# Module Overview

<table>
<thead>
<tr>
<th>Day</th>
<th>Lecture</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>DNA library synthesis (PCR)</td>
</tr>
<tr>
<td>2</td>
<td>SELEX I: Building a Library</td>
<td>DNA library purification (agarose gel electrophoresis)</td>
</tr>
<tr>
<td>3</td>
<td>SELEX II: Selecting RNA with target functionality</td>
<td>RNA library synthesis <em>(In vitro transcription = IVT)</em></td>
</tr>
<tr>
<td>4</td>
<td>SELEX III: Technical advances &amp; problem-solving</td>
<td>RNA purification and heme affinity selection</td>
</tr>
<tr>
<td>5</td>
<td>Characterizing aptamers</td>
<td>RNA to DNA by RT-PCR</td>
</tr>
<tr>
<td>6</td>
<td>Introduction to porphyrins: chemistry &amp; biology</td>
<td>Post-selection IVT; Journal Club 1</td>
</tr>
<tr>
<td>7</td>
<td>Aptamer applications in biology &amp; technology</td>
<td>Aptamer binding assay</td>
</tr>
<tr>
<td>8</td>
<td>Aptamers as therapeutics</td>
<td>Journal Club 2</td>
</tr>
</tbody>
</table>
SELEX II

Selecting RNA with target functionality

20.109 Lecture 3
10 February, 2011
Overall architecture of ds DNA library

- **Fixed sequence Region 1**
- **Fixed sequence Region 2**
- **Variable Region**
  - (at population level)
  - Sequence distinguishes one library member from the other!

- **T7 promoter**

- **How do we achieve variability between individual library members?**

- **Each library member has a unique, defined sequence**

- **Members differ from each other in the variable region**

- **Technical constraints dictate this architecture**
SELEX: The process (simply)

Target

*e.g. heme*

Interact

Eliminate “undesirables”

(Non-binders)

Iterate

Recover “desirables”

(Binders = Aptamers)

Now, that we understand the technical constraints for manipulating our library:

Library Design Principles given these technical constraints?
The RNA Library

Building a Library

Design Principles

- One library per target or one library for all targets
- Balance between “useful” and “useless” library members
- Maximizing “useful” collection within space constraints
- Now, let’s think about what we want in our library!

Technical Considerations

- Stability during storage
- Synthesizing library at reasonable costs
- Availability of efficient methods for manipulating library

✔
One master library or many libraries?

Molecular targets
e.g. heme

• Known target with a general idea about what its partner RNA should look like --> “custom build” library

• In absence of this data, build “generic” library
Library design principles

Co-optimize several competing variables:

• **Diversity**
  – Maximize the number of distinct RNA sequences present

• **Space limitations**
  – Maximize the total number of RNA molecules present
  – Practical limitations exist (i.e. How much RNA can you reasonably prepare?)

• **Representation**
  – Each possible RNA sequence is present at least once

• **Adaptability**
  – Have an easy way for increasing the representation of “popular” RNA molecules = **SELEX**!

• **Easily replenished:** Chemical synthesis; PCR; *in vitro* transcription
Diversity

• How can you increase diversity in your RNA library?
  – Increase:
    • The length of the variable region;
    • The number of nucleotides from which to choose;
    • The molar quantity of library available (sometimes)

• How do you calculate your library diversity?
  – Distinguish theoretical versus actual
Calculating theoretical diversity

- Let’s fix the nucleotides available = 4 (A, G, T, C)
  - 8 nucleotide variable region:
    • Maximum Diversity = Number of distinct sequences possible
      • = $(4)^8 \sim 6.6 \times 10^4$ unique sequences
  - 20 nucleotide variable region:
    • Maximum Diversity = $(4)^{20} \sim 1 \times 10^{12}$ unique sequences possible!
  - 50 nucleotide variable region:
    • Maximum Diversity = $(4)^{50} \sim 1.3 \times 10^{30}$ unique sequences possible!!

- Enormous theoretical diversity possible with nucleic acid libraries!
Alas, there’s only so much practical and affordable space for your library

- How many unique sequences can be represented in this space?

The Avogadro Constant:
\[ = 6.022 \times 10^{23} \text{ molecules/mol} \]

\[ (1 \text{ nmol} = 1 \times 10^{-9} \text{ mol}) \]

Number of molecules in 1 nmol
\[ \sim (1 \times 10^{-9} \times 6.022 \times 10^{23}) \]
\[ \sim 6 \times 10^{14} \text{ molecules!} \]

- 1 µmol scale synthesis
  - Nice compromise between cost and library mass obtained
  - On larger scale, downstream steps in library prep become limiting

- From this scale synthesis:
  - Obtain \( \sim 1 \text{ nmol} \) full-length, useable library
So, what size library (diversity) fits comfortably into the practical space available?

- **Total space** = $6 \times 10^{14}$ molecules

- 8 nucleotide variable region:
  - Number of distinct sequences possible
  - $= (4)^8 \sim 6.6 \times 10^4$ unique sequences

- 20 nucleotide variable region:
  - Maximum Diversity $= (4)^{20} \sim 1 \times 10^{12}$ unique sequences possible!

- 50 nucleotide variable region:
  - Maximum Diversity $= (4)^{50} \sim 1.3 \times 10^{30}$ unique sequences possible!!

- **In which of these libraries can the theoretical diversity be fully represented given our space constraints?**
Representation

- **Total space =** $6 \times 10^{14}$ molecules

- 8 nucleotide variable region:
  - Maximum Diversity = $(4)^8 \sim 6.6 \times 10^4$ unique sequences
  - Each sequence present @ $(6 \times 10^{14}/6.6 \times 10^4) \sim 1 \times 10^{10}$ copies/library

- 20 nucleotide variable region:
  - Maximum Diversity = $(4)^{20} \sim 1 \times 10^{12}$ unique sequences possible!
  - Each sequence present @ $(6 \times 10^{14}/1 \times 10^{12}) \sim 6 \times 10^2$ copies/library

- 50 nucleotide variable region:
  - Maximum Diversity = $(4)^{50} \sim 1.3 \times 10^{30}$ unique sequences possible!!
  - Each sequence present @ $(6 \times 10^{14}/1.3 \times 10^{30})$: 0 or 1 copy/library!
How do you co-optimize across these parameters

**Scenario I**

- Maximize diversity
- Achieve full representation by ensuring you have the available space.
  - Choose 50-nucleotide variable region (assume 100-base oligo)
  - Require \(~3 \times 10^5\) metric tons of oligonucleotide!!!
    - And that’s to have each possible sequence represented once!
  - How much diversity is enough?
    - 8, 20 or 50 (or more?)-nucleotide variable region?
    - Can you determine this ahead of time for every possible target?
How do you co-optimize across these parameters

**Scenario II**

- Set space limit (i.e. reasonable cost)
- Maximize diversity (within this limit)
- Preserve representation at some acceptable (read: arbitrary) limit?
  - You’ll saturate your space at ~ 23-nucleotide variable region (~ $10^{14}$ maximum diversity)
  - (Recall: For 1 µmol synthesis (yield: ~1 nmol) --> ~ $10^{14}$ molecules present)
  - Is this enough diversity?
How do you co-optimize across these parameters

**Scenario III**

- Set space limit (i.e. reasonable cost)
- Maximize diversity
- Sacrifice representation
  - A given sequence present only once (if at all) in library
  - Is this problematic?
  - What does this mean for library reuse?
    - Sampling without replacement
What’s the best strategy for assembling your library?

**Scenario I**
- Maximize diversity
- Achieve full representation by ensuring you have the available space

**Scenario II**
- Set space limit (i.e. reasonable cost)
- Maximize diversity (within this limit)
- Preserve representation at some acceptable (read: arbitrary) limit?

**Scenario III**
- Set space limit (i.e. reasonable cost)
- Maximize diversity
- Sacrifice representation

*The Answer? In the end, it’s really up to you!*
SELEX: The process (simply)

Now, that we understand:

- Target selection
- Library construction & manipulation
  - How do we enrich for binders?
  - *How do we put this all together into a workflow?*
Enriching your library for binders

• Need a partitioning strategy:
  – Separate target bound RNA from unbound fraction
  – Selectively release target bound RNA

• Most commonly involves immobilizing target on:
  – A membrane (e.g. nitrocellulose)
  – Solid support (usually some kind of bead)
    • Column format
    • Magnetic separation

• Other approaches, for e.g.:
  – Electrophoretic methods to separate \{RNA:target complex\} from free RNA [J. Club].
Putting it all together: A typical SELEX workflow

10^{13-15} unique molecules
RNA Library

Selection column
Immobilized target

Wash to eliminate non-binders
Elute survivors

Aptamer enriched RNA library

RT-PCR
DNA library

In vitro transcription

Deconvolution of the selected library
SELEX à la Tuerk & Gold

- Target known to interact with RNA from prior work
  - Sequence below found in the mRNA encoding the T4 DNA polymerase
  - Regulatory mechanism:
    - T4 DNA polymerase binds its own mRNA decreases its own synthesis
- 8 nucleotides [AAUAACUC] are critical for the interaction
  - What underlies the preference for this loop sequence?

Based on objective, what library design would you choose?

C. Tuerk and L. Gold; Science; 249 (4968), 505-510, 1990
RNA library

T7 promoter

Variable Region:

= 8 nucleotides

• **Total space = 6 x 10^{14} molecules**

• 8 nucleotide variable region:
  – Maximum Diversity = (4)^8 \sim 65,556 unique sequences
  
  – Each sequence present @ (6 \times 10^{14}/\sim 6.6 \times 10^4) \sim 1 \times 10^{10} copies/library
  
  – The known RNA target present @ 2 in 10^5 molecules!

C. Tuerk and L. Gold; *Science*; 249 (4968), 505-510, 1990
SELEX à la Tuerk & Gold

**Immobilize on nitrocellulose**

- Works well for many protein targets

**T4 DNA polymerase**

*Structure for residues 1-388 from the PDB (www.rcsb.org)*

**Advantages**

- Very easy and inexpensive!
- Well-developed and straightforward protocols available

**Disadvantages**

- Protein can denature during immobilization step
  - Selected aptamers cannot recognize native protein
- Not all proteins stick strongly enough to survive washing steps to remove unbound library

C. Tuerk and L. Gold; *Science*; 249 (4968), 505-510, 1990
SELEX à la Ellington & Szostak

- Discover RNA binding to small molecule organic dyes
  - No prior knowledge of their RNA binding capacity
- Can RNA specifically interacting with these molecules be discovered?

Cibracon Blue  Reactive Blue 4

www.sigmaaldrich.com

A.D. Ellington and J.W. Szostak; *Nature*; 346 (6287), 818-822, 1990
SELEX à la Tuerk & Gold

**Fixed sequence Region 1**

**Variable Region:**

= 100 nucleotides!

**Fixed sequence Region 2**

- **Total space ~ 6 \times 10^{14} molecules**
- **100 nucleotide variable region:**
  - Maximum Diversity = (4)^{100} \approx 2 \times 10^{60} unique sequences possible!
  
  - Each sequence present @ (6 \times 10^{14}/\sim 2 \times 10^{60}): Absent or 1 copy/library
  
  - The known RNA target present @ ??? frequency

A.D. Ellington and J.W. Szostak; *Nature*; 346 (6287), 818-822, 1990
SELEX à la Ellington & Szostak

**Immobilize on agarose beads**

- Very common strategy
  - Low molecular weight compounds
  - Macromolecules (e.g. proteins)

**Advantages**
- Extremely convenient and adaptable to many formats (e.g. column)
- Better define how your target is displayed for binding (though not completely)

**Disadvantages**
- Not all immobilized molecules will be able to interact (even with its cognate RNA)
- Immobilized form recognized is distinct from the free form of the target

A.D. Ellington and J.W. Szostak; *Nature*; 346 (6287), 818-822, 1990
Summary

• Developed a conceptual framework for SELEX

• Library diversity
  – Calculations
  – Maximizing diversity within technical constraints
  – Choosing the appropriate library for your needs!

• Examined some key steps involved in the process:
  – Target selection
  – RNA library construction
  – Partitioning strategies

• SELEX can be successfully executed on:
  – Very distinct targets
  – Using distinct library design (diversity, representation, etc)
  – Using distinct partitioning strategies
  – Fairly robust and generally applicable strategy
Next time…

• Determining the sequence identity of individual aptamers in the selected library

• Determining that your library truly contains RNA with affinity for your target!

• Modifying your SELEX strategy to more efficiently achieve your desired outcome