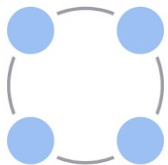


# Developing method for filtering false positive CNVkit results of CNV presence in tumor tissue samples

Anastasiya Vinogradova

*Artem Kasianov*  
*Skoltech*

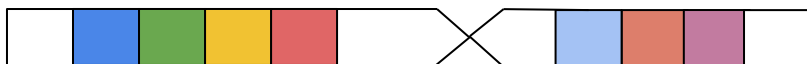
*Atlas Oncodiagnostics*



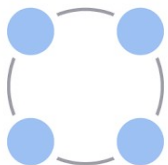
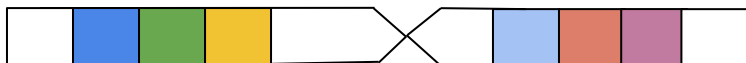
# Copy Number Variation



Duplication



Deletion



# Background

There are treatments for certain CNV-associated cancers

Trastuzumab Pertuzumab Lapatinib	Breast cancer	<i>HER2</i> amplification
Gefitinib, erlotanib Cetuxumab	Lung cancer Bowel cancer	<i>EGFR</i> mutation
PKC412, SU11248, CMT53518	AML, ALL	<i>FLT3</i> mutation, tandem duplication
PARP inhibitors	Breast cancer Ovarian cancer	<i>BRCA1/2</i> mutation

BUT:

Treatment is expensive, Treatment has side effect  
So false positive results takes time,  
health and money from patient

# Possible solution

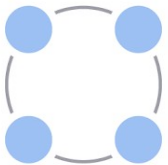
Let's just compare health tissue samples and tumor tissue samples from patient

BUT:

Taking health tissue samples is expensive and this is extra traumatic for the patient organ

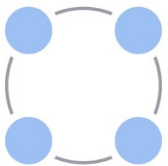
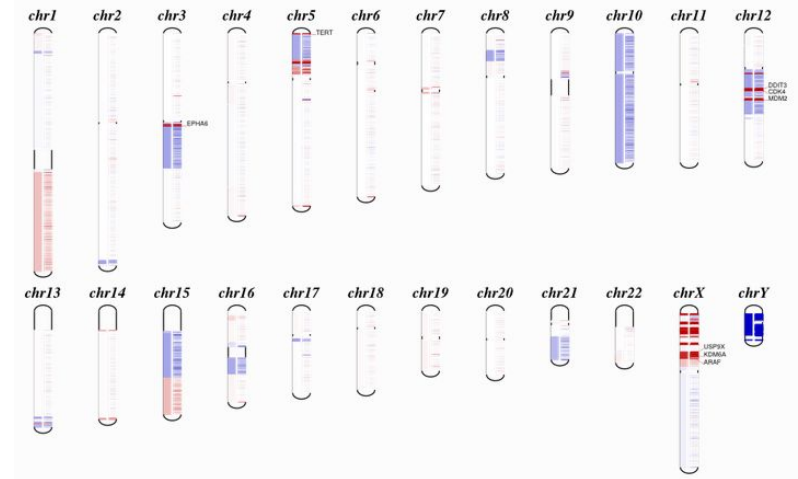
So:

Use only tumor samples from various tissue of different patients



# Goal of the project

Develop a method for filtering false positive CNVkit results of CNV (Copy Number Variation) presence in tumor tissue samples



# Bayesian model

$$P(cnv|data) = \frac{P(cnv)(data|cnv)}{P(data)}, P(nocnv|data) = \frac{P(nocnv) \cdot P(nocnv|data)}{P(data)}$$

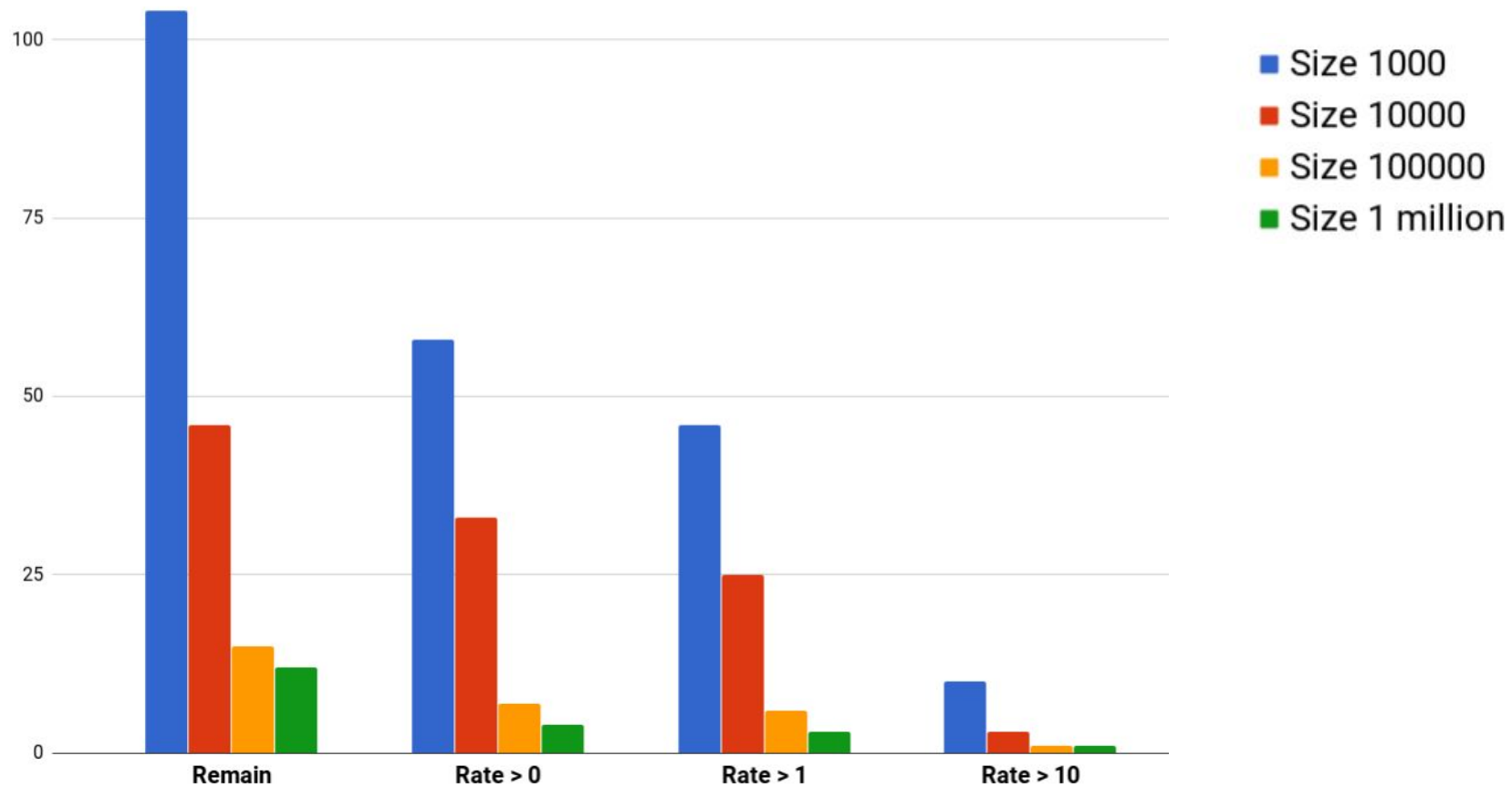
$$Rate = \frac{P(cnv|data)}{P(nocnv|data)} = \frac{P(cnv)}{P(nocnv)} \cdot \frac{FC \cdot M \cdot e^{-M \cdot FC}}{M \cdot e^{-M}}$$

$$C = \frac{Rate}{\frac{P(cnv)}{P(nocnv)}} = Rate \cdot \frac{P(nocnv)}{P(cnv)}$$

FC - Fold Change

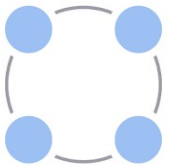
M - Mean region coverage

# CNV count ~ region size



# Filter method

- 1) Set region size
- 2) Filter regions according the coverage
- 3) Use Bayesian model
- 4) Set Rate for filtering CNV





# Rate choose

C	2	5	10	20	40	60
Remain cnv	19	6	3	2	2	2
Error	5	1	1	1	1	0
Filtered 2	74	78	78	78	78	79

CNV after Filter 2	19	6	3	2	2	2
CNV after Filter 1	187	187	187	187	187	187
All CNVkit cnv	210	210	210	210	210	210

# Result

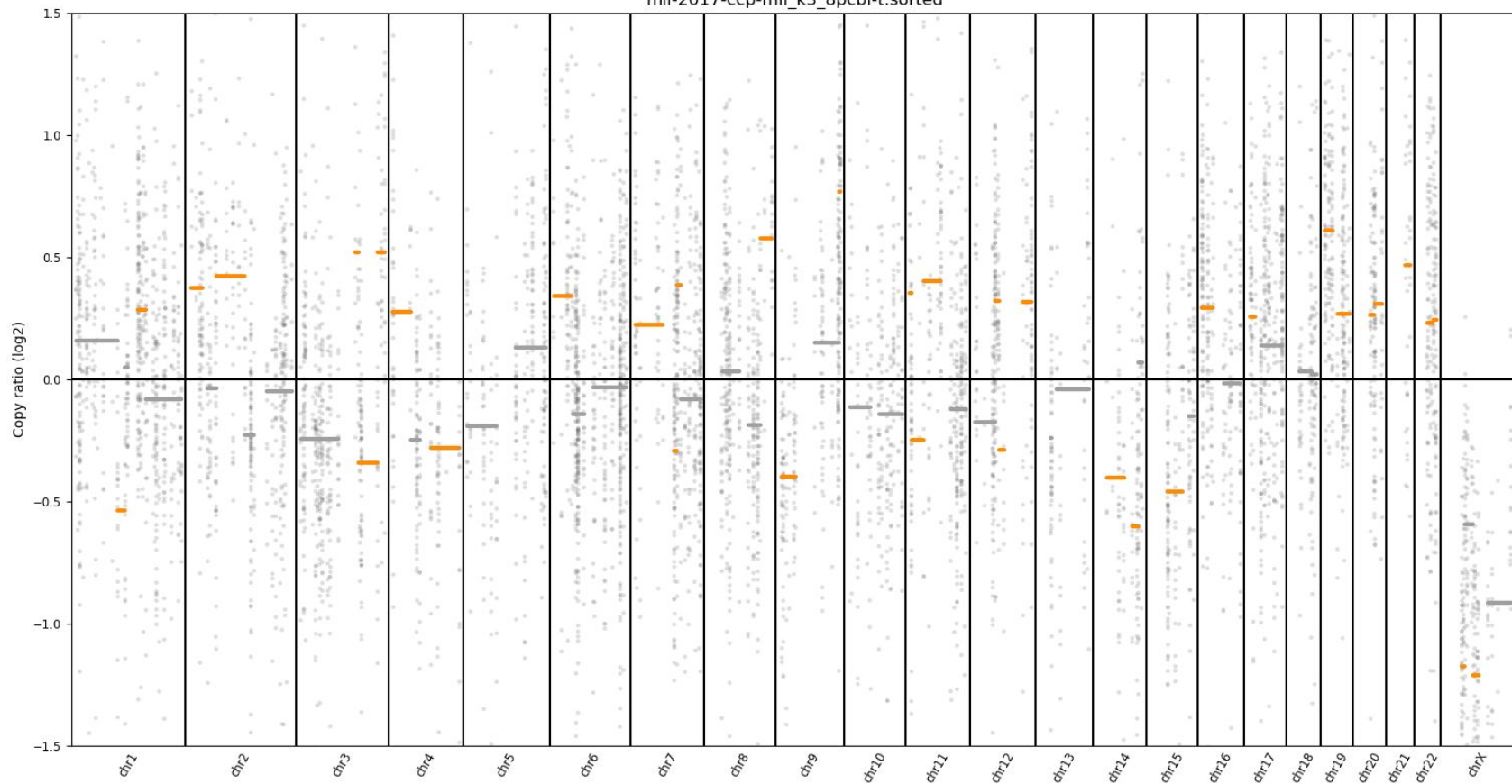
$$\left. \begin{array}{l} size = 10000 \\ rate = 60 \end{array} \right\} + 815 \text{ samples' regions} \Rightarrow 2 \text{ cnv}$$

$$P_{pred}(cnv) = 2/815 \approx 0.002$$

$$P_{pred}(cnv) = P_{true}(cnv)$$

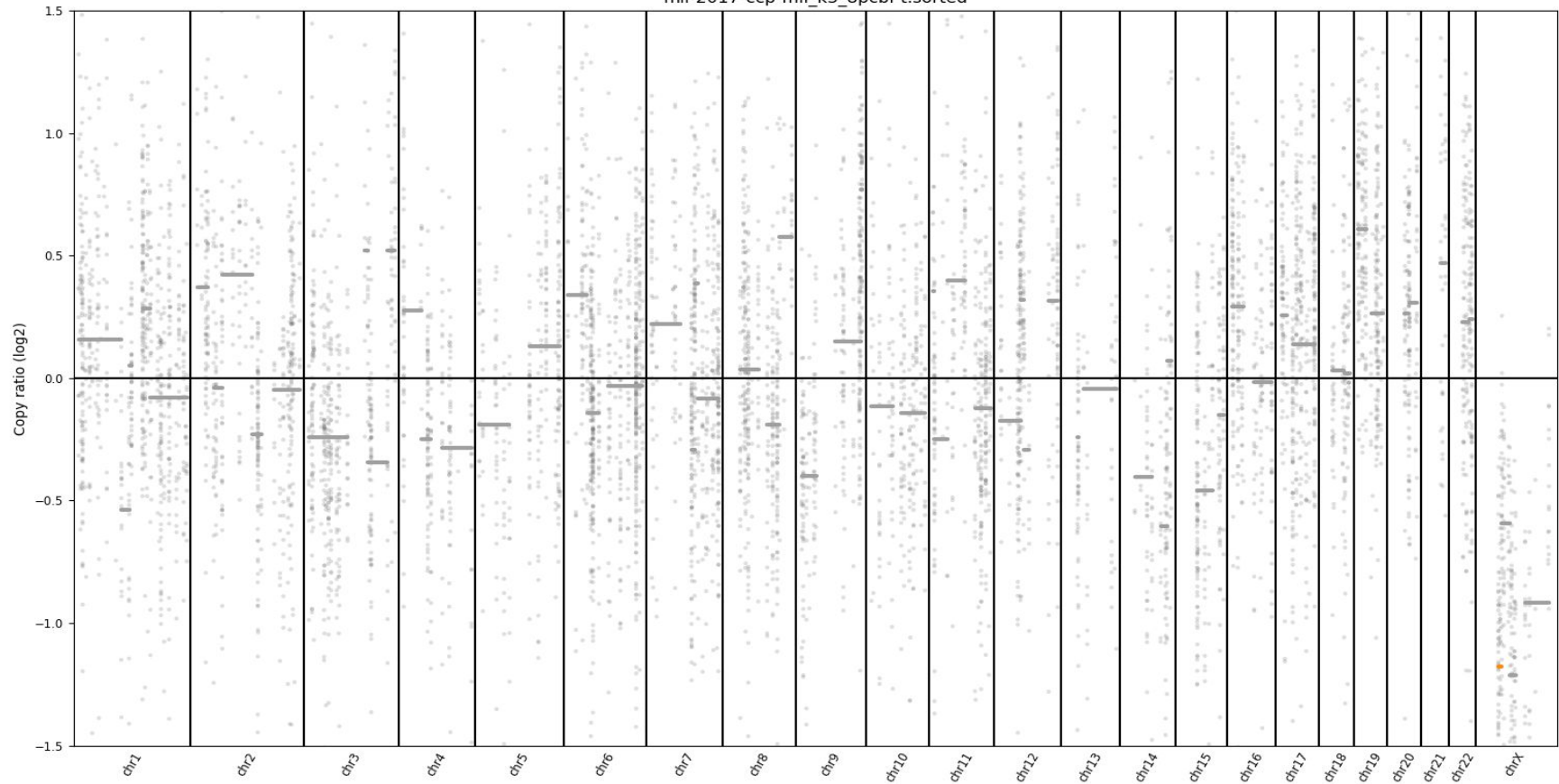
# Before

mil-2017-ccp-mil\_k3\_8pcbl-t.sorted

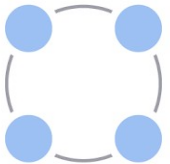


# After

mil-2017-ccp-mil\_k3\_8pcbl-t.sorted



Thank you for attention!



# Link to project

<https://github.com/newCarroll/Filtering-false-positive-results-of-cnv-presence>

