

Drug discovery: The Process



Rabia Khan

- 하나의 약을 생산하는데 있어 매우 painful, 오랜 시간이 걸림

AI for Drug Discovery,
Biomarker Development
and Advanced R&D
Landscape / 2020



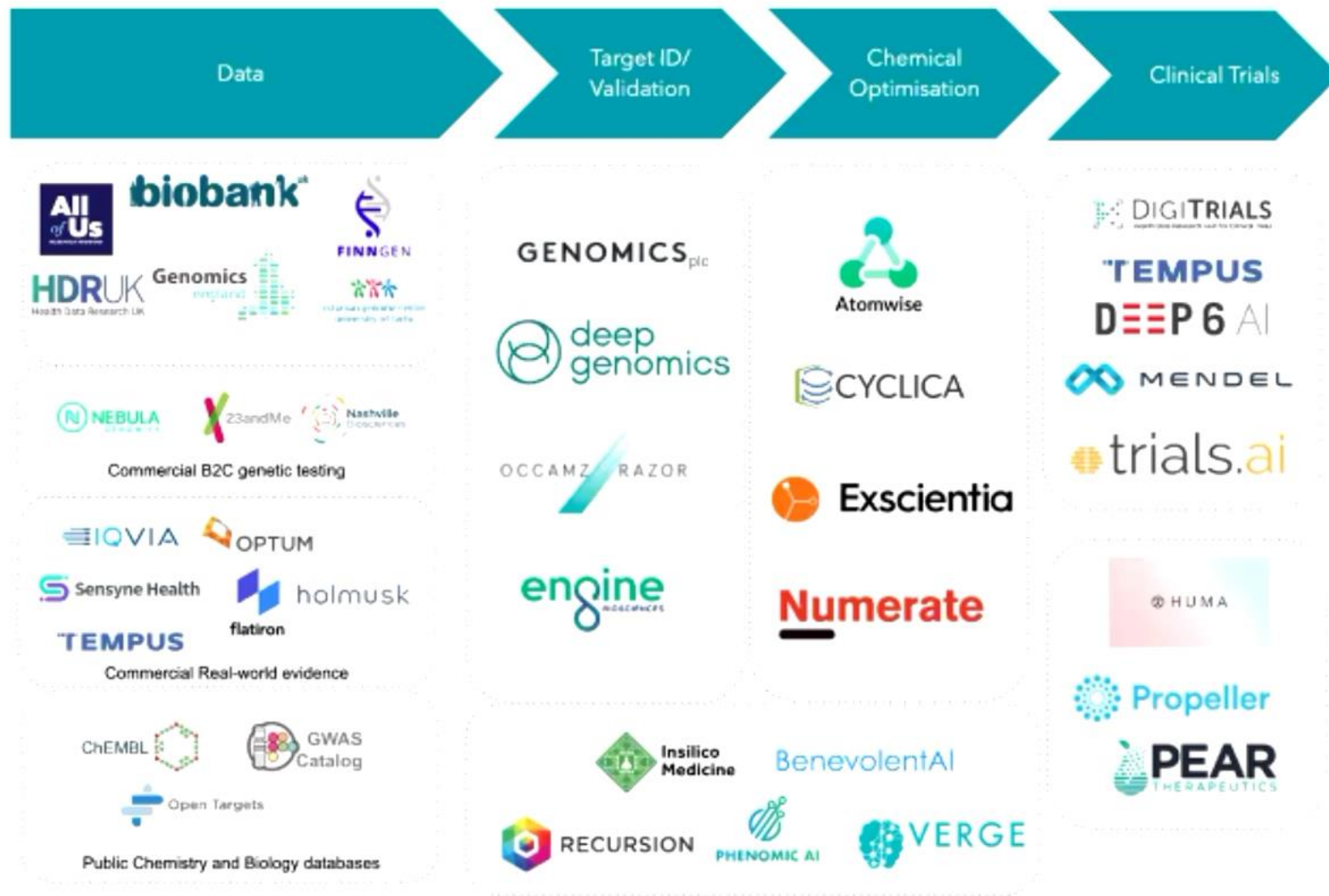
AI Companies - 240
Investors - 600
Corporations - 90



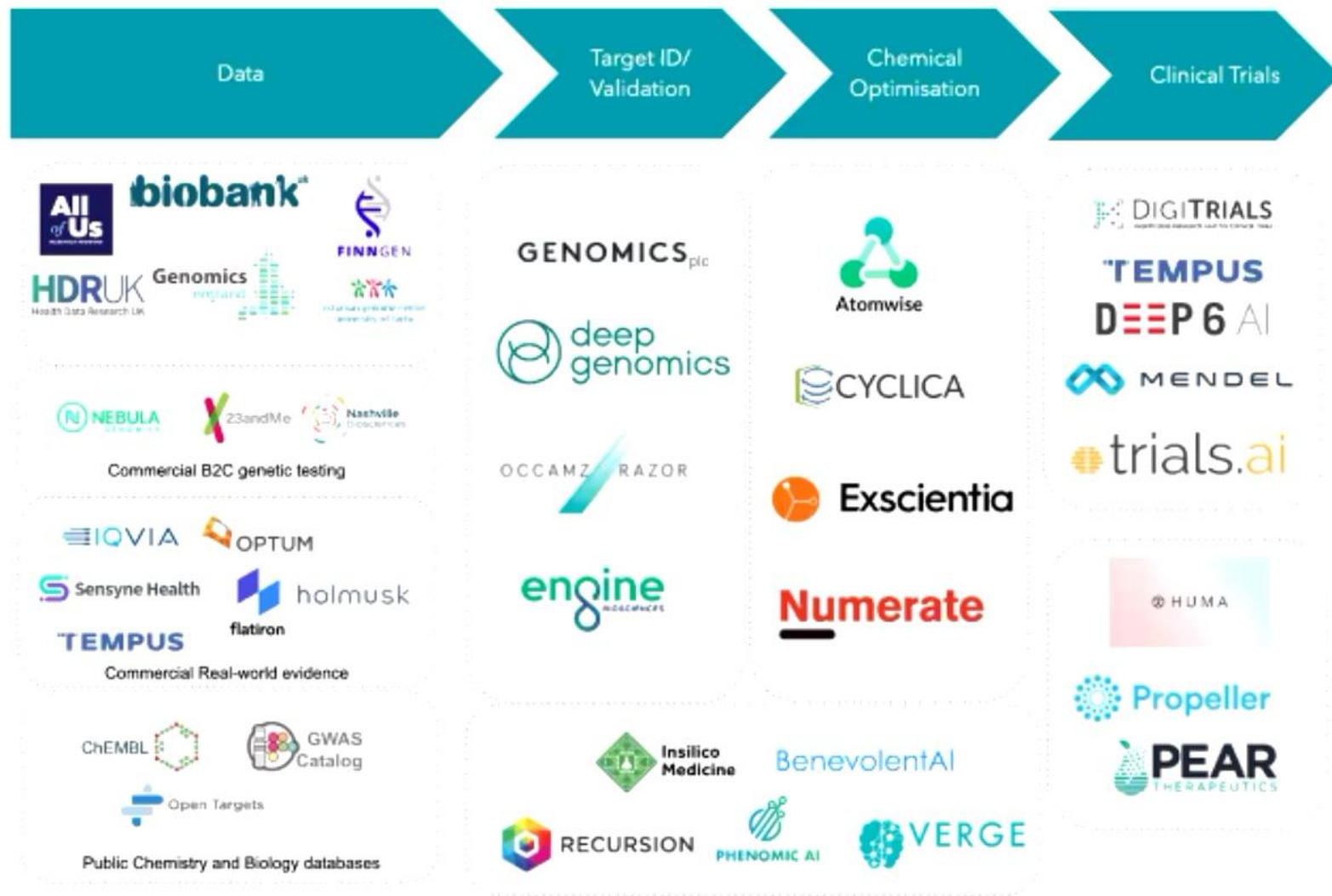
Rabia Khan

Source: <https://www.biopharmatrend.com/post/302-new-report-artificial-intelligence-for-drug-discovery-biomarker-development-and-advanced-rd-landscape-overview-2020/>

- 매우 많은 제약 회사들이 박차를 가하고 있으나, Machine Learning 을 기반으로 하지 않는 이상 힘들다는 것이 요즘의 트렌드



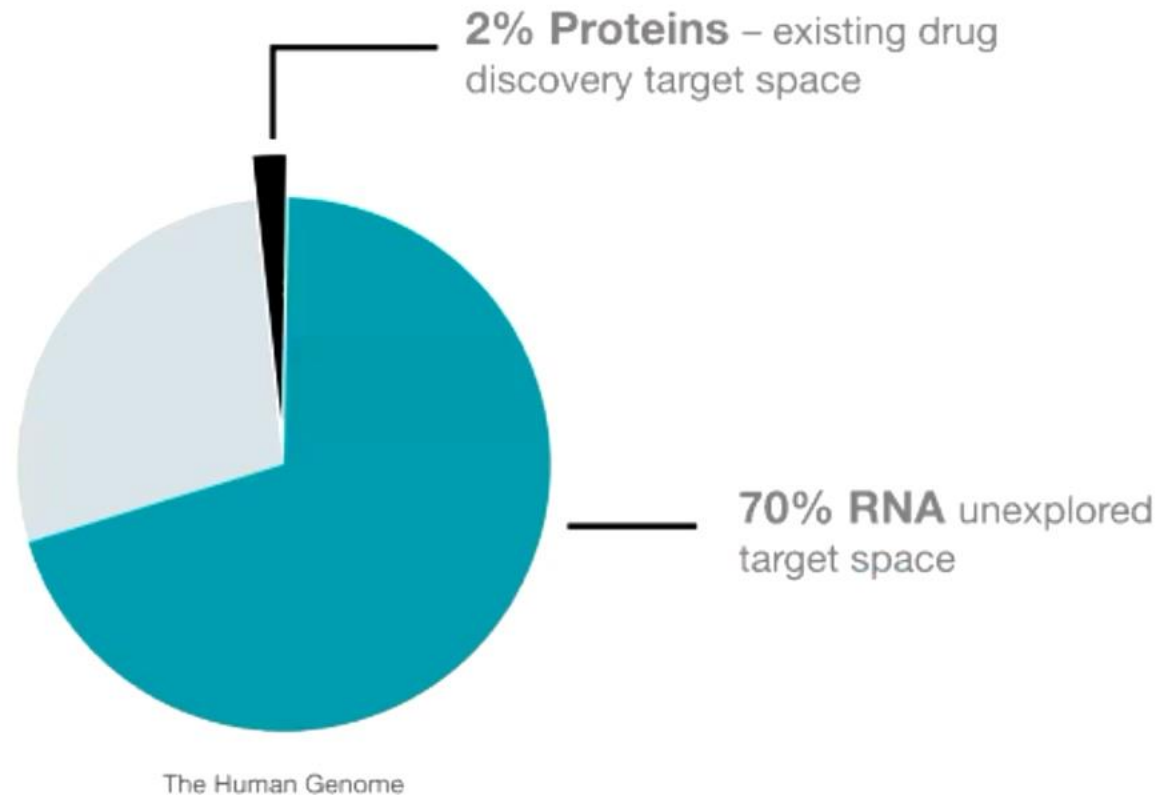
- ML 을 적용하는 것이야 간단하겠지. 대용량의 Data 를 가지고, 생물학적인 지식을 이용해 target 을 세팅하고, 화학적 지식을 이용해 실제 약을 만들어주고, 임상 단계를 거치면, 짜잔~ 약이 완성ㅋ



Rabia Khan

- 이름있는 유망한 회사 조차, data 에 여전히 머물러 있고 (sit 이라는 표현)
- They use structural 3D, structural models, to identify small molecules

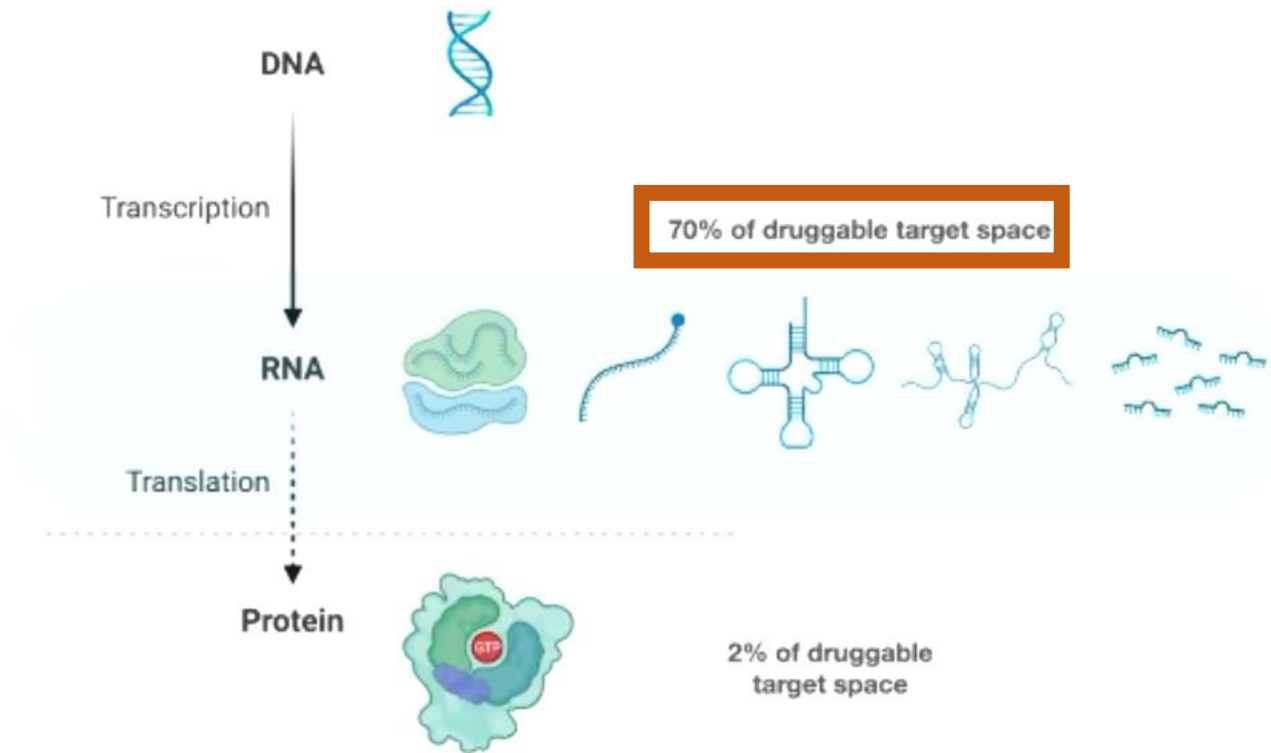
70% of the human genome: unexplored target space



Rabia Khan

- 현재, 사람 전체 유전체의 30%만이 밝혀졌으며 그 중 2%가 drug 의 타겟이 되는 protein 으로 코딩된다
- 여기 있는 70% 는 junk DNA 라고 불림...하지만 이는 좀 예전의 자료니까 지금은 어느정도 밝혀졌을지도 모릅니다

Targeting RNA: Expanding the target space



Created with BioRender.com

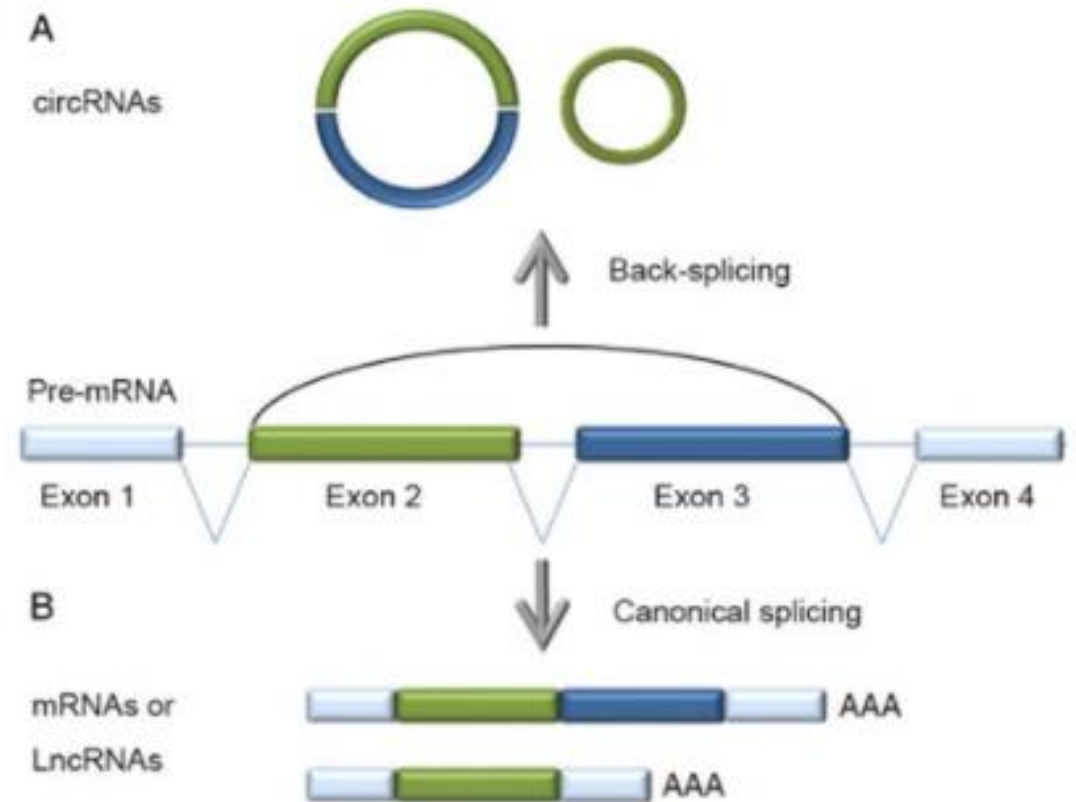


Rabia Khan

- Not just mRNA, but non-coding RNA, Micro RNA, circular RNA,,,, 역시 타겟팅이 되고 있는 추세
- Micro RNA – 전사되지 않는 RNA 로써, RNA 발현 억제 (RNA silencing) 의 역할을 함

• Circular RNA

- Exon 들이 뭉쳐서 원형을 이루어 만들어진 RNA
- Splicing 의 과정과, miRNA 의 양을 조절
- 몇몇 protein 의 Transcription 을 조절
- 최근 뜨고 있다고 생각해도 될 듯함??
- 나도 뭔가 생소함...



But....it's junk DNA?



- Junk DNA 의 경우는? 이걸 왜 신경쓰지? Drug 와 관련이 있다고?

90% of disease
associated
variants are in the
non-coding
genome*

Small molecule
drug discovery
efforts have
targeted **~0.15%** of
the human genome

Source Giral et al, Into the Wild: GWAS Exploration of Non-coding RNAs: Front.
Cardiovasc. Med., 17 December 2018



- 그에 대한 정답 – 질병의 90%는 non-coding genome 과 관련되어 있다
- 그 이유는, 애네들이 regulatory elements 라고 불리는 단백질을 조절하는데 있어서 매우 중요한 역할을 하기 때문

RNA: Novel target space



Rabia Khan

- 즉, RNA 는 중요한 drug 의 중요한 target space 가 될 수 있다



Ladder Therapeutics

Mapping the druggable transcriptome



- 여기 회사에서는 그렇게 말하긴 하는데, protein 뿐 아니라, RNA 에 좀 더 집중하는데 druggable transcriptome map 을 만드는데 집중하는 중
- 그냥 처음부터 끝까지 RNA 가 중요하다고 말하는 중

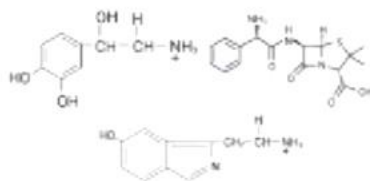
Structure Discovery



"Functional" RNA
Structures

<1% of structures in
public domain are RNA

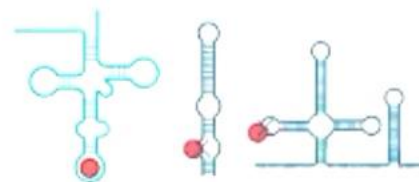
Chemical Discovery



Public Dataset:
~1500 Compounds

HTS Screening: 0.01% - 0.1%
Success Rate

Specificity



No Commercially
available assays

Rabia Khan

Targeting RNA

- Targeting 하는 RNA 를 3가지 bucket 으로 구분 가능 하다

1. Structure discovery (similar to target identification)

- Functional RNA structure 은 전체 public domain 의 1% 미만

2. 1,500 개의 알려진 compound 가 있음 (= Small molecule) binding to RNA

- High throughput Screening 을 수행하면 매우 낮은 success rate 를 보임...

Definition

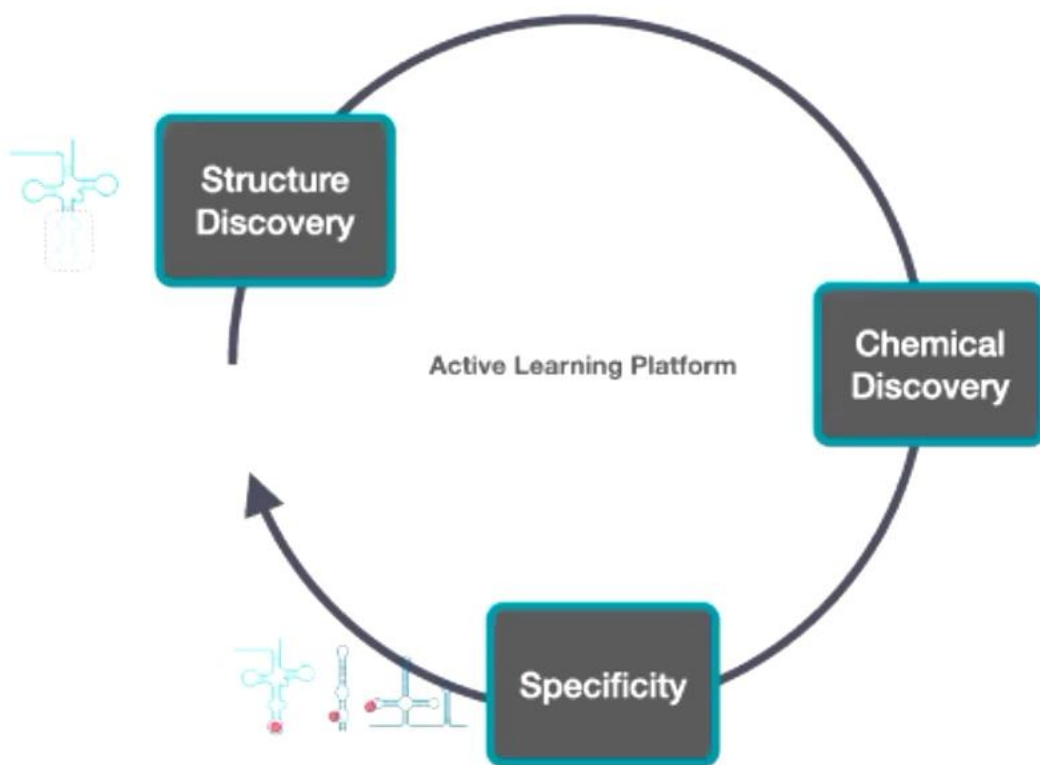
High throughput screening (HTS) is the use of automated equipment to rapidly test thousands to millions of samples for biological activity at the model organism, cellular, pathway, or molecular level.

3. Specificity

- No commercial available assays (올바른 측정 법이 없다)

We are building **the tool kit** for RNA drug
discovery





Public Dataset:
~1500 Compounds

HTS Screening: 0.01% - 0.1%
Success Rate

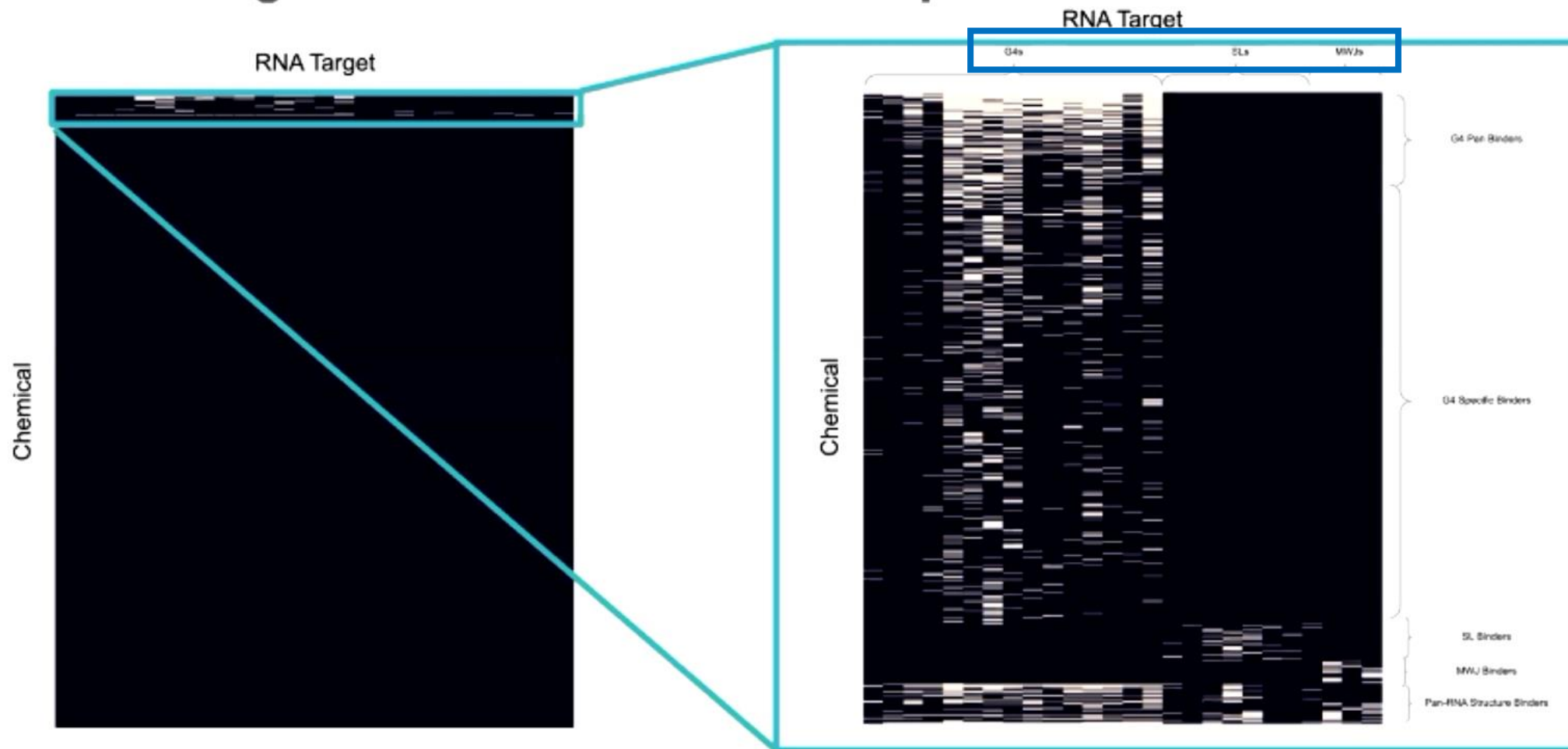
- Can we define RNA structure specific chemical scaffolds?
- Can we learn from public datasets?
- Do RNA binding small molecules sit within drug-like chemical space?



Rabia Khan

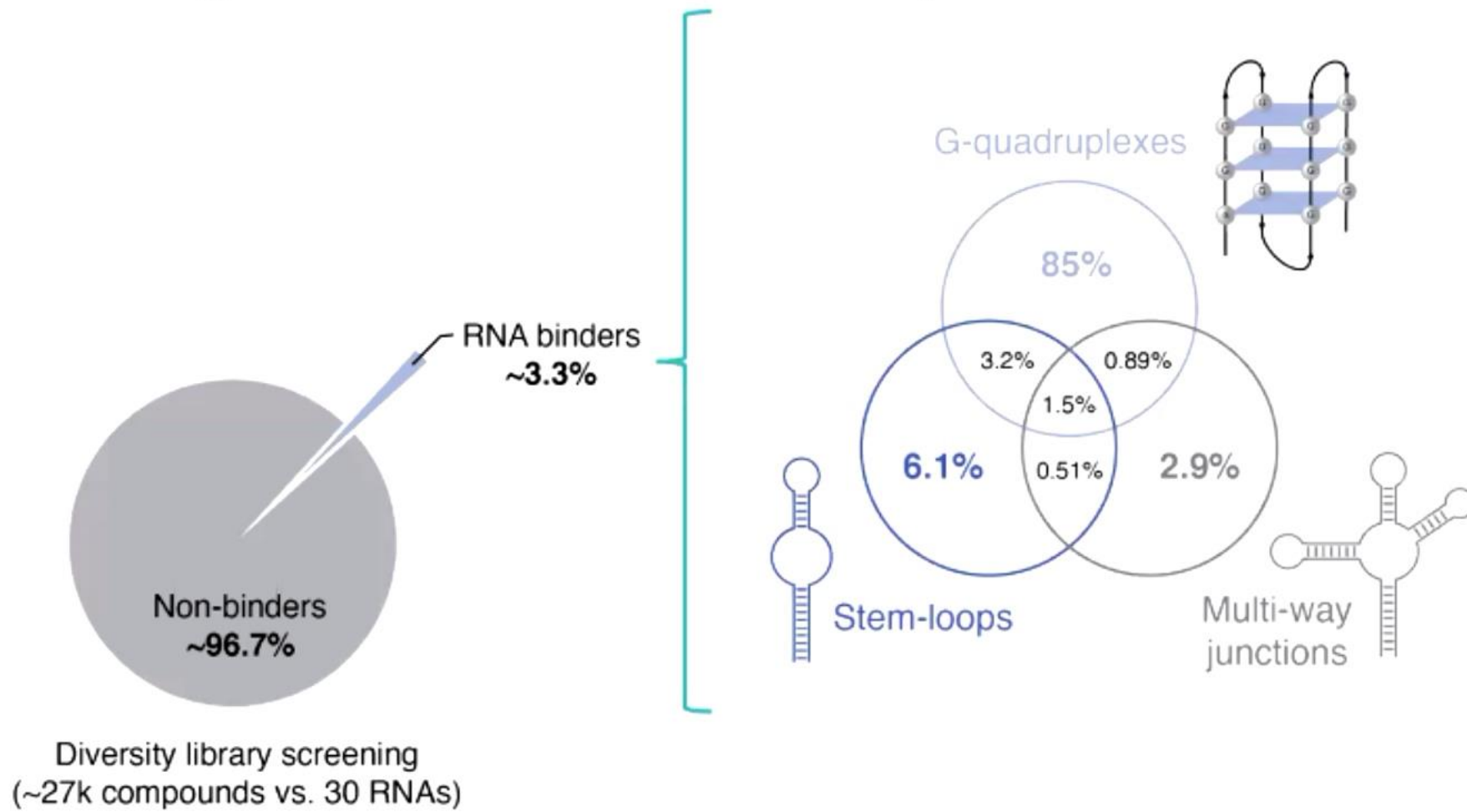
- 3가지의 중요한 질문을 이용해서 drug generation 을 한다

Learning Chemical Scaffolds Unique to RNA Motifs



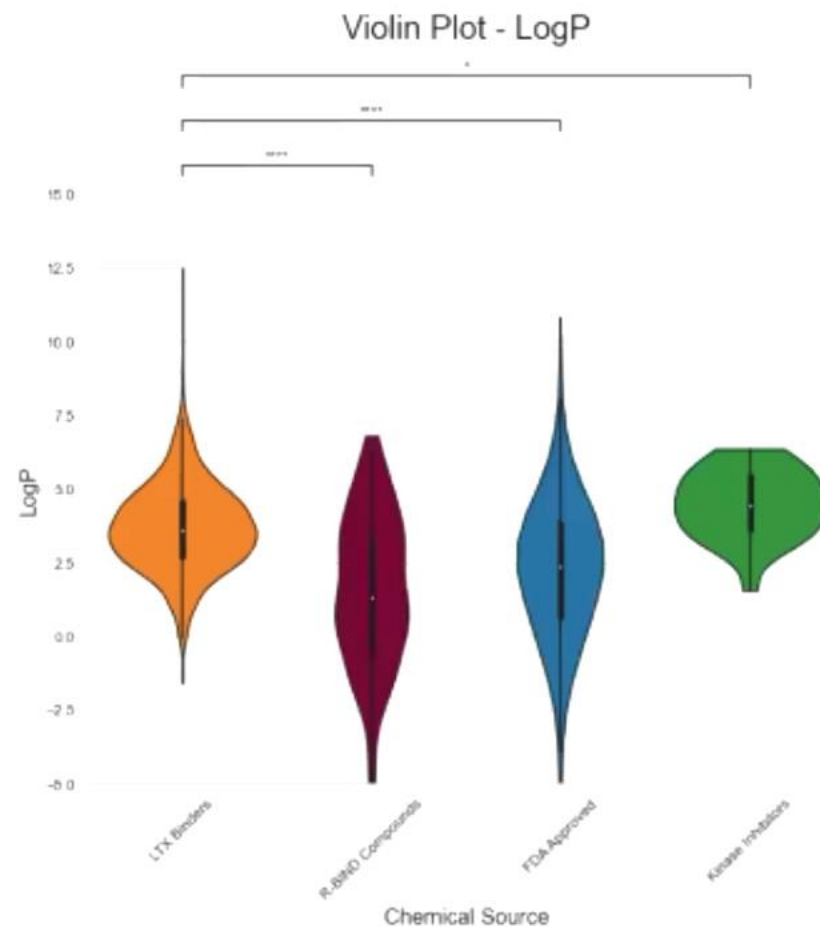
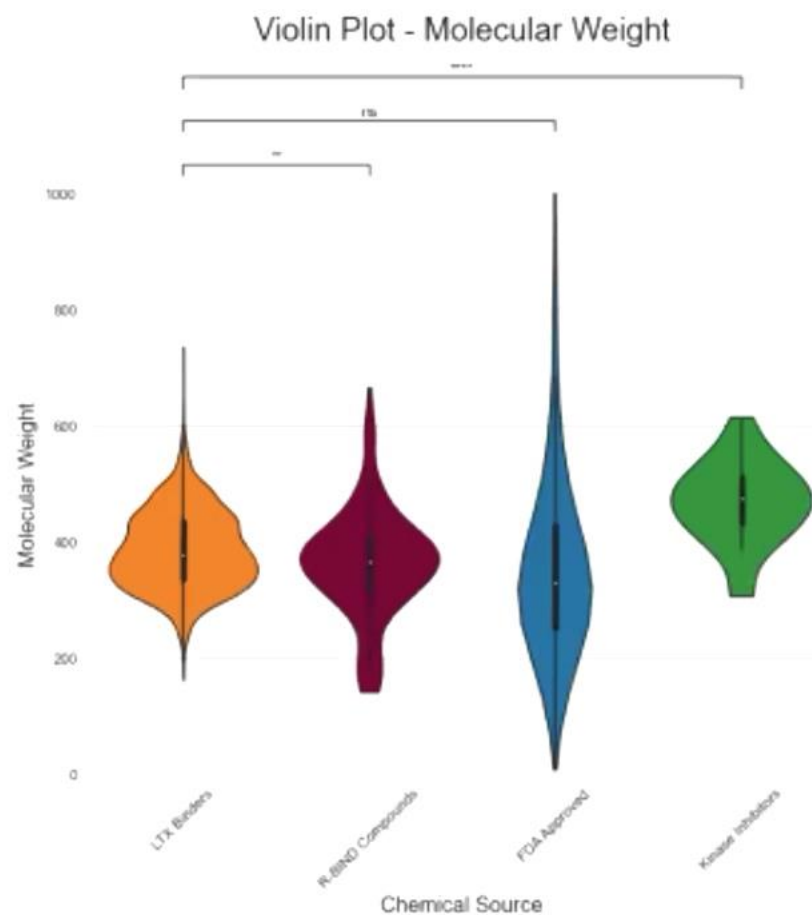
- 하얀색은 실제로 target RNA 에 binding 한 것
- Three different RNA structures (G-quadruplexes, Stem-loops, Multi-way junctions) → Very Specific chemical scaffold (or types)

Learning Chemical Scaffolds Unique to RNA Motifs



- 3가지 structure 에 모두 결합하는 binder 가 3.3% 중에서 1.5%가 됩니다

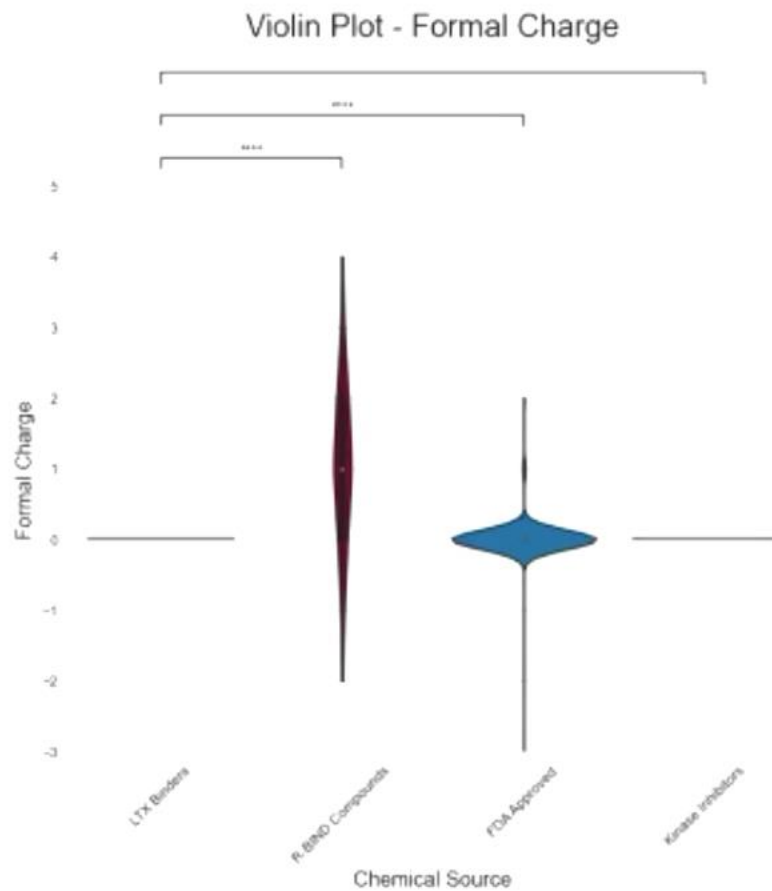
Learning from Chemical Descriptors



Rabia Khan

- Can we actually make drug binding RNA? – LogP 가 뭘 나타내는지 모르겠지만 서로 비슷한 경향을 띄는 것을 볼 수 있다

Learning from Chemical Descriptors

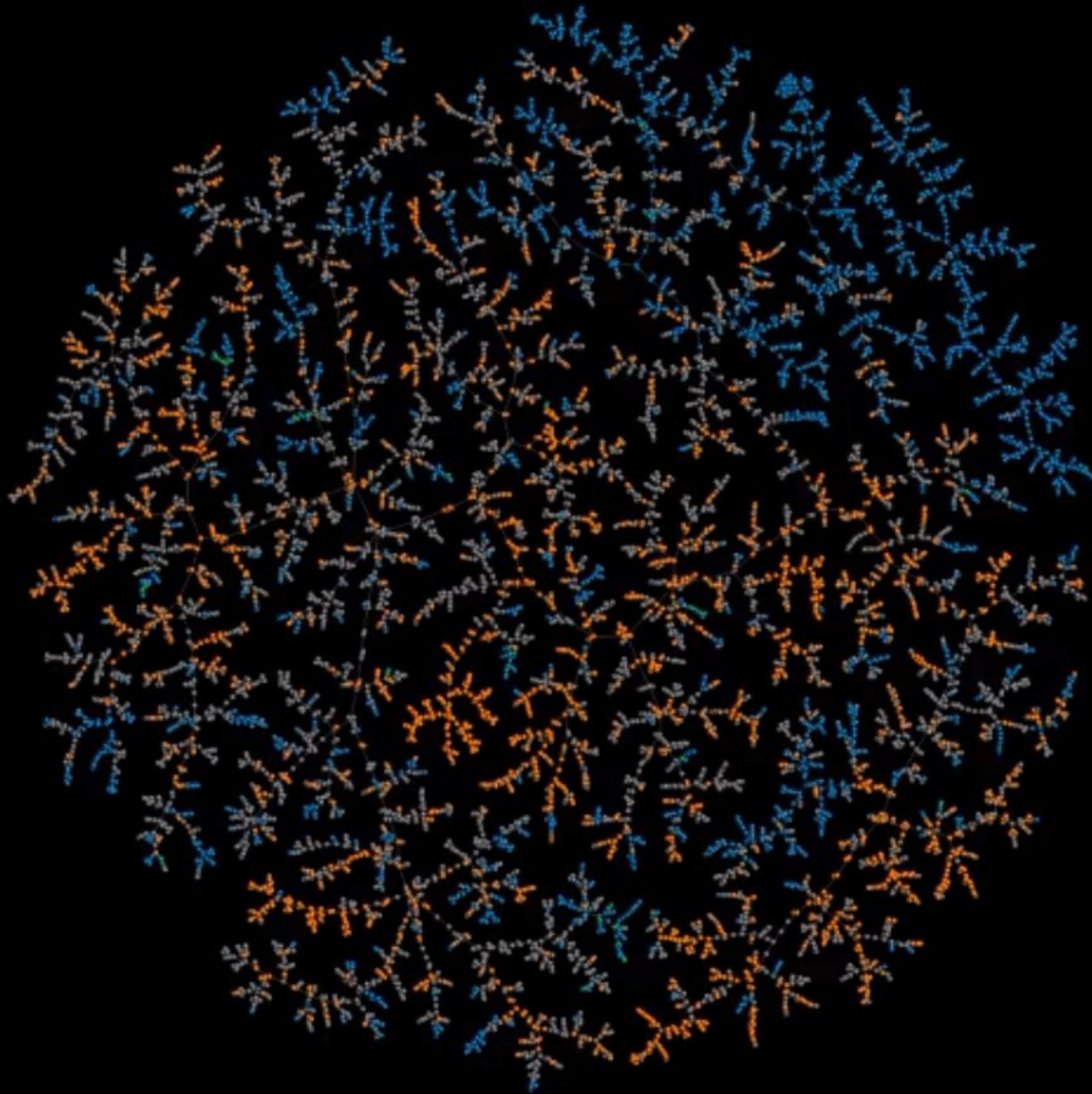


Public data sets are charged - likely binding in a non-structure specific manner



Rabia Khan

- Public dataset 으로는 학습이 많이 어렵다. So, have to generate our own dataset



Compound Library

- FDA
- Hit
- Kinase
- Not-Hit

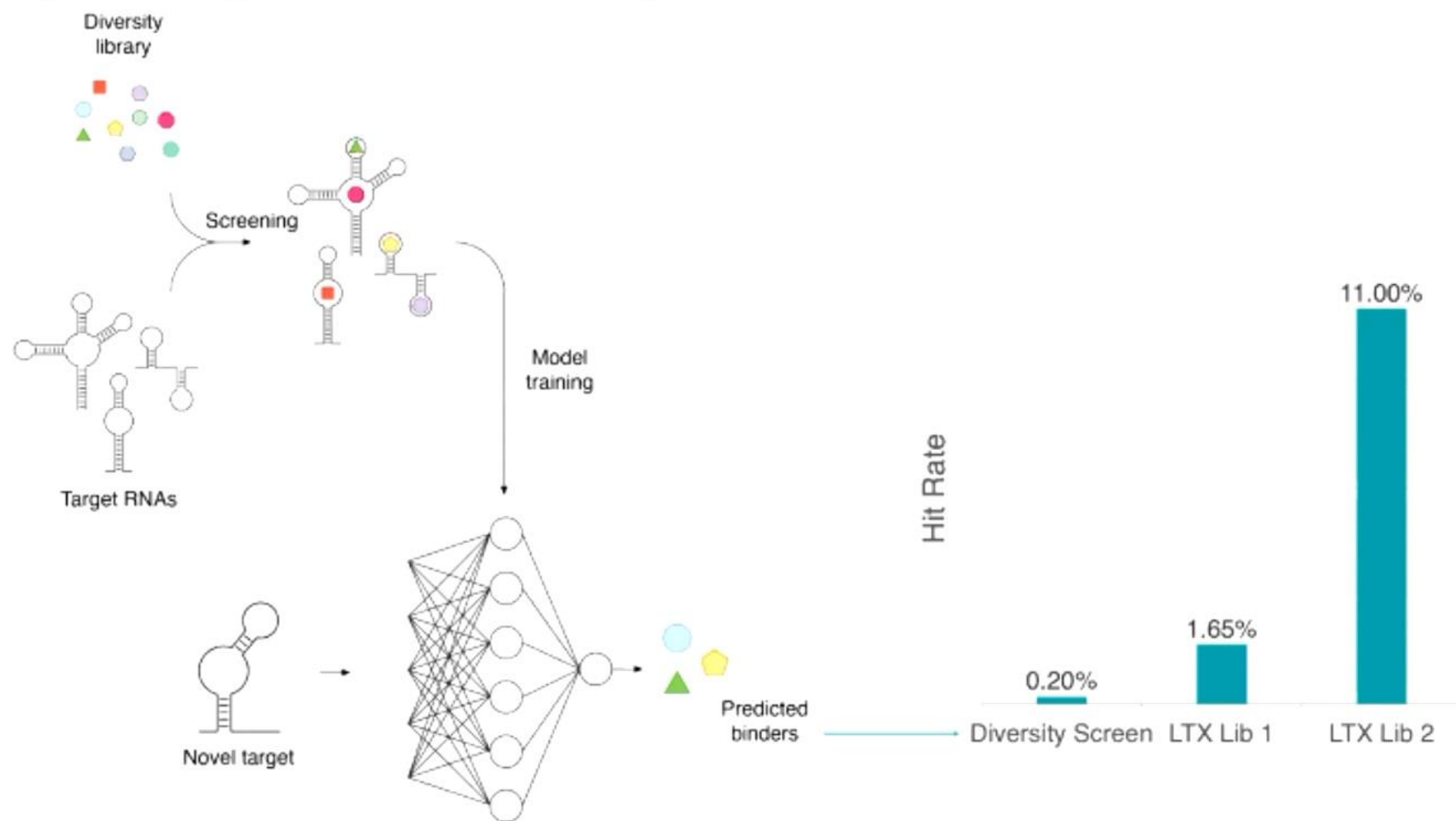
- 애네들이 만든 binder – Orange
- Non-binder – Gray



Rabia Khan

- Finger print 를 이용해 분자를 특징시키고, 그래프로 나타냄 ← Visualize mechanism

Exploring Chemical Space In Silico



Computationally **learning** the features of small molecules that lead to RNA structure-based binding



Rabia Khan

- Learn the features of small molecules that lead to RNA structure based binding

Challenges up ahead

- We do not understand biology
- Speaking both “languages”: biology and ML
- “Proving” the value will be based on what matters: **Impact to patients**

