Phylogenetic ctDNA analysis depicts early-stage cancer evolution

C. Abbosh et al. (TRACERx consortium)
Nature 2017

Presented by Vaishnavi S.

What is ctDNA?

- Circulating tumor DNA (ctDNA) is found in the bloodstream
- Refers to DNA that comes from cancerous cells and tumors. More generally cell free DNA (cfDNA).
- The dead cells get broken down and their contents, including DNA, are released into the bloodstream.
- ctDNA are small pieces of DNA, usually comprising fewer than 200 nucleotides in length.
- Normal: apoptotic and necrotic remains are cleared by infiltrating phagocytes.
- Tumor: This does not happen efficiently, accumulation of cellular debris and its inevitable release into the circulation

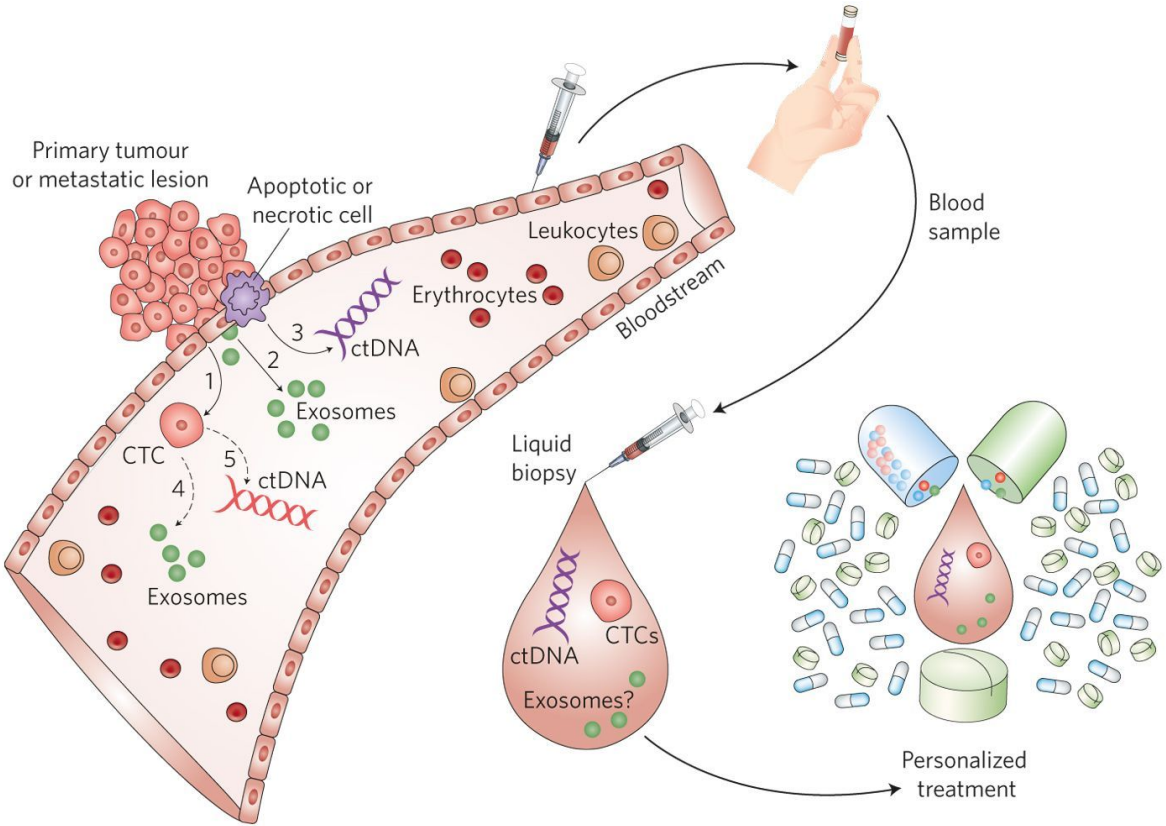


Figure Source: Alix-Panabières, C., & Pantel, K. (2017). Clinical prospects of liquid biopsies. Nat Biomed Eng, 1, 0065.

Liquid biopsy for management

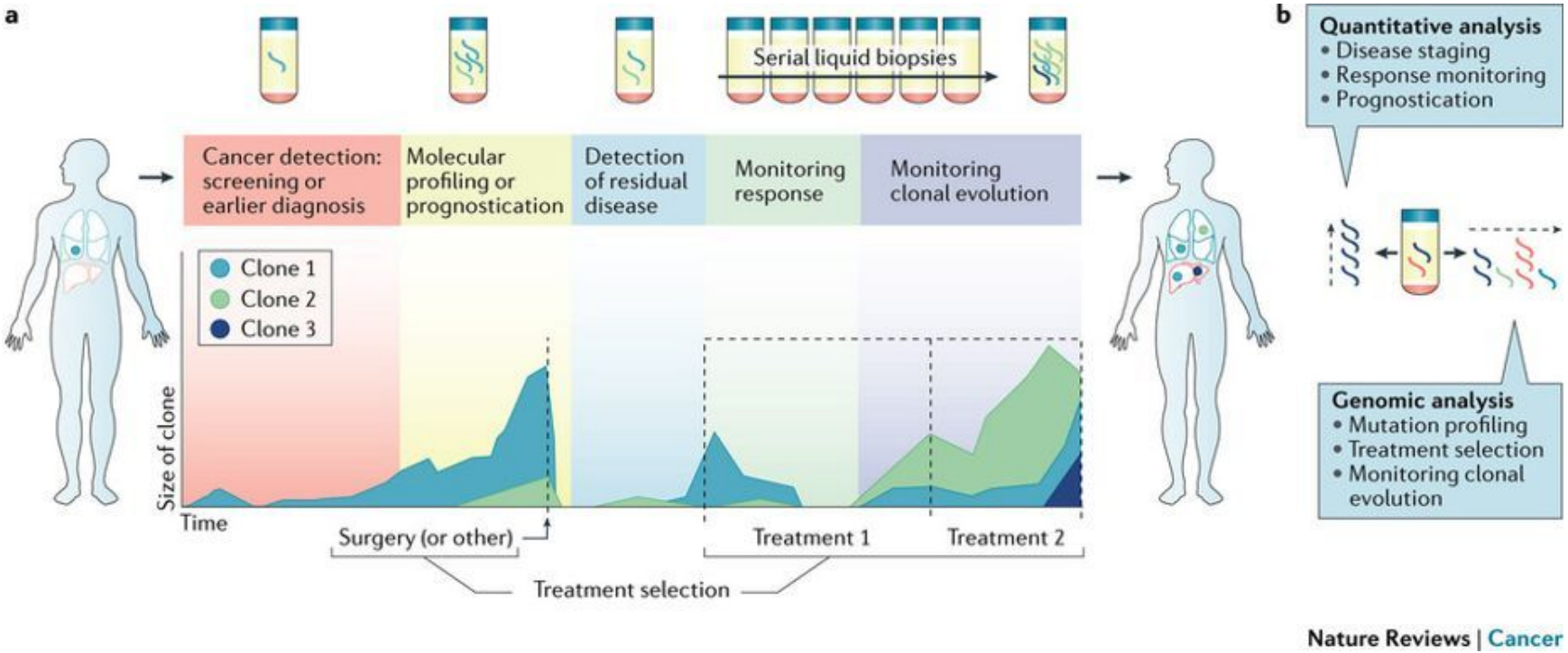


Figure Source: Liquid biopsies come of age: towards implementation of circulating tumour DNA

Targets of this paper:

1. What factors determine ctDNA detection?
2. Can you identify evidence of therapy resistance and relapse using ctDNA?
3. Can you track the sub-clonal nature of relapse for ctDNA-driven therapeutics?

Overview of Phylogenetic ctDNA tracking

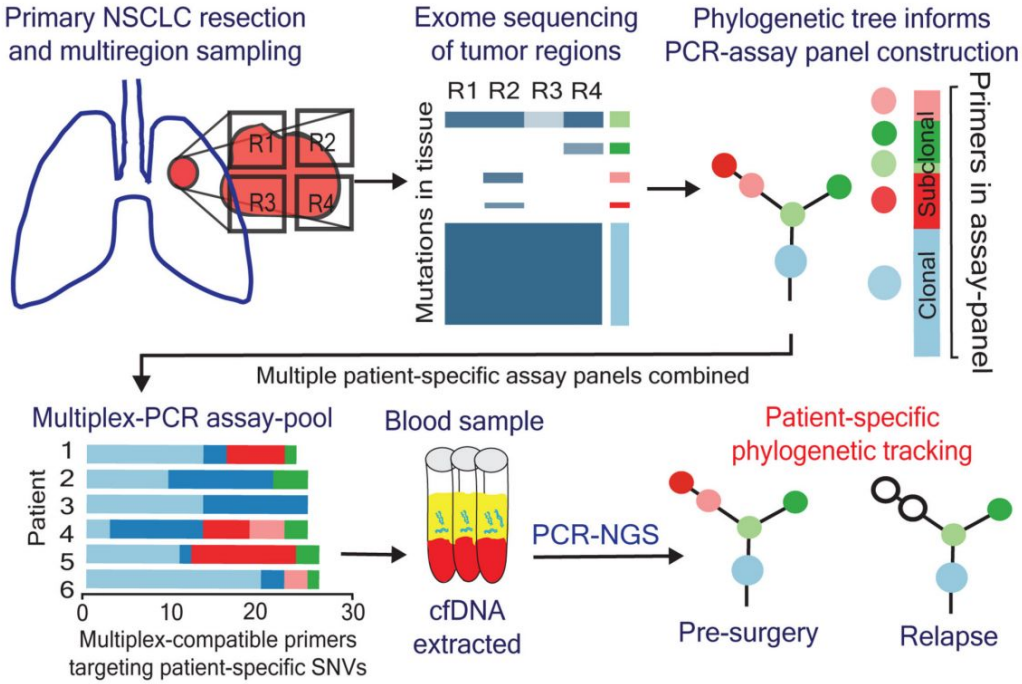


Figure 1. Phylogenetic ctDNA tracking

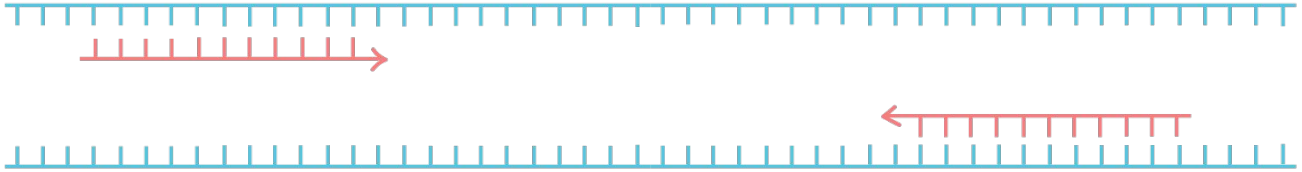
1. Multi-region sequencing
2. Reconstruct phylogenetic tree using CITUP tool
3. **Design PCR-assay panels**
4. **Extract cf-DNA and perform multiplex-PCR**
5. **Integrate with multi region exome-sequence**

1., 2. Part of previous paper

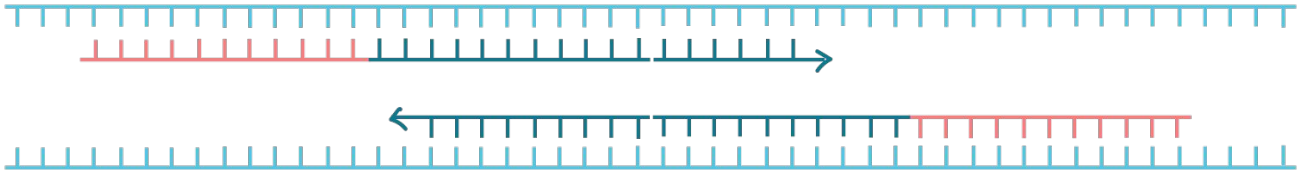
Polymerase Chain Reaction

Template DNA

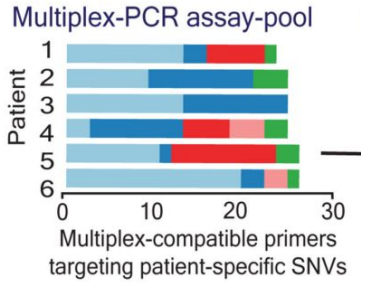
Primers bind to template



Taq polymerase extends primers



Design of PCR assay panel



- Generate forward and reverse PCR primers for somatic SNVs detected in tumour
- For every pair of primers, the probability of forming a primer-dimer was calculated and assays were combined into pools such that any primer combination in a pool is not predicted to form primer-dimers.

- For each patient, assays were prioritized such that
  - Assays covering **driver SNVs** had highest priority
  - There was **uniform sampling of the phylogenetic tree**.

- For the **baseline** cohort 10 balanced pools were created, each containing on average 18 assays for **10 patients**' SNVs.
- For the **longitudinal (14 patients)** cohort, up to 10 extra assays were generated for samples.

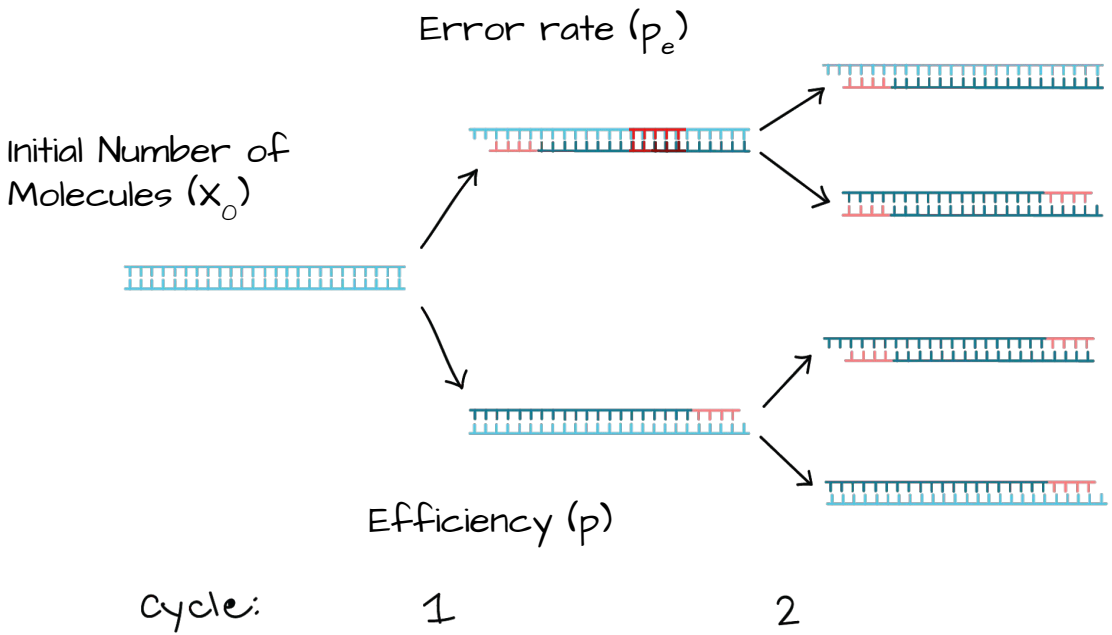
- Assay panel size: median 18 SNVs (range 10-22)
  - Median 11 clonal SNVs (2-20), median 6 sub-clonal SNVs

How does error propagate across cycles?

- The PCR process was modelled as a **stochastic process**
  error parameters estimated using 28–30 control plasma samples

For each target SNV, a target-specific **background-error model** was built by estimating the following from the control samples:

- **PCR efficiency (*p*):** probability that each molecule is replicated in a PCR cycle
- **Error rate (p_e):** error rate per cycle for mutation type *e*
  - For example, wild-type allele A to mutant allele G
- **Initial number of molecules:** X_0



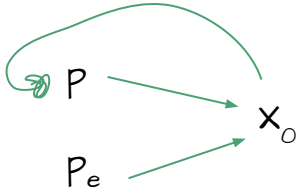
How does error propagate across cycles?

- The PCR process was modelled as a **stochastic process**
  - estimating the error parameters using a set of 28–30 control plasma samples

Error propagation model: As a molecule is replicated over the course of the PCR, more errors occur

- Error in cycle *i* with X_i wild-type molecules in the system
- Error molecule is duplicated in next cycle with probability *p*
- New errors occur according to a binomial process $B(X_i, p_e)$
- Can define **recurrence relations** for mean and variance for total number of molecules X_n and number of error molecules E_n
  - $\mathbf{E}(E_{i+1}) = \mathbf{E}(E_i)(1 + p) + X_i p_e$

Plasma SNV calling: Algorithm



1. Estimate PCR efficiency *(p)* and per-cycle error rate *(p_e)* using normal samples
2. Using the efficiency estimate, compute the starting number of molecules in the test set
3. Use the starting number of molecules and the prior efficiency distribution from the training set to **estimate the PCR efficiency in the test sample**
4. For a range of **potential real mutant fraction** values θ between 0 and 1 (they used 0.15 as upper bound), using the error propagation model and parameters, estimate
   a. the mean and variance for the total number of molecules
   b. the mean and variance for the background error molecules
   c. the mean and variance for the real mutation molecules

5. Use the mean and variance estimated in step (4) to compute the **likelihood L(θ)** for each potential real mutant fraction.
   a. Select the value of θ that maximizes this likelihood ($\hat{\theta}_{\mathrm{MLE}}$)
   b. Compute the confidence score = $\dfrac{L(\hat{\theta}_{\mathrm{MLE}})}{L(0) + L(\hat{\theta}_{\mathrm{MLE}})}$

L(0) : No mutations
Confidence score: higher the better

6. Call a mutation positive if the confidence score passes a predefined threshold
   a. (threshold of 95% for transitions and 98% for transversions)

Phylogenetic analysis of relapse

- Analysis is based on the CCF determined for the SNVs, clustered across tumor regions using a modified version of Pyclone
    - Clusters with similar CCF values
- **Mutation clusters were assumed to represent tumour subclones**, either current or ancestral, and were used as input for construction of the phylogenetic trees, primarily using CITUP

Relapse tree construction was performed as follows. (CRUK0063, as an example)

- Clustering was performed twice, once across five primary tumour regions and once across five primary, one relapse, and six autopsy regions.
- When deriving a phylogenetic tree based on all tumour regions, for CCF clusters based on clustering, only the primary tumour regions were maintained for mutations not involved in metastatic relapse. A phylogenetic tree was constructed based on 17 mutation clusters.

Followed different strategies for different patients, depending on the clusters.

Targets of this paper:

1. **What factors determine ctDNA detection?**
2. Can you identify evidence of therapy resistance and relapse using ctDNA?
3. Can you track the sub-clonal nature of relapse for ctDNA-driven therapeutics?

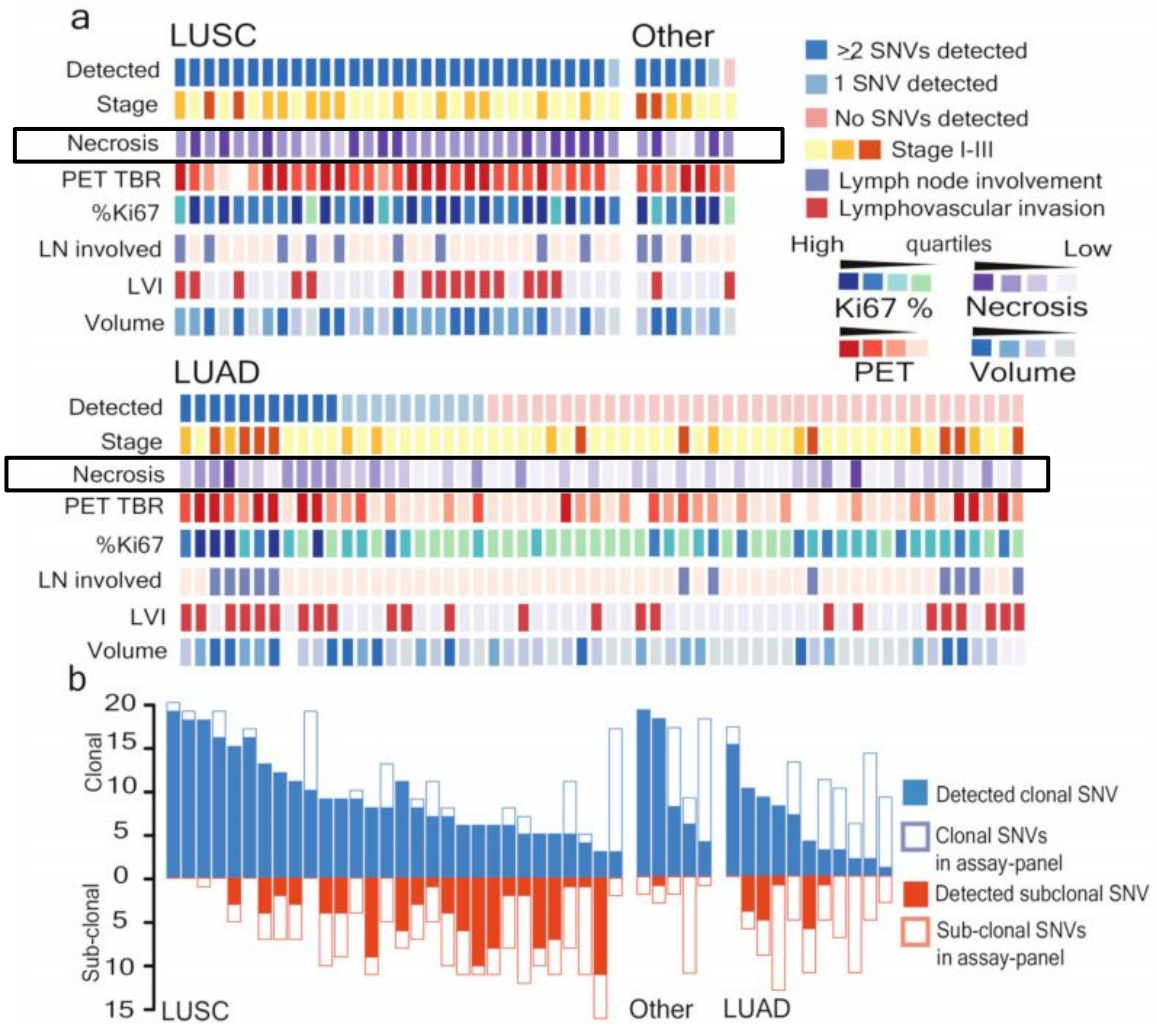


Figure 2. Clinicopathological predictors of ctDNA detection

- If at least 2 SNVs were detected from ctDNA, consider sample ctDNA positive.
- Squamous cell more likely to be detected:
  - Higher necrosis
- Clonal SNVs in **all 46** ctDNA-positive patients
- A **median of 94%** of clonal SNVs targeted by assay panels were detected in the ctDNA of these patients.
- Subclonal SNVs were detected in 27 (68%) of these patients
- A median of 27% of subclonal SNVs within individual assay panels were detected in ctDNA-positive patients

- Tumor volume and clonal VAF are well correlated
- Linear modelling based on the TRACERx data predicted that a primary tumour burden of 10cm^3 would result in a mean clonal plasma VAF of 0.1% (95% confidence interval, 0.06–0.18%)
- Estimated effective subclone size influences subclonal SNV detection.
- Estimated effective subclone size correlates with subclonal plasma VAF

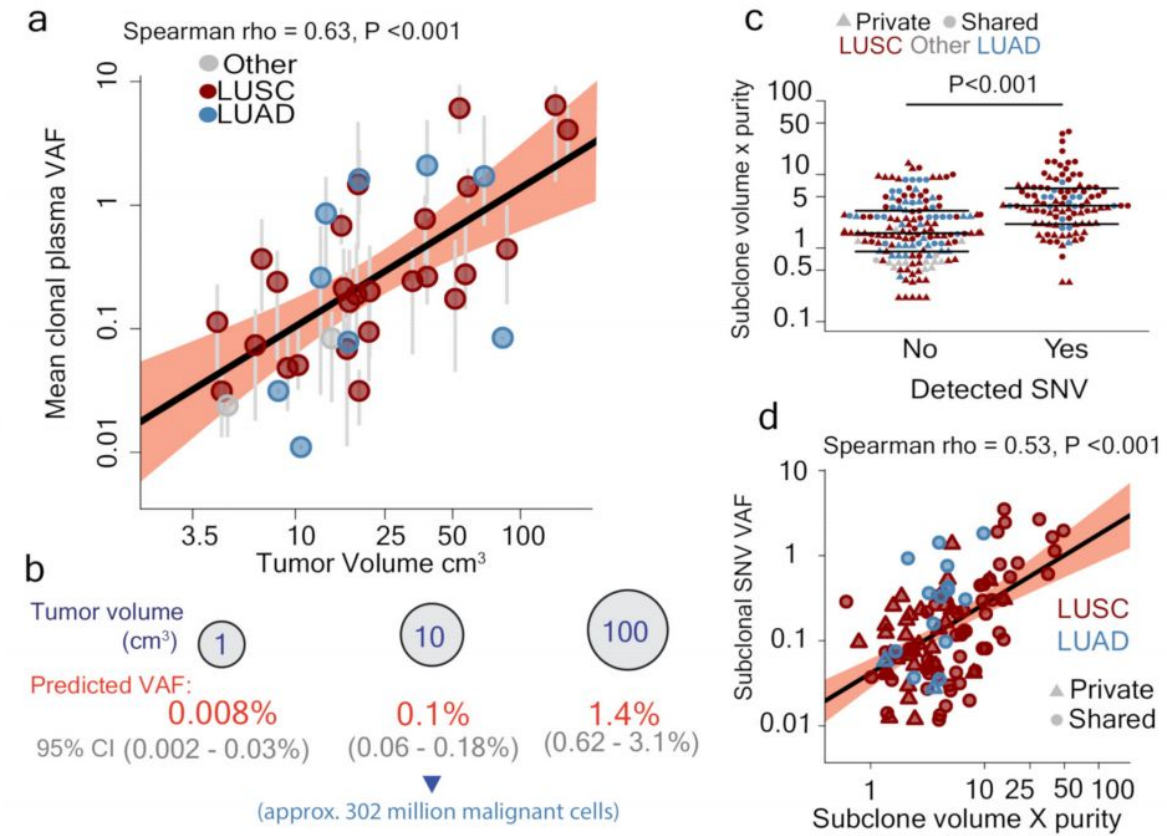
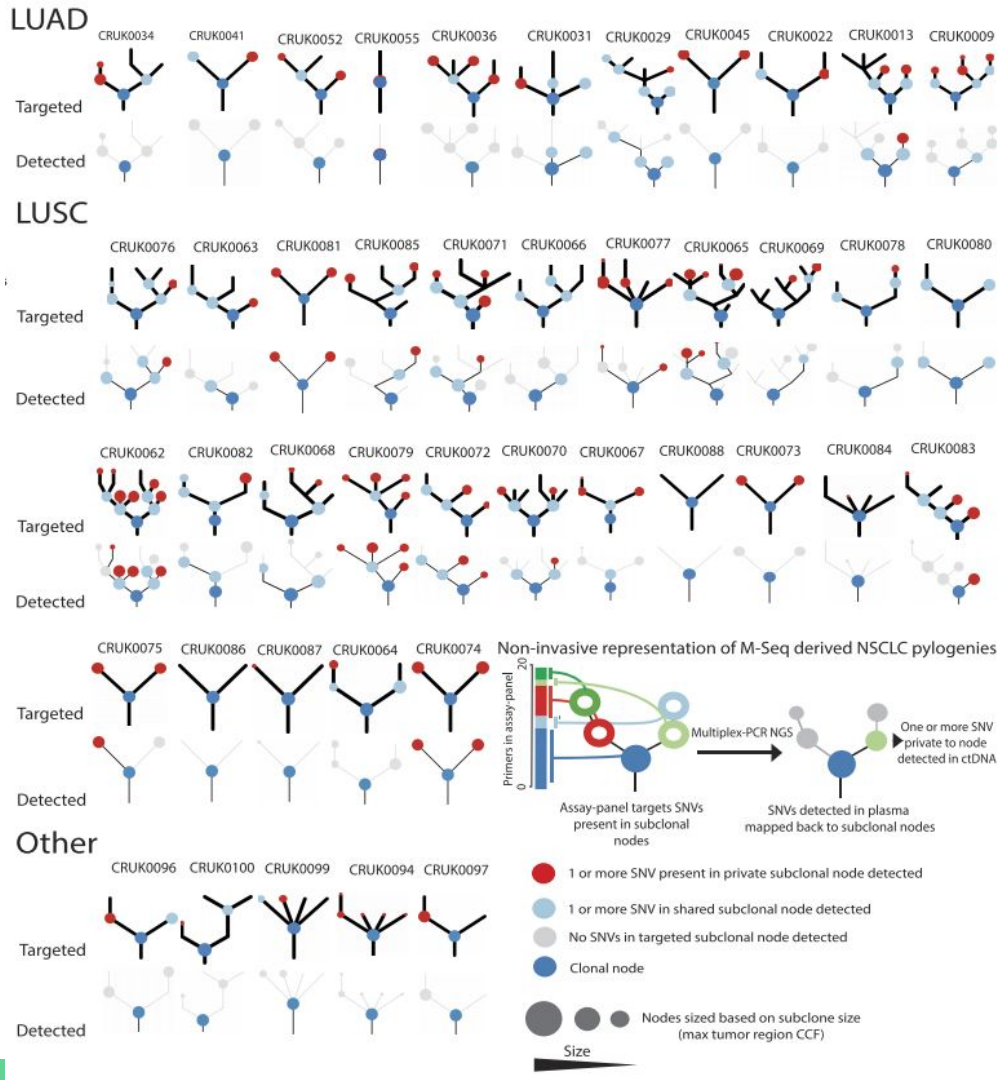


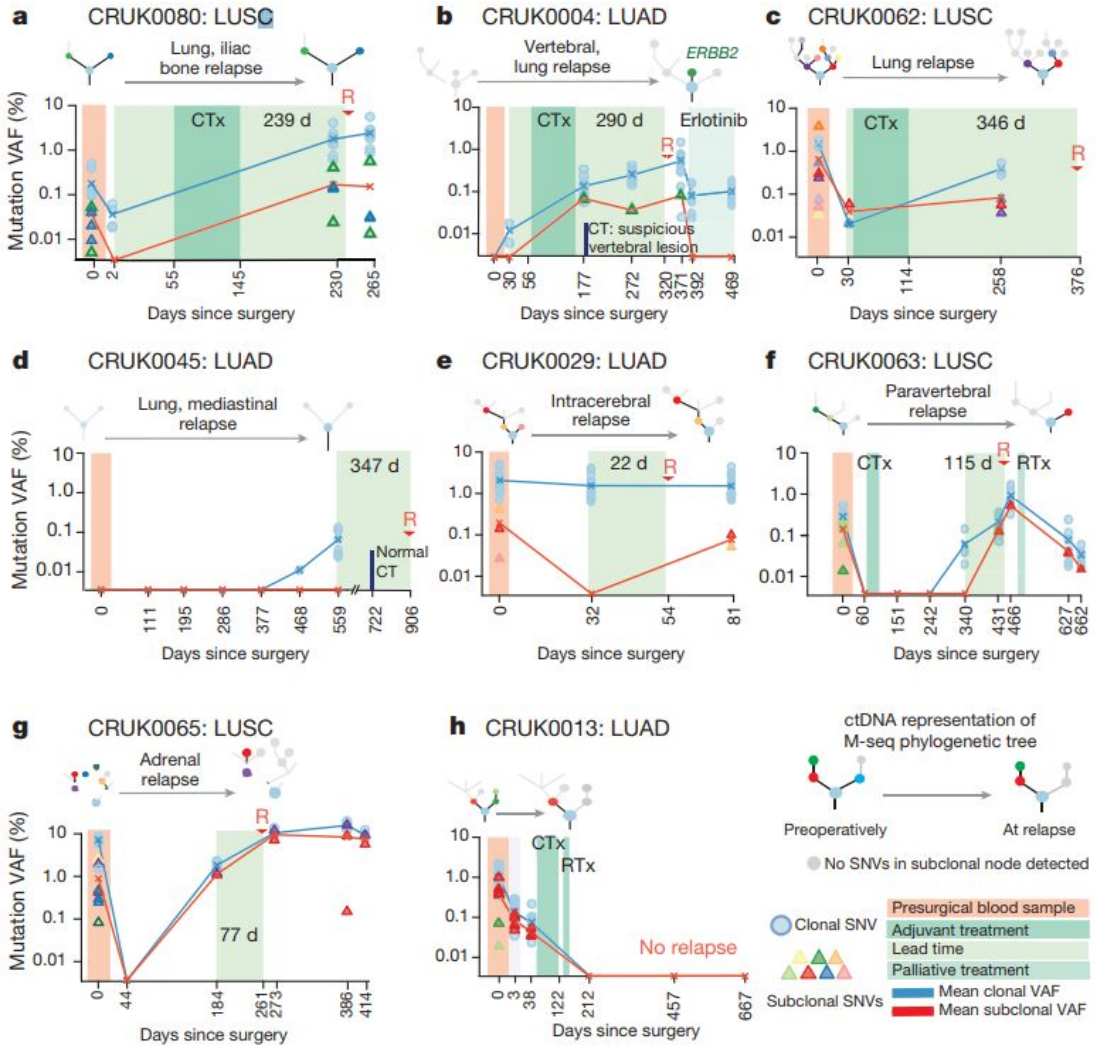
Figure 3. Tumor volume predicts plasma variant allele frequency

- Detected subclonal SNVs were mapped back to M-seq-derived tumour phylogenetic trees
- Detected private subclones (subclones identified within only a single tumour region) are coloured red. Shared subclones (subclones detected in more than one tumour region) are light blue.
- The top row of the phylogenetic trees represent subclonal nodes targeted by primers within that patient's assay panel
- The bottom row represent subclonal nodes detected in ctDNA
    - Grey subclonal nodes => not detected in ctDNA.

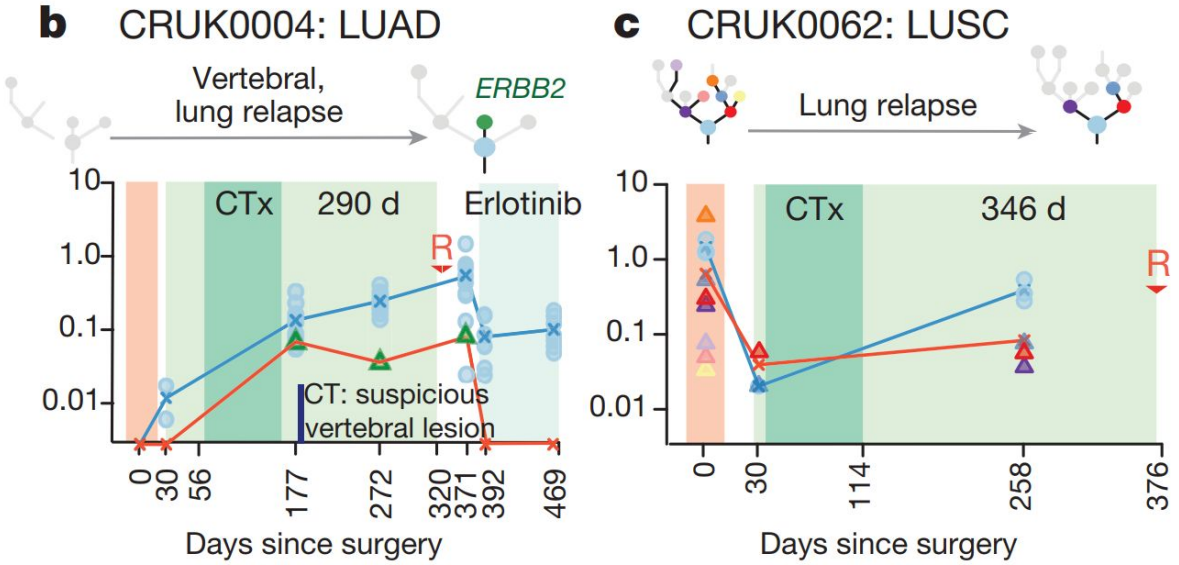


Targets of this paper:

1. What factors determine ctDNA detection?
2. **Can you identify evidence of therapy resistance and relapse using ctDNA?**
3. Can you track the sub-clonal nature of relapse for ctDNA-driven therapeutics?



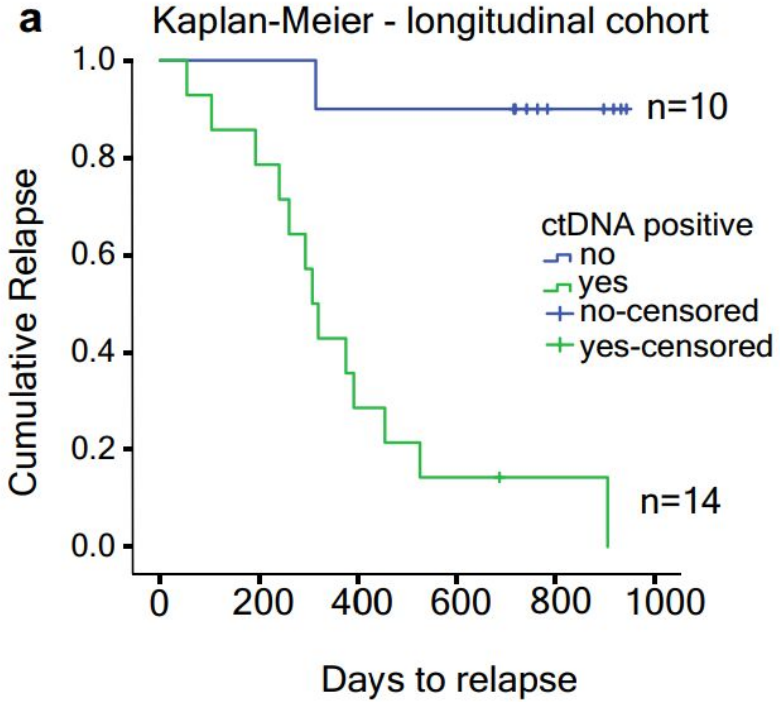
- ctDNA detection in plasma was defined as the detection of two tumour-specific SNVs.
- Preoperative and relapse phylogenetic trees represented by ctDNA are illustrated above each graph.
- The median interval between ctDNA detection and NSCLC relapse that was confirmed by CT imaging indicated by clinical and chest radiograph follow-up (lead time) was 70 days (range, 10–346 days).



- ctDNA detection in plasma was defined as the detection of two tumour-specific SNVs.
- Preoperative and relapse phylogenetic trees represented by ctDNA are illustrated above each graph.
- The median interval between ctDNA detection and NSCLC relapse that was confirmed by CT imaging indicated by clinical and chest radiograph follow-up (lead time) was 70 days (range, 10–346 days).

Stratification

Kaplan–Meier curve demonstrating relapse-free survival for patients in whom ctDNA was detected versus patients in whom ctDNA was not detected.



Targets of this paper:

1. What factors determine ctDNA detection?
2. Can you identify evidence of therapy resistance and relapse using ctDNA?
3. **Can you track the sub-clonal nature of relapse for ctDNA-driven therapeutics?**

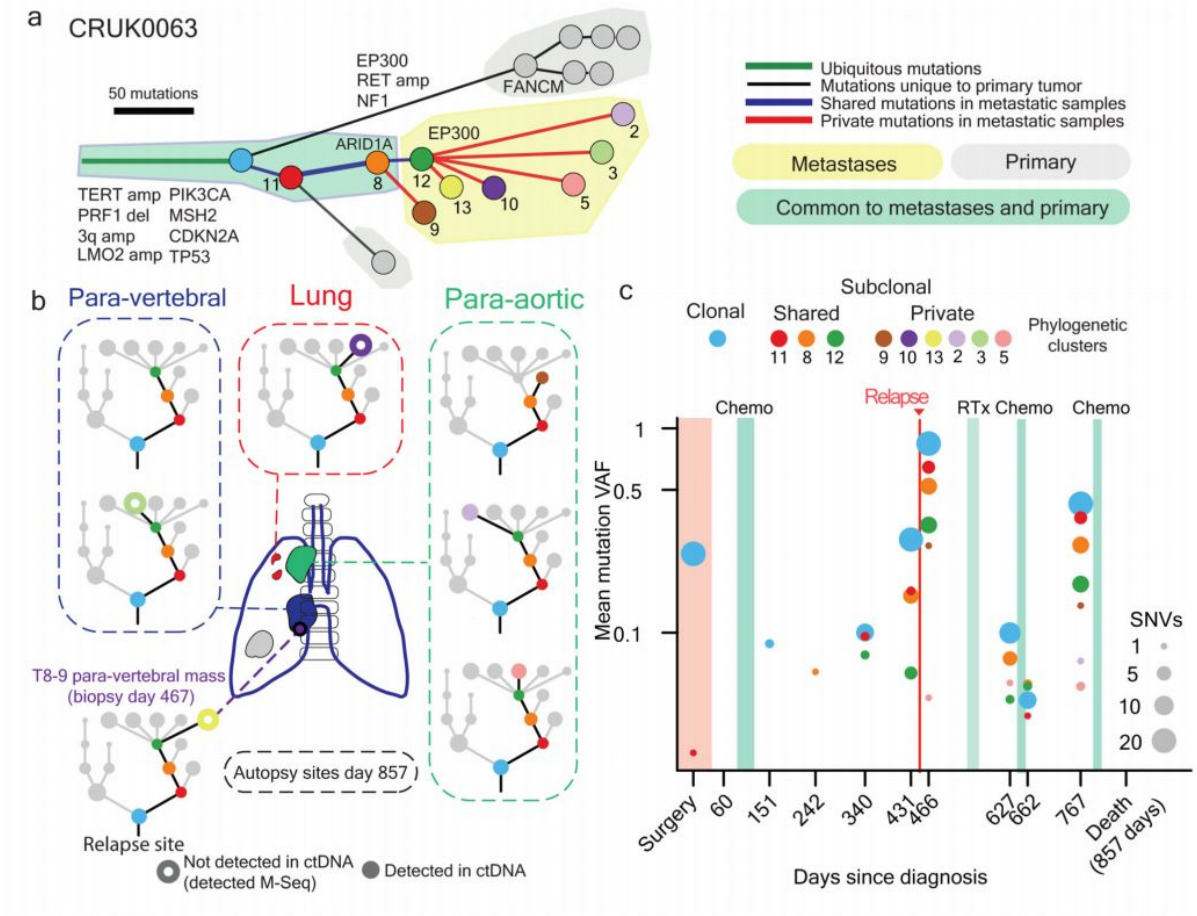


Figure 6. ctDNA tracking of lethal cancer subclones in CRUK0063

- Patient CRUK0063 was examined through the PEACE post-mortem study 24h after death
- Single SNVs from two private subclones (phylogenetic clusters 5 and 9) were also detectable in ctDNA at day 466
- The mean plasma VAF of the SNVs detected in phylogenetic clusters 11, 8, 12, 9 and 5 mirrored their proximity to the clonal cluster (light blue) in the M-seq-derived phylogenetic tree
- The mean clonal VAF decreased in response to palliative radiotherapy and chemotherapy, but increased at day 767

Feasibility (for primary tumor detection)

- A primary NSCLC tumour volume of 10 cm^3 predicted a ctDNA plasma VAF of 0.1%.
    - The sensitivity of the multiplex-PCR NGS platform was in excess of 99% at VAFs of 0.1% and above, suggesting optimum platform sensitivity with tumour burdens in excess of 10cm^3 .
    - Low-dose CT lung screening can identify lung nodules with diameters from 4mm = volume of 0.034 cm^3.
    - Based on this study, a tumour volume of 0.034 cm^3 would equate to a plasma VAF of 1.8×10^{-4} %, which is at the extreme of detection limits of current ctDNA platforms.

  Therefore, using current technologies, the sensitivity of clonal-SNV ctDNA-directed early-NSCLC screening may be constrained by tumour size.

- Cost: Estimated at US$1,750 per patient for sequencing of a single tumour region, synthesis of a patient-specific assay panel and profiling of five plasma samples.

Conclusions

- Predictors of ctDNA detection were characterized by non-adenocarcinoma histology, necrosis, increased proliferative indices and lymphovascular invasion.
- Tumour volume correlated with the mean plasma VAF of clonal SNVs in ctDNA-positive NSCLCs
- Demonstrated the ability to detect relapse, often prior to CT

To explore:

- Effect of non-uniqueness of phylogenetic trees
- Longitudinal analysis of clonal and subclonal composition variation
- More efficient modelling of PCR errors

Questions?

Thank you for your attention!