Modeling cancer evolution from genomic data

Niko Beerenwinkel
Cancer is an evolutionary process

First mutant

Normal cells

Genetic progression

Carcinoma
Clonal expansion

- Normal cell
- MRCA cell
- Driver mutations
- Time point X: diagnosis and treatment initiation
- Time point Y: distant and local relapse
- Distant metastasis

Yates et al 2012
Intra-tumor heterogeneity

Marusyk et al 2012
Some challenges

1. Mutation calling
2. Predicting the phenotypic effects of mutations
3. Reconstructing the evolutionary history of a tumor
4. Predicting cancer evolution and progression
Phylogenetic vs. oncogenetic models

Phylogenetic models

Oncogenetic models
SCITE: Tree inference for single-cell data

Katharina Jahn, Jack Kuipers (RECOMB 2016)
Intra-tumor phylogeny

Infinite sites assumption
Observation error

\[ E = \begin{array}{ccccccc}
 & s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \\
M_1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
M_2 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
M_3 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
\end{array} \]

\[ D = \begin{array}{ccccccc}
 & s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \\
M_1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
M_2 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
M_3 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
\end{array} \]

\[ \alpha = \text{false positive rate} \]
\[ \beta = \text{false negative rate} \]

\[ P(D_{ij} = 1|E_{ij} = 0) = \alpha, \quad P(D_{ij} = 0|E_{ij} = 0) = (1 - \alpha) \]
\[ P(D_{ij} = 0|E_{ij} = 1) = \beta, \quad P(D_{ij} = 1|E_{ij} = 1) = (1 - \beta) \]
Likelihood

- $n$ mutations, $m$ samples
- Tree topology, $T$
- Attachment of samples, $\sigma$
- Error rates, $\theta = (\alpha, \beta)$
- Likelihood:

$$P(D|T, \sigma, \theta) = \prod_{i=1}^{n} \prod_{j=1}^{m} P(D_{ij}|E_{ij})$$

- Posterior:

$$P(T, \sigma, \theta|D) \propto P(D|T, \sigma, \theta)P(T, \sigma, \theta)$$

$$P(\sigma|T, \theta)P(T, \theta)$$
Attachment of samples

\[
\frac{P(T, \theta | D)}{P(T, \theta)} \propto \sum_\sigma \prod_{j=1}^m \left[ \prod_{i=1}^n P(D_{ij} | A(T)_{i\sigma_j}) \right] P(\sigma_j | T, \theta) \\
= \prod_{j=1}^m \sum_{\sigma_j=1}^{n+1} \left[ \prod_{i=1}^n P(D_{ij} | A(T)_{i\sigma_j}) \right] P(\sigma_j | T, \theta) \quad O(nm)
\]
Inference

- For \( n \) mutations and \( m \) samples, the search space is 
  \[ ((n + 1)^{(n-1)}) \times [(n + 1)^m] \times \mathbb{R}^2 \]
- and \( ((n + 1)^{(n-1)}) \times \mathbb{R}^2 \) after marginalization
- MCMC:
Single-cell sequencing of a myeloproliferative neoplasm (Hou et al., Cell 2012)

- WES of 58 cancer cells
- 18 selected mutations
- 45% missing data
- \( \alpha = 6.04 \times 10^{-5}, \, \beta = 0.43 \)
Single-cell sequencing of an ER$^+$ breast tumor (Wang et al., Nature 2014)

- nuc-seq of 47 cells
- 40 mutations
- 1.4% missing data
- $\alpha = 1.24 \times 10^{-6}$
- $\beta = 0.097$
past clonal expansions

current, co-existing subclones
pathTiMEx: Mutually exclusive cancer pathways and their dependencies in tumor progression

Simona Cristea, Jack Kuipers (RECOMB 2016)
Partial order among pathways
Individual progression

![Diagram showing individual progression of tumors with nodes A, B, C, and D, and independent realizations leading to diagnosis of tumors 1, 2, 3, and 4.]

Partial order

Individual progression
Genotypes at diagnosis

Partial order

Individual progression

Genotypes X

Tumor 1
Tumor 2
Tumor 3
Tumor 4

A B C D
Tumor 1 0 1 0 0
Tumor 2 1 1 1 0
Tumor 3 1 1 0 1
Tumor 4 1 1 1 0
Tumor 5 1 1 1 1
Tumor 6 1 1 0 0
Tumor 7 1 1 1 0
...
Observed genotypes

Partial order

Individual progression

Genotypes X

Noisy observations Y
Waiting time distribution

\[ T_S \sim \text{Exp}(\lambda_S) \]

\[ T_A \sim \text{Exp}(\lambda_A) \]
\[ T_B \sim \text{Exp}(\lambda_B) \]
\[ T_C \sim T_A + \text{Exp}(\lambda_C) \]
\[ T_D \sim \max(T_A, T_B) + \text{Exp}(\lambda_D) \]

Hidden conjunctive Bayesian network (H-CBN)

Time to occurrence of mutations

Time to diagnosis

censoring

Genotype

Observed genotype

noise
Can we find pathways de novo?

- For example, as groups of mutually exclusive genes:

\[ \mu_N P \left( T_K < \min_{i \in N \setminus K} (T_i, T_{obs}) \right) + (1 - \mu_N) P \left( T_K \leq T_{obs} < \min_{i \in N \setminus K} T_i \right) \]

\[ P(g_{\theta} \mid \theta_{ME}) = P \left( T_{obs} < \min_{i \in N} T_i \right) \]

Constantinescu, Szczurek, et al. 2015
Pathways and their dependencies

- A pathway is altered as soon as one of its genes is altered.
- Genes depend on upstream pathways.
Joint inference
Progression in colorectal cancer

Vogelstein et al., Science 2013
Progression in glioblastoma

CDKN2A(D)
CDK4(A)

TP53
MDM4(A)
MDM2(A)

NF1
FAF1(D)
SPTA1
OBSCN
CNTNAP2

PTEN
PTEN(D)
PIK3CA
IDH1

PDGFRA(A)
LRP2

EGFR
RB1(D)
TP53(D)
PAOX

$\lambda = 412$
$\lambda = 0.89$
$\lambda = 0.85$
$\lambda = 1.38$
$\lambda = 0.48$
$\lambda = 0.64$
Conclusions

- **SCITE**, https://gitlab.com/jahnka/SCITE
  - Single-cell sequencing data can be used to reconstruct the evolutionary history of individual tumors.
  - Two intra-tumor phylogenies support an evolutionary model of successive clonal expansions in which subclones co-exist until one of them reaches fixation.

- **pathTiMEx**, https://github.com/cbg-ethz/pathTiMEx
  - Tumor evolution is constrained by (partial) orders of gene and pathway alterations.
  - Mutually exclusive gene groups and their dependencies can be inferred jointly from observed mutation profiles.
Acknowledgements

CBG
Mathias Cardner
Simona Cristea
Madeline Diekmann
Christos Dimitrakopoulos
Simon Dirmeier
Monica Golumbeanu
Ariane Hofmann
Katharina Jahn
Vinay Jethava
Jack Kuipers
Brian Lang
Hesam Montazeri
Susana Posada Cesped
Hans-Joachim Ruscheweyh
Fabian Schmich
David Seifert
Jochen Singer
Ewa Szczurek
Thomas Thurnherr

Funding
ETH Zurich, SNSF, SHCS, SystemsX.ch, ERASysAPP, Swiss Cancer League, Horizon 2020, ERC Synergy

Collaborators (oncology)
Gerhard Christofori (U Basel), Mike Hall (U Basel), Markus Heim (U Basel), Willy Krek (ETHZ), Markus Manz (USZ), Florian Markowetz (CRUK), Holger Moch (USZ), Jörg Rahnenführer (TU Dortmund), Alejandro A. Schäffer (NIH), Bert Vogelstein (Johns Hopkins), Peter Wild (USZ)

www.cbg.ethz.ch