



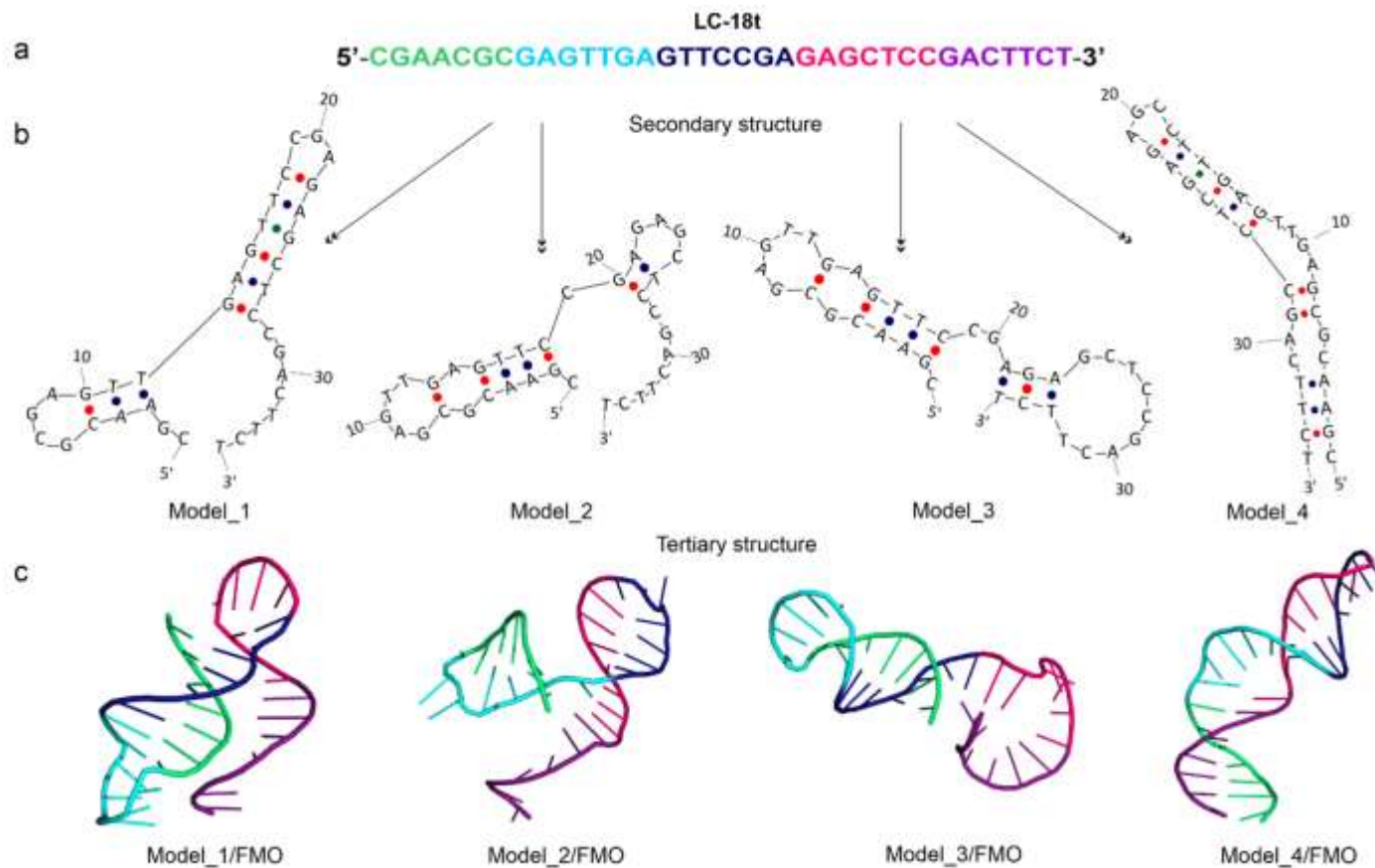
Высокопроизводительное секвенирование повышает качество селекции аптамеров (цифровых лекарств)

Дмитрий Вепринцев

к.ф.-м.н., с.н.с. лаборатории цифровых
управляемых лекарств и тераностики
ФИЦ КНЦ СО РАН

27 октября 2021

Аптамеры – короткие одноцепочечные ДНК



Molecular Therapy
Nucleic Acids
Original Article

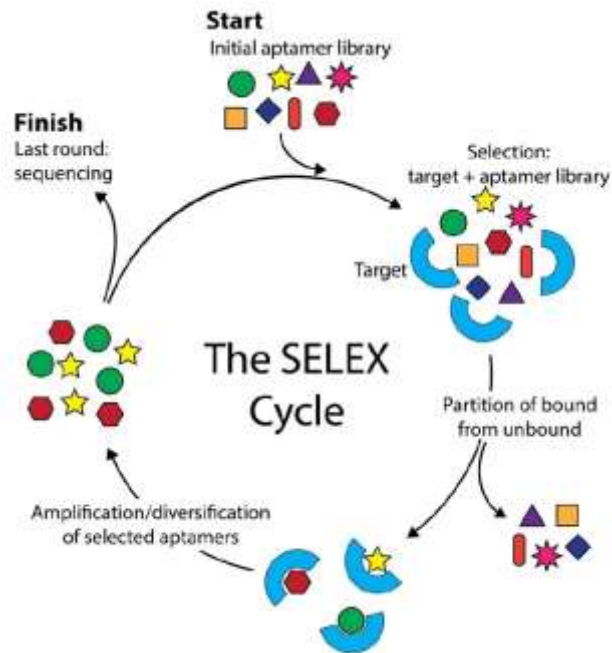
The role of SAXS and molecular simulations
in 3D structure elucidation of a DNA aptamer
against lung cancer

Dmitry Shvetsov,¹ Vladimir Mamonov,¹ Roman V. Mironchikov,^{1,2} Irina A. Bolshagina,^{1,2,3}
Svetlana V. Artyushenko,^{1,2,3} Galina S. Zolotareva,^{1,2} Olga S. Sokolovskaya,^{1,2} Ekaterina N. Zolotareva,¹ Alexey V. Kozlov,¹
Dmitry S. Malozemov,¹ Vladislav M. Zabludov,¹ Dmitriy V. Vorobiov,¹ Alexey E. Solov'ev,^{1,2} Radka A. Zakharova,¹
Marina V. Berezinskaya,¹ Elena N. Turchik,^{1,2} Dmitriy G. Polozov,¹ Yuriy Mironov,^{1,2} and Anna S. Kochubei^{1,2}

<https://doi.org/10.1016/j.omtn.2021.07.015>

SELEX – эволюция «в пробирке»

SELEX - Systematic evolution of ligands by exponential enrichment



Next-Generation Sequencing



Data processing



Kinghorn A, Fraser L, Lang S, Shiu S and Tanner J. 2017. Aptamer Bioinformatics. *Int J Mol Sci*, 18, 2516.

Обработка NGS-данных SELEX включает несколько этапов

1. Демультиплексирование

2. Фильтрация по длине

3. Фильтрация по качеству

4. Удаление праймеров

5. Подсчет числа копий каждой последовательности

6. Анализ обогащения

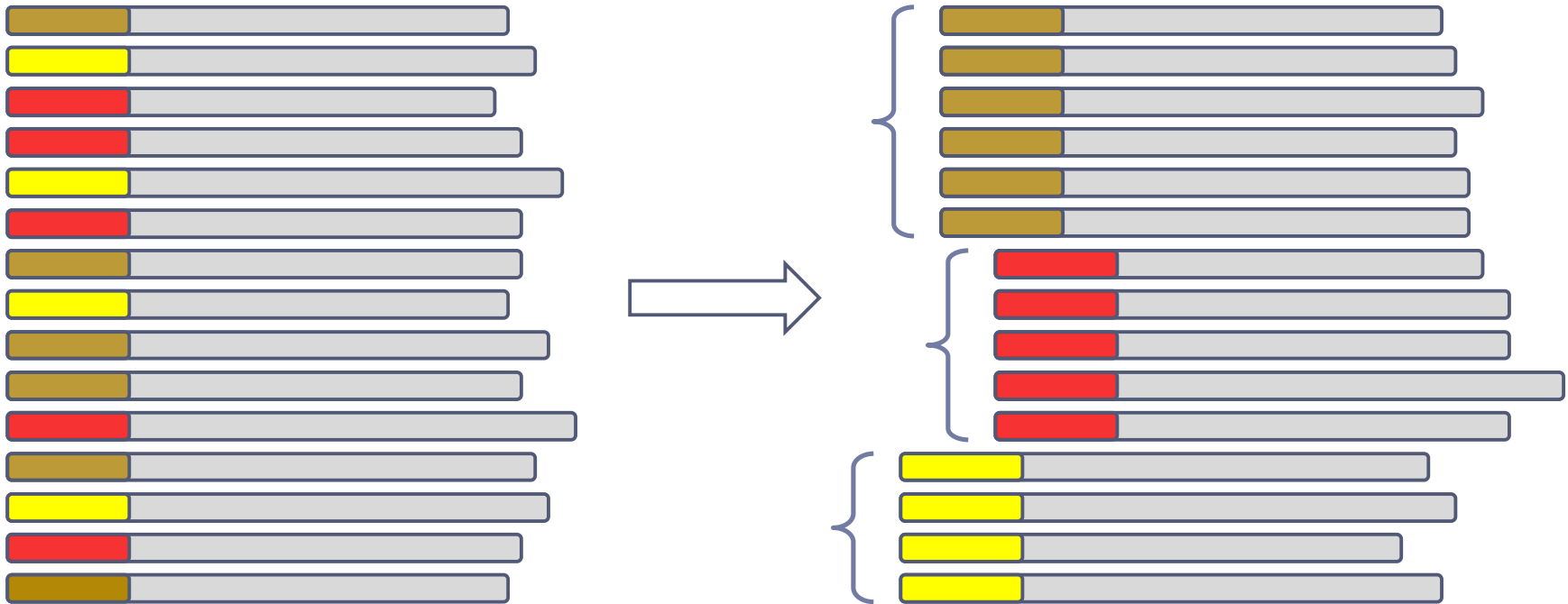
7. Кластеризация

1. Демультимплексирование

2/2

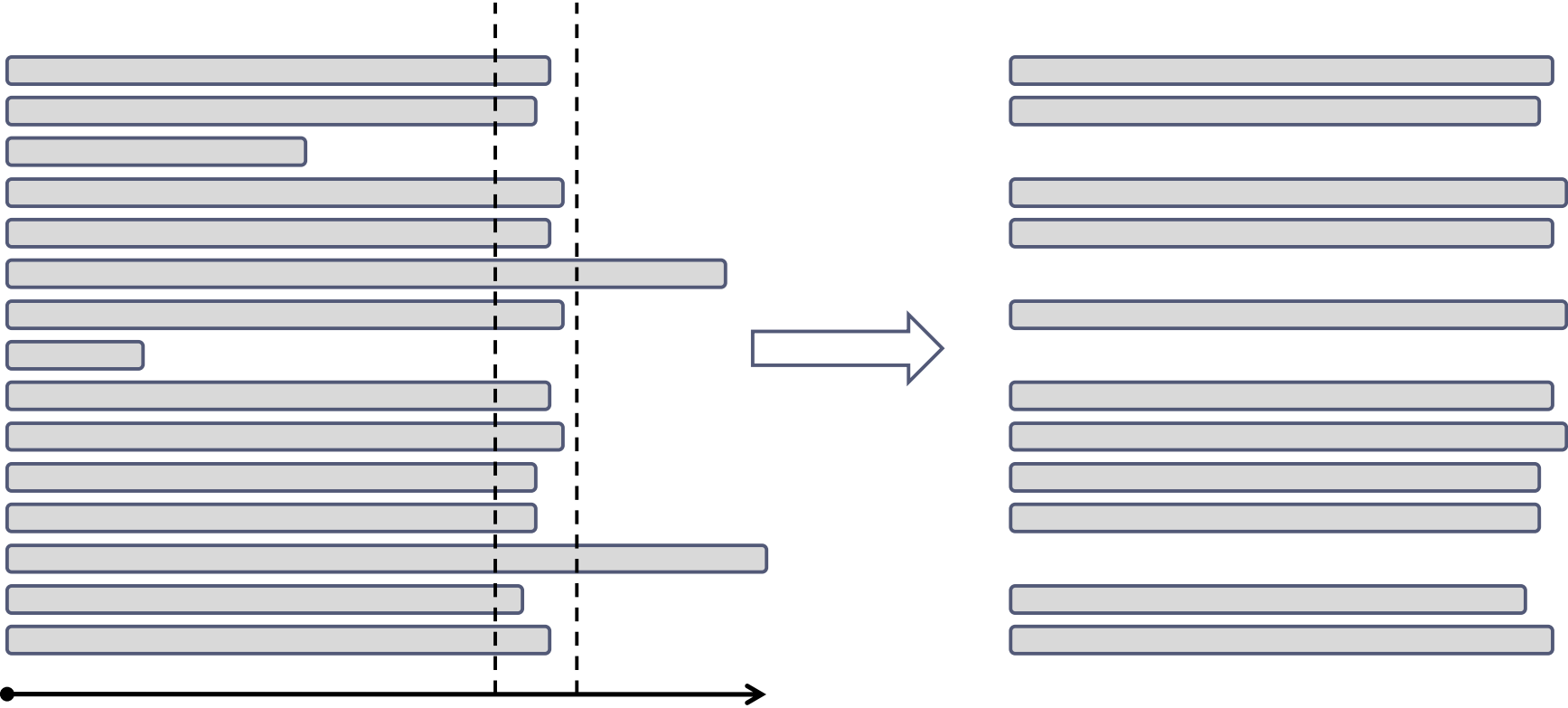
GCACG CTCCTCTGACTGTAACCACGGTGCACAGGGGGGAGCGCGCGTGGGGTGTGCGGGTTCGGTGGCATAGGTAGTCCAGAAGCC

barcode



2. Фильтрация по длине

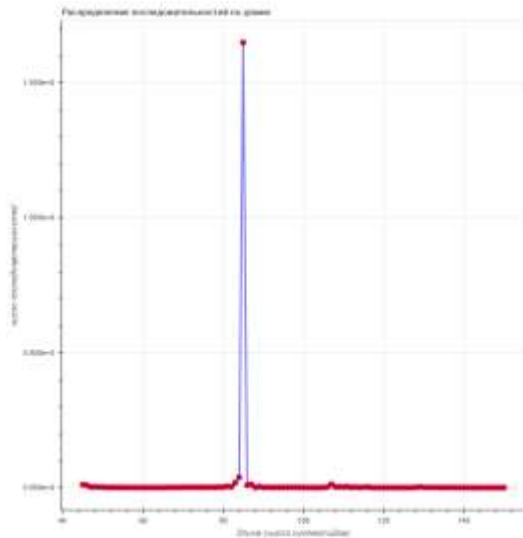
1/2



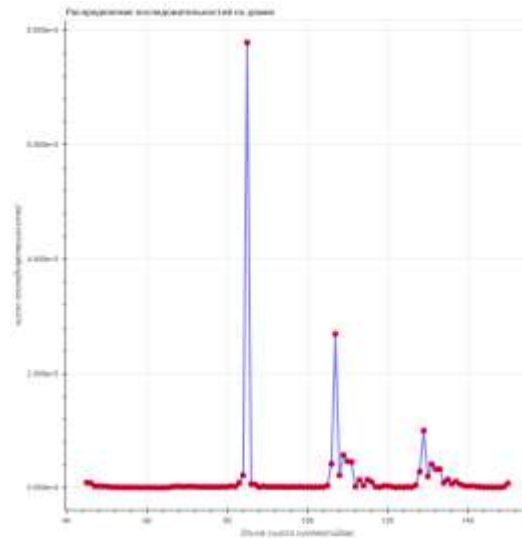
2. Фильтрация по длине

2/2

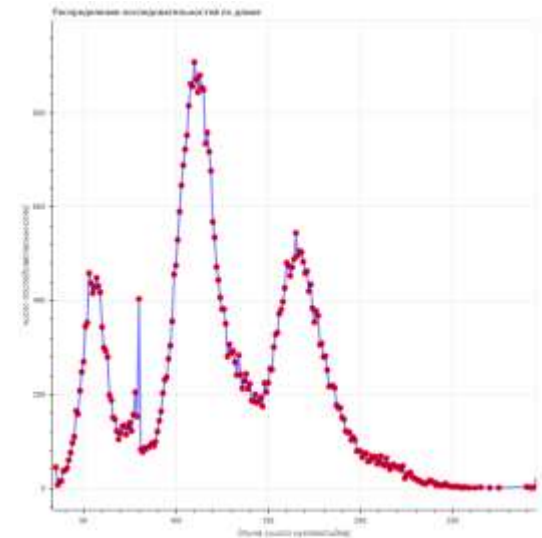
Идеальное распределение



Хорошее распределение

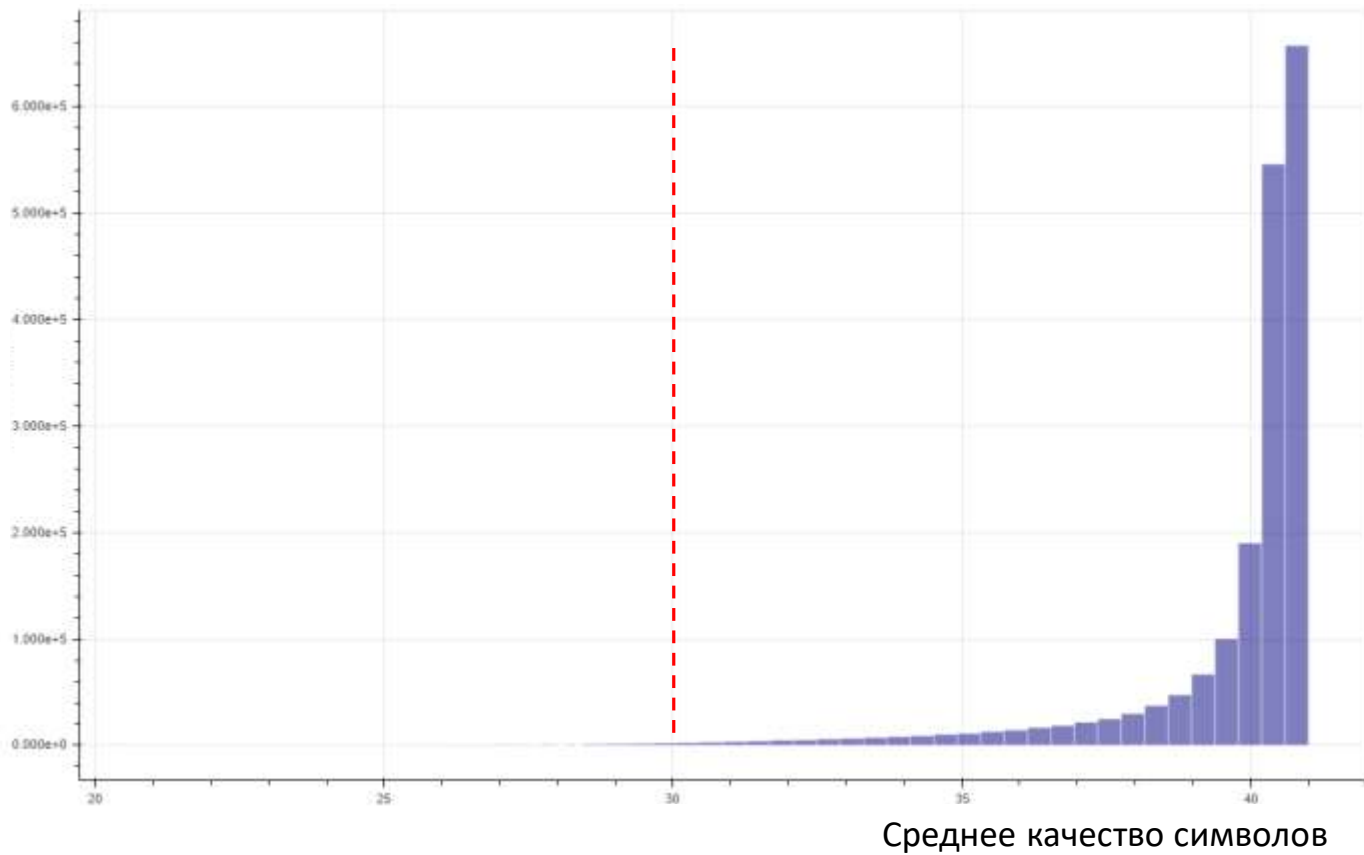


Плохое распределение



3. Фильтрация по качеству

Удаление цепочек со средним качеством символов ниже порога (например, Q30)



4. Удаление праймеров



5. Подсчет числа копий каждой последовательности

Пример: TOP типичного пула аптамеров

№	последовательность	число копий, шт	доля	доля, ppm
1	ACTGAGCTTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	86074	0,14992	149920,2
2	TGCCCCAACGCGAGTTGAGTTCGAGAGCTCCGACTTCTT	21050	0,036664	36664,0
3	CGCGGTGAAGGGTATATCCACTGCGTCCCGTGCCGTCGGT	2149	0,003743	3743,0
4	ACTGAGCTTTGGTGGGTCGAGCTATGTGCCGCTCTTCAC	749	0,001305	1304,6
5	ACTGAGCTTTGGCGGGTCGAGCTATGTGCTGCCTCTTCAC	619	0,001078	1078,1
6	ACTGAGCTTTGGTGGGTTGAGCTATGTGCTGCCTCTTCAC	553	0,000963	963,2
7	GCTGAGCTTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	421	0,000733	733,3
8	ACTGAGCTTTGGTGGGTCGAGCTATGTGCTGTCTCTTCAC	397	0,000691	691,5
9	ACTGAGCTTTGGTGGGTCGAGCTATGTGTTGCCTCTTCAC	381	0,000664	663,6
10	ACTGAGCCTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	376	0,000655	654,9
11	ACTGAGCTTTGGTGGGTCGAGCTACGTGCTGCCTCTTCAC	349	0,000608	607,9
12	ACTGAGCTTTGGTGGGTCGAGTTATGTGCTGCCTCTTCAC	340	0,000592	592,2
13	AAGGTGTCGGCCTTAGTAAGGCTACAGCCAAGGGAACGTA	334	0,000582	581,7
...				

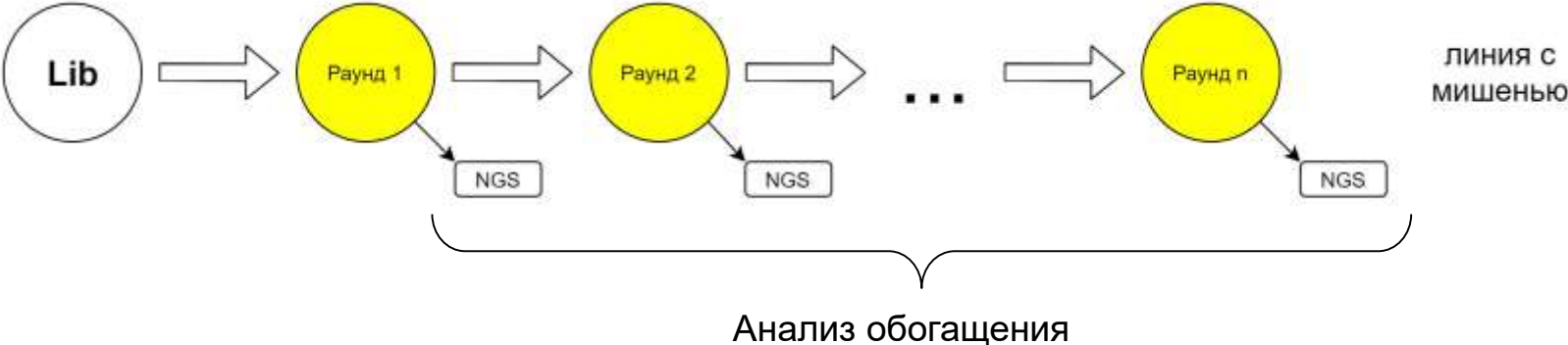
Всего различных последовательностей: 361217

Суммарное число копий: 574132

NGS-SELEX уровень 0



NGS-SELEX уровень 1



6. Анализ обогащения

Мера стабильности обогащения:

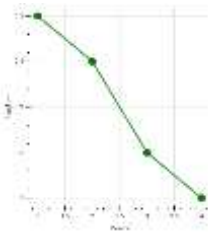
n - число раундов

f_k - доля аптамера A в раунде k

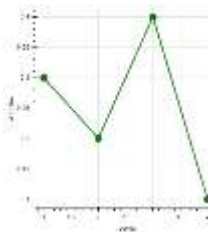
$$R(A) = \frac{2 \sum_{i < j} I[f_i < f_j]}{n(n-1)}, \quad 0 \leq R(A) \leq 1$$

$$I[f_i < f_j] = \begin{cases} 1, & \text{если } f_i < f_j, \\ 0, & \text{если } f_i \geq f_j \end{cases}$$

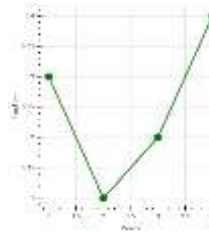
R=0



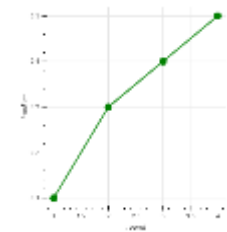
R=2/6



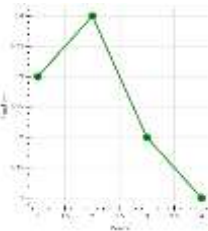
R=4/6



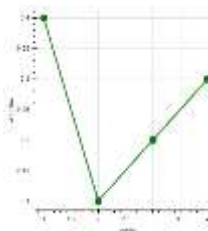
R=1



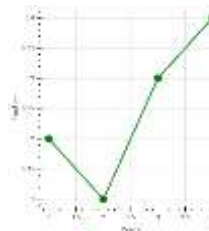
R=1/6



R=3/6

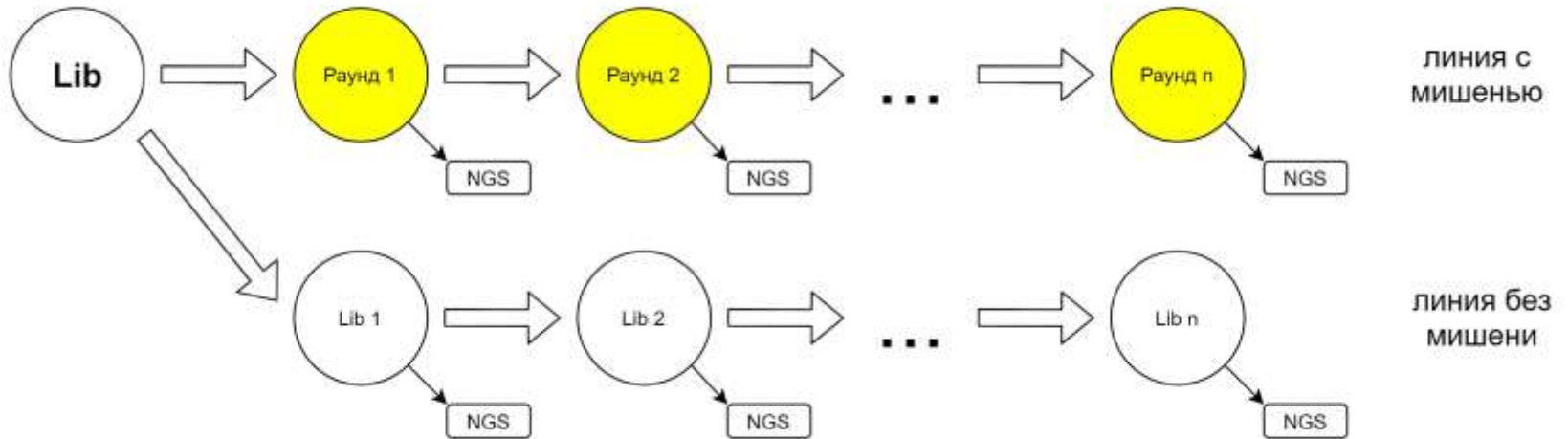


R=5/6

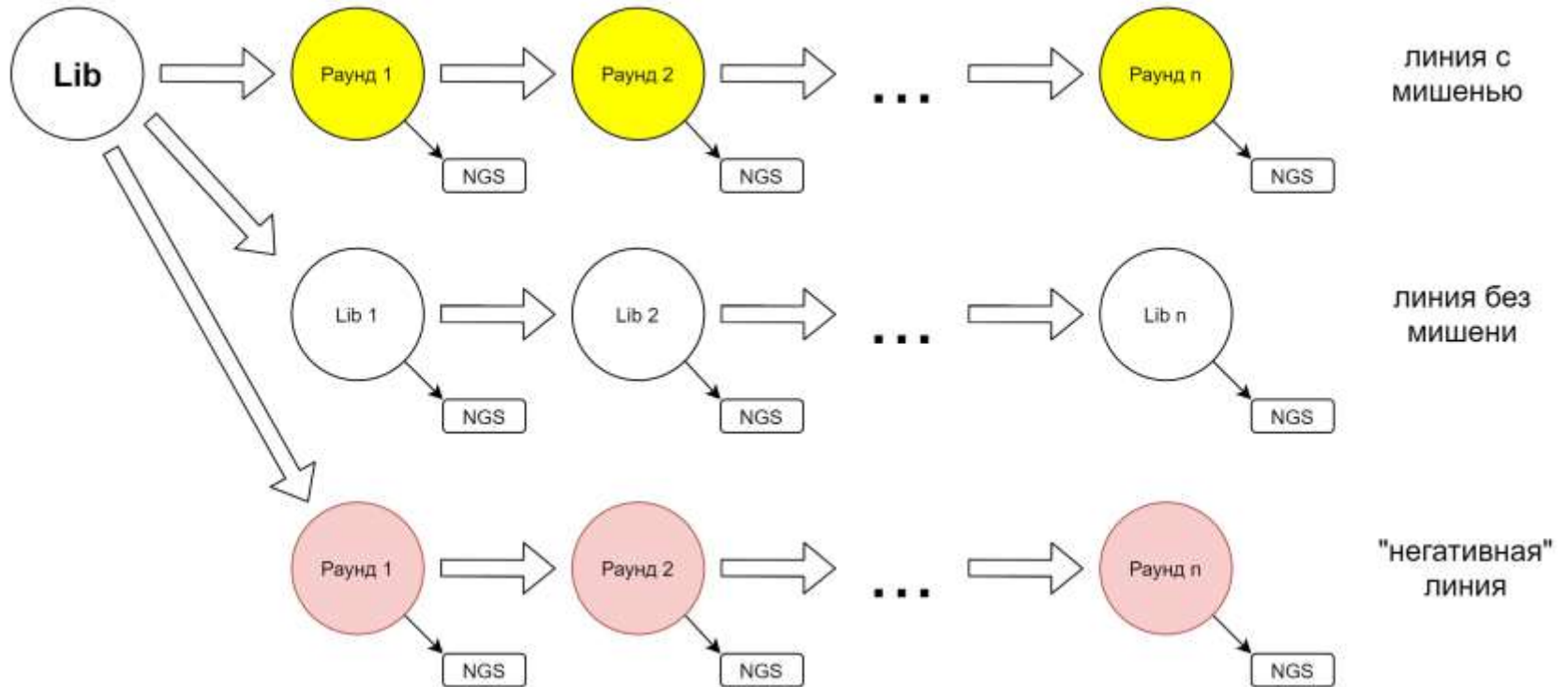


Мера скорости обогащения: $v(A) = \text{median}_{f_j \neq 0} \left\{ \frac{f_{j+1}}{f_j} \right\}$

NGS-SELEX уровень 2



NGS-SELEX уровень 2+



Что мы ждем от «хорошего» аптамера?

1. Его достаточно много в итоговом пуле селекции
2. Его достаточно мало в соответствующем пуле пустой селекции
3. Он стабильно обогащается в линии с мишенью
4. Он заметно хуже обогащается в пустой линии или в линии селекции к другой мишени (негативная фаза)

7. Кластеризация

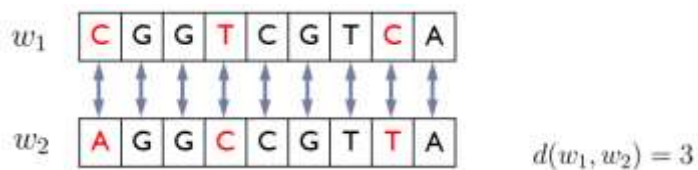
1/2

Кластер – набор «похожих» аптамеров

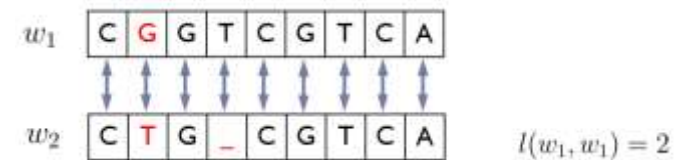
№	последовательность	число копий, шт	доля	доля, ppm
1	ACTGAGCTTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	86074	0,14992	149920,2
2	TGCCCCAACGCGAGTTGAGTTCCGAGAGCTCCGACTTCTT	21050	0,036664	36664,0
3	CGCGGTGAAGGGTATATCCACTGCGTCCCCTGCCGTCGGT	2149	0,003743	3743,0
4	ACTGAGCTTTGGTGGGTCGAGCTATGTGCCGCCTCTTCAC	749	0,001305	1304,6
5	ACTGAGCTTTGGCGGGTCGAGCTATGTGCTGCCTCTTCAC	619	0,001078	1078,1
6	ACTGAGCTTTGGTGGGTTGAGCTATGTGCTGCCTCTTCAC	553	0,000963	963,2
7	GCTGAGCTTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	421	0,000733	733,3
8	ACTGAGCTTTGGTGGGTCGAGCTATGTGCTGTCTCTTCAC	397	0,000691	691,5
9	ACTGAGCTTTGGTGGGTCGAGCTATGTGTTGCCTCTTCAC	381	0,000664	663,6
10	ACTGAGCCTTGGTGGGTCGAGCTATGTGCTGCCTCTTCAC	376	0,000655	654,9
11	ACTGAGCTTTGGTGGGTCGAGCTACGTGCTGCCTCTTCAC	349	0,000608	607,9
12	ACTGAGCTTTGGTGGGTCGAGTTATGTGCTGCCTCTTCAC	340	0,000592	592,2
13	AAGGTGTCGGCCTTAGTAAGGCTACAGCCAAGGGAACGTA	334	0,000582	581,7
...				

Как измерить «похожесть»?

Расстояние Хэмминга:



Расстояние Левенштейна:



7. Кластеризация

Исследование «ландшафта» кластера

Core sequence →

ID	N40 sequence	Round 5 prevalence
#411-J	AGCCATGACGATGTCGTTACGTAGATGCAGAGACTCCTAA	6311
#411-J-A1G	GGCCATGACGATGTCGTTACGTAGATGCAGAGACTCCTAA	47
#411-J-G7T	AGCCATTACGATGTCGTTACGTAGATGCAGAGACTCCTAA	25
#411-J-A19C		none
#411-J-A19G	AGCCATGACGATGTCGTTGCGTAGATGCAGAGACTCCTAA	3
#411-J-A19T	AGCCATGACGATGTCGTTTCGTAGATGCAGAGACTCCTAA	2
#411-J-C28T	AGCCATGACGATGTCGTTACGTAGATGTAGAGACTCCTAA	17
#411-J-G30A	AGCCATGACGATGTCGTTACGTAGATGCAAGACTCCTAA	9
#411-J-G30C	AGCCATGACGATGTCGTTACGTAGATGCACAGACTCCTAA	3
#411-J-G30T	AGCCATGACGATGTCGTTACGTAGATGCATAGACTCCTAA	46
#411-J-A39G	AGCCATGACGATGTCGTTACGTAGATGCAGAGACTCCTGA	5
#411-J-A39C		none

← Frequent

← Average frequency

← Rare

Identifying high-affinity aptamer ligands with defined cross-reactivity using high-throughput guided systematic evolution of ligands by exponential enrichment

ARTICLE | NUCLEIC ACIDS RESEARCH - MAY 2015
Impact Factor: 8.11 | DOI: 10.1093/nar/gkv434 | Source: PubMed

8 AUTHORS, INCLUDING:

 **Russhel Brantman**
University of Miami Miller School of Medicine
20 PUBLICATIONS | 375 CITATIONS
[SEE PROFILE](#)

 **Teresa Przytycka**
National Institutes of Health
150 PUBLICATIONS | 2,557 CITATIONS
[SEE PROFILE](#)

 **Alexey Derzhavsky**
University of Miami
13 PUBLICATIONS | 43 CITATIONS

NGS-SELEX уровень 3

машинное обучение ?



to be continued...