Discovering cooperating oncogenes
by statistical analysis of
Retroviral Insertional Mutagenesis Data
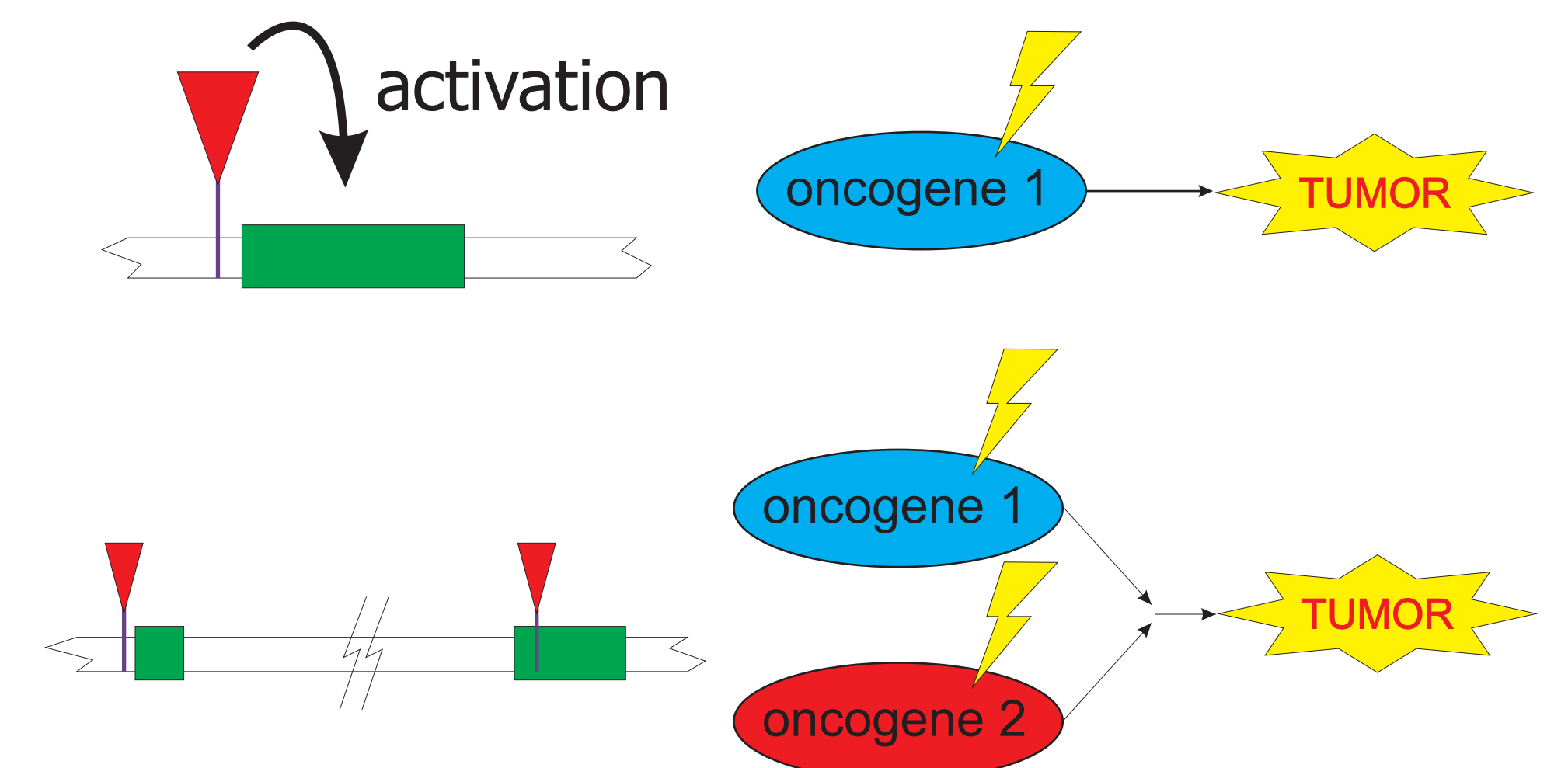
Jeroen de Ridder[1,2], **Lodewyk Wessels**[1,2], **Anthony Uren**[3], **Jaap Kool**[3] **& Marcel Reinders**[1]

[1]Information & Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, +31 (0)15 27 83418, J.deRidder@TUDelft.nl
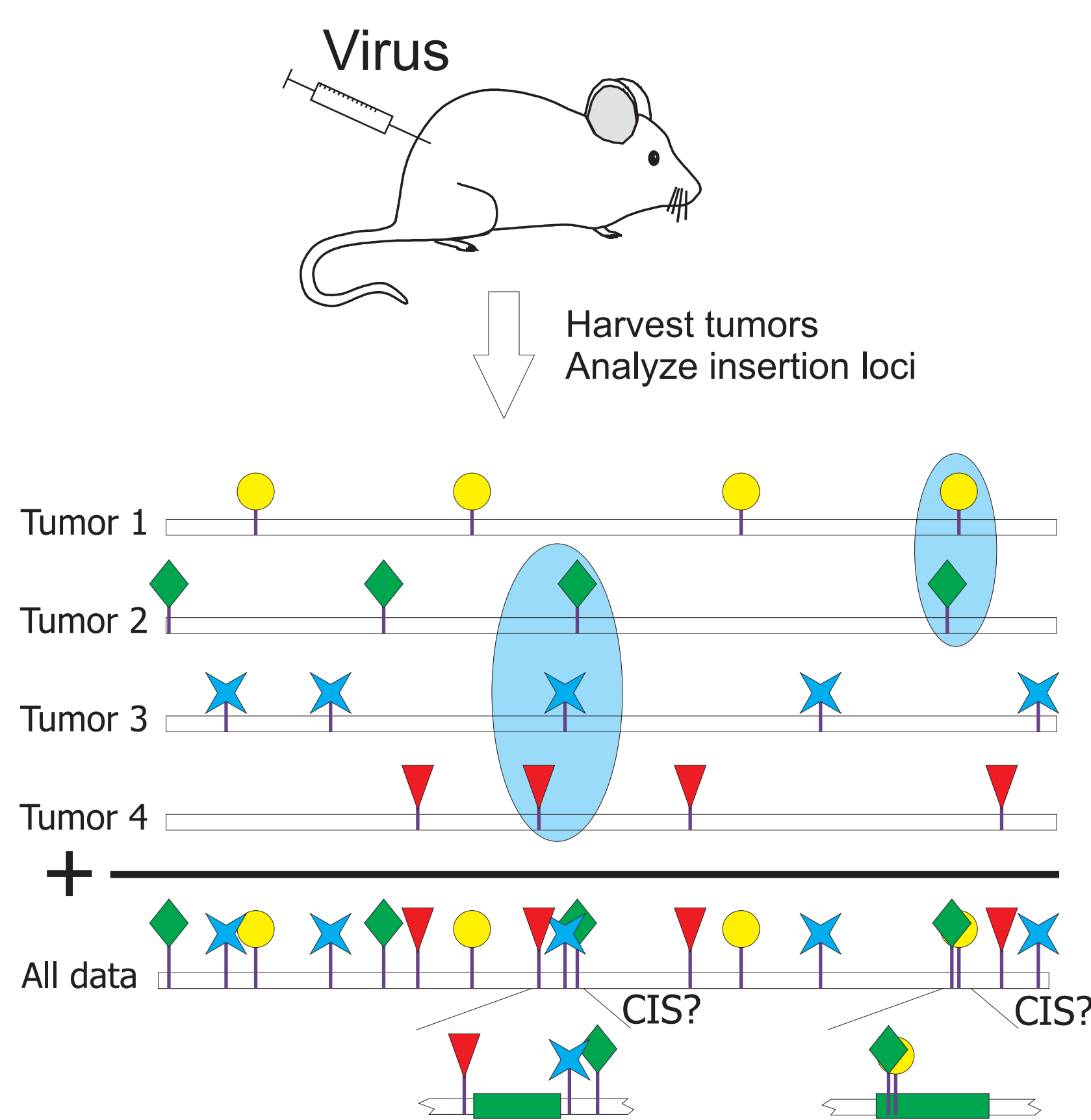
[2]Division of Molecular Biology, [3]Division of Molecular Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Abstract / Problem description

Viruses can induce oncogenic mutations when inserted near (or within) proto-oncogenes. Cancer genes can be identified by determining the loci of viral insertions from tumors induced by retroviruses. Most often, multiple co-occurring mutations are needed for a cell to develop into a tumor.

We propose a 2D Gaussian Kernel Convolution method to discover the cooperating oncogenes from publicly available retroviral insertional mutagenesis data.
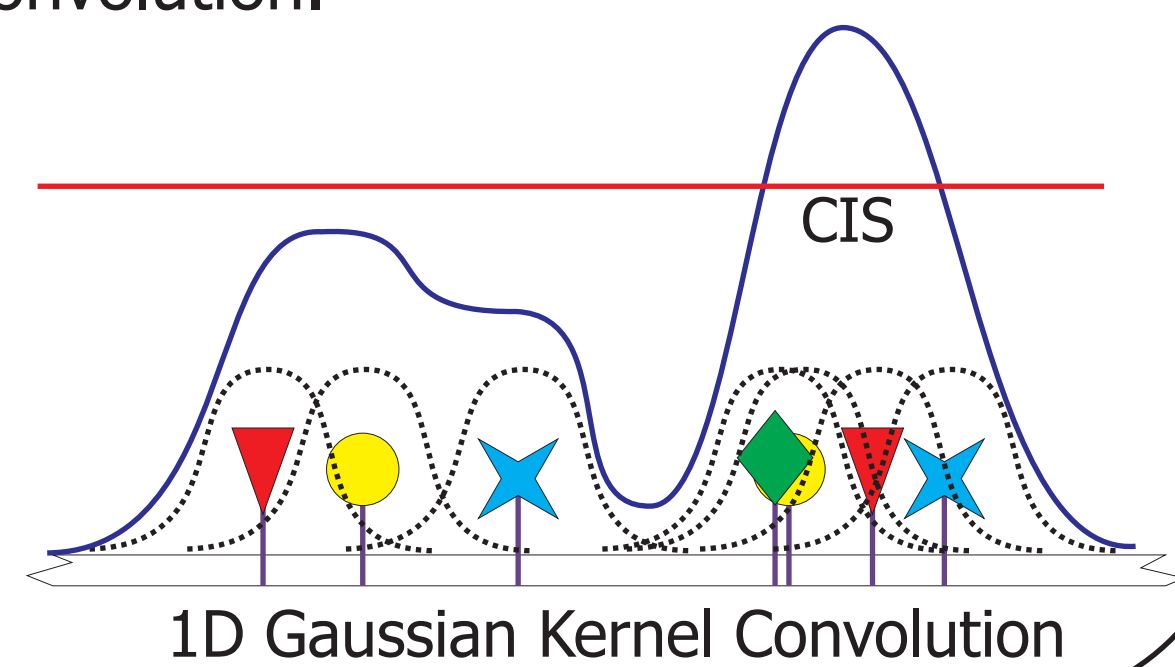
The data

An extensive amount of insertional mutagenesis data is compiled in the public Retroviral Tagged Cancer Gene Database (RTCGD). The database contains approximately 4000 insertions from 1076 tumors.

Common Insertion Site (CIS):
A region in the genome that carries insertions in multiple independent tumors significantly more frequent than expected by chance. CISs can be detected by using a 1D Gaussian kernel convolution.
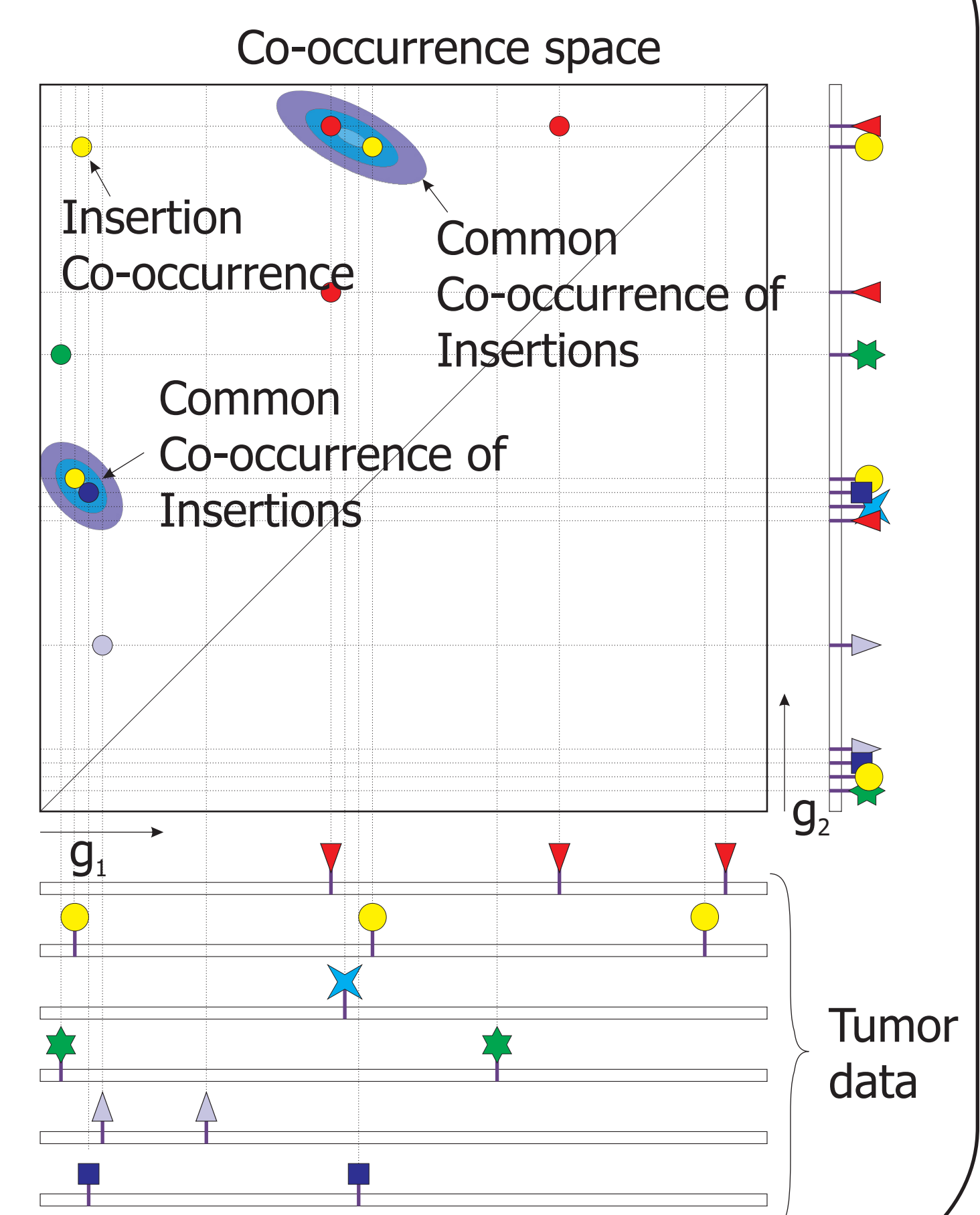
1D Gaussian Kernel Convolution

Procedure (1)

To identify cooperation between oncogenes we transform the data to the co-occurrence space.
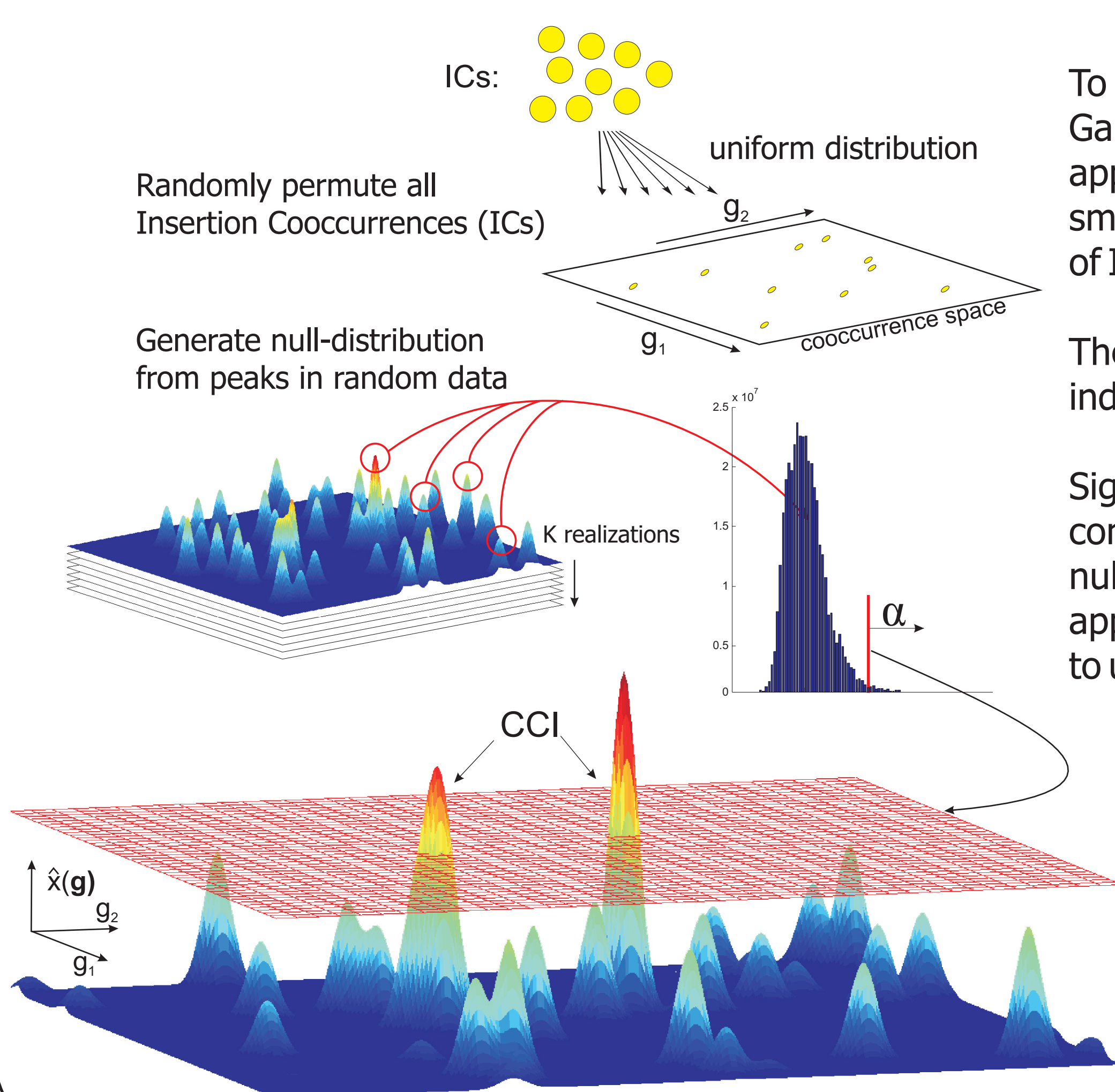
Insertion Co-occurrence (IC):
A unique combination of insertions observed to co-occur in the same tumor.

Common Co-occurrence of Insertions (CCI):
A combination of two loci in the genome, observed to be inserted in the same tumor significantly more frequent than expected by chance, across multiple tumors.

CIS-CIS interaction:
A combination of insertions that occur in the same tumor and both fall in a CIS. When considering only CIS-CIS interactions significant co-occurring insertions may be missed.

Procedure (2)

ICs:

Randomly permute all Insertion Cooccurrences (ICs)

uniform distribution

Generate null-distribution from peaks in random data
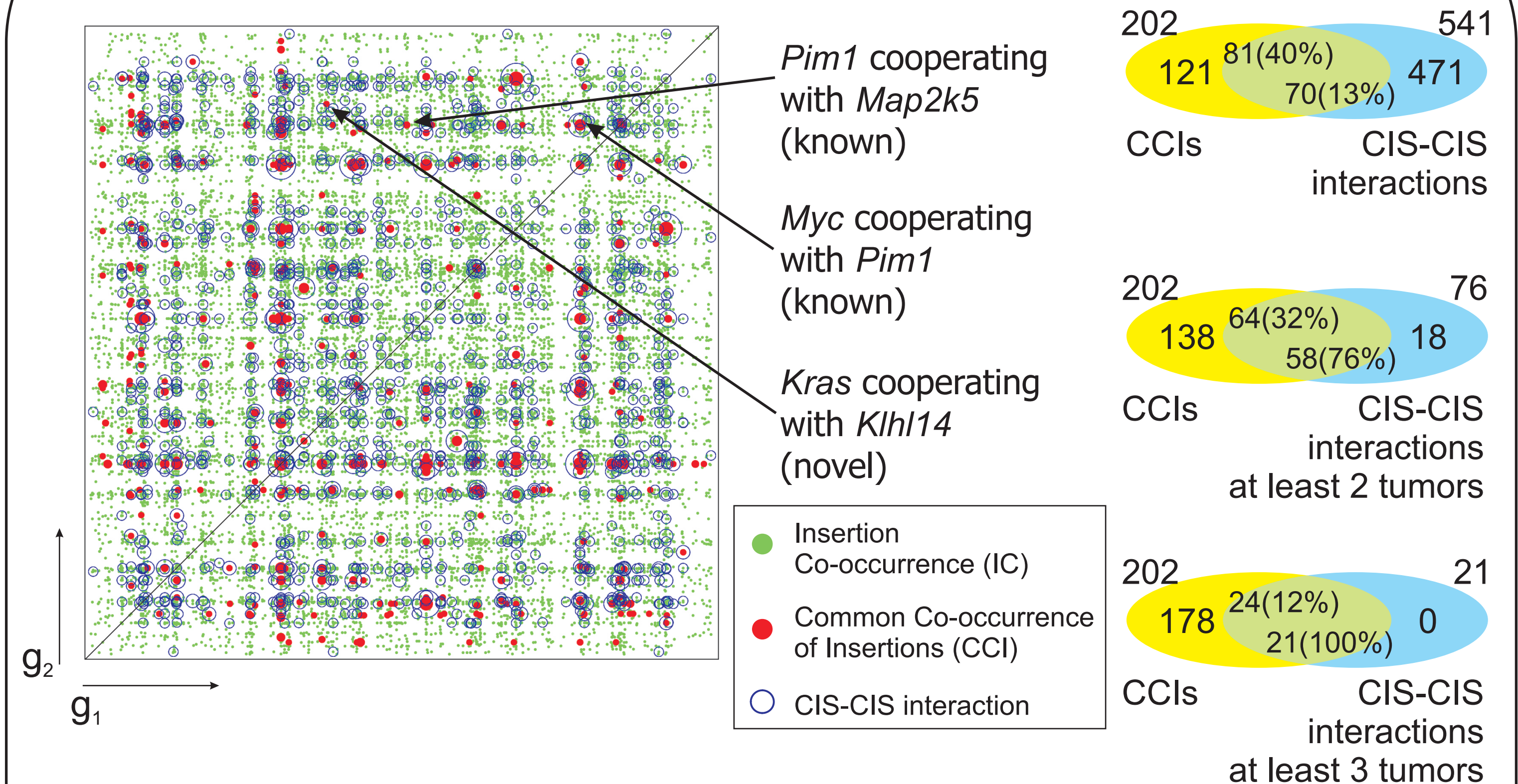
K realizations

CCI

To detect the CCIs, a 2D Gaussian kernel convolution is applied to the data resulting in smooth estimate of the number of ICs

The peaks in the estimate of ICs, indicate putative CCIs.

Significance is evaluated by comparing the peak heights to a null-distribution, resulting from applying the kernel convolution to uniform random data.

Conclusions

Pim1 cooperating with *Map2k5* (known)

Myc cooperating with *Pim1* (known)

Kras cooperating with *Klhl14* (novel)

- Insertion Co-occurrence (IC)
- Common Co-occurrence of Insertions (CCI)
- CIS-CIS interaction

202		541
121	81(40%) / 70(13%)	471
CCIs		CIS-CIS interactions

202		76
138	64(32%) / 58(76%)	18
CCIs		CIS-CIS interactions at least 2 tumors

202		21
178	24(12%) / 21(100%)	0
CCIs		CIS-CIS interactions at least 3 tumors

Multidimensional analysis of retroviral insertional mutagenesis data results in the discovery of cancer genes that may cooperate in tumorigenesis. Apart from known interactions, also novel interactions are identified.

The Venn diagrams show that the interactions could only be discovered by doing a multidimensional analysis, since the CCIs are not fully contained in the set of CIS-CIS interactions, that resulted from a 1D analysis.

NKI-AVL
The Netherlands Cancer Institute
Antoni van Leeuwenhoek Hospital

Information and Communication Theory Group $[I,C]^T$

TUDelft

Delft University of Technology