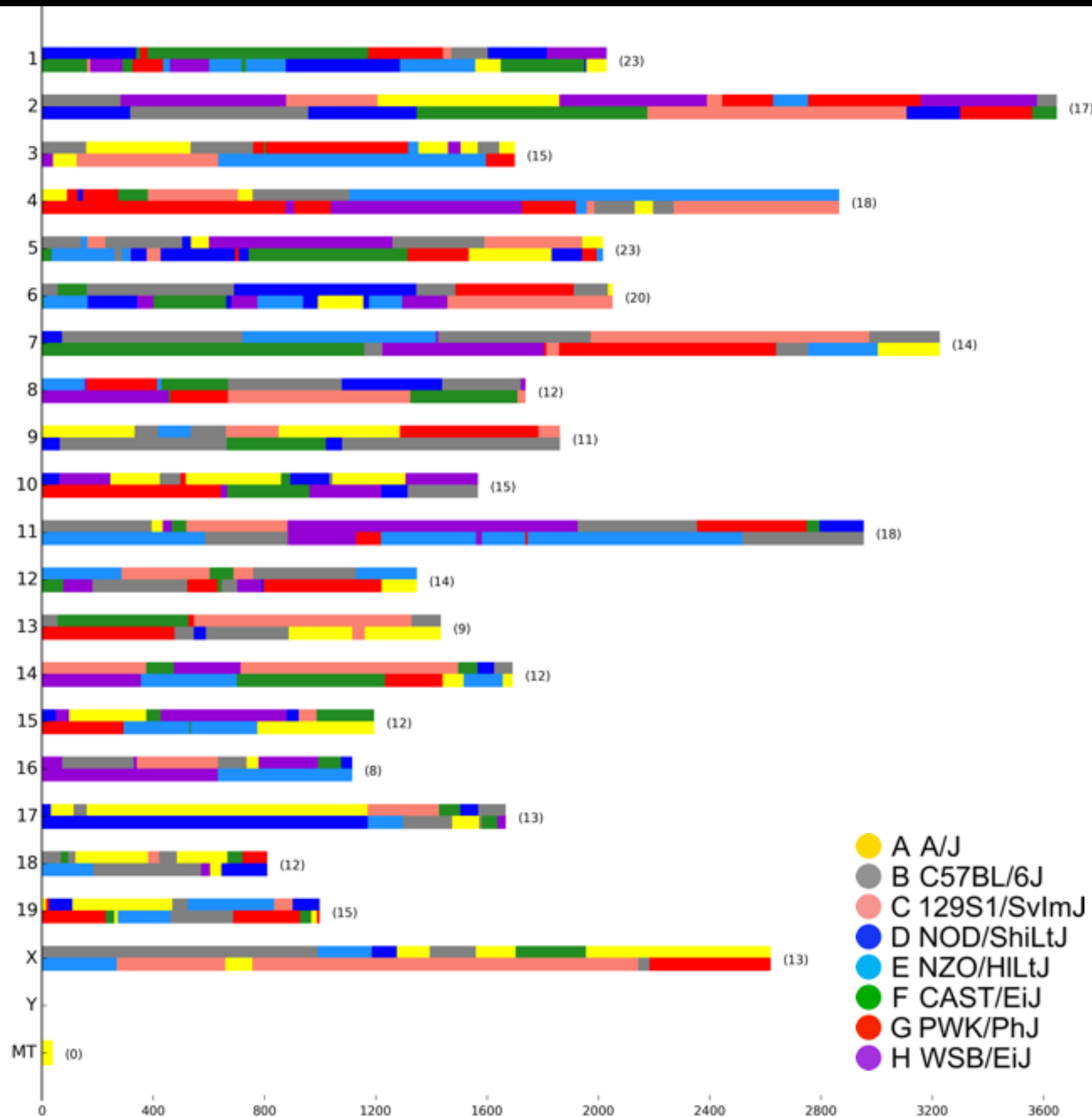
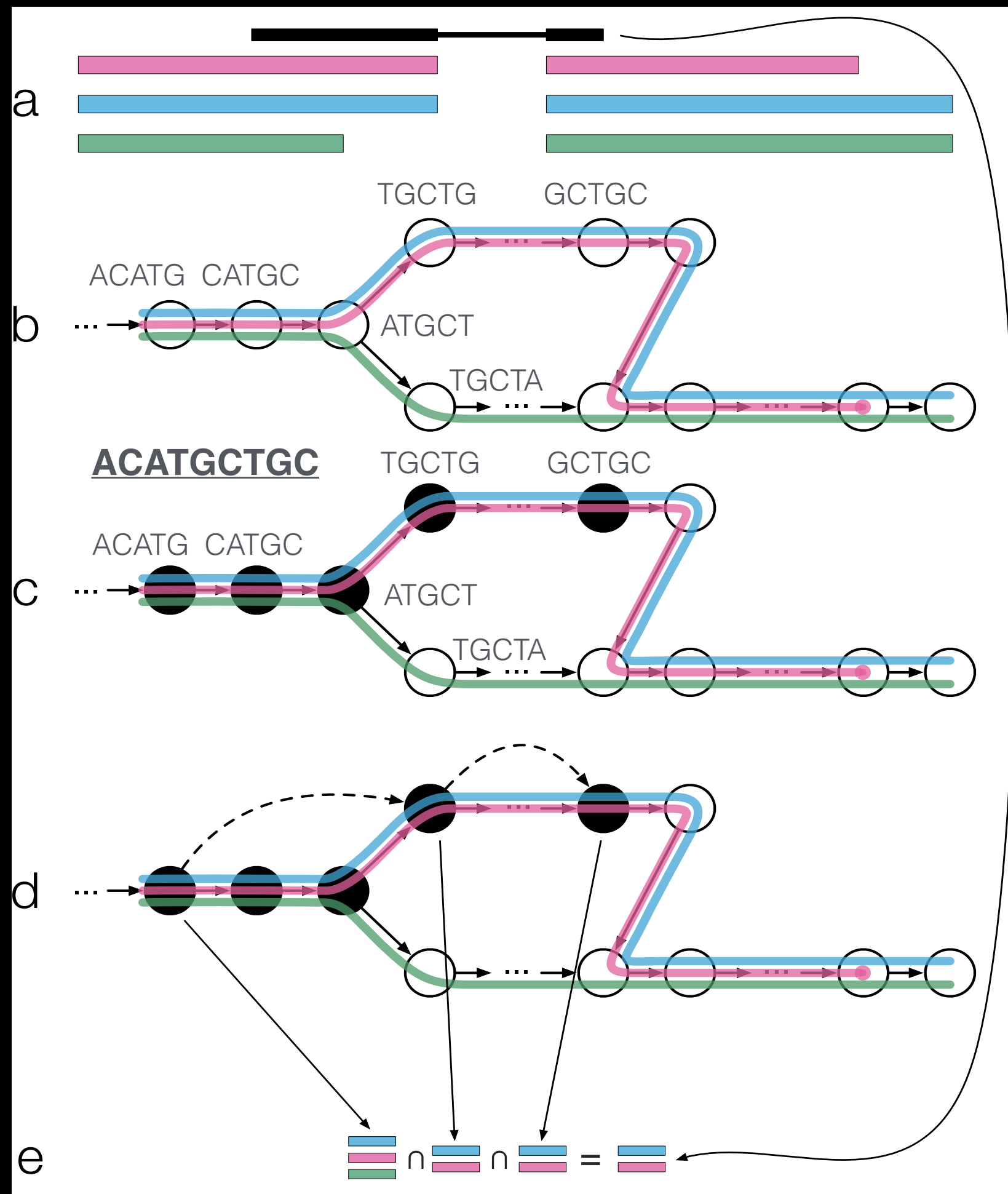


Quantification of Allele-Specific Expression



Indexing target sequences

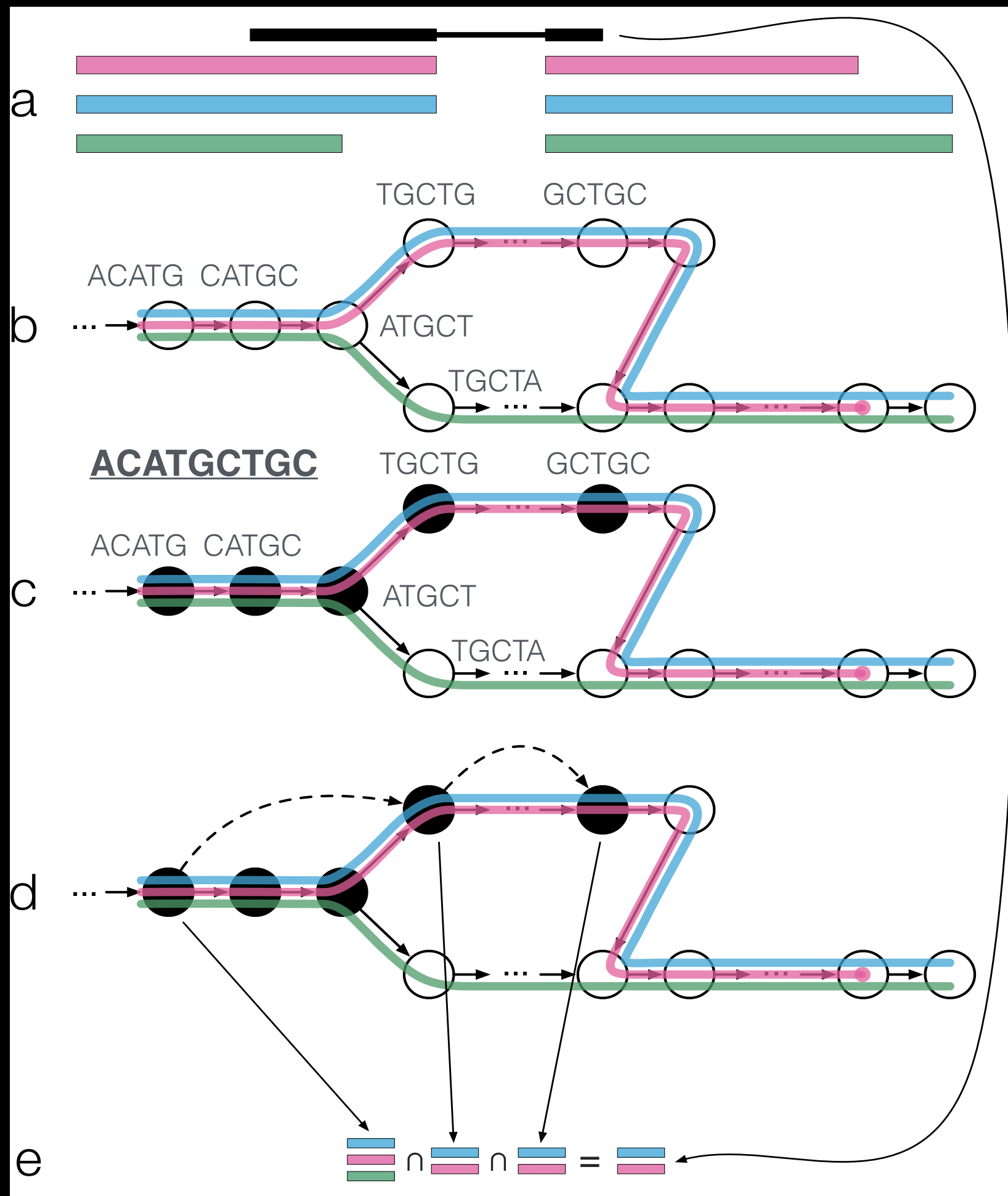
```
$ kallisto index \  
  -i individualized.transcriptome.idx \  
  /kbdata/individualized.transcriptome.fa.gz
```



```
[build] loading fasta file individualized.transcriptome.fa.gz
[build] k-mer length: 31
[build] warning: clipped off poly-A tail (longer than 10)
           from 824 target sequences
[build] warning: replaced 6 non-ACGUT characters in the input
sequence with pseudorandom nucleotides
[build] counting k-mers ... done.
[build] building target de Bruijn graph ... done
[build] creating equivalence classes ... done
[build] target de Bruijn graph has 1162115 contigs and contains
92155258 k-mers
```

Identifying the origin of raw reads

```
$ kallisto-align \  
  -f /kbdata/rawreads.fastq.gz \  
  -i individualized.transcriptome.idx \  
  -b pseudo-alignments.EC.bin
```



```
[kallisto-align] Creating pseudo-alignment.EC.bin...
[index] k-mer length: 31
[index] number of targets: 184699
[index] number of k-mers: 92155258
[index] number of equivalence classes: 545307
[kallisto-align] Mode: single-end
[kallisto-align] Processing 1: rawreads.fastq.gz
[kallisto-align] Creating (1) Equivalence Class File...
[kallisto-align] Pseudo-Aligning...
[kallisto-align]   Total Targets: 184699
[kallisto-align]   Main Targets: 94121
[kallisto-align]   Haplotypes: 8
[kallisto-align]   Reads: 1995221
[kallisto-align]   Mapped Reads: 1782790
[kallisto-align]   Alignments: 4218347
[kallisto-align] Done
```


Quantifying allele specific expression

```
$ emase-zero -v --model 4 \  
  -b pseudo-alignments.EC.bin \  
  -o emase.m4.expected_read_counts
```

emase-zero Version 0.2.1

Alignment File: pseudo-alignments.EC.bin

Grouping File: None

Transcript Length File: None

EM Model: 4

Output File: emase.m4.expected_read_counts

Loading pseudo-alignment.EC.bin...

File had the following haplotype names:

A B C D E F G H
81155 alignment classes loaded (1782790 total reads)
94121 transcripts

Time for initializing stack sum = 0.026706s

Beginning EM Iterations...

Iter No	Time(s)	Change
-----	-----	-----
1	0.0369	275232.6
2	0.035	130600.5
3	0.035	81413.0
4	0.035	58510.9
5	0.035	45375.6
6	0.035	36858.0
...		
49	0.035	1921.7
50	0.035	1849.9
51	0.035	1781.7

Time for 51 iterations = 1.9s

Time per iteration 0.04s

Saving results to emase.m4.expected_read_counts

Done.

Loading alignments onto the EMASE framework

```
$ kallisto-to-emase \  
  -i pseudo-alignments.EC.bin \  
  -a pseudo-alignments.EC.h5
```

Loading alignments onto the EMASE framework

```
$ ipython
Python 2.7.9 (default, Mar 1 2015, 12:57:24)

Type "copyright", "credits" or "license" for more information.

IPython 2.3.0 --An enhanced Interactive Python.
?                -> Introduction and overview of IPython's features.
%quickref        -> Quick reference.
help             -> Python's own help system.
object?         -> Details about 'object', use 'object??' for extra details.


In [1]: from emase import AlignmentPropertyMatrix as APM

In [2]: alnmat = APM(h5file='pseudo-alignments.EC.h5')

In [3]: alnmat.hname
Out[3]: ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H']
```

Computation with pseudo-alignments

```
In [4]: alnmat.print_read(2)  # The 3rd equivalence class
```

```
2
```

```
--
```

```
ENSMUST00000153787 [ 0.  0.  0.  0.  1.  0.  0.  1.]
ENSMUST00000107904 [ 0.  0.  0.  0.  1.  0.  0.  1.]
ENSMUST00000107903 [ 0.  0.  0.  0.  1.  0.  0.  1.]
ENSMUST00000018572 [ 0.  0.  0.  0.  1.  0.  0.  1.]
ENSMUST00000143720 [ 0.  0.  0.  0.  1.  0.  0.  1.]
```

```
In [5]: alncnt = alnmat.count_alignments()
```

```
In [6]: unqcnt = alnmat.count_unique_reads()
```

```
In [7]: unqcnt.sum()
```

```
Out[7]: 10132.0
```

```
In [8]: alncnt.sum()
```

```
Out[8]: 7491317.0
```

```
In [9]: alnmat.count.sum()
```

```
Out[9]: 1782790
```