DTU

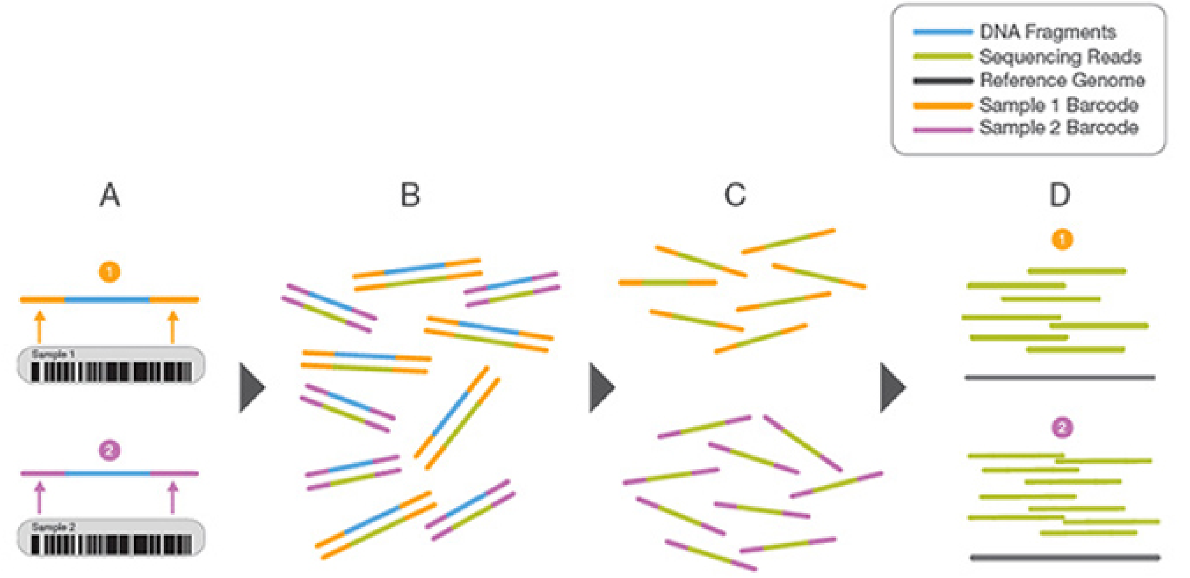
DTU Health Technology
Bioinformatics

Barcoding

Gisle Vestergaard
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gisves@dtu.dk

Multiplex sequencing

- Multiplex sequencing allows sequencing libraries to be pooled and sequenced simultaneously during a single run on most high-throughput instrument



A. Two representative DNA fragments from two unique samples, each attached to a specific barcode sequence that identifies the sample from which it originated.

B. Libraries for each sample are pooled and sequenced in parallel. Each new read contains both the fragment sequence and its sample-identifying barcode.

C. Barcode sequences are used to de-multiplex, or differentiate reads from each sample.

D. Each set of reads is aligned to the reference sequence.

- NovaSeq system outputs 20 billion paired-end reads yielding up to 3000 Gbases
- Running several samples on one sequencing run
  - Saves money & time
  - Eliminates sequencing run as variable

De-multiplexing

- De-multiplexing is generally done automatically before you receive the samples
- If sequencing locally always check how much on the run ended up unassigned
  - Bcl2fastq2 stringency can be adjusted if barcodes are very different

- How would you realise that demultiplexing was not done?

Index hopping and bleeding

- Failure to remove free adapters leads to index hopping
- Very similar barcodes and sequencing errors can lead to index bleeding

- Dual indexing solves both problems at a cost