

**Rongxin Fang**

Bioinformatics, UC San Diego

E-mail: [r3fang@ucsd.edu](mailto:r3fang@ucsd.edu)

<https://github.com/r3fang/tfc>

## Object:

Develop a gene fusion detection tool for clinical purpose with following features:

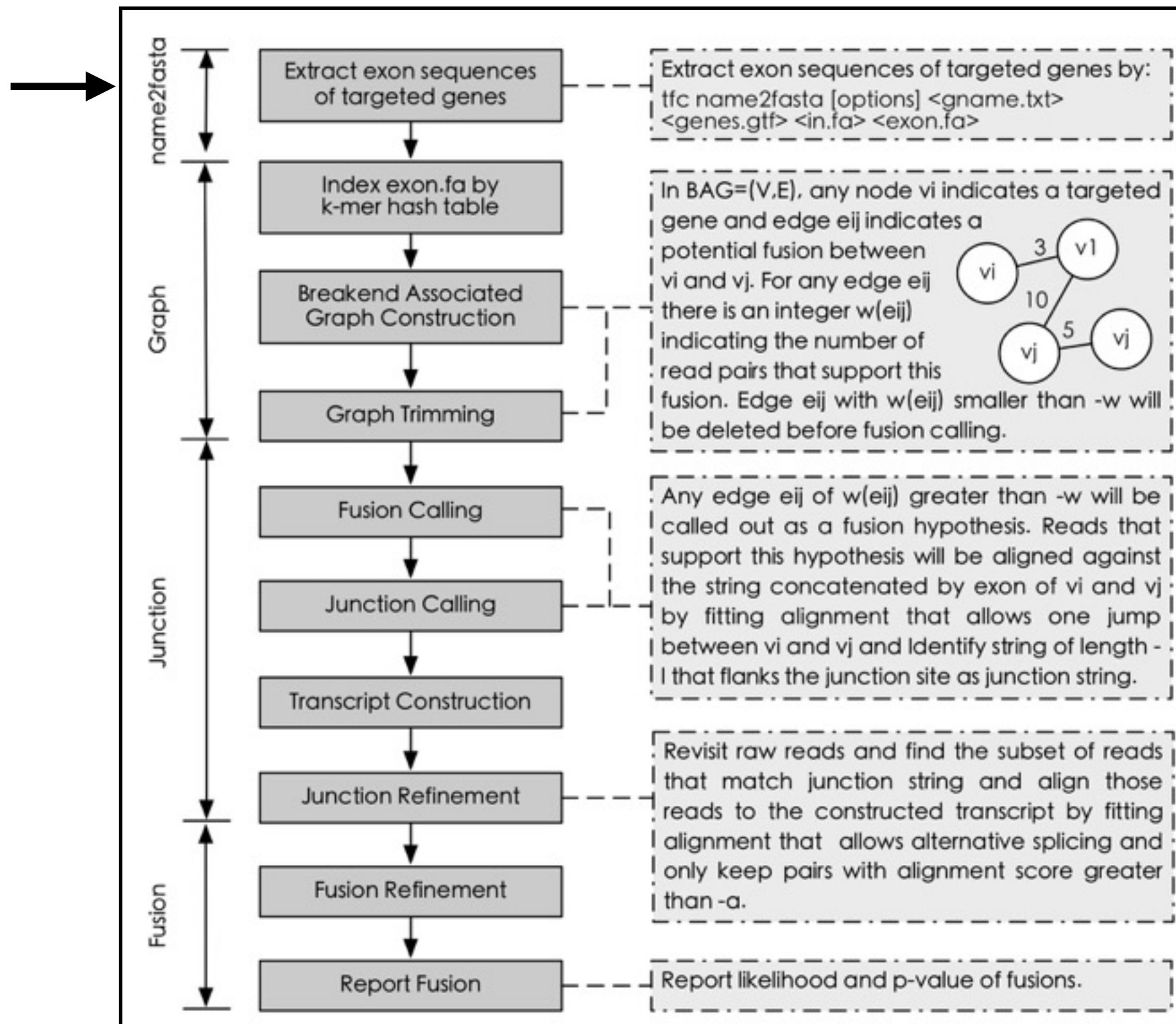
1. Precise
2. Fast
3. User friendly
4. Affordable

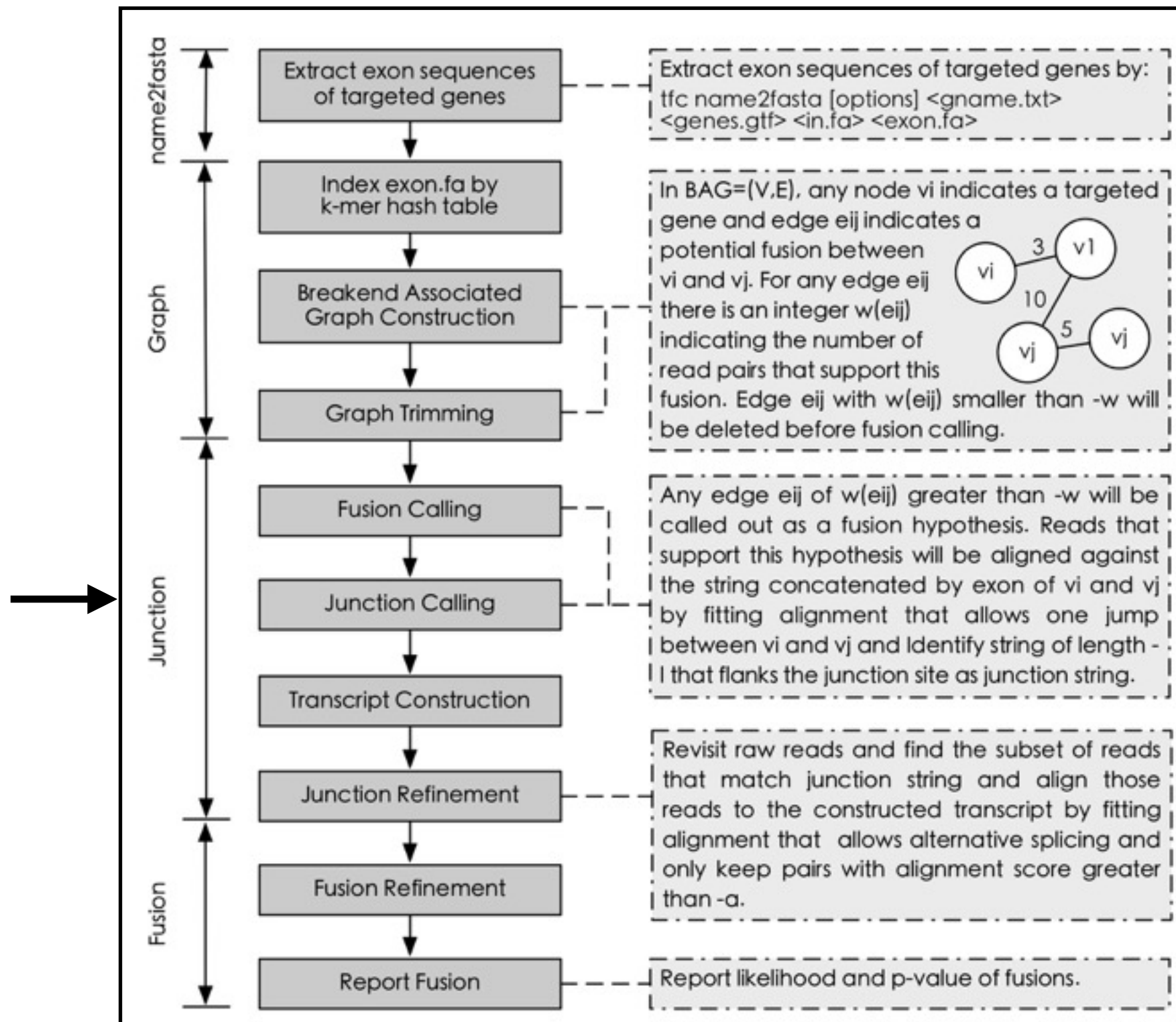
## Input:

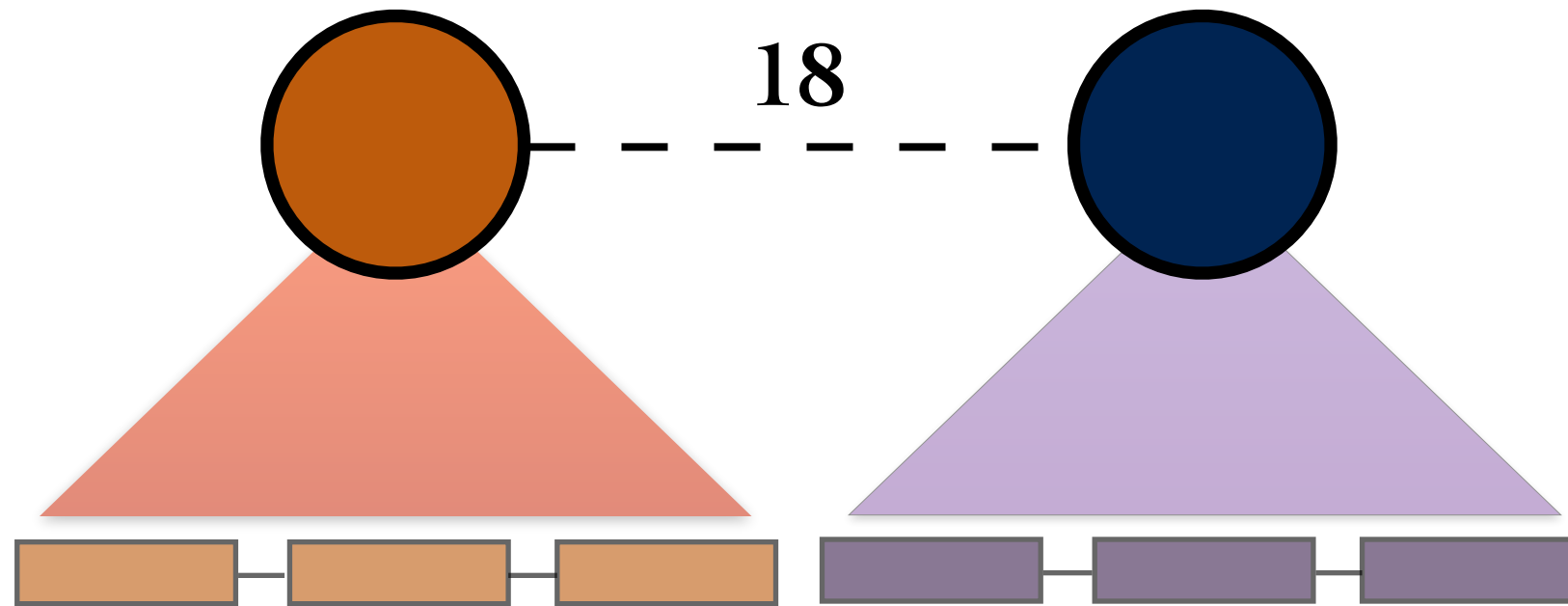
RNA-seq (.fq)

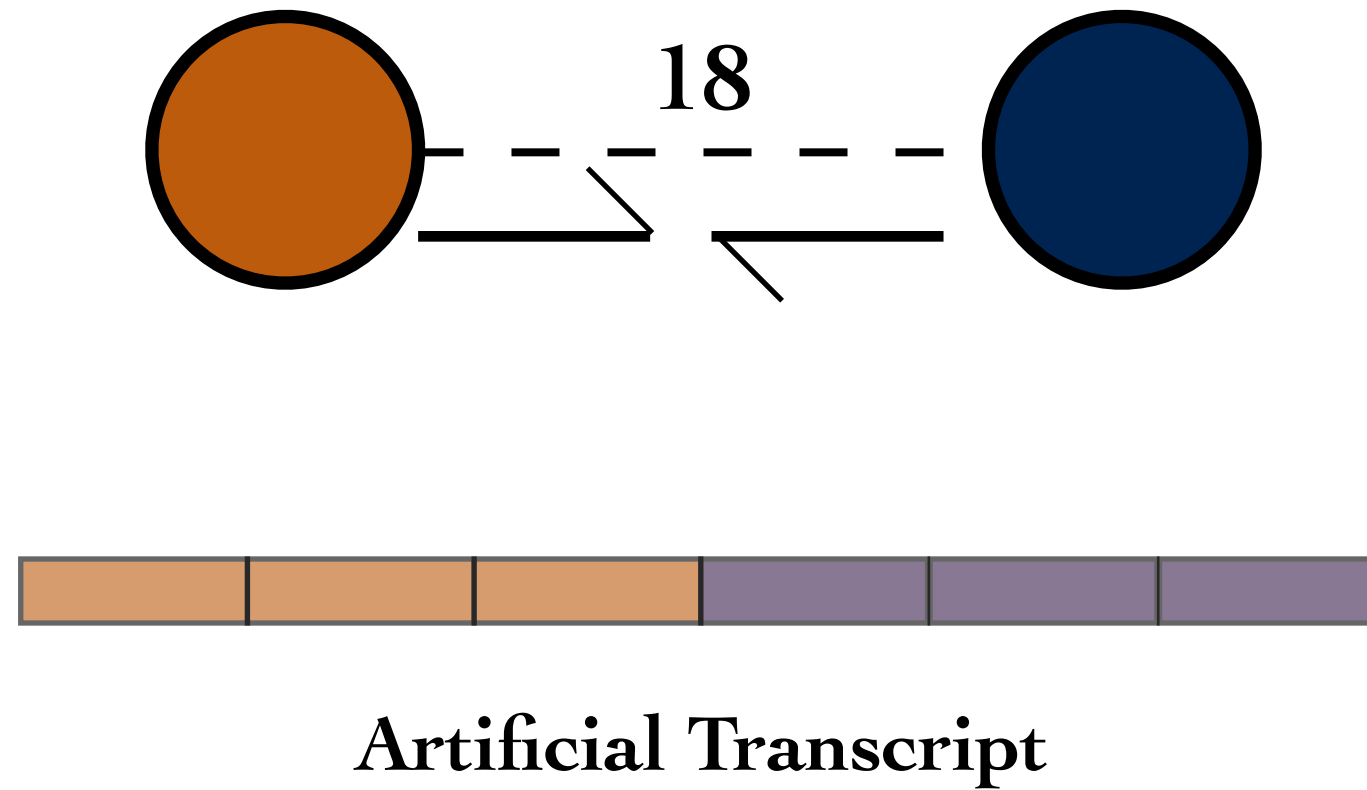
## Output:

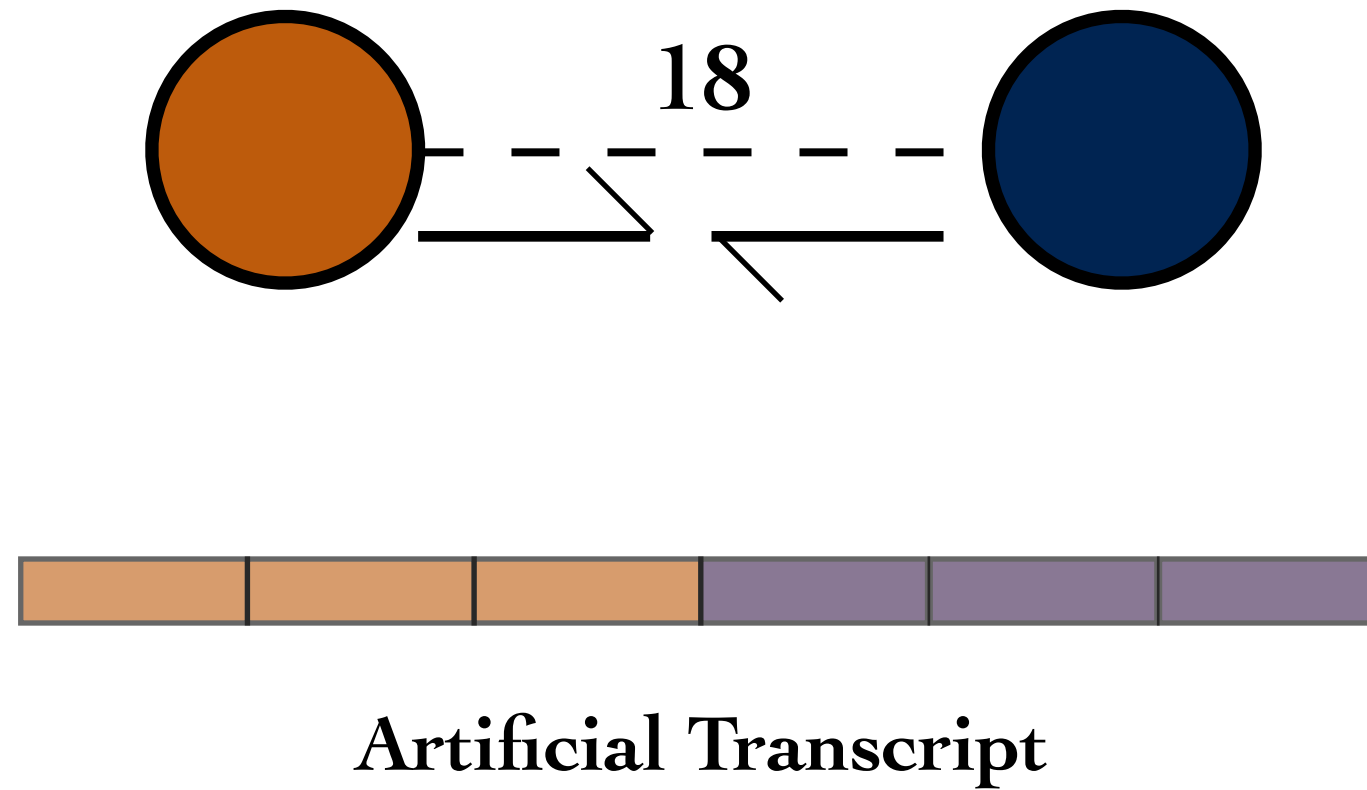
A list of gene fusion candidates with likelihood and p-value.

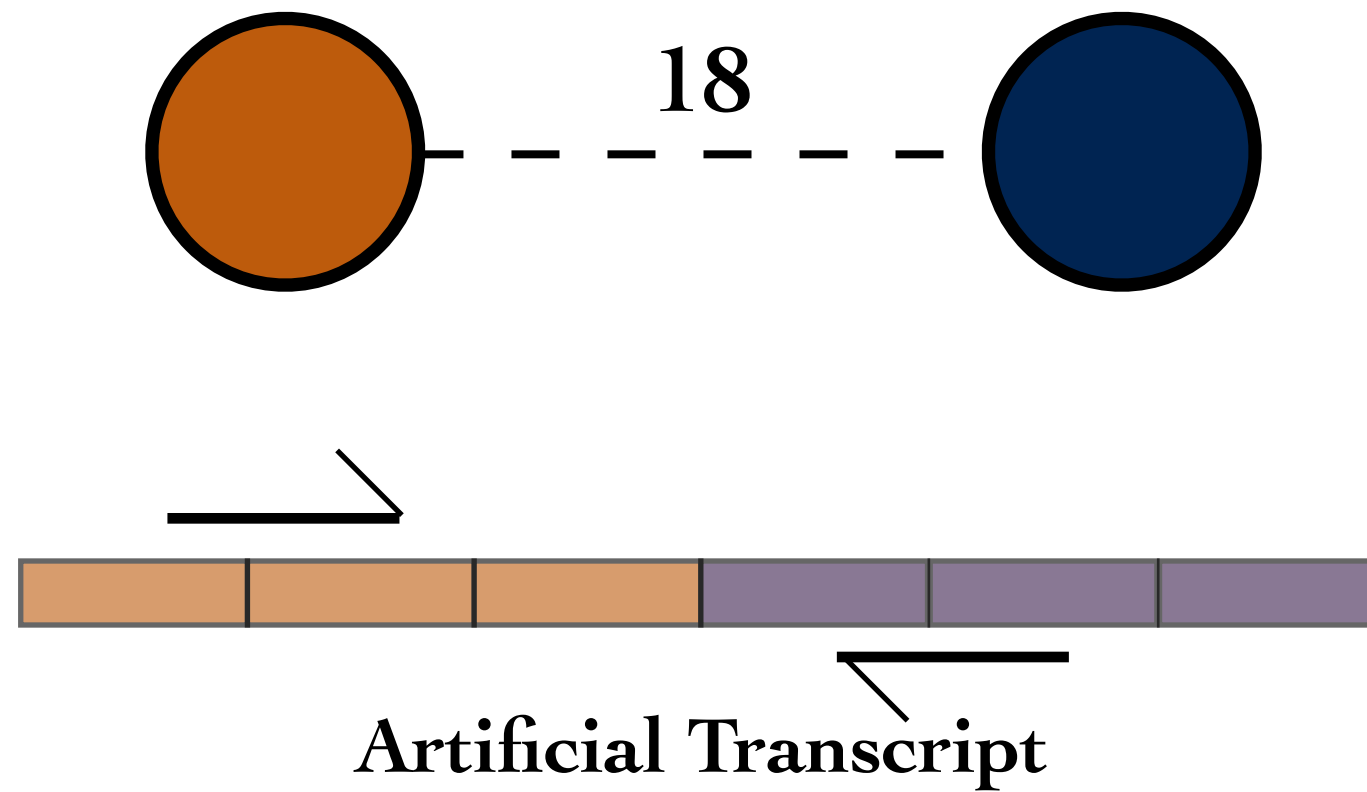




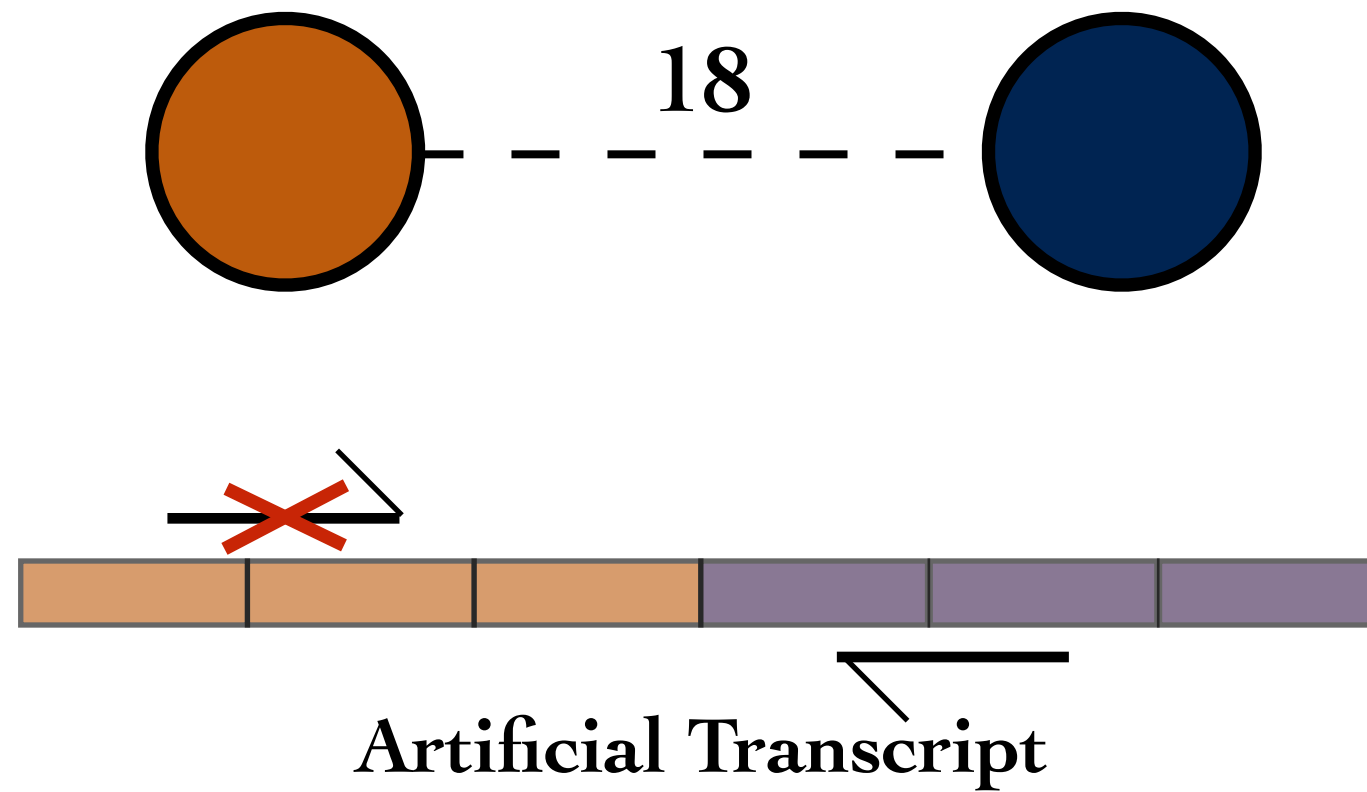


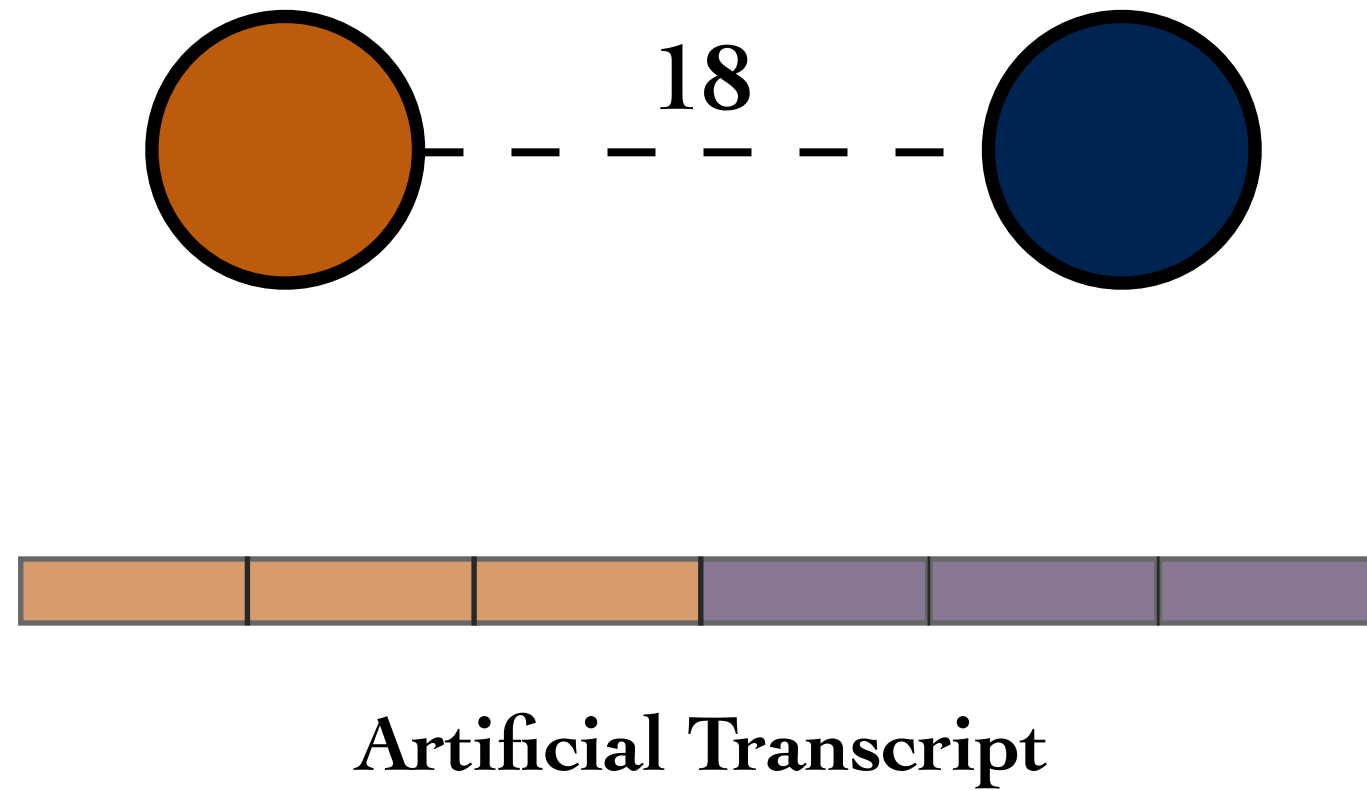




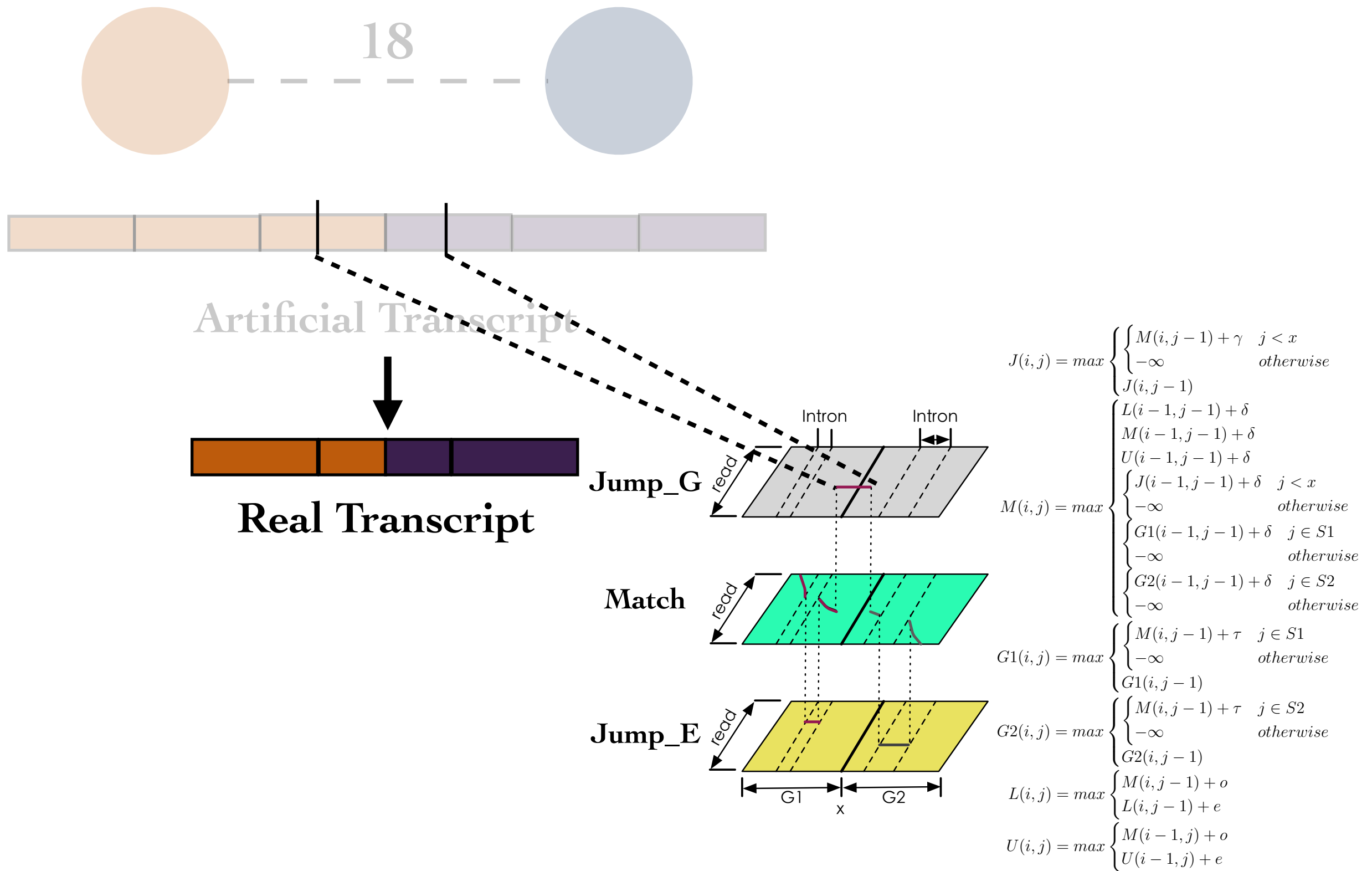


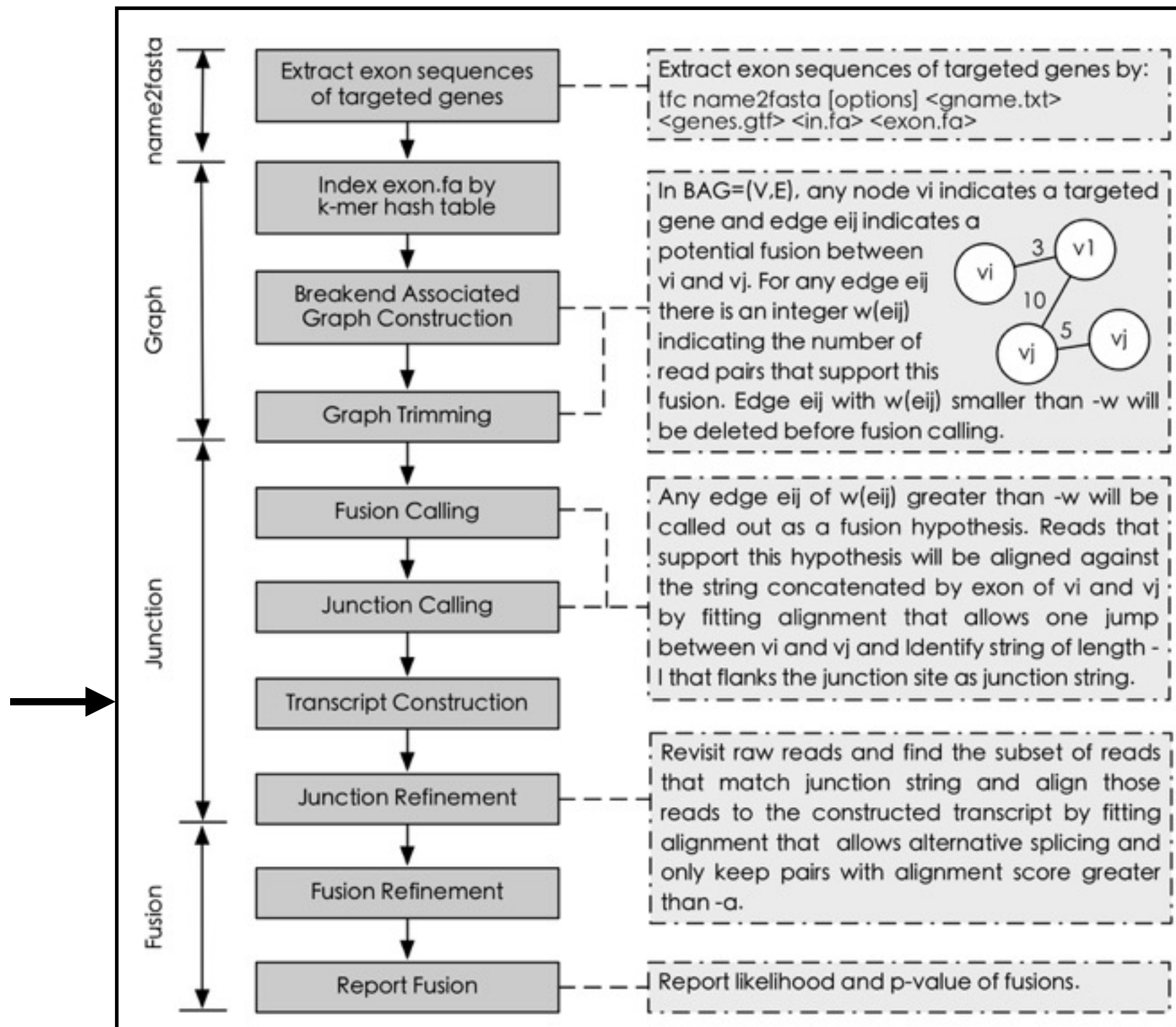




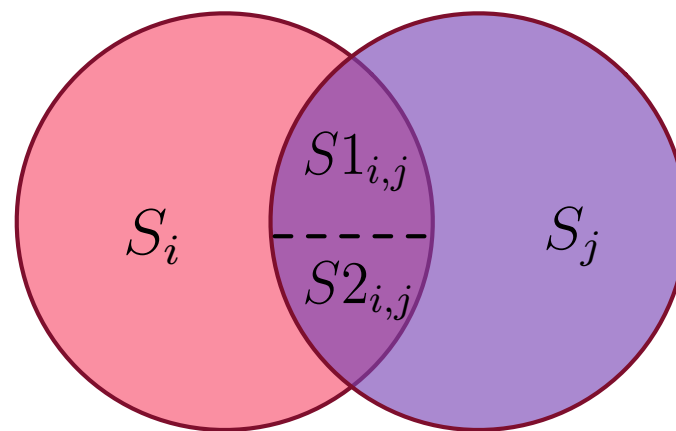


# Fitting Alignment with Affine Gap and 2 Jump States





likelihood equals the product of alignment probability of the reads that support the fusion normalized by the sequencing depth.

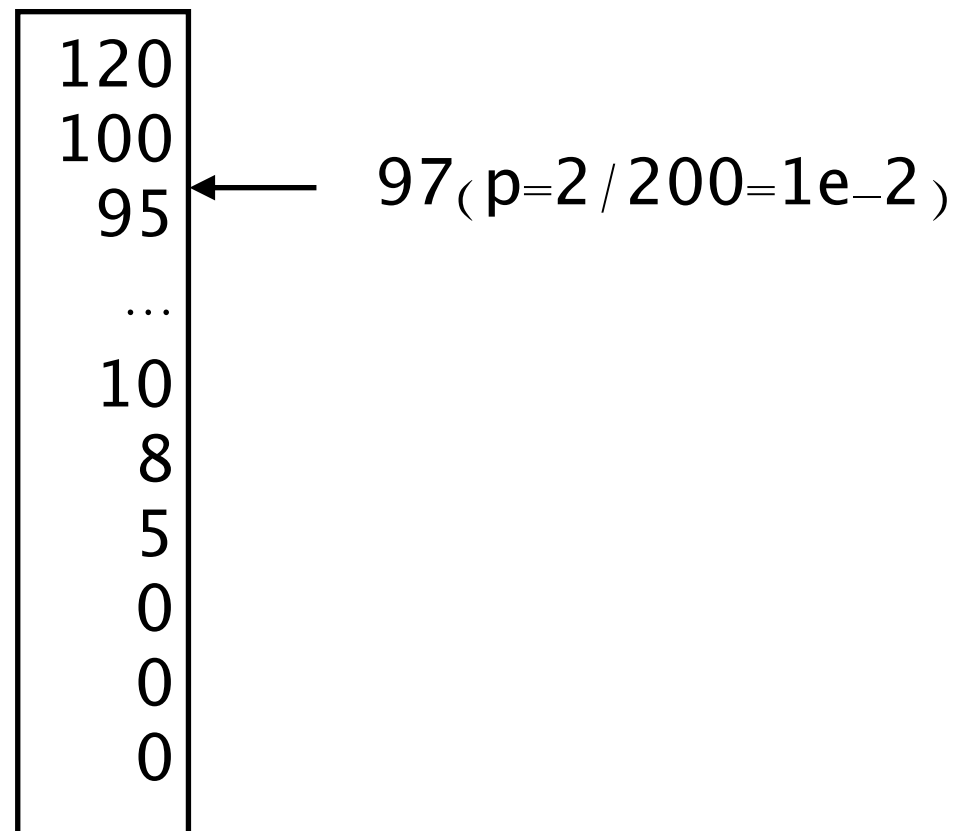


$$L_{i,j} = \frac{-\alpha \sum_s^{S1_{i,j}} \log(1 - f(s)) - \beta \sum_s^{S2_{i,j}} \log(1 - f(s))}{|S_i| + |S_j|}$$

in which  $f(s)$  is alignment probability,  $\alpha = 3$  and  $\beta = 1$ .

p-value is the probability of observing given likelihood in the normal sample.

gene\_i vs gene\_j



Fusion	Hits	Likelihood	P-value
TMPRSS2-ETV1	207	173166.20	1E-10

Se=90%

Sp=100%

Fusion	Hits	Likelihood	P-value
TMPRSS2-ETV1	207	173166.20	1E-10
BCAS4-BCAS3	70	20250.15	1E-10
BRD4-NUTM1	38	3903.03	1E-10
EWSR1-FLI1	35	2608.38	1E-10
ETV6-NTRK3	34	17625.09	1E-10
EML4-ALK	28	4238.51	1E-10
EWSR1-ATF1	24	1860.94	1E-10
HOOK3-RET	23	6593.92	1E-10
CD74-ROS1	6	4468.12	1E-10
AKAP9-BRAF	NA	NA	NA



## Object:

Develop a fusion detection tool for clinical usage with following features:

1. Precise
2. Fast
3. User friendly
4. Affordable

## How precise is TFC?

~0.85 and ~0.99 for Se and Sp on the simulated data.

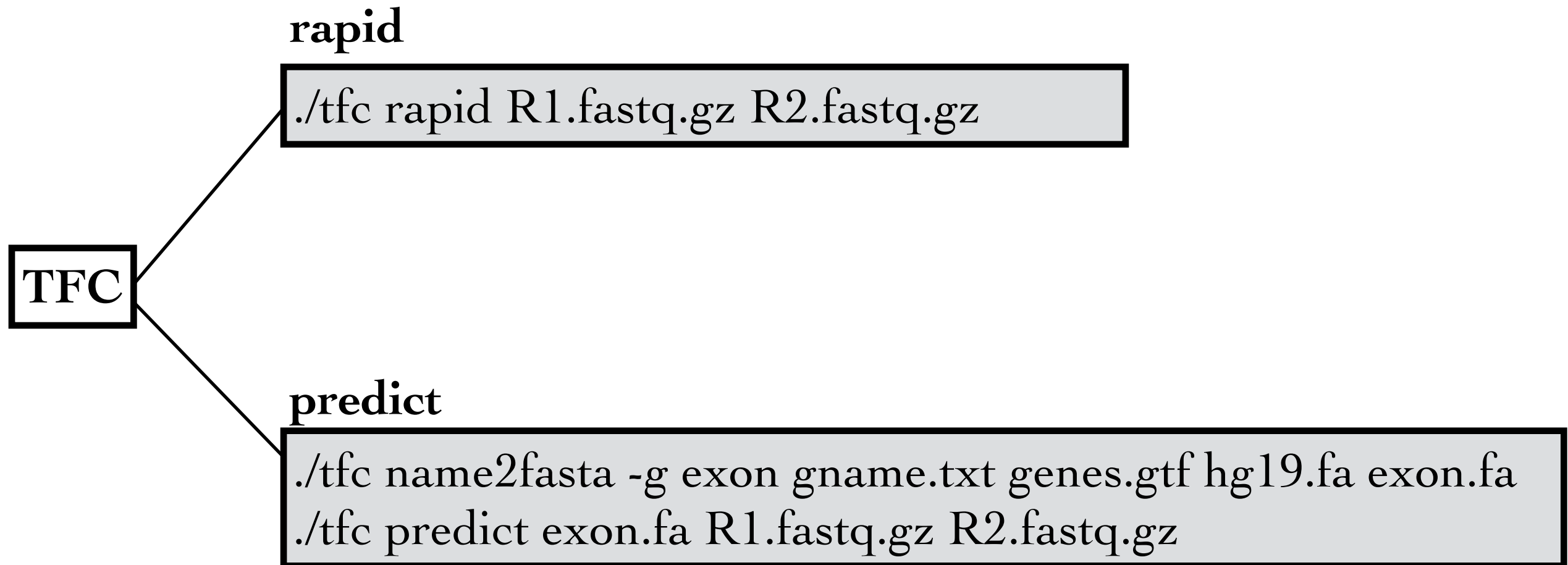
## How fast is TFC?

Implemented by C. ~5min per million read pairs using a single x86\_64 32-bit 2000 MHz GenuineIntel processor.

## What's the maximum memory requirement for TFC?

1GB would be the up limit for predicting against 500 genes.

## Is it easy to use?



Is it easy to use?

**Demo**

**Thanks!**