




Iso-Seq Clustering

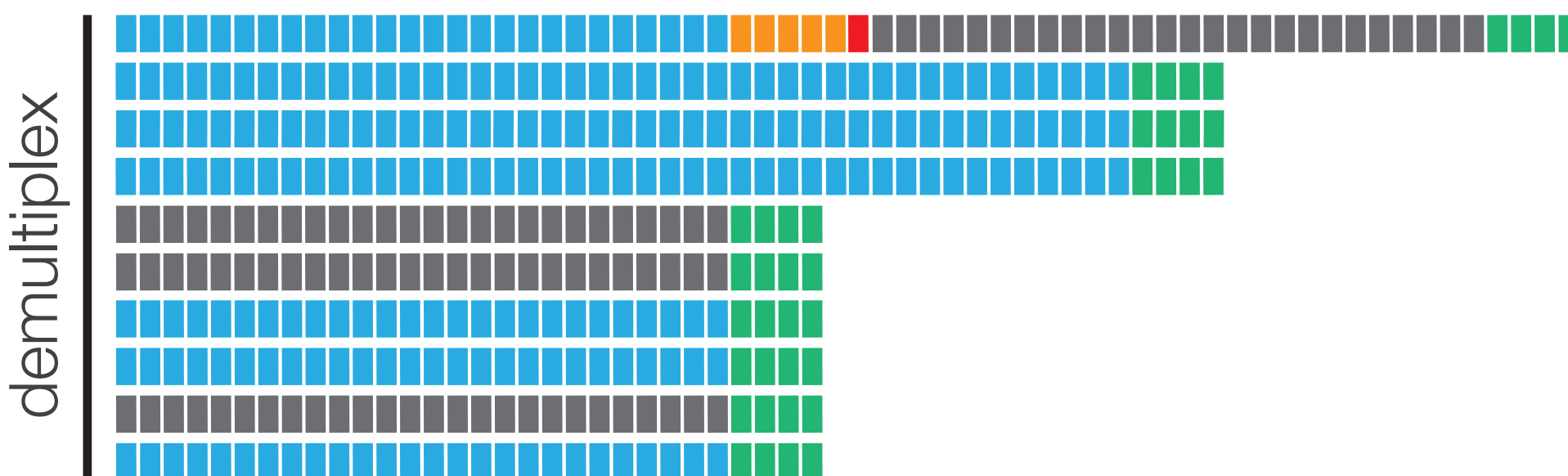
transcript gene A 
transcript gene B 
optional polyA 

3' cDNA primer 
5' cDNA primer 
sample barcode 

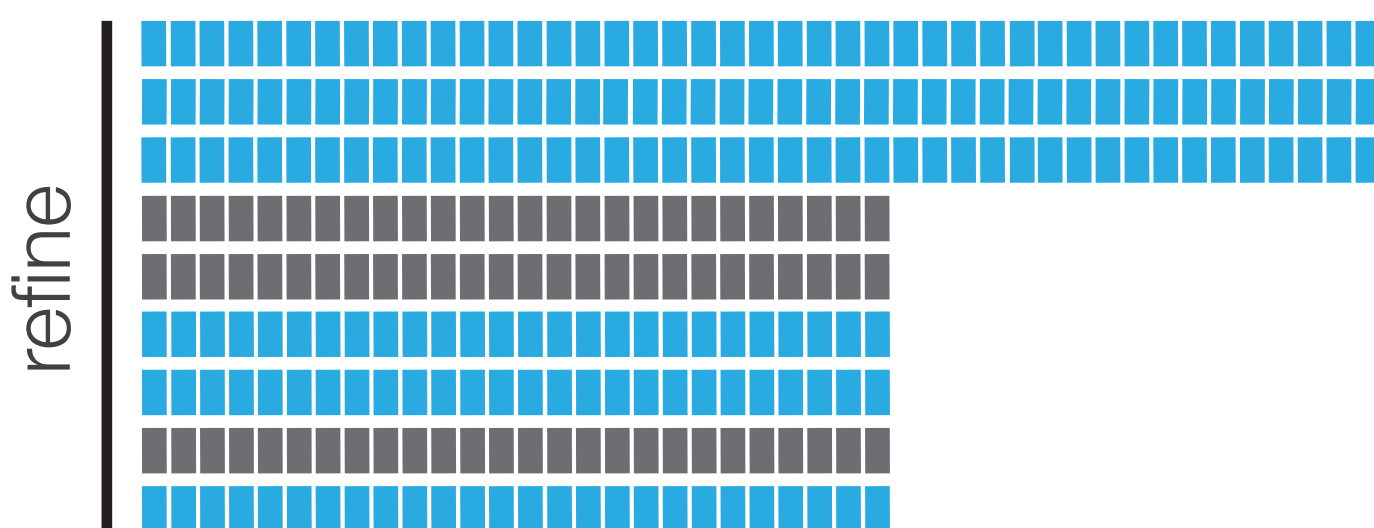
Version 9, Dr. Armin Töpfer



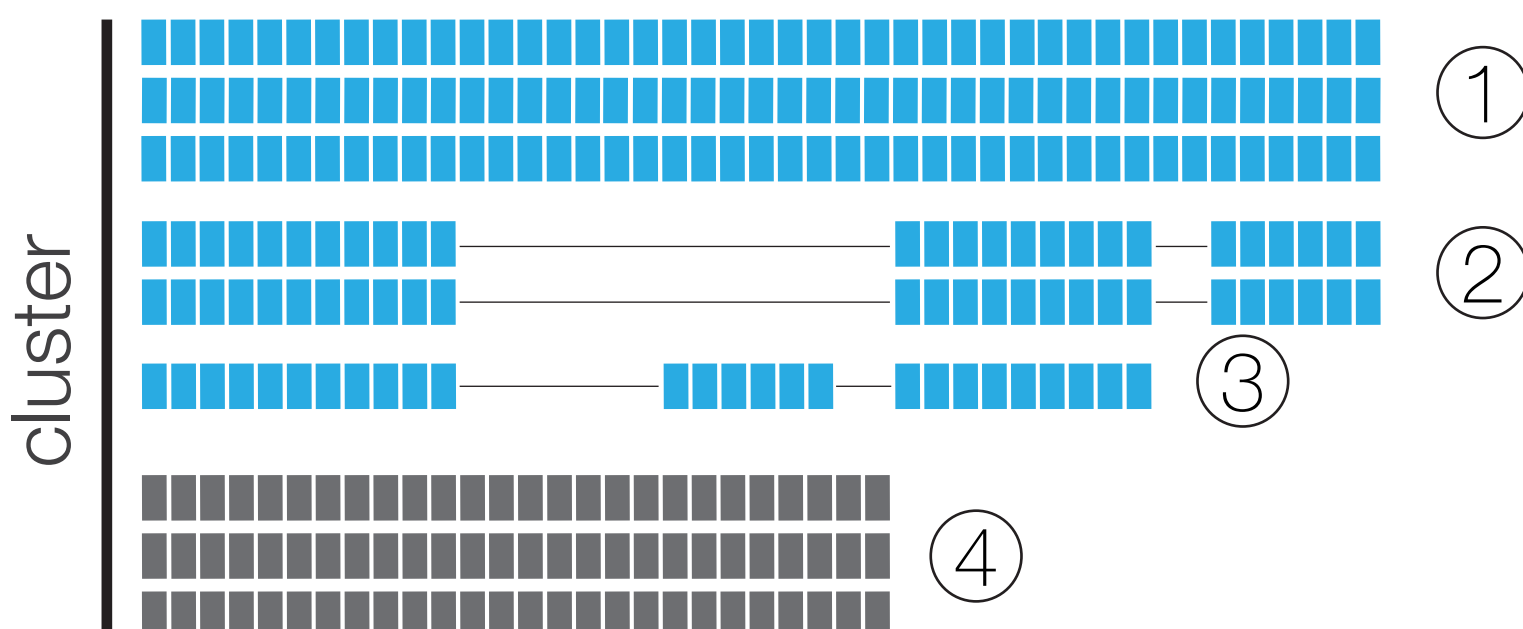
- Use **polished** CCS reads
- Only full-pass ZMWs



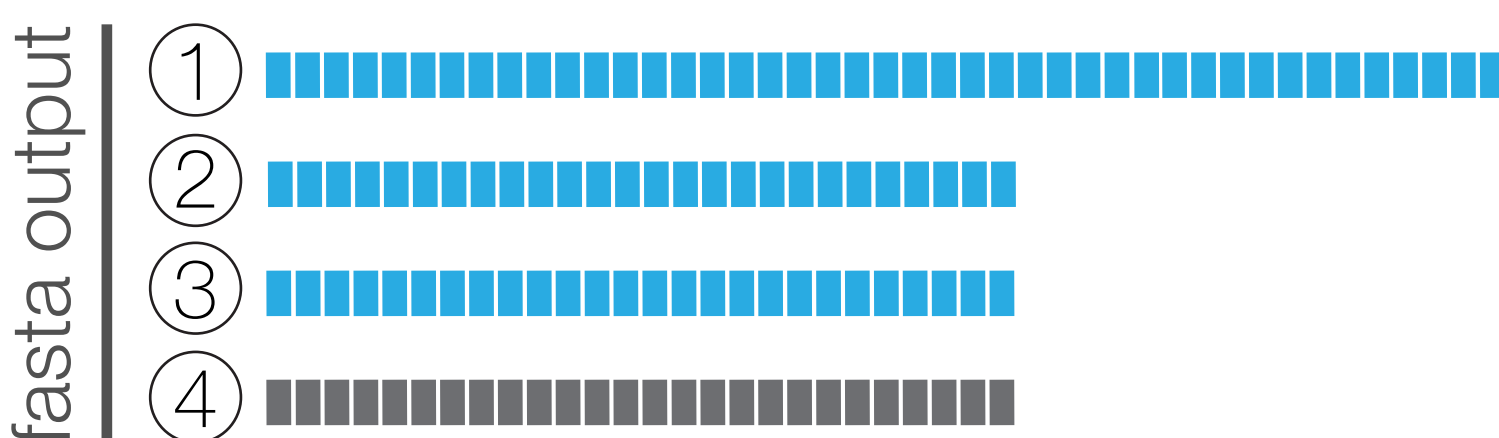
- Barcoded and unbarcoded cDNA primer removal
- Orientation
- Unwanted primer combination removal



- PolyA tail trimming
- Concatemer removal



- Hierarchical, $n \cdot \log(n)$ clustering, alignment of shorter to longer sequences
- Iterative cluster merging
- Generate consensus for each read cluster using QV guided PoA



- Fasta output is split into HQ and LQ reads
- One consensus per read cluster