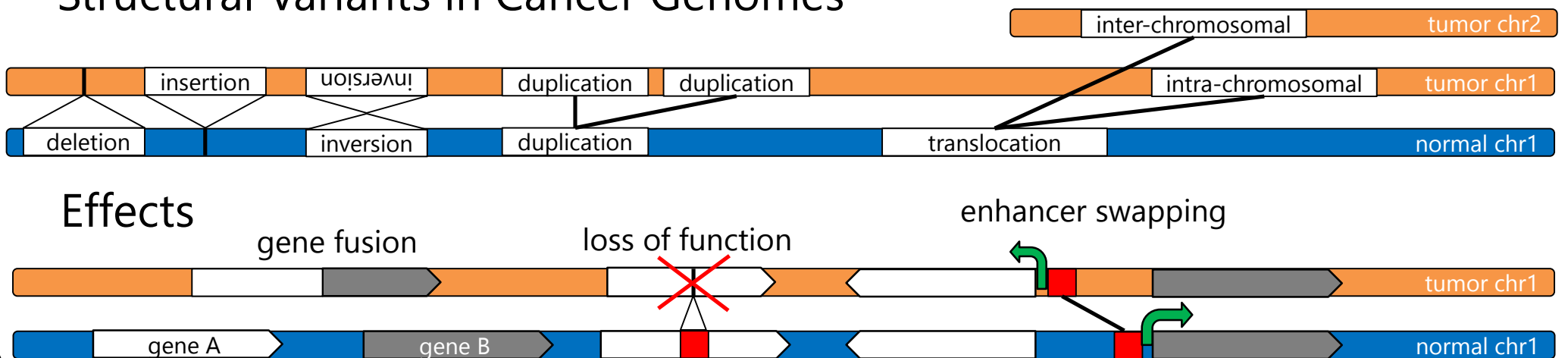


# DeepSV: Somatic Structural Variant Detection with Deep Learning

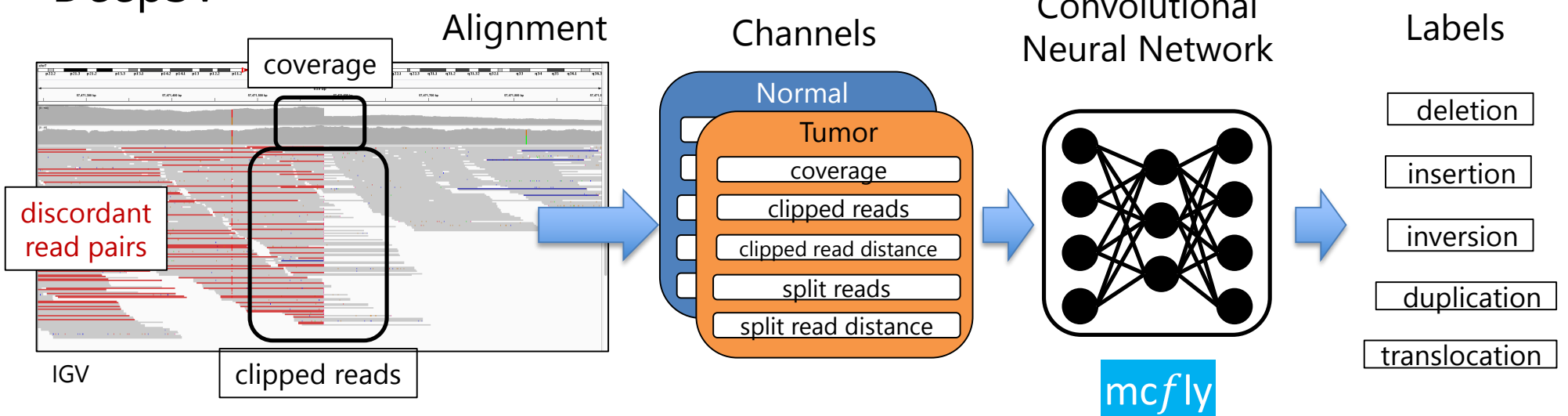
Luca Santuari<sup>1</sup>, Sonja Georgievska<sup>2</sup>, Carl Shneider<sup>1</sup>, Arnold Kuzniar<sup>2</sup>, Tilman Schaefer<sup>1</sup>, Wigard Kloosterman<sup>1</sup> and Jeroen de Ridder<sup>1</sup>

1 Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, NL  
2 Netherlands eScience Center, Science Park 140, 1012 WX Amsterdam, NL

## Structural Variants in Cancer Genomes



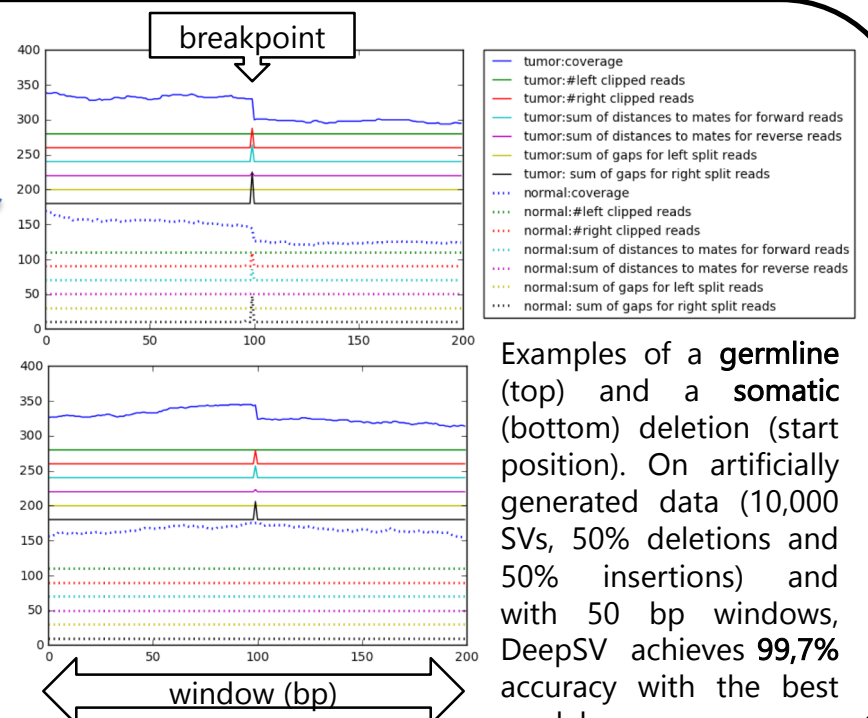
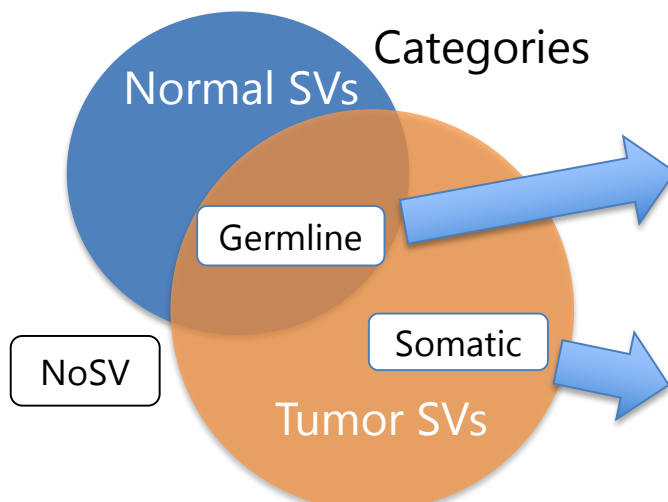
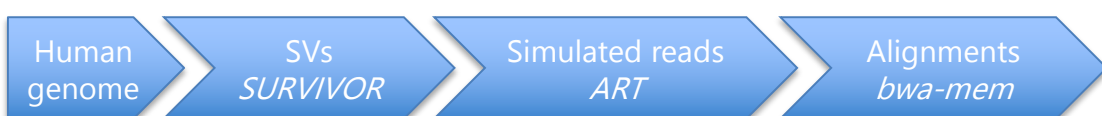
## DeepSV



## Methods

We generated a dataset of somatic and germline SVs (deletions and insertions) to train DeepSV and assessing its accuracy in detecting SV breakpoints, before applying it to real data.

Training set:  
Artificial data workflow



Examples of a **germline** (top) and a **somatic** (bottom) deletion (start position). On artificially generated data (10,000 SVs, 50% deletions and 50% insertions) and with 50 bp windows, DeepSV achieves **99,7%** accuracy with the best model.

## Perspectives

DeepSV has high accuracy for detecting SV breakpoints in artificial WGS data using the current channel specification. In the next step, we will test the model on multiple real tumor/normal sample pairs.

GitHub repositories  
GTCG: <https://github.com/GooglingTheCancerGenome>  
McFly: <https://github.com/NLeSC/mcfly>