

ИНСТИТУТ
БИОИНФОРМАТИКИ

Joint analysis of Rep-seq and mass spectra data

Noskova Ekaterina

Scientific advisor: Safonova Yana

**Center for Algorithmic Biotechnology,
Saint Petersburg State University.**

NGS & MS

Существует два способа получения данных о репертуаре антител:

1. **Иммуносеквенирование** РНК В-клеток.

(+) можно покрыть полную длину антитела

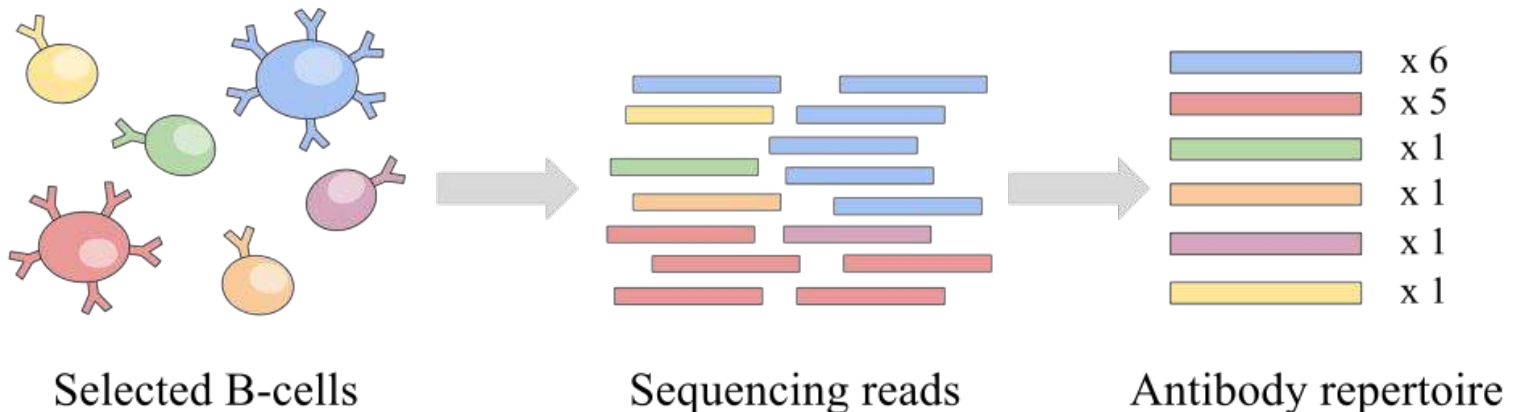
(-) могут быть биологически нерелевантные результаты

2. Данные **масс спектрометрии** репертуара антител в виде белков.

(+) отражает биологически важные антитела

(-) данные обычно очень короткие (10 аа против требуемых 120 аа)

Immunoproteogenomics - комбинация двух подходов.



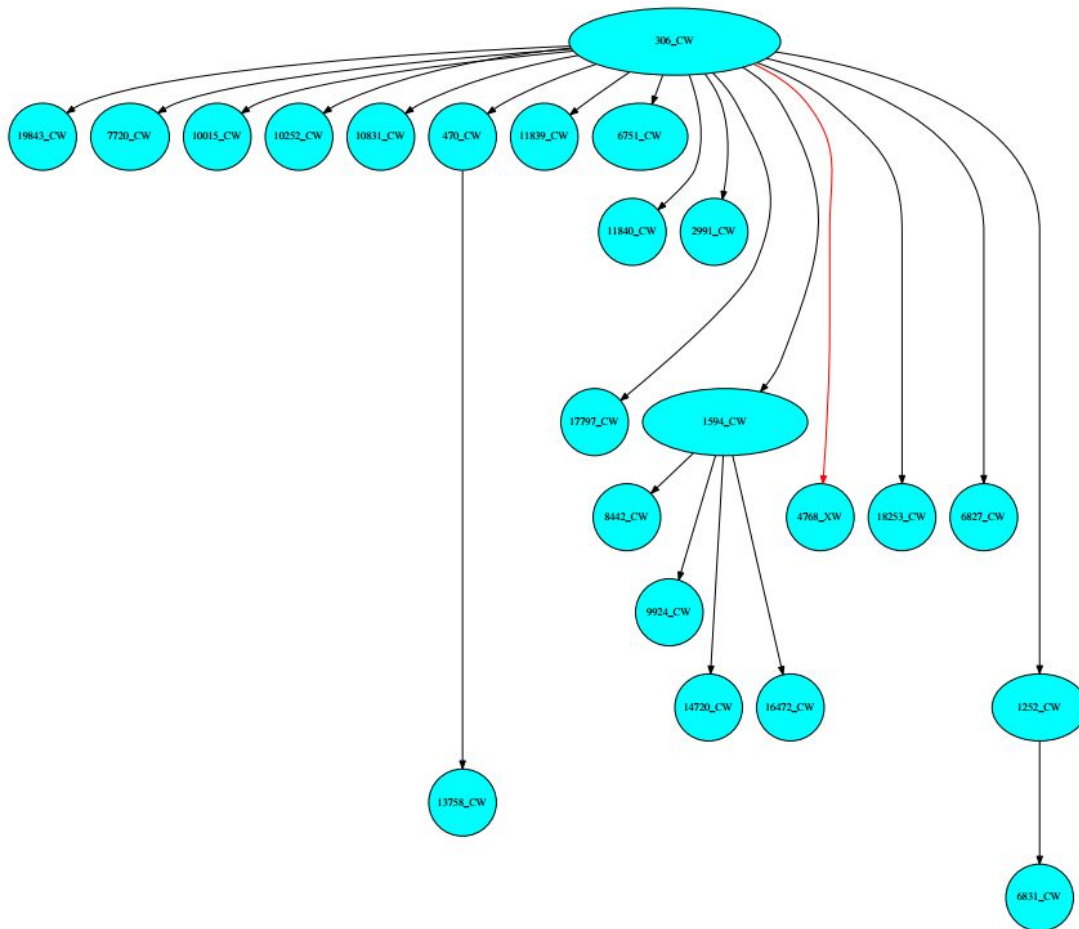
Постановка задачи

1. С помощью комбинации данных иммуносеквенирования и масс-спектрометрии найти полные последовательности биологически **важных** белков
2. Стандартный метод – иммунопротеоеномный поиск – матчинг масс спектров против репертуара, построенного по данным секвенирования.
 - a. Cheung et al., *Nat Biotech*, 2012: monoclonal antibodies, 454 sequencing technology
 - b. Safonova et al., *Bioinformatics*, 2015: polyclonal antibodies, Illumina MiSeq;

“Our analysis revealed an alarming lack of correlation between two approaches to quantification: more than half of identified peptides come from antibodies that are represented by a single read in the antibody repertoire!”

3. Разработать новый подход для иммунопротеоеномного поиска

Репертуар антител как набор клональных деревьев



Репертуар антител - результат эволюционного процесса

Используя информацию о разбиении репертуара антител на клональные деревья, мы ожидаем увидеть более хорошее соответствие между представленностями антител в Rep-seq и MS.

Project pipeline

2 NGS datasets -> IgRepertoireConstructor, merge two repertoires

Merged repertoire -> AntEvolvo -> a set of clonal trees (# trees, size of largest tree)

Separate repertoire by cluster size: (5, 9), (10, 99), (100, 999), ≥ 1000 .

{ 4 separated repertoires,
8 MS files (2×AspN, 2×chymo, 2×trypsin, 2×elastase) } -> MS-GF+ -> matching in mzld

Finding correspondence between clonal trees and matched peptides

Match result

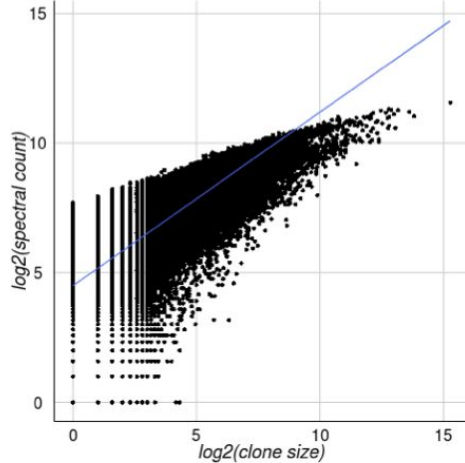
Получилось всего 34 дерева из 5117, которым было найдены соответствия в данных MS.

В отличие от матчинга на кластера, **на самое большое дерево попало больше всего пептидов!**

Clonal trees sorted by size (matched trees):

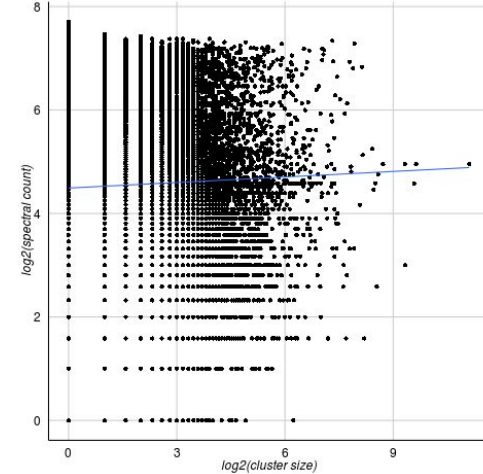
```
clonal_tree_117-444-0_Vsize_601_Esize_596.tree
clonal_tree_248-13-0_Vsize_313_Esize_312.tree
clonal_tree_151-16-0_Vsize_281_Esize_239.tree
clonal_tree_18-49-0_Vsize_161_Esize_154.tree
clonal_tree_129-12-0_Vsize_144_Esize_143.tree
clonal_tree_144-6-0_Vsize_124_Esize_111.tree
clonal_tree_281-13-0_Vsize_121_Esize_120.tree
clonal_tree_4-20-0_Vsize_105_Esize_101.tree
clonal_tree_130-17-0_Vsize_103_Esize_101.tree
clonal_tree_17-1-0_Vsize_99_Esize_92.tree
clonal_tree_202-50-0_Vsize_97_Esize_85.tree
clonal_tree_3-231-0_Vsize_96_Esize_87.tree
clonal_tree_294-16-0_Vsize_96_Esize_95.tree
clonal_tree_6-472-0_Vsize_95_Esize_90.tree
clonal_tree_283-940-0_Vsize_89_Esize_88.tree
clonal_tree_96-62-0_Vsize_89_Esize_88.tree
clonal_tree_151-317-0_Vsize_88_Esize_79.tree
clonal_tree_159-10-0_Vsize_85_Esize_84.tree
clonal_tree_36-157-0_Vsize_77_Esize_76.tree
clonal_tree_202-44-0_Vsize_76_Esize_75.tree
clonal_tree_267-3-0_Vsize_76_Esize_75.tree
clonal_tree_183-11-0_Vsize_76_Esize_69.tree
clonal_tree_36-147-0_Vsize_74_Esize_73.tree
clonal_tree_117-956-0_Vsize_70_Esize_68.tree
clonal_tree_295-142-0_Vsize_69_Esize_68.tree
clonal_tree_4-683-0_Vsize_63_Esize_58.tree
clonal_tree_36-388-0_Vsize_62_Esize_61.tree
clonal_tree_264-6-0_Vsize_62_Esize_61.tree
```

Correspondence between “clonal” size and matched peptides

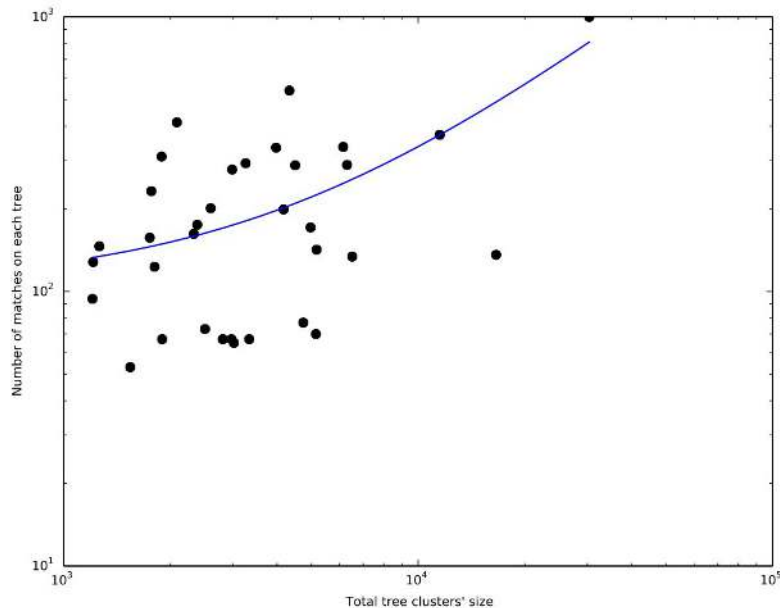


Pearson correlation $\rho = 0.5687614$.

Correspondence between cluster size and matched peptides



Pearson correlation $\rho = 0.007544031$.

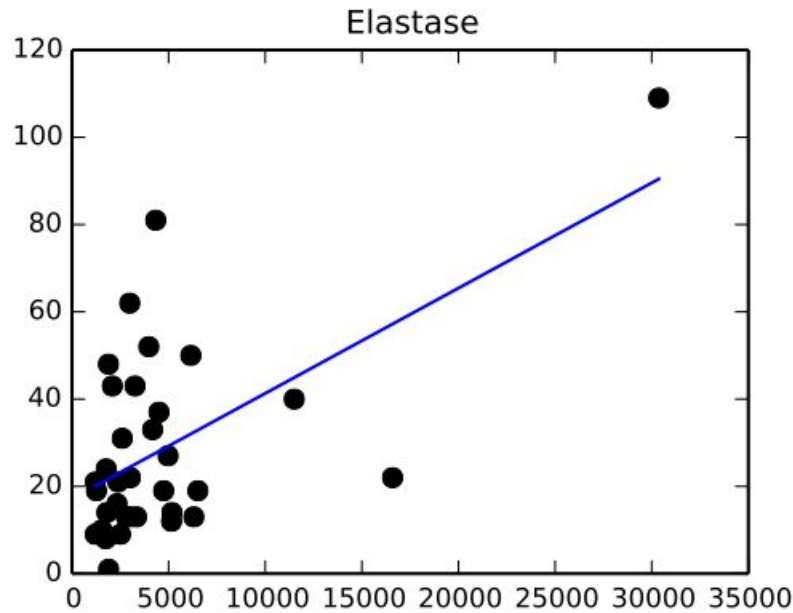


Pearson correlation $\rho = 0.693951$.

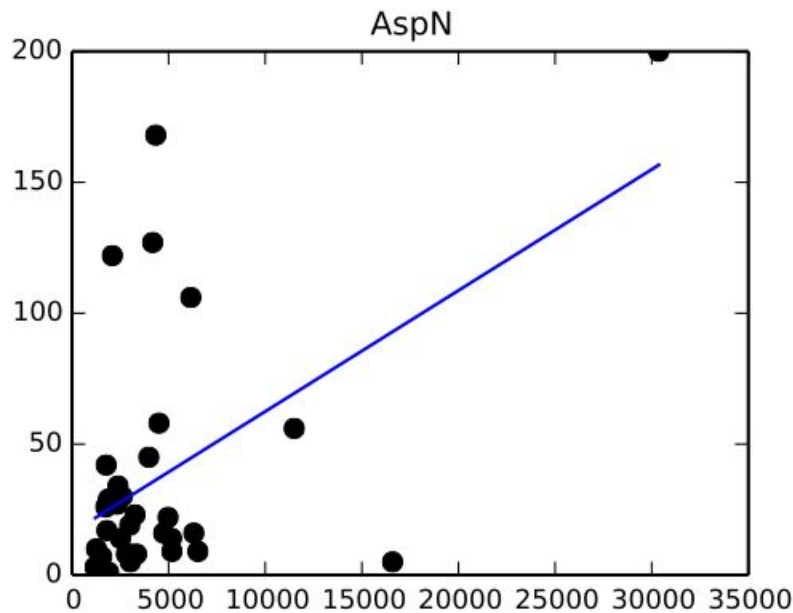
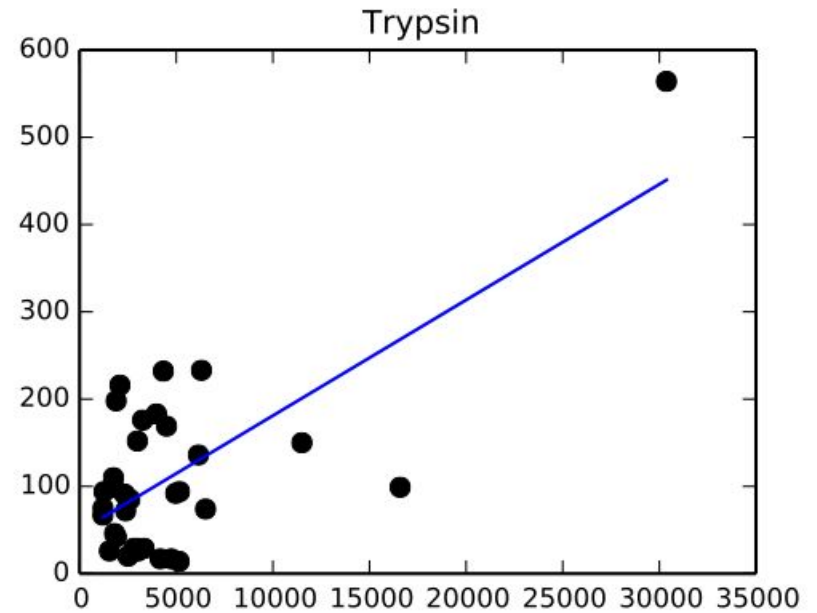
Мы смогли улучшить корреляцию между NGS и MS данными.

Самое большое дерево производит больше всего белков в репертуаре антител.

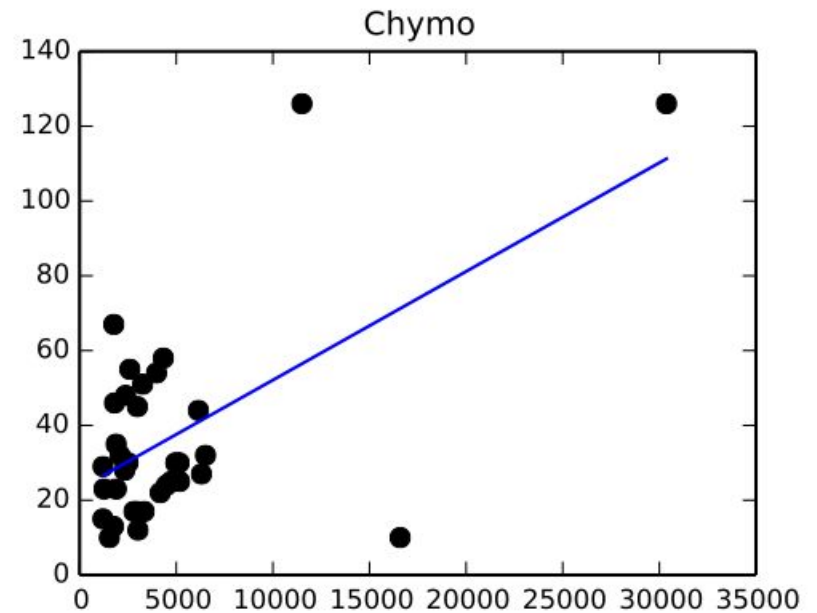
Pearson correlation $\rho = 0.5815353$



Pearson correlation $\rho = 0.6966817$



Pearson correlation $\rho = 0.5150688$



Pearson correlation $\rho = 0.5870782$

Спасибо за внимание