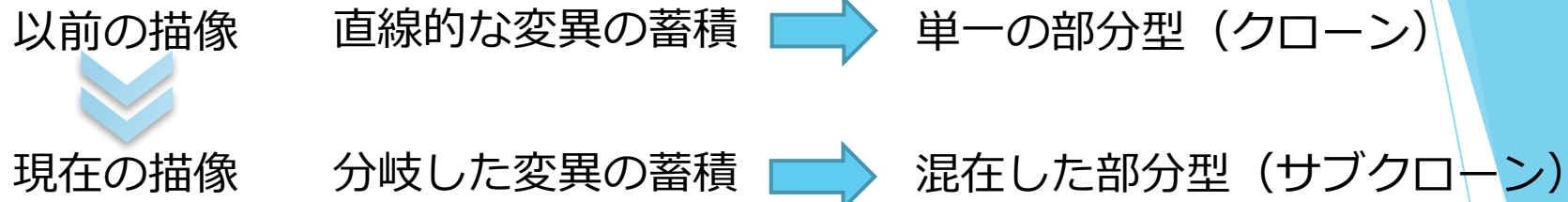


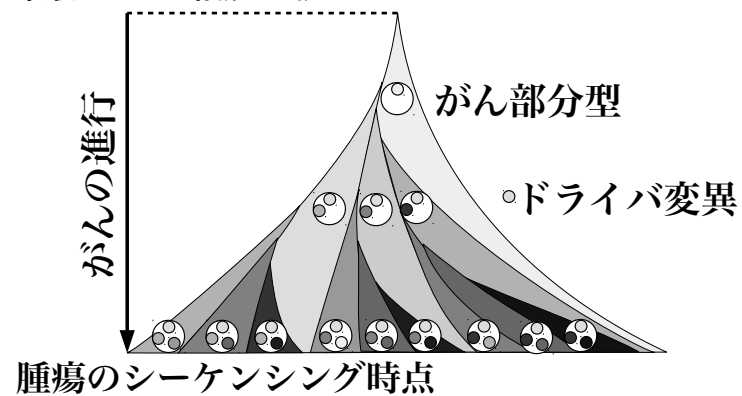
集団遺伝学を用いた腫瘍内不均一性の解析

東京大学 メディカル情報生命 今田雄太郎

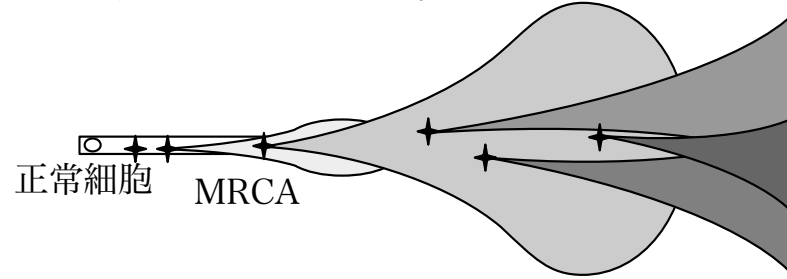
がん進行過程



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

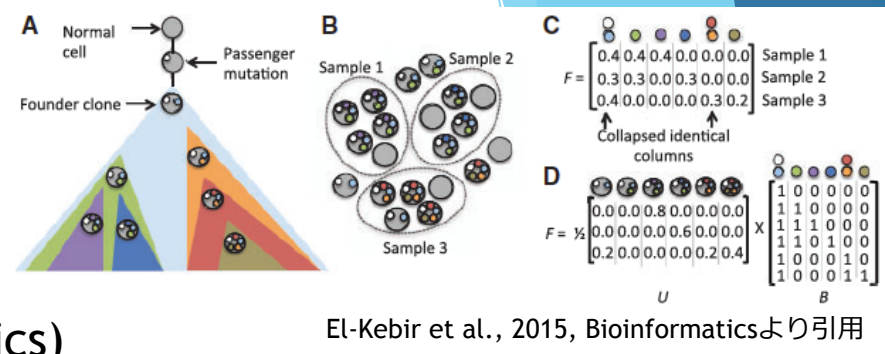
部分型構成比・増殖率の推定

- ・ 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ジアンクラスタリング
- ▶ AncesTree (El-Kebir et al., 2015, Bioinformatics)



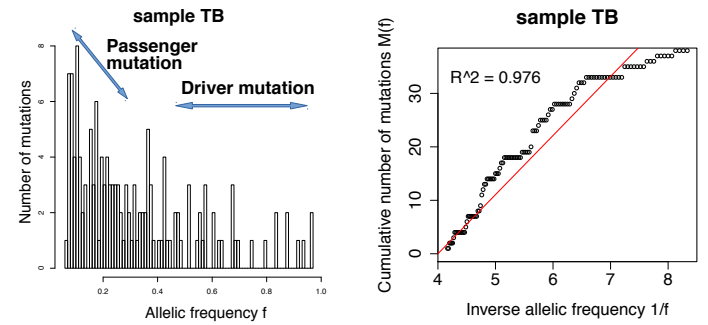
El-Kebir et al., 2015, Bioinformaticsより引用

- 腫瘍内の複数箇所のサンプルを利用
- 変異アレル頻度 **因数分解** 部分型頻度 **×** 部分型系統樹
- ▶ いずれも ~~部分型誕生時刻推定~~ ~~増殖率推定~~

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** **→** **中立進化モデルを仮定**

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



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

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

提案手法

集団遺伝学モデル

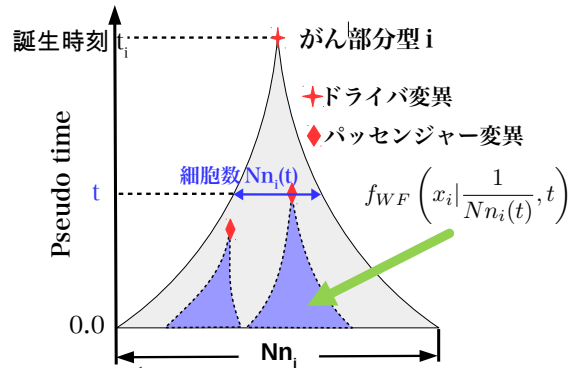
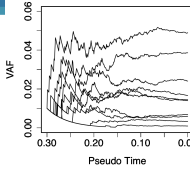
- ▶ パッセンジャー変異を持つ細胞と持たない細胞の分裂はランダム

➡ 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 $p(x_i|i, n_i, t_i) \propto \int_0^{t_i} dt f_{WF}\left(x_i \middle| \frac{1}{Nn_i(t)}, t\right) (Nn_i(t) - 1)$



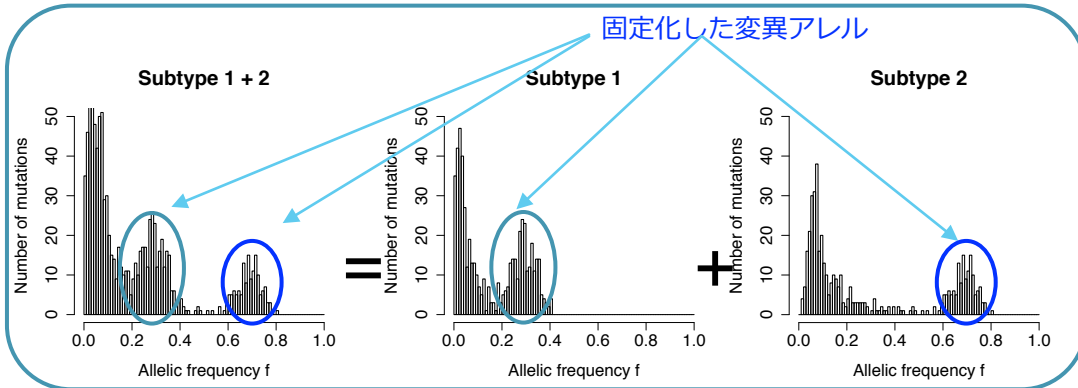
構成比 $n_i = \frac{1}{N} e^{\alpha_i t_i}$, N : 総細胞数, α_i : 増殖率, t_i : 誕生時刻

混合分布モデル

- 混合分布モデルのパラメータ推定問題として、各部分型の構成比と誕生時刻を最尤推定

VAF分布形状 ➡ 誕生時刻を反映

固定化したVAF ➡ 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

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

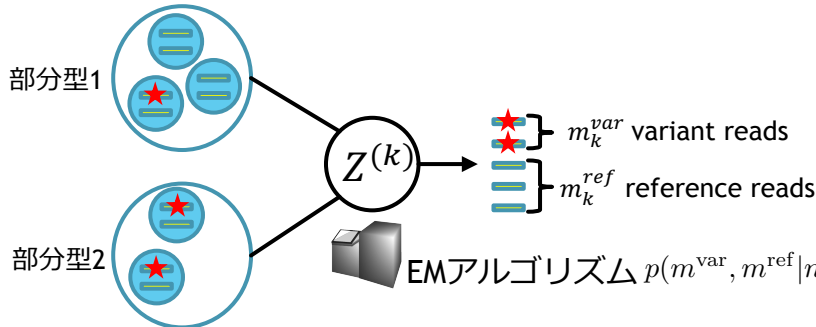
各部分型 i_k の変異率

リードの出力確率

K : #SNVs

上の集団遺伝学から計算される
変異アレル頻度分布

最尤推定問題 $(n, t) = \arg \max_{n, t} \ln p(m^{var}, m^{ref} | n, t)$



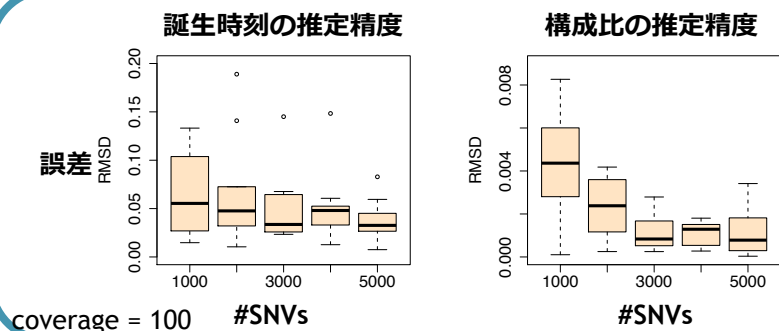
$$p(m^{var}, m^{ref} | n, t) = \prod_{k=1}^K \sum_{Z^{(k)}} \prod_{i_k=1}^I \prod_{x_{i_k}} \prod_{q_k=1}^I \left[p(i_k) p(x_{i_k} | i_k, n_{i_k}, t_{i_k}) \prod_{l=1}^{m_k^{var}} (n_{q_{kl}} x_{q_{kl}}) \prod_{l=m_k^{var}+1}^{m_k^{var}+m_k^{ref}} (n_{q_{kl}} (1-x_{q_{kl}})) \right] Z^{(k)}$$

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

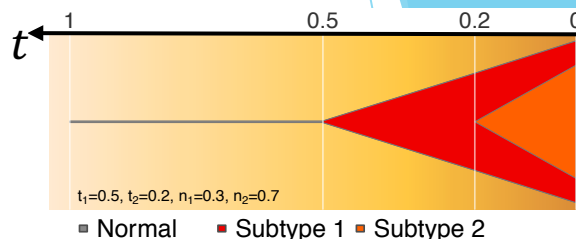
結果・考察 及び 今後の課題

- ▶ 正解の誕生時刻 t , 構成比 n を用意
- ▶ 正解 t, n に基づき計算される変異アレル頻度 → リードをシミュレーション
- ▶ シミュレーションしたリード → t, n をEMアルゴリズムにより最尤推定

SNV数と推定精度

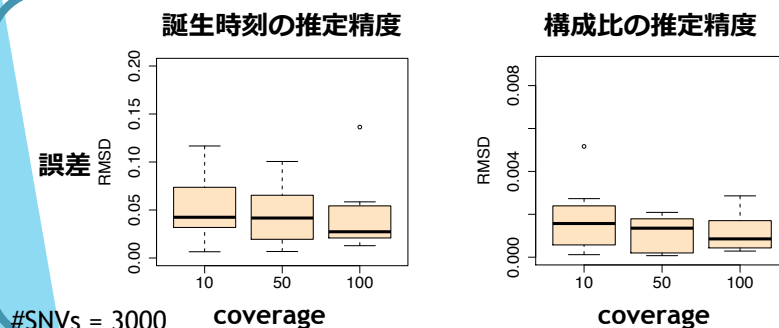


正解パラメータ



- ▶ SNVの数が多いほど、またcoverageが高いほど、誕生時刻 t と構成比 n の推定精度が向上
- ▶ 構成比推定 → 比較的少ない数のSNVでも推定可能
- ▶ 誕生時刻推定 → 多くのSNVが必要

coverageと推定精度



今後の課題

- ▶ シーケンシング実データへの適用
- ▶ 抗がん剤治療中のクローン進化のダイナミクスの解明
- ▶ 血中循環腫瘍DNAの時系列クリニカルシーケンスを用い、部分型の構成比および増殖率を逐次推定しながら、予後の予測、および抗がん剤投薬計画を最適化する手法を、制御工学の最適制御を応用することにより開発