A Novel Genetic Algorithm for Detecting FLT3 Internal Tandem Duplications in Acute Myeloid Leukemia Patients

Jonathan King

Undergraduate, U.C. Berkeley

BS Bioengineering BA Computer Science

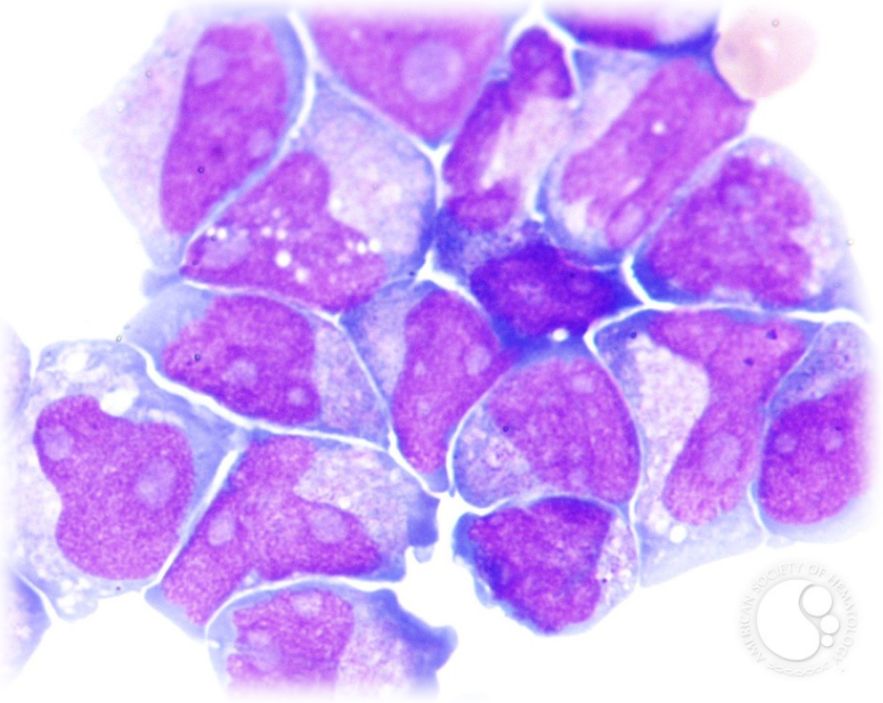
Intern, Plexxikon Inc.

NCCB

Berkeley
UNIVERSITY OF CALIFORNIA

Plexxikon

Background



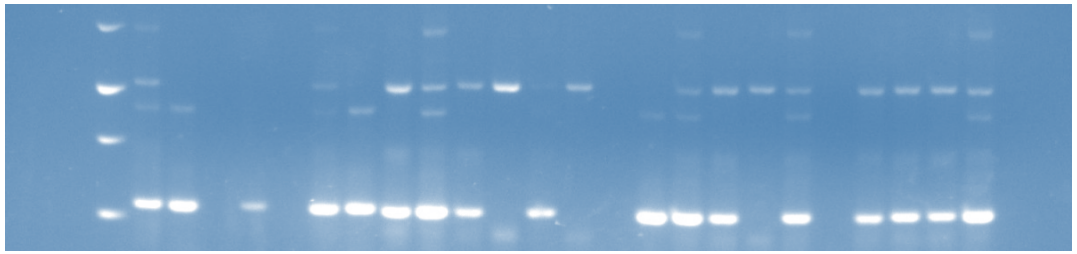
- Treating Acute Myeloid Leukemia
  - 20,000 expected cases in 2016
- Target: “Unfavorable” FLT3-ITD^{+} subtype
  - ITD = Internal Tandem Duplication*
- **Goal:** Identify ITD^{+} individuals for future clinical trials

*

Identifying ITD^{+} Drug Candidates

PCR

- Lacks potentially useful info:
  - Sequence
  - Precise location, length
- Inaccessible



Next Generation Sequencing

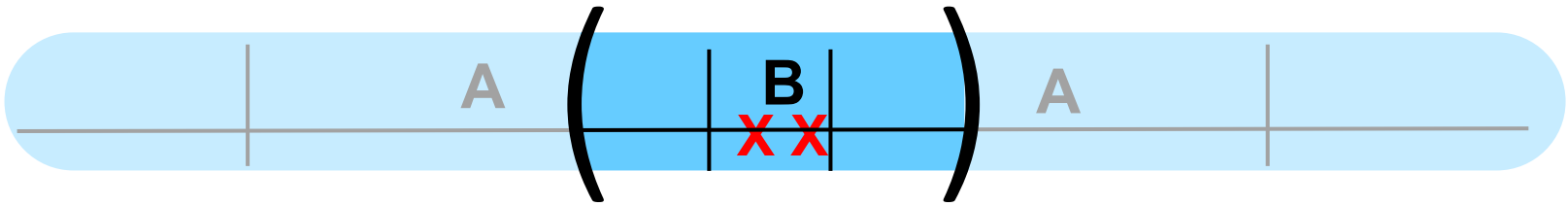
- Best at finding small variants:
  - SNPs
  - Short insertions & deletions
- Limited software for ITDs

```
ACTGTACCAGGACGCGCGCAGGGGGGAGGGGTACCTTCGGTCGACA
ACACCAGGTGGAGCGGGGGCAATCTTACGCTGCGGGGCAACTCCCC
TGTGTATAGTAGTATAGCGGCTAATGATCTACGGTTTACGTTATAA
AGGGAGCAGCGTTCGCTTATTGGAAGTAGTATAGGTTCGGCGATAT
CTCCGCGATAGGAGCAGTATCGAGGGTCACTGAACGTTCAGGGGGC
```

Challenges to ITD Discovery

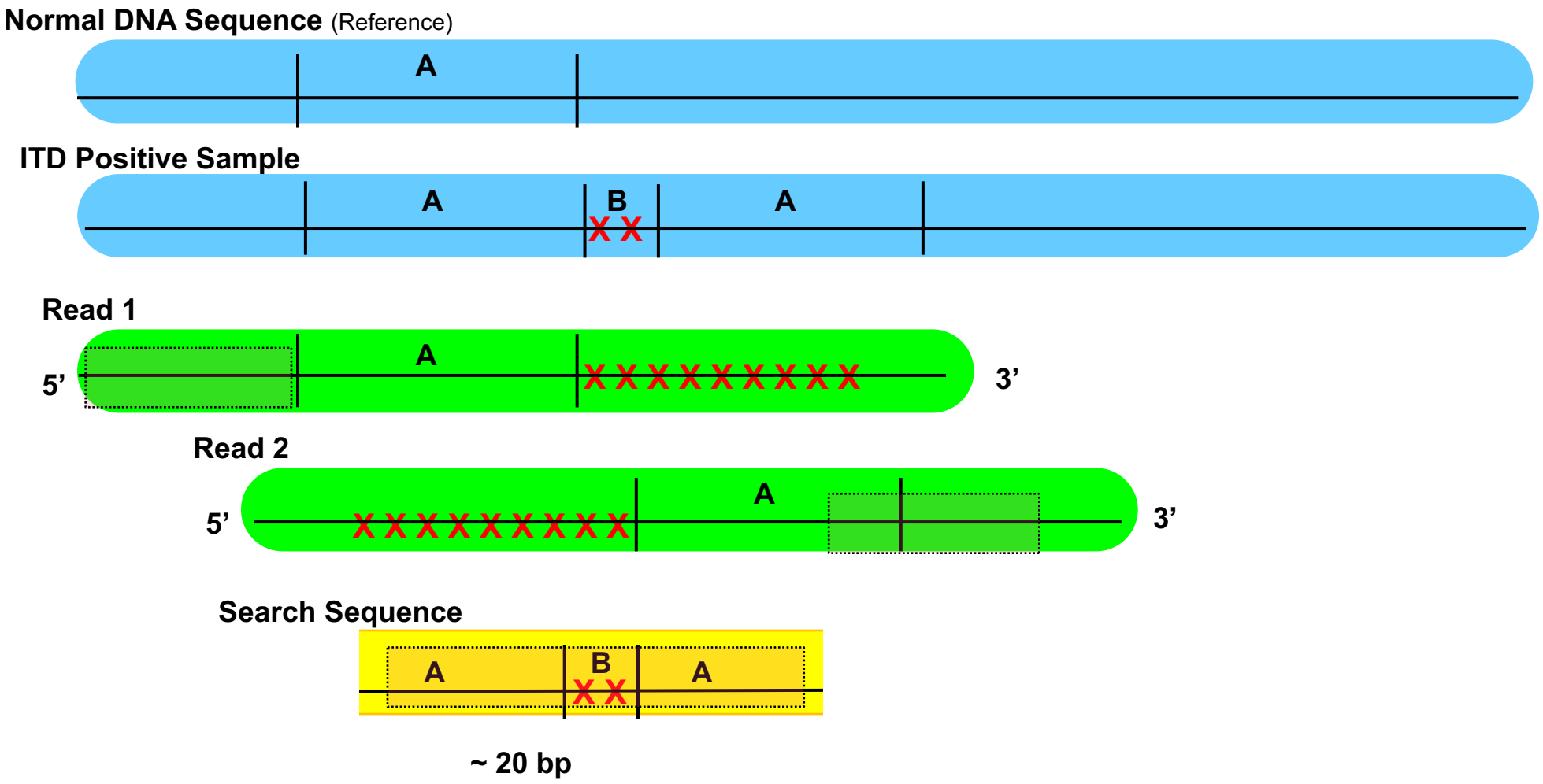
- Vary in location, length, and frequency!

Claim: The ~ 20bp "Junction of Duplication" uniquely identifies FLT3-ITDs.

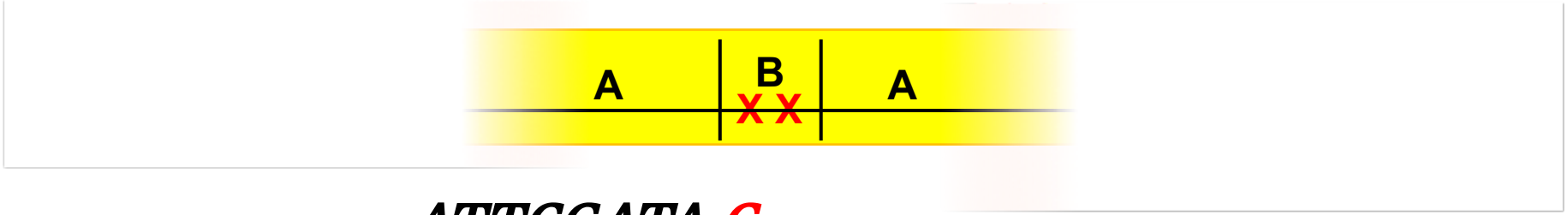


Probability of same sequence occurring by random chance in the region of interest is: $\frac{1}{4^{20}} \times 10^3 \approx \frac{1}{10^9}$

An Algorithm for Detecting ITDs in FLT3



Flexible Searching with Regular Expressions



$$ATTCGATA\ C\ldots$$

$$\ldots A\ CGAAGTTC$$

$$ATTCGATA\ (\ C\ \ [ACTG]\{0,m\}\ \ A\ )\ CGAAGTTC$$

Left Probe $[0:n-1]$ *LP* $[n]$ *between 0 and* m *bases* *RP* $[0]$ *Right Probe* $[1:n]$

Example parameters:

$$n = 13$$

$$m = 30$$

Plexxikon

FLT3 ITD Identification Results

07/19/2016 11:45:39 AM

`9996_S7` is **ITD positive**, with `168` supporting reads and `2` possible duplications.

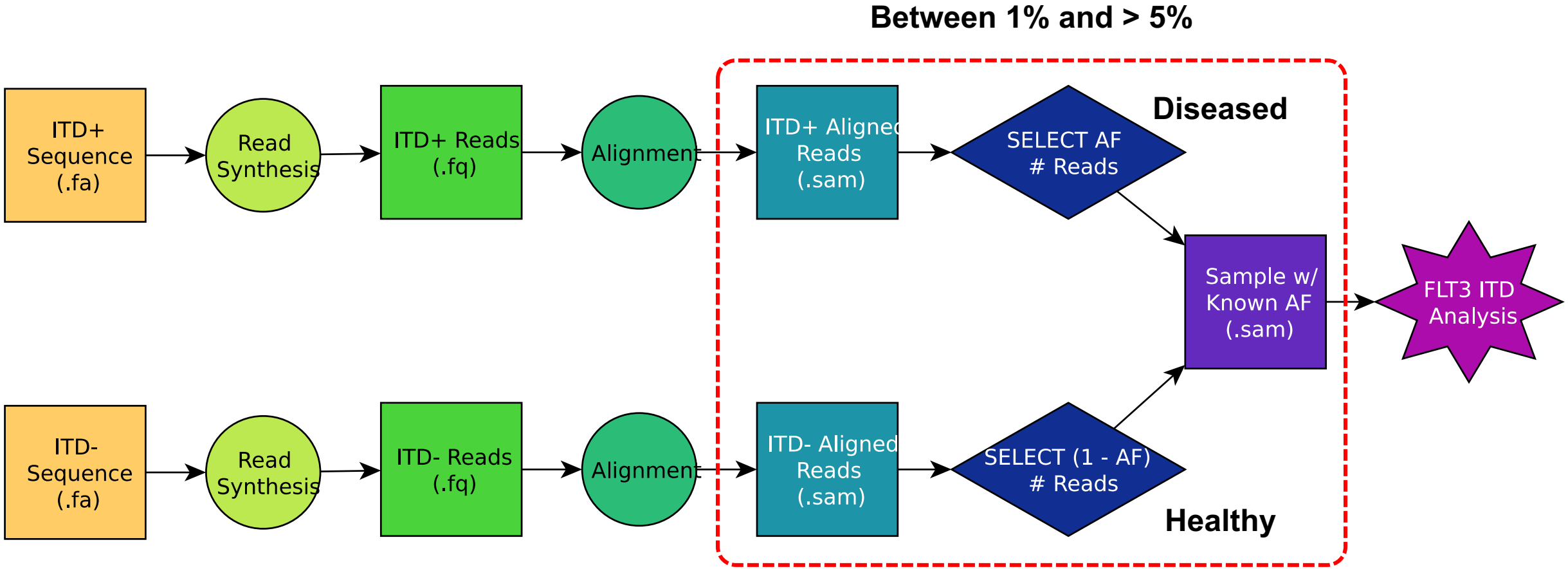
Duplication 1

Location	Length	Reads	Non-ITD Reads	Allele Frequency	Reads with entire ITD
28608262	51	118	769	0.13303	80

Sequence:

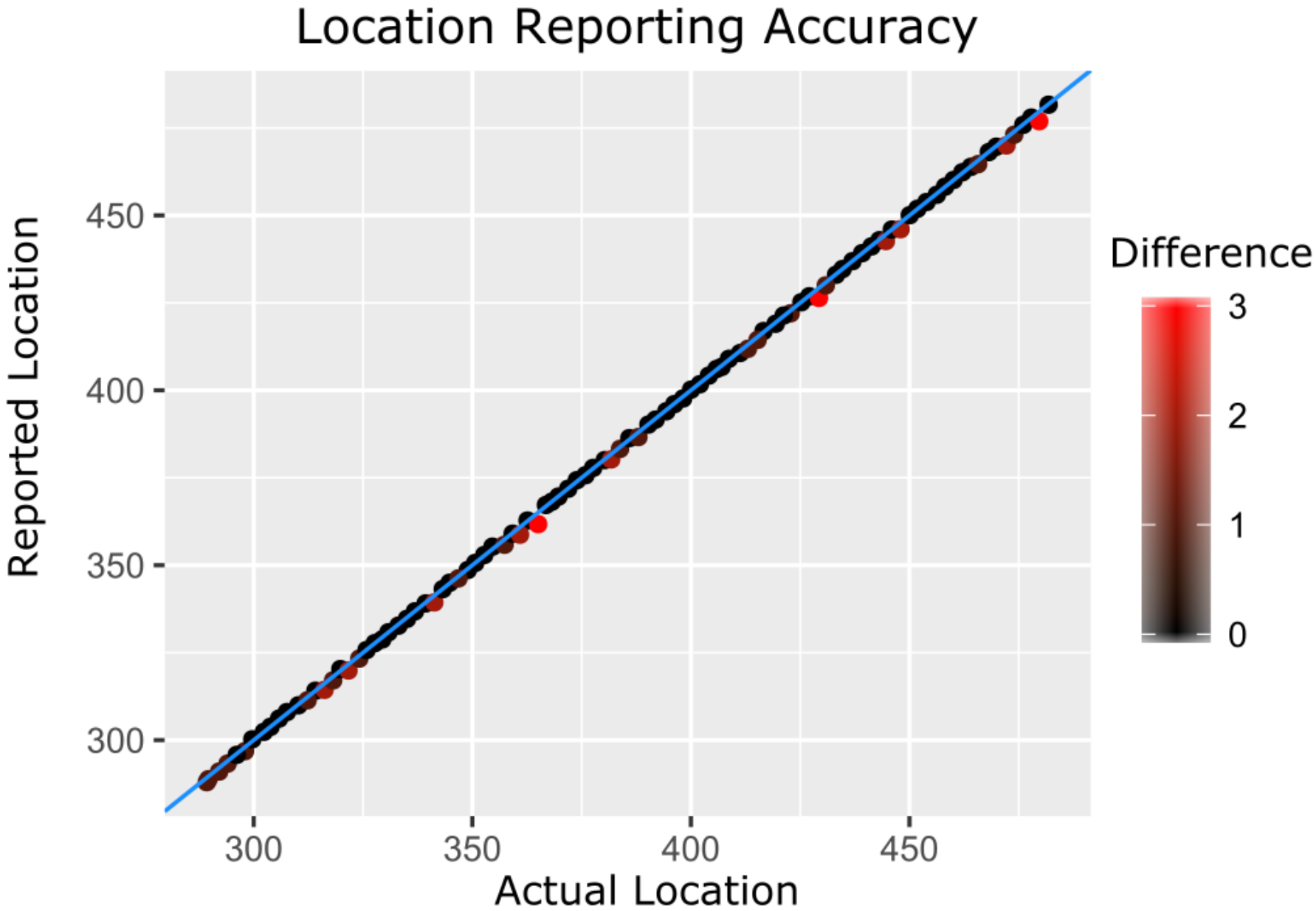
```
ATATGATCTCAAATGGGAGTTTCCAAGAGAAAATTTAGAGTTTGGattccc
```

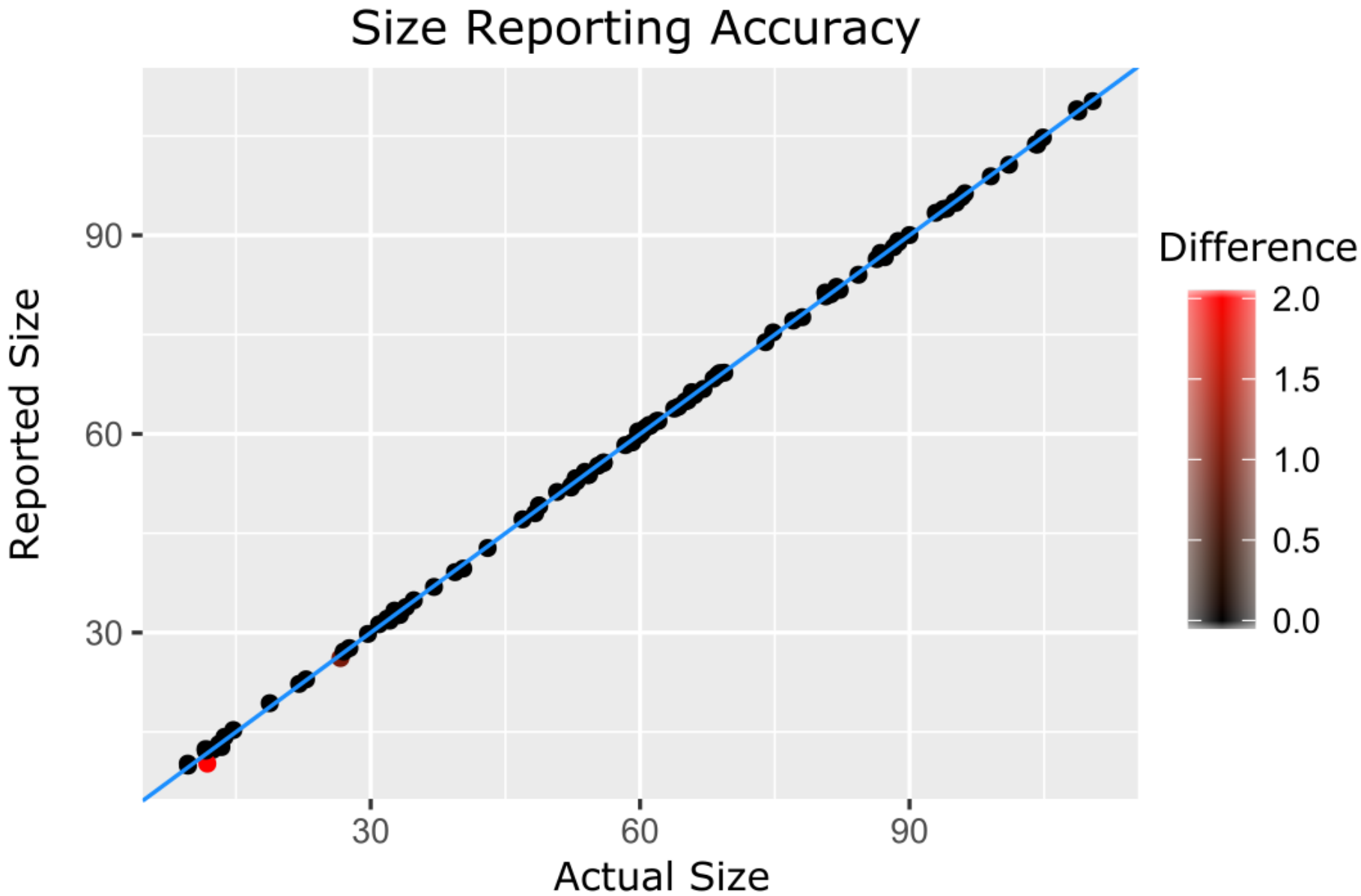
in silico Testing Procedure



Specificity:
99.9%

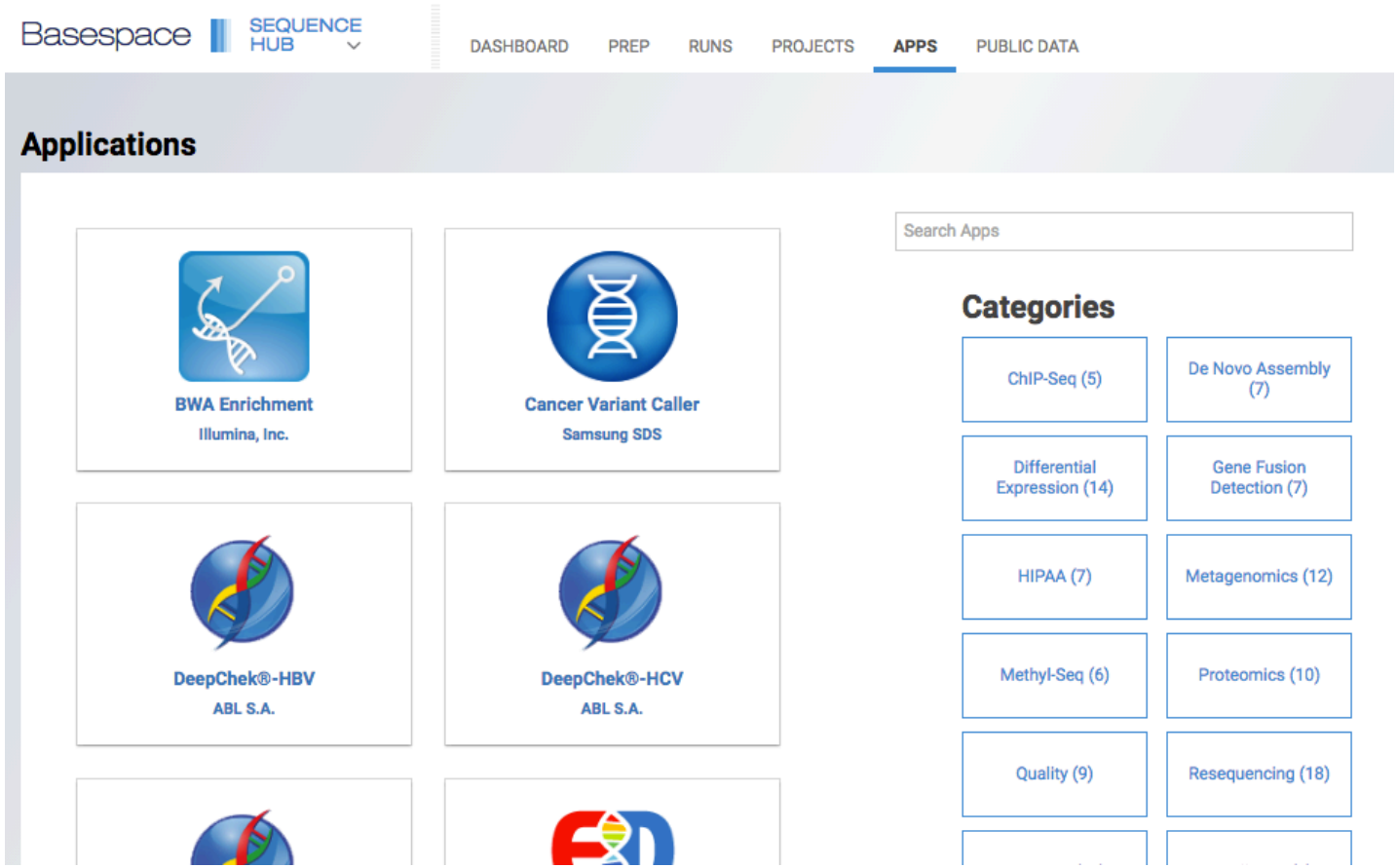
Sensitivity:
(assuming AB > 2%)
100%


Location Reporting Accuracy
Reported Location
450
400
350
300
300
350
400
450
Actual Location
Difference
3
2
1
0


Size Reporting Accuracy
Reported Size
90
60
30
30
60
90
Actual Size
Difference
2.0
1.5
1.0
0.5
0.0

Using the Algorithm Offsite

BaseSpace®



Thank you!

Plexxikon

Paul Severson

Brian West

Gideon Bollag, Chao Zhang, Paul Lin

Marguerite Hutchinson, Robert Bernstein

Interns!

qb3 ucb·ucsc·ucsf

California Institute for Quantitative Biosciences

Prof. Susan Marqusee
Donna Hendrix

NCCB

Berkeley EECS
ELECTRICAL ENGINEERING & COMPUTER SCIENCES

Friends and family!

BIOENGINEERING
UNIVERSITY of CALIFORNIA, BERKELEY