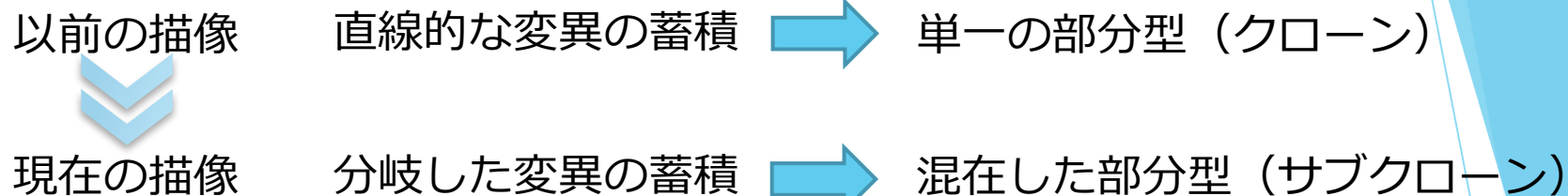
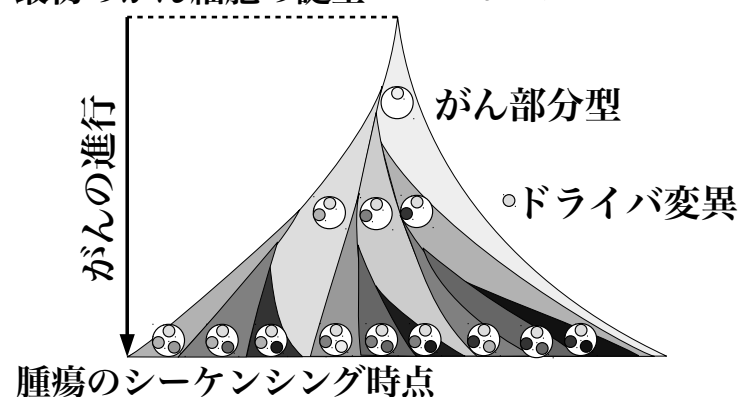


# 集団遺伝学を用いた腫瘍内不均一性の解析

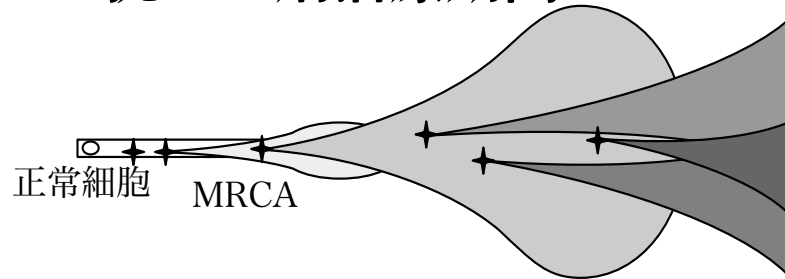
## がん進行過程



最初のがん細胞の誕生 ○ 正常細胞



抗がん剤治療期間



近年のがん治療      生化学検査による部分型の同定      ➡      部分型に応じた標準治療

## 目標

患者のクリニカルシーケンス

部分型構成比・増殖率の推定

- ・ 予後の予測
- ・ 最適な抗がん剤投薬計画の提案

# 先行研究

## 構成比推定

- ▶ PyClone (Roth et al., 2014, Nat Methods)

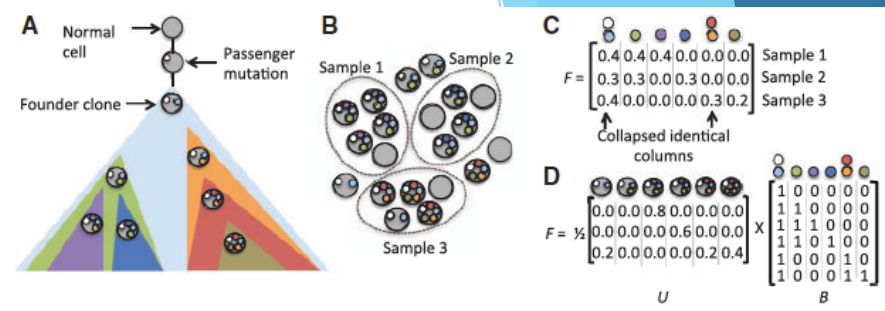
変異アレル頻度のベイジアンクラスタリング

- ▶ AncesTree (El-Kebir et al., 2015, Bioinformatics)

- 腫瘍内の複数箇所のサンプルを利用

変異アレル頻度  $\xrightarrow{\text{因数分解}}$  部分型頻度  $\times$  部分型系統樹

- ▶ いずれも ~~部分型誕生時刻推定~~ ~~増殖率推定~~



El-Kebir et al., 2015, Bioinformaticsより引用

## 増殖率推定

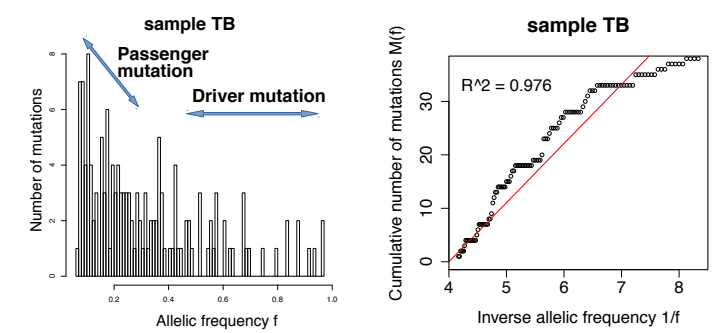
- ▶ がんの中立進化の同定 (Williams et al., 2016, Nat Genet)

がんの発展に直接寄与しない **パッセンジャー変異**  $\rightarrow$  **中立進化モデルを仮定**

腫瘍内の変異の累積度数  $M(f) = \mu_e \left( \frac{1}{f} - \frac{1}{f_{max}} \right)$

$\mu_e$  増殖率の情報

- ▶ NGSデータから変異コール  $\rightarrow$  線形回帰
- ▶ 単一の部分型  $\rightarrow$  線形回帰で増殖率を推定可能
- ▶ However, 複数の部分型  $\rightarrow$  ~~増殖率推定~~



Williams et al., 2016, Nat Genet のNGSデータを用いた追試

# 提案手法

## 集団遺伝学モデル

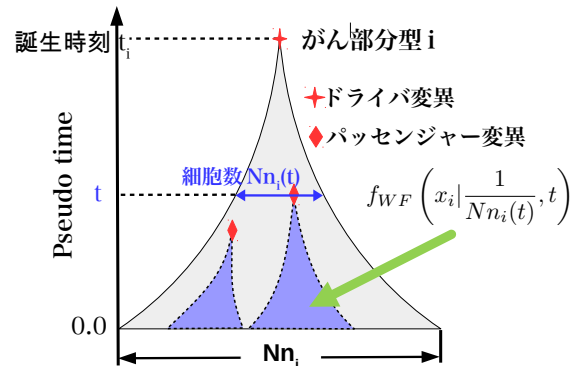
- ▶ パッセンジャー変異を持つ細胞と持たない細胞の分裂はランダム

➡ 変異アレル頻度(VAF)  $x$  は細胞分裂とともに拡散

拡散方程式

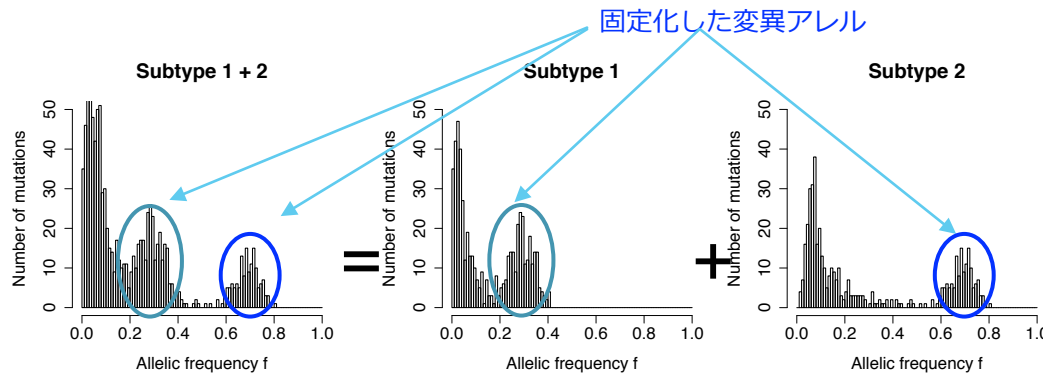
$$\frac{\partial f_{WF}(x|y, t)}{\partial t} = \frac{1}{2} \frac{\beta}{n(t)} \frac{\partial}{\partial x^2} (x(1-x) f_{WF}(x|y, t))$$

変異アレル頻度分布  $p(x_i|i, n_i, t_i) \propto \int_0^{t_i} dt f_{WF} \left( x_i | \frac{1}{N n_i(t)}, t \right) (N n_i(t) - 1)$



構成比  $n_i = \frac{1}{N} e^{\alpha_i t_i}$ ,  $N$ : 総細胞数,  $\alpha_i$ : 増殖率,  $t_i$ : 誕生時刻

## 混合分布モデル



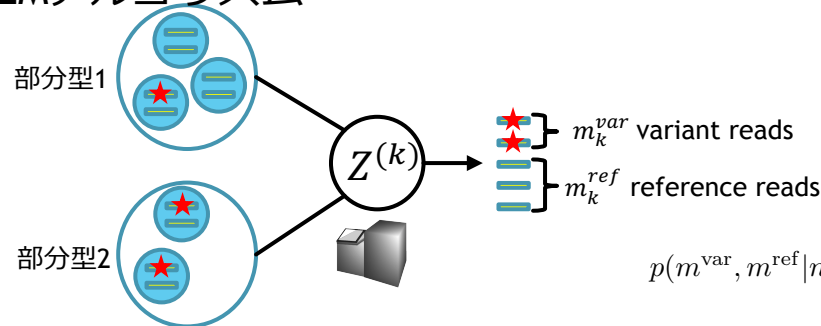
VAF分布形状

固定化したVAF

部分型誕生時刻を反映

部分型構成比を反映

誕生時刻  $t$ , 構成比  $n$  の下で、変異リード  $m^{var}$  本、  
リファレンスリード  $m^{ref}$  本が観測される確率  $p(m^{var}, m^{ref} | n, t)$   
EMアルゴリズム



$$p(m^{var}, m^{ref} | n, t) = \prod_{k=1}^K \sum_{i_k=1}^I p(i_k) \sum_{x_{i_k}} p(x_{i_k} | i_k, n_{i_k}, t_{i_k}) p(m_k^{var}, m_k^{ref} | n, x)$$

$K$ : #SNVs

各部分型  $i_k$  の変異率

リードの出力確率

先のWright-Fisher modelから計算される  
変異アレル頻度分布

$$p(m^{var}, m^{ref} | n, t) = \prod_{k=1}^K \sum_{Z^{(k)}} \prod_{i_k=1}^I \prod_{x_{i_k}} \prod_{q_k=1}^I \left[ p(i_k) p(x_{i_k} | i_k, n_{i_k}, t_{i_k}) \prod_{l=1}^{m_k^{var}} (n_{q_{kl}} x_{q_{kl}}) \prod_{l=m_k^{var}+1}^{m_k^{var}+m_k^{ref}} (n_{q_{kl}} (1-x_{q_{kl}})) \right] Z^{(k)}$$

変異リードが部分型  $q_{kl}$  から出力される確率

リファレンスリードが部分型  $q_{kl}$  から出力される確率

# Results and Discussion

- ▶ 正解の誕生時刻 $t$ , 構成比 $n$ を用意
- ▶ 上記 $t, n$ に基づき計算される変異アレル頻度からリードを生成
- ▶ 生成したリードから、 $t, n$ を最急降下法により最尤推定

$$p(m|M, n, t) = \prod_{k=1}^K \int_0^1 dx_{i_k} p(m_k|M_k, n, x) p(x_{i_k}|i_k, n_{i_k}, t_{i_k})$$

$K = \text{\#SNV loci}$

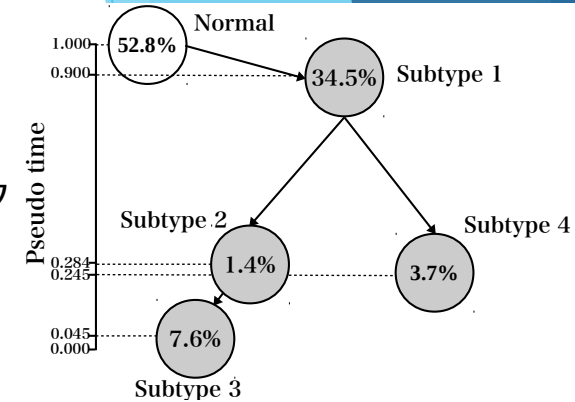
$$(t, n) = \arg \max_{t, n} \ln p(m|M, n, t)$$

- ▶ 集団遺伝学を応用し、変異アレル頻度分布から、各部分型の構成比および誕生時刻が推定できた。

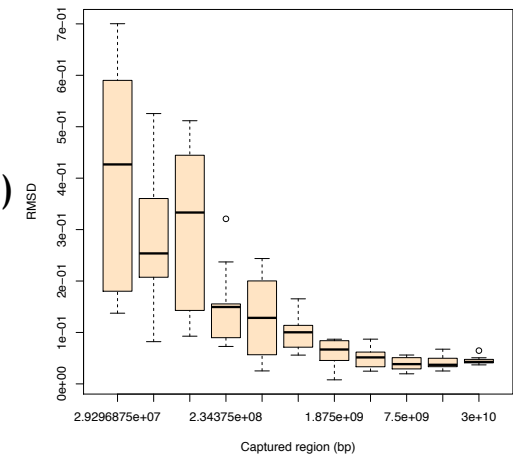
## Future work

- ▶ 今回のパラメータ推定ではシミュレーションデータを用いたが、シーケンシング実データに対して適用する。
- ▶ 1時刻のシーケンシングデータから推定できるのは、部分型間の相対的なPseudo timeと増殖率である。時系列シーケンシングデータから、絶対時刻と増殖率を推定し、がんのクローン進化過程のダイナミクスを解析できる手法を開発する。

正解パラメータ



推定精度 (t)



推定精度 (n)

