max planck institut informatik

Clinical Relevance of Genetic Tumor Progression Scores

Jörg Rahnenführer

Computational Biology and Applied Algorithmics
Max Planck Institute for Informatics
D-66123 Saarbrücken
Germany

gmds · LEIPZIG 2006

GMDS 2006, Leipzig
September 11, 2006

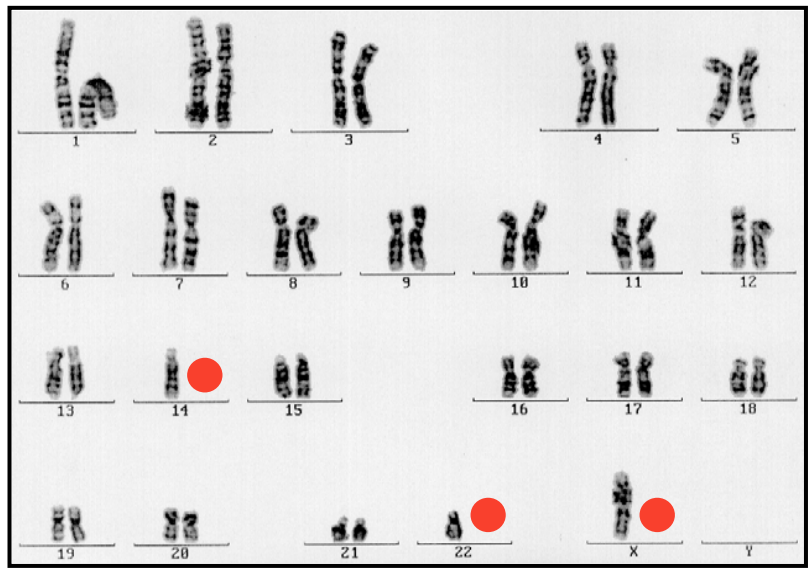
Outline

- Genetic tumor progression score
  - Oncogenetic trees
  - Estimation of tree mixture models with EM algorithm
  - Definition of the GPS (Genetic Progression Score)
- Clinical relevance in cancer
  - GPS as a marker for human cancer data
    - Prostate cancer: Time until PSA relapse
    - Glioblastoma: Time until death
    - Meningioma: Time until recurrence
  - Intratumoral cytogenetic heterogeneity
  - Extension to genome-wide measurements (arrayCGH data)

Evolutionary Pathways

- Consider as genetic events chromosomal alterations (gains/losses) in tumor cells
- Data
  - CGH measurements
  - arrayCGH measurements
- Goals
  - Reconstruction of the preferred order in which events occur
  - Assignment of measure of progression to a tumor based on estimated order

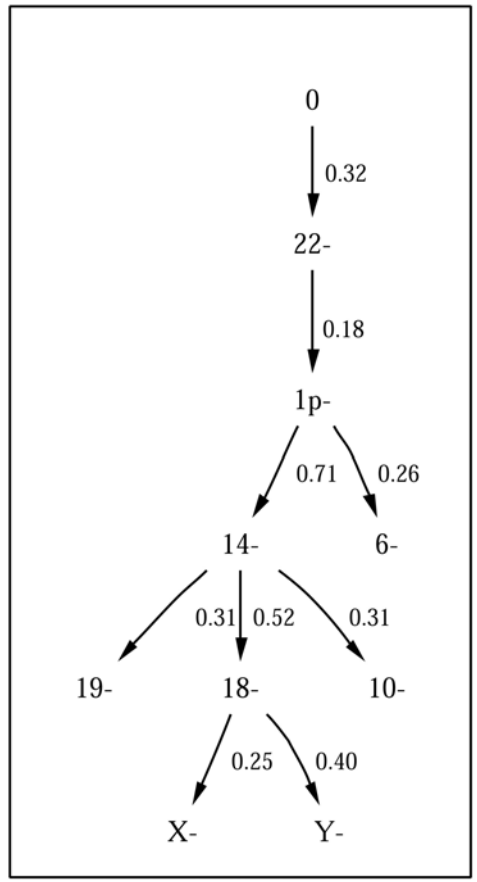
Analysis of tumor cells from meningioma (benign brain cancer)



X-	Y-	22-	19-	18-	14-	10-	6-	1p-
1	0	1	0	0	1	0	0	0
0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	0	1	0	0	1	0	1	1
...								

Evolutionary Pathways

- Consider as genetic events chromosomal alterations (gains/losses) in tumor cells
- Data
  - CGH measurements
  - arrayCGH measurements
- Goals
  - Reconstruction of the preferred order in which events occur
  - Assignment of measure of progression to a tumor based on estimated order



Estimated model for carcinogenesis in meningioma

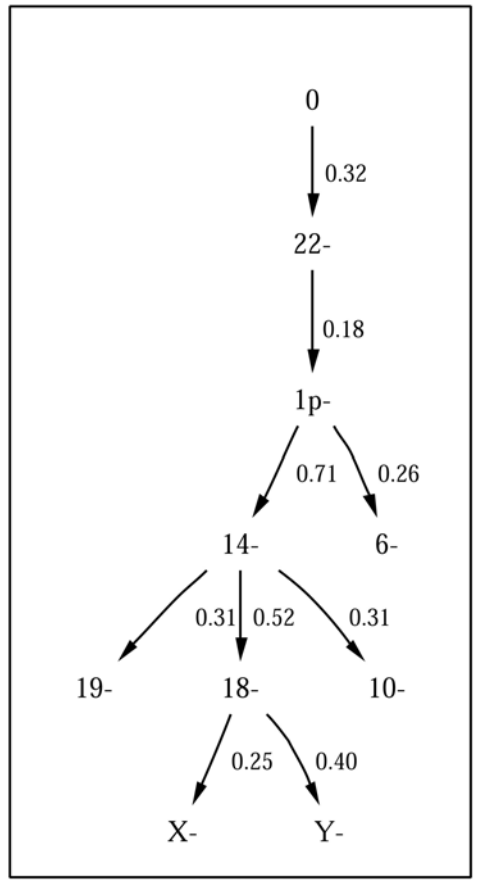
Evolutionary Pathways

Tree estimation

- Add root vertex *0* representing the null event: $P(X_0) = 1$.
- Start with complete graph G on $\ell+1$ vertices and weigh edge $e=(u,v)$ with

$$w(u,v) = \log\left(\frac{p(u,v)}{p(u)\,p(v)} \cdot \frac{p(u)}{p(u)+p(v)} \right)$$

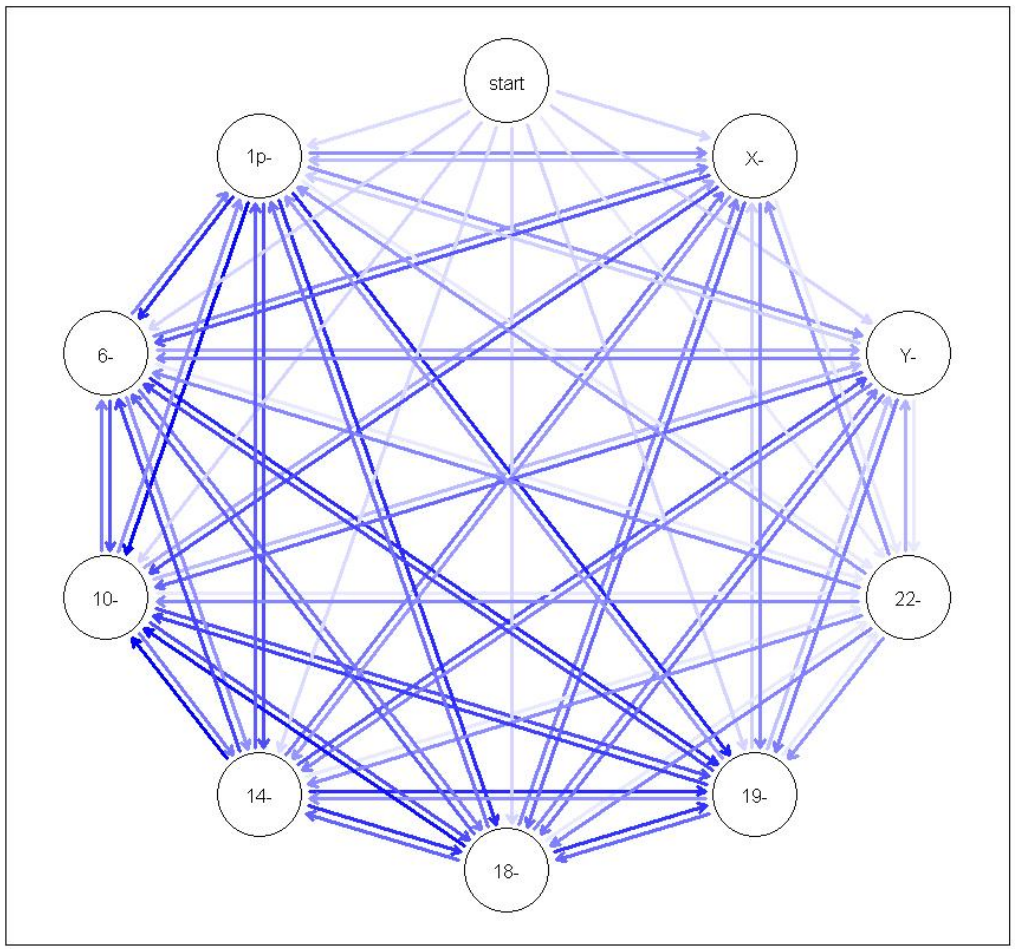
- Find maximum weight branching in complete graph G (with Edmond's branching algorithm).



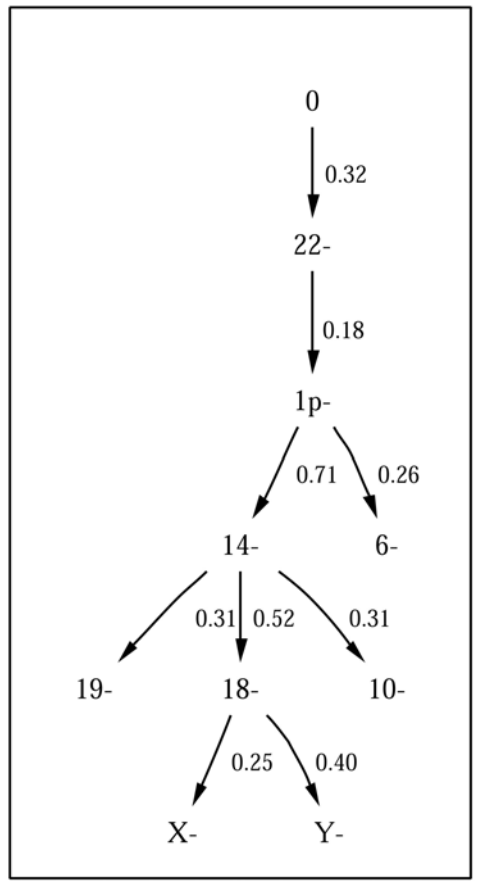
Estimated model for carcinogenesis in meningioma

Evolutionary Pathways

Weights: $w(u,v) = \log\left(\frac{p(u,v)}{p(u)\,p(v)} \cdot \frac{p(u)}{p(u)+p(v)}\right)$



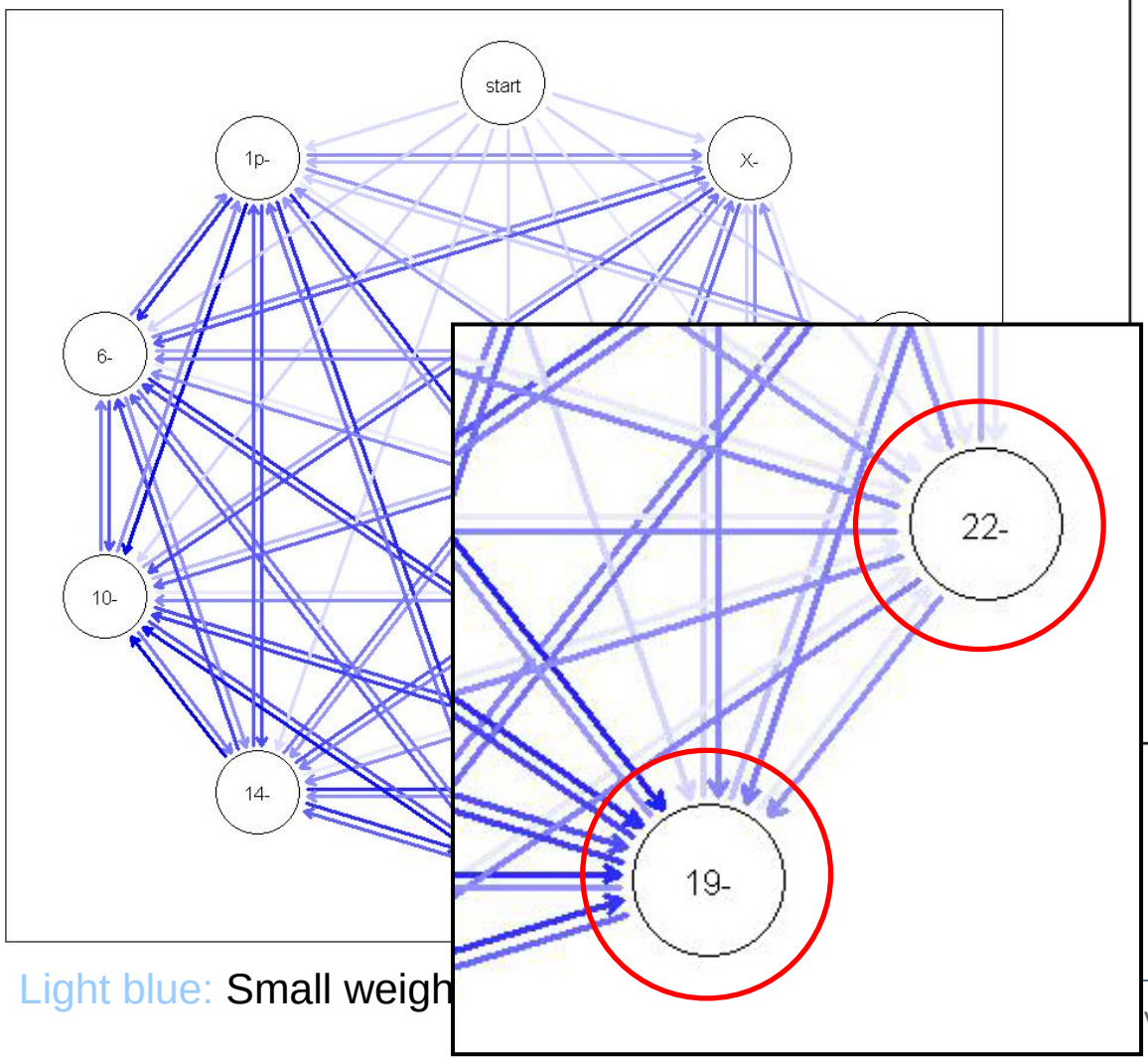
Light blue: Small weight Dark blue: Large weight

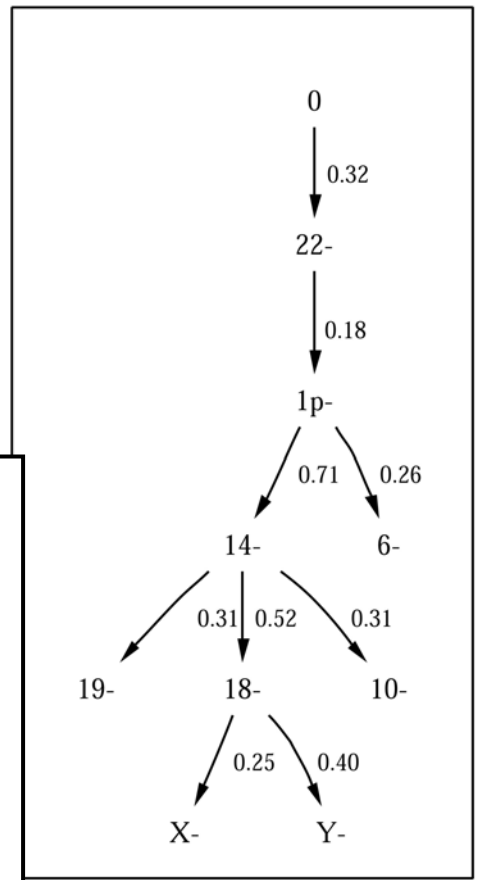


Estimated model for carcinogenesis in meningioma

Evolutionary Pathways

Weights: $w(u,v) = \log\left(\frac{p(u,v)}{p(u)\,p(v)} \cdot \frac{p(u)}{p(u)+p(v)}\right)$





Estimated model for carcinogenesis in meningioma

Light blue: Small weigh

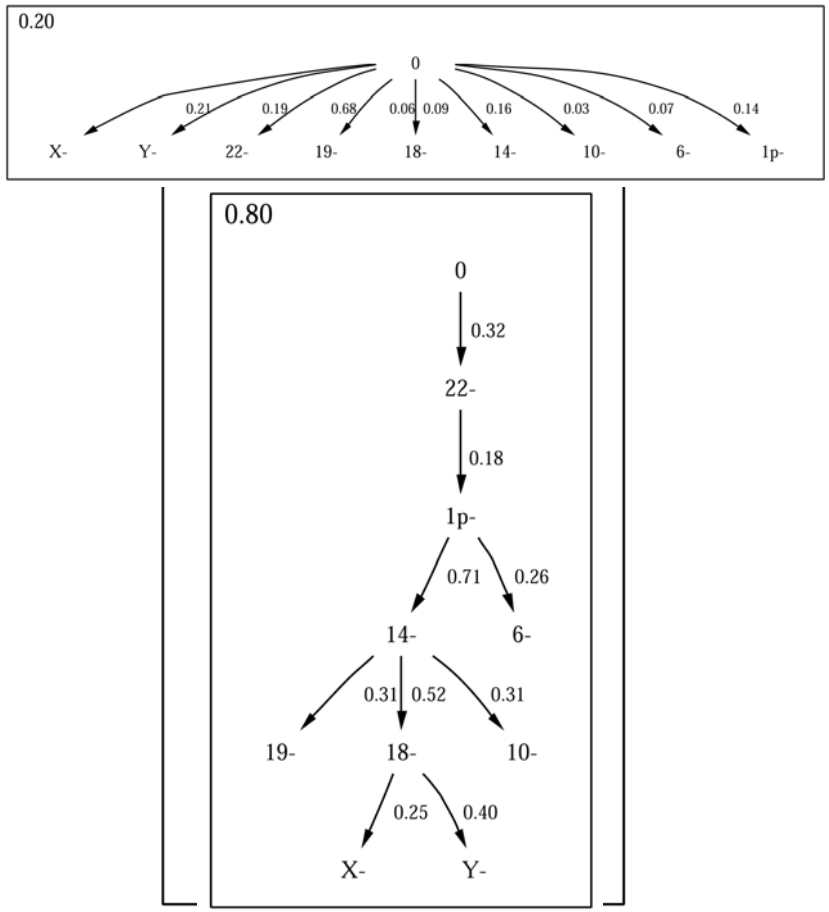
Evolutionary Pathways

- Problem of a single tree:
  - Many subsets of events (patterns) are not represented.
  - In a statistical framework such patterns have likelihood 0.
- Solution:
  - Consider K trees $T_1,\ldots,T_K$
  - Let T_1 be a star representing the noise component.
  - Define tree mixture model as

$$M = \sum_{k=1}^{K} \alpha_k T_k \quad \left(\sum_{k=1}^{K} \alpha_k = 1\right)$$

  - Likelihood of a sample is given by

$$L(x \mid M) = \sum_{k=1}^{K} \alpha_k L(x \mid T_k)$$

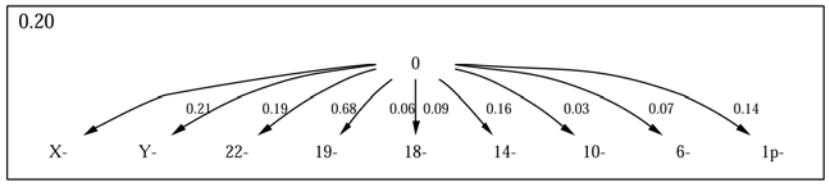


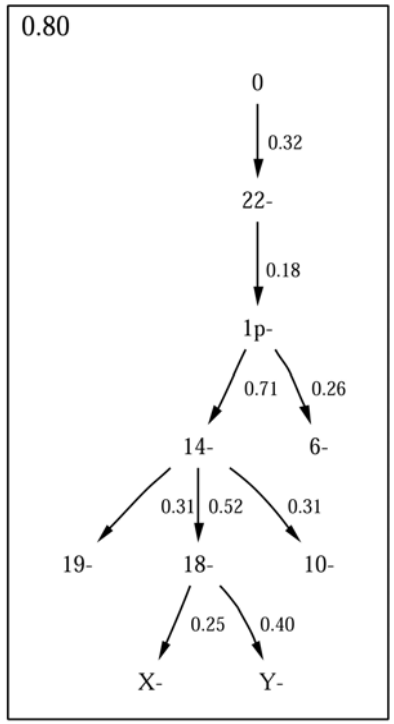
Estimated model for carcinogenesis in meningioma

Evolutionary Pathways

EM-like learning algorithm for estimating mixture model *M*:

1. Initial solution:
   Group samples with k-means clustering algorithm.
2. M-step: Update parameters:
   Estimate pairwise probabilities within tree components and reconstruct trees T_k.
3. E-step: Compute responsibilities:
   Assign samples to tree mixture model components.
4. Iterate steps 2 and 3 until convergence.





Estimated model for carcinogenesis in meningioma

Genetic Progression Score

- Idea: Replace conditional probabilities on tree edges by expected waiting times.
- Add waiting times on edges to obtain waiting time W of a genetic patterns *x*.
- The genetic progression score (GPS) is defined by

$$\mathrm{GPS}(x) = E_T(W(x))$$

where the expectation is taken with respect to the underlying tree model.

- Let $T_i \sim \exp(\lambda_i)$ be the waiting time for event *i* given that *pa(i)* has occurred.
- Let $T_s \sim \exp(\lambda_S)$ be the random sampling time (age) of the tumor.
- Age of tumor unknown, thus set $1/\lambda_S = E(T_s) = 1$.
- Then the expected waiting time for event *i* is given by

$$p_i = \frac{\lambda_i}{\lambda_i + \lambda_S}$$

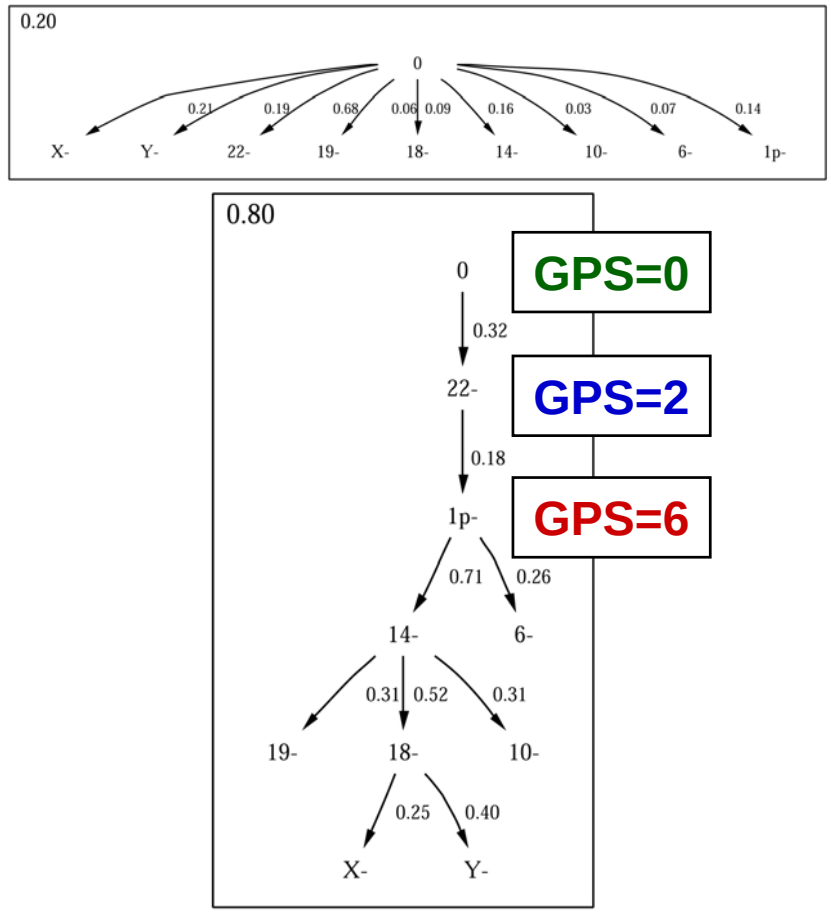
$$\Rightarrow \quad E(T_i) = \frac{1}{\lambda_i} = \frac{1-p_i}{p_i}\lambda_S = \frac{1-p_i}{p_i}$$

Genetic Progression Score

- Idea: Replace conditional probabilities on tree edges by expected waiting times.
- Add waiting times on edges to obtain waiting time W of a genetic patterns *x*.
- The genetic progression score (GPS) is defined by

$$\mathrm{GPS}(x) = E_T(W(x))$$

where the expectation is taken with respect to the underlying tree model.



Estimated model for carcinogenesis in meningioma

Clinical Relevance

Prostate cancer

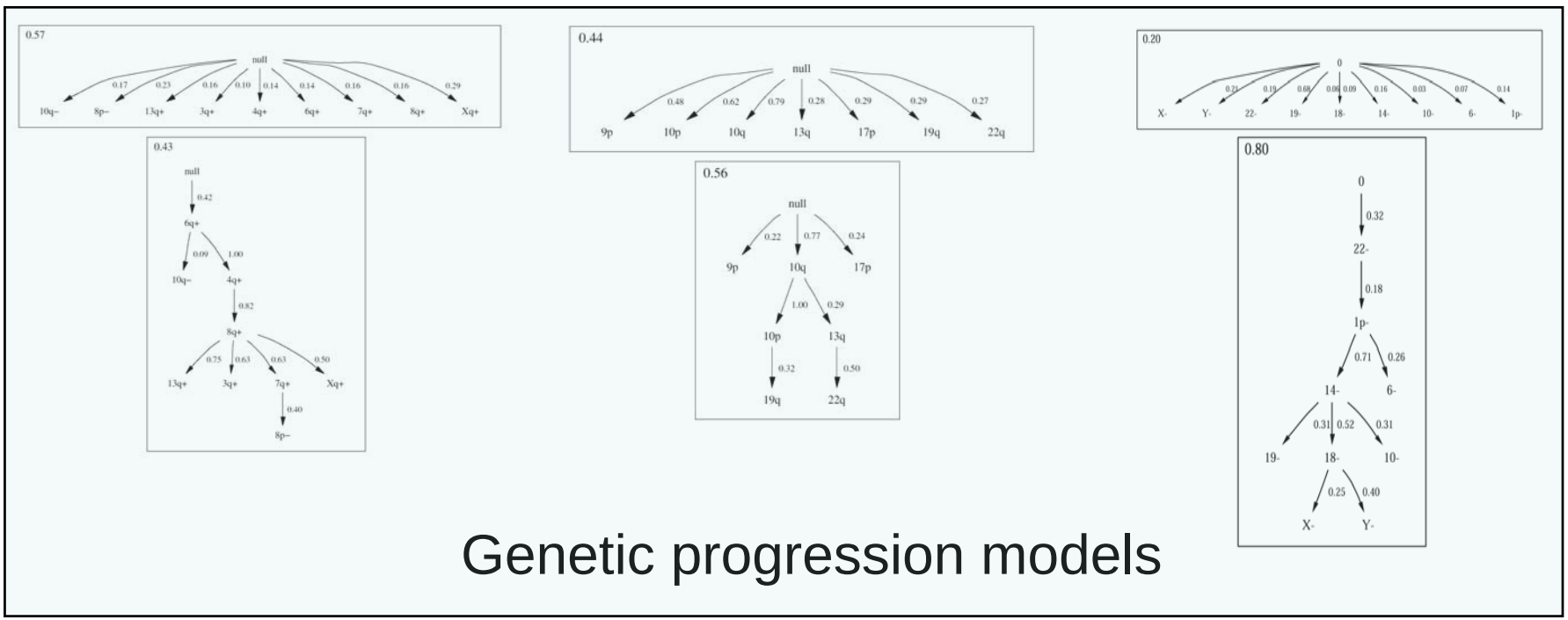
- Time until PSA relapse (*n=54*, *30* with *Gleason=7*)

Glioblastoma

- Survival time (*n=75* patients)

Meningioma

- Time until recurrence (*n=661* patients)

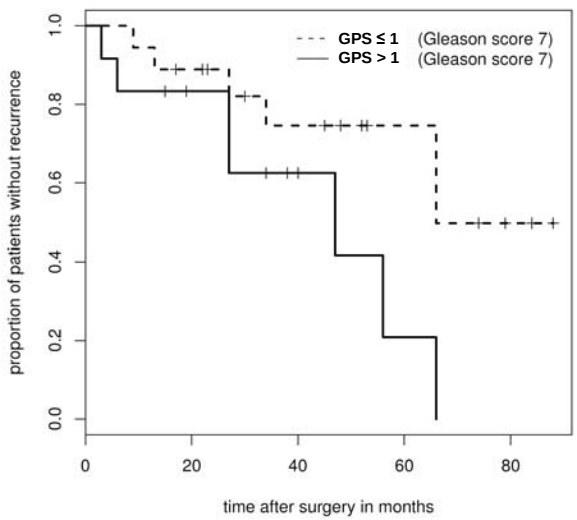


Genetic progression models

Clinical Relevance

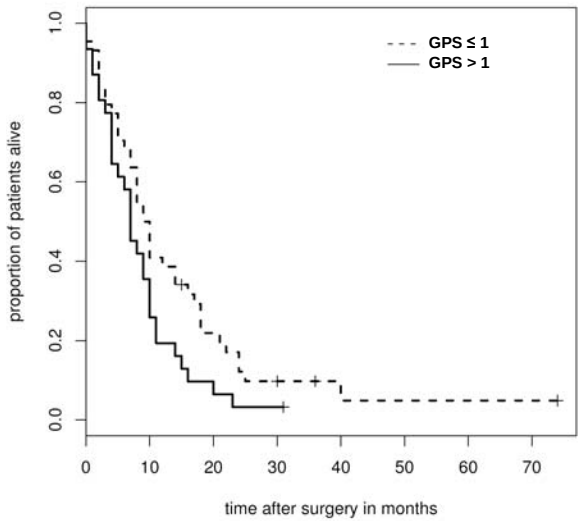
Prostate cancer

- Time until PSA relapse (*n=54*, *30* with *Gleason=7*)
- Significant shorter time for group with *GPS>1* (*p=0.040*).



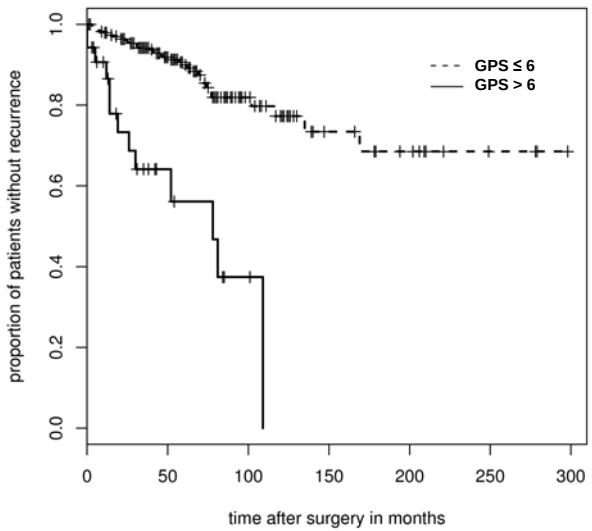
Glioblastoma

- Survival time (*n=75* patients)
- Significant shorter time for group with *GPS>1* (*p=0.015*).



Meningioma

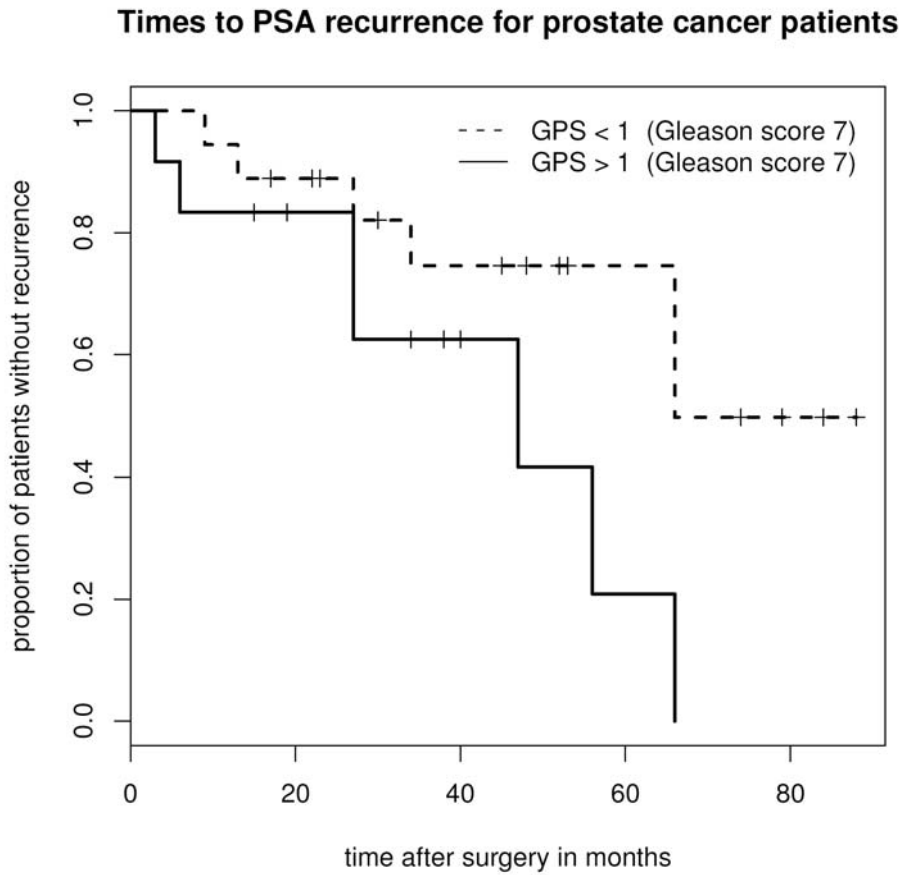
- Time until recurrence (*n=661* patients)
- Significant shorter time for group with *GPS>6* (*p<10^{-6}*).



Clinical Relevance

Prostate cancer

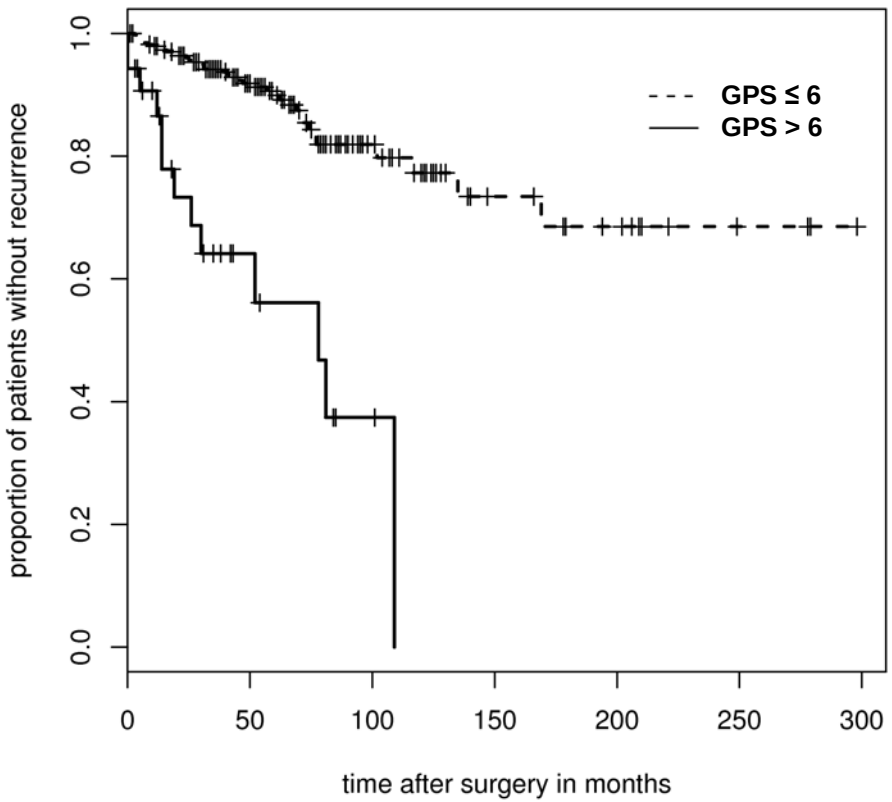
- Analysis of time to recurrence (54 patients)
- Gleason score is widely used clinical grading system – most tumors are scored with a value of Gleason=7.
- We split patients into *2* groups according to average GPS.
- Significant longer time to PSA recurrence for group with GPS<1 with p-value *p=0.040.*
- GPS improves diagnostics.



Clinical Relevance

Meningioma

- Analysis of time to PSA recurrence (661 patients)
- We split patients into *2* groups according to average GPS.
- Significant longer survival time for group with GPS<6 with p-value $p<10^{-6}$.
- Significant correlation between GPS and
  - time to recurrence: $p<10^{-6}$
  - WHO grade: $p<10^{-10}$
  - tumor location: $p<10^{-8}$



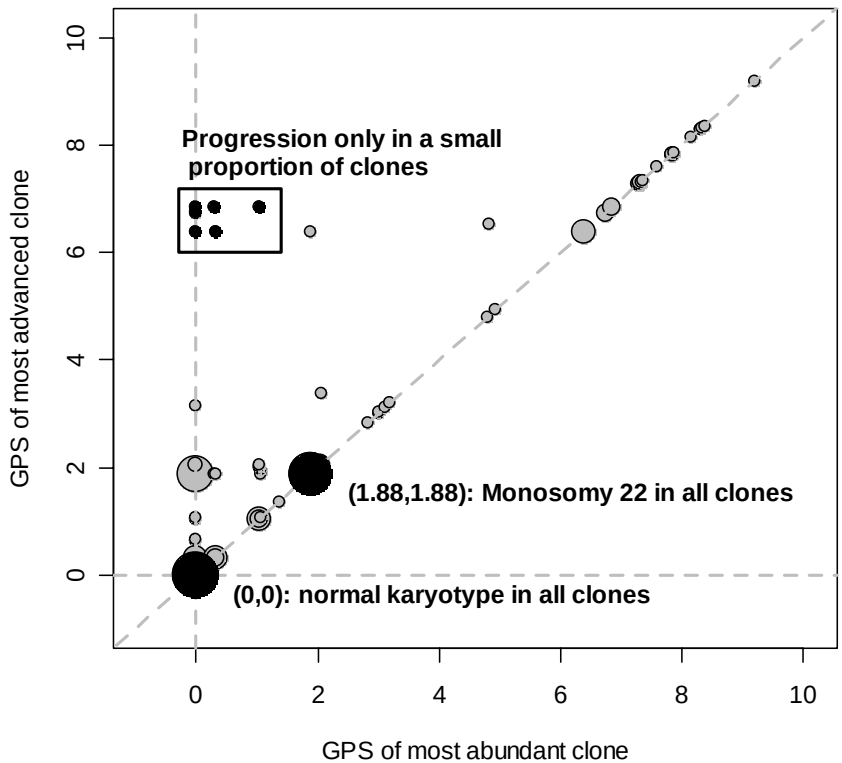
Intratumoral Heterogeneity

Meningioma

- Analysis of time to PSA recurrence (661 patients)
- Intratumoral heterogeneity observed in 221 out of 661 patients (33.4%).
- Number of different genetic patterns detected in single tumor between 1 and 10.
- GPS of most advanced and most abundant clone different in 120 tumors (18.2%), almost always with GPS=0 for most abundant clone.

- Analysis of genetic progression in single tumor cells

Size of circle indicates number of tumors with respective combination

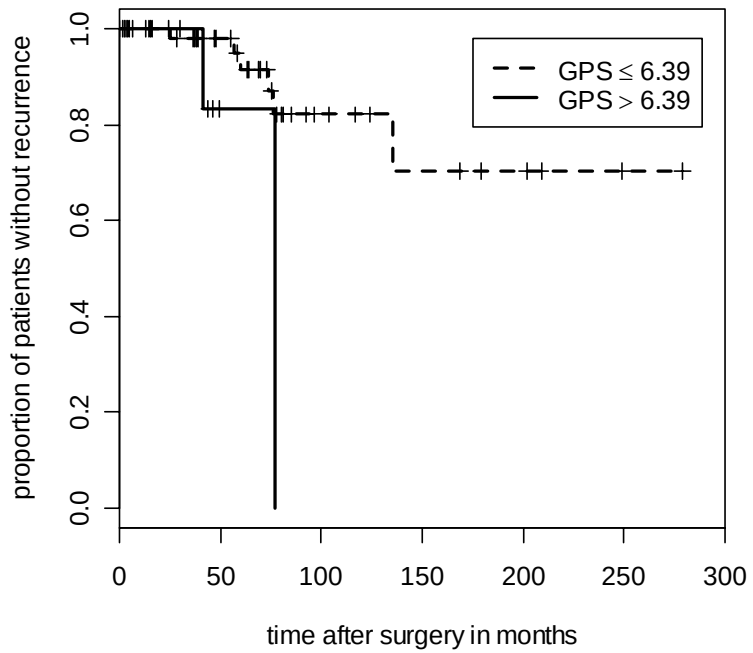


Intratumoral Heterogeneity

Meningioma

- Analysis of time to PSA recurrence (661 patients)
- Intratumoral heterogeneity observed in 221 out of 661 patients (33.4%).
- Number of different genetic patterns detected in single tumor between 1 and 10.
- GPS of most advanced and most abundant clone different in 120 tumors (18.2%), almost always with GPS=0 for most abundant clone.

- Restrict to patients with different progression in single cells.
- Significant longer survival time for group with GPS<6 with p-value *p=0.05.*

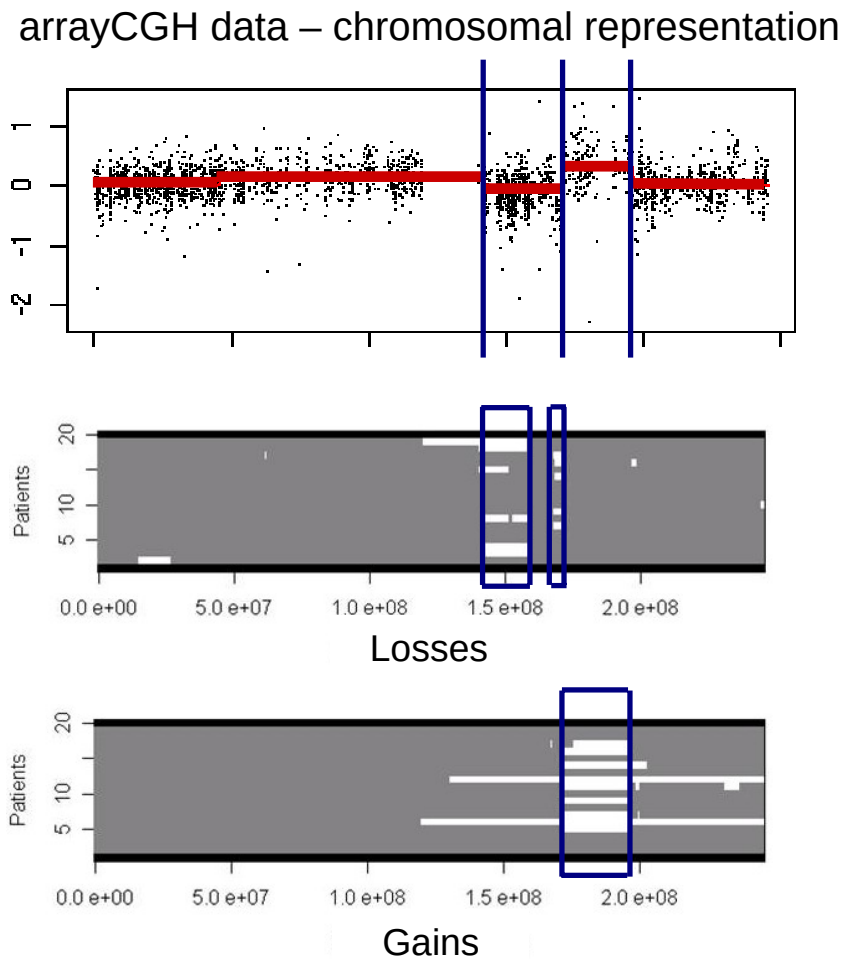


ArrayCGH data

- Analyze data on gene level (~30,000 events)

Analysis pipeline:

- Regions with constant copy number estimated with GLAD algorithm (breakpoints on x-axis)
- Gained/lost regions determined with robust fitting of normals (cutoff on y-axis)
- Genetic events: Recurrent regions with significant number of counts



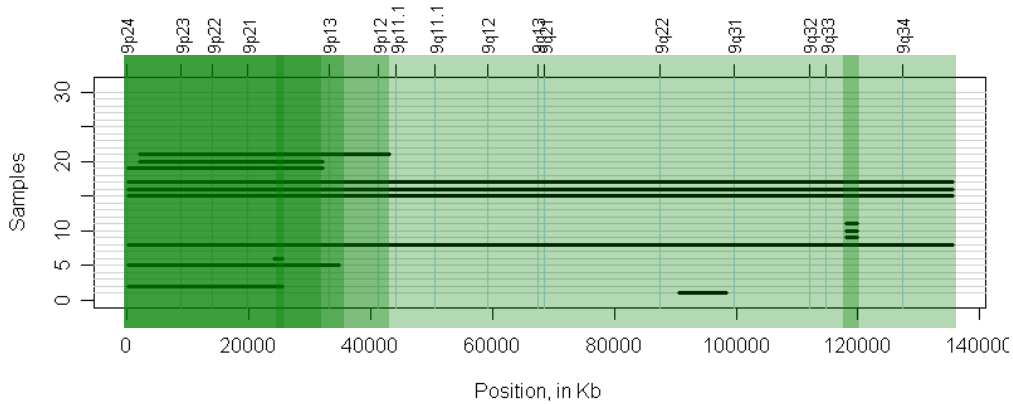
ArrayCGH data

- Analyze data on gene level (~30,000 events)

Analysis pipeline:

- Regions with constant copy number estimated with GLAD algorithm (breakpoints on x-axis)
- Gained/lost regions determined with robust fitting of normals (cutoff on y-axis)
- Genetic events: Recurrent regions with significant number of counts

- Problem: How to adjust breakpoints?

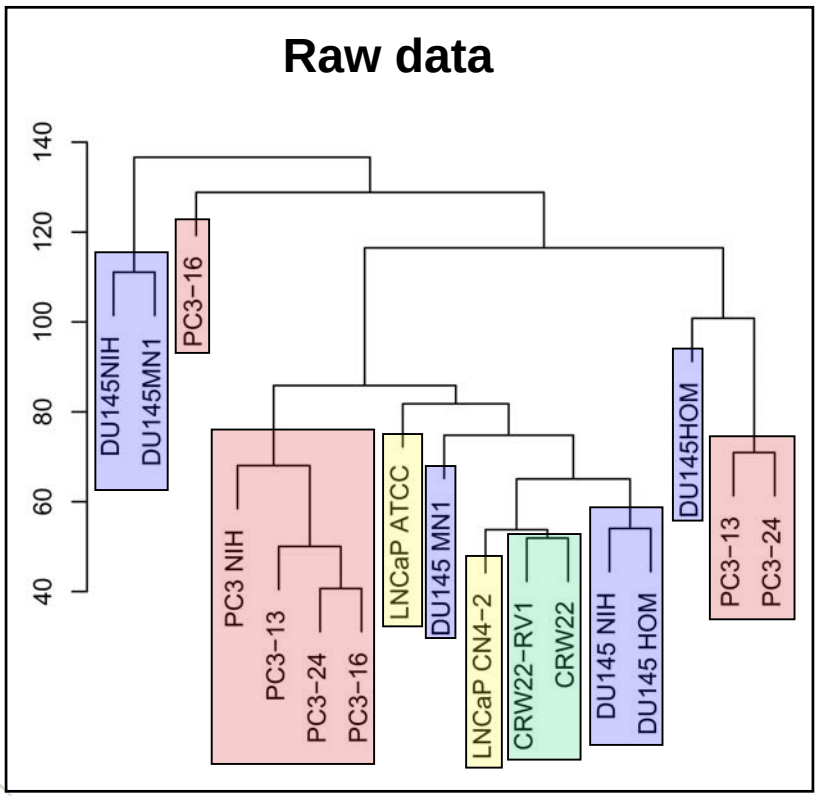


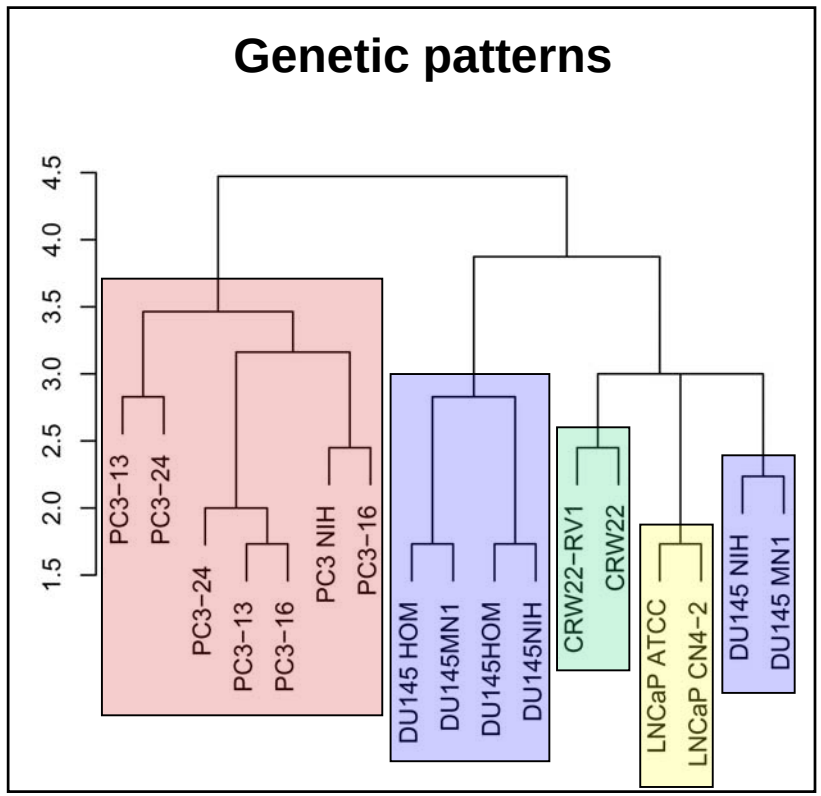
Losses along chromosome 9 in arrayCGH data obtained from a leukemia patient

- Ongoing research
  Select regions such that the correlation to clinical variables is maximized!

ArrayCGH data

- Clustering of prostate cancer cell lines:
  PC3 (7), DU145 (6), LNCaP (2), CRW22 (2)
- Comparison of results for raw data (Euclidean distance) and for binary genetic patterns (Manhattan distance)





Literature

- Niko Beerenwinkel, Jörg Rahnenführer, Martin Däumer, Daniel Hoffmann, Rolf Kaiser, Joachim Selbig, Thomas Lengauer: **Learning multiple evolutionary pathways from cross-sectional data**, *Journal of Computational Biology*, 12(6): 584-598, 2005.

- Jörg Rahnenführer, Niko Beerenwinkel, Wolfgang A. Schulz, Christian Hartmann, Andreas von Deimling, Bernd Wullich, Thomas Lengauer: **Estimating cancer survival and clinical outcome based on genetic tumor progression scores**, *Bioinformatics* 21(10): 2438-2446, 2005.

- Ralf Ketter, Steffi Urbschat, Wolfram Henn, Yoo-Jin Kim, Wolfgang Feiden, Niko Beerenwinkel, Thomas Lengauer, Wolf-Ingo Steudel, Klaus D. Zang, Jörg Rahnenführer: **Application of oncogenetic trees mixtures as a biostatistical model of the clonal evolution of tumors in the postoperative management of meningiomas**, *submitted*, 2006.

Acknowledgements

- Max-Planck-Institute for Informatics, Saarbrücken

  Prof. Thomas Lengauer

  Dr. Niko Beerenwinkel (now Harvard)

  Junming Yin
  Adrian Alexa
  Laura Tolosi

- Dept. of Urology and Pediatric Urology, University of the Saarland, Homburg

  Prof. Bernd Wullich, Dr. Volker Jung, Dr. Jörn Kamradt

- Institute of Human Genetics, University of the Saarland, Homburg

  Prof. Klaus Zang, Dr. Steffi Urbschat

- Department of Neurosurgery, University of the Saarland, Homburg

  Dr. Ralf Ketter

- Department of Urology, Heinrich-Heine-University, Düsseldorf

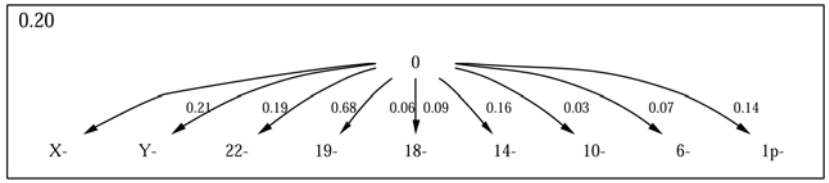
  Prof. Wolfgang A. Schulz

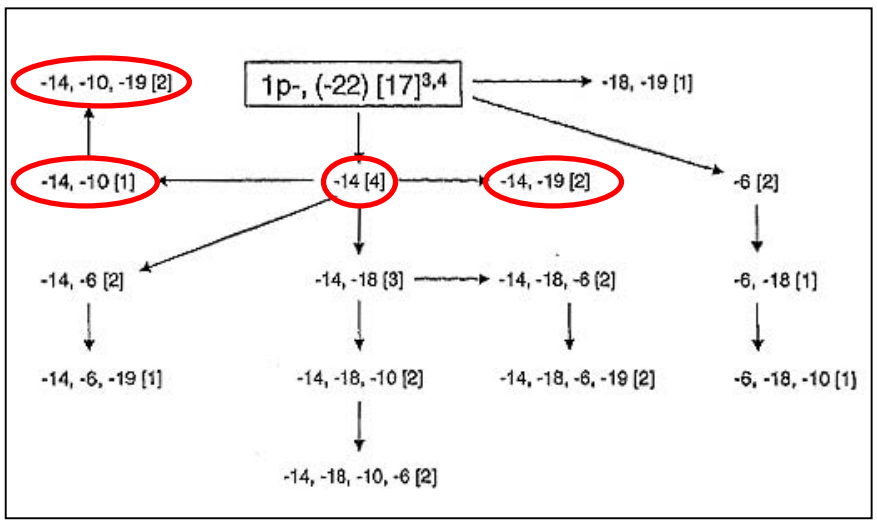
- Department of Neuropathology, Charité, Humboldt University, Berlin

  Prof. Andreas von Deimling

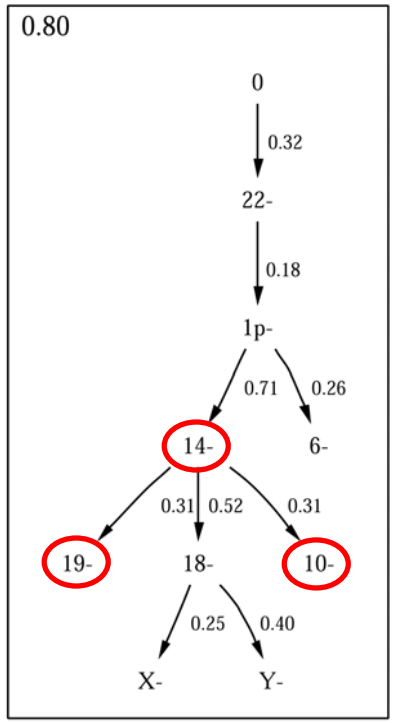
Evolutionary Pathways

- Meningioma
  - Well-studied benign brain tumor.
  - Expert knowledge: Initial events are 22- and 1p- (Zang, 2001)





Expert handcrafted model
for carcinogenesis
in meningioma

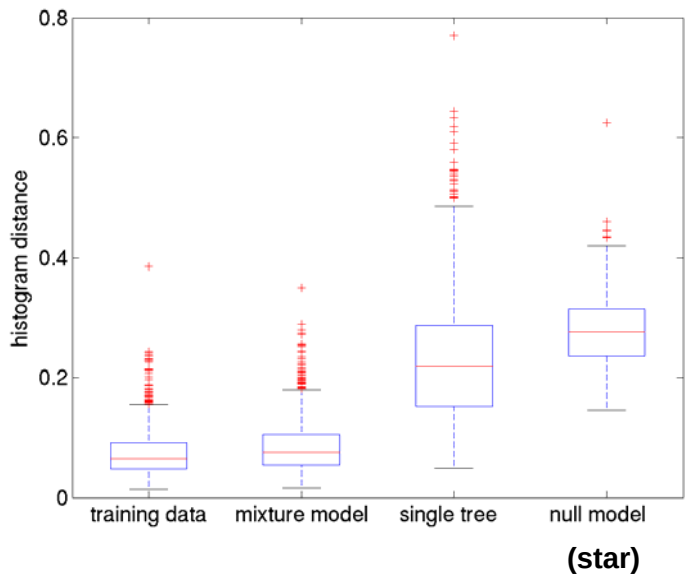


Estimated model
for carcinogenesis
in meningioma

Evolutionary Pathways

- Model selection
  - Modified BIC criterion
  - Redundancy defined as maximum similarity between two tree components
  - Penalize both complexity and redundancy
- Tree Stability
  - Bootstrap analysis
  - Draw bootstrap replicates from the induced distribution and estimate a new tree
  - Count how often an edge appears in reconstructed tress

- Performance as Density Estimator
  - Quantify how closely a trained model reproduces the empirical distribution
  - Distance measures to compare distributions: Cosine distance
  - Compute distance to test data with CV

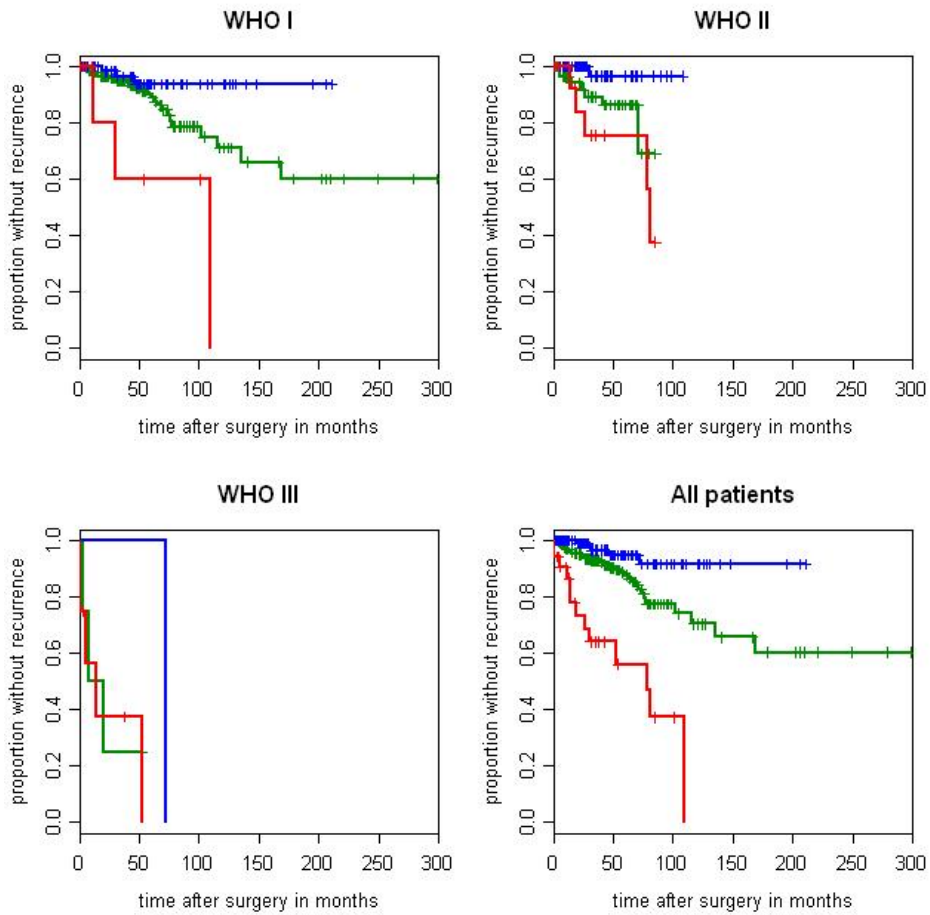


Diagnostic Relevance

GPS=0 GPS=2 GPS=6

Meningioma

- Analysis of time to recurrence
- WHO grade is widely used clinical grading system, tumors are classified in grades WHO I, WHO II, WHO III with increasing malignancy.
- We split patients into three groups according to GPS.
- Significant shorter time to recurrence for group with GPS>6.
- GPS improves diagnostics!



Diagnostic Relevance

Meningioma

- Analysis of time to recurrence
- WHO grade is widely used clinical grading system, tumors are classified in grades WHO I, WHO II, WHO III with increasing malignancy.