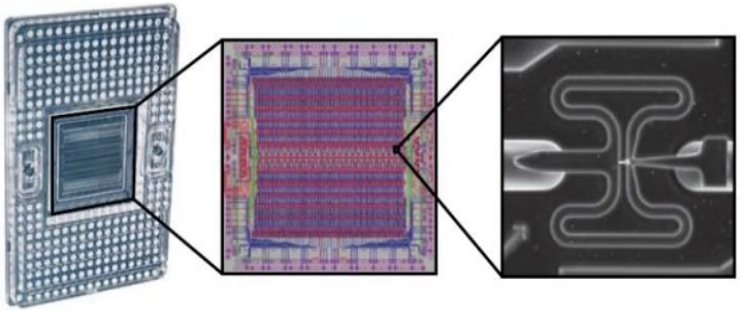
Single cell separation



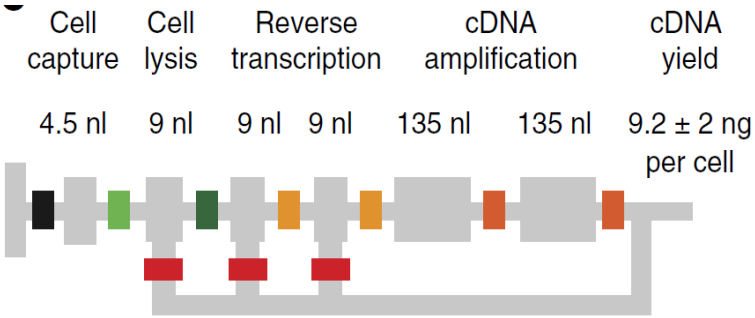
The Human Cell Atlas White Paper, 2017

C1 for single cell isolation and RNA amplification





Average capture: 72 ± 5 single cells per chip



Nature Biotech, 2014

How many single cells needed?

Box 1: How many single cells to sequence?

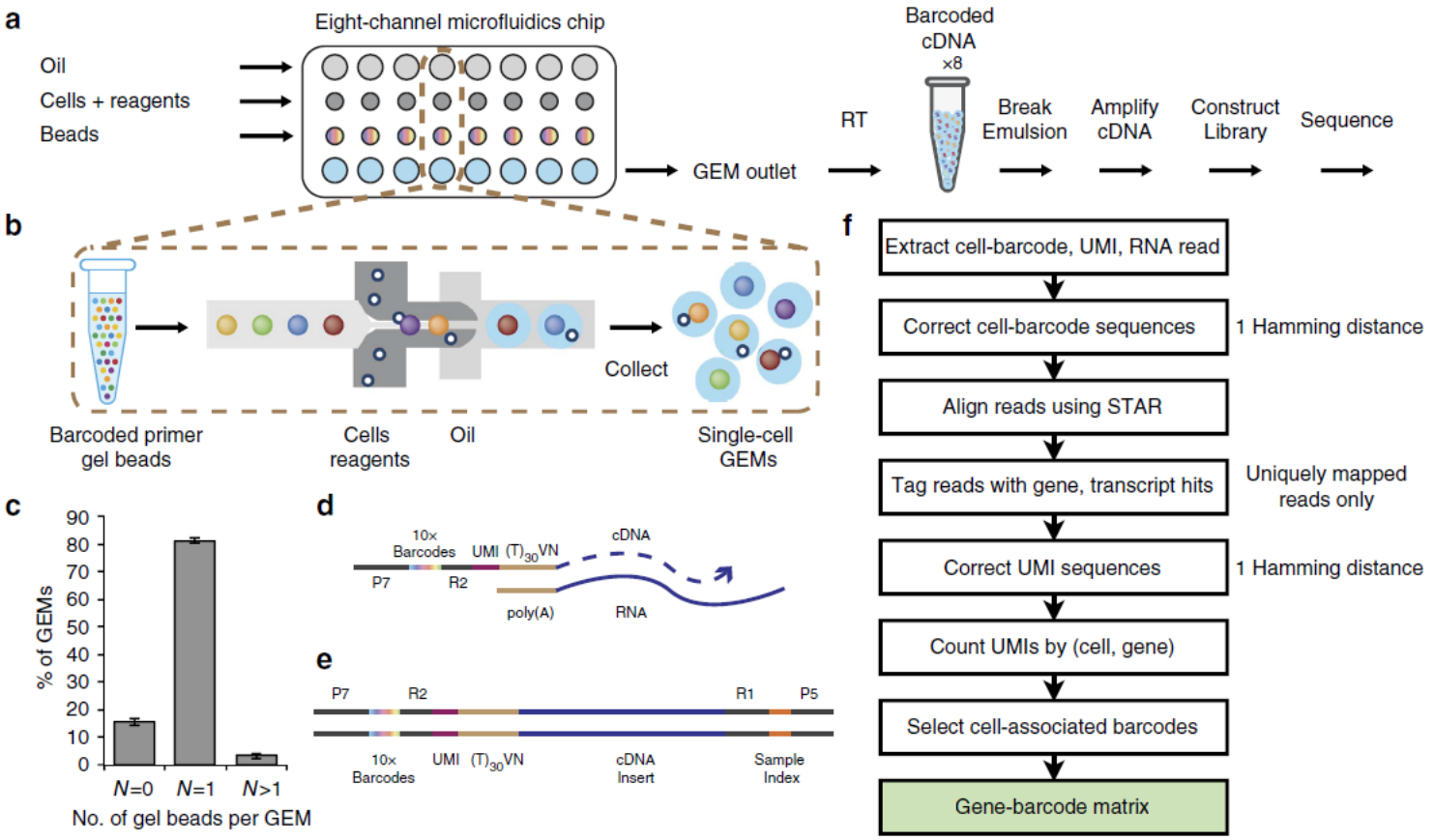
Most SCS studies to date have analyzed an arbitrary number of cells, ranging from about 50–100 cells from each patient, a number that is mainly dictated by cost. The difficulty in estimating sample sizes (N) for SCS experiments is that these calculations rely on several important factors that are often difficult to measure, including (1) the diversity of the underlying population; (2) the sensitivity required for detecting rare clones (e.g., 1%); and (3) the technical variability of the SCS method. The statistical methods for estimating sample numbers from SCS experiments can be broadly classified into two categories: (1) retrospective estimations; and (2) prospective estimations. Retrospective methods involve back-calculations, in which an arbitrary number of single cells are sequenced (DNA or RNA) in a series of tumors, and a cumulative number of new mutations or transcripts are plotted as a function of the cumulative number of single cells, resulting in a saturation curve. This calculation is analogous to rarefaction calculations in ecology, which are derived from species capture–recapture experiments. In these plots, the cumulative number of species versus the cumulative number of samples are plotted in random order to determine the point at which the number of species detected reaches a plateau in the curve (Gotelli and Colwell 2001). Statistical methods for prospective calculations are more challenging to estimate; however, they do not require empirical data sets. A simple power calculation can be formalized as

$$P(d) = 1 - (1 - s)^n,$$

where s is equivalent to the subclonal frequency, and n is equal to the number of single cells that need to be sequenced. Based on this equation, the detection of a 10% subclone would require sequencing at least $n = 25$ single cells to achieve a 0.9 detection power, whereas a 1.0 detection power would require $n = 50$ single cells. Similarly, $n = 250$ single cells are require to detect a 1% subclone with 0.9 detection power, and $n = 500$ single cells are required for 1.0 power. However, this calculation does not incorporate the number of replicate cells needed to mitigate false positive (FP) and false negative (FN) error rates and eliminate random errors, which may range from $r = 2$–5 replicate cells depending on the SCS technology ($N * r$). More sophisticated sample number estimations can be calculated using probabilistic methods from population genetics that were originally developed for discovering rare disease variants based on expected minor allele frequencies (MAFs) in the human population (Wendl and Wilson 2009). In summary, a number of statistical methods from ecology and population genetics are useful for estimating the sample sizes required for SCS experiments and should be considered before starting any single-cell sequencing study.

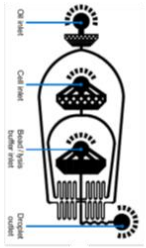
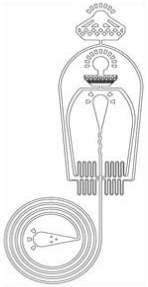
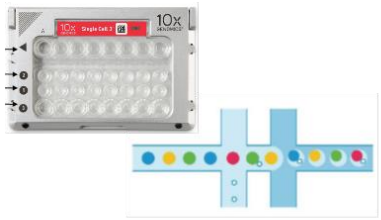
Navin NE, Genome Reserch, 2015

3'RNA profiling of thousands of single cell



Zheng G et al Nature Communications, 2017

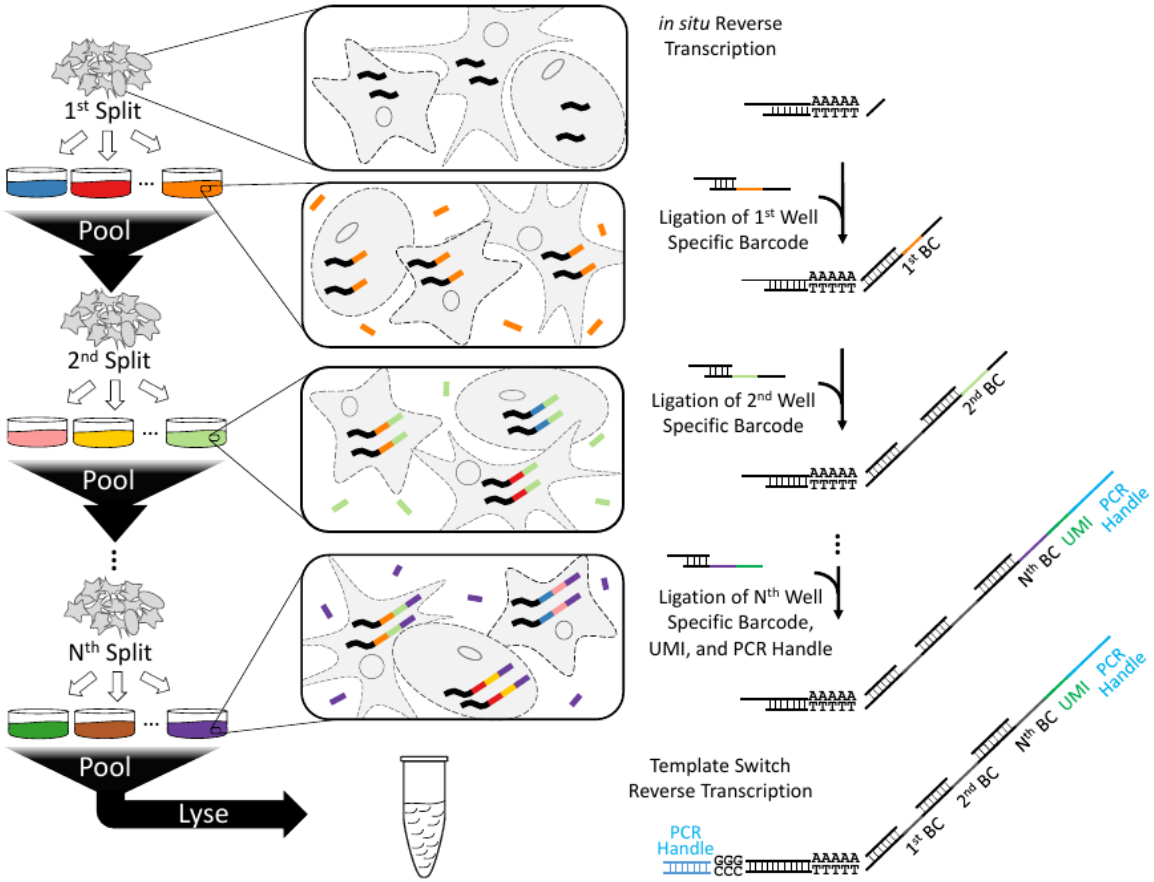
Drop-seq single cell analysis

		Harvard McCarroll group	Ordered encapsulation	10x Chromium™ single cell 3'
				
Input	Cell	12.5/uL, 50/uL (in PBS)	100~1000/uL (in PBS)	42~831/uL (in buffer+master mix)
	Index	PMMA bead: 120/uL	PMMA bead: ~1500/uL	Gel bead : 15000/uL (load 40uL)
Cell yield		1-5% (estimated)	~16%	~50%
Throughput		<10^4/h (@100cell/uL)	~10^5 /h (@1000cell/uL)	6000 cells per lane in 1run (15min) Max 8 lane
Doublet	Cell	0.36% (@12.5/uL) 11.3% (@100/uL)	1.2% (@500/uL) 2.5% (@1000/uL)	~1.1% @ 2600cell load ~5.3% @12800cell load
	Bead	5% (@120/uL)	0.9% (@1500/uL)	

Moon HS and Je K et al Lab Chip, 2018

SPLiT-Seq

- Split-pool ligation-based transcriptome sequencing (SPLiT-seq), a single-cell RNA-seq (scRNA-seq) method that labels the cellular origin of RNA through combinatorial barcoding.



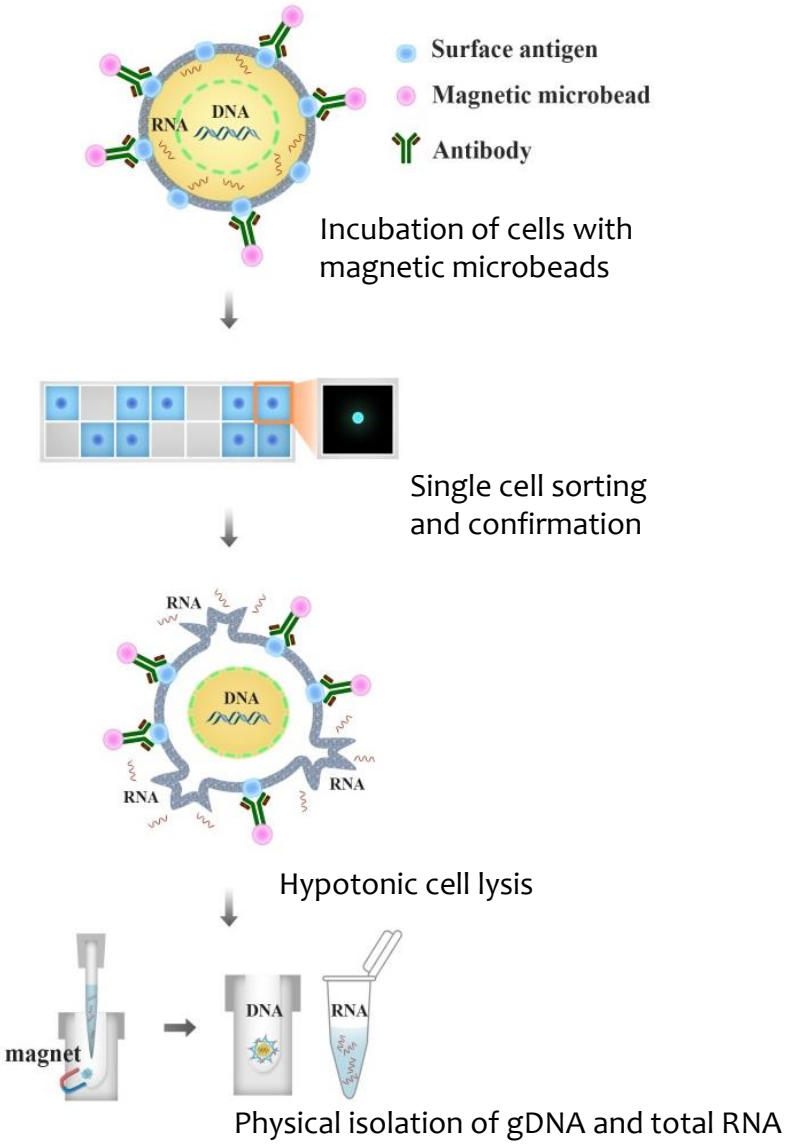
Rosenberg AB et al. bioRxiv, 2017, Rosenberg AB et al. Science, 2018

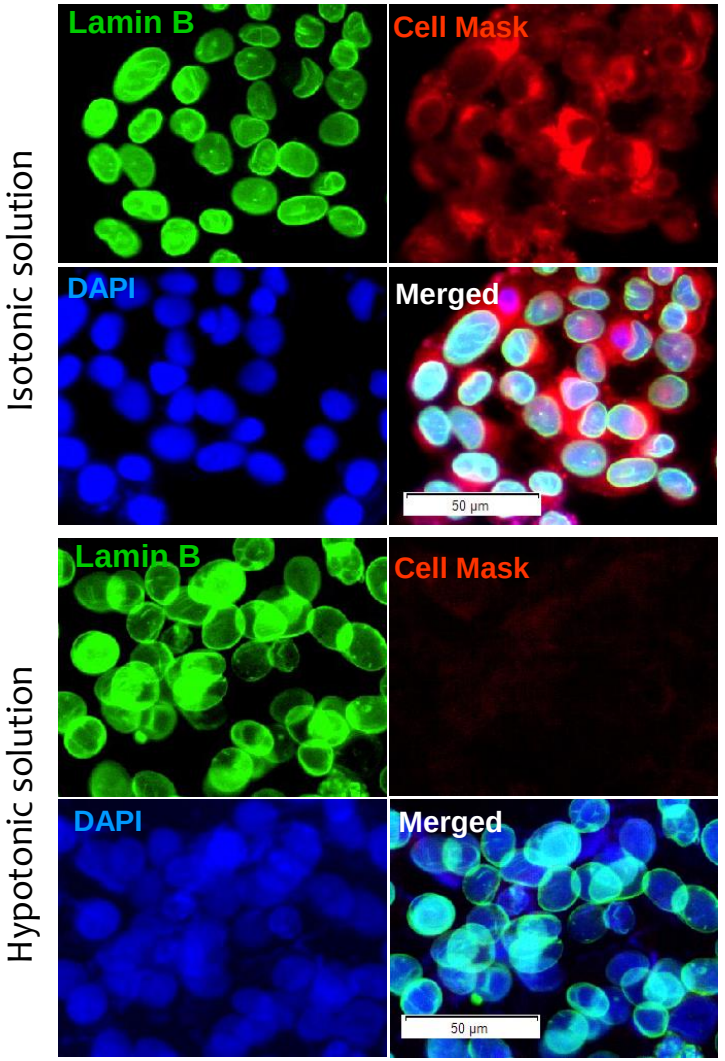
scRNA-seq methods

Protocol example	*C1 (SMARTer)*	*Smart-seq2*	*MATQ-seq*	*MARS-seq*	*CEL-seq*	*Drop-seq*	*InDrop*	*Chromium*	*SEQ-well*	*SPLIT-seq*
Transcript data	Full length	Full length	Full length	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (number of cells)	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^5
Typical read depth (per cell)	10^6	10^6	10^6	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4
Reaction volume	Nanoliter	Microliter	Microliter	Microliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Microliter
Reference	[63]	[57]	[39]	[10]	[64]	[45]	[46]	[47]	[101]	[38]

Genome Medicine, 2017

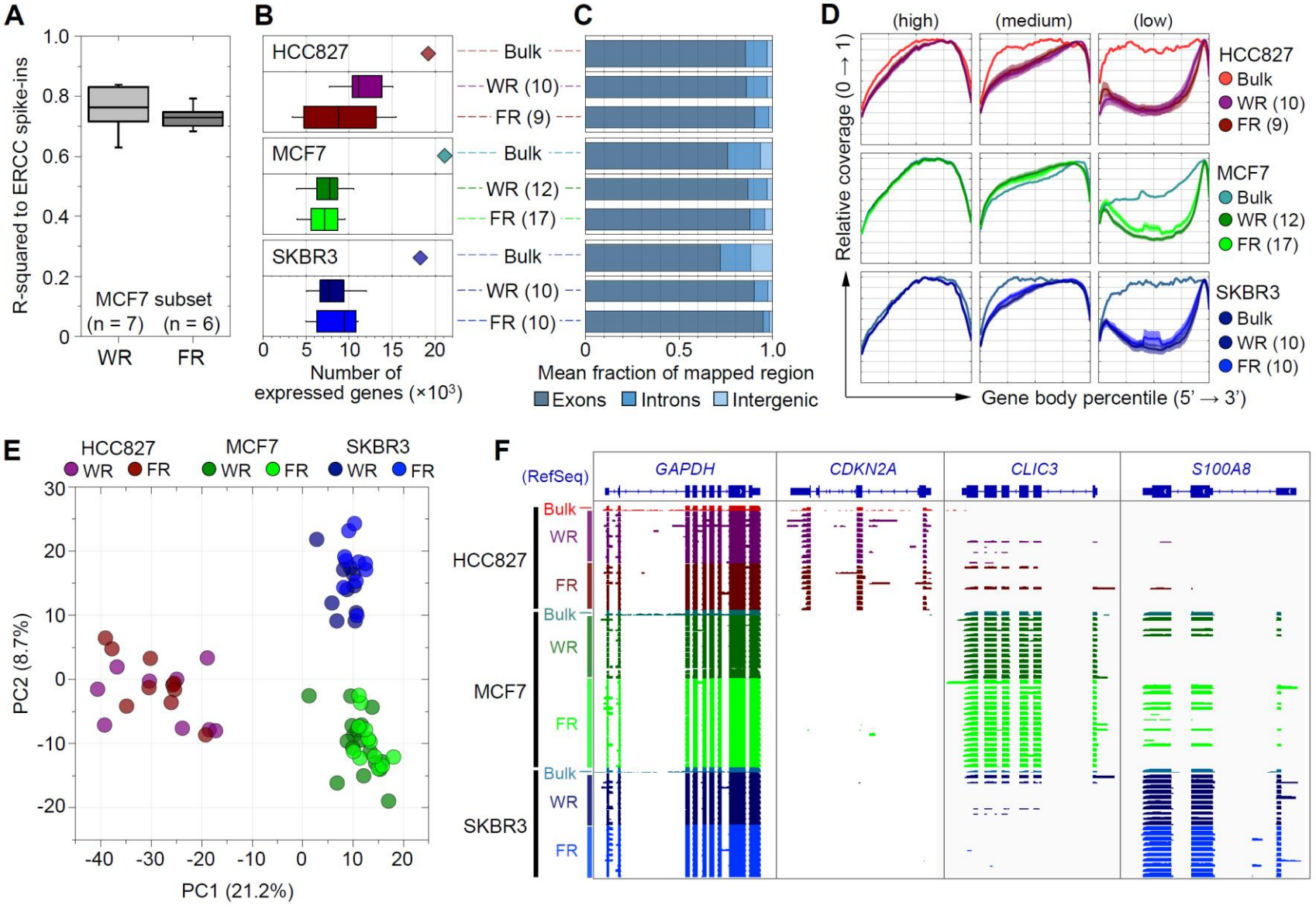
Simultaneous isolation of DNA/RNA from single cells



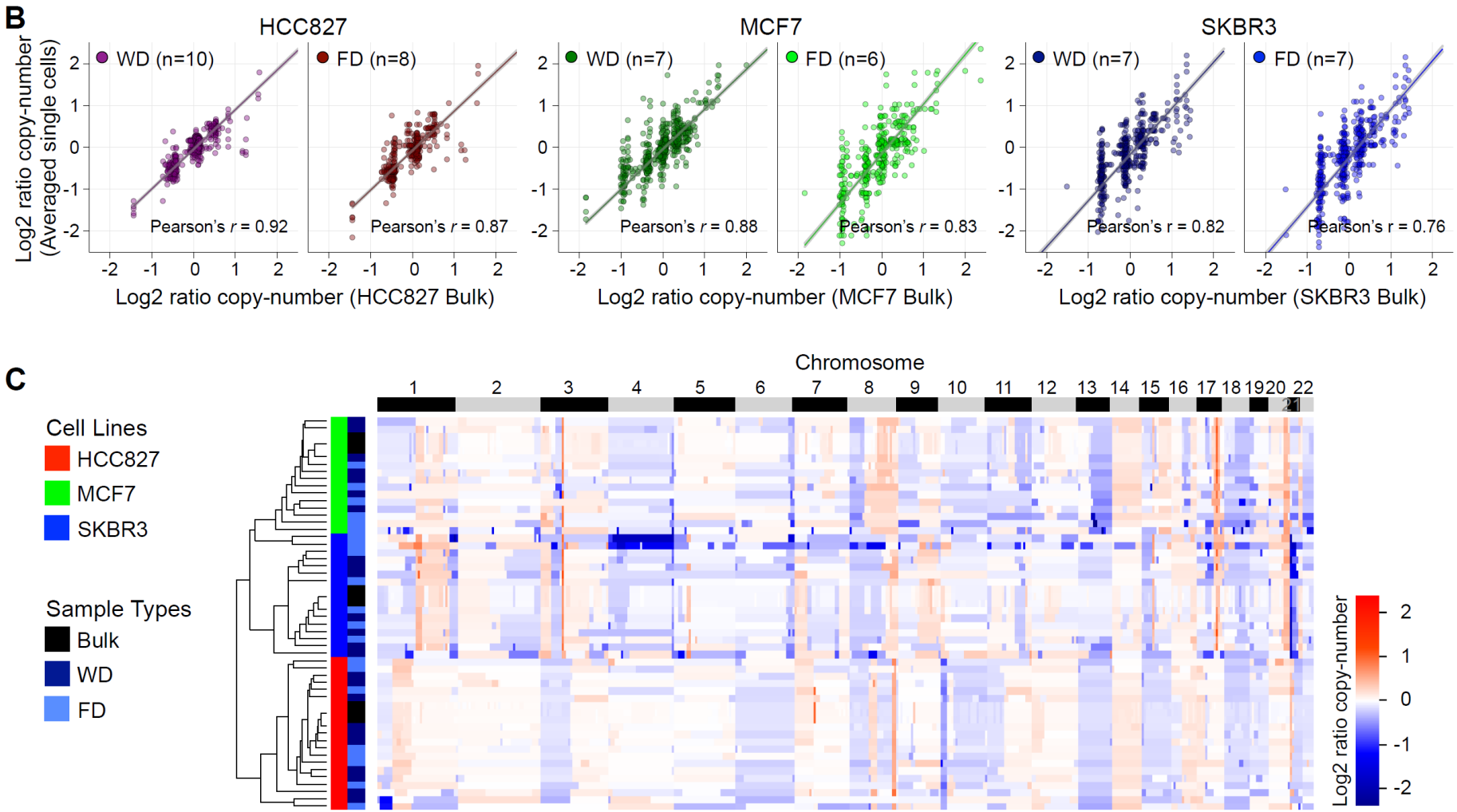


Han KY et al. Genome Research, 2018

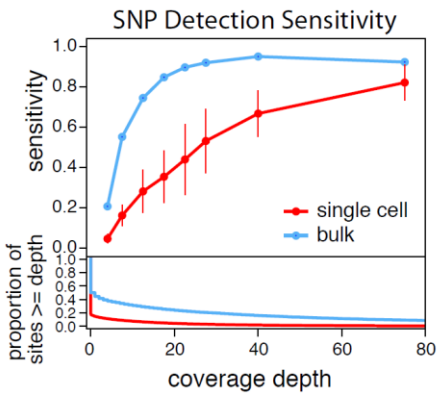
Compatible quality of fractionated RNA sequencing

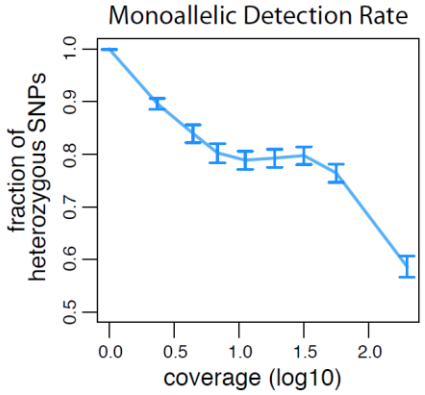


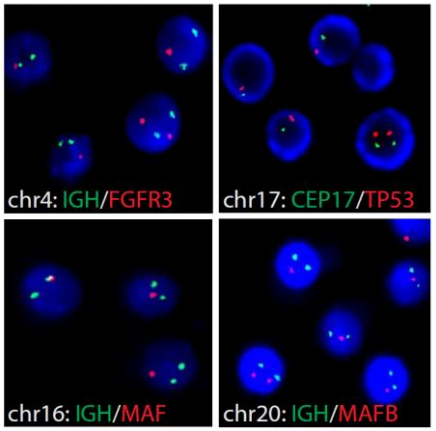
CNV detection using fractionated DNA WGS

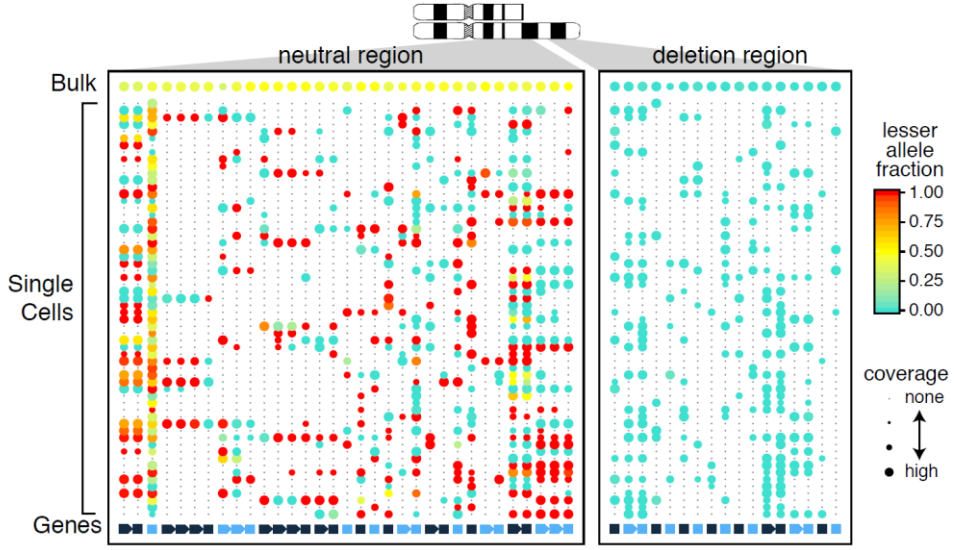


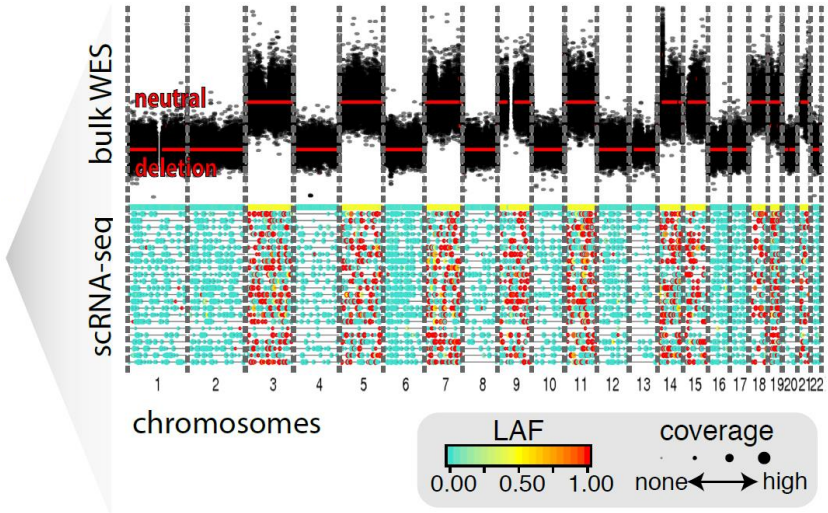
Linking transcriptional and genetic tumor heterogeneity





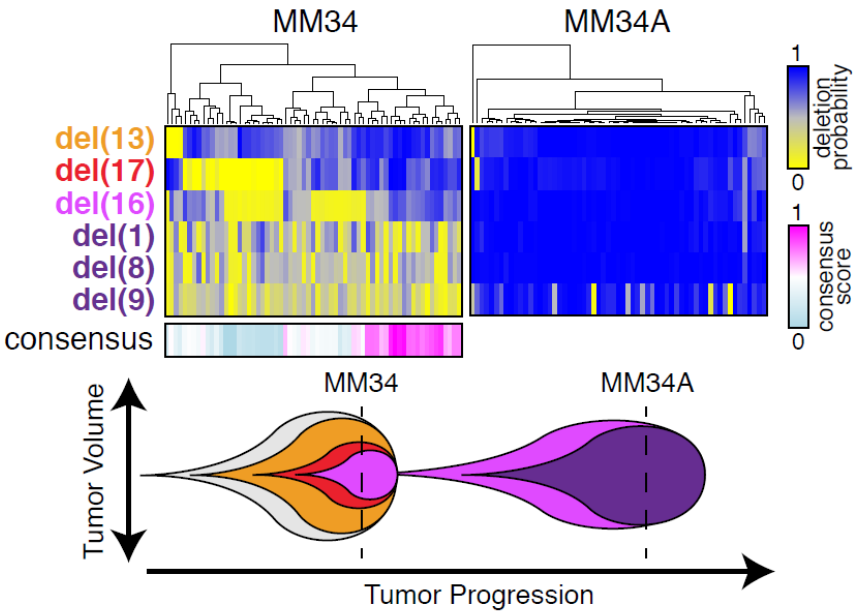


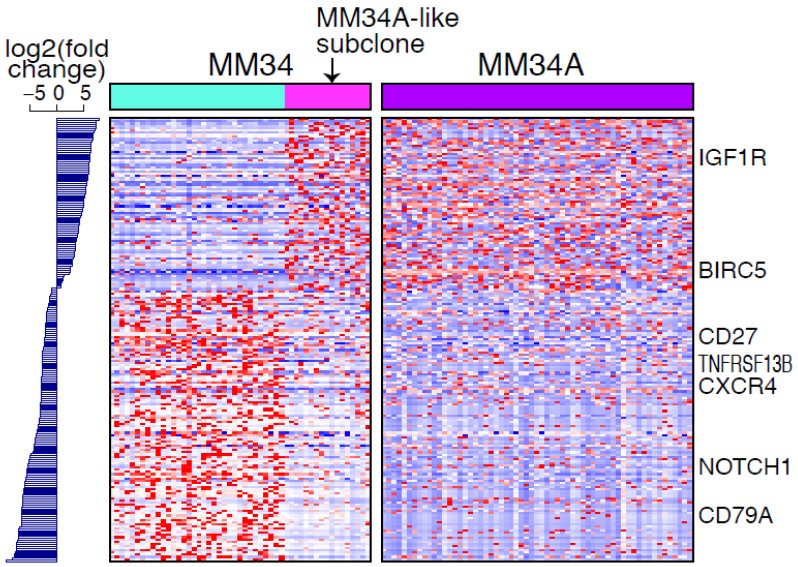




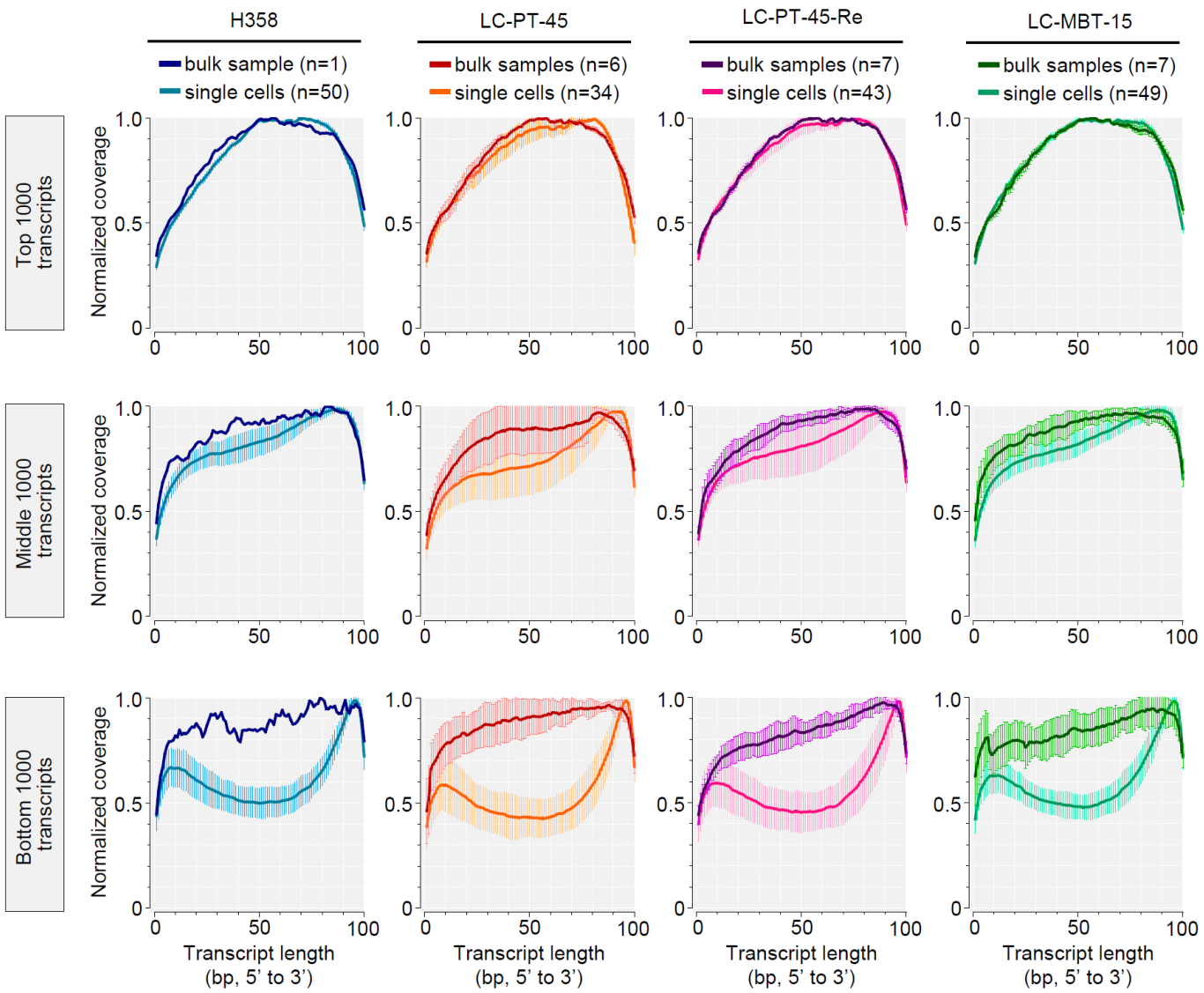
Wang G et al. Genome Research, in revision

Subclonal evolution in progressive multiple myeloma patient



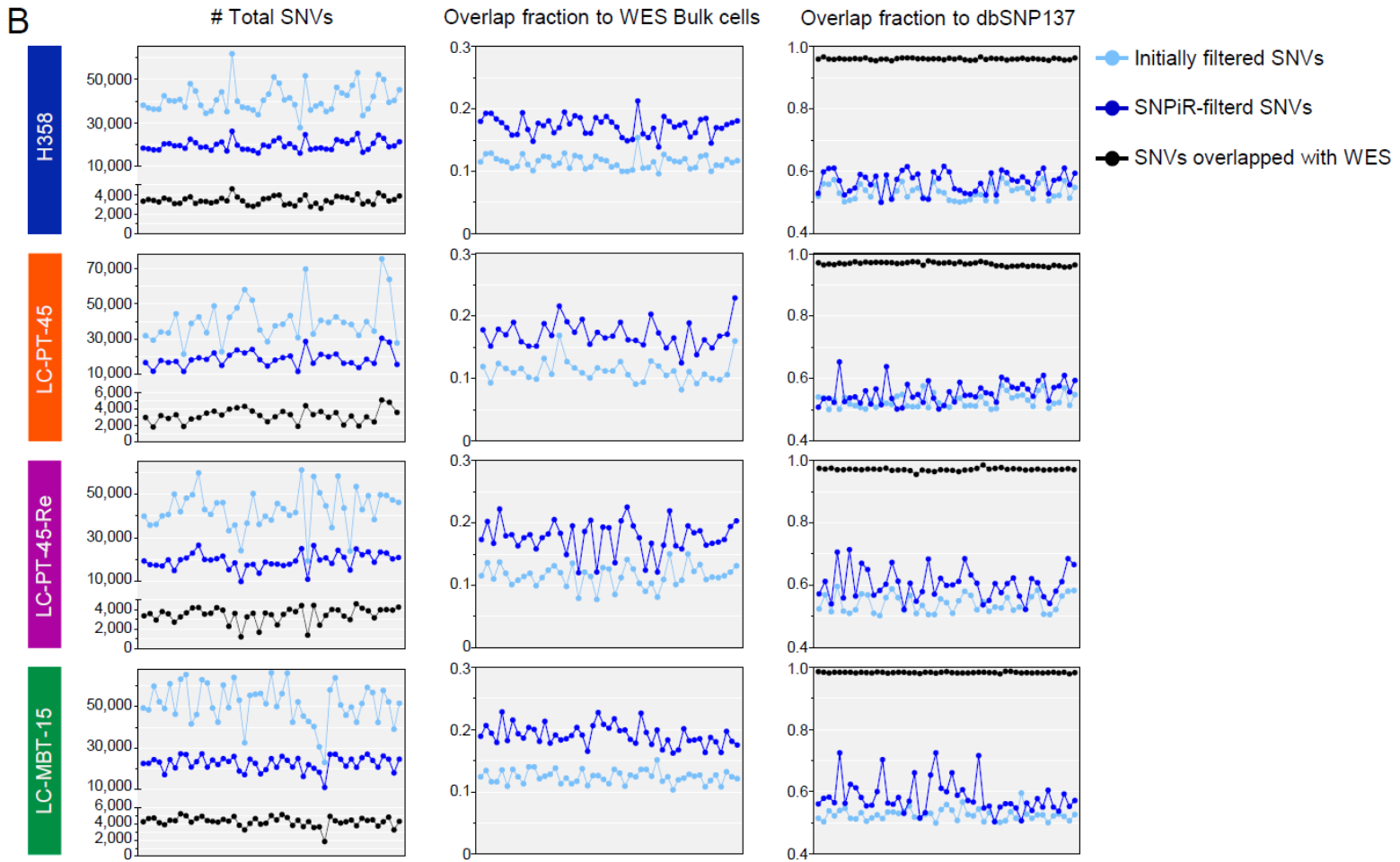


Skewed coverage at the 3' end of transcripts

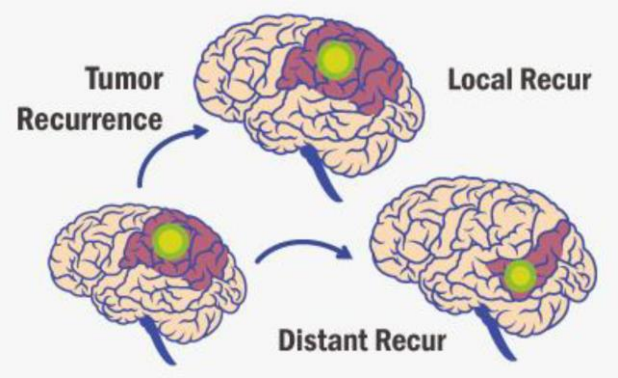


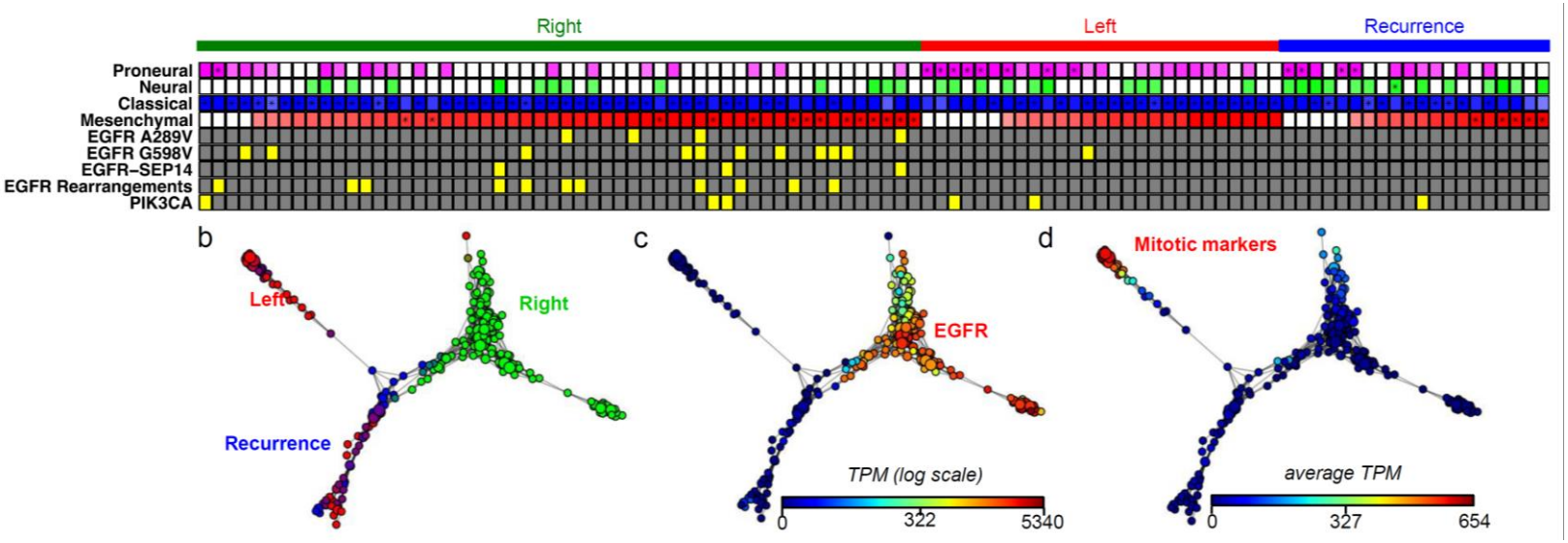
Kim KT *et al.* Genome Biology 2015

- SNV call from single-cell RNA-seq data



Single cell analysis to understand tumor evolution



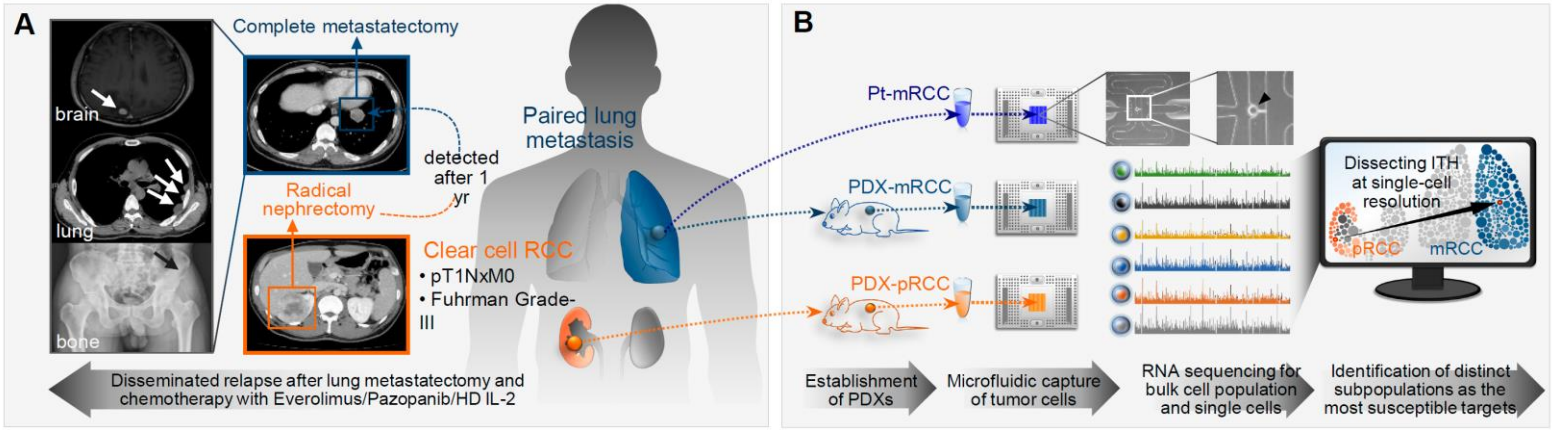


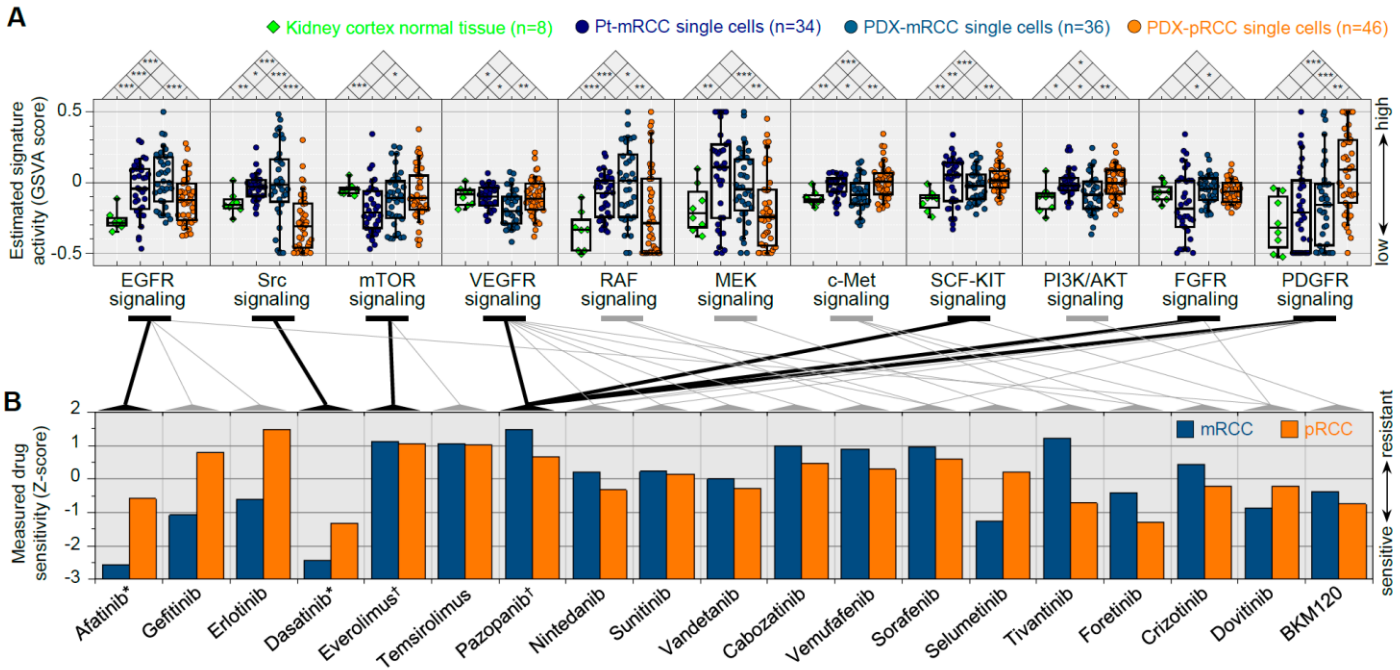
Sa JK et al. Nature Genetics, 2017

Clinical utility of single cell genome analysis

1. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. Genome Biol. 2016 Apr 29;17:80.
2. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017 May 5;8:15081.

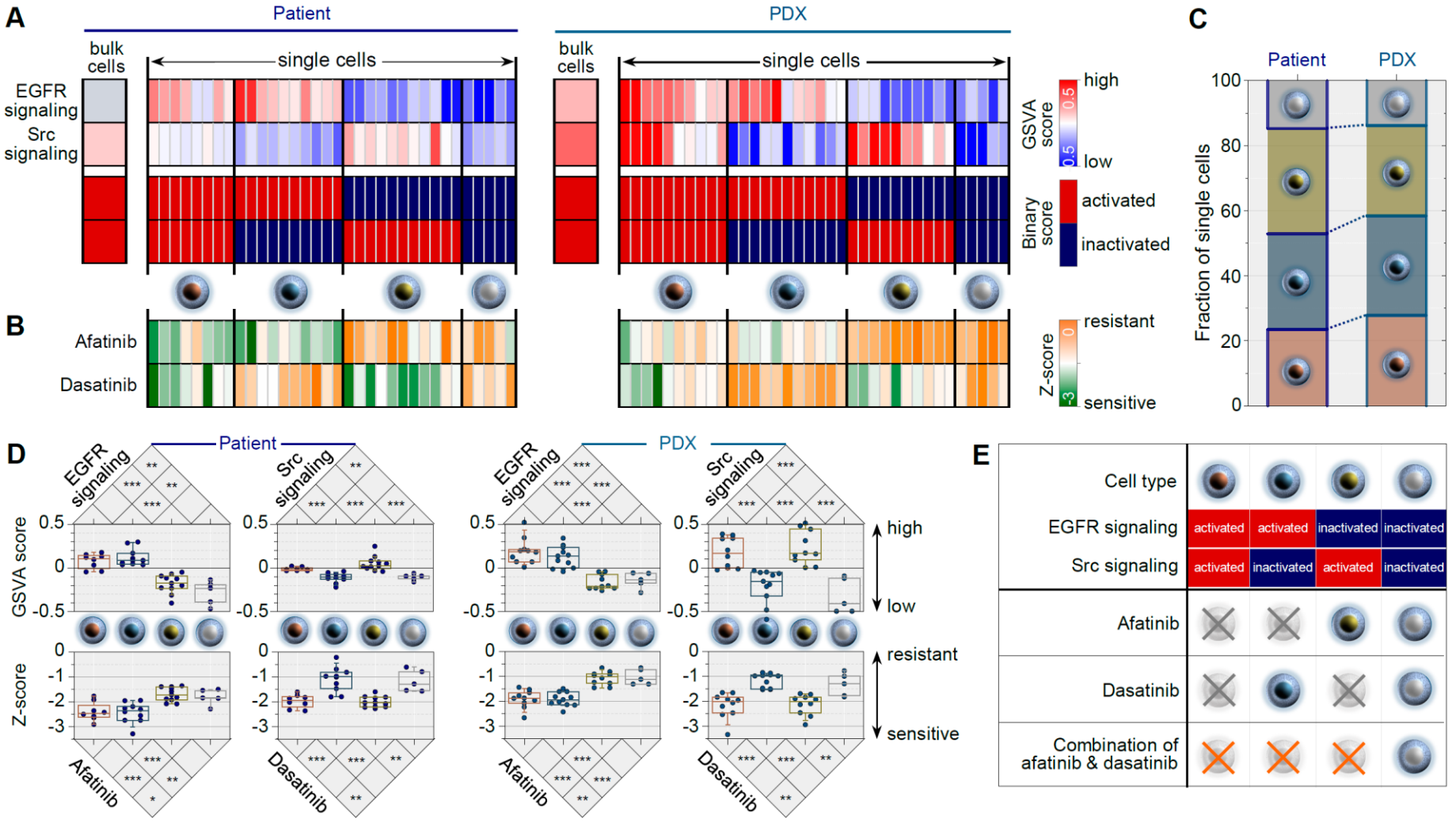
Combination therapy based on subclone analysis



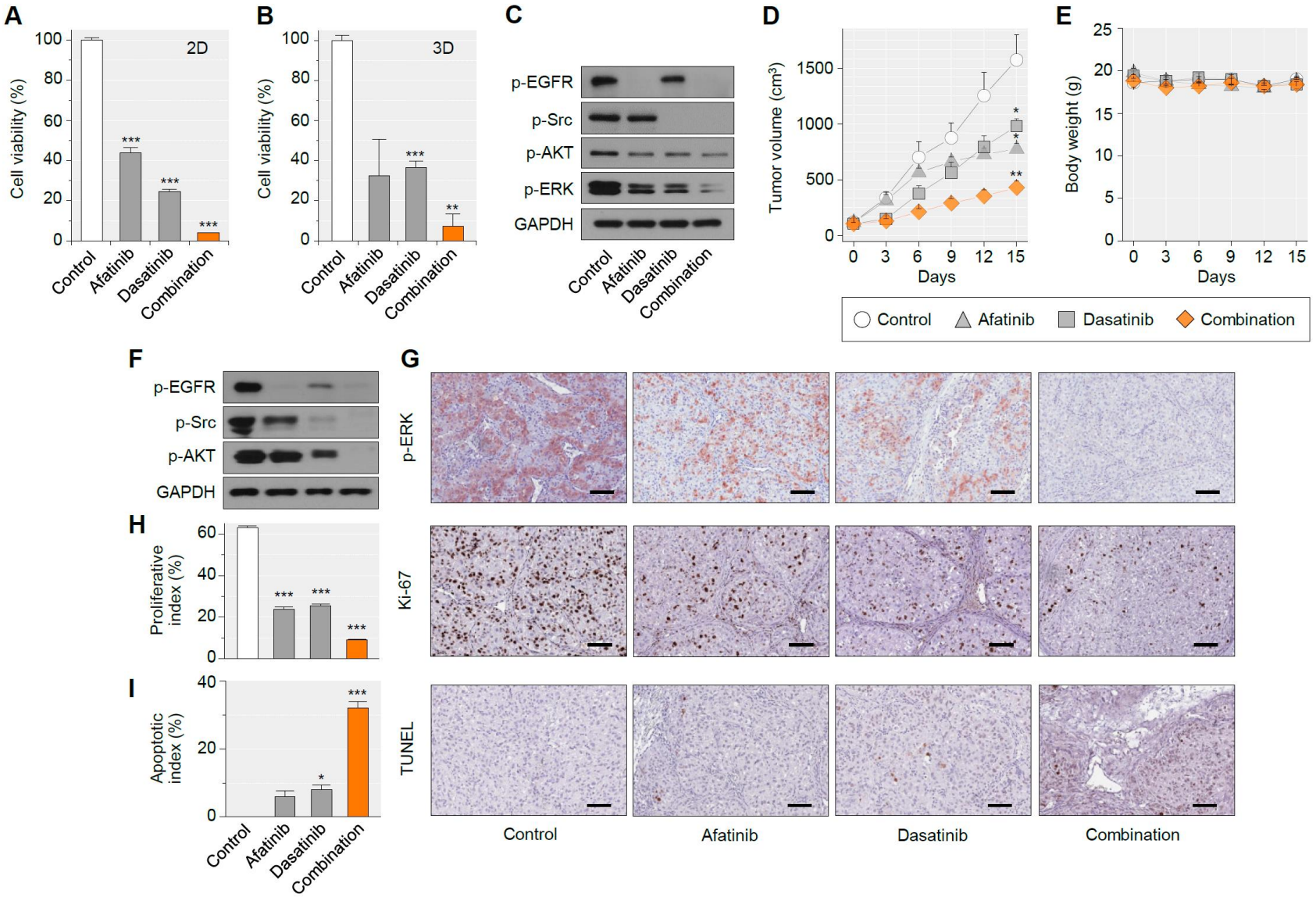


Kim KT, Lee HW et al. Genome Biology, 2016

Combinatorial therapy based on single cell profile



Validation of combinatorial therapy in PDC and PDX

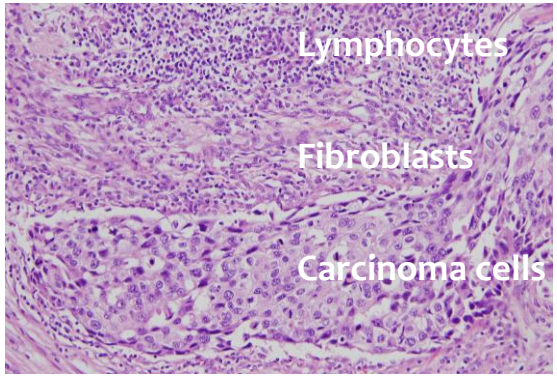


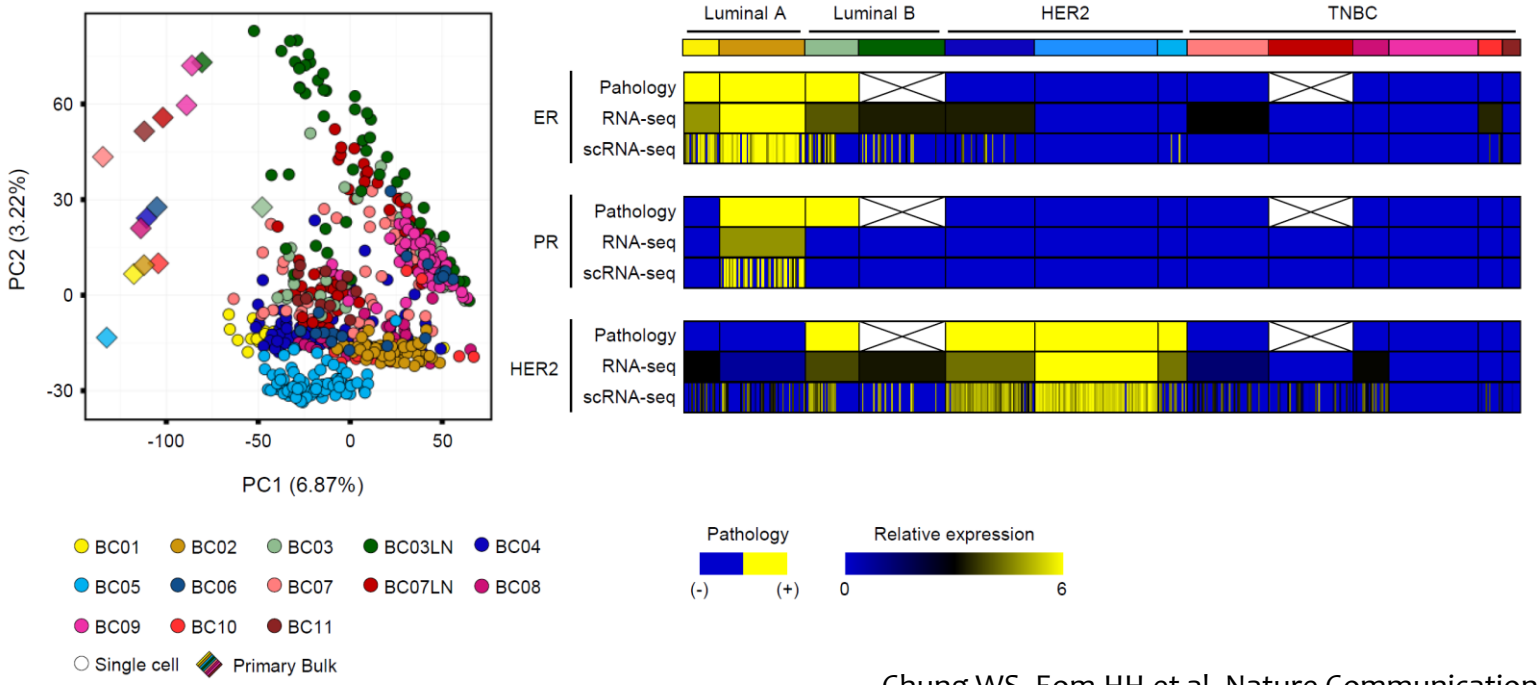
Subclonal pattern of drug response

- Intratumoral heterogeneity hampers the success of marker-based anticancer treatment because the targeted therapy may eliminate a specific subpopulation of tumor cells while leaving others unharmed.
- Using single-cell RNA sequencing (RNA-seq), we examine the intratumoral heterogeneity of a pair of primary renal cell carcinoma and its lung metastasis.
- Activation of drug target pathways demonstrates considerable variability between the primary and metastatic sites, as well as among individual cancer cells within each site.
- Based on the prediction of multiple drug target pathway activation, we derive a combinatorial regimen co-targeting two mutually exclusive pathways for the metastatic cancer cells.
- This combinatorial strategy shows significant increase in the treatment efficacy over monotherapy in the experimental validation using patient-derived xenograft platforms in vitro and in vivo.

scRNA-seq of breast cancer single cells

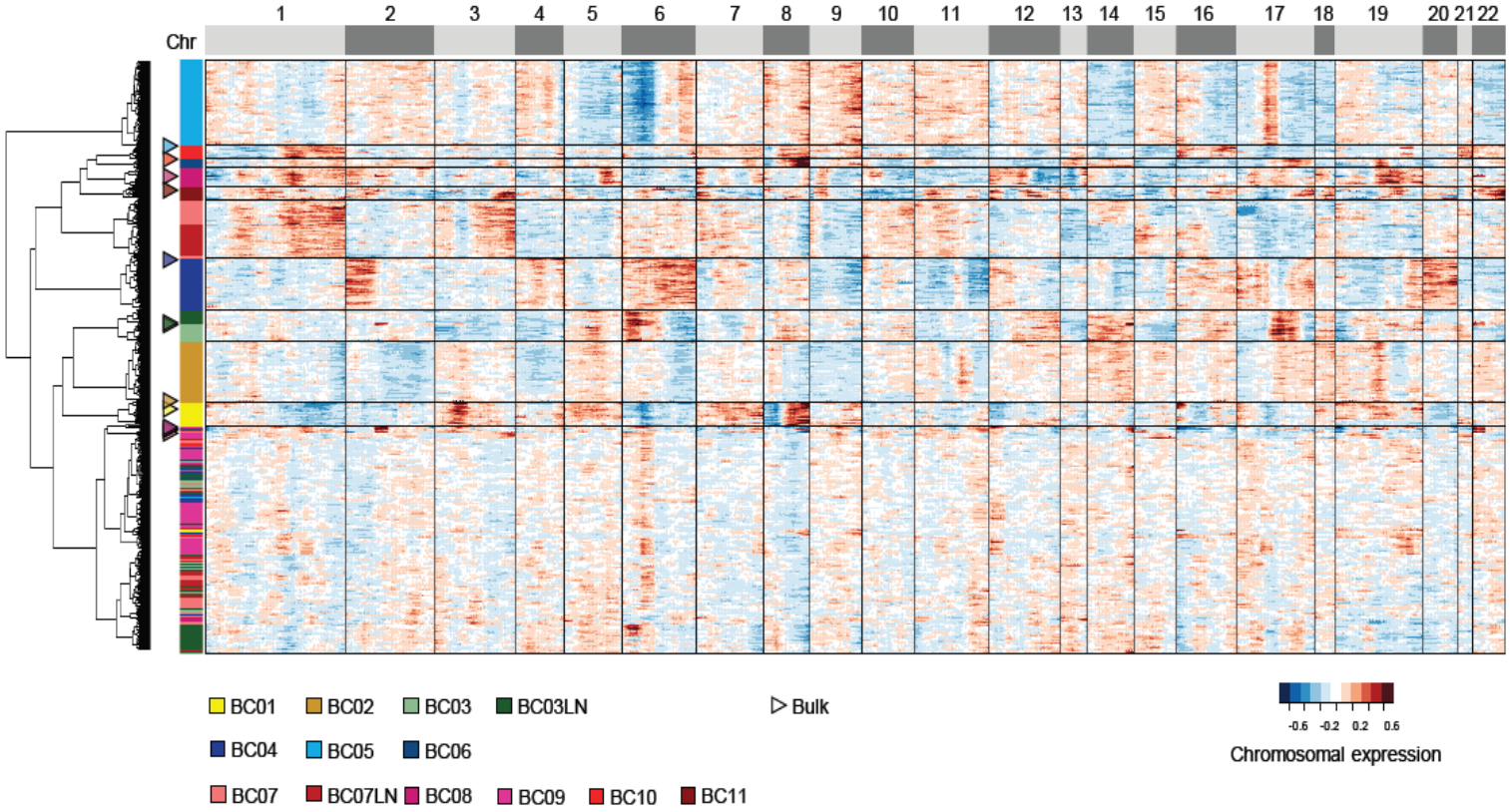
- Surgically resected primary chemo-naïve breast cancer
  - Luminal A (2): BC01, BC02
  - Luminal B (1): BC03 (BC03LN)
  - HER2+ (3): BC04, BC05, BC06
  - TNBC (5): BC07 (BC07LN), BC08, BC09, BC10, BC11
- RNA-seq analysis in 515 single cells from 11 patients
  - WES and RNA-seq for bulk tumor
  - RNA-seq for single cell analysis



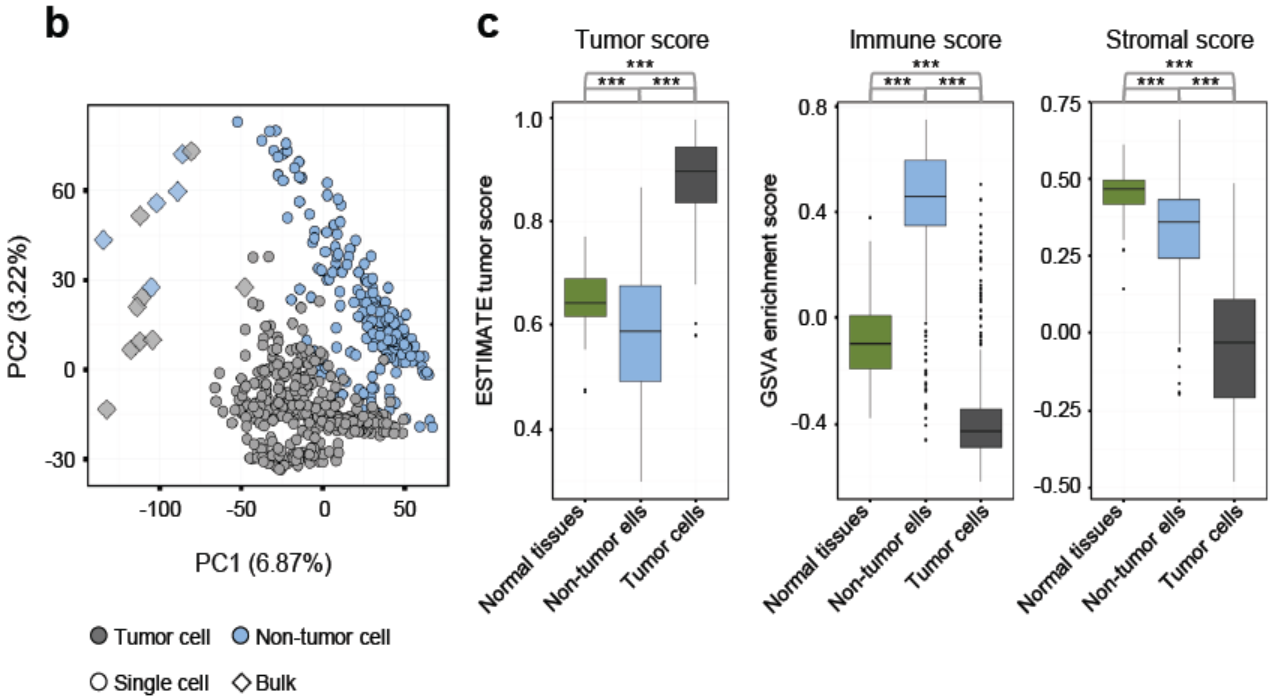


Chung WS, Eom HH et al. Nature Communications, 2017

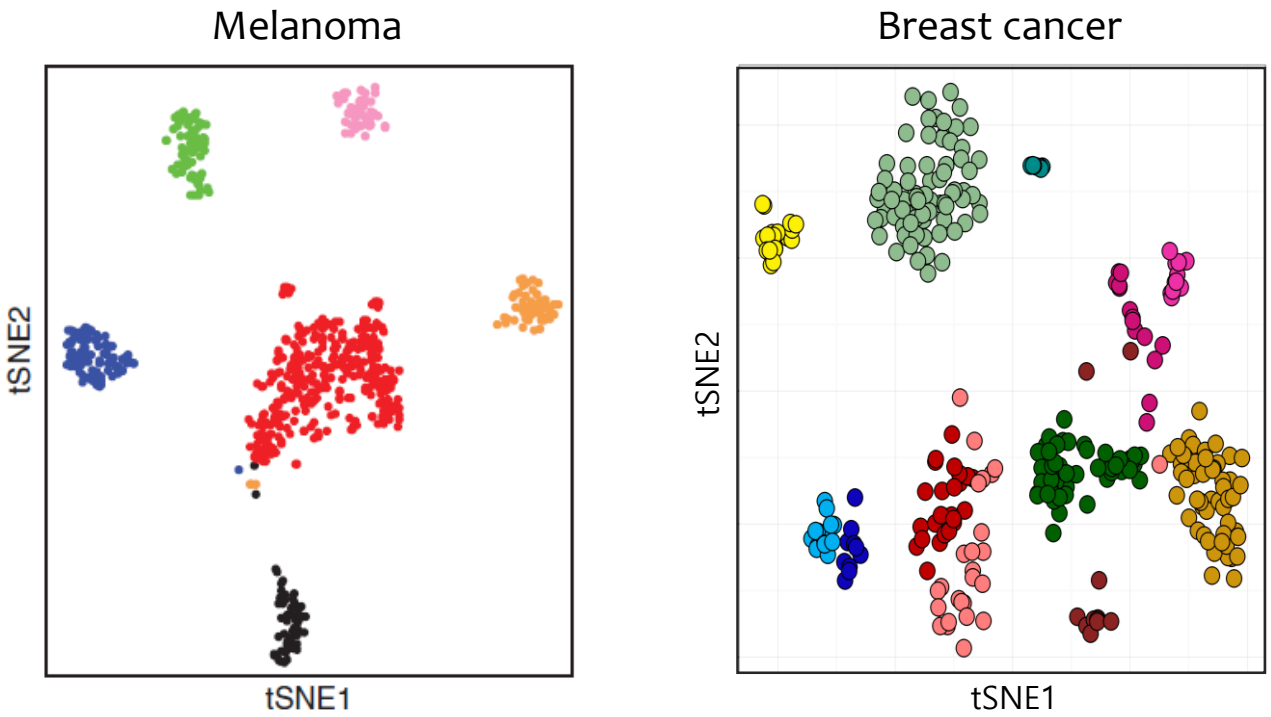
Inferred genomic features for tumor and non-tumor cells



Inferred CNV, is it reliable?

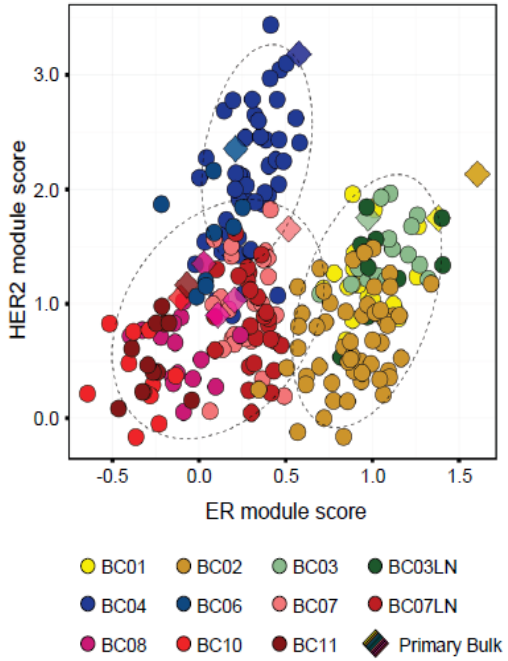


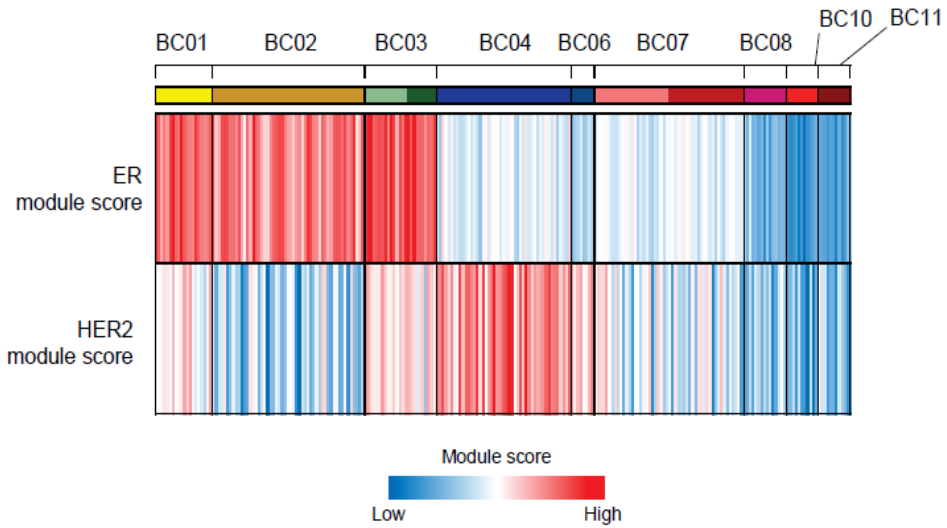
Inter-tumoral heterogeneity in tumor cells

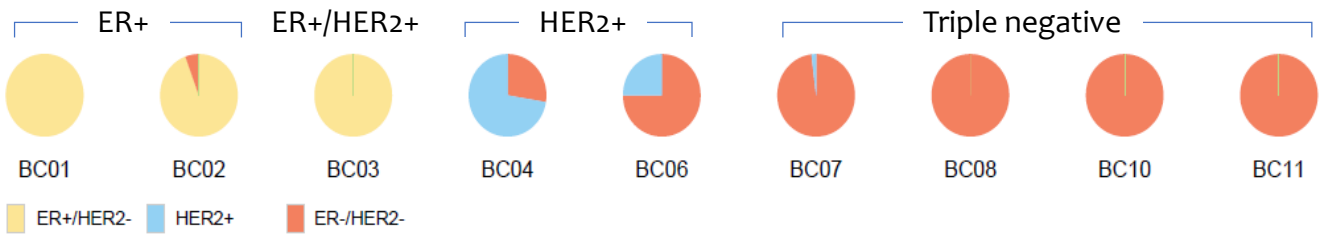


Tirosh et al. (2016) Science

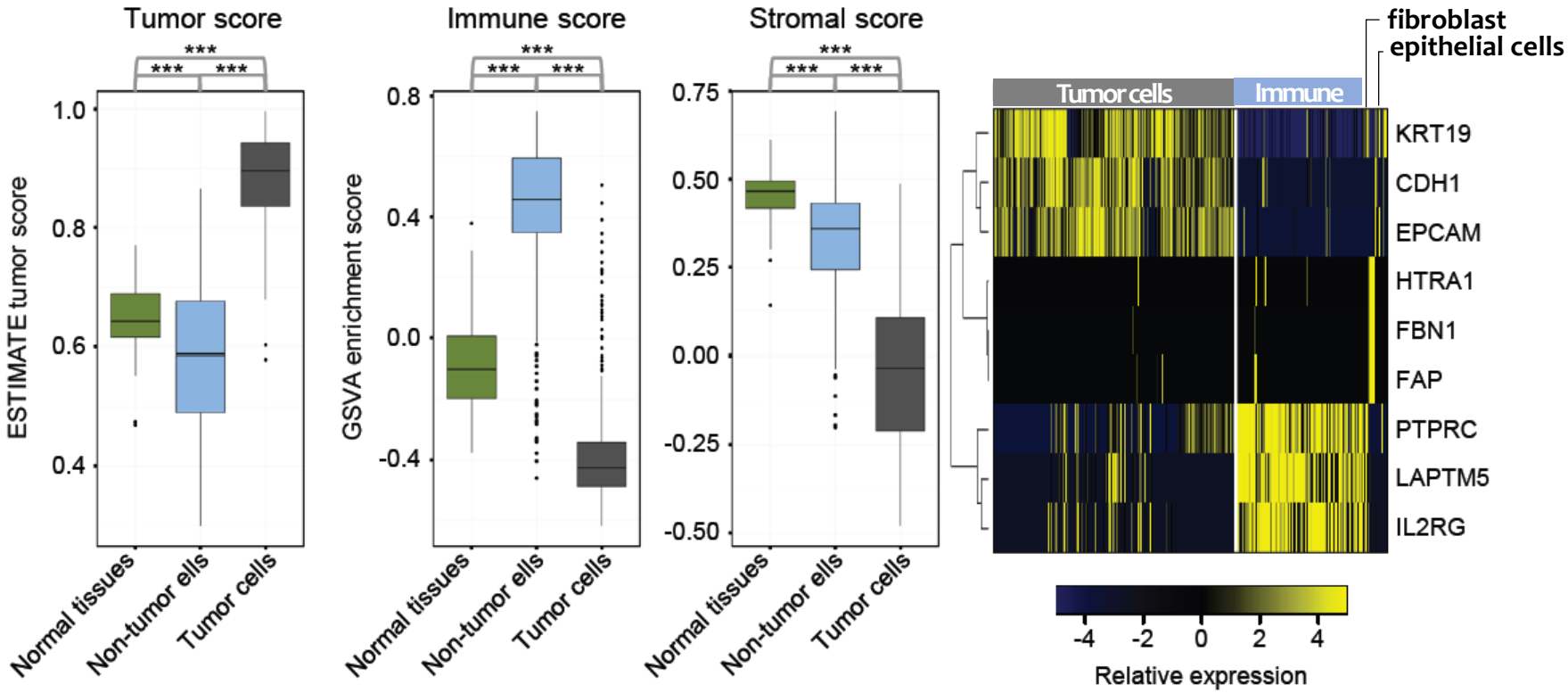
Mixed subtypes in breast cancer

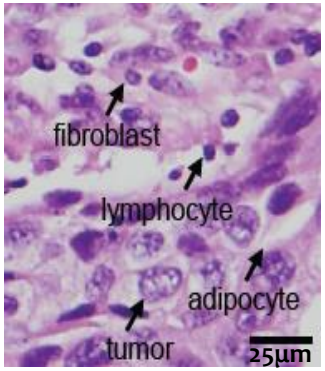




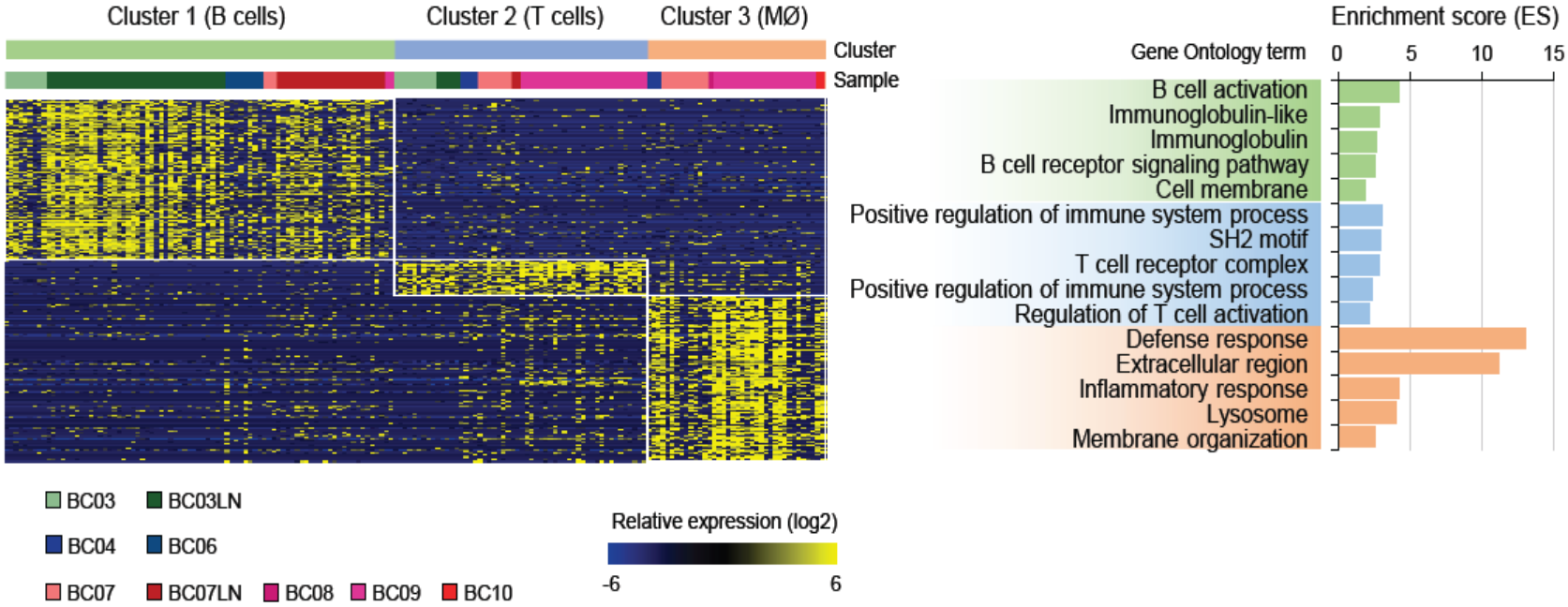


Recovery of immune cells from single cell isolation

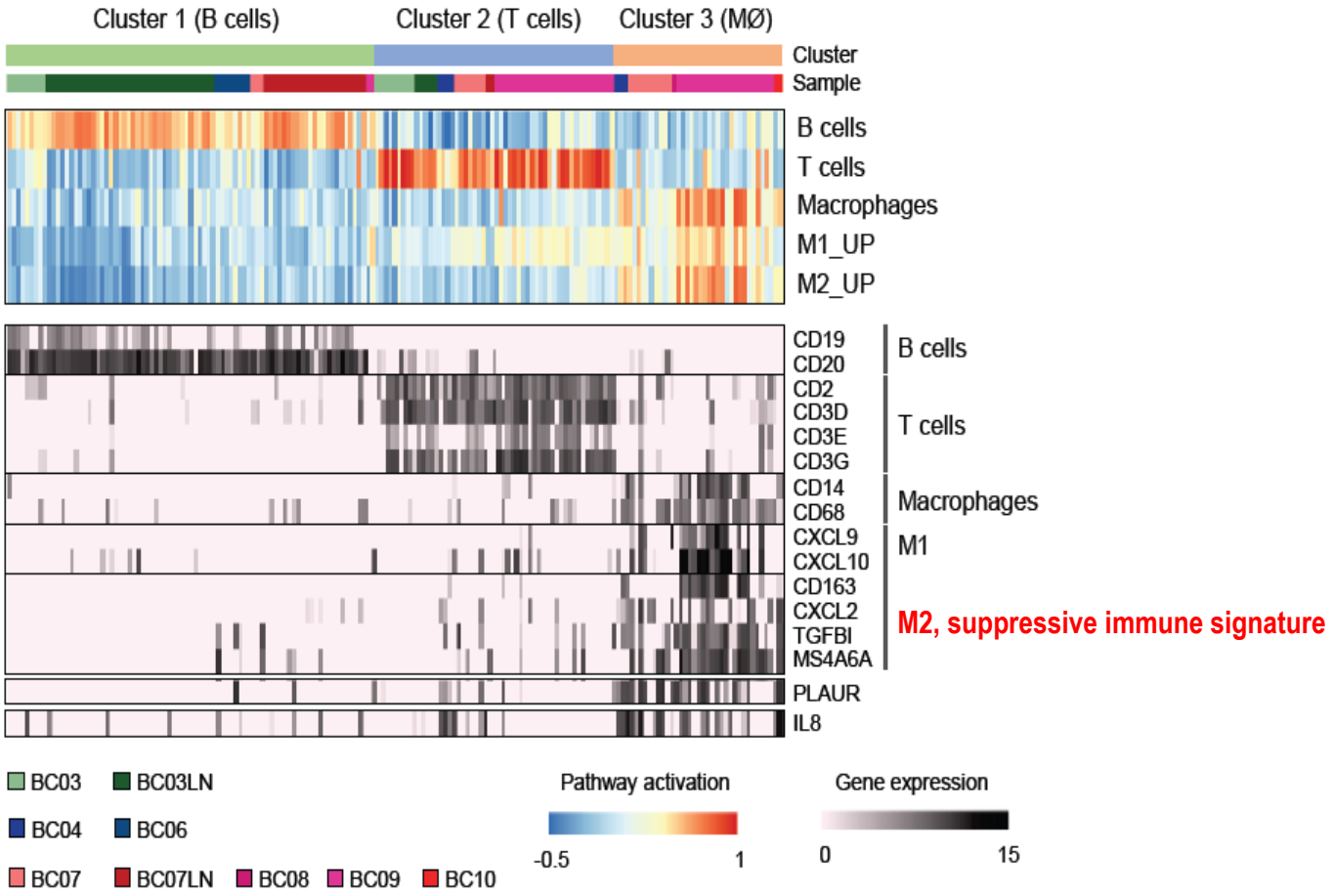




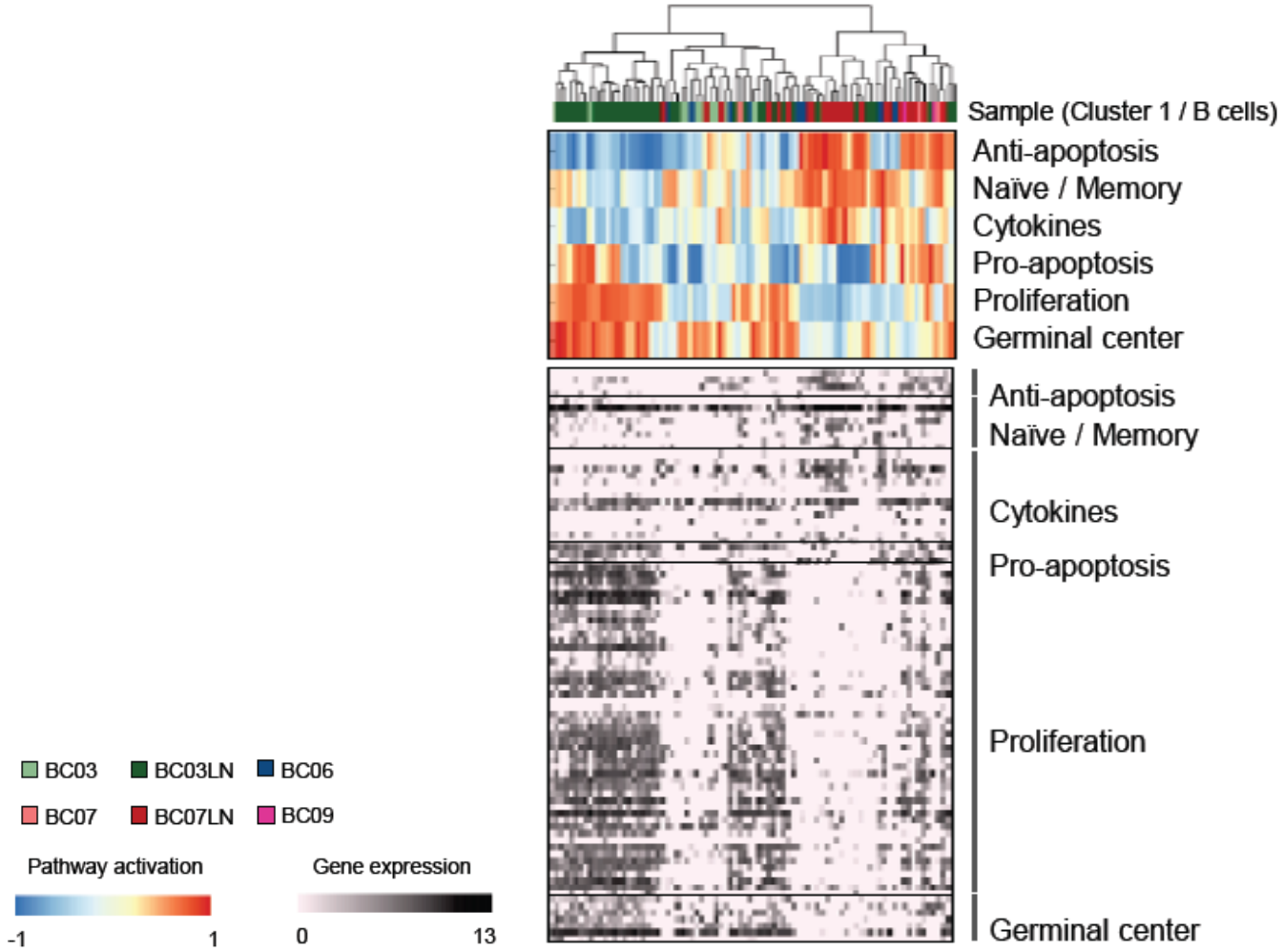
Three types of immune cells



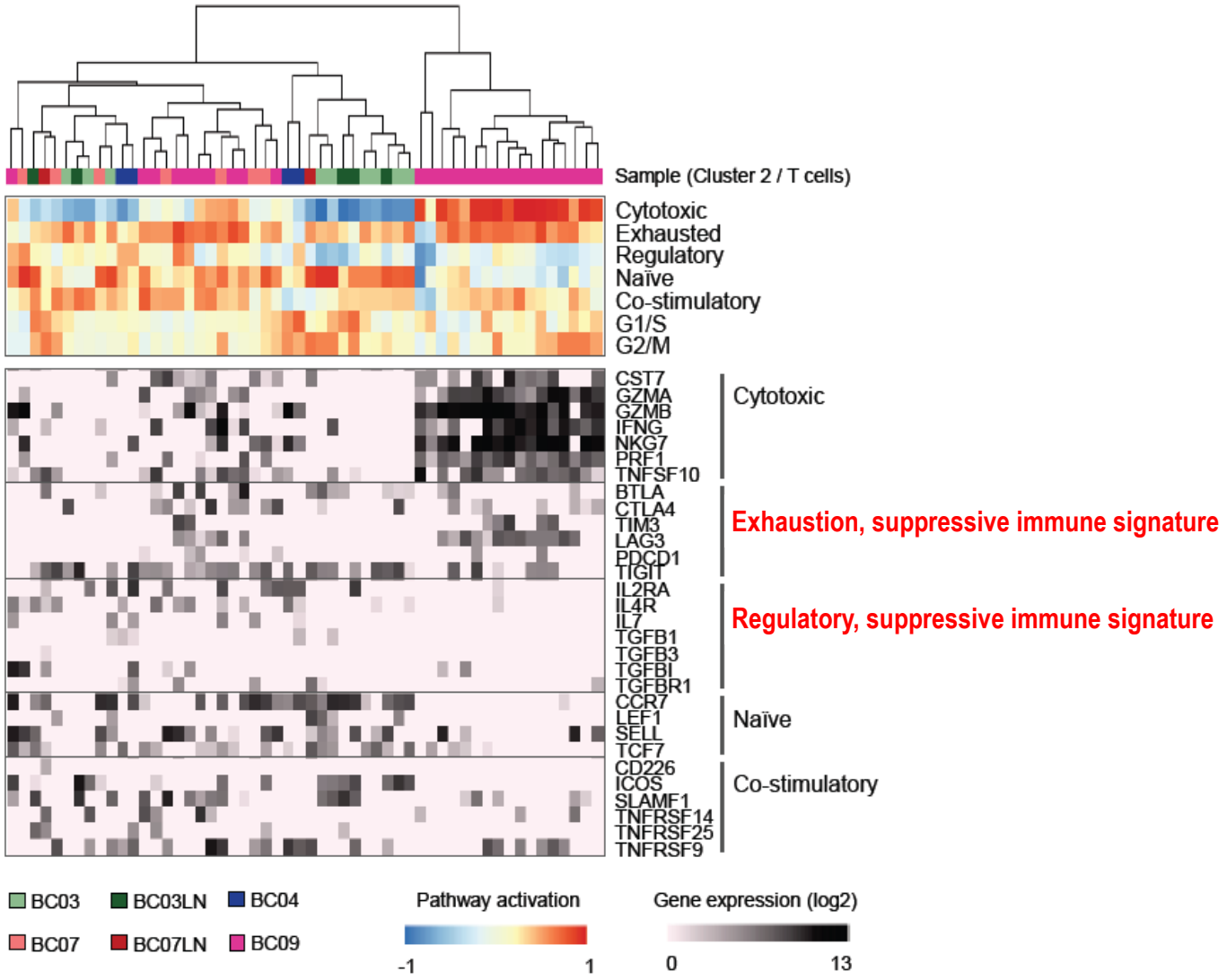
Tumor associated macrophages in breast cancer



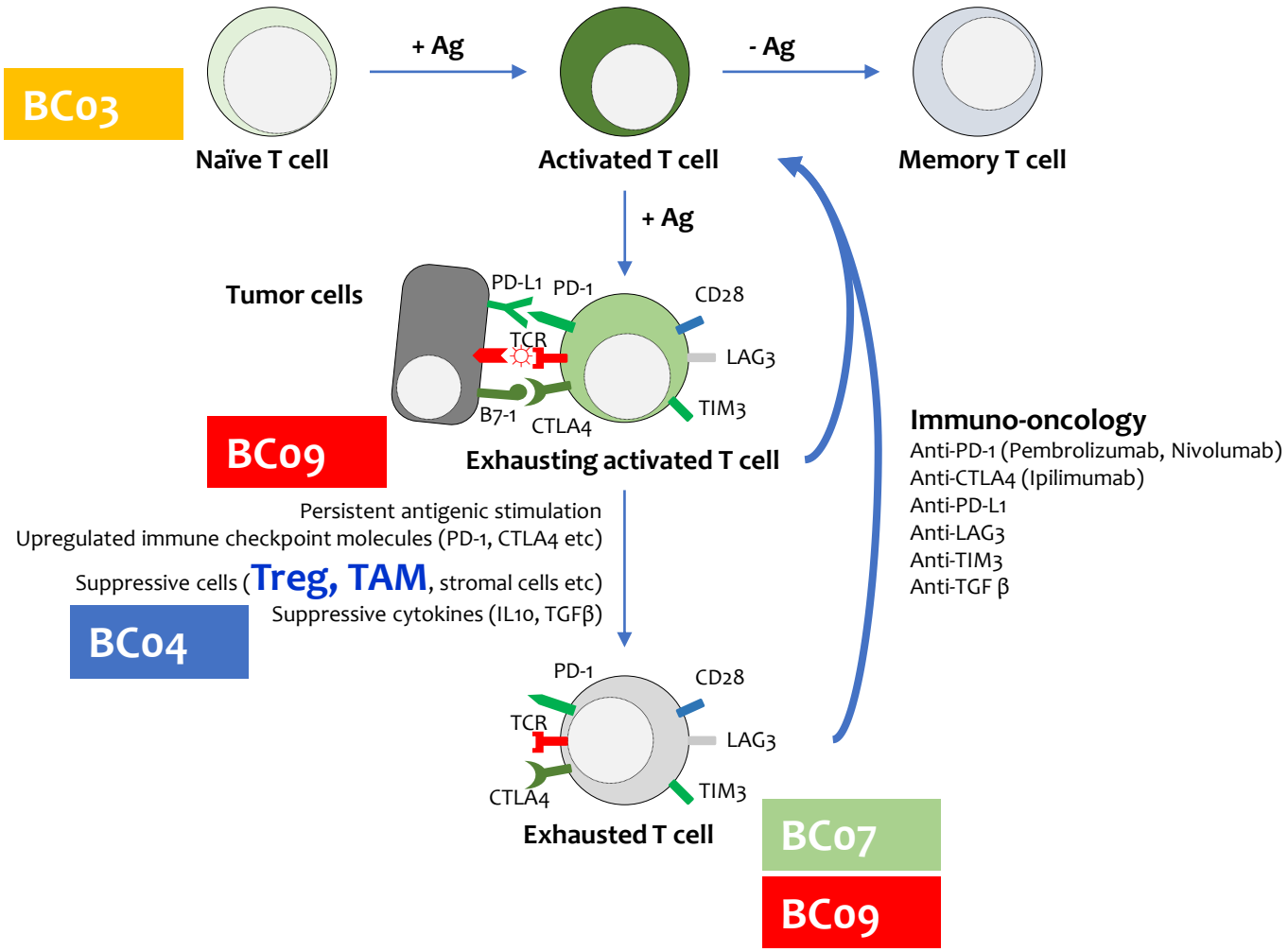
Tumor-infiltrating B cells in breast cancer



Tumor-infiltrating T cells in breast cancer



Exhaustion of tumor-infiltrated T lymphocytes



Personal tumor immune profiling

- Single-cell transcriptome profiling of tumour tissue isolates allows the characterization of heterogeneous tumour cells along with neighbouring stromal and immune cells.
- We adopt this powerful approach to breast cancer and analyse 515 cells from 11 patients. Inferred copy number variations from the single-cell RNA-seq data separate carcinoma cells from non-cancer cells.
- At a single-cell resolution, carcinoma cells display common signatures within the tumour as well as intratumoral heterogeneity regarding breast cancer subtype and crucial cancer-related pathways.
- Most of the non-cancer cells are immune cells, with three distinct clusters of T lymphocytes, B lymphocytes and macrophages.
- T lymphocytes and macrophages both display immunosuppressive characteristics: T cells with a regulatory or an exhausted phenotype and macrophages with an M2 phenotype.