Session 9: Tumor Evolution

Emerging Approaches for Tumor Analyses in Epidemiological Studies

March 28, 2023

9:30 AM- 12:00 PM

Session Overview

- ➔ Introduction to tumor evolution
- ➔ Tumor heterogeneity
- ➔ Cancer subclonal reconstruction from DNA sequencing
- ➔ Timing somatic events in the evolution of cancer
- ➔ Evolutionary dynamics extrachromosomal DNA (ecDNA) in human cancers

Introduction to tumor evolution

Definition of tumor evolution

Tumor evolution refers to the **dynamic process** by which tumors **change and adapt** over time in response to various **biological, environmental, and therapeutic factors**. Tumor evolution encompasses the **genetic and epigenetic alterations** that occur within a tumor, as well as the interactions between the **tumor and its microenvironment**. This process results in the acquisition of new biological properties and functions that drive tumor growth, progression, and resistance to treatment. Tumor evolution is a **complex and heterogeneous process** that contributes to the diverse and evolving nature of cancer, and understanding its mechanisms is critical for developing effective strategies for cancer diagnosis, treatment, and prevention.

The Clonal Evolution of Tumor Cell Populations

Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression.

Peter C. Nowell

Peter C. Nowell, Science, 1976

The Journal of Heredity 68:3–10. 1977.

"Nothing in biology makes sense except in the light of evolution"

Theodosius Dobzhansky: 1900–1975

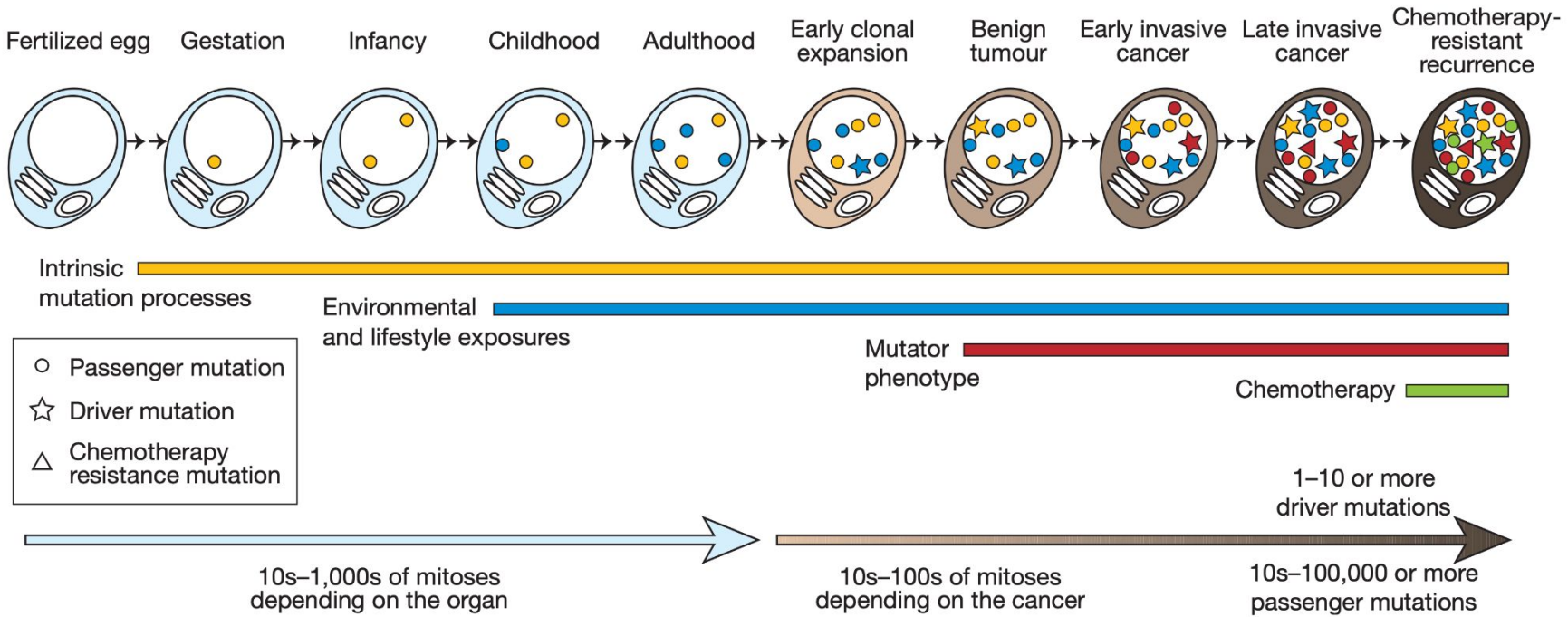
FRANCISCO J. AYALA

Dobzhansky, T. *Am. Biol. Teach.* 35, 125–129 (1973).

Importance of understanding tumor evolution

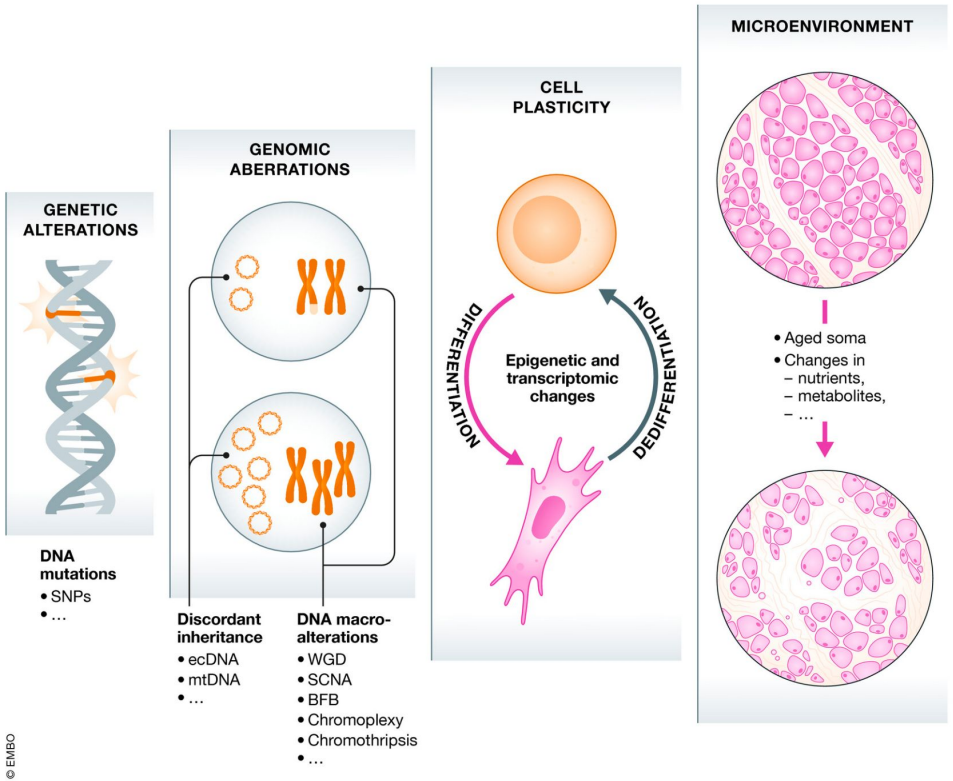
- Tumor evolution -> **heterogeneity**; Identify and distinguish the clone heterogeneous population -> **Improving Cancer Diagnosis**
- Tumor evolution -> **treatment resistant**; Inform the development of new effective treatments -> **Improving Cancer Treatment**
- Tumor evolution -> **diverse subpopulation**; guide the development of personalized medicine approaches -> **Personalized Medicine**
- Tumor evolution -> **cancer recurrence and metastasis**; Predict the likelihood of recurrence or response to therapy -> **Improving Cancer Prognosis**

Overview of tumor evolution process



Stratton et al., *Nature*, 2009

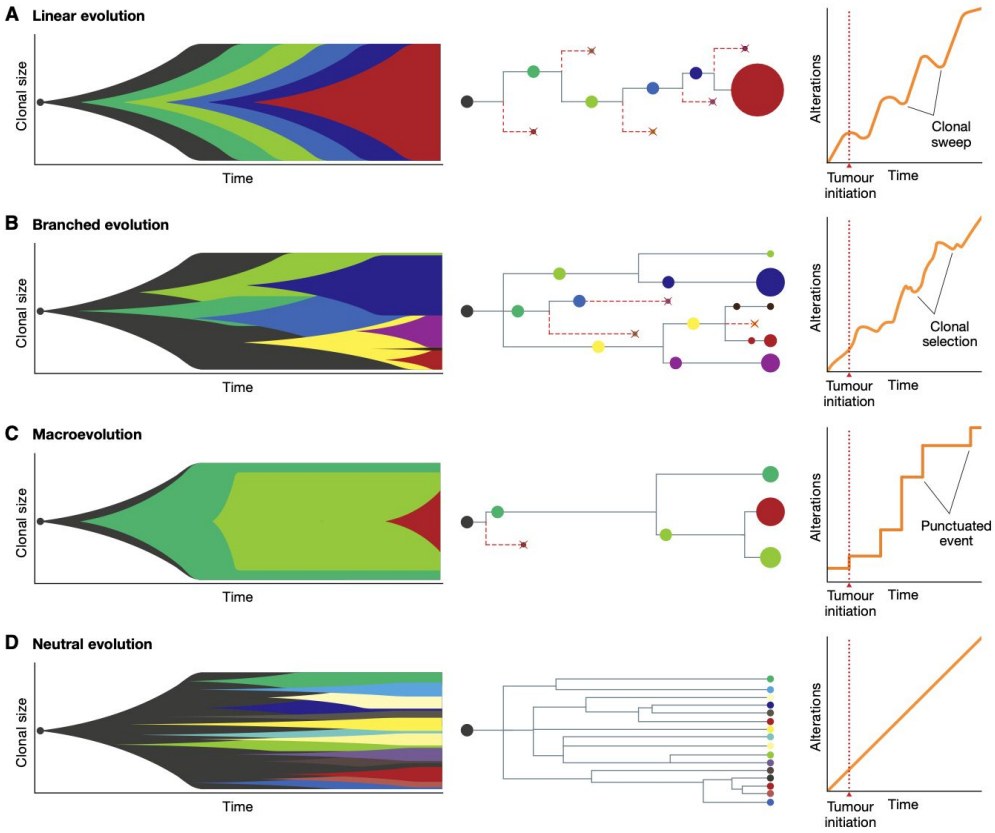
Mechanisms of tumor evolution



Schematic illustration of the different determinants of tumour evolution, which influence evolutionary trajectories through highly interdependent mechanisms, from a microscopic (left) to a macroscopic (right) scale.

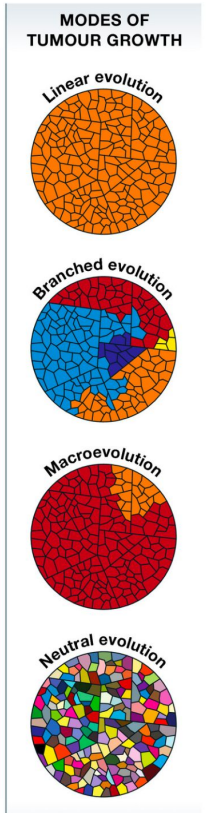
Vendramin et al. 2021

Models of tumor evolution



Models of linear evolution (A), branched evolution (B), macroevolution (C) and neutral evolution (D) described by Muller plots representing dynamic changes in clonal size over time (left), clonal lineages and phylogenetic trees (centre) and changes in the number of alterations over time (right). Colours indicate different clones.

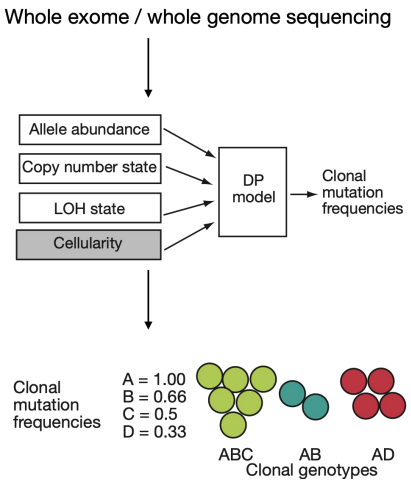
[Vendramin et al. 2021]



Observation at the time of the diagnosis

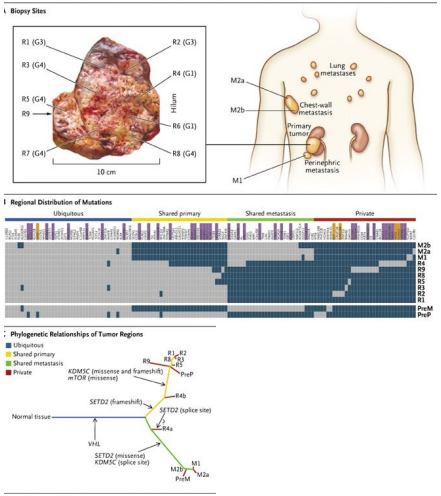
Approaches for studying tumor evolution

Bulk sequencing and clone decomposition



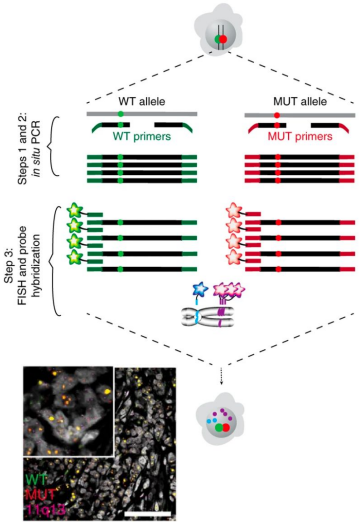
[(Shah et al. 2012)]

Multi-regional sequencing



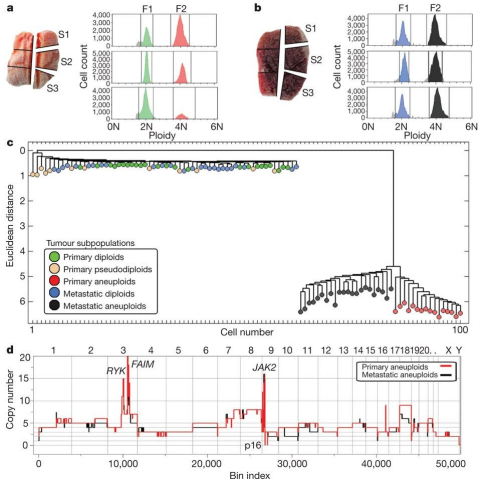
[(Gerlinger et al. 2012)]

In situ mutation detection

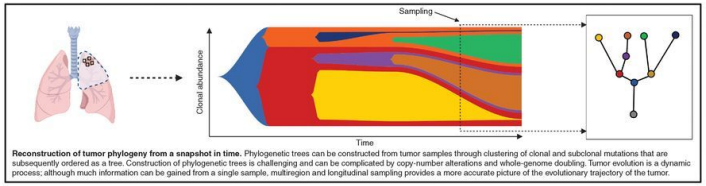


[(Janiszewska et al. 2015)]

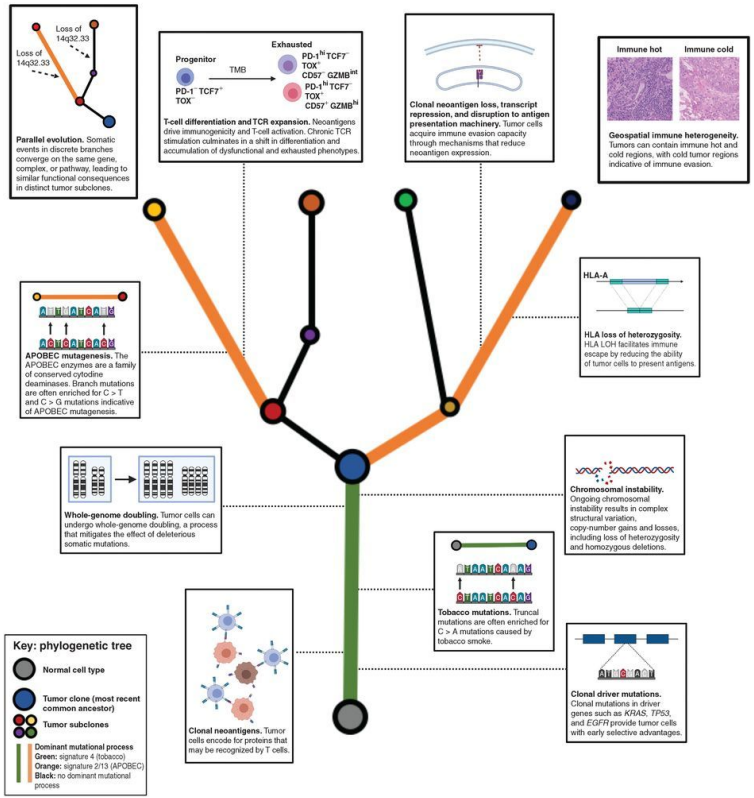
Single cell sequencing



[(Navin et al. 2011)]



Reconstruction of tumor phylogeny from a snapshot in time. Phylogenetic trees can be constructed from tumor samples through clustering of clonal and subclonal mutations that are subsequently ordered as a tree. Construction of phylogenetic trees is challenging and can be complicated by copy-number alterations and whole-genome doubling. Tumor evolution is a dynamic process; although much information can be gained from a single sample, multiregion and longitudinal sampling provides a more accurate picture of the evolutionary trajectory of the tumor.



Tracking Cancer Evolution through Therapy (The TRACERx Study)

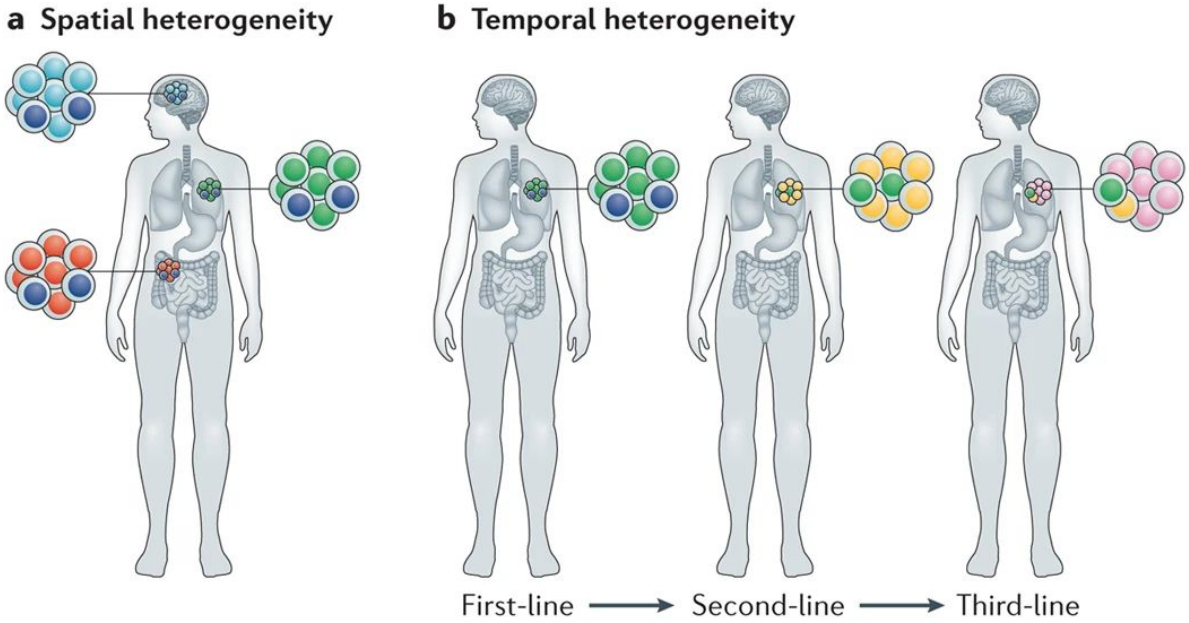
By integrating multiregion sequencing of primary tumors with longitudinal sampling of a prospectively recruited patient cohort, cancer evolution can be tracked from early- to late-stage disease and through therapy.

Tumor evolution in NSCLC. Evolutionary processes in NSCLC are outlined. Top, subclonal dynamics over time can be represented by a fish plot; however, a single sample in time provides only a snapshot. From this snapshot, tumor phylogeny can be inferred. Bottom, evolutionary processes generating immune and genomic heterogeneity are described as part of a "tree." Events that occur in the "trunk" are clonal, i.e., they occur within every cell in the tumor. Through tumor evolution, subclones can emerge through selection; events that occur in these subclones are known as "branch" events.

(Bailey et al. 2021)

Tumor heterogeneity

Intratumoural heterogeneity: Spatial vs Temporal

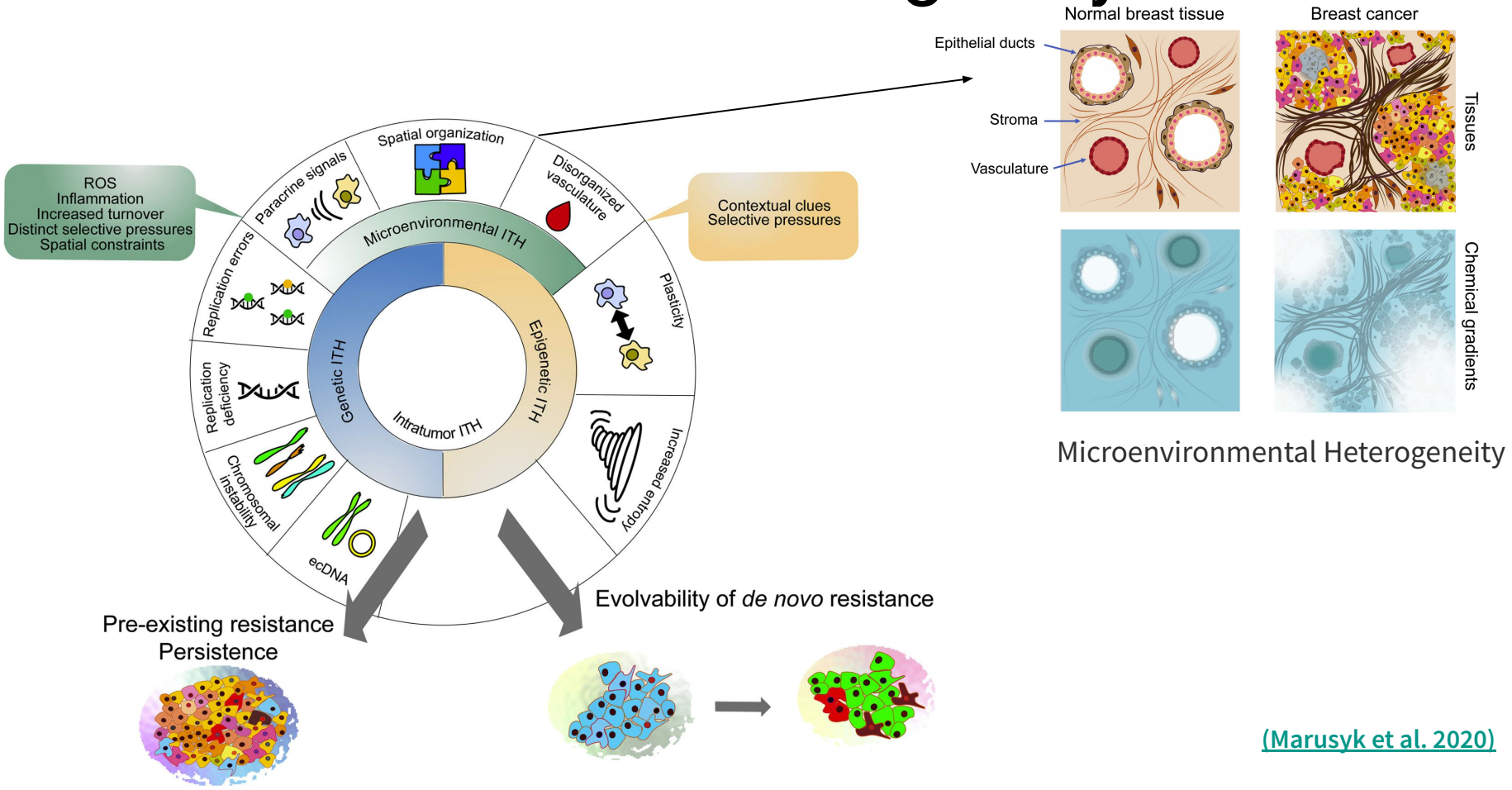


a | Spatial heterogeneity denotes an uneven distribution of cancer subclones across different regions of the primary tumour and/or metastatic sites

b | Temporal heterogeneity refers to variations in the molecular makeup of a single lesion over time

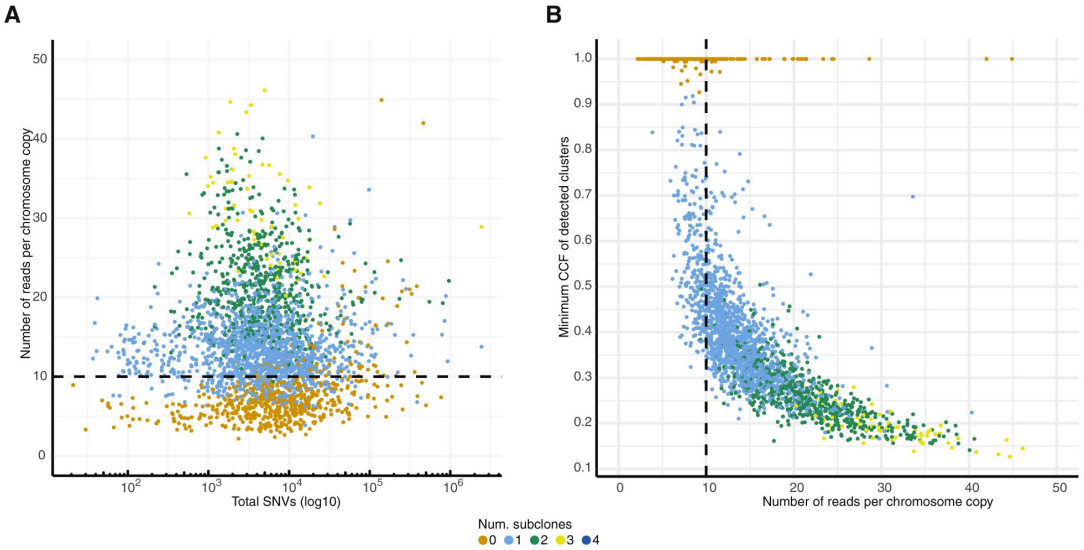
(Dagogo-Jack and Shaw 2018)

Source of intratumor heterogeneity



(Marusyk et al. 2020)

Power analysis for study intra-tumor heterogeneity

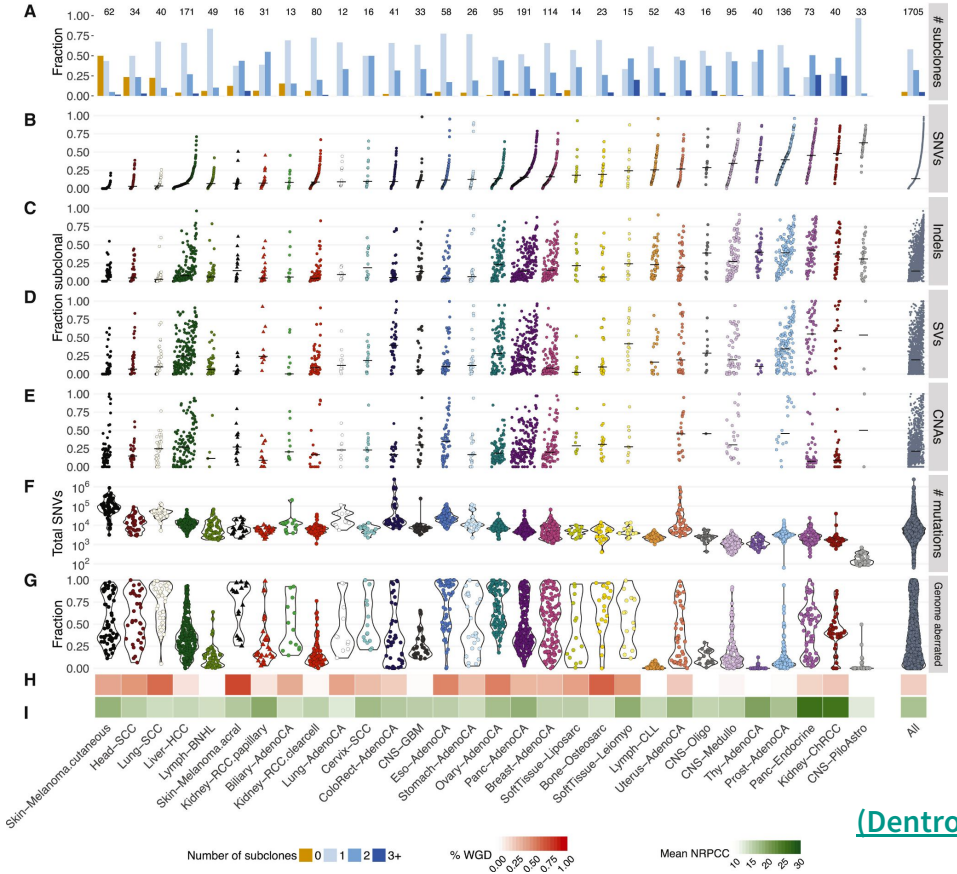


To account for varying purity, ploidy and sequencing depth in the analyzed samples, we calculated the **number of reads per tumor chromosomal copy (nrpcc)** to uniformly quantify the power to detect subclonal mutation clusters.

$$\text{nrpcc} = \frac{\rho \cdot \text{cov}}{\rho\psi_t + (1-\rho)\psi_n}$$

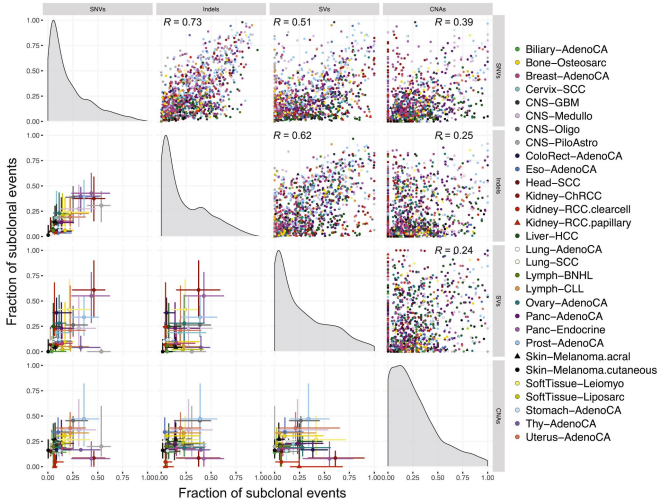
where ρ is the determined purity of the sample and ψt and ψn denote the ploidy of the matching tumor and normal sample, respectively. As we assume all germline samples to be diploid, ψn is set to two by default. We verified that the nrpcc is a strong factor influencing the number of identified subclones in a sample, whereas the total number of mutations identified does not impact the reconstruction

Overview and characterization of ITH across cancer types



Pervasive ITH across cancer types
Evidence of ITH is shown for 1,705 samples with sufficient power to detect subclones at a CCF of more than 30%

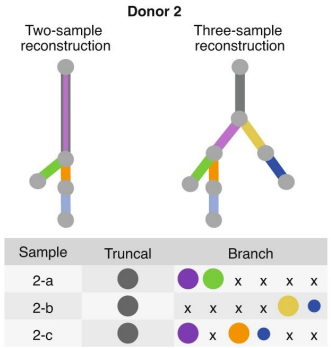
Correlation in ITH between SNVs, indels, CNAs, and SVs

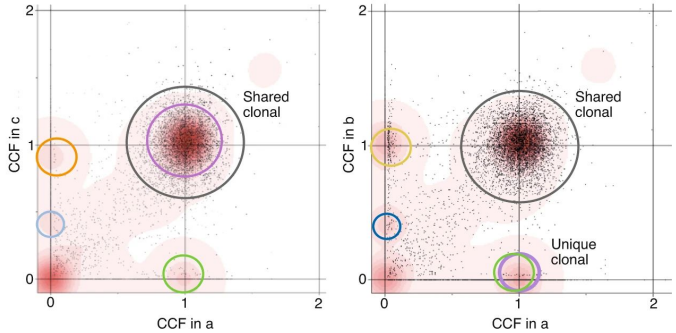


(Dentro et al. 2021)

Illusion of clonality

A mutation that is clonal in the sequenced tumor sample but is not clonal in the whole tumor



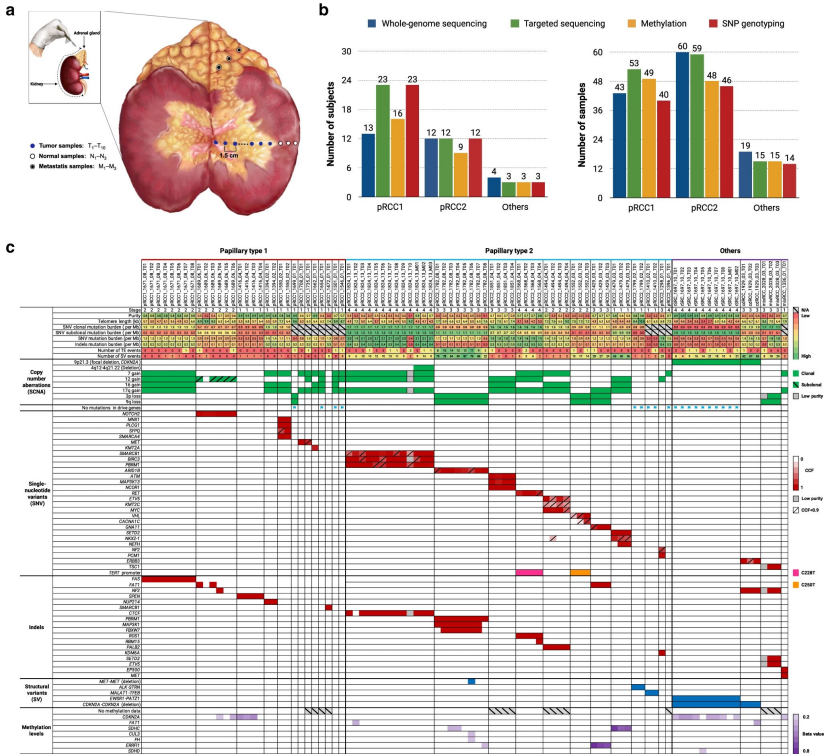


[(Tarabichi et al. 2021)]

Cancer cell fraction (CCF)

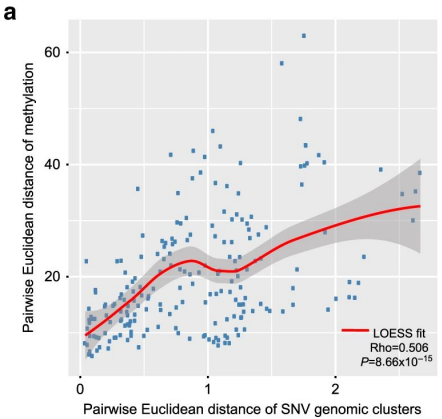
In tumor level, clone mutations are mutations with CCF =1 among all sequenced regions

Multi-regions DNA sequencing for studying ITH



a single biopsy would be sufficient to identify the important genetic drivers

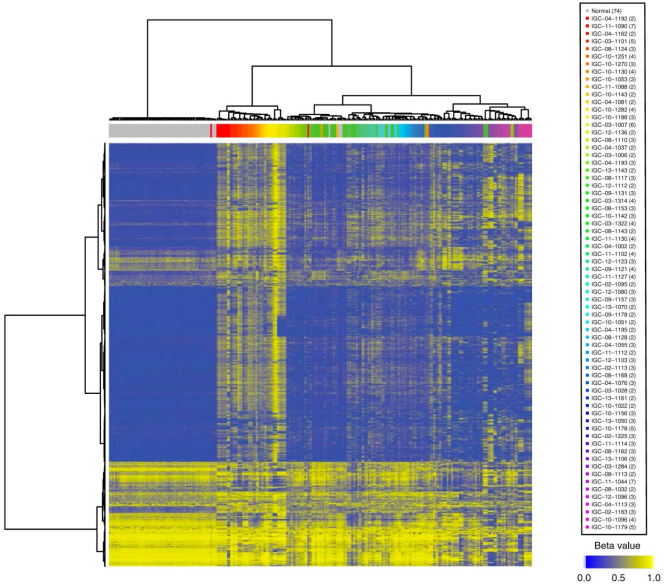
[(Zhu et al. 2020)]

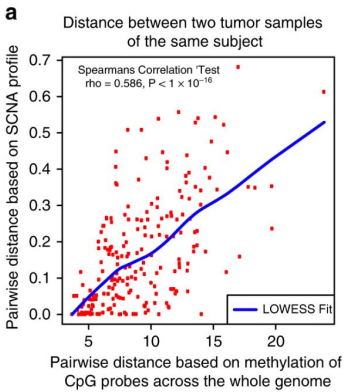


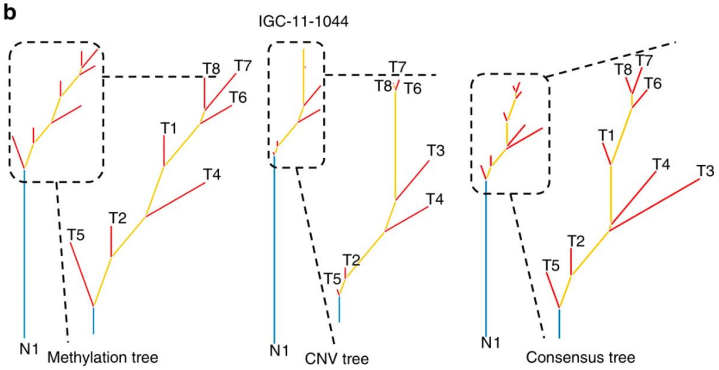
congruent patterns of genomic and epigenomic evolution

ITH by methylation

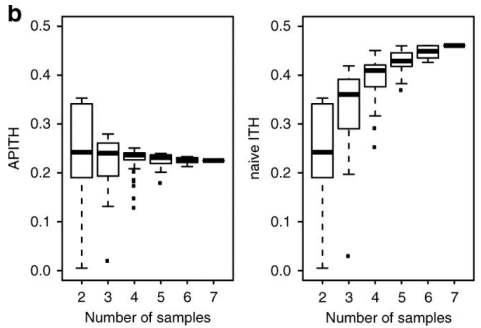
Intratumoral heterogeneity
of DNA methylation profiles







congruent patterns of genomic and epigenomic evolution



Developed an average pairwise ITH index (APITH), which does not depend on the number of samples per tumor.

(Hua et al. 2020)

Cancer subclonal reconstruction from DNA sequencing

Tumor purity and ploidy

Purity (a.k.a. cellularity, or aberrant cell fraction)

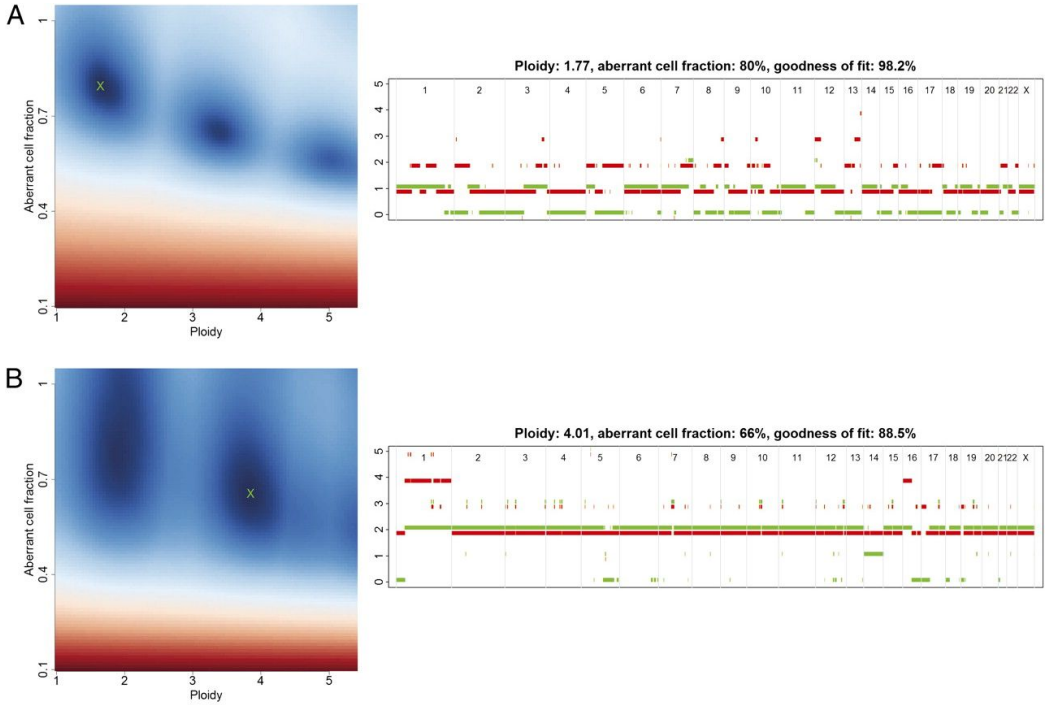
The proportion of cells in a sample that are tumor cells

Ploidy

The average total copy number across the genome

Tumor purity and ploidy

- Simulate a range of purity and overall ploidy values
- Calculate the major/minor copy number of each locus
- Determine which purity/ploidy values give the most optimal solution, using a metric such as:
  - sum of Euclidean distances to integer (i.e. clonal) values (*ASCAT*)
  - proportion of aberrant genome that is clonal (*Battenberg*)



Allele-specific copy number analysis of tumors, Van Loo et al., PNAS 2010
https://doi.org/10.1073/pnas.1009843107

Tumor purity and ploidy

Calculating the major/minor copy number of each locus:

- i = genomic locus
- r = Log R (log-transformed total read depth)
- b = BAF (B-allele frequency, i.e. relative presence of two alternative nucleotides)
- ρ = purity
- ψ = ploidy
- γ = constant: drop in Log R in case of a deletion in a 100% pure sample
- n_A = major copy number
- n_B = minor copy number

$$r_i = \gamma \log_2 \left(\frac{2(1-\rho) + \rho(n_{A,i} + n_{B,i})}{\psi} \right) \quad [1]$$

$$b_i = \frac{1 - \rho + \rho n_{B,i}}{2 - 2\rho + \rho(n_{A,i} + n_{B,i})} \quad [2]$$

Allele-specific copy number analysis of tumors, Van Loo et al., PNAS 2010
https://doi.org/10.1073/pnas.1009843107

Variant allele fraction or frequency (VAF)

The fraction of mutated reads for a given variant, which is a readout of the proportion of DNA mutated in the sequenced tissue.

Cellular prevalence (CP)

The fraction of all cells (both tumor and admixed normal cells) from the sequenced tissue carrying a set of SNVs.

Cancer cell fraction (CCF)

The fraction of cancer cells from the sequenced sample carrying a set of SNVs, that is, CCF = CP/purity. It can be inferred from the VAF (f) given a sample purity (ρ), the local copy number (NT) and the inferred multiplicity m of the mutations:

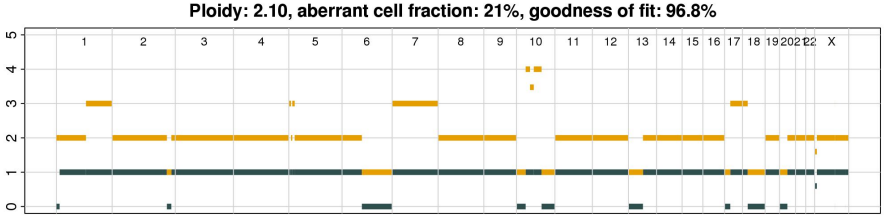
$$\mathrm{CCF} = \frac{f}{m\rho}(\rho N_T + 2(1-\rho))$$

Examples:

	chr	pos	WT count	mut count	total copy number	multiplicity
a	1	63040670	73	11	2	1
b	1	155951020	96	22	3	?

Purity: 0.21189

VAF:



CCF:

Variant allele fraction or frequency (VAF)

The fraction of mutated reads for a given variant, which is a readout of the proportion of DNA mutated in the sequenced tissue.

Cellular prevalence (CP)

The fraction of all cells (both tumor and admixed normal cells) from the sequenced tissue carrying a set of SNVs.

Cancer cell fraction (CCF)

The fraction of cancer cells from the sequenced sample carrying a set of SNVs, that is, CCF = CP/purity. It can be inferred from the VAF (f) given a sample purity (ρ), the local copy number (NT) and the inferred multiplicity m of the mutations:

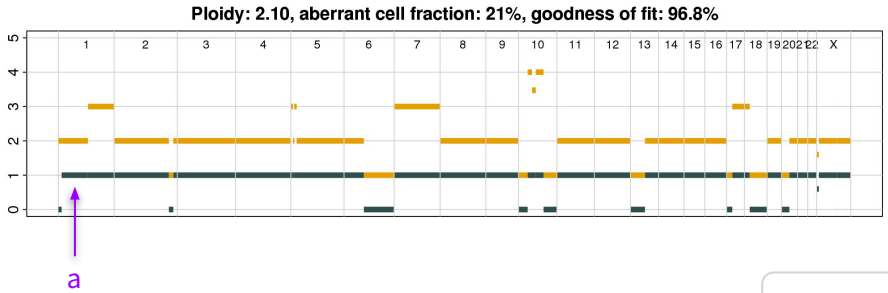
$$\mathrm{CCF} = \frac{f}{m\rho}(\rho N_T + 2(1-\rho))$$

Examples:

	chr	pos	WT count	mut count	total copy number	multiplicity
a	1	63040670	73	11	2	1
b	1	155951020	96	22	3	?

Purity: 0.21189

VAF: $\frac{11}{73 + 11} = 0.130952$



CCF:

Color coding: WT Count | Mut Count | Purity | VAF | Total Copy Number | Multiplicity

Variant allele fraction or frequency (VAF)

The fraction of mutated reads for a given variant, which is a readout of the proportion of DNA mutated in the sequenced tissue.

Cellular prevalence (CP)

The fraction of all cells (both tumor and admixed normal cells) from the sequenced tissue carrying a set of SNVs.

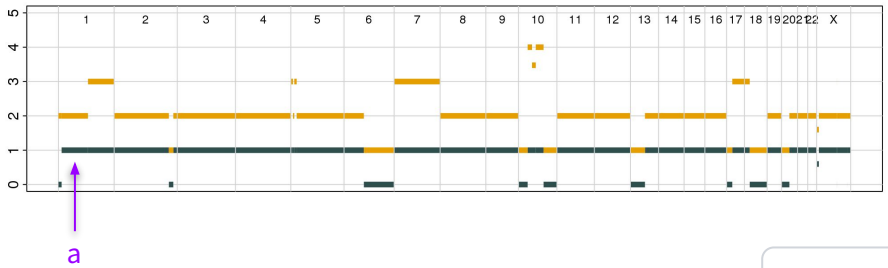
Cancer cell fraction (CCF)

The fraction of cancer cells from the sequenced sample carrying a set of SNVs, that is, CCF = CP/purity. It can be inferred from the VAF (f) given a sample purity (ρ), the local copy number (NT) and the inferred multiplicity m of the mutations:

$$\mathrm{CCF} = \frac{f}{m\rho}(\rho N_T + 2(1-\rho))$$

Examples:

	chr	pos	WT count	mut count	total copy number	multiplicity
a	1	63040670	73	11	2	1
b	1	155951020	96	22	3	?



Purity: 0.21189

VAF: $\frac{11}{73 + 11} = 0.130952$

CCF: $\frac{0.130952}{1 \times 0.21189}\left(0.21189 \times 2 + 2(1 - 0.21189) \right) = 1.236041$

Color coding: WT Count | Mut Count | Purity | VAF | Total Copy Number | Multiplicity

Variant allele fraction or frequency (VAF)

The fraction of mutated reads for a given variant, which is a readout of the proportion of DNA mutated in the sequenced tissue.

Cellular prevalence (CP)

The fraction of all cells (both tumor and admixed normal cells) from the sequenced tissue carrying a set of SNVs.

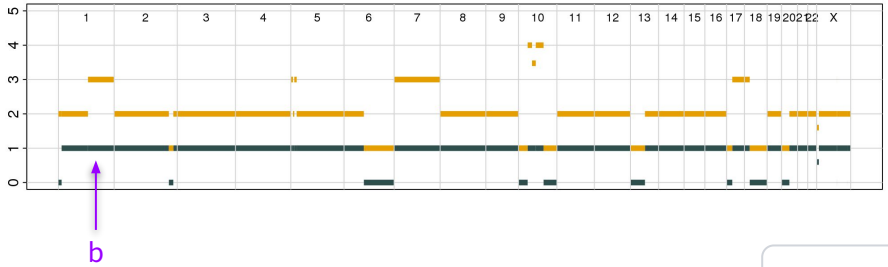
Cancer cell fraction (CCF)

The fraction of cancer cells from the sequenced sample carrying a set of SNVs, that is, CCF = CP/purity. It can be inferred from the VAF (f) given a sample purity (ρ), the local copy number (NT) and the inferred multiplicity m of the mutations:

$$\mathrm{CCF} = \frac{f}{m\rho}(\rho N_T + 2(1-\rho))$$

Examples:

	chr	pos	WT count	mut count	total copy number	multiplicity
a	1	63040670	73	11	2	1
b	1	155951020	96	22	3	?



Purity: 0.21189

VAF: $\frac{22}{96 + 22} = 0.186441$

CCF:

Color coding: WT Count | Mut Count | Purity | VAF | Total Copy Number | Multiplicity

Variant allele fraction or frequency (VAF)

The fraction of mutated reads for a given variant, which is a readout of the proportion of DNA mutated in the sequenced tissue.

Cellular prevalence (CP)

The fraction of all cells (both tumor and admixed normal cells) from the sequenced tissue carrying a set of SNVs.

Cancer cell fraction (CCF)

The fraction of cancer cells from the sequenced sample carrying a set of SNVs, that is, CCF = CP/purity. It can be inferred from the VAF (f) given a sample purity (ρ), the local copy number (NT) and the inferred multiplicity m of the mutations:

$$\mathrm{CCF} = \frac{f}{m\rho}(\rho N_T + 2(1-\rho))$$

Examples:

	chr	pos	WT count	mut count	total copy number	multiplicity
a	1	63040670	73	11	2	1
b	1	155951020	96	22	3	2

Purity: 0.21189

VAF: $\frac{22}{96 + 22} = 0.186441$

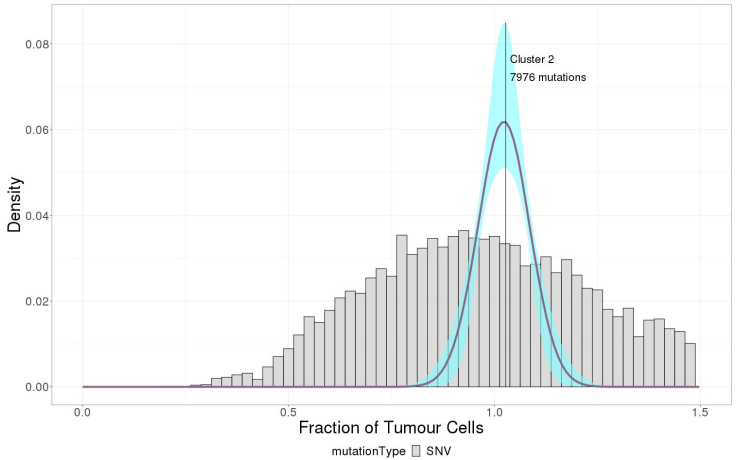
CCF: $\frac{0.186441}{2 \times 0.21189} (0.21189 \times 3 + 2(1 - 0.21189)) = 0.973114$

Color coding: WT Count | Mut Count | Purity | VAF | Total Copy Number | Multiplicity

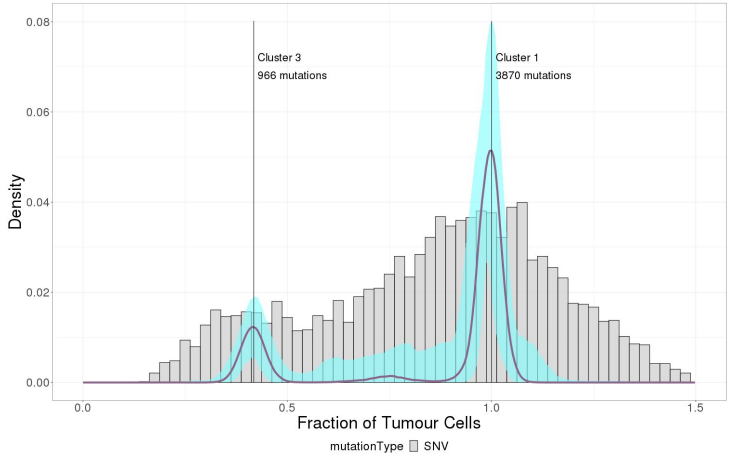
Clustering SNVs by CCF

Examples from DPClust

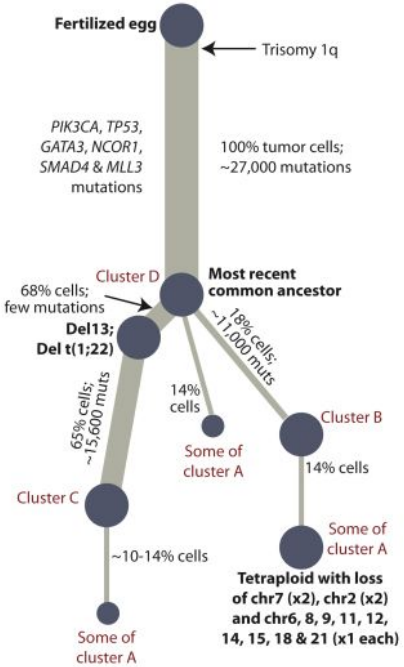
Sample with one clonal cluster identified:



Sample with two clusters identified - one clonal and one subclonal:



Principles for creating a phylogenetic tree

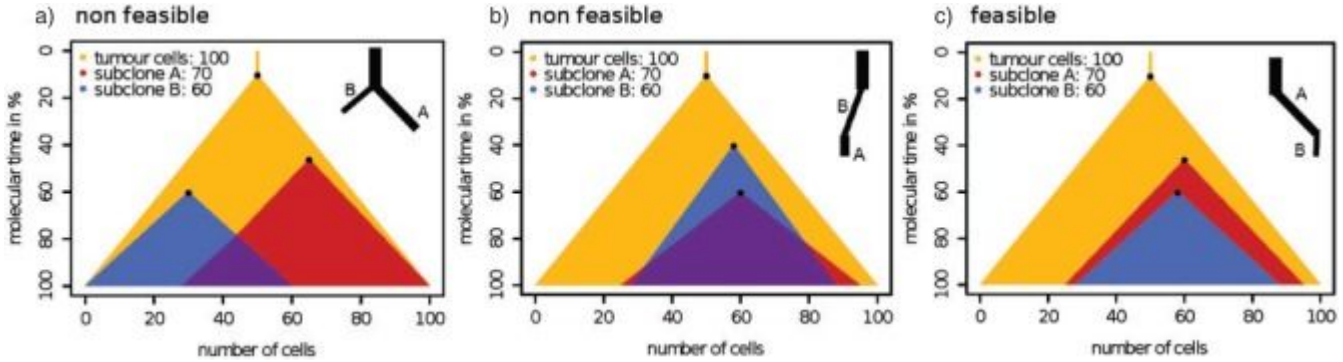


Most recent common ancestor (MRCA). The MRCA is the most recent cell that spawned a set of cells. By extension, the MRCA also refers to the genotype of that ancestor cell. The MRCA of a given tumor is sometimes used to implicitly refer to the MRCA of all cells in a set of sequenced samples. Note that the MRCA of a tumor sample (or set of samples) is not necessarily the MRCA of the whole tumor, owing to the illusion of clonality.

The Life History of 21 Breast Cancers, Nik-Zainal et al., 2012, Cell, doi: 10.1016/j.cell.2012.04.023

Principles for creating a phylogenetic tree

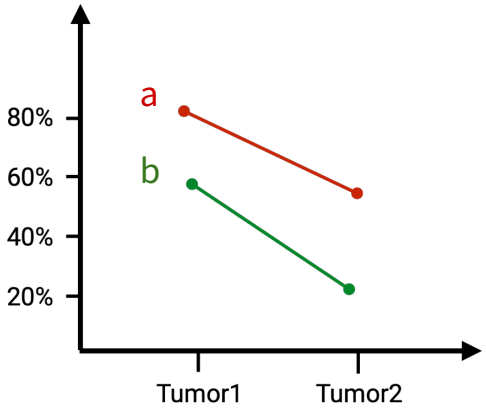
Pigeonhole principle. In the context of subclonal reconstruction, the sum of CCFs of branching subclones should be less than or equal to the CCF of their parent clone. Indeed, if it was greater, this would mean that mutations have occurred independently in branching lineages. However, according to the infinite sites hypothesis, the same set of random mutations is unlikely to have happened twice independently. Therefore, the smaller subclone must be a descendant of the bigger subclone; that is, they are linear subclones, which is compatible with the infinite sites hypothesis.



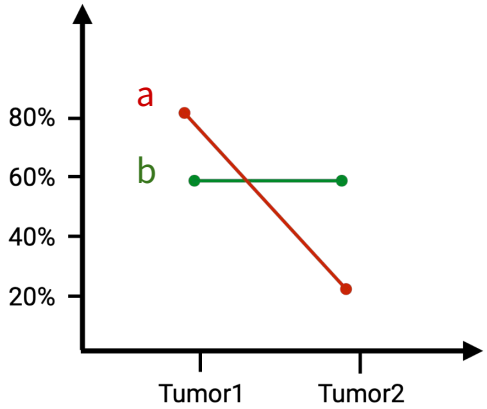
Cancer Evolution: Mathematical Models and Computational Inference, Beerenwinkel et al., 2015, Systems Biology, doi: 10.1093/sysbio/syu081

Principles for creating a phylogenetic tree

Crossing rule. When performing multisample or multiregion sequencing, when clone A and B are descendant of clone C and the CCF of clone A is higher than the CCF of clone B in one sample but the opposite is true in another sample, then clone A and B must be branching subclones. This rule stems from the more general rule that the shared subclones across samples must have arisen from the same phylogeny, which further constrains the possible phylogenetic relationships between subclones.



Yes



No

a descendent clone must exhibit a smaller cellular prevalence than its ancestor within each and every tumor region

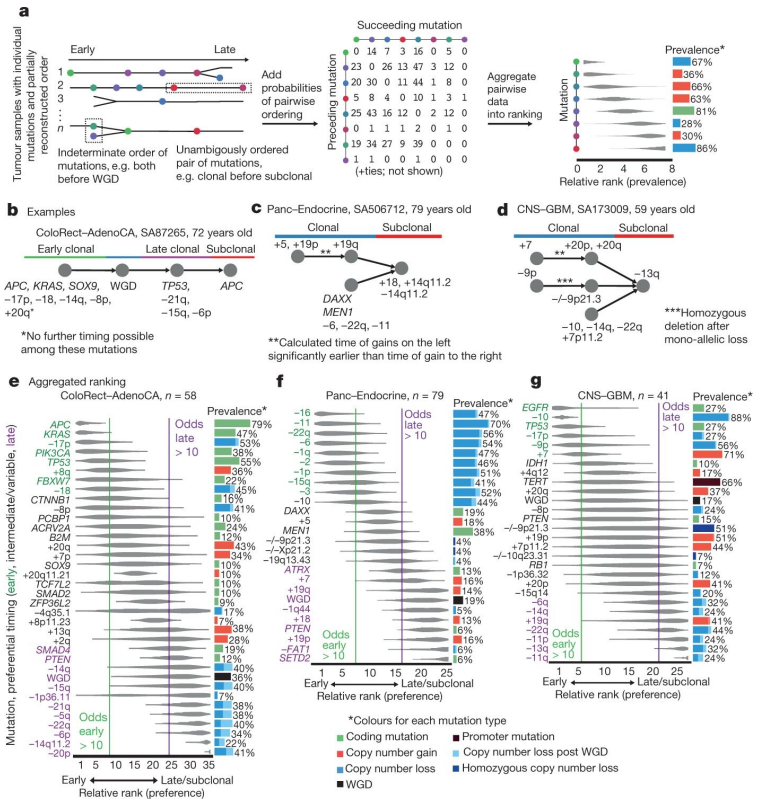
Standard workflow for subclonal reconstruction

- Sequence
  - …tumours at high depth, multiple regions if possible
  - …matched normal tissue
- Call somatic variants
- Reconstruct allele-specific copy number profiles
  - Consider using multiple copy number callers, to handle ambiguity
  - Check solutions (e.g. for correct CP and WGD status), refit if necessary
- Reconstruct subclonal SNV clustering
- Where possible, reconstruct phylogenies
  - Multi-region sampling helps
  - Phasing information and single-cell sequencing can provide further evidence

Recommended reading for further details: "A practical guide to cancer subclonal reconstruction from DNA sequencing", Tarabichi et al., nature methods, 2021

Timing somatic events in the evolution of cancer

Timing of somatic driver events



Events in a sample ordered based on:

- CCF (clonal before subclonal)
- Number of copies
  - Mutations, e.g.:
    - On 2 copies in a 2+2 region: before WGD
    - On 1 copy in a 2+2 region: after WGD
  - SCNAs, e.g.:
    - 2+0 region in WGD sample: loss occurred before WGD
    - 2+1 region in a WGD sample: loss occurred after WGD

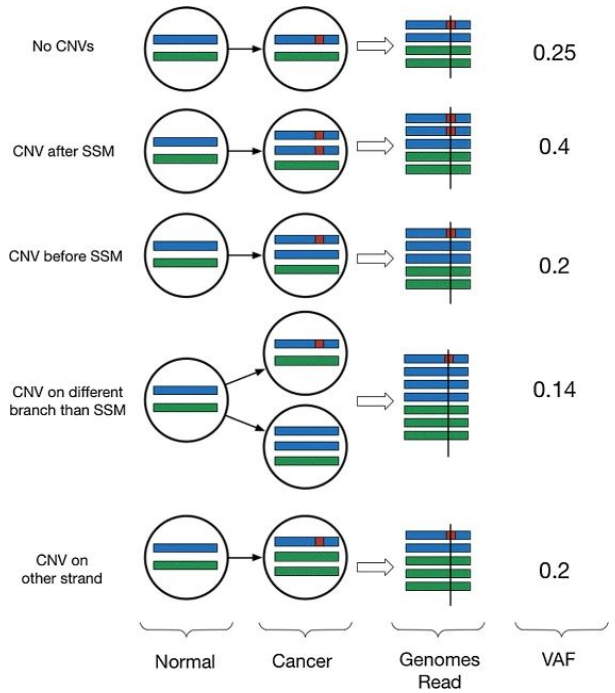
Orderings aggregated across samples…

Some options:

- PhylogicNDT league model
- Bradley-Terry model
- Plackett-Luce model

The evolutionary history of 2,658 cancers, Gerstung et al., 2020, Nature, doi: [10.1038/s41586-019-1907-7](https://doi.org/10.1038/s41586-019-1907-7)

Timing of copy number gains

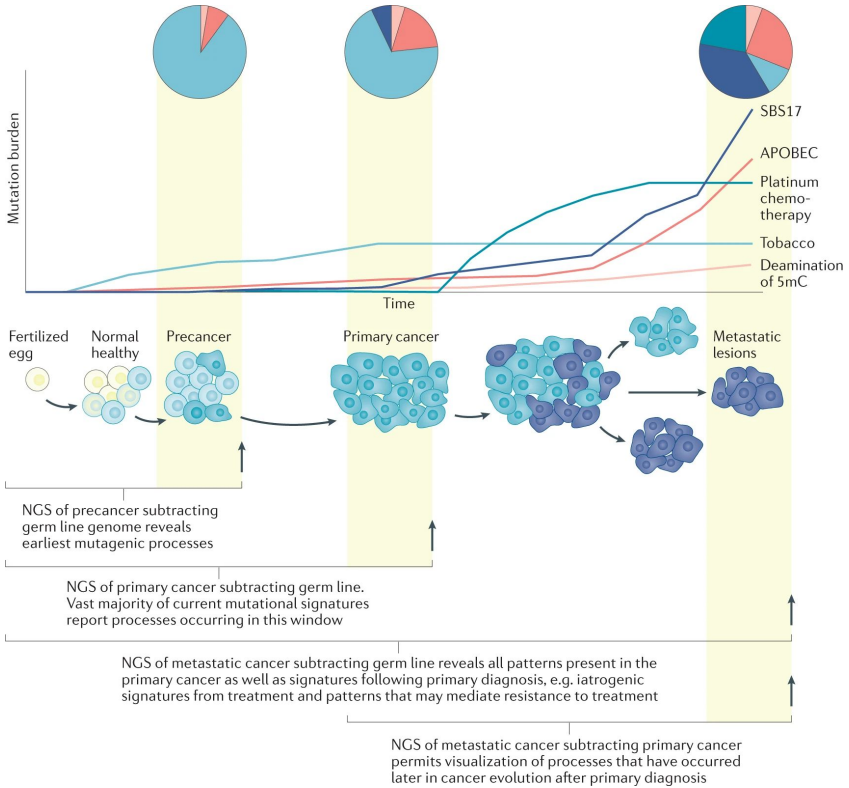


We can use mutation VAFs to infer:

- Whether a SCNA (aka CNV) has occurred
- Relative order of the SCNA and mutation
- Whether the SCNA occurs in a different subclone (i.e. set of cells) than the mutation
- Strand of SCNA relative to mutation

PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors, Deshwar et al., Genome Biology, 2015, doi: 10.1186/s13059-015-0602-8

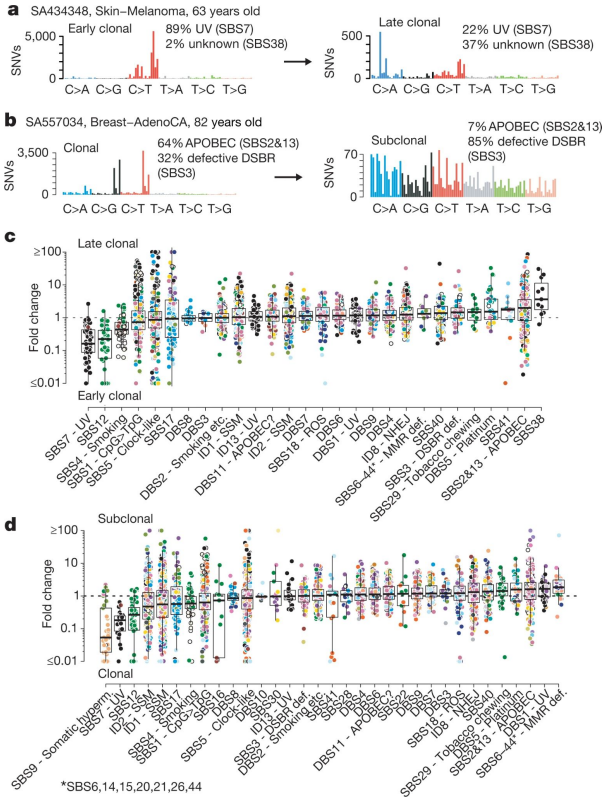
Timing of mutational signatures



Sequencing at multiple time points can be highly informative

Koh Gene, et al., Nature Reviews Cancer, 2021

Timing of mutational signatures

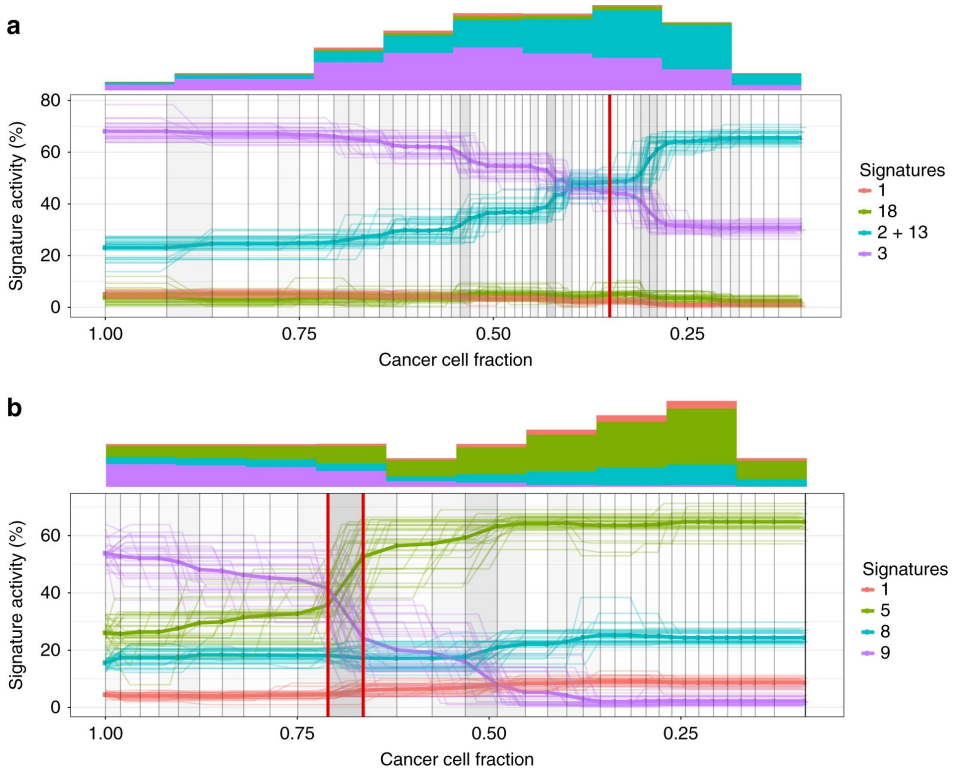


From single samples/time points, we can compare signatures by timing category:

- Early (pre-WGD) clonal
- Late (post-WGD) clonal
- Clonal (All clonal mutations; note that early/late cannot always be specified e.g. if no WGD)
- Subclonal

The evolutionary history of 2,658 cancers, Gerstung et al., 2020, Nature, doi: [10.1038/s41586-019-1907-7](https://doi.org/10.1038/s41586-019-1907-7)

Timing of mutational signatures

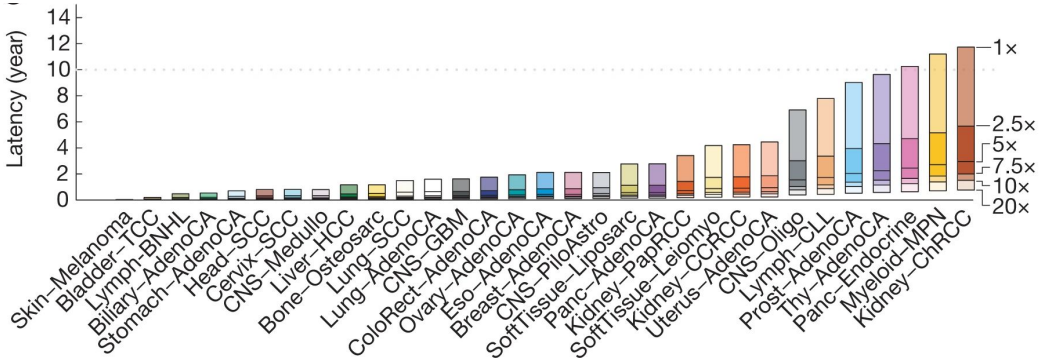


We can track signature activity (approximately) by mutation CCFs

Reconstructing evolutionary trajectories of mutation signature activities in cancer using TrackSig, Rubanova et al., Nature Communications, 2020, doi: [10.1038/s41467-020-14352-7](https://doi.org/10.1038/s41467-020-14352-7)

Chronological time estimates

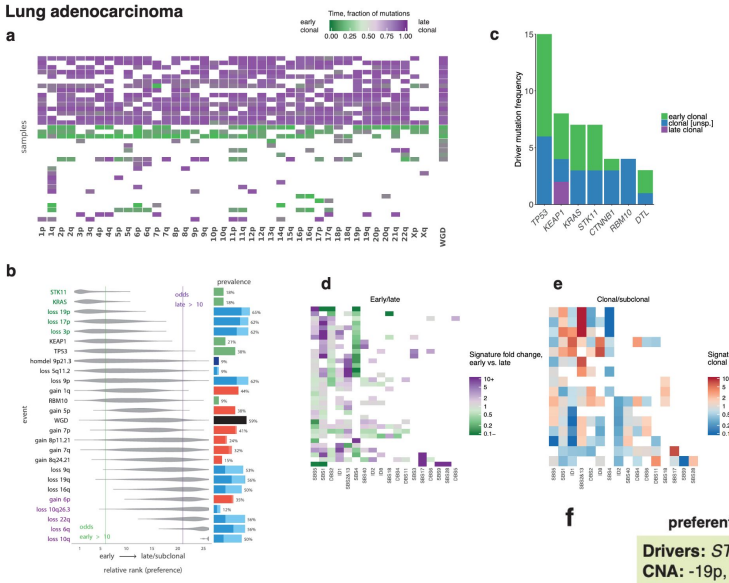
- Previous “timings” are relative, not chronological
- Mutation rates change over time in cancers
- CpG>TpG are relatively stable, clock-like mutations
- However, they have also been shown to increase modestly in cancer cells vs normal
- We can count these mutations to estimate chronological timing
- We should account for a range of possible increases in CpG>TpG mutation rate



Median latency between the MRCA and the last detectable subclone before diagnosis for different CpG>TpG mutation rate changes in *n* = 1,921 non-hypermutant samples with low tumour in normal contamination and at least 5 cases per cancer type.

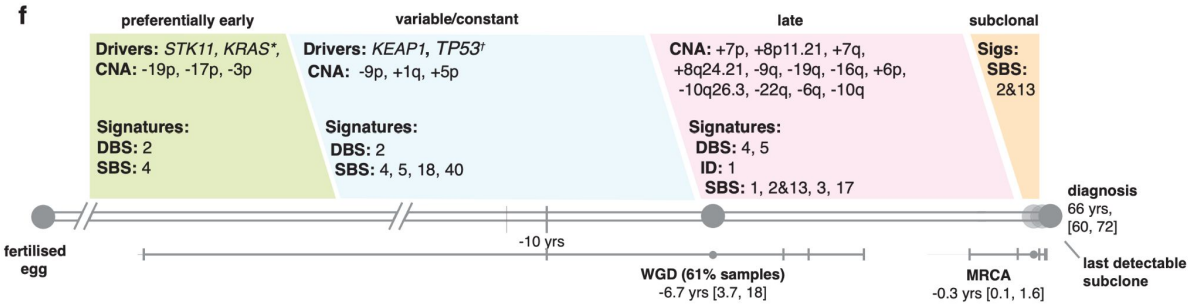
The evolutionary history of 2,658 cancers, Gerstung et al., 2020, Nature, doi: 10.1038/s41586-019-1907-7

Reconstructing the life history of tumors



The evolutionary history of 2,658 cancers

Gerstung et al., 2020, Nature, doi: [10.1038/s41586-019-1907-7](https://doi.org/10.1038/s41586-019-1907-7)



Recommended reading for an in-depth study using many of these evolutionary timing methods:
"The evolutionary history of 2,658 cancers", Gerstung et al., 2020, Nature, https://www.nature.com/articles/s41586-019-1907-7

A review of cancer genomic evolution:
"Evolution of the cancer genome", Yates & Campbell, 2012, Nature Reviews Genetics, https://www.nature.com/articles/nrg3317

Evolutionary dynamics extrachromosomal DNA (ecDNA) in human cancers

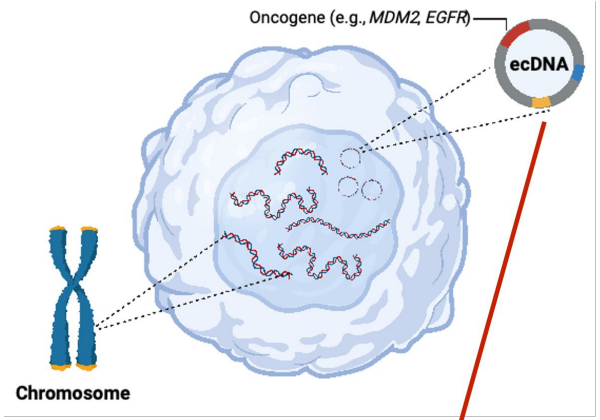
Extrachromosomal DNA (ecDNA)

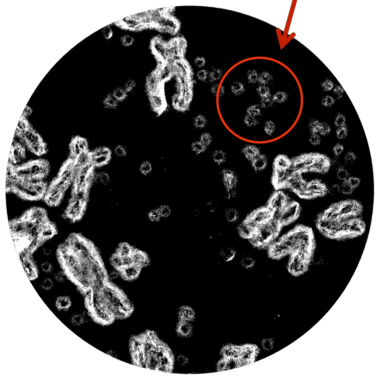
ecDNA are large units of **circular DNA** that reside within the nuclei of cells yet are physically distinct from chromosomal DNA.

They often range in size from **1-5 mega base pairs** in length and can encode **one or more full-length genes** and regulatory regions.

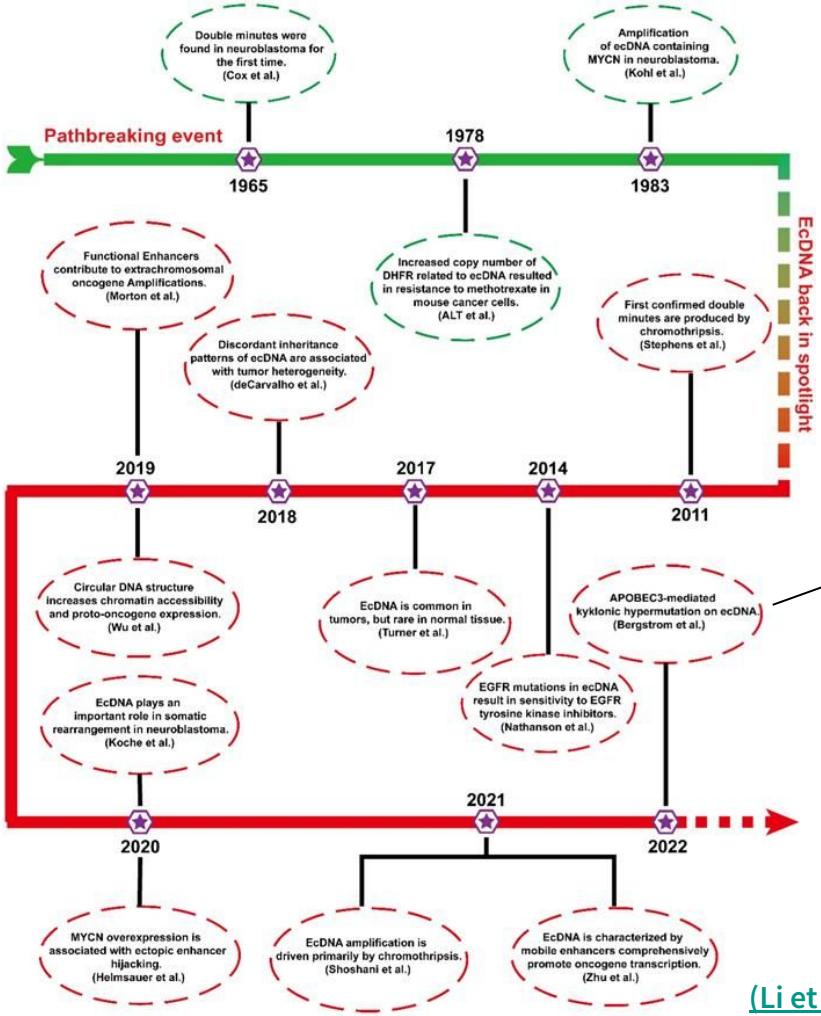
ecDNA have accessible chromatin and are highly transcribed, meaning they are fully functional and often more active than chromosomally located genes.

ecDNA are one of the primary locations for **high copy number focal oncogene amplifications** in cancer cells; in fact, more than half of all high copy number amplifications in cancer occur on ecDNA.





ecDNA observed by scanning electron microscope image



Timeline of landmark ecDNA explorations

- More recurrent APOBEC3 kataegis was observed across circular ecDNA regions compared to other forms of structural variation

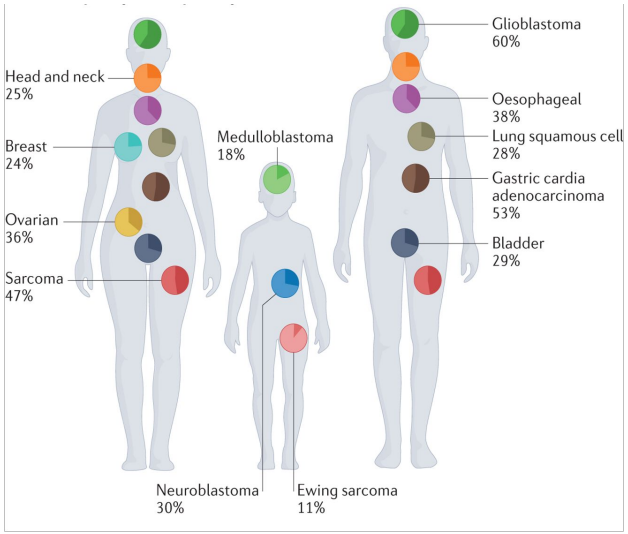
- Recurrent kyklonic events were increased within or near known cancer-associated genes including *TP53*, *CDK4* and *MDM2*, etc.

(Li et al. 2022)

ecDNA is a cancer specific phenomenon

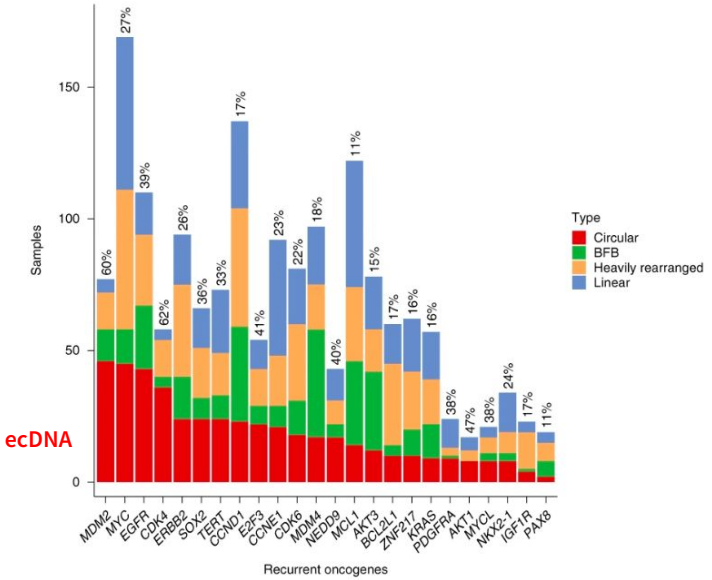
Absent in normal healthy tissue, ecDNA are found in 14% of primary cancers and >40% of metastatic cancers.

ecDNA frequency across primary cancers



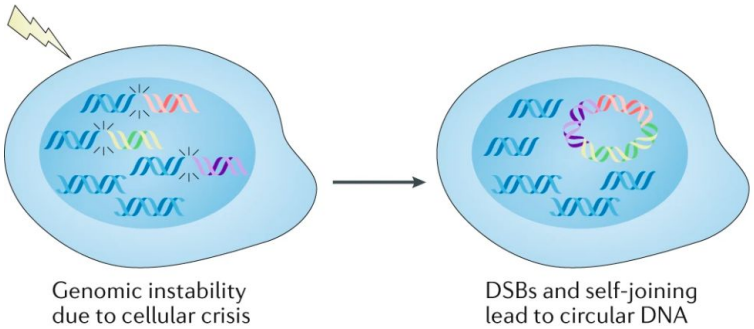
Yi et al. 2022

Oncogene amplification driven by ecDNA



Kim et al. 2020

The origin of ecDNA: Chromosome instability

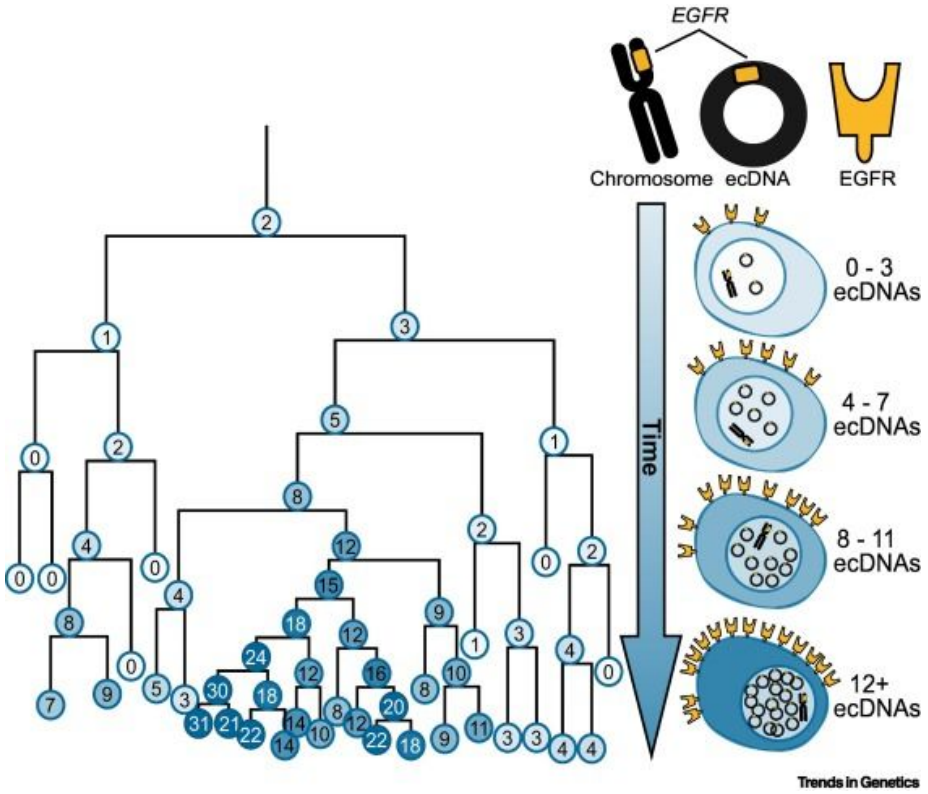


(Verhaak et al. 2019)

- Chromothripsis
- Breakage-fusion-bridge (BFB) cycles
- Slight damage to DNA and relegation
- Replication fork stalling and template switching

(Li et al. 2022)

Model of the rapid accumulation of ecDNA in cancer



(Noer et al. 2022)

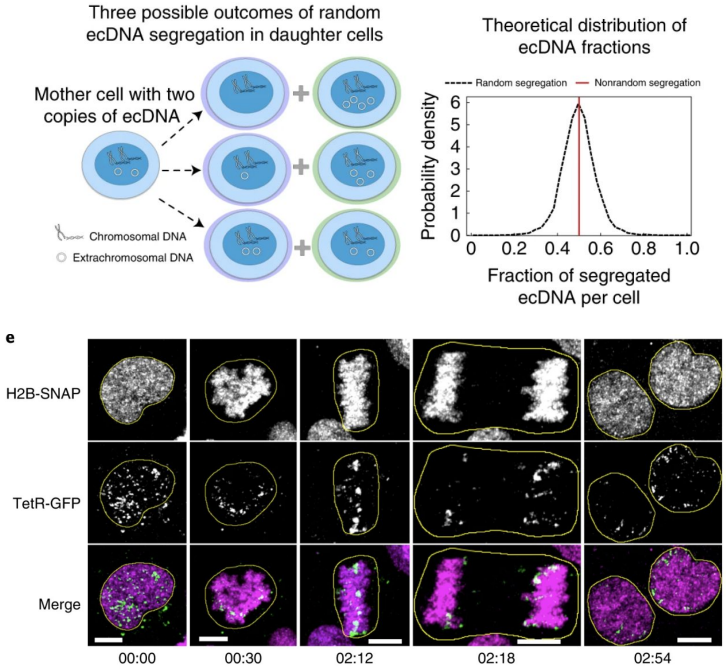
Driving high copy number gene amplifications and non-Mendelian genomic adaptation, ecDNA enable tumors to rapidly evolve and switch their oncogene dependency when under therapeutic pressure, thereby rendering current targeted and immunotherapy approaches largely ineffective in patients with gene amplified cancers.

The evolutionary dynamics of extrachromosomal DNA in human cancers

- Integrating theoretical models of random segregation, unbiased image analysis, CRISPR-based ecDNA tagging with live-cell imaging and CRISPR-C, we demonstrate that random ecDNA inheritance results in extensive intratumoral ecDNA copy number heterogeneity and rapid adaptation to metabolic stress and targeted treatment.

- These results show how the nonchromosomal random inheritance pattern of ecDNA contributes to poor outcomes for patients with cancer.

(Lange et al. 2022)



THANKS FOR YOUR ATTENTION!
Questions?

Next: Practical session 9 (10:45 am)

- **Tumor evolution analysis using NGSpurity and Palimpsest**

Invited speaker

Date: Thursday, April 20, 2023

Time: 10:30 AM – 11:30 AM

Speaker: David Wedge, Ph.D., University of Manchester

Title: Tumour evolution in diverse human populations