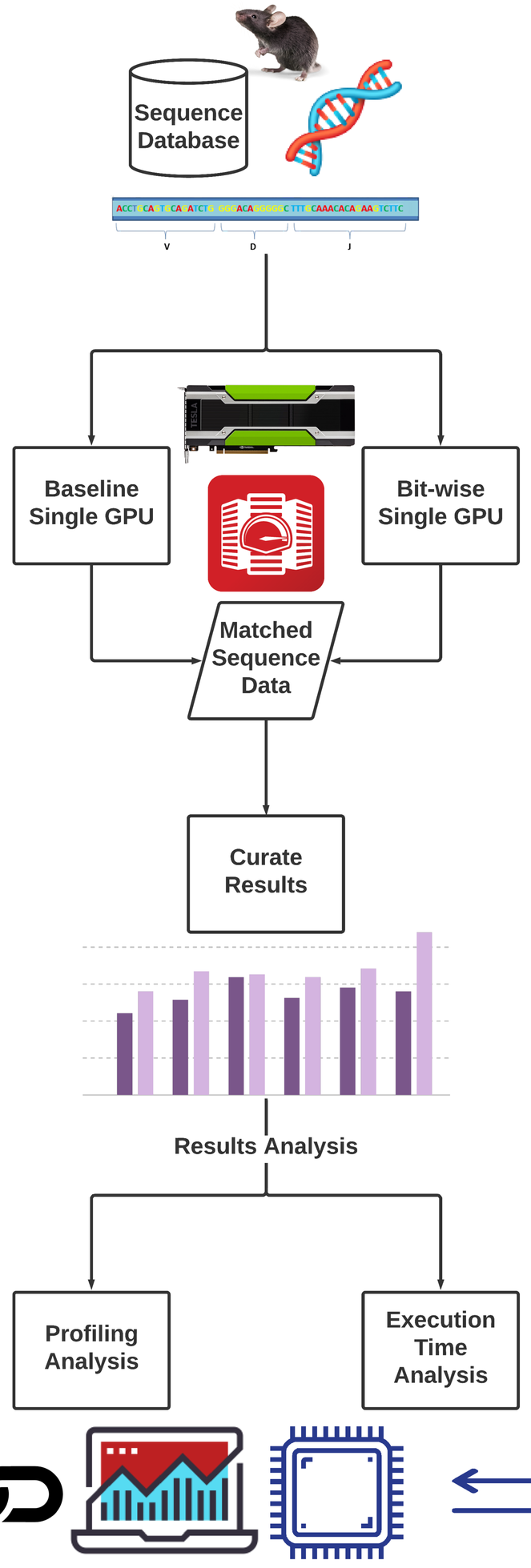
OVERCOMING THE LIMITATIONS POSED BY TCR-BETA REPERTOIRE MODELING THROUGH IN-SILICO DNA RECOMBINATION ALGORITHM

Sreevatsa Nukala[1], Ali Akoglu[2], Nirmal S Kumbhare[2], Joshua Andrew Mack[2]
KEYS Internship[1], Department of Electrical and Computer Engineering[2]

Scope

- Immune system uses TCRs to detect and neutralize antigens
- V(D)J recombination generates diverse TCR repertoire to detect variety of antigens
- Two GPU versions of V(D)J Recombination algorithm to model mouse TCRβ repertoire
- Algorithm generates unique pathways to create an in-vivo sequence
  - Counting of the recombination pathways is a peta scale process
- Timing and profiling analysis for baseline and memory-optimized, bit-wise, version of algorithm
- Research Question: Are there commonly occurring pathways in sequences that can be generated with many unique pathways?

Materials & Methods



Results

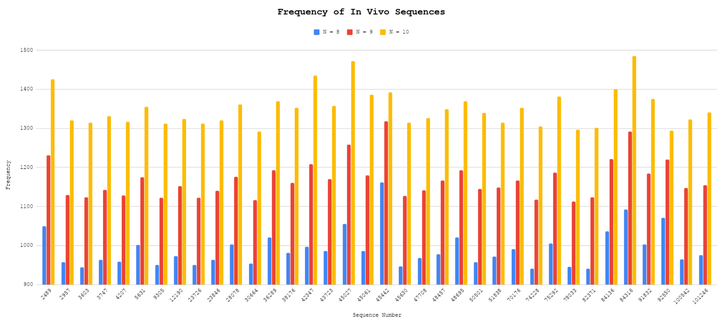


Fig. 1: High Matched in vivo Sequences for N-nucleotide Lengths of N = 8, 9, 10
The chart above is a representation of ~50 of the most frequently matched sequences for each N-length from 8 to 10.

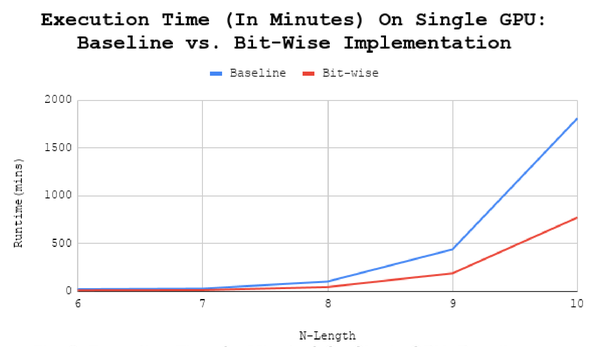
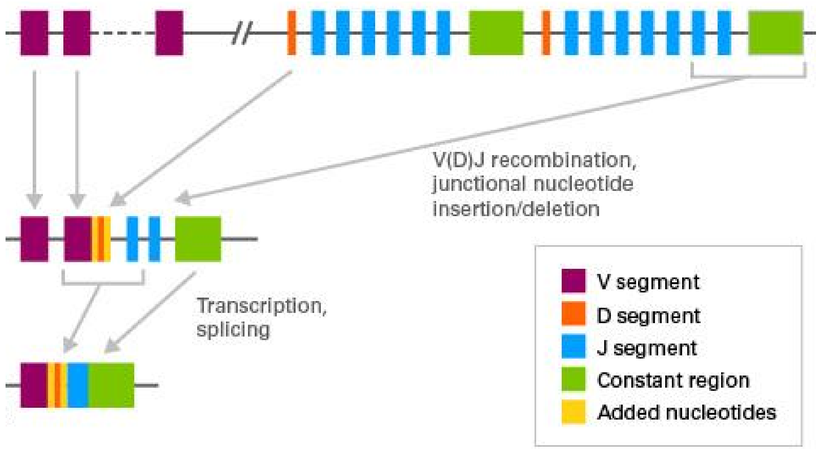


Fig. 2: Execution Time (In Minutes) On Single GPU: Baseline vs. Bit-Wise Implementation for Each N-nucleotide Length (6 to 10)

DNA Recombination

- V(D)J Recombination requires peta scale level data processing



Algorithm : Pseudocode for V(D)J Recombination Algorithm

Input : V, J, D and $n-nucleotide$ sequences
Output : Number of times each unique *in vivo* sequence is generated (Counter)

```
for i = 0 to number of V sequences do
    for j = 0 to number of J sequences do
        for k = 0 to number of D sequences do
            for m = 0 to number of n − nucleotide sequences do
                for n = 0 to number of in vivo sequences do
                    Combination = CombineString(V[i], n[m], D[k], n[m], J[j]);
                    for p = n − m to 0 do
                        if Combination == invivo[n] then
                            Counter[n]= Counter[n]+1;
                        move (N[m][p] → T[n − m − p])
```

Outcomes

- V(D)J Recombination uses in TCR Research:
  - Analyze diversity of the TCR repertoire
  - TCR application in immune response
  - Develop tools for methods such as machine learning
- Evaluation of the TCR repertoire needed to develop immunotherapy solutions and improve quality of life.
- Execution time reduction
  - General-purpose processor -> baseline: 50 weeks -> ~1.7 days
  - Baseline -> bit-wise: ~1.7 days -> ~18.5 hours
- Pathway profiling shows high degree of correlation between n-nucleotide length and the number of pathways
- Further research to understand relationship between V, D, J sequences, and n-nucleotides
- Our research aims to enable immunologists and other domain experts to rapidly visualize and identify these genes that play a significant role in the immune response.

College of Pharmacy
THE UNIVERSITY OF ARIZONA
KEYS
Research Internship