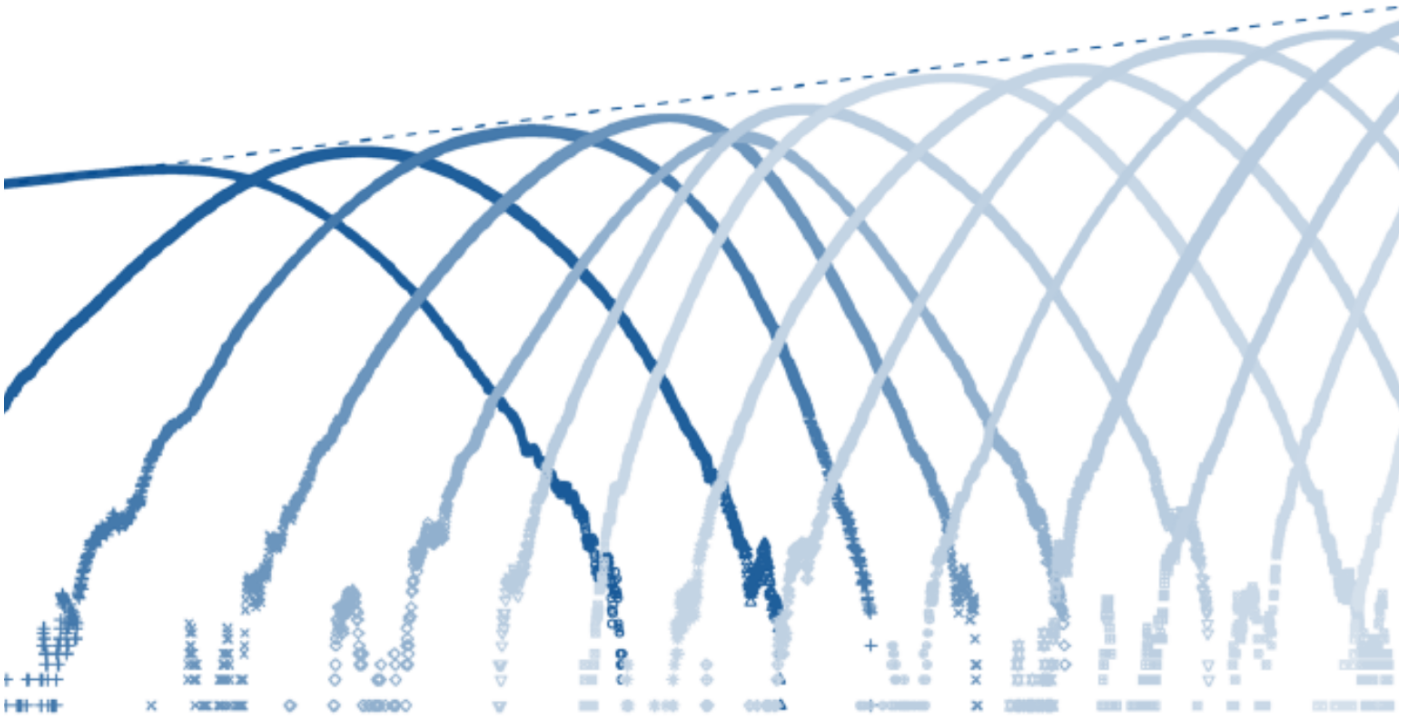
ETH zürich



D-BSSE
Department of Biosystems
Science and Engineering

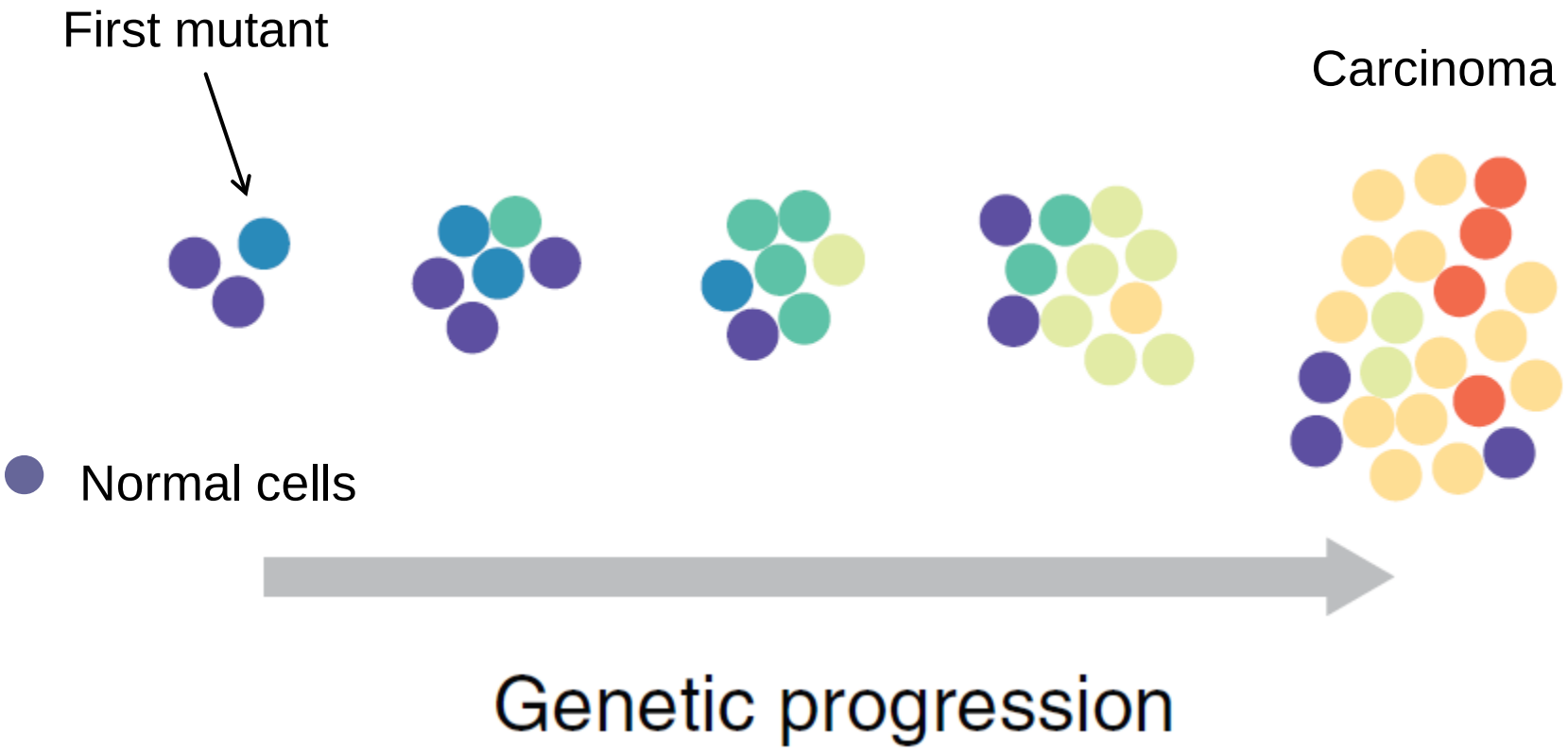
Competence Center
Personalized Medicine
ETH/UZH

SIB
Swiss Institute of
Bioinformatics

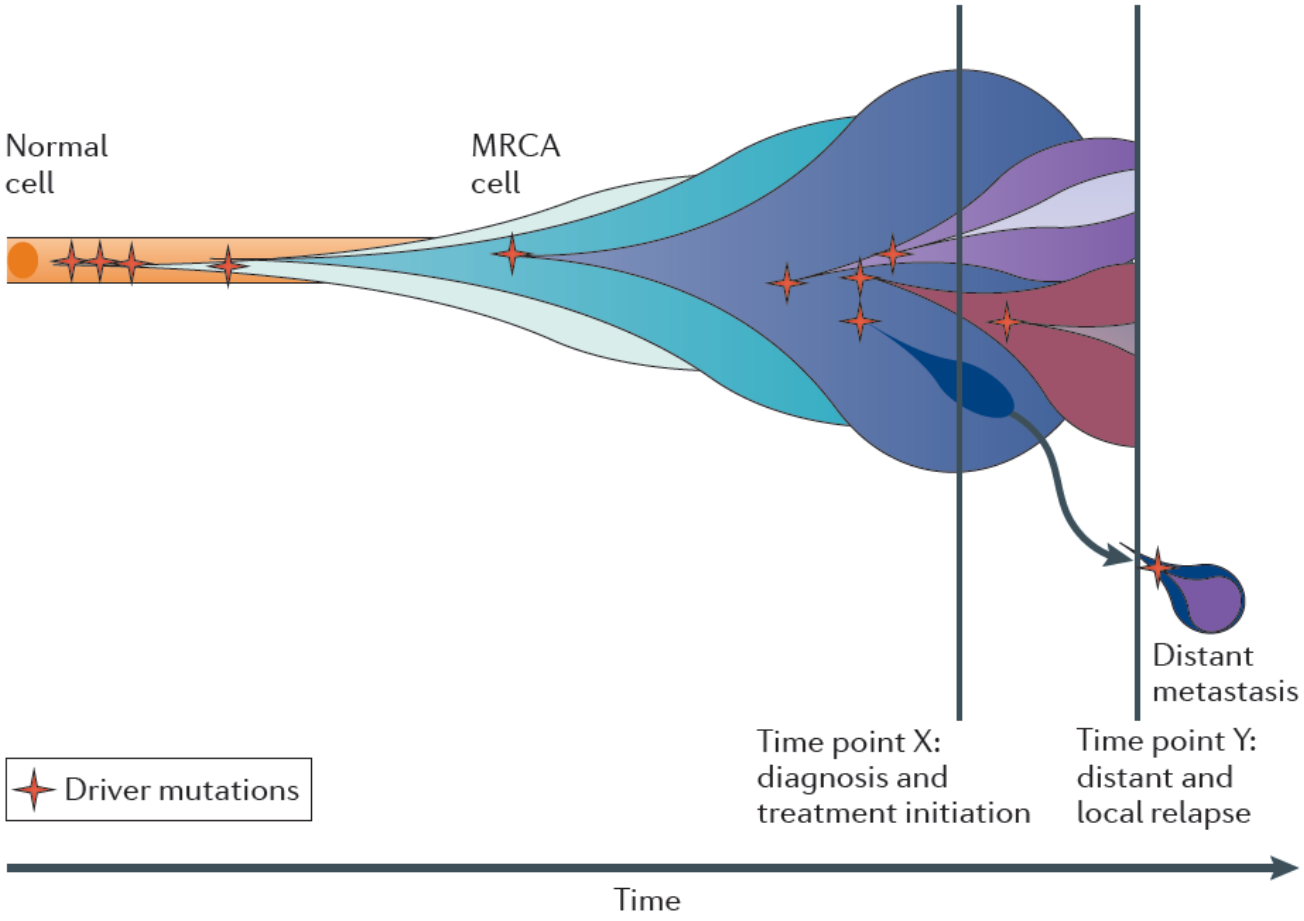
Modeling cancer evolution from genomic data

Niko Beerenwinkel

Cancer is an evolutionary process

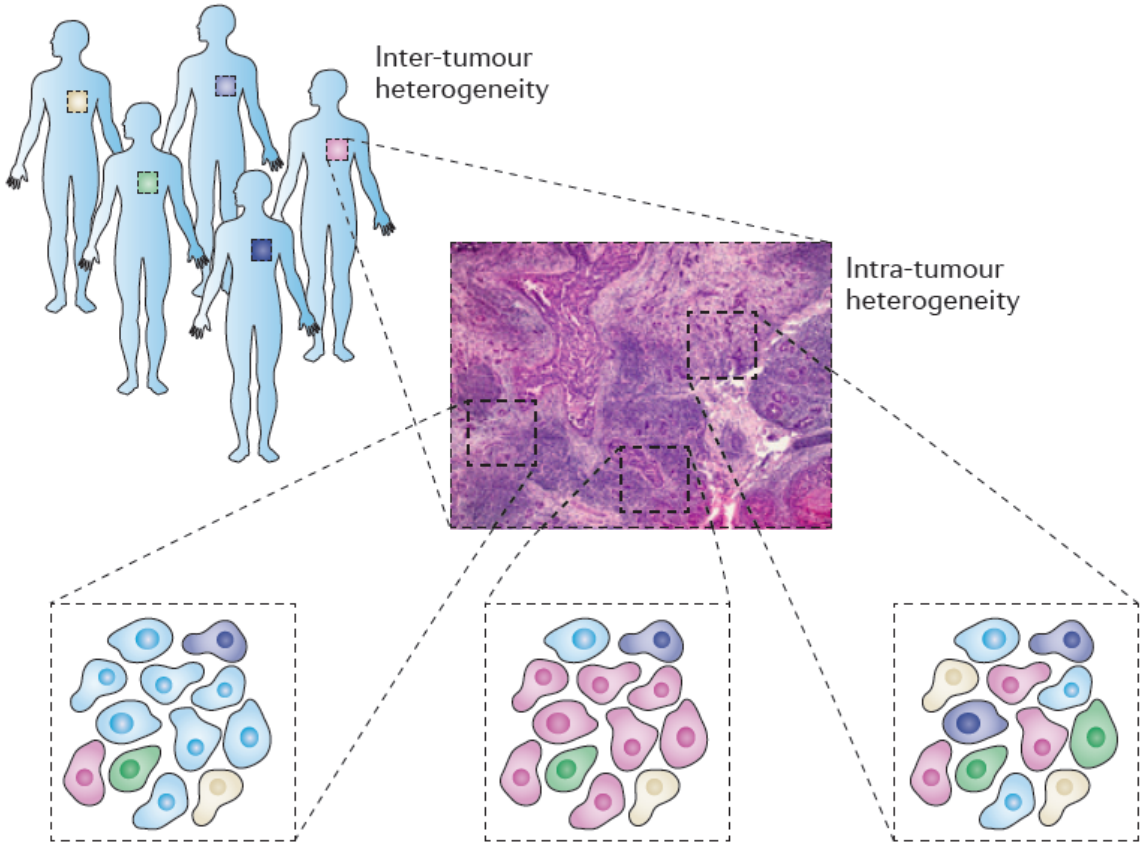


Clonal expansion



Yates et al 2012

Intra-tumor heterogeneity



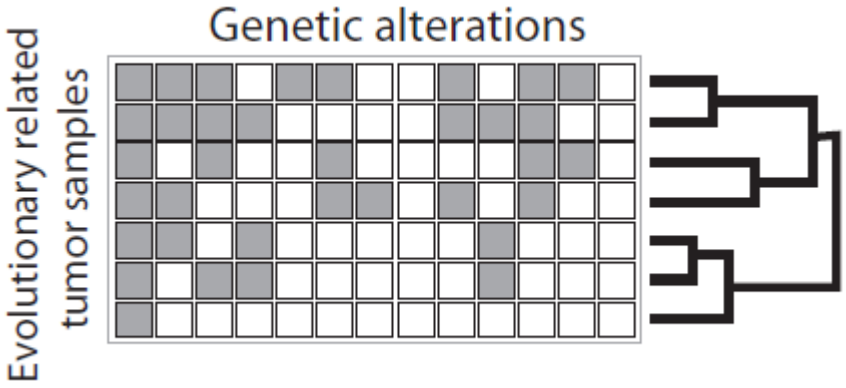
Marusyk et al 2012

Some challenges

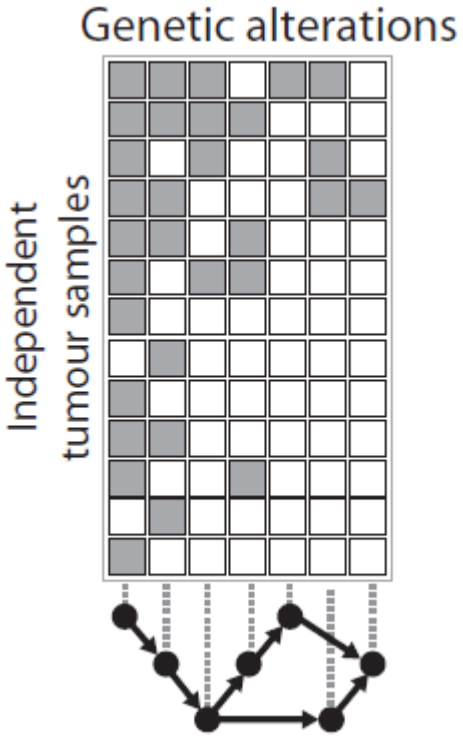
1. Mutation calling
2. Predicting the phenotypic effects of mutations
3. Reconstructing the evolutionary history of a tumor
4. Predicting cancer evolution and progression

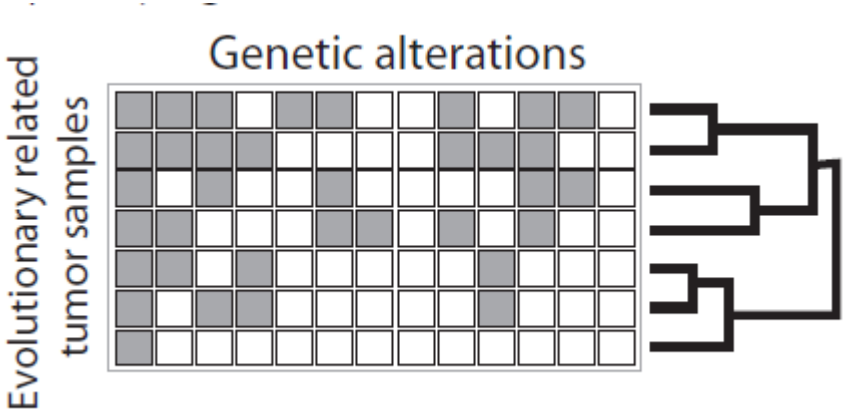
Phylogenetic vs. oncogenetic models

Phylogenetic models



Oncogenetic models

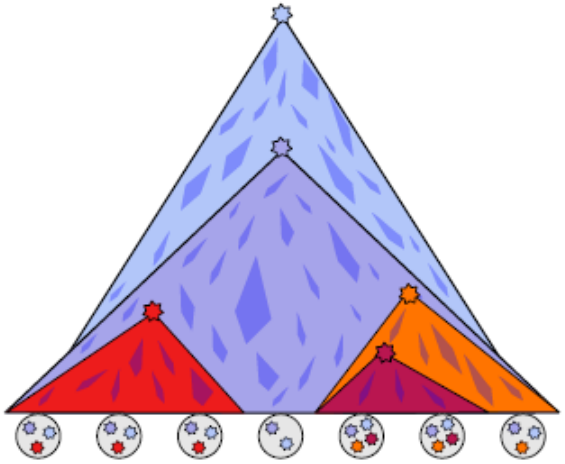




SCITE: Tree inference for single-cell data

Katharina Jahn, Jack Kuipers (RECOMB 2016)

Intra-tumor phylogeny

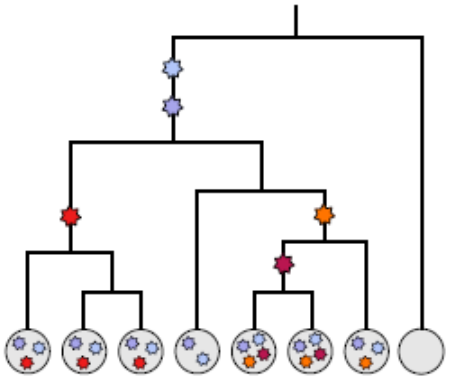


cells

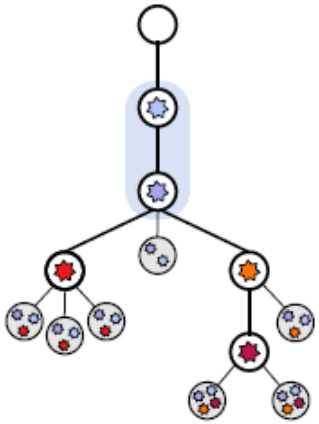
mutations							
	1	1	1	1	1	1	1
	1	1	1	1	1	1	1
	1	1	1	0	0	0	0
	0	0	0	0	1	1	1
	0	0	0	0	1	1	0

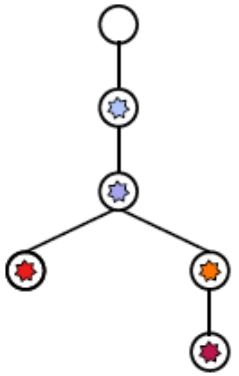
cells

mutations							
	1	1	–	1	1	–	1
	1	1	1	1	0	1	1
	1	1	1	0	0	0	0
	1	0	0	0	1	1	1
	0	0	–	0	0	1	0



infinite sites assumption





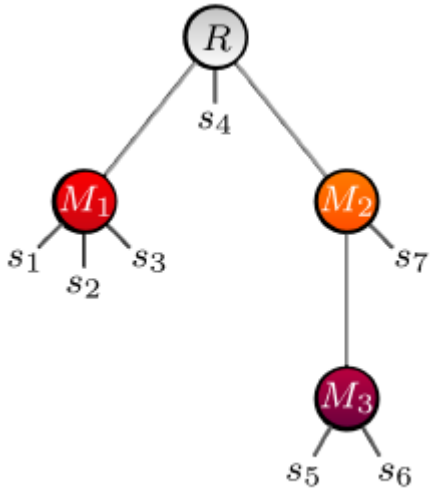
Observation error

$E =$

	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

$D =$

	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

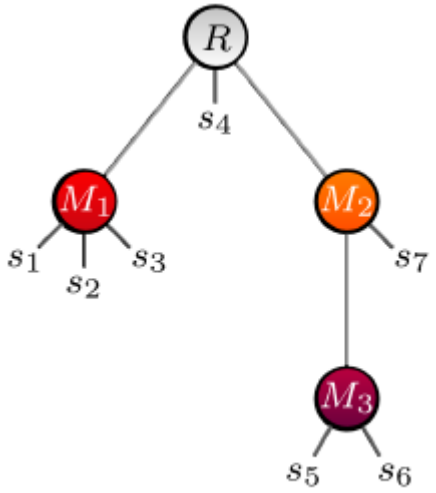


α = false positive rate
β = false negative rate

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

Likelihood

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

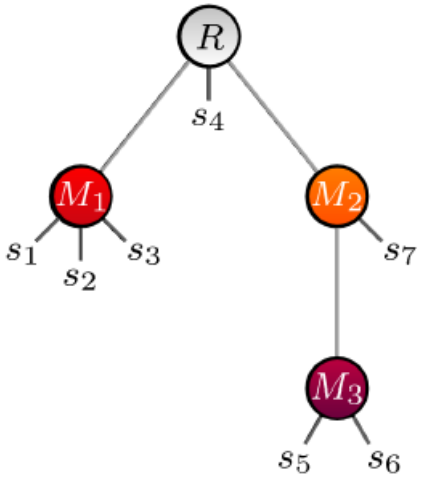


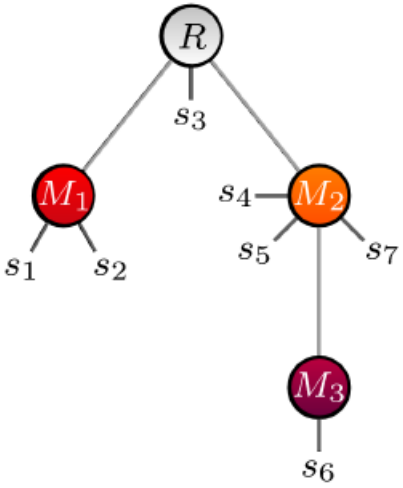
$$P(D|T,\boldsymbol{\sigma},\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{m} P(D_{ij}|E_{ij})$$

- Posterior:

$$P(T,\boldsymbol{\sigma},\boldsymbol{\theta}|D) \propto P(D|T,\boldsymbol{\sigma},\boldsymbol{\theta})\underbrace{P(T,\boldsymbol{\sigma},\boldsymbol{\theta})}_{P(\boldsymbol{\sigma}|T,\boldsymbol{\theta})P(T,\boldsymbol{\theta})}$$

Attachment of samples



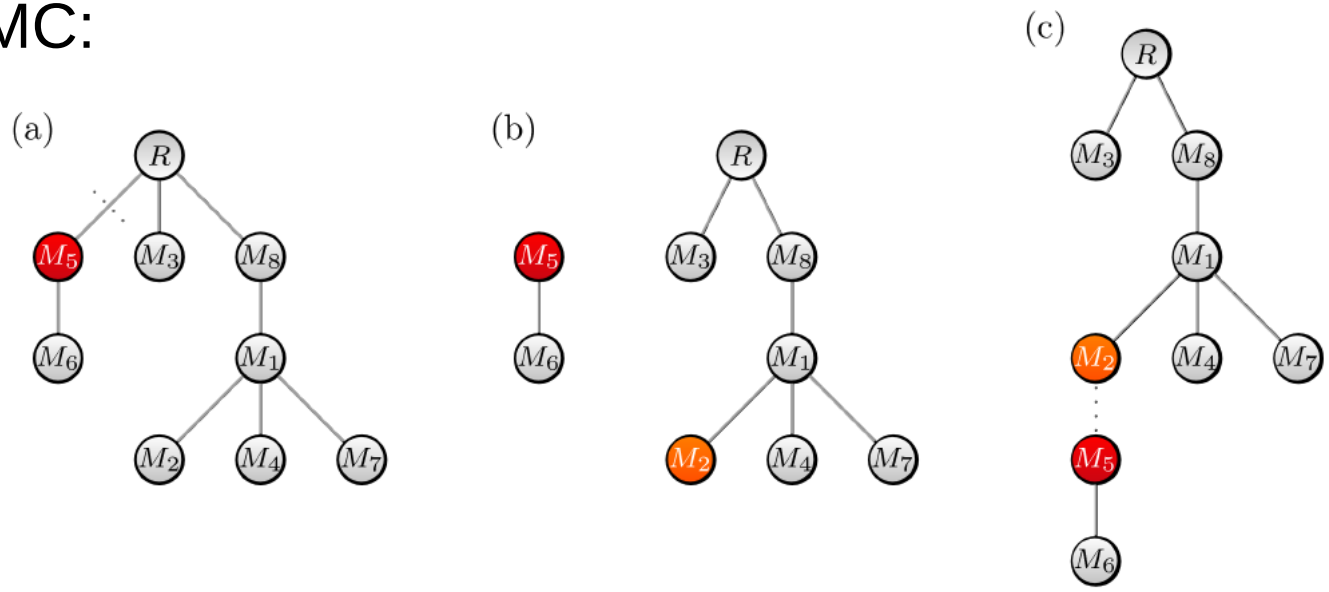


$$\frac{P(T,\boldsymbol{\theta}|D)}{P(T,\boldsymbol{\theta})} \propto \sum_{\boldsymbol{\sigma}} \prod_{j=1}^{m} \left[\prod_{i=1}^{n} P(D_{ij}|A(T)_{i\boldsymbol{\sigma}_j}) \right] P(\boldsymbol{\sigma}_j|T,\boldsymbol{\theta})$$

$$= \prod_{j=1}^{m} \sum_{\boldsymbol{\sigma}_j=1}^{n+1} \left[\prod_{i=1}^{n} P(D_{ij}|A(T)_{i\boldsymbol{\sigma}_j}) \right] P(\boldsymbol{\sigma}_j|T,\boldsymbol{\theta}) \qquad O(nm)$$

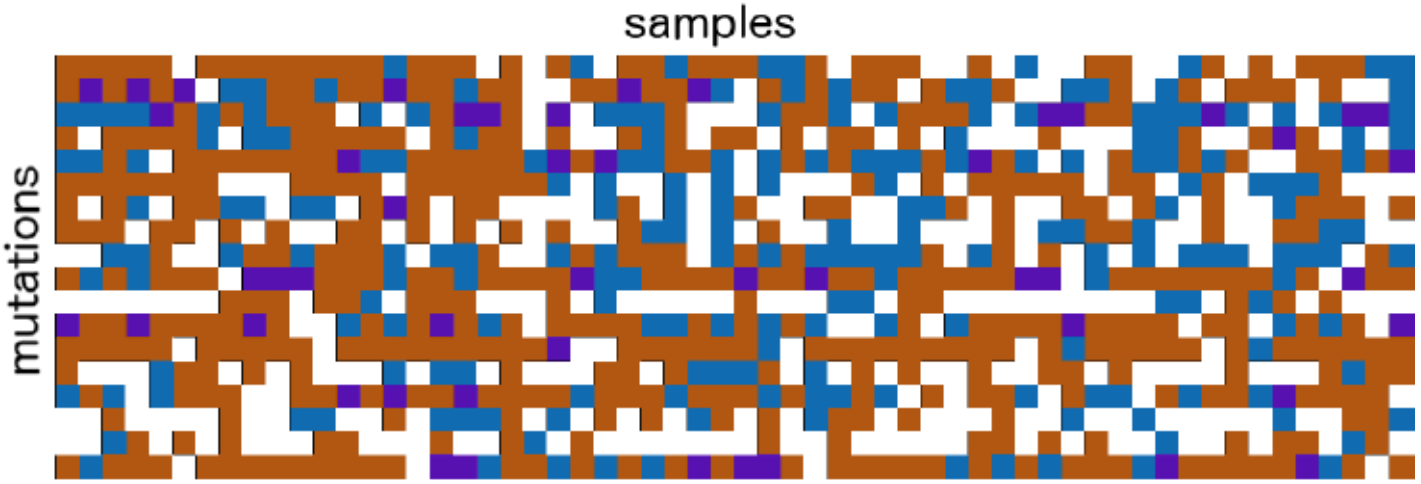
Inference

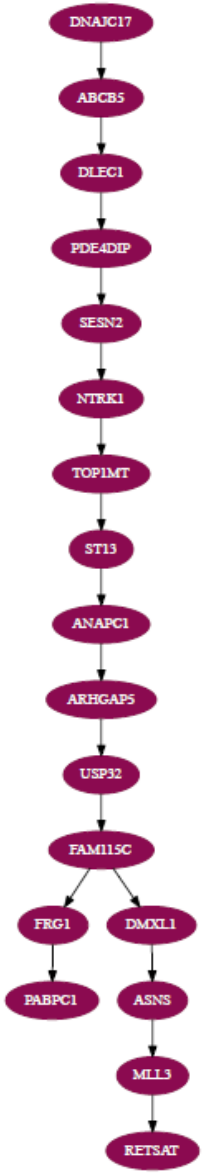
- For n mutations and m samples, the search space is
  $[(n+1)^{(n-1)}] \times [(n+1)^m] \times \mathbb{R}^2$
- and $[(n+1)^{(n-1)}] \times \mathbb{R}^2$ after marginalization
- MCMC:

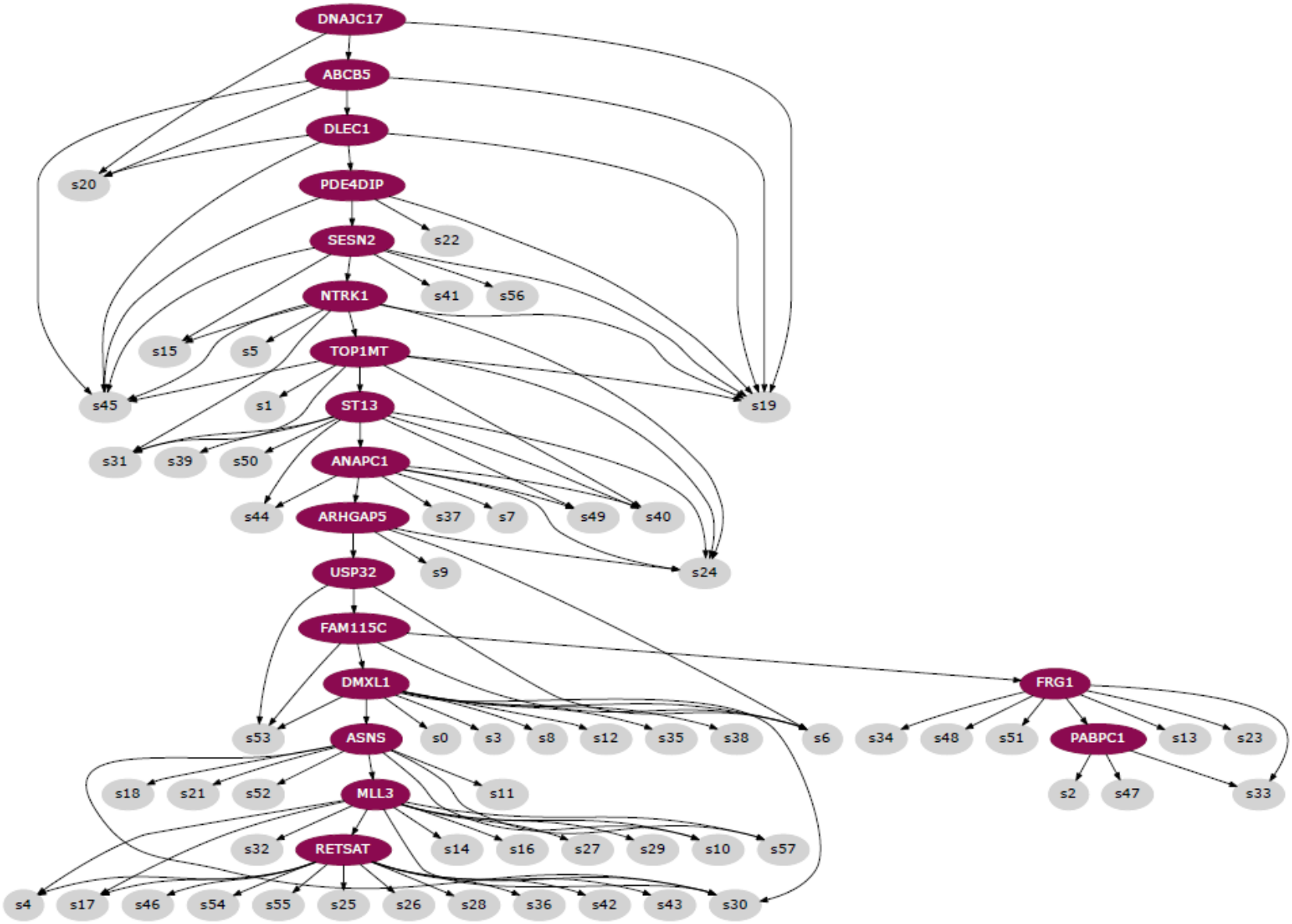


Single-cell sequencing of a myeloproliferative neoplasm (Hou et al., Cell 2012)

- WES of 58 cancer cells
- 18 selected mutations
- 45% missing data
- $\alpha = 6.04 \times 10^{-5}$, $\beta = 0.43$

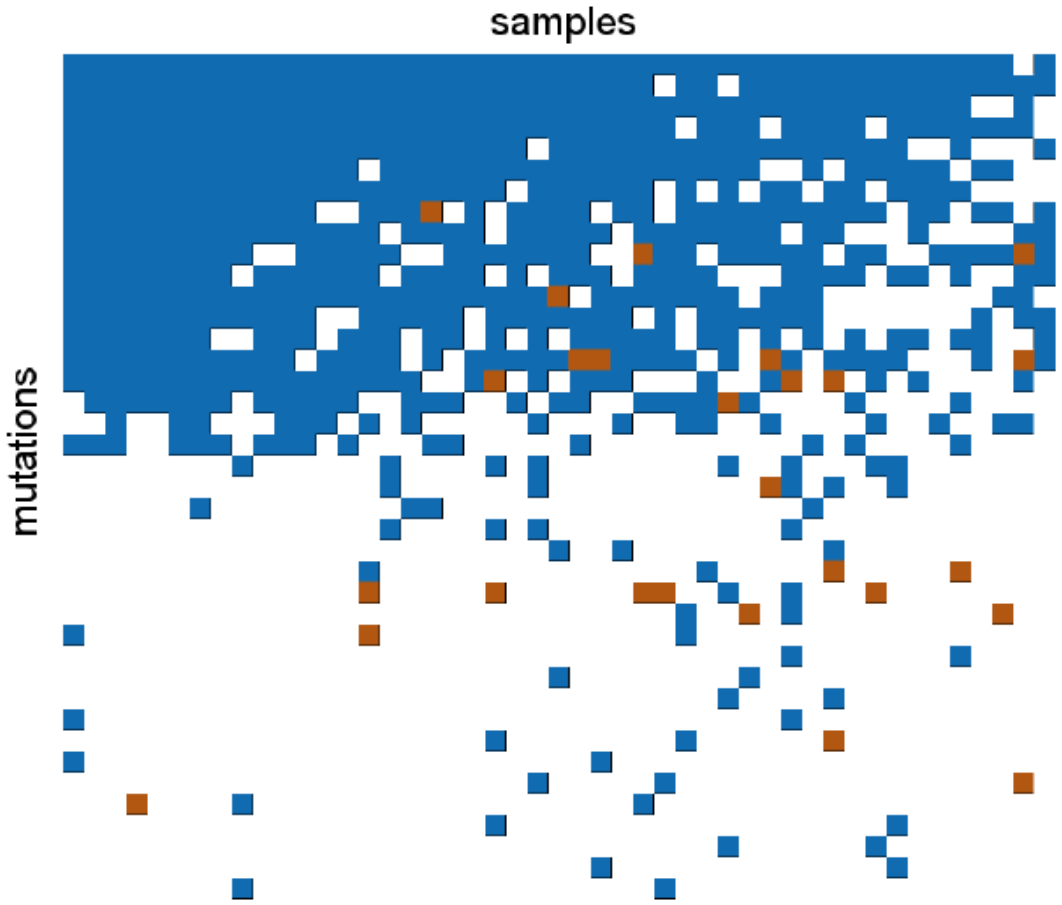


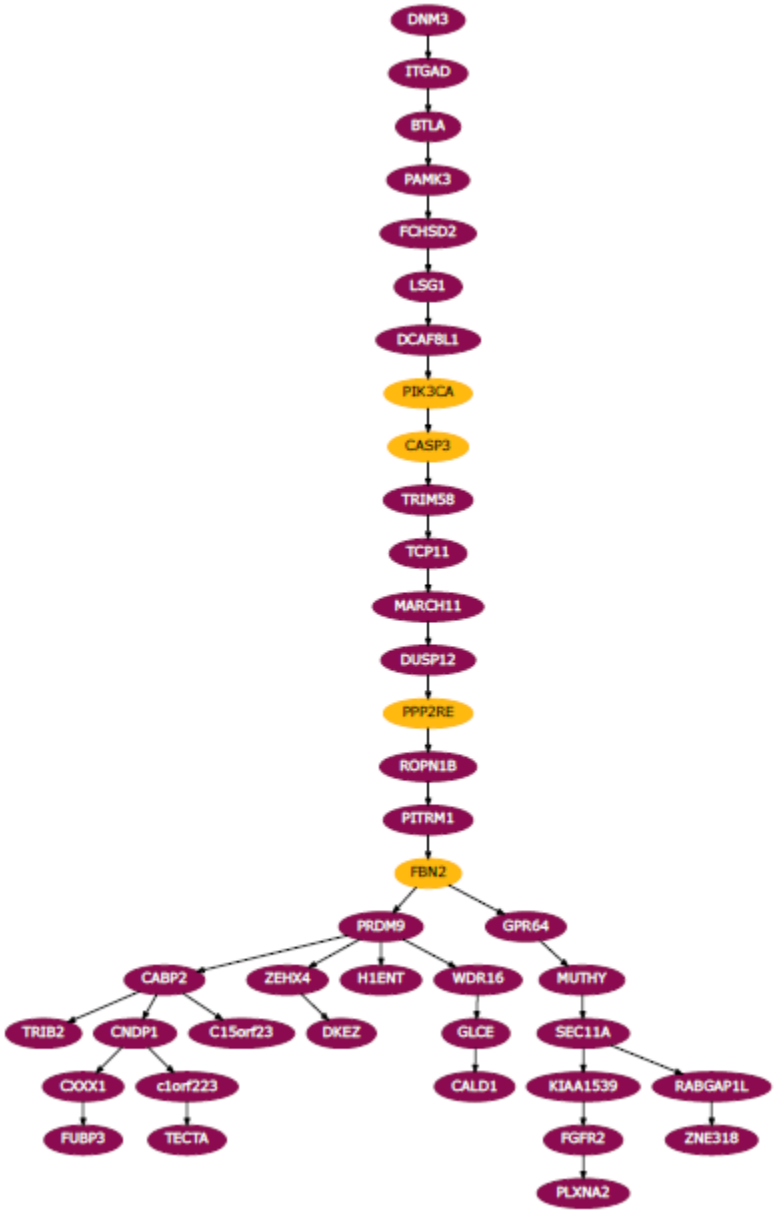

DNAJC17
ABCB5
DLEC1
PDE4DIP
SESN2
NTRK1
TOP1MT
ST13
ANAPC1
ARHGAP5
USP32
FAM115C
FRG1
DMXL1
PABPC1
ASNS
MLL3
RETSAT

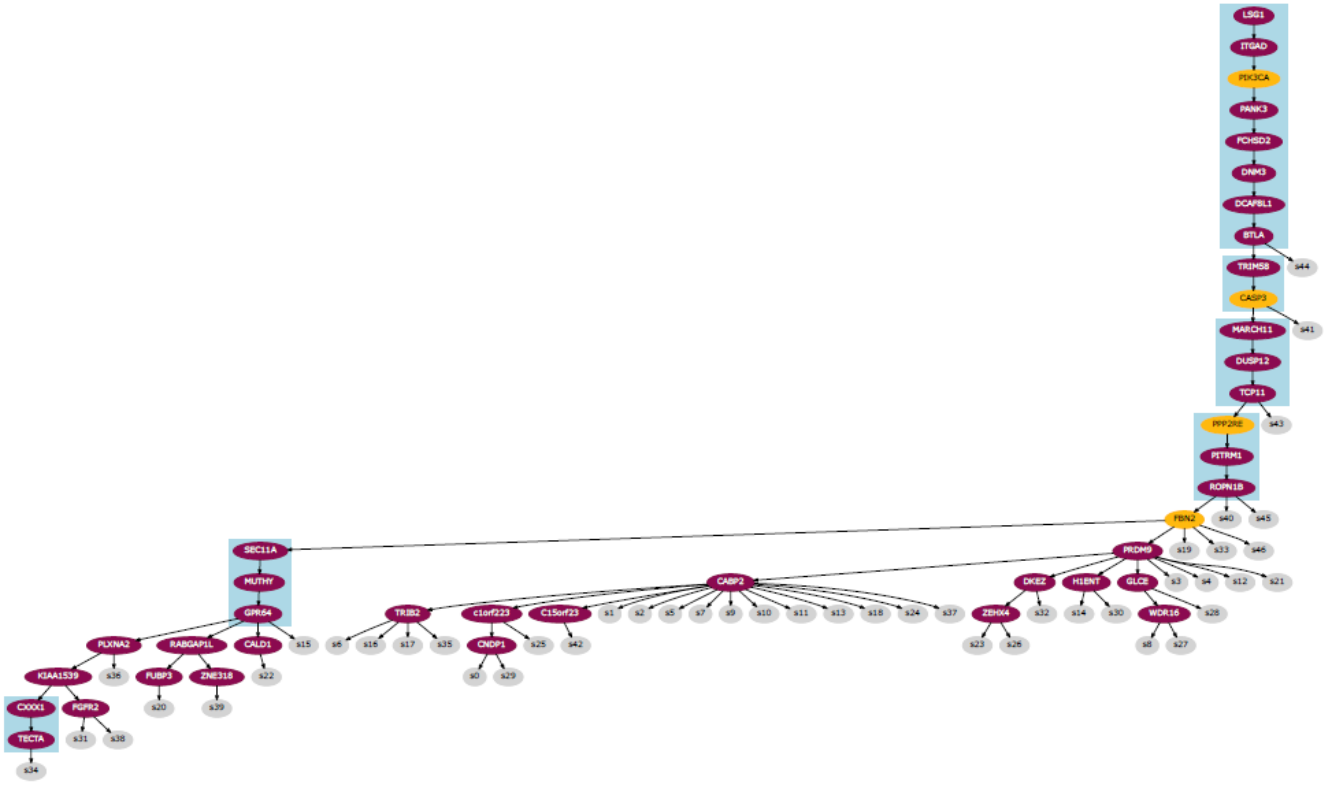

DNAJC17
ABCB5
DLEC1
s20
PDE4DIP
SESN2
s22
NTRK1
s41
s56
s15
s5
TOP1MT
s45
s1
ST13
s19
s31
s39
s50
ANAPC1
s44
ARHGAP5
s37
s7
s49
s40
s24
USP32
s9
FAM115C
DMXL1
FRG1
s53
ASNS
s0
s3
s8
s12
s35
s38
s6
s34
s48
s51
PABPC1
s13
s23
s18
s21
s52
MLL3
s11
s2
s47
s33
s32
RETSAT
s14
s16
s27
s29
s10
s57
s4
s17
s46
s54
s55
s25
s26
s28
s36
s42
s43
s30

Single-cell sequencing of an ER$^+$ breast tumor (Wang et al., Nature 2014)

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

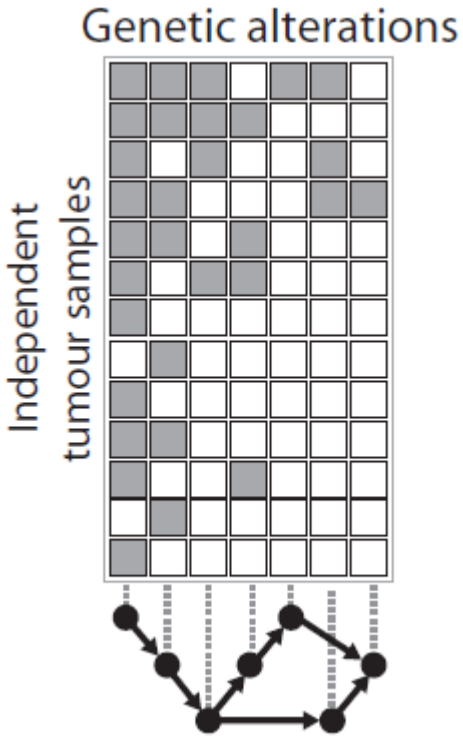



DNM3
ITGAD
BTLA
PAMK3
FCHSD2
LSG1
DCAFBL1
PIK3CA
CASP3
TRIM58
TCP11
MARCH11
DUSP12
PPP2RE
ROPN1B
PITRM1
FBN2
PRDM9
GPR64
CABP2
ZEHX4
H1ENT
WDR16
MUTHY
TRIB2
CNDP1
C15orf23
DKEZ
GLCE
SEC11A
CXXX1
c1orf223
CALD1
KIAA1539
RABGAP1L
FUBP3
TECTA
FGFR2
ZNE318
PLXNA2



past
clonal
expansions

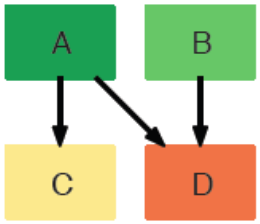
current,
co-existing
subclones



pathTiMEx: Mutually exclusive cancer pathways and their dependencies in tumor progression

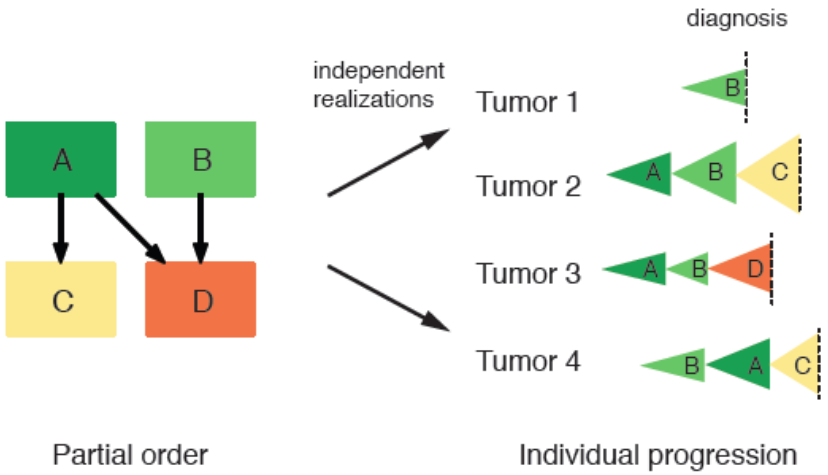
Simona Cristea, Jack Kuipers (RECOMB 2016)

Partial order among pathways

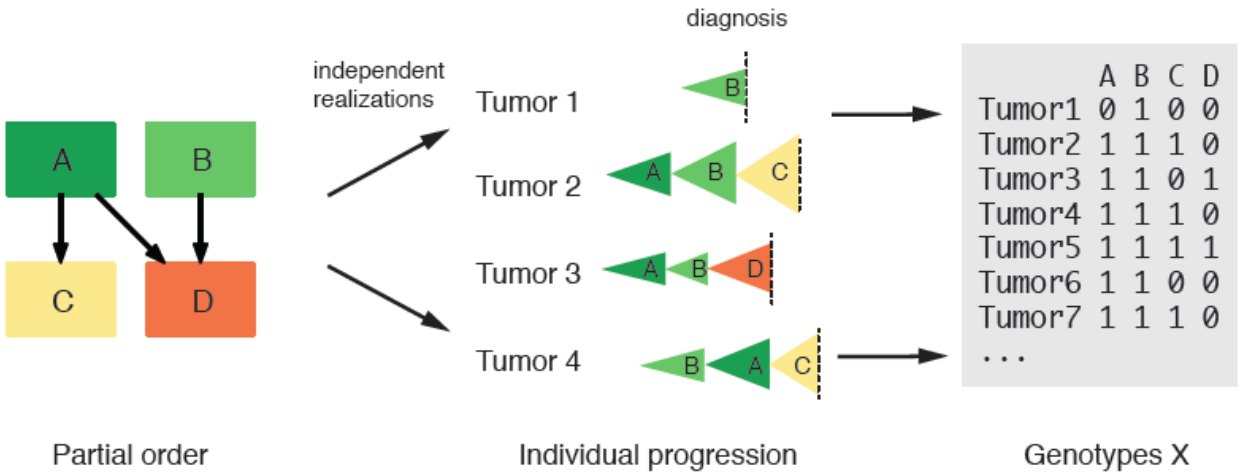


Partial order

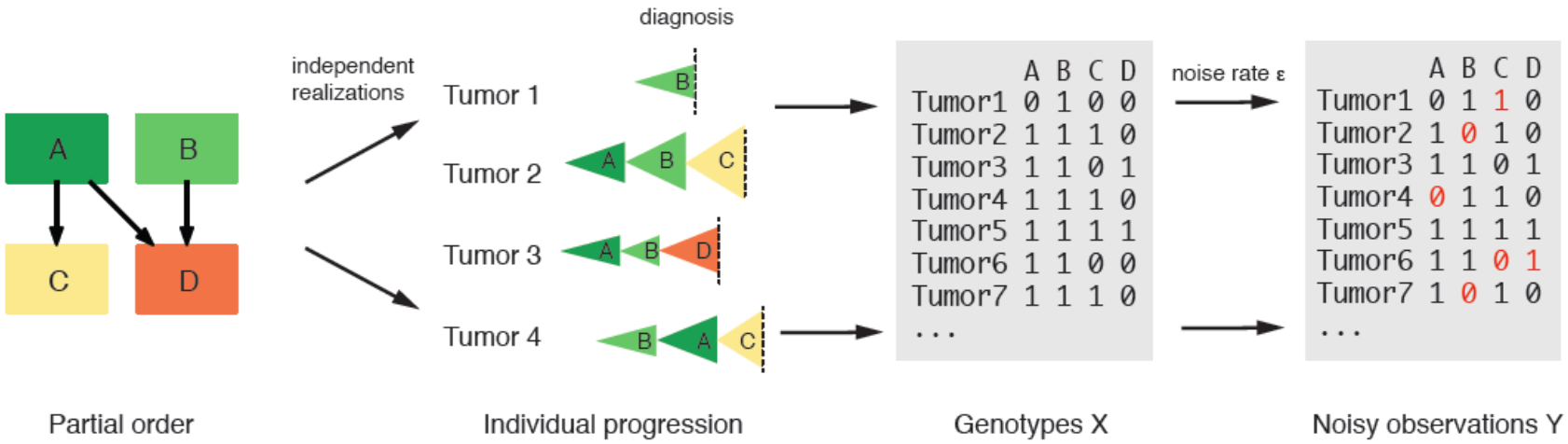
Individual progression



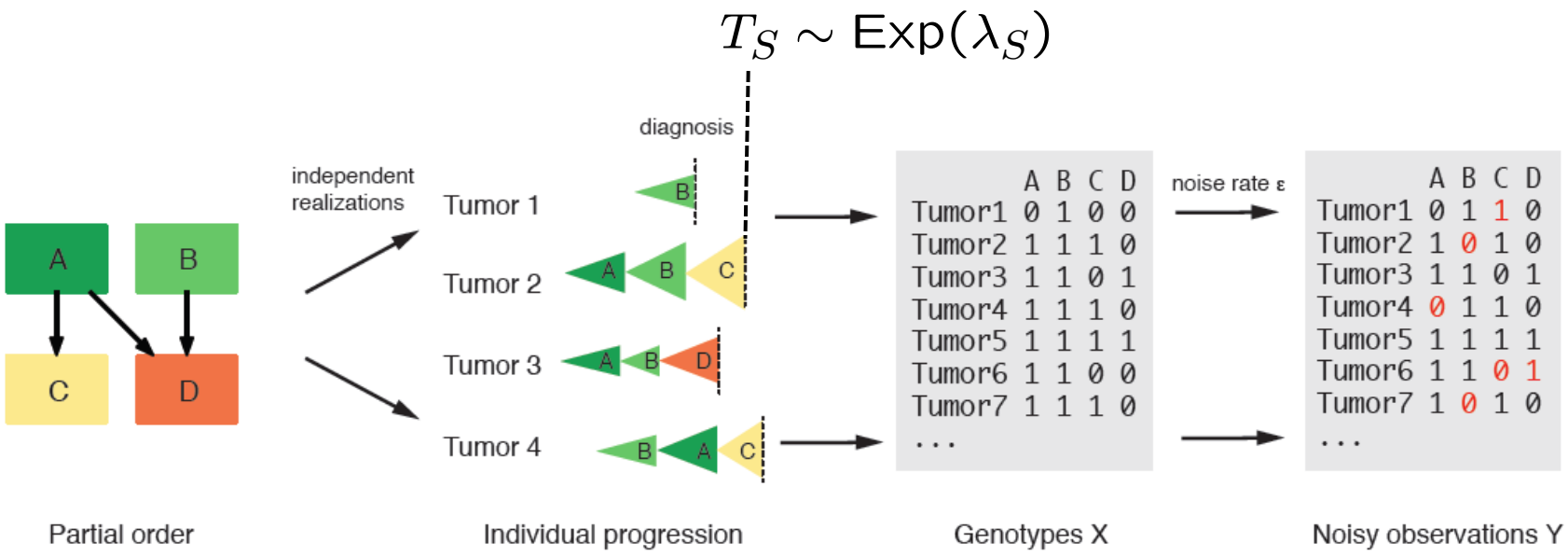
Genotypes at diagnosis



Observed genotypes



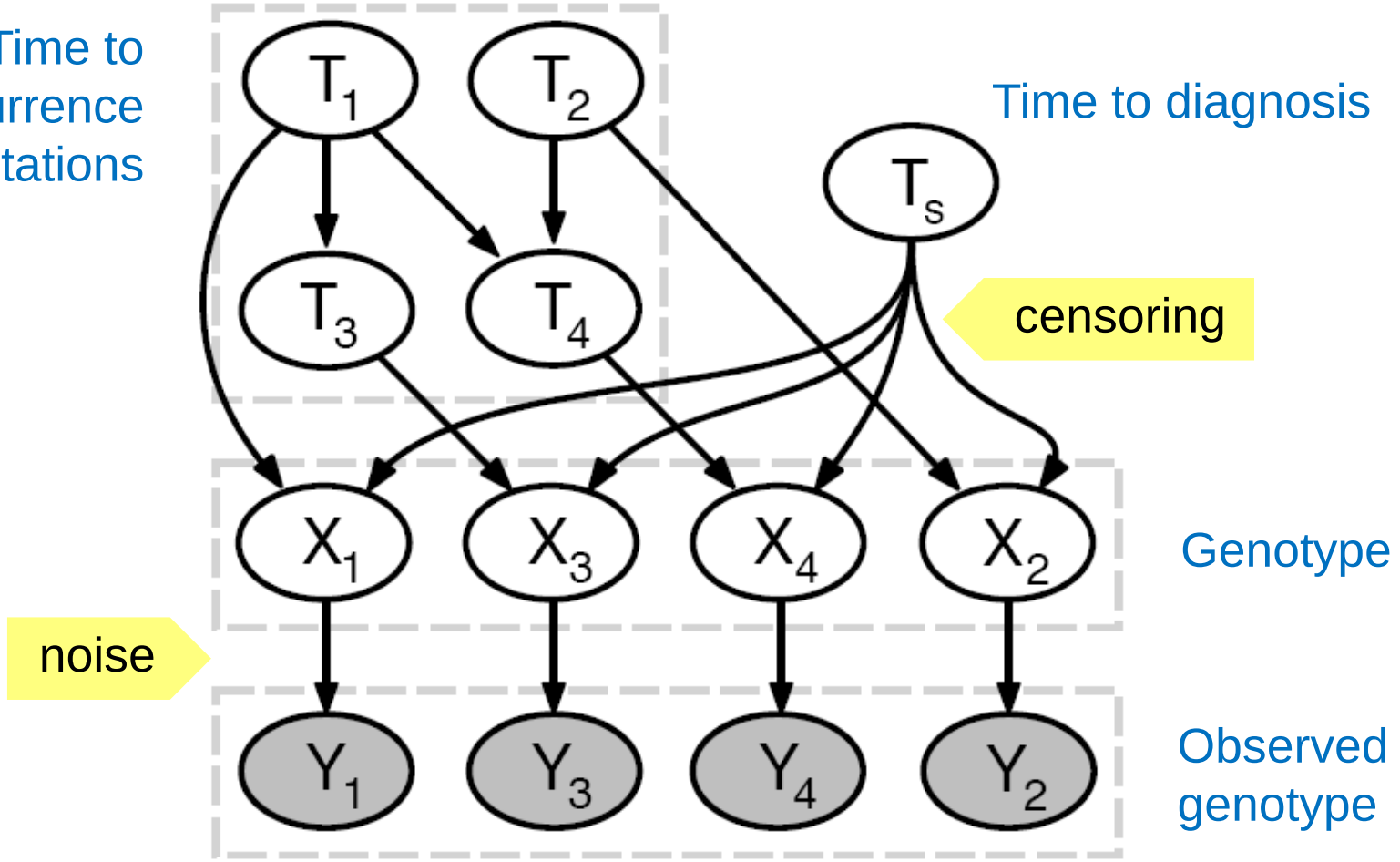
Waiting time distribution



$$
\begin{aligned}
T_A &\sim \text{Exp}(\lambda_A) \\
T_B &\sim \text{Exp}(\lambda_B) \\
T_C &\sim T_A + \text{Exp}(\lambda_C) \\
T_D &\sim \max(T_A, T_B) + \text{Exp}(\lambda_D)
\end{aligned}
$$

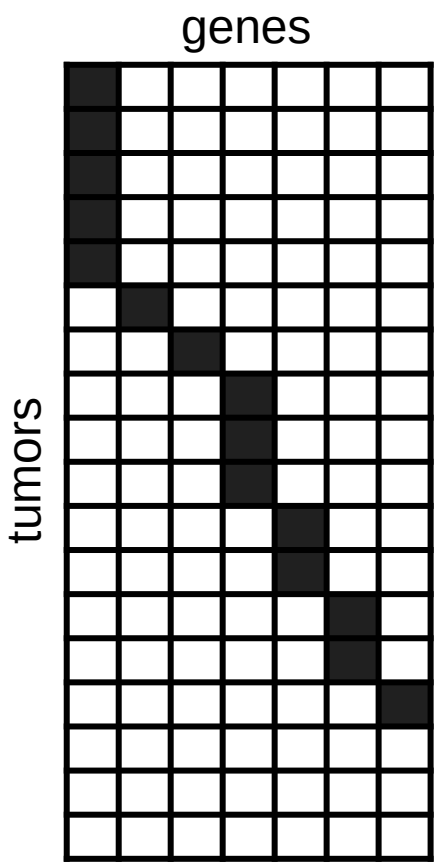
B. & Sullivant (2009), Gerstung et al (2012)

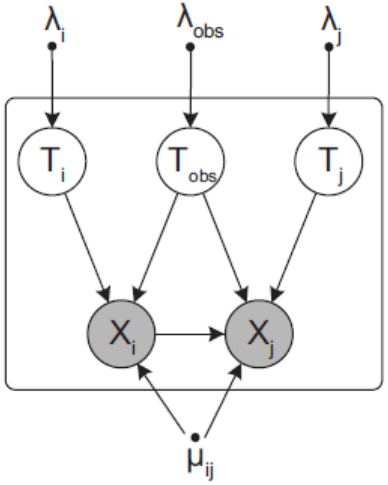
Hidden conjunctive Bayesian network (H-CBN)



Can we find pathways de novo?

- For example, as groups of mutually exclusive genes:



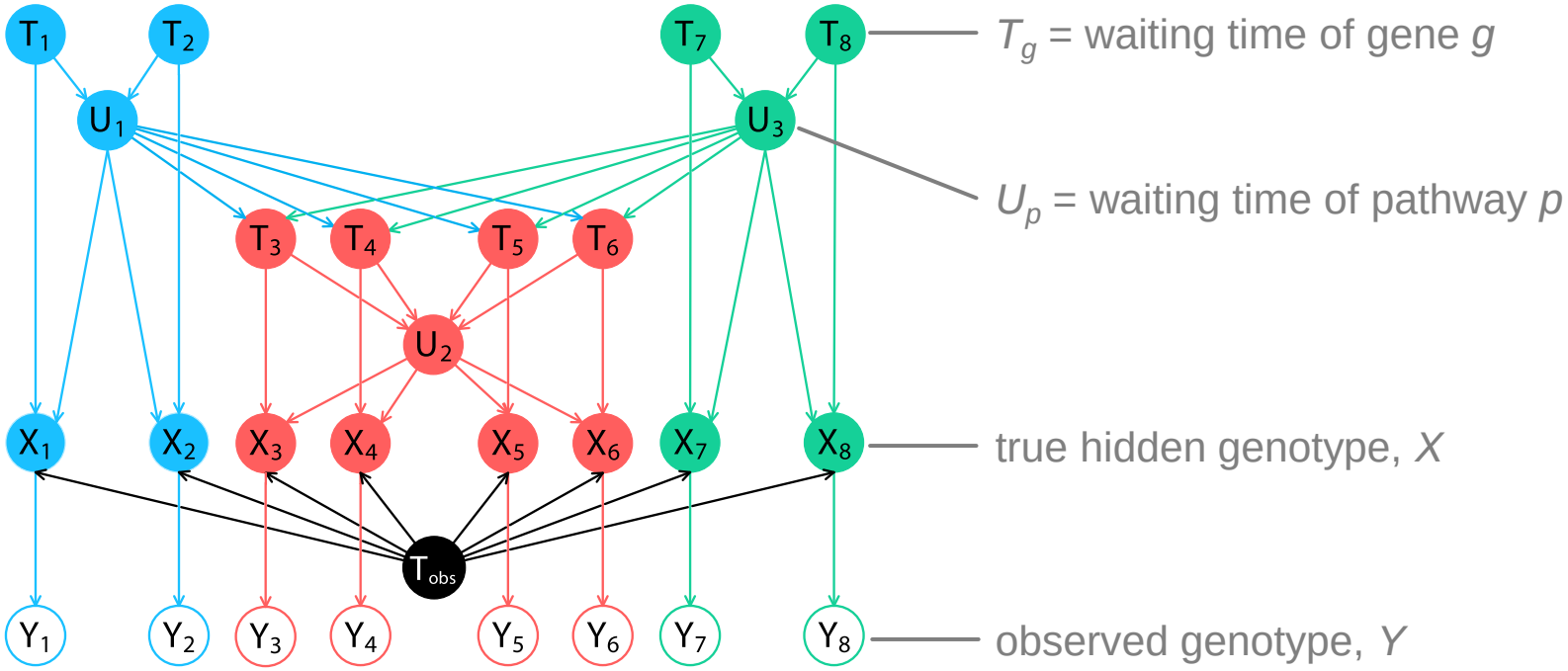


$$\mu_N P\left(T_K < \min_{i \in N \setminus K} (T_i, T_{\text{obs}})\right) + (1 - \mu_N) P\left(T_K \leq T_{\text{obs}} < \min_{i \in N \setminus K} T_i\right)$$

$$P\left(g_\emptyset \mid \theta_{\text{ME}}\right) = P\left(T_{\text{obs}} < \min_{i \in N} T_i\right)$$

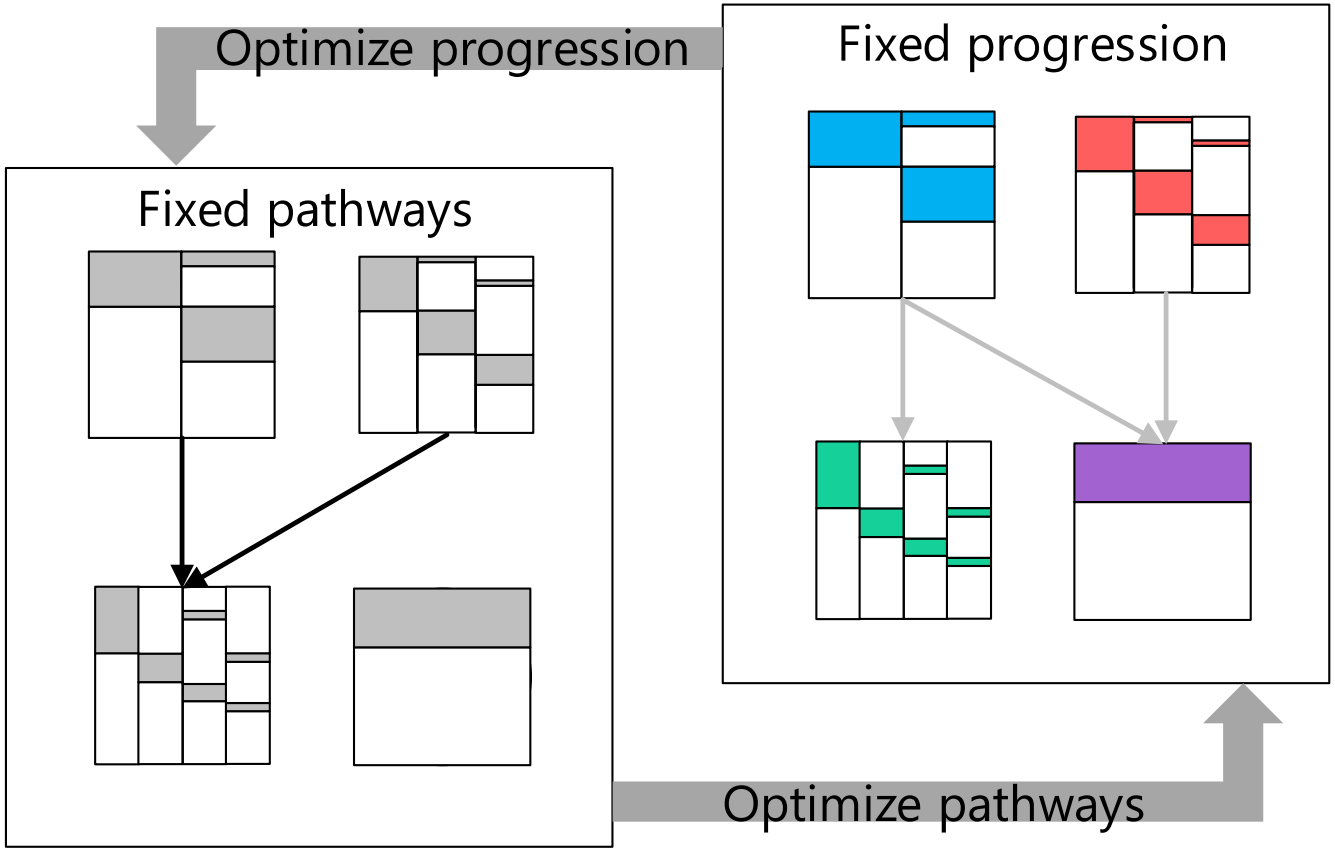
Constantinescu, Szczurek, et al. 2015

Pathways and their dependencies

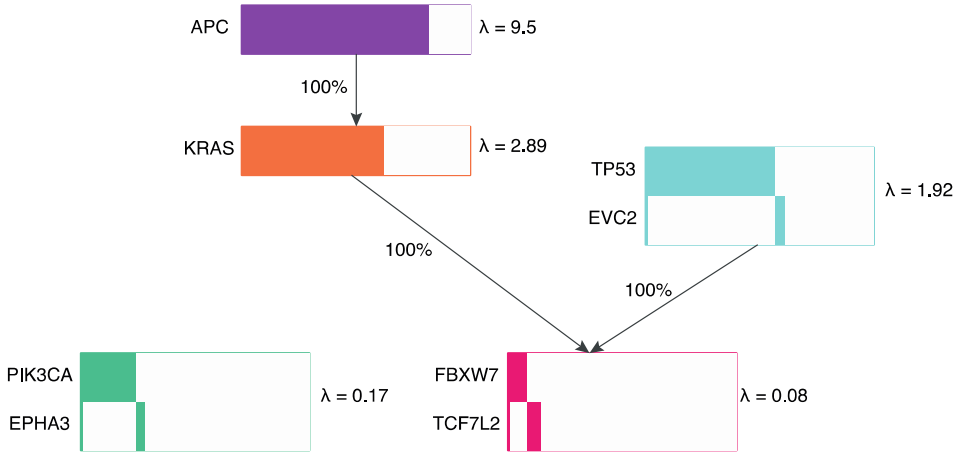


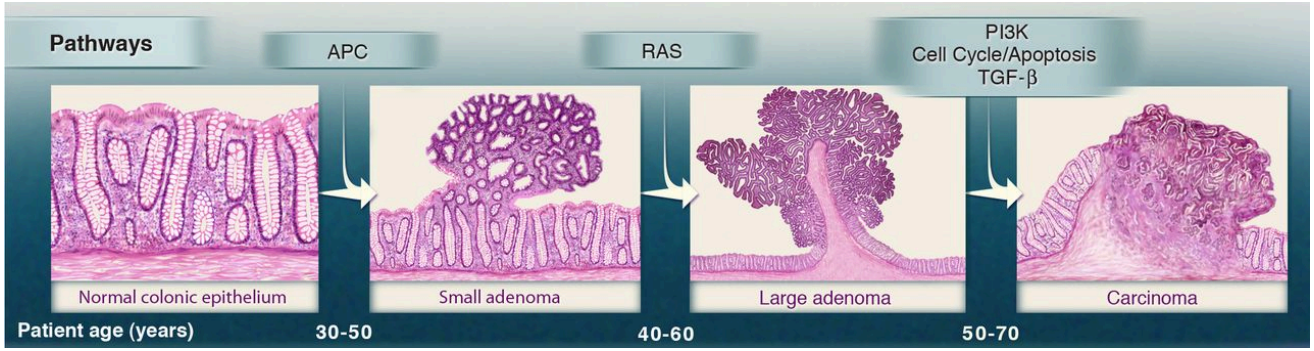
- A pathway is altered as soon as one of its genes is altered.
- Genes depend on upstream pathways.

Joint inference



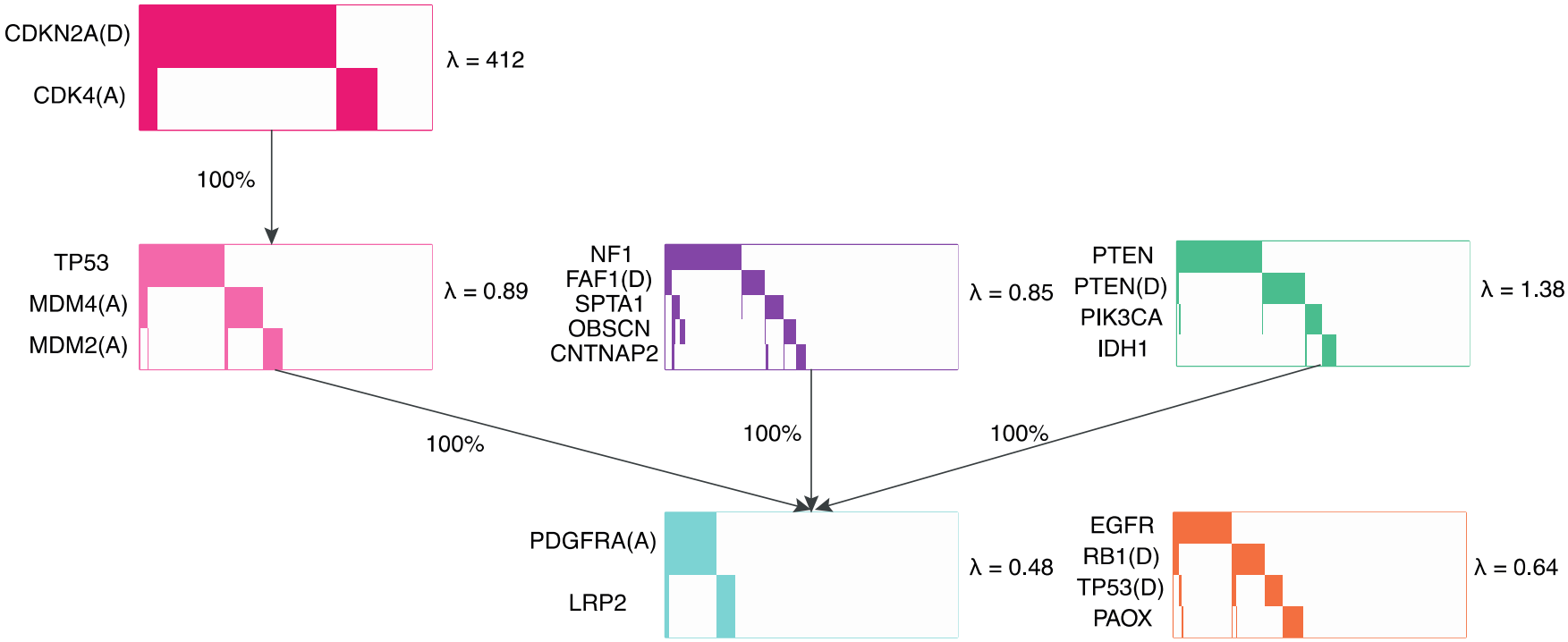
Progression in colorectal cancer





Vogelstein *et* al., Science 2013

Progression in glioblastoma



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