

## **Рассмотрение ВИЧ в контексте аутореактивности методом сравнительного анализа с СКВ**

Мыров Владислав,  
магистрант 2-го курса,  
Санкт-Петербургский Академический Университет

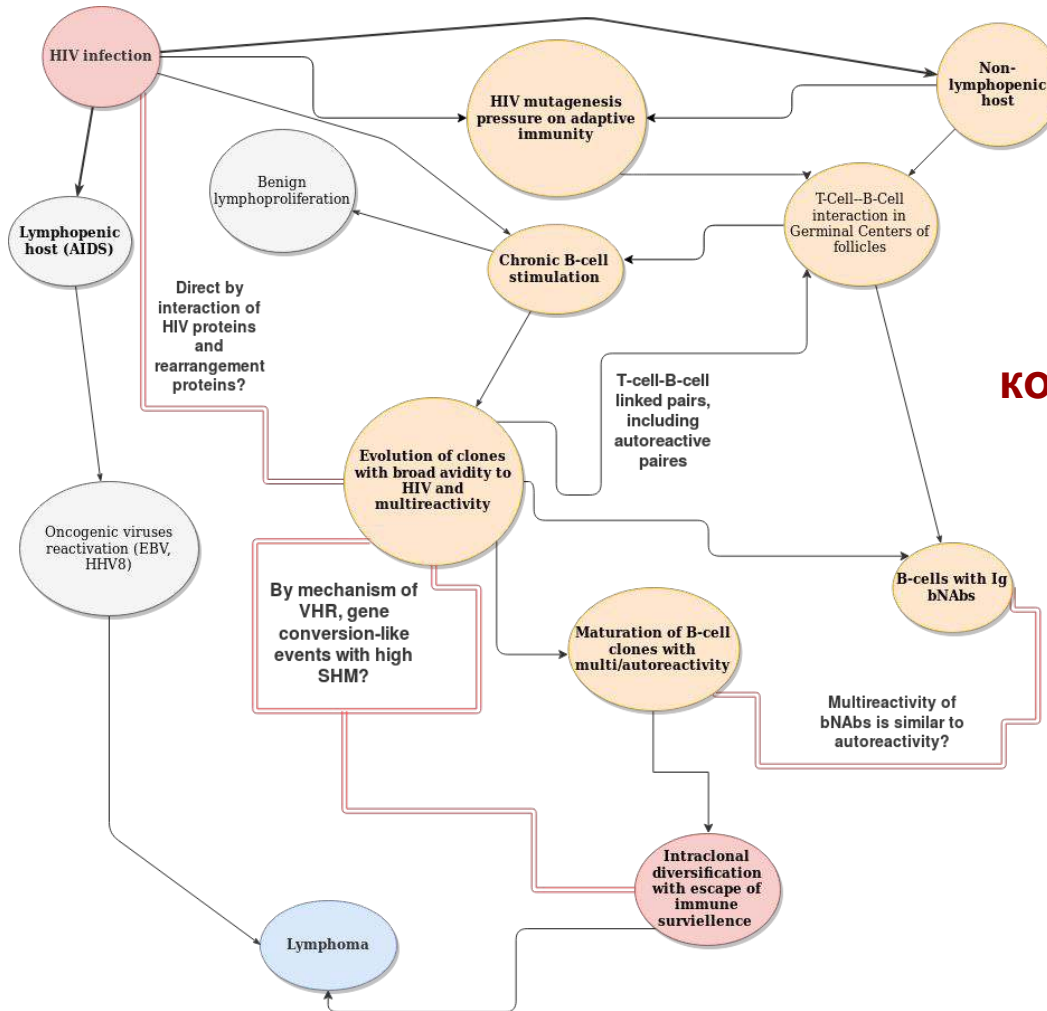
Айзсилниекс Оксана,  
врач-инфекционист,  
ПСПбГМУ им. Павлова  
НИИ ДОГиТ им. Р.М. Горбачевой

### **Руководитель**

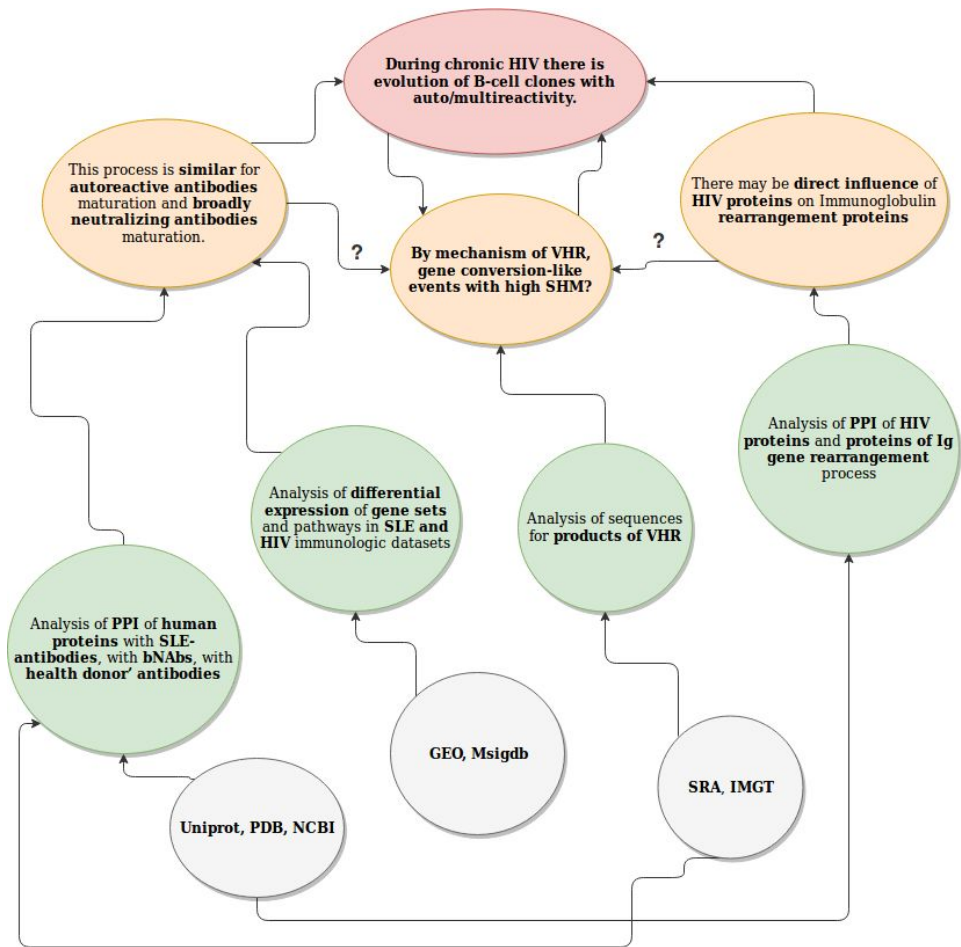
Лиознов Дмитрий Анатольевич, д.м.н.,  
заведующий кафедрой инфекционных болезней ПСПбГМУ

**Общая**

**КОНЦЕПЦИЯ**

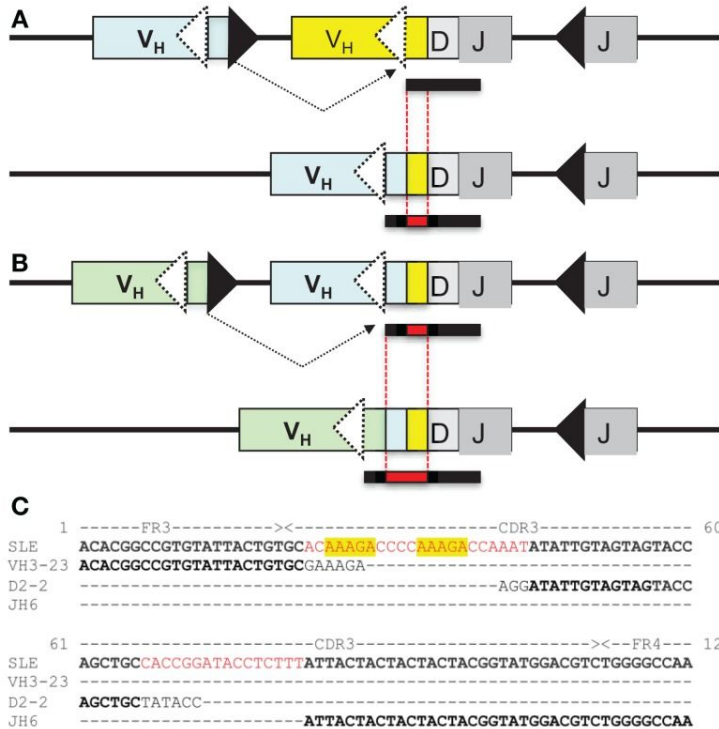


# Основные гипотезы, подходы для их проверки in silico, базы данных



**Анализ последовательностей  
антител на наличие продуктов VH  
replacement (footprint-  
последовательностей)**

# VH replacement



- впервые были обнаружены в knock-in мышинных моделях СКВ (3Н9 и 56R) и в клеточных линиях pre-B-ALL
- высокая вероятность ложноположительной footprint-последовательности (по числу случайных пентамеров из 4 нуклеотидов)
- относительные критерии: 1) отношение к N1 зоне CDR3, 2) отсутствие отношения к сегменту VH6-1

## Анализ VHR



**IgBlast (VDJ разметка, поиск аллелей)**

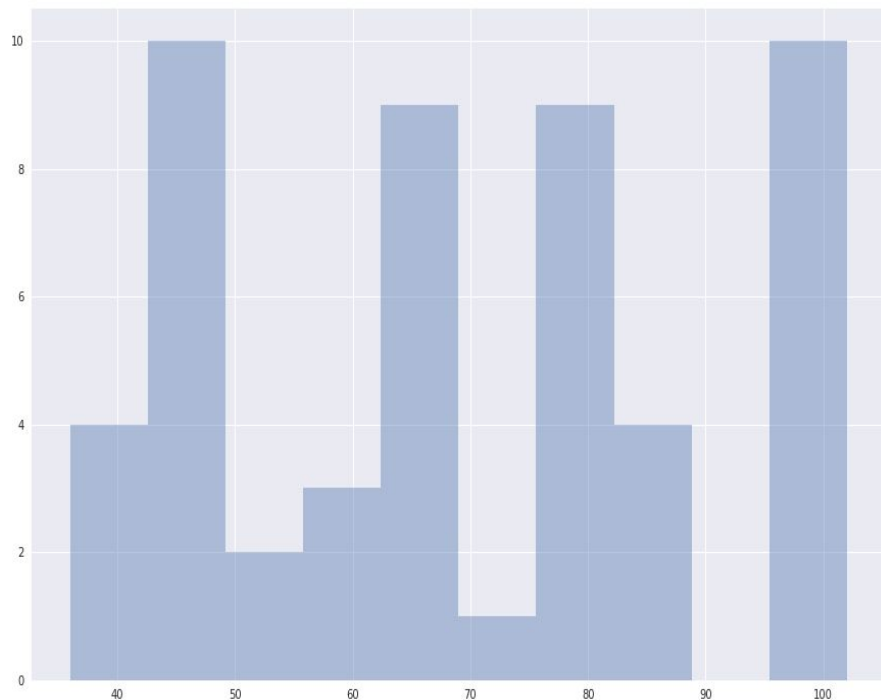


**IMGTV-QUEST (разметка CDR3, N1/2)**

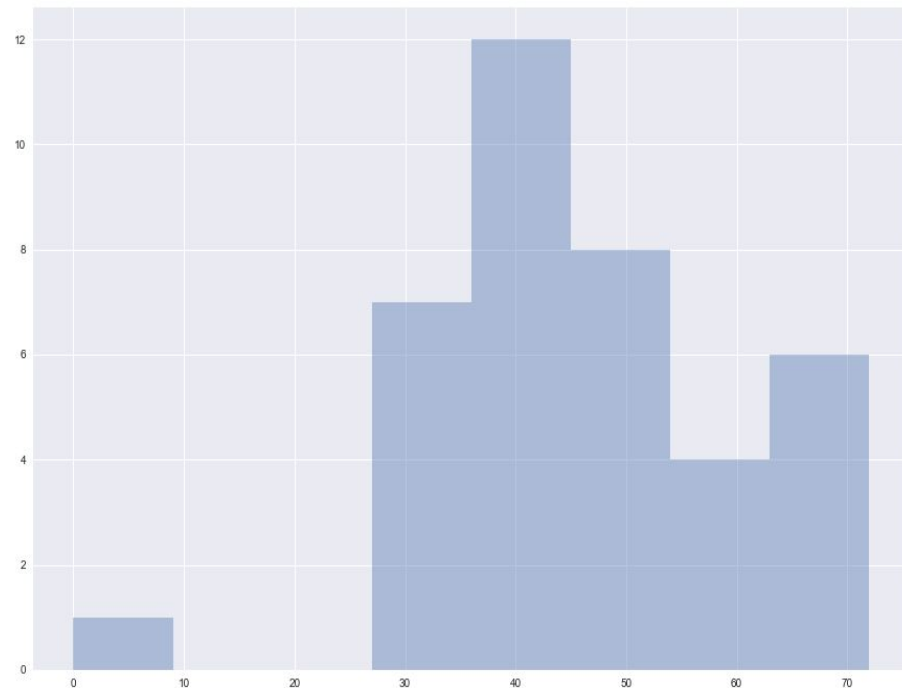


**Сборка IGH контрольной группы (IgConstructor)**

# Длины CDR3 в bNAbs и антителах при СКВ



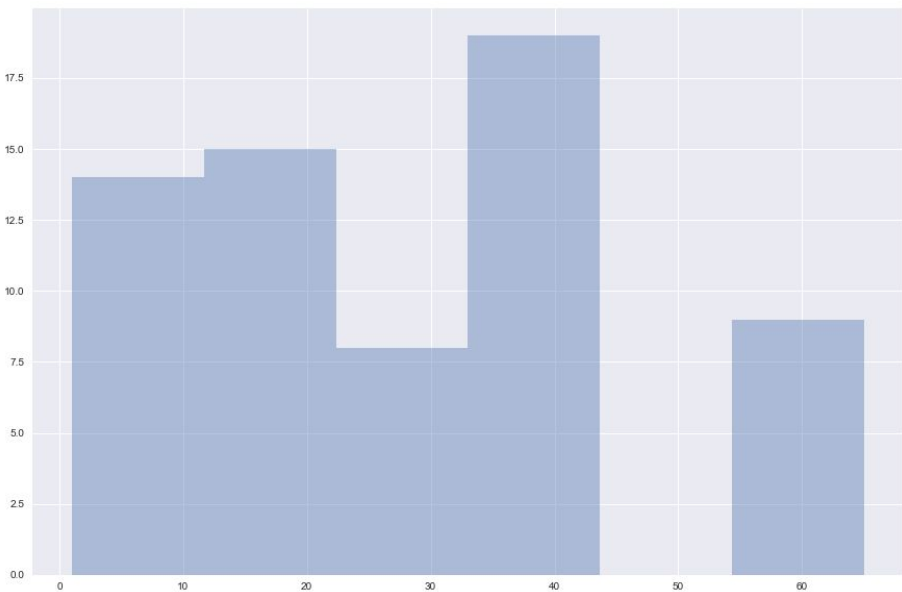
bNAbs



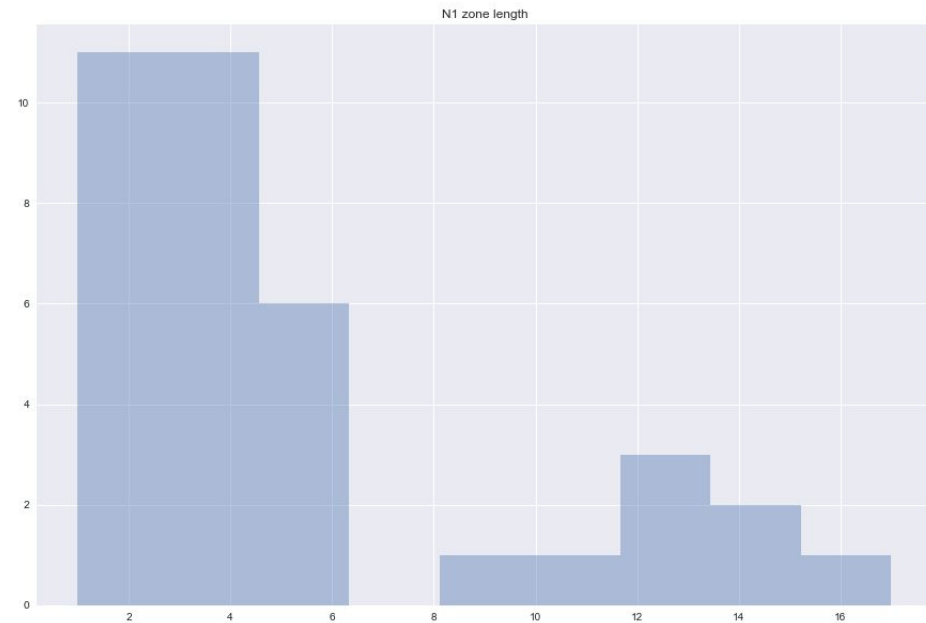
SLE

pvalue=0.0047

# Длины N1 зоны в bNAbs и антителах при СКВ



bNAbs

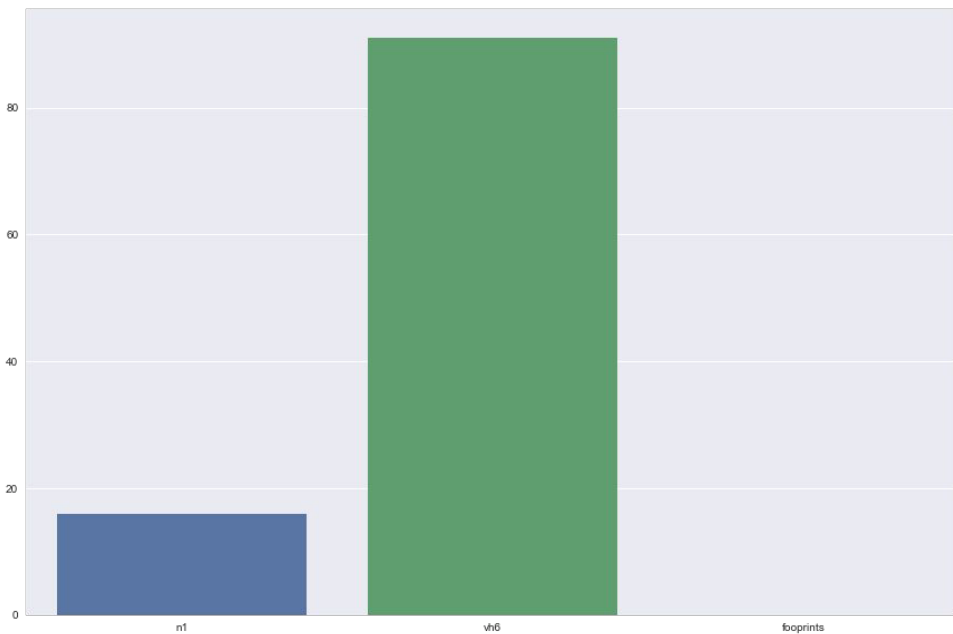


SLE

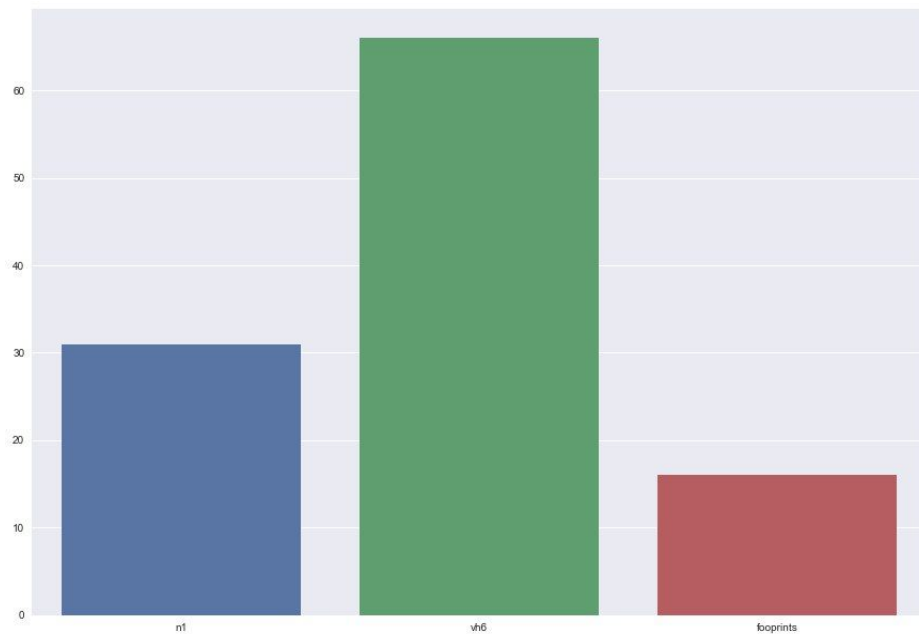
pvalue ~ 8.1e-13



# Критерии обнаружения footprint-последовательности



bNAbs



SLE

**Анализ РРІ трех групп антител (при СКВ, bNAbs, здоровый контроль) с белками человеческого организма**

**Анализ РРІ белков HIV и белков, участвующих в процессе реаранжировки генов Ig**

## Белки человека, к которым вырабатываются антитела при СКВ

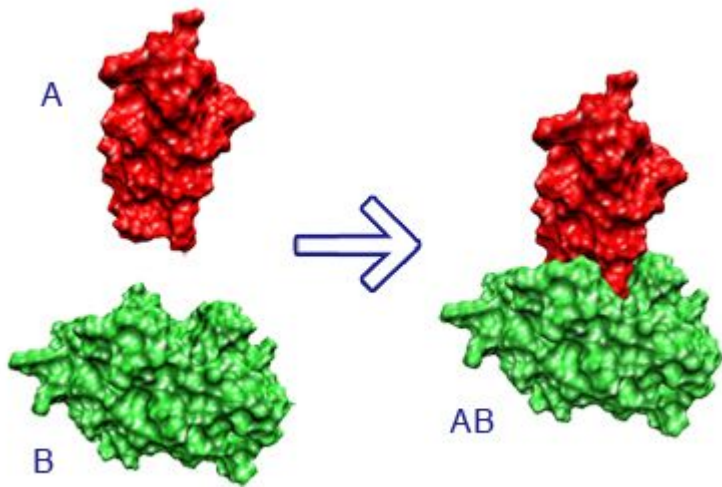
- SmB protein (<http://www.uniprot.org/uniprot/P14678>)
- SmD1 (<http://www.uniprot.org/uniprot/P62314>)
- SmD2 (<http://www.uniprot.org/uniprot/P62316>)
- SmD3 (<http://www.uniprot.org/uniprot/P62318>)
- SmE (<http://www.uniprot.org/uniprot/P62304>)
- SmF (<http://www.uniprot.org/uniprot/P62306>)
- SmG (<http://www.uniprot.org/uniprot/P62308>)
- Beta-2-glycoprotein 1 (<http://www.uniprot.org/uniprot/P02749>)
- SS-A/Ro ribonucleoprotein (<http://www.uniprot.org/uniprot/P10155>)
- Lupus La protein (SSB gene) (<http://www.uniprot.org/uniprot/P05455>)
- U3 small nucleolar ribonucleoprotein protein MPP10 (<http://www.uniprot.org/uniprot/O00566>)
- Major centromere autoantigen B (<http://www.uniprot.org/uniprot/P07199>)
- Centromere protein H (<http://www.uniprot.org/uniprot/Q9H3R5>)
- SCL protein (<http://www.uniprot.org/uniprot/Q9NY94>)
- Histidyl-tRNA synthetase (<http://www.uniprot.org/uniprot/Q52NV4>)

## Белки HIV-1: Env, Nef, Gag, Gag-Pol, Rev, Tat, Vif, Vpr, Vpu

## Белки реаранжировки гена Ig в В-лимфоцитах

- RAG-1 & RAG-2 (<http://www.uniprot.org/uniprot/P15918>, <http://www.uniprot.org/uniprot/P55895>, <http://www.rcsb.org/pdb/explore.do?structureId=4WWX>)
- Ku (это гетеродимер, состоящий из XRCC5/Ku70:XRCC6/Ku80) (XRCC5/Ku70 - <http://www.uniprot.org/uniprot/P13010>, <http://www.rcsb.org/pdb/explore.do?structureId=1jeY>, XRCC6/Ku80 - <http://www.uniprot.org/uniprot/P12956>, <http://www.rcsb.org/pdb/explore.do?structureId=1jeY>)
- Artemis (<http://www.uniprot.org/uniprot/Q96SD1>, <http://www.rcsb.org/pdb/gene/DCLRE1C?range=14964980&v=hg37>)
- DNA ligase IV (<http://www.rcsb.org/pdb/explore.do?structureId=3w1b>)
- XRCC4 (<http://www.rcsb.org/pdb/explore.do?structureId=1fu1>, <http://www.uniprot.org/uniprot/Q13426>)
- chromatin protein HMGB1 & HMGB2 (HMGB1 - <http://www.uniprot.org/uniprot/P09429>, HMGB2 - <http://www.uniprot.org/uniprot/P26583>)
- DNA-PK subunit (<http://www.uniprot.org/uniprot/P78527>, <http://www.rcsb.org/pdb/ngl/ngl.do?pdbid=3KGV&bioNumber=1>)
- Terminal deoxynucleotidyl transferase (TdT) (<http://www.uniprot.org/uniprot/P04053>)
- MSH2 and MSH6 (MSH2/6) (<http://www.uniprot.org/uniprot/P43246>, <http://www.uniprot.org/uniprot/P52701>)
- uracil-DNA glycosylase (UNG) (<http://www.uniprot.org/uniprot/P13051>)
- apurinic/apyrimidinic endonuclease 1 (APE1) (<http://www.uniprot.org/uniprot/P27695>)
- polymerase Pol η (<http://www.uniprot.org/uniprot/O75417>)
- BAFF (<http://www.uniprot.org/uniprot/Q5H8V1>)

## Два подхода



```
      10      20      30      40      50      60
MANLGCWMLV LFMATWSDLG LCKKRPKPGG WNTGGSRYPG QGSPGGNRYP PQGGGGWGQP
      70      80      90     100     110     120
HGGGWGQPHG GGWGQPHGGG WGQPHGGGWG QGGGTHSQWN KPSKPKTNMK HMAGAAAAGA
     130     140     150     160     170     180
VVGGLGGYML GSAMSRPIIH FGSDYEDRYY RENMHRYPNQ VYYRPMDEYS NQNNFVHDCV
     190     200     210     220     230     240
NITIKQHTVT TTKGENFTE TDVKMMERVV EQMCITQYER ESQAYYQRGS SMVLFSSPPV
     250
ILLISFLIFL IVG
```

Структурный

+ Точный

+ Информативный

- Очень медленный

Основанный на сиквенсе

+ Простой

+ Быстрый

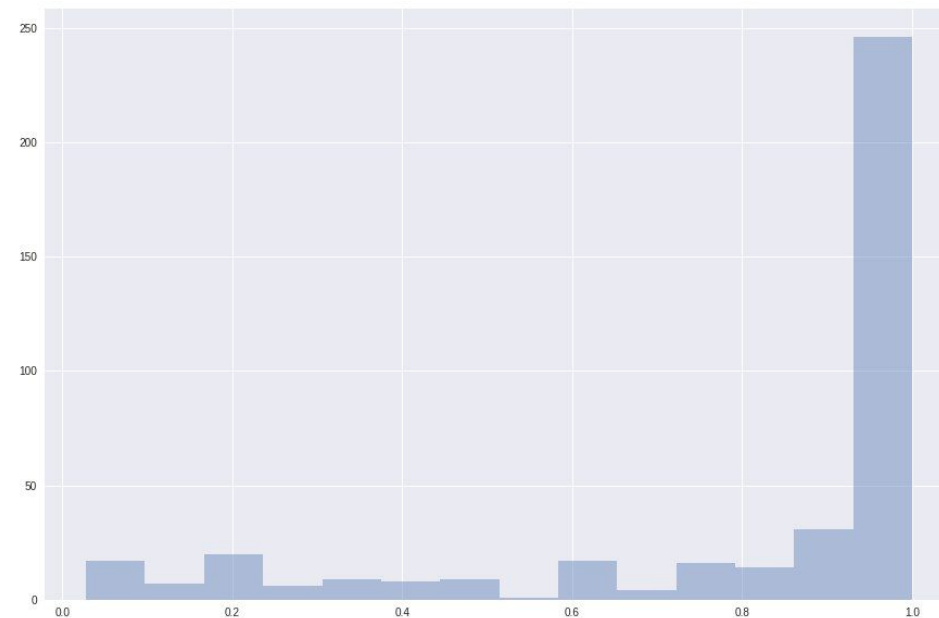
- Мало информации

# Тулы

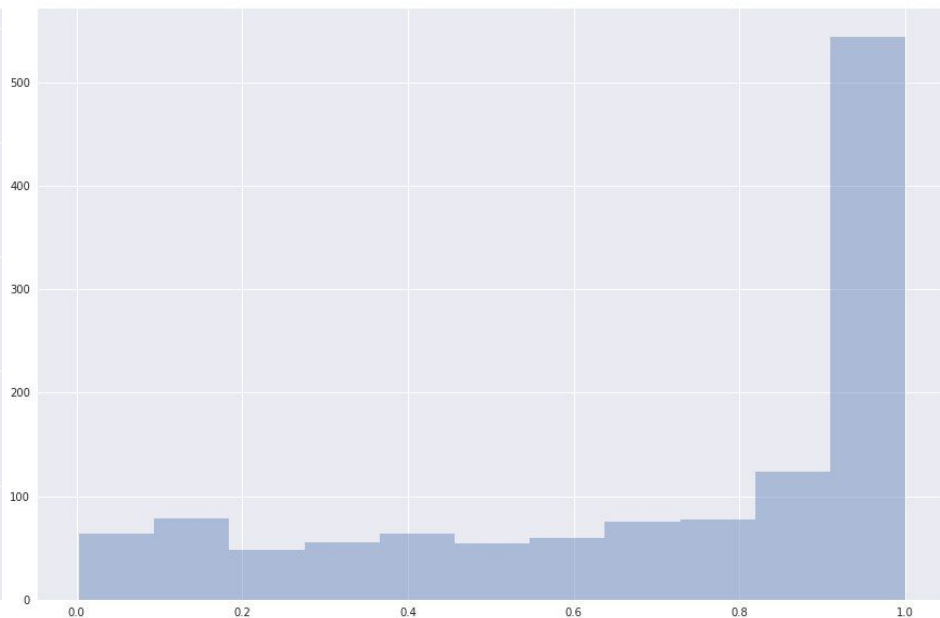
## Omictools\* - 102 программы!

Zdock	Структурный анализ	Считается state-of-art, но слишком долго	Структура
Prisma	Структурный анализ	самый информативный, но слишком долго	Возможные структуры + энергия
struct2net	Основан на структуре/сиквенсе	Только предподсчитанные взаимодействия	Logistic score
iLoops	Основан на анализе последовательностей	Быстро и довольно точно, но сервис работает в 20% случаев	Вероятность взаимодействия
LR_PPI	Основан на анализе последовательностей	Основан на LDA, довольно точно, использовали его	Вероятность взаимодействия

# Гистограмма вероятностей

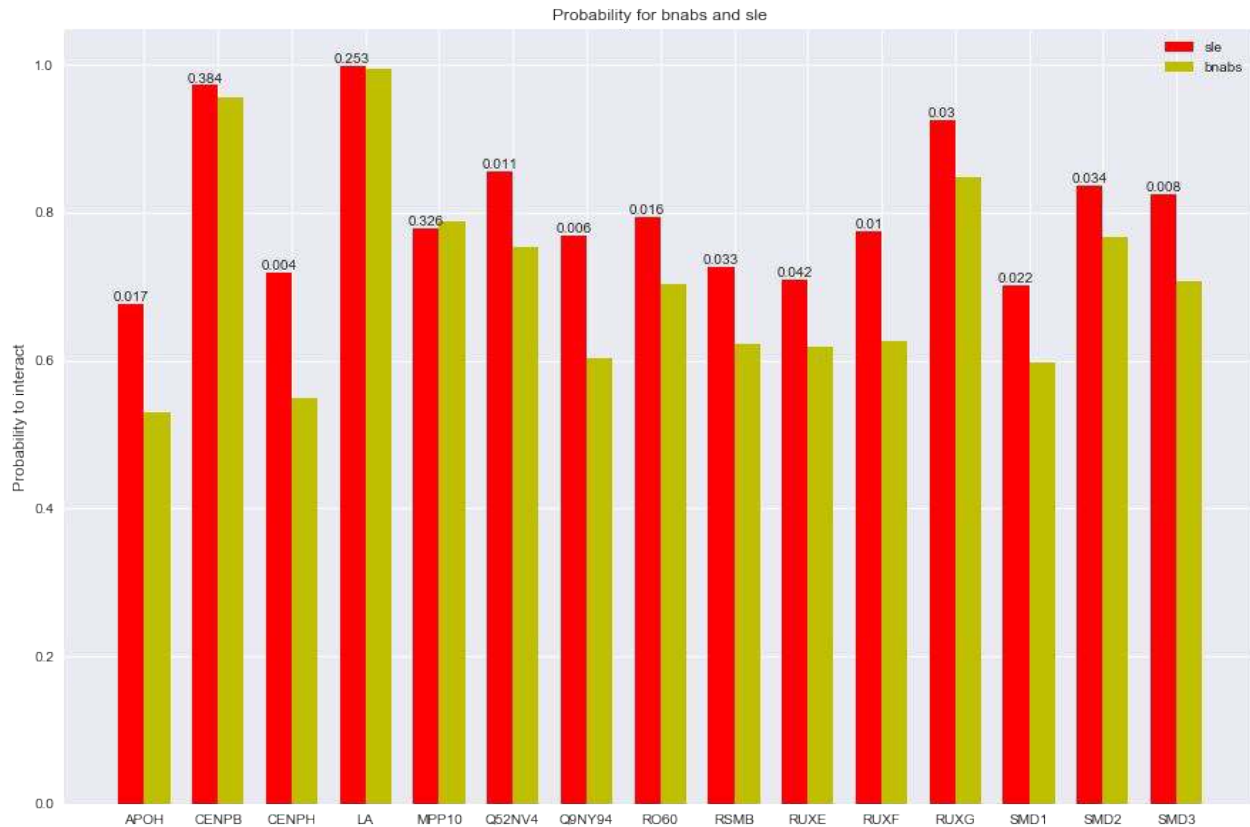


bNAbs

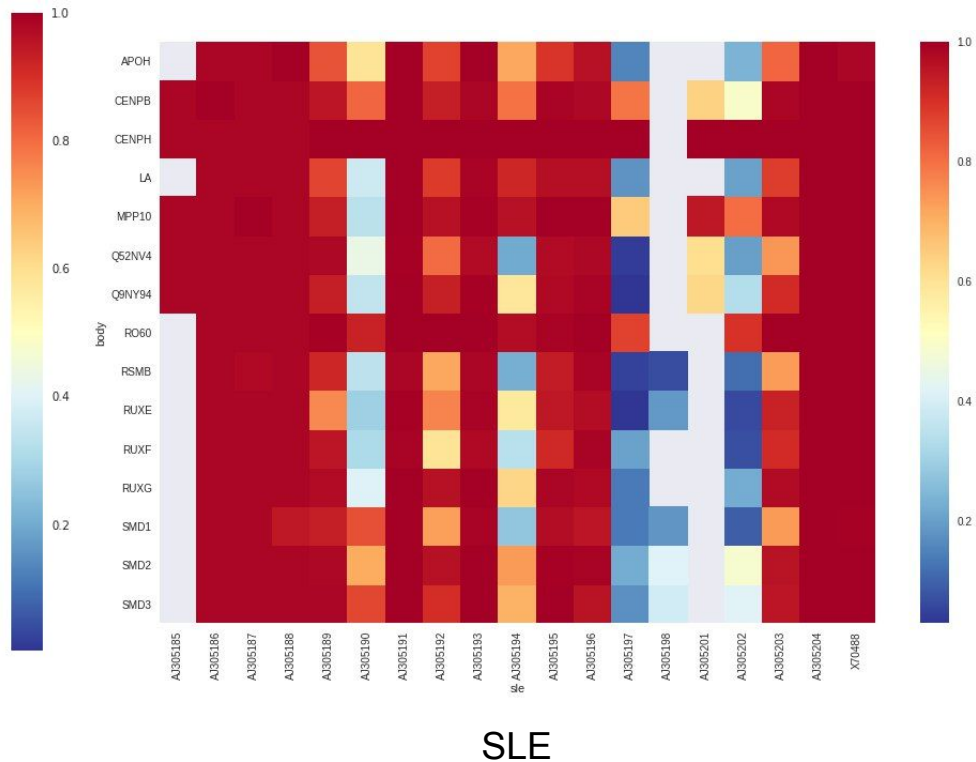
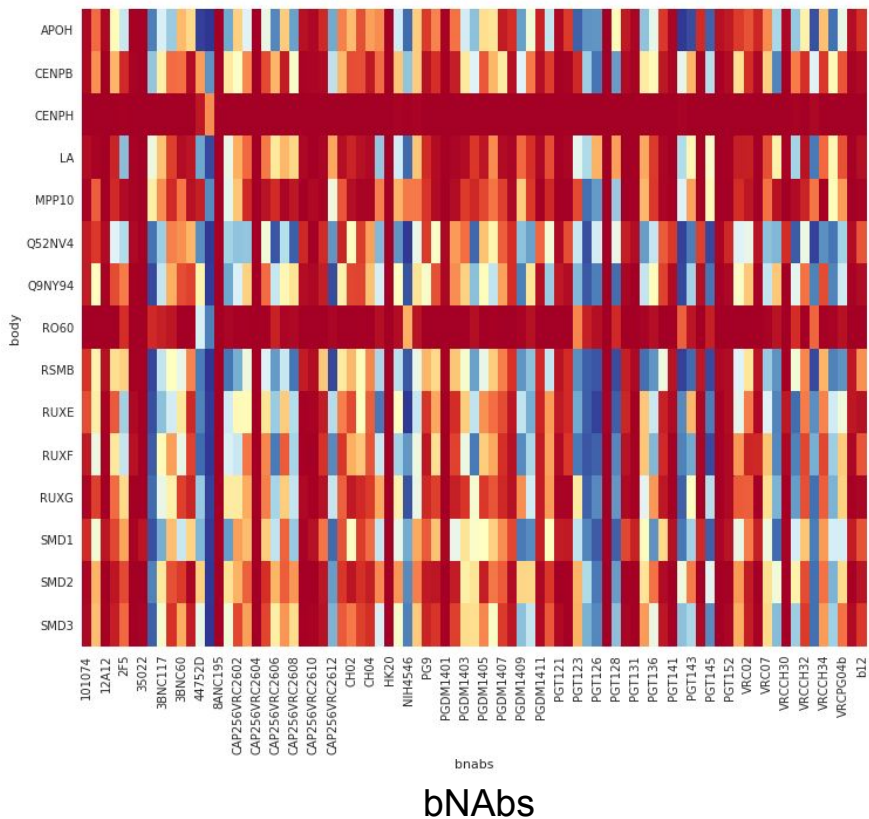


SLE

# Вероятности по белкам



# Попарные вероятности





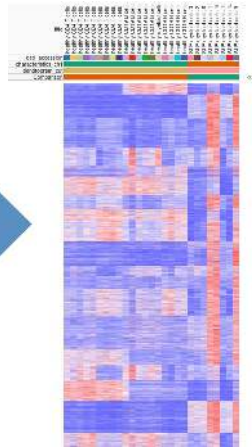
**Анализ дифференциальной  
экспрессии множеств генов и генов  
сигнальных путей в СКВ и ВИЧ  
иммунологических датасетах**

## Оценка схожести заболеваний путем сравнения ассоциированных списков

1. Формируется **множество заболеваний**, для которых существуют доступные данные microarray для дифференциальной экспрессии
2. Для каждого заболевания рассчитывается **список TOP-N генов** с наиболее выраженной дифференциальной экспрессией (напр. при помощи Phantasus)
3. При помощи взвешенного индекса Жаккара находятся **попарные расстояния между списками** (заболеваниями)
4. Попадут ли SLE и EC HIV в один кластер?

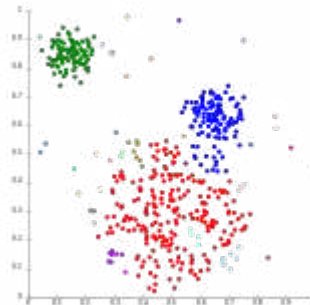
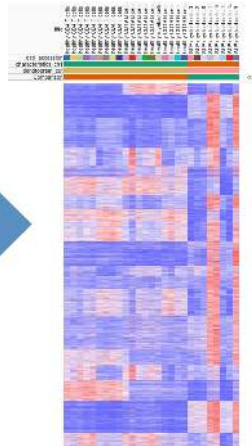
# Оценка схожести заболеваний путем сравнения ассоциированных списков

1. Формируется **множество заболеваний**, для которых существуют доступные данные microarray для дифференциальной экспрессии
2. Для каждого заболевания рассчитывается **список TOP-N генов** с наиболее выраженной дифференциальной экспрессией (напр. при помощи Phantastus)
3. При помощи взвешенного индекса Жаккара находятся **попарные расстояния между списками** (заболеваниями)
4. Попадут ли SLE и EC HIV в один кластер?



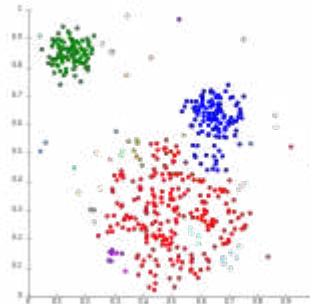
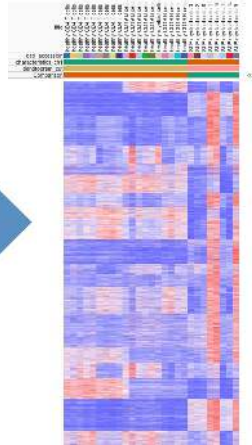
# Оценка схожести заболеваний путем сравнения ассоциированных списков

1. Формируется **множество заболеваний**, для которых существуют доступные данные `microarray` для дифференциальной экспрессии
2. Для каждого заболевания рассчитывается **список TOP-N генов** с наиболее выраженной дифференциальной экспрессией (напр. при помощи Phantasus)
3. При помощи взвешенного индекса Жаккара находятся **попарные расстояния между списками** (заболеваниями)
4. Попадут ли SLE и EC HIV в один кластер?



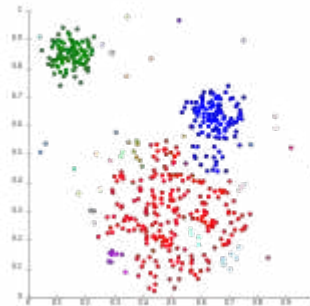
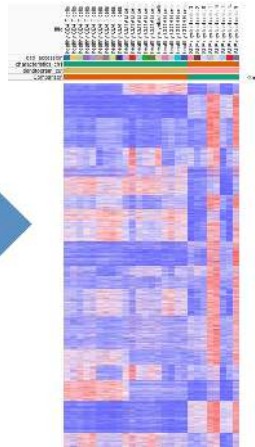
# Оценка схожести заболеваний путем сравнения ассоциированных списков

1. Формируется **множество заболеваний**, для которых существуют доступные данные microarray для дифференциальной экспрессии
2. Для каждого заболевания рассчитывается **список TOP-N генов** с наиболее выраженной дифференциальной экспрессией (напр. при помощи **Phantasus**)
3. При помощи взвешенного индекса Жаккара находятся **попарные расстояния между списками** (заболеваниями)
4. Попадут ли SLE и EC HIV в один кластер?



# Оценка схожести заболеваний путем сравнения ассоциированных списков

1. Формируется **множество заболеваний**, для которых существуют доступные данные microarray для дифференциальной экспрессии
2. Для каждого заболевания рассчитывается **список TOP-N пасвеев** с наиболее выраженной дифференциальной экспрессией (напр. при помощи **GSEA**)
3. При помощи взвешенного индекса Жаккара находятся **попарные расстояния между списками** (заболеваниями)
4. Попадут ли SLE и EC HIV в один кластер?



## Предстоящие задачи

- 1) Предсказать binding sites (сейчас считаются), посмотреть на возможные кластера
- 2) Попробовать другие тулы, более сложные (NN, считать структурно)
- 3) Возможное влияние footprint-последовательностей на структуру белков
- 4) Нетривиальный анализ последовательностей антител на VHR, gene conversion-like events и D-D fusion
- 5) Разработать валидный статистический тест для оценки получившегося сходства дифференциальной экспрессии множеств генов при разных заболеваниях
- 6) Проанализировать целевые датасеты из закрытой базы данных dbGap
- 7) Дизайн лабораторного эксперимента и клинического исследования на основе полученных данных

## **Благодарности:**

- **Институту Биоинформатики** за умение мотивировать
- **Евгению Бакину** за консультирование по сложным вопросам
- **ЦАБ СПбГУ** за привитый вкус к вопросам нетривиальной иммунологии