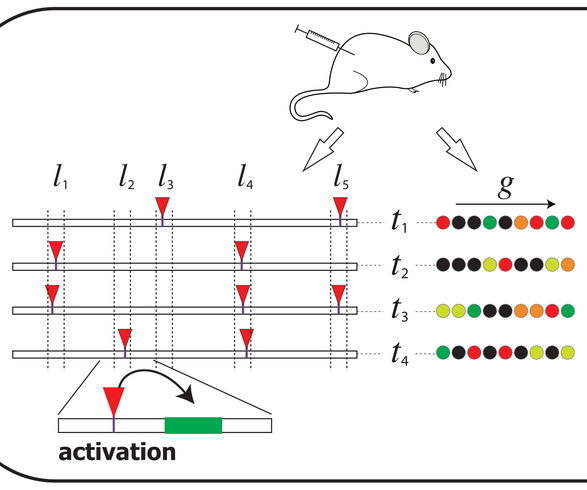
Discovering cancer pathways
by inferring combinatorial association logic

Jeroen de Ridder[1,2,*], Jan Bot[1], Jaap Kool[3], Anthony Uren[3], Lodewyk Wessels[1,2], Marcel Reinders[1]

*J.deRidder@TUDelft.nl

[1]Information & Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands. [2]Division of Molecular Biology, [3]Division of Molecular Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands



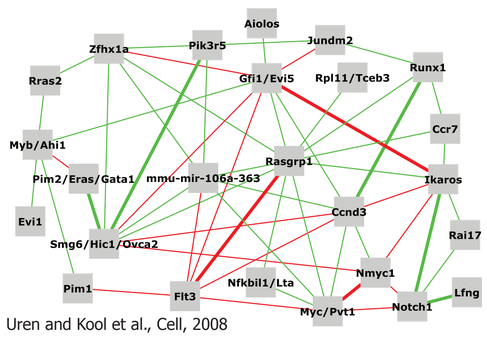
Summary

In this study, 43 tumors that were induced by retroviral insertional mutagenesis are expression profiled, resulting in a dataset for which both the initiating events (the viral integration sites) as well as the consequent expression profiles are available.

To capture complex associations that arise due to interaction among insertion target genes, we infer small Boolean logic networks that explicitly incorporate operators to model the potential parallel alternatives ('exclusive-or' gates) as well as the potential cooperation between mutations ('and' gates).

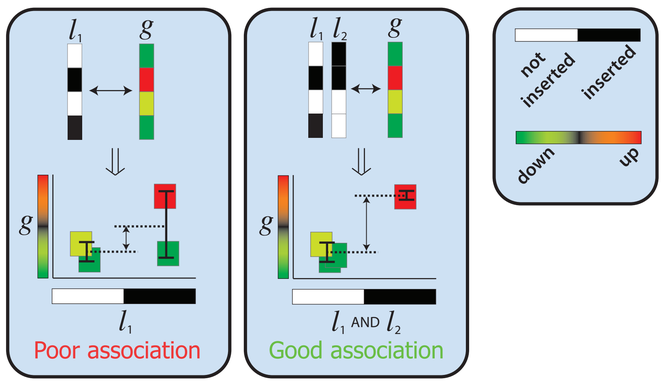
Co-occurrence and mutual exclusiveness

Interaction network



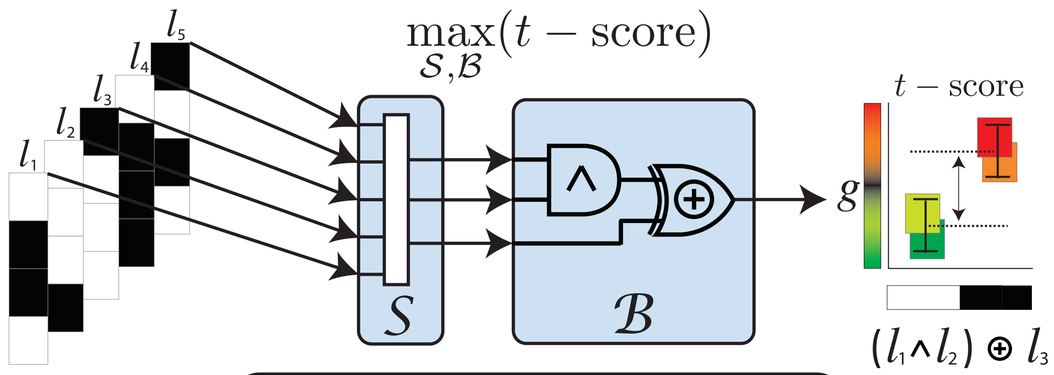
Uren and Kool et al., Cell, 2008

Significantly co-occurrent
Significantly mutually exclusive



Interactions destroy correlation

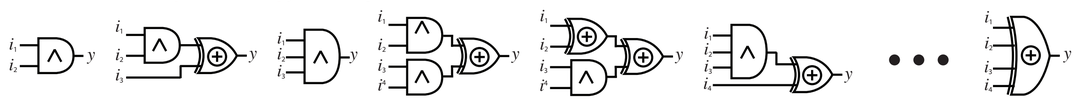
Combinatorial Association Logic (CAL)



Challenging optimization problem

Observation 1 - limited number of networks

- Due to the risk of overtraining, only small networks are considered
- Due to symmetry, many networks are not considered



Solve for each topology seperately

Observation 2 - optimize approximate t - score

$$\hat{t} - \text{score} = \sum_{i \in \text{all tumors}} \mathbf{w}_i (\mathbf{y}_i \cdot \mathbf{y}_i^{\text{opt}})$$

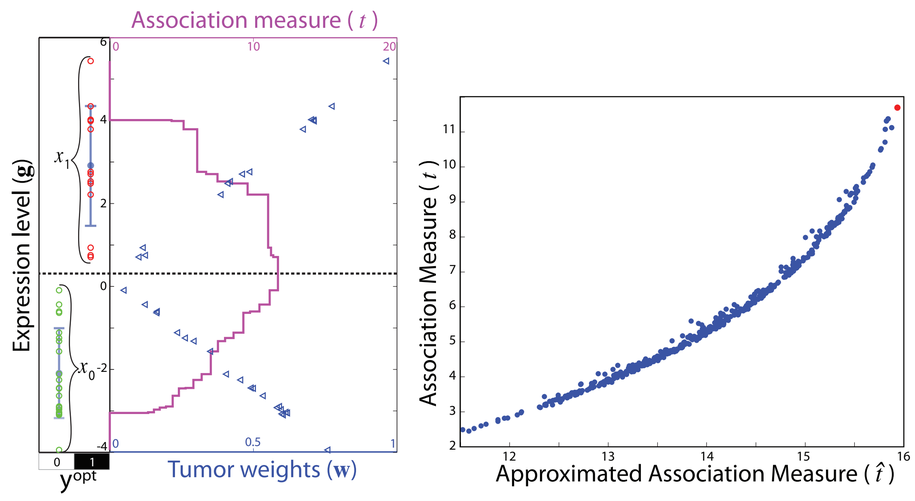
$\mathbf{y}^{\text{opt}}$ → **Best possible solution** (independent of network topology or inputs)

$\mathbf{w}$ → **Tumor** weights, found by minimizing difference between t and $\hat{t}$

Efficiently optimized by branch-and-bound

Estimating weights

$$\mathbf{w} = \min_{\mathbf{w}} ||t - \hat{t}||_2$$



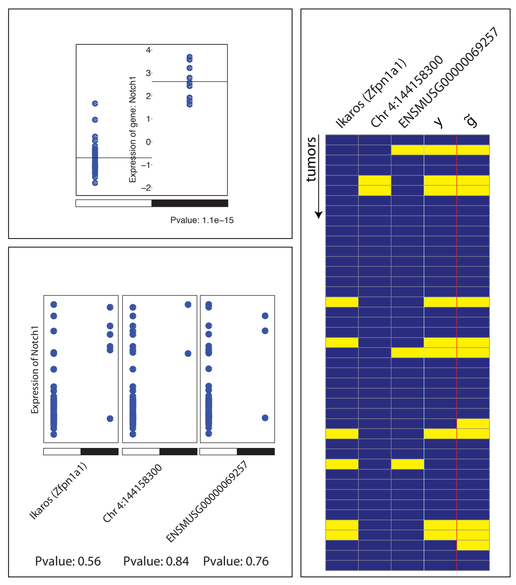
Very good approximation can be obtained

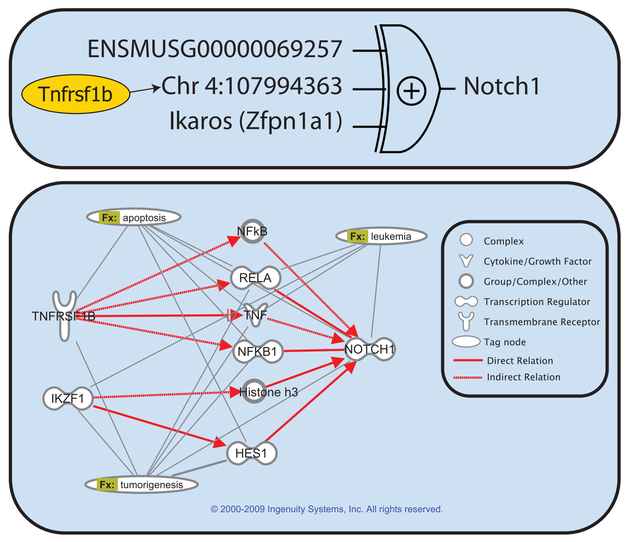
Estimating significance

$$p - \text{value} = \mathbb{P}(t \geq t^{\text{opt}} | \mathbf{L}, \mathbf{g}, \mathcal{B})$$

Determined by means of permutation procedure

Results





New pathways discovered

NKI-AVL
The Netherlands Cancer Institute
Antoni van Leeuwenhoek Hospital

nbic
Netherlands Bioinformatics Centre

TU Delft
Delft University of Technology