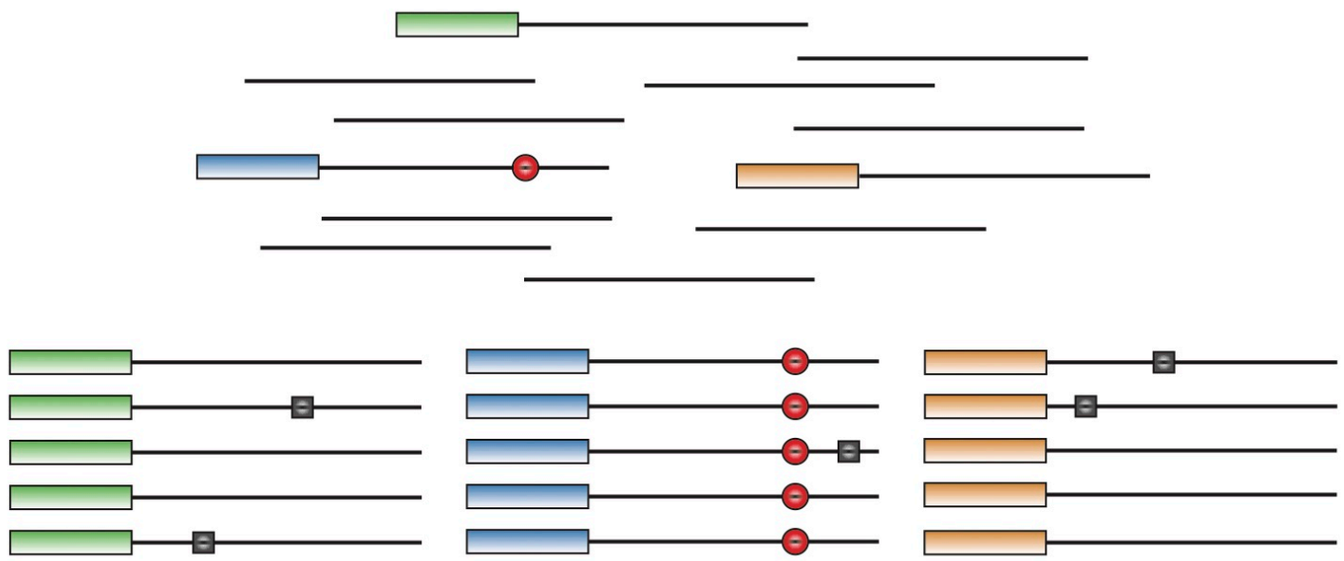
Unique Molecular Identifiers



Library Barcoding vs. UMIs

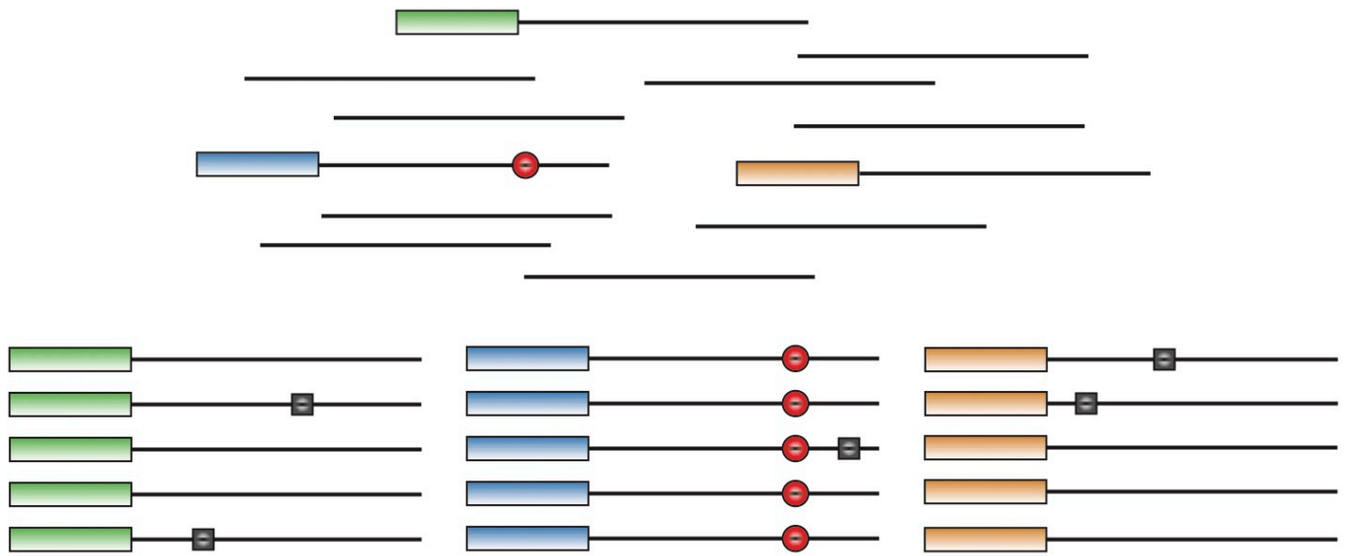
Library Index Sequences: Barcodes used to assign reads to their **sample or library** of origin
Example: Standard i5 and i7 Illumina indexing

Unique Molecular Identifiers: **Random** barcodes used to assign reads to their **original biological molecule** of origin

Applications for UMIs

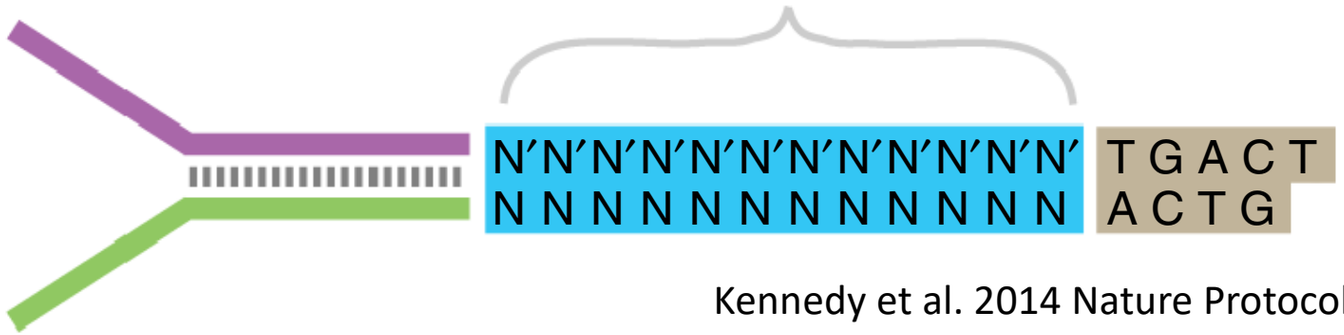
Remove PCR Duplicates: Library amplification results in pseudoreplicated data points, causing problems for applications that rely on read counting for quantifying gene expression, identifying variants, etc. UMIs can be used to identify and remove duplicated copies.

Increase Sequencing Accuracy: Redundant reads from the same original biological molecule can be used to generate consensus basecalls and filter out errors introduced during library amplification and sequencing.



Incorporation of UMIs into Libraries

Adjacent to Insert: Sequenced as first bases in the insert sequencing read(s)

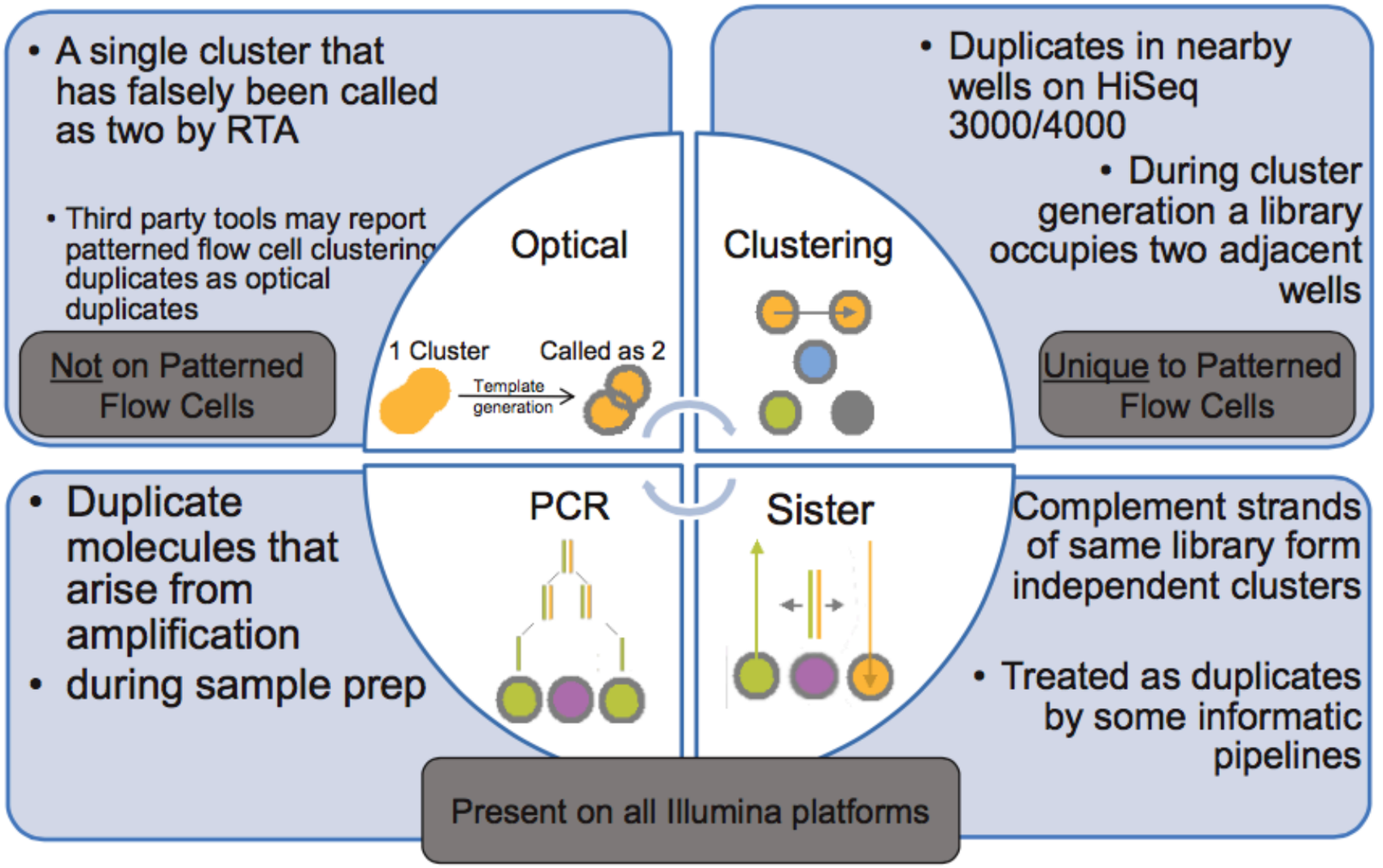


Kennedy et al. 2014 Nature Protocols

Adjacent to i7: Sequenced as part of the i7 library index read

(commercially available from IDT)

Duplicate Reads

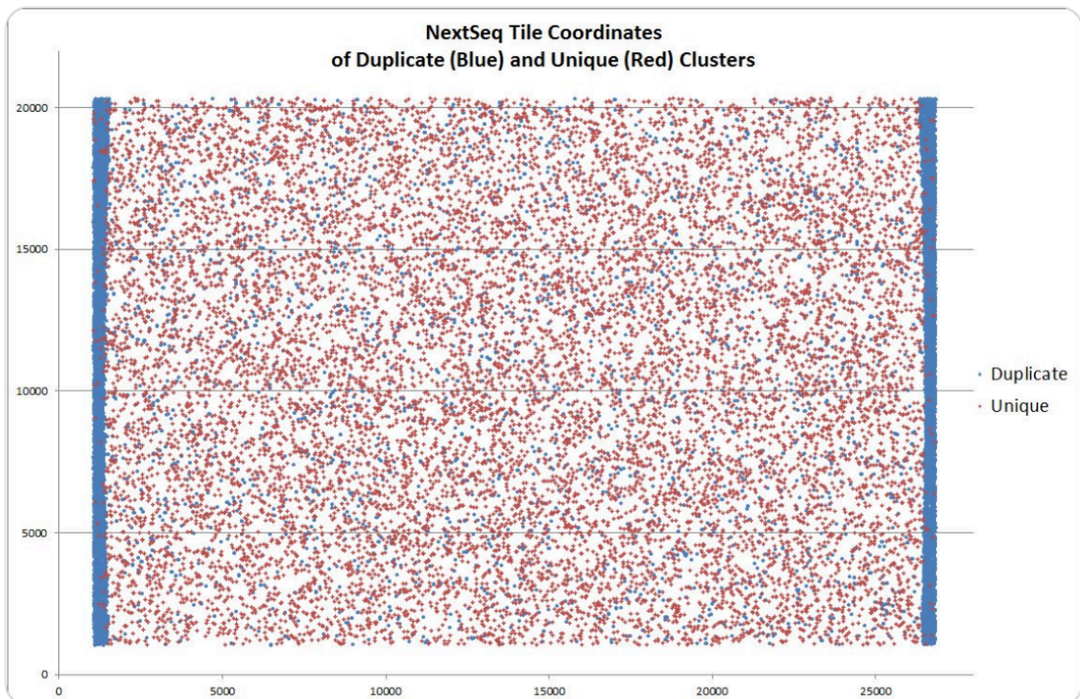


Duplicate Reads

Brian Bushnell
@BBToolsBio

Follow

New Illumina duplicate mode! Tile-edge duplicates account for >80% of our NextSeq duplicates.
More details:
biostars.org/p/229842/#2299…



Deduplicating Reads

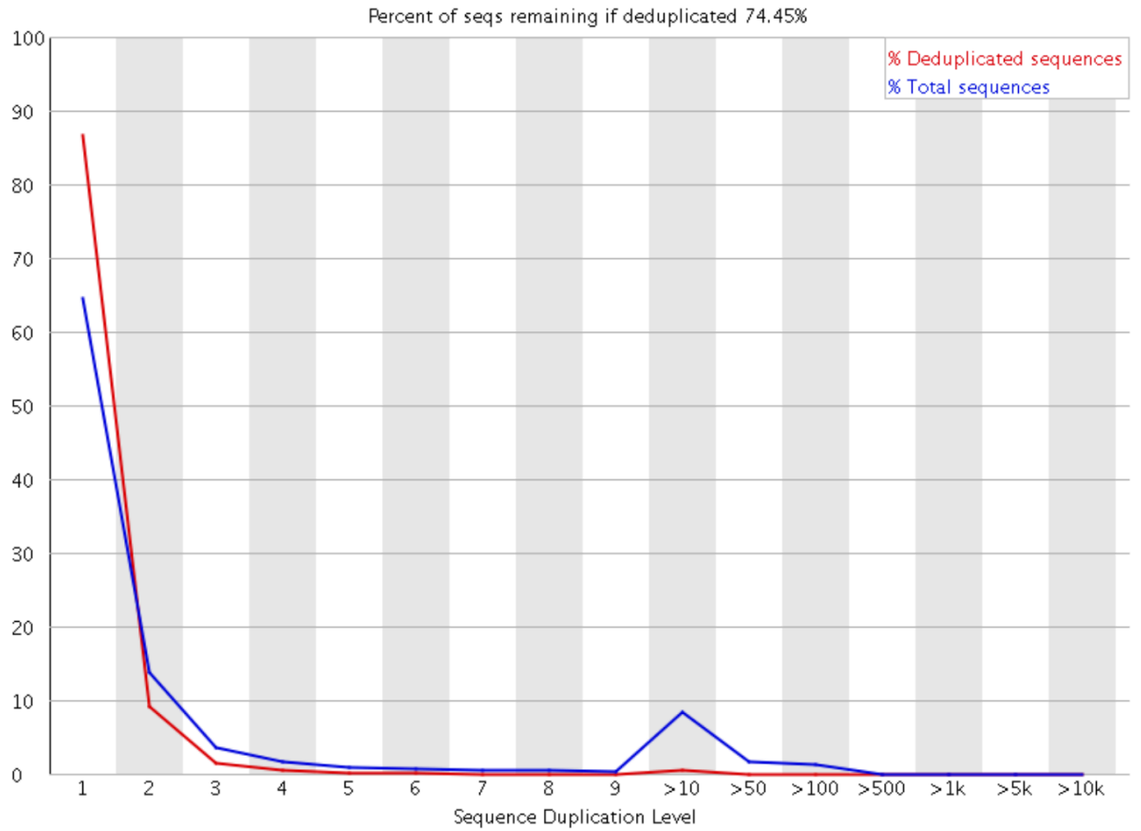
Sequence Identity, Flow-Cell Postion, and Mapping Position:

FastQC Report

Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Kmer Content

Sequence Duplication Levels

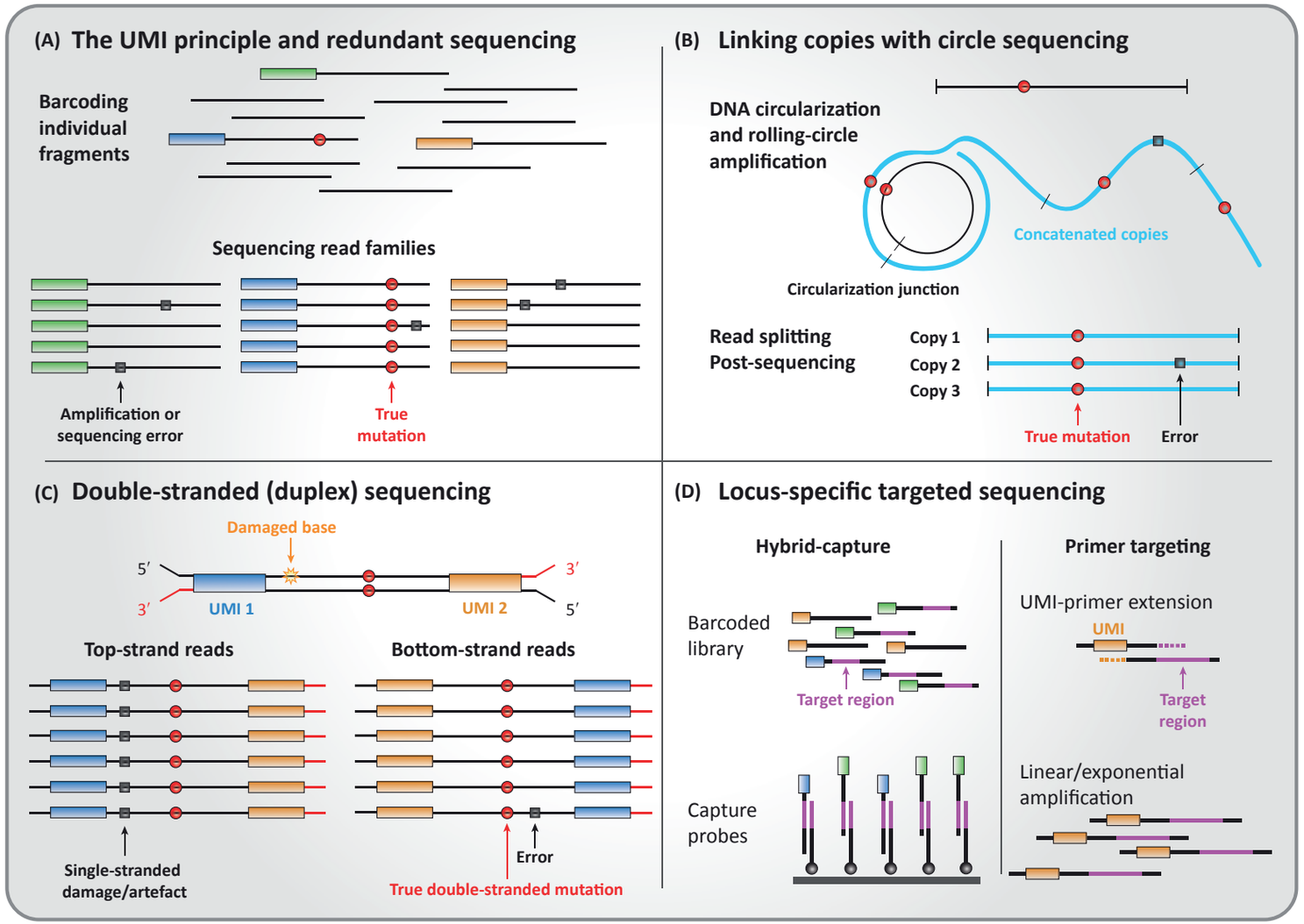


Deduplicating Reads

Additional Resolution with UMIs:

- 16S sequencing and other amplicon sequencing methods
- RNA-seq (regions of extremely high coverage)
- Variant calling in pooled/heterogeneous samples
- Low-input methods

High-Fidelity Sequencing



Sloan et al. 2018 Trends in Biotechnology

Applications of High-Fidelity Sequencing

Application	Use of sequencing technique allows
Design cancer treatments	• Identification of subclonal variants • Better understanding of causes and progression of disease
Non-invasive medical screening	• Accurate sequencing of small amounts of circulating cell-free DNA • Assessment of cancer therapy success (circulating tumor DNA) • Monitoring of organ transplant rejection (graft-derived cell free DNA) • Identification of genetic abnormalities (circulating fetal DNA in mother)
Inform immunization strategies	• Understanding risk of mutagenesis in existing viruses • Strategies to reduce risk of vaccine mutagenesis/virulence potential
Identify disease risk	• Determine DNA repair capacity in individuals as an indicator of disease risk
Measure toxin effects	• Assessment of mutagenicity in environmental samples • Quantification of mutagenic potential of known substances
Study mechanisms and rates of mutation	• Measurements of mutation rate and spectrum • Determination of error rates in polymerases and DNA repair enzymes • More accurate characterization of diversity at molecular level
Create informed models of evolution	• Precise determination of patterns of viral mutation and sequence evolution

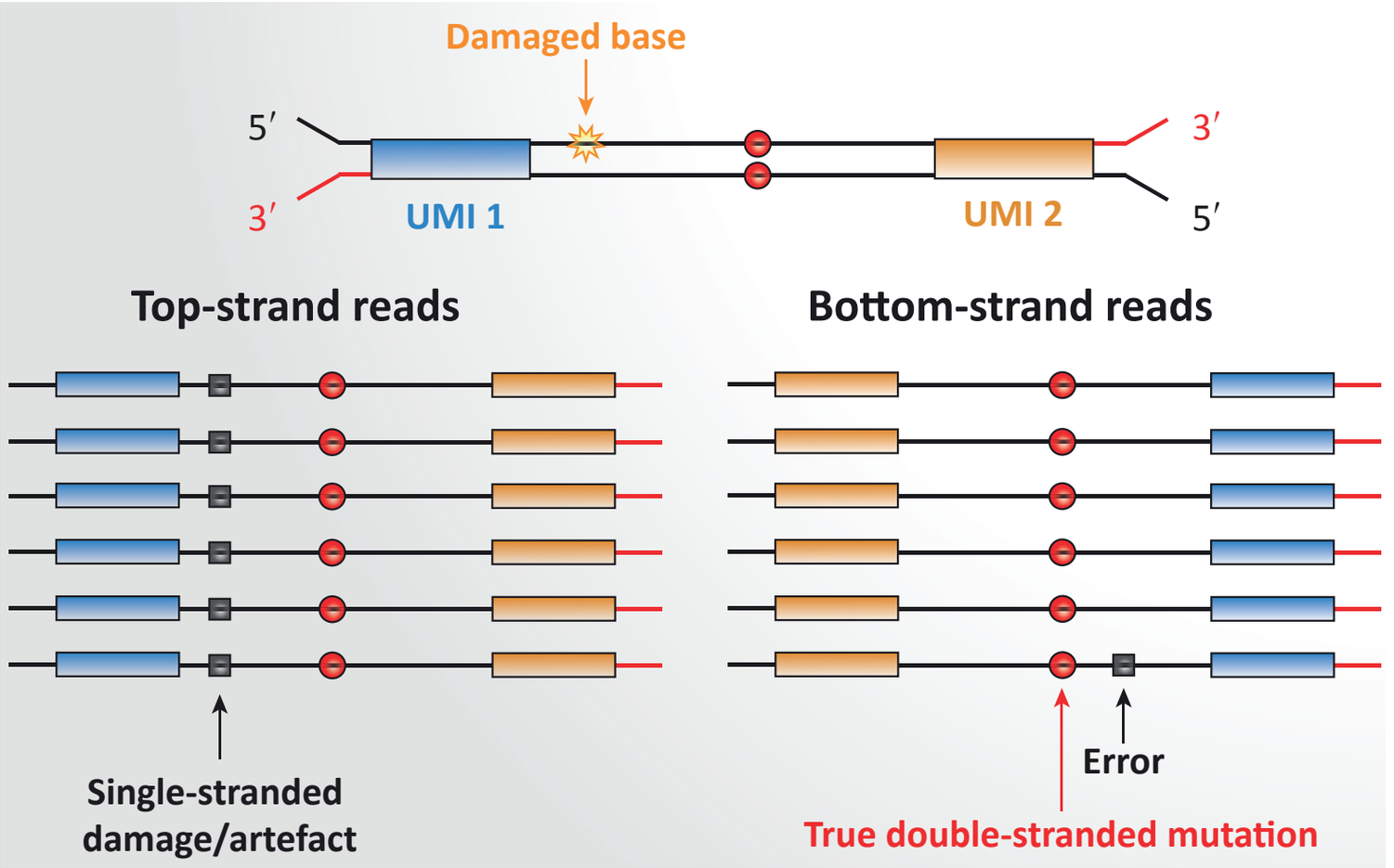
"We expect applications to further expand into other disciplines…
…the field of forensics…
…ancient DNA sequencing"

High-Fidelity Sequencing Methods

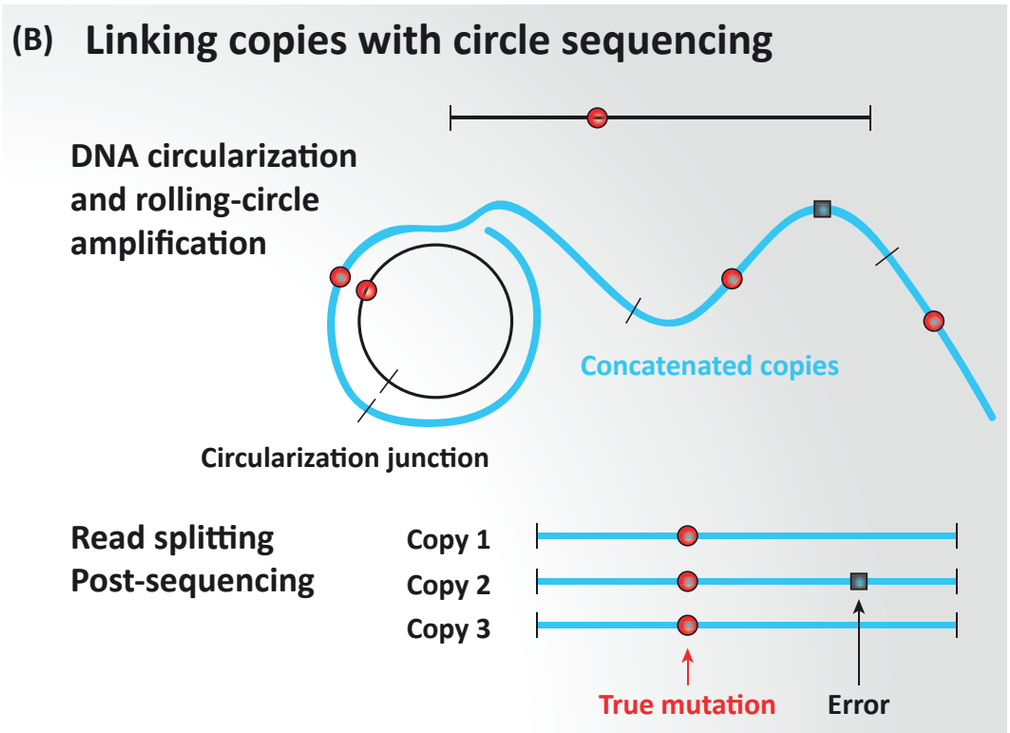
A few options…

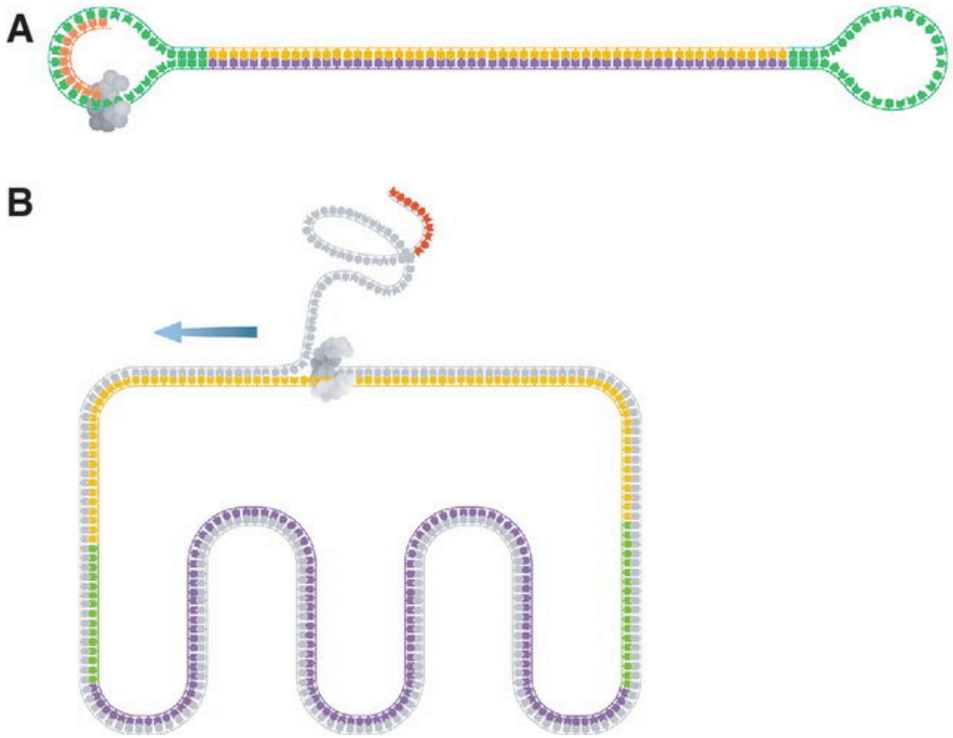
Method	Linked reads	Double-stranded	Targeted[a]
SMRTbell	Yes	Yes	Optional
Safe-SeqS	No	No	Optional: hybrid-capture
Duplex sequencing	No	Yes	Optional: hybrid-capture
Narayan 2012	No	No	Yes: PCR targeting
Circle sequencing	Yes	No	No
NOIR	No	No	Yes: PCR targeting
Barcoding-coupled SMRT	No	No	Yes: PCR targeting
BotSeqS	No	Yes	No
CAPP-Seq and iDES	No	Yes	Optional: hybrid-capture
CypherSeq	No	No	Optional: RCA enrichment
Droplet-CirSeq	Yes	No	No
Lee 2016	No	No	Yes: PCR targeting
MDS	No	No	Yes: PCR Targeting
Alcaide 2017	No	Yes	Optional: hybrid capture
Asymmetrical barcode adapters	No	No	Optional: hybrid-capture
BiSeqS	No	Yes	Yes: PCR targeting
DEEPER-seq	No	No	Optional: hybrid-capture
o2n-seq	Yes	No	Optional: hybrid-capture
Pro-Seq	Yes	Yes	Optional: PCR targeting
SiMSen-Seq	No	No	Yes: PCR targeting
TEC-Seq	No	No	Optional: hybrid-capture
Yamashita 2017	No	No	Yes: PCR targeting
muSeq	No	Yes	Yes: restriction digest

Duplex Sequencing



Physically Connecting Copies as an Alternative to UMIs





Travers et al. 2010 NAR