



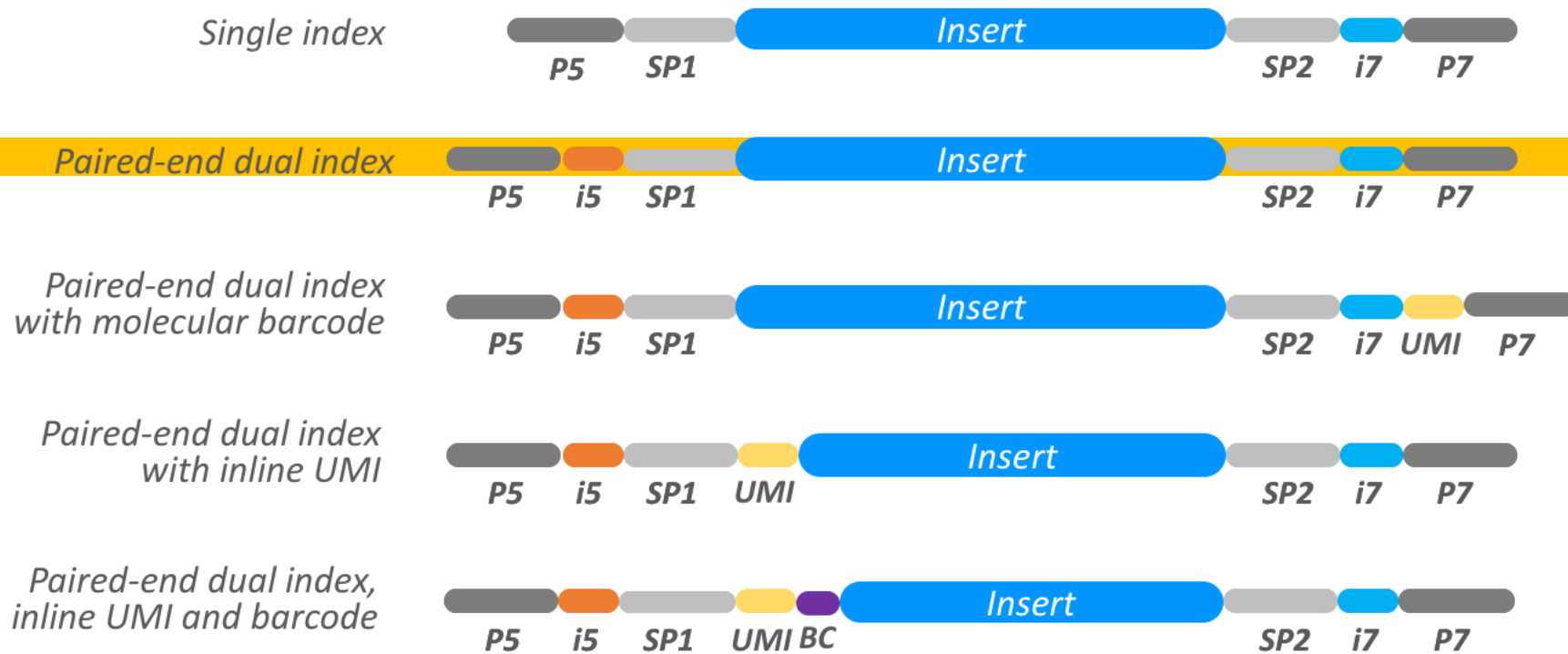
Ligated sequences

- Adapters
- Primers
- Tags
- Barcodes
- UMIs
- Spacers
- Linkers

Why to ligate anything?

- 1) Bind to a flow cell for next generation sequencing
- 2) Allow for PCR enrichment of adapter-ligated DNA fragments only
- 3) Allow for indexing or “barcoding” of samples so multiple DNA libraries can be mixed together into 1 sequencing lane (known as multiplexing)
- 4) Error discovery with tags

Adapters



- P5/P7:** Flow cell binding sequences (platform-specific)
- SP1/2:** Sequencing primer binding sites (common for all libraries)
- i5/i7:** Sample Indexes (specific to a particular library)
- UMI:** Unique molecular index (barcode tag for individual molecules)
- BC:** User-defined barcode (unique per sample, single cell, etc.)
- Insert:** Target DNA or cDNA fragment (library-specific)

Ligated sequences

Have to be present:

P5/P7 – adapters for flowcell binding

SP1/SP2 – sequencing primer binding site

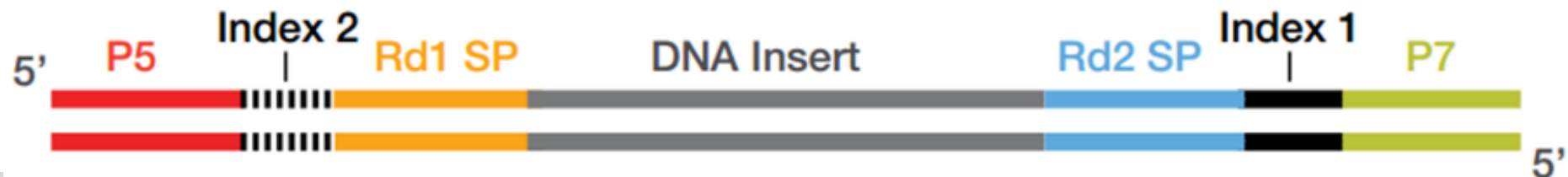
Optional – but often used

i5/i7 – Sample index – to recognize sequenced libraries

Optional:

Barcode – unique for sample, cell

UMI – Unique Molecular Identifier – to identify technical duplicates



Ligated sequences

Spacers

- If combining different lengths libraries

Linkers

- for better merging of sequences

Demultiplexing

Sorting reads according to sample. Based on index/barcode.

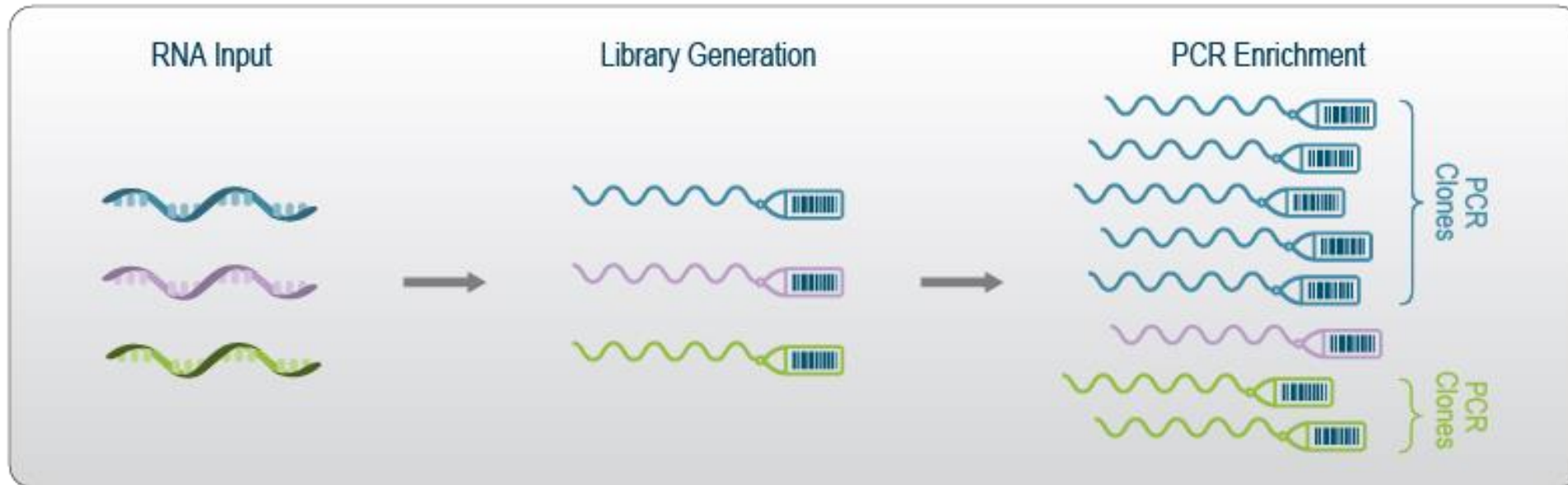
Forward barcode: **AGGCT**

Reverse barcode: **CAATG**

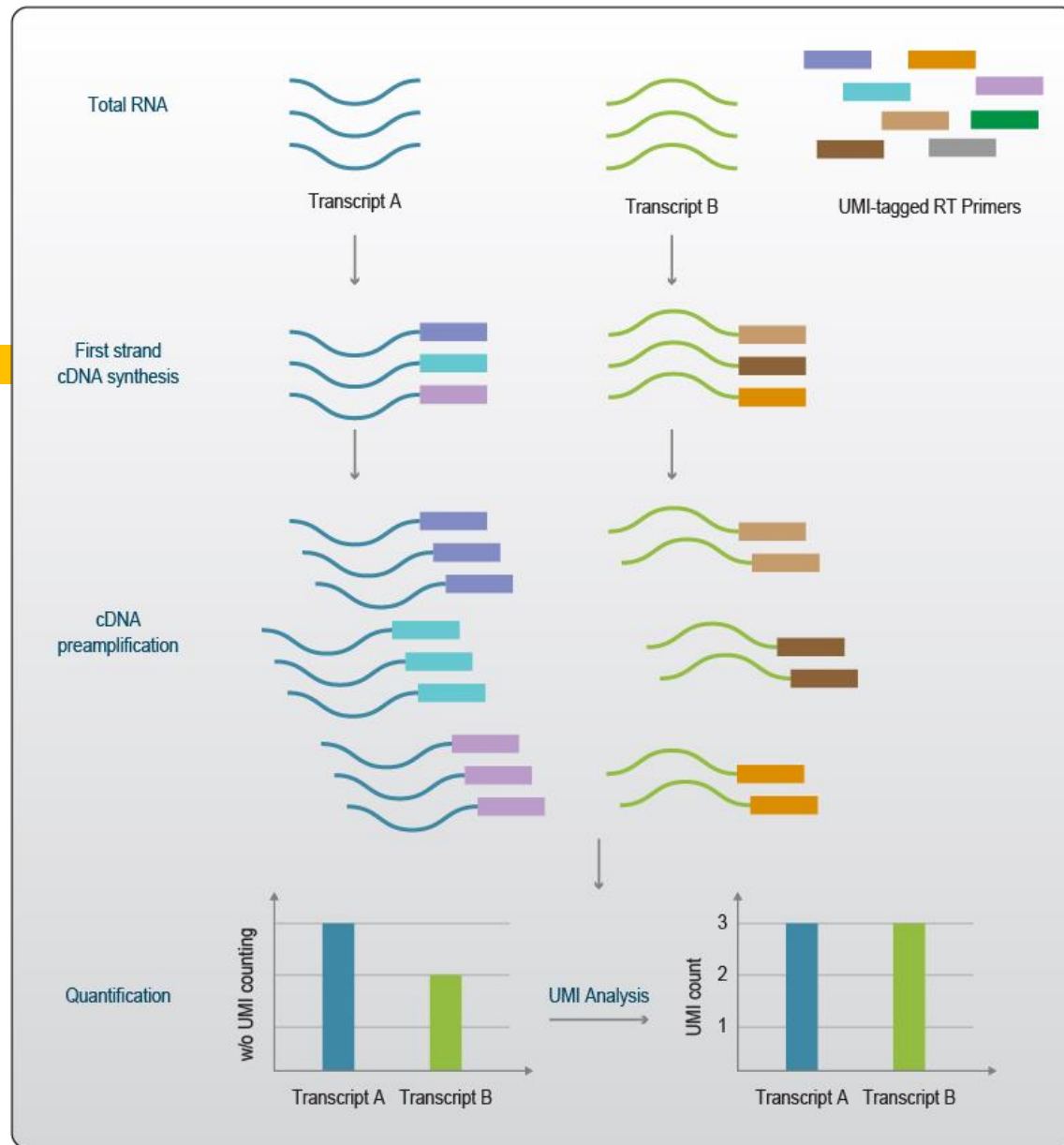
3' – **T C C G A** A T G G C C A T C C G **G T A A C** – 5'
5' – **A G G C T T A C C G G** T A G G C **C A T T G** – 3'

```
>read_1
AGGCTATTTAGCGCTACGTAATTTAGCCAATG
>read_2*
AGGCTATTTAATTTAATATTTAGCATTTAGCCATTG
>read_3
CATTGATTTAGCATTTAGCATATTTAGCAAGCCT
> read_4
AGGCTATATTTATTTAGCATATTTAGCATTCAATG
> read_5
CAATGATTTAGCATTTAGCATTTAGCATTAGGCT
> read_6
CAATGATTTAGCATATTTAATTTAGCATAGGCT
> read_7
AGCCTATTTAGCATTATTTAGCATTTAGCCATTG
> read_8*
CAATGATTTAATTTAGCATATTTAGCAGCCT
```

UMI

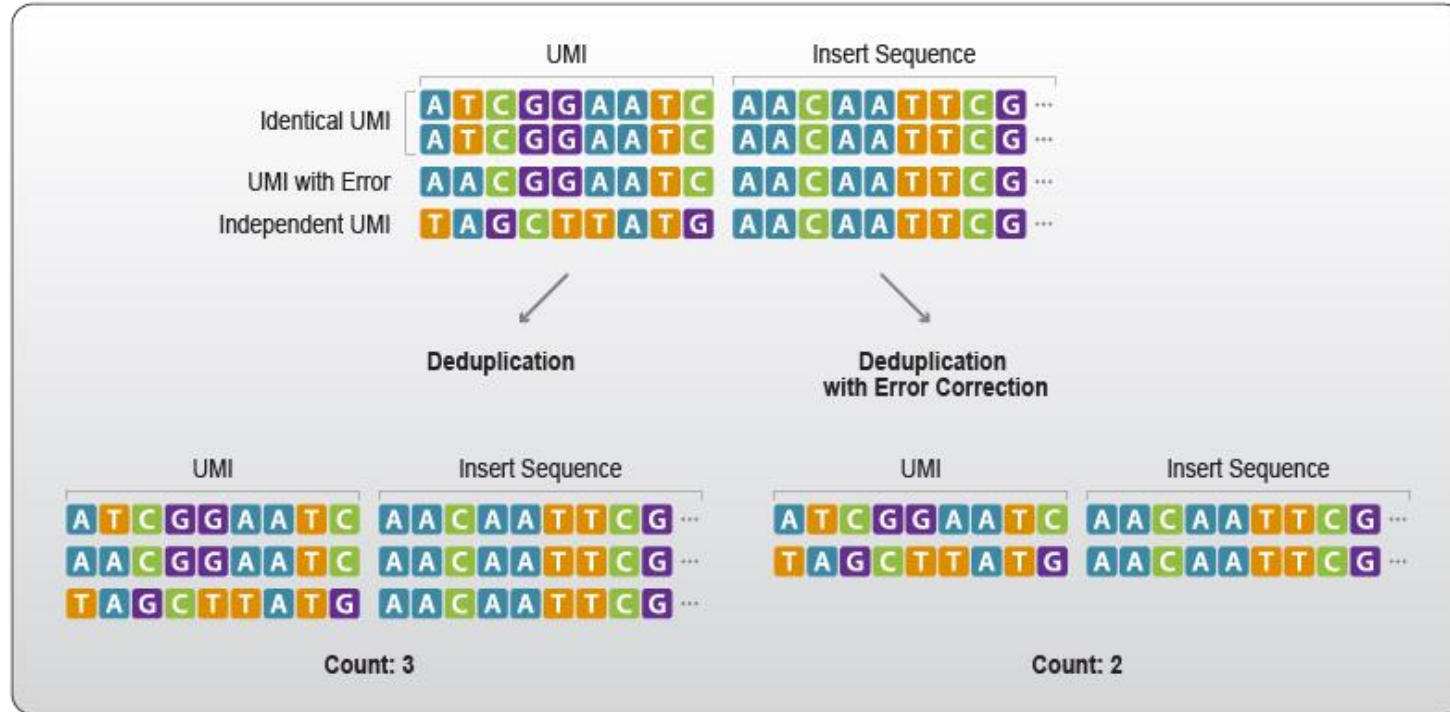


UMI



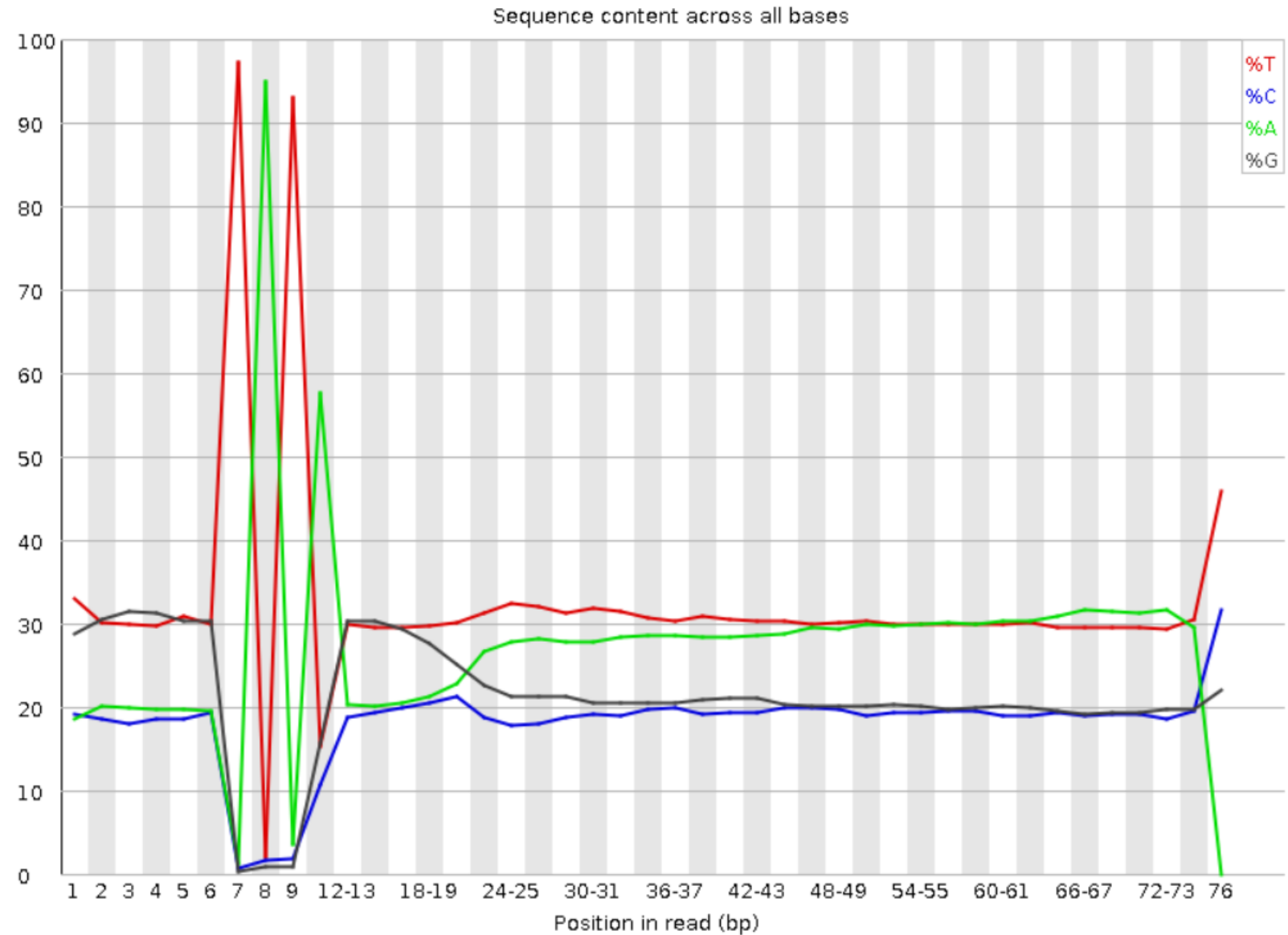
*<https://www.lexogen.com/rna-lexicon-what-are-unique-molecular-identifiers-umis-and-why-do-we-need-them/>

UMI



Linker

✖ Per base sequence content



Linker

Per Base Sequence Content

18

Help

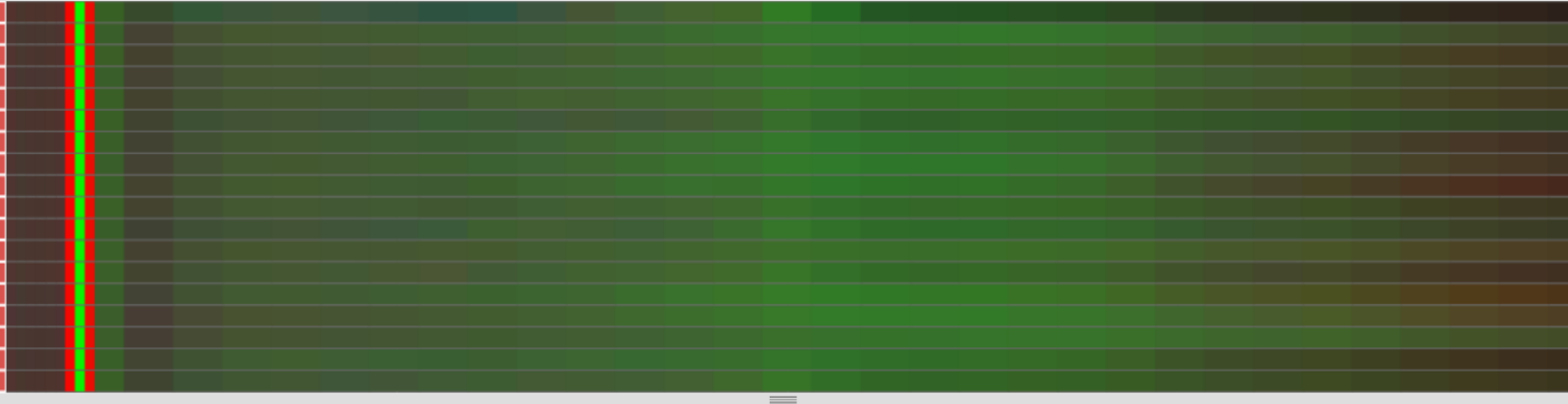
The proportion of each base position for which each of the four normal DNA bases has been called.

Click a sample row to see a line plot for that dataset.

Rollover for sample name

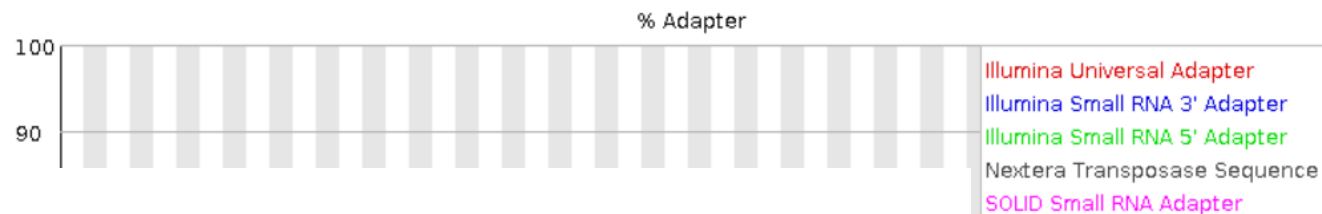
Position: - %T: - %C: - %A: - %G: -

Export Plot



Adapters

✖ Adapter Content



! Overrepresented sequences

Sequence	Count	Percentage	Possible Source
TATACCTTCTGCATAATGAATTAAGTAACTAGAAATAACTTTGCAAGGAGAGCC	33970	0.26410651026046406	No Hit
TATACTTACAACAATCCTAATCCTAATACCAACTATCTCCCTAATTGAA	21045	0.16361853130501813	No Hit
TATACTGTTAACGATTGTCTGCCCATGTCCTGCCTGAAATACCATGATTG	19574	0.15218194971558208	No Hit
TATACTCCTTACACTATTCCTCATCACCCAATAAAAAATATTAACACAA	17287	0.13440121409692796	No Hit
TATACCGCAGATTCAGGCATGGATTCCGTGAAGGAACAACACCTAAACCC	14760	0.11475455082262141	No Hit
TATAGGAAGAGCACACGTCTGAACTCCAGTCACAGCTTCTTTTCCATCTC	14187	0.11029964854475136	TruSeq Adapter, Index 11 (96% over 33bp)

