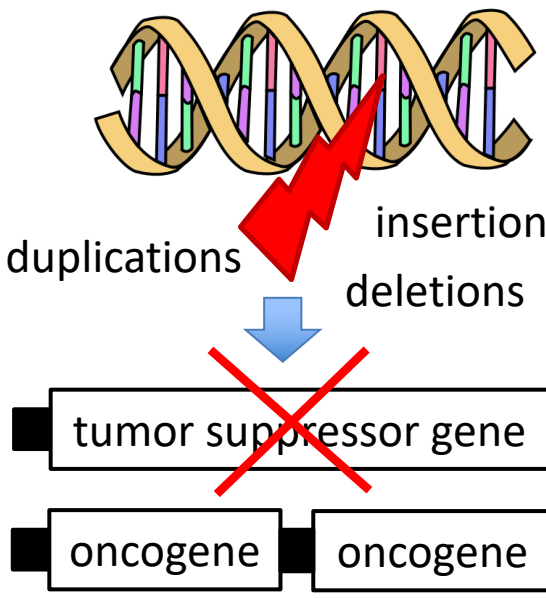
DeepSV: somatic structural variant detection using deep learning

UMC Utrecht

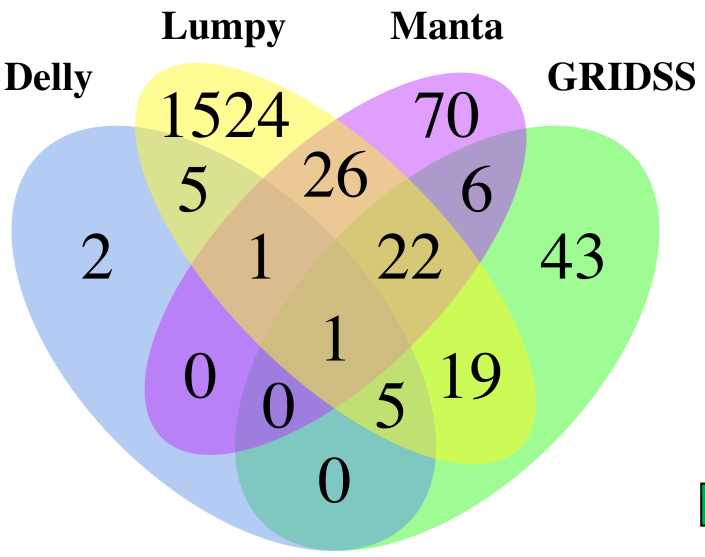
Luca Santuari[1*], Sonja Georgievska[2*], Carl Shneider[1], Arnold Kuzniar[2], Tilman Schaefers[1], Wigard Kloosterman[1] and Jeroen de Ridder[1]

1 Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, Netherlands
2 Netherlands eScience Center, Science Park 140, 1012 WX Amsterdam, Netherlands
* These authors have contributed equally to this work.

Structural **V**ariants

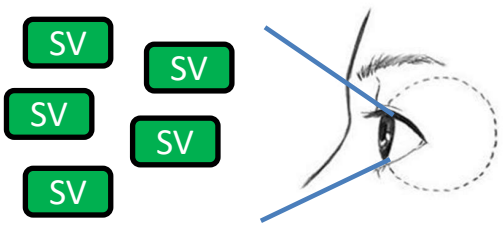


Lack of consensus among SV callers



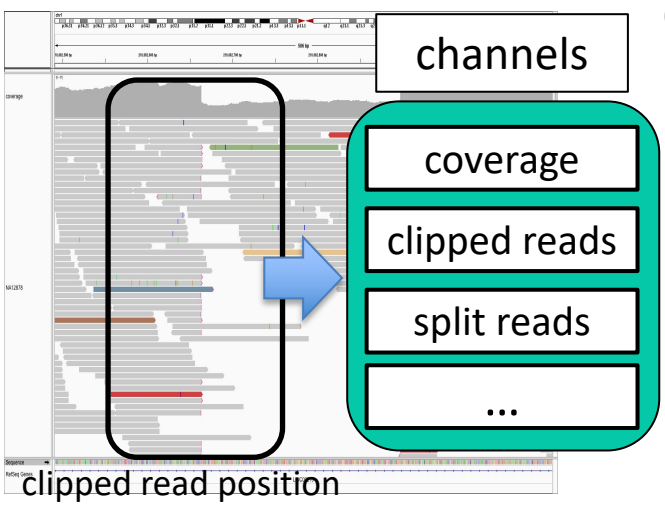
COLO829 cell line

Internal parameters set by trial and error



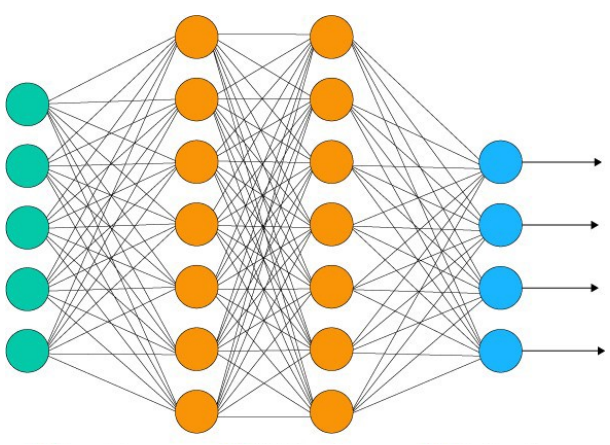
Learn parameters automatically from the data

Read alignment



Breakpoint detection

Convolutional Neural Network



Output generation

SV breakpoints

postprocessing

SVs

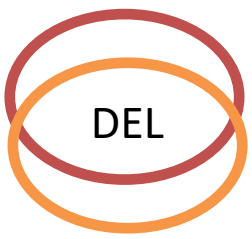
DeepSV workflow

Read alignments at genomic intervals (200bp) centered on clipped read positions are converted to channels (one-dimensional representations). A deep neural network is trained on windows labelled (SV breakpoints or not) to classify SV breakpoints in an independent test set. Clipped read and paired end read information are then used to generate the final SV callset.

DeepSV performance on germline deletions (DEL)

Genome in a Bottle NA12878

Mills et al. 2011



nanoSV callset **Nanopore**

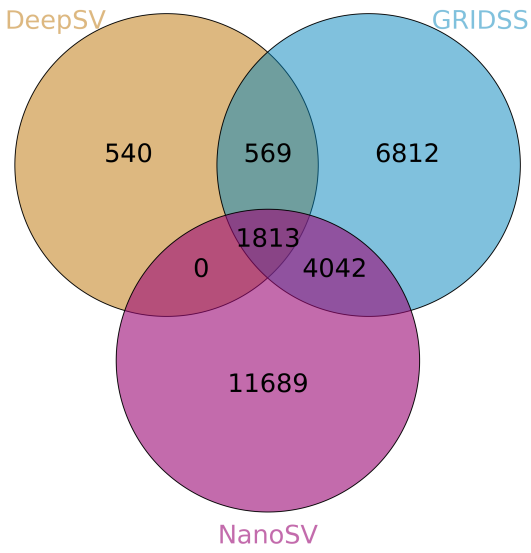
Ground truth

Predicted

	start	end	no DEL	Recall
start	175	2	4	96.7%
end	1	194	6	96.5%
no DEL	0	3	188	98.4%
Precision	99.4%	97.5%	95%	

Training: 2400
Test: 573

Performance on Patient1



GitHub https://bit.ly/2CtJXVk

References
Mills et al., Nature, 2011
Cretu Stancu et al., Nat. Commun., 2017

Oncode Institute

DeepSV aims at bringing machine-learned SV detection closer to clinical applications, by learning relevant parameters automatically from the data.

L.Santuari@umcutrecht.nl