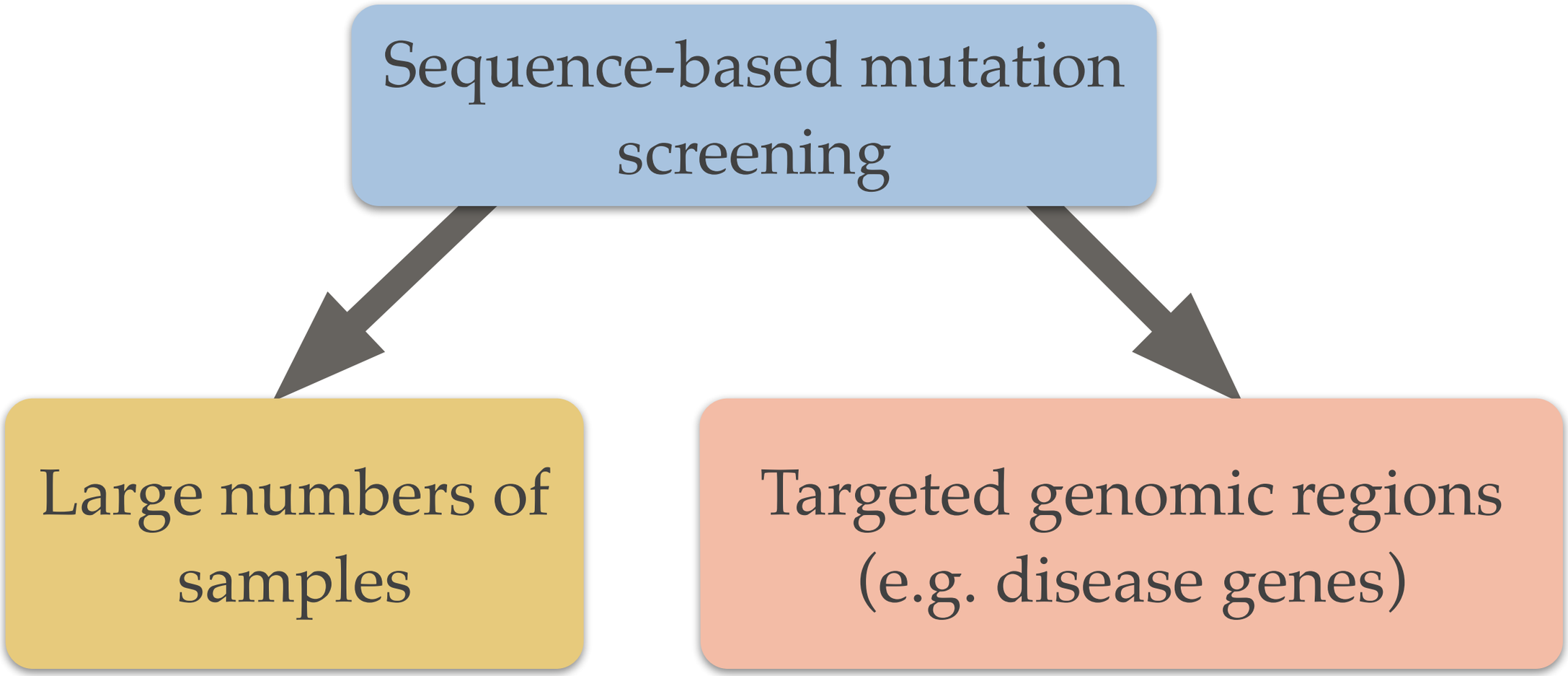
Monday 28 September 2015

Hi-Plex

Simple, low-cost, modular targeted DNA sequencing

Bernie Pope, bjpope@unimelb.edu.au
Lead Bioinformatician, Cancer Genomics

Motivation


Sequence-based mutation screening
Large numbers of samples
Targeted genomic regions (e.g. disease genes)

Example application

Cancer Research Malaysia performs 90% of breast cancer screening in Malaysia

Collaboration to use Hi-Plex for screening BRCA1, BRCA2, PALB2 and TP53 mutations

Competing technologies

Haloplex

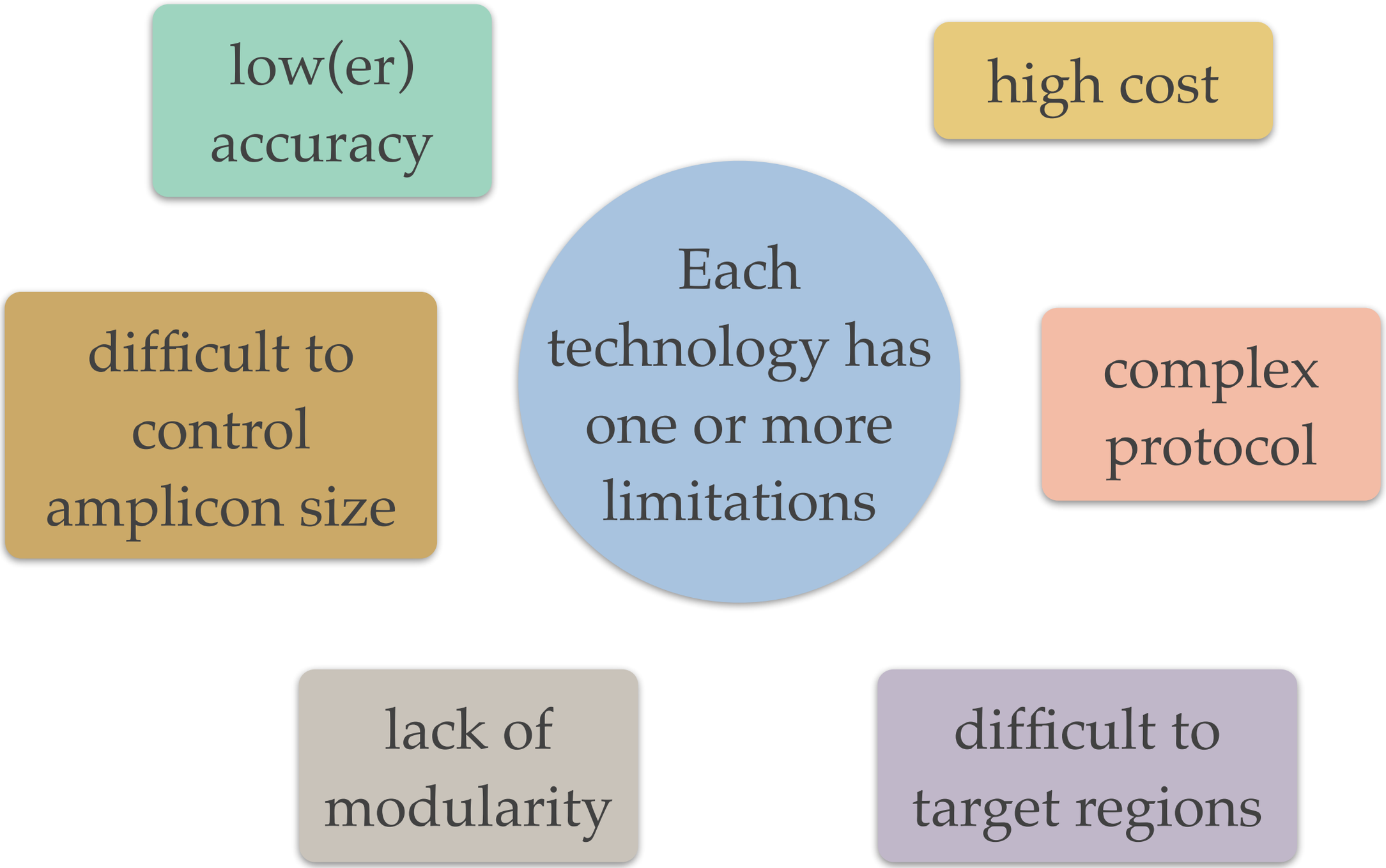
TruSeq Amplicon

Fluidigm Access Array

Ion Ampliseq

Have been proven to work with high amplicon plexity, but …

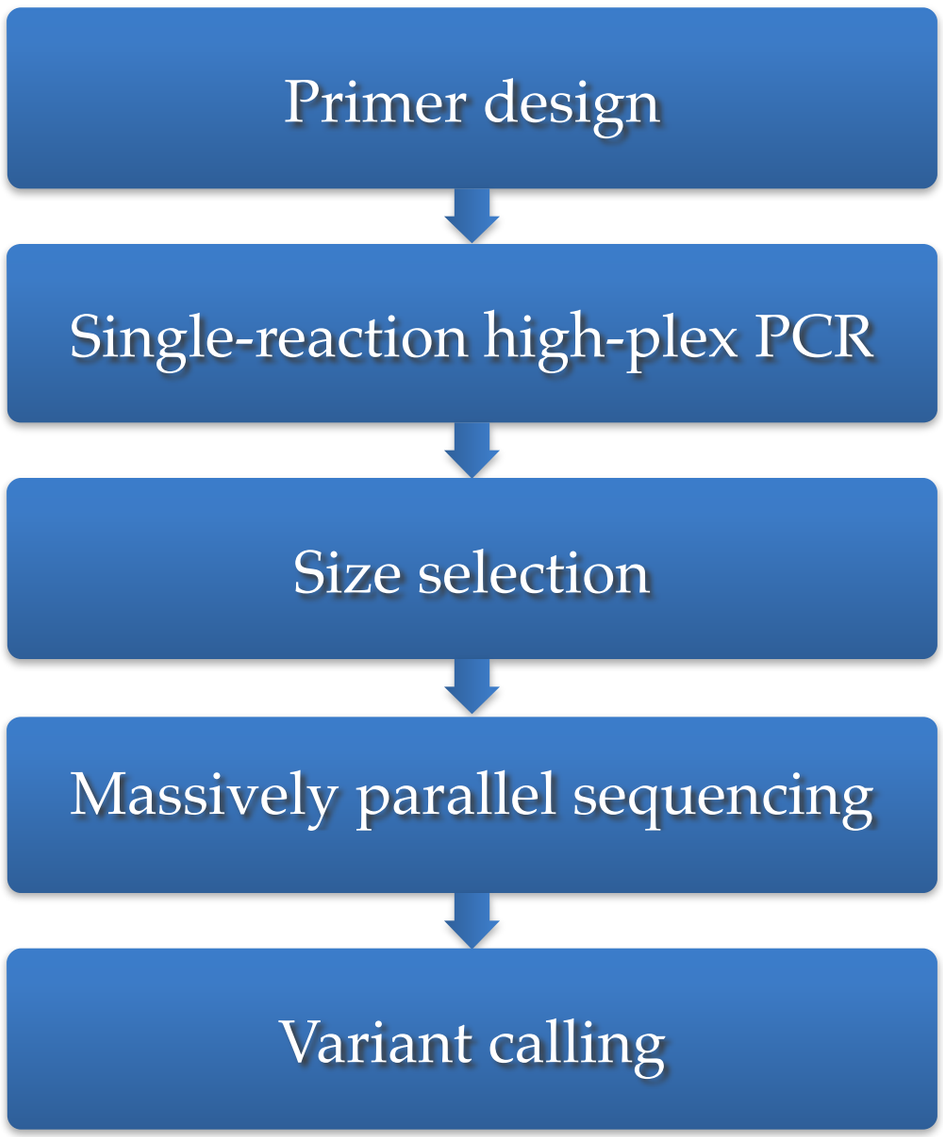
Competing technologies



Hi-Plex objectives

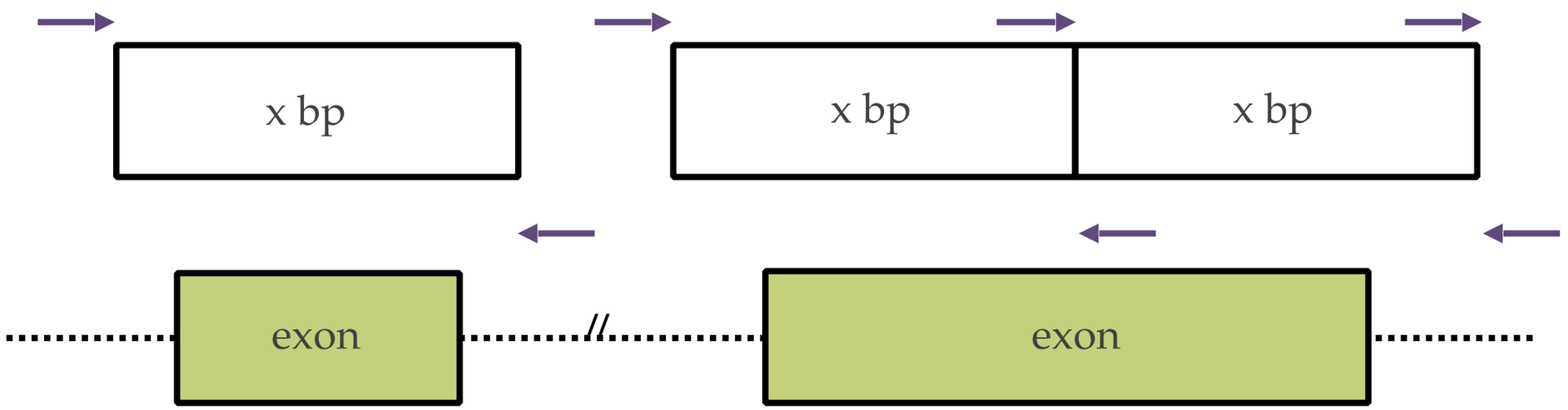
- Low cost
- Accurate
- Scalable
- Enable high-throughput bench work
- Portable to different sequencing technologies
- Modular
- Applicable to low quality DNA source material

Workflow


Primer design
Single-reaction high-plex PCR
Size selection
Massively parallel sequencing
Variant calling

Hi-Plex principles

Region tiling with controlled tile size



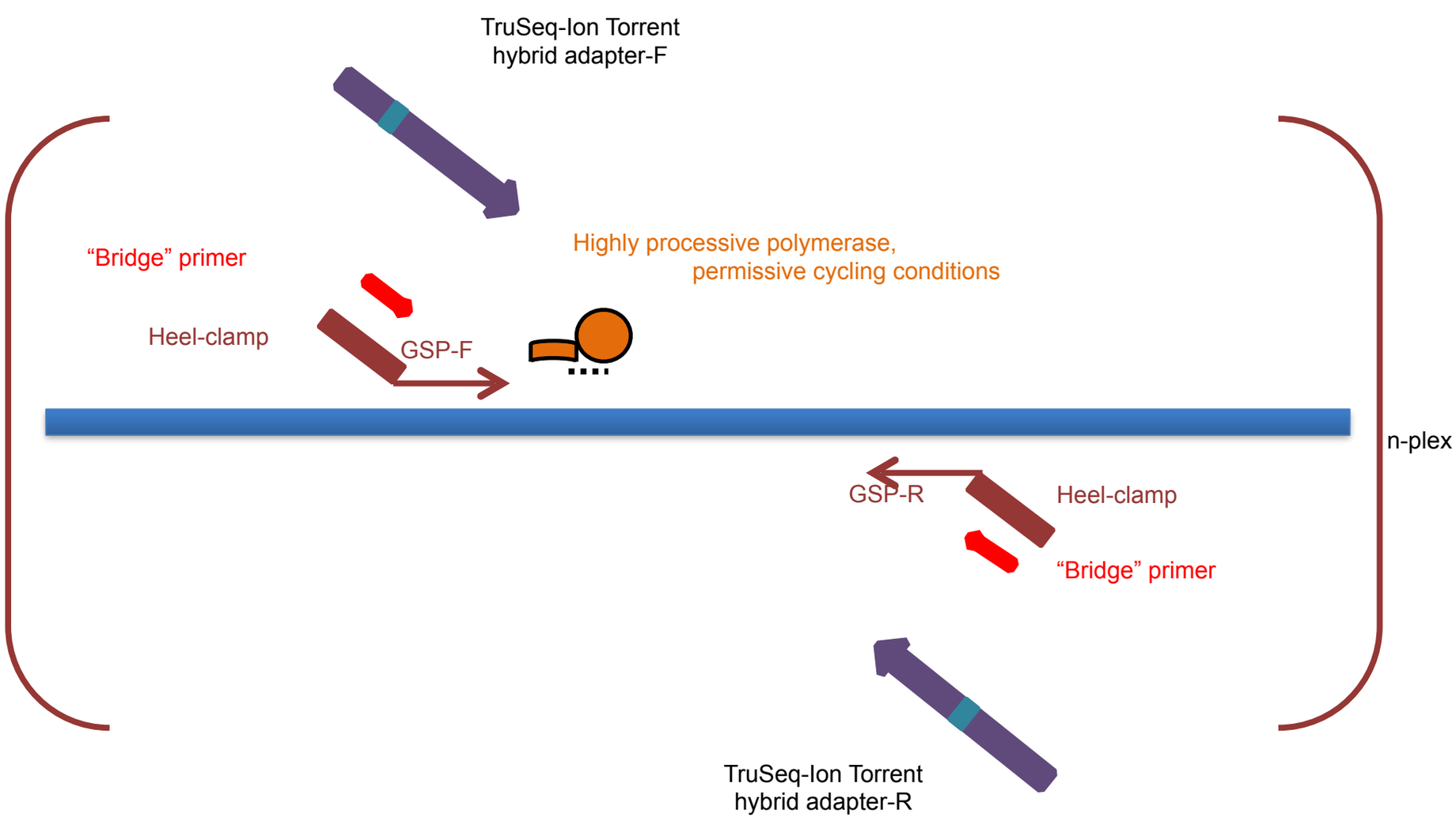
Hi-Plex principles

Region tiling with controlled tile size

- Precise definition of amplicon size
- PCR products size-selected on a single GEL lane
    - removes off-target products
- Allows completely overlapping reads
    - can stringently filter chemistry artefacts because each read pair measures the same sequence location twice
- Allows more permissive annealing temperature

Hi-Plex principles

Primers and polymerase



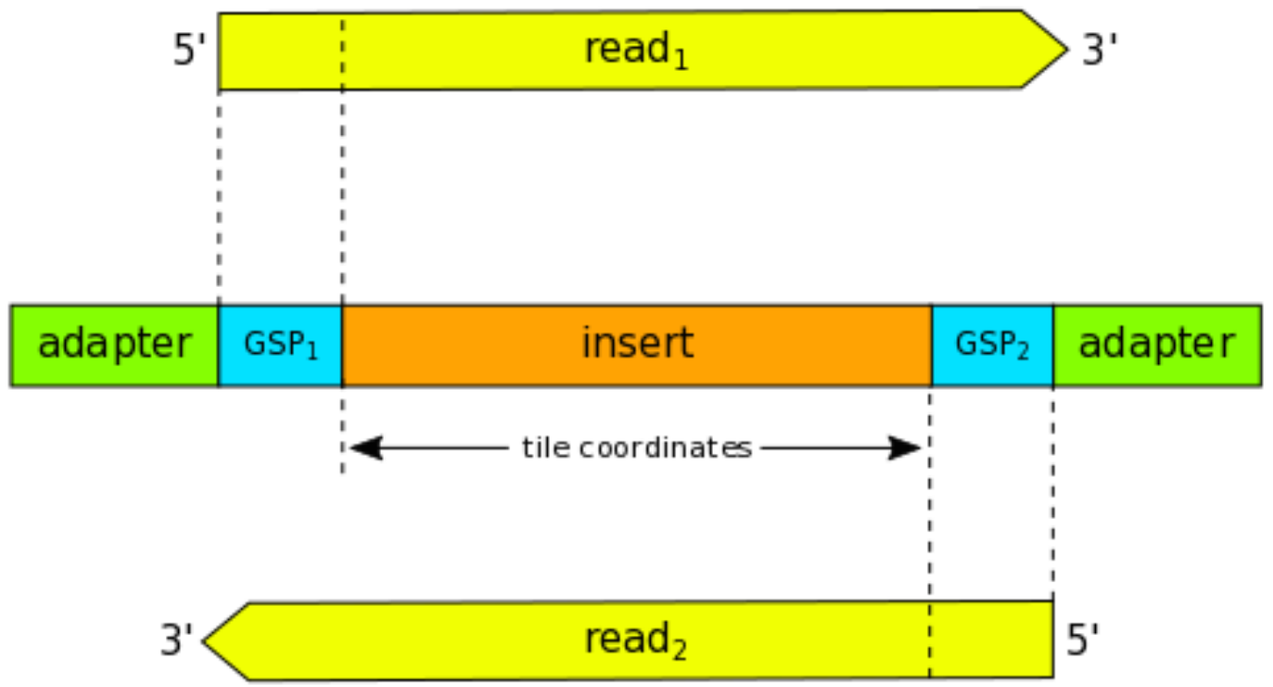
Hi-Plex principles

Primers and polymerase

- Highest processivity and fidelity DNA polymerase
- Gene-specific primers (GSPs) seed the PCR
- Universal primers drive the reaction
- GSPs are in low concentration compared to universal primers
- Lowers the chance that off-target amplification will overwhelm the reaction

Hi-Plex principles

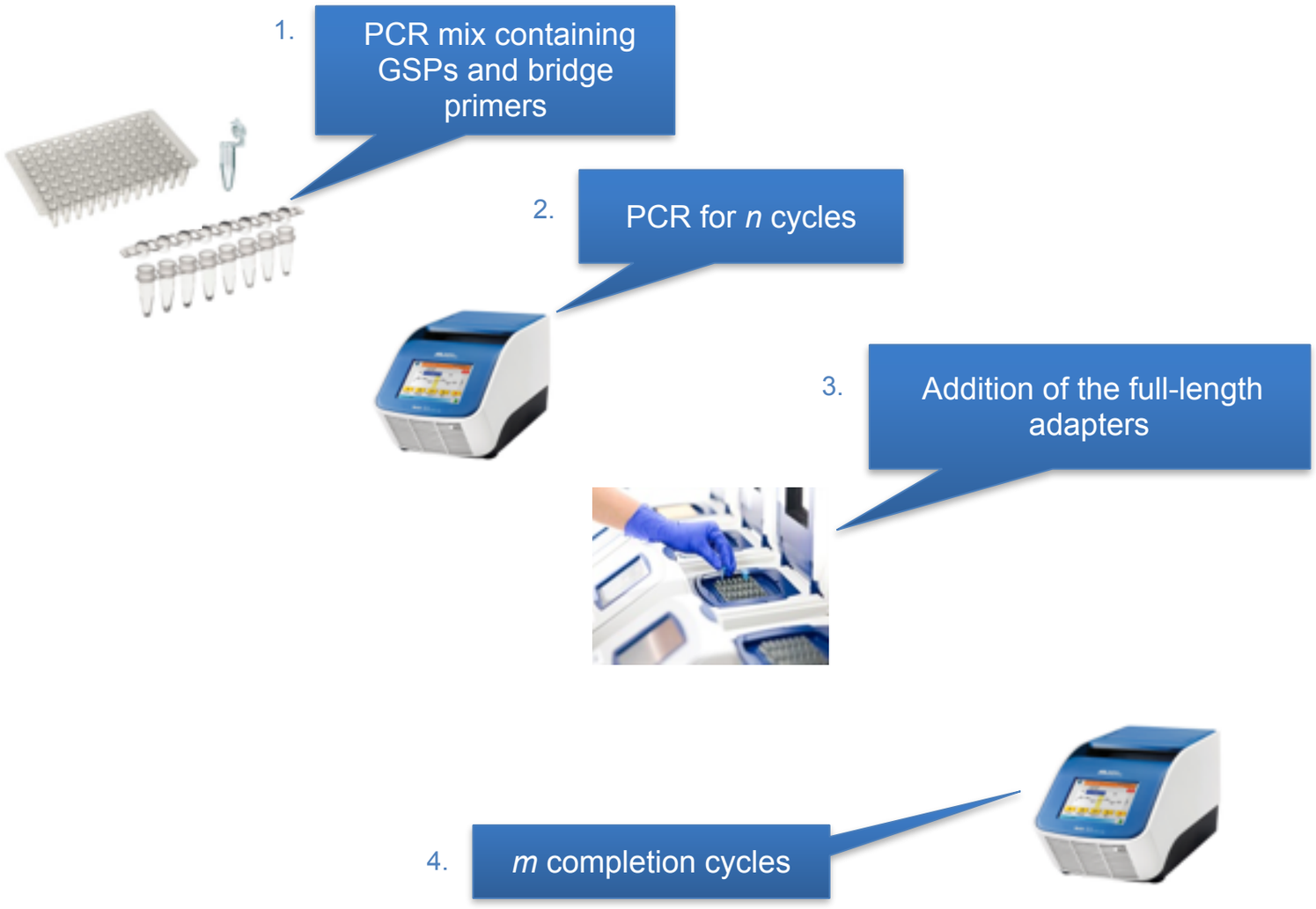
Completely overlapping reads



ROVER variant caller requires a variant to appear in both reads for a threshold number of pairs.

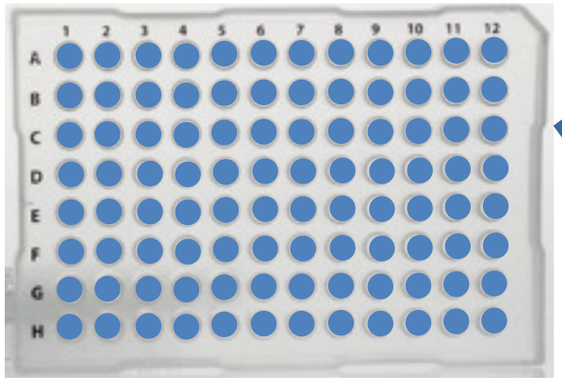
Hi-Plex principles

Single tube, single type of DNA polymerase, per sample



Hi-Plex principles

Single tube, single type of DNA polymerase, per sample

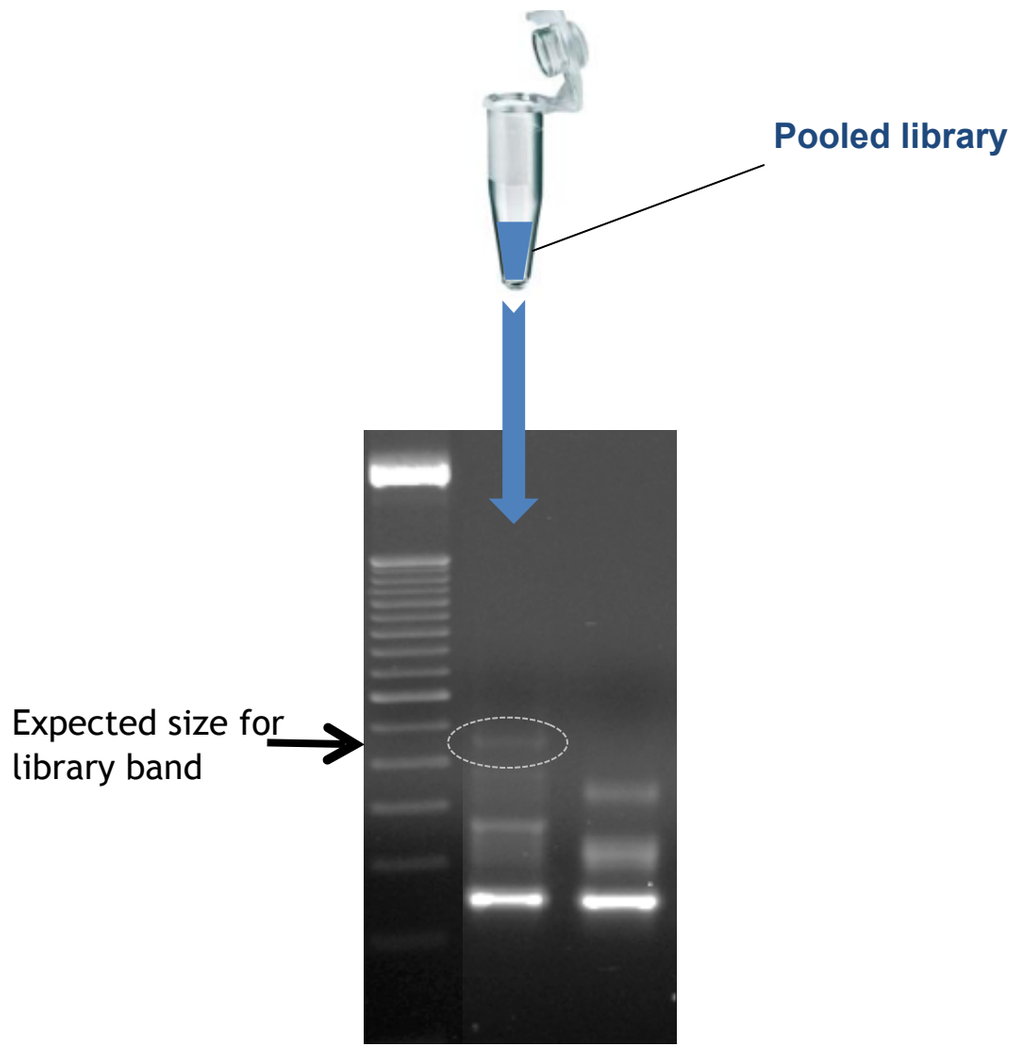


dual indexing

Library pooling

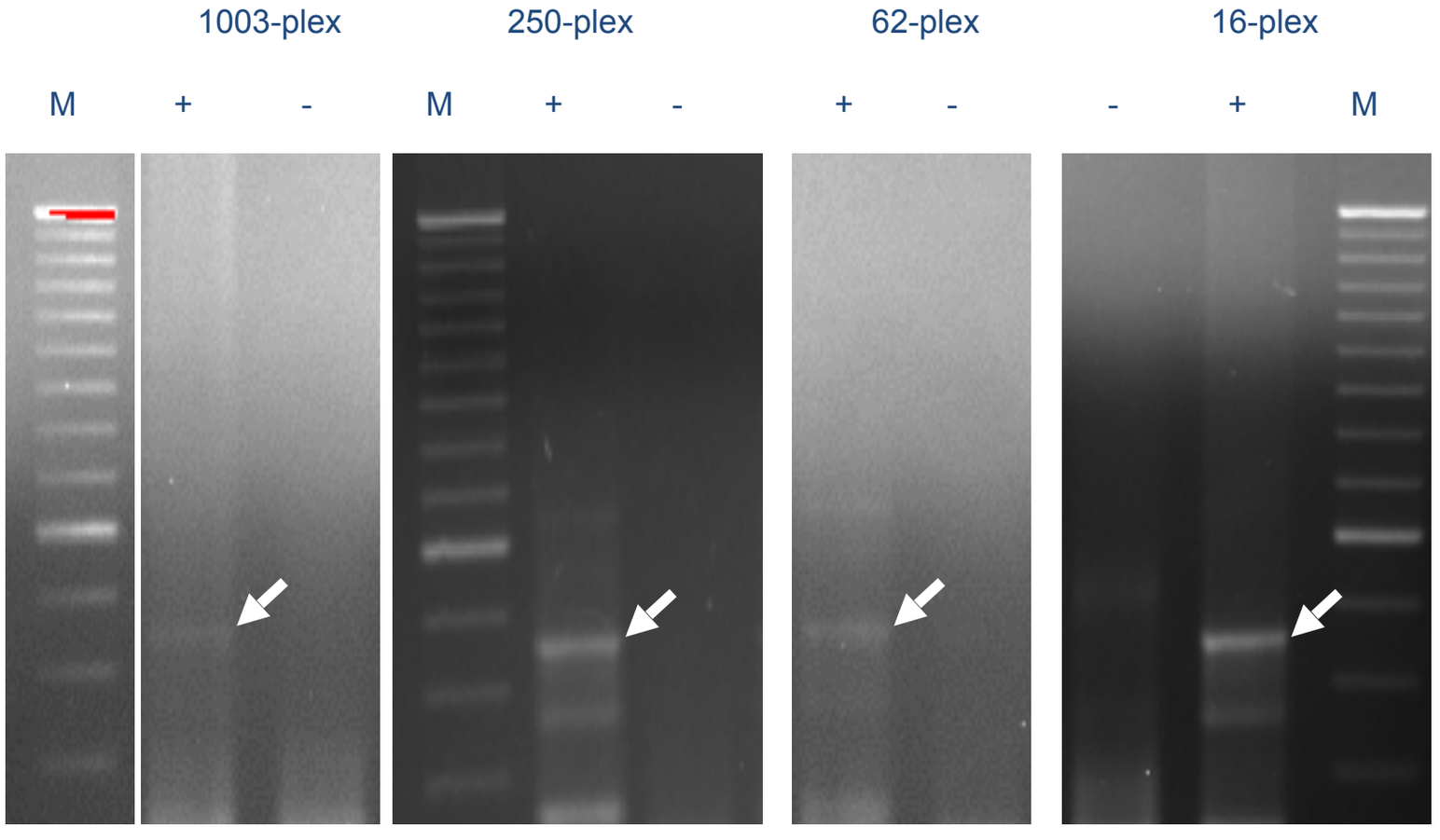
Hi-Plex principles

Single tube, single type of DNA polymerase, per sample



Hi-Plex results

Gel bands across a range of amplicon *plexity*



Hi-Plex results

	1003 plex	250 plex	62 plex	16 plex
% amplicons within 25 fold from median depth	94.12	95	96.77	100

Hi-Plex results

- *XRCC2* and *PALB2* breast cancer screen
- 95 blood-derived DNAs from women affected by breast cancer
- previously characterised by various screening and genotyping methods
- Hi-Plex + ROVER accurately called 56 previously reported variants
- 4 new additional variant calls were made, which were later validated by Sanger sequencing
- No false positive calls

Hi-Plex results

- Hi-Plex has been demonstrated to work with a variety of DNA source material:
  - whole blood and lymphoblastoid cell line
  - FFPE tumours
  - Guthrie cards

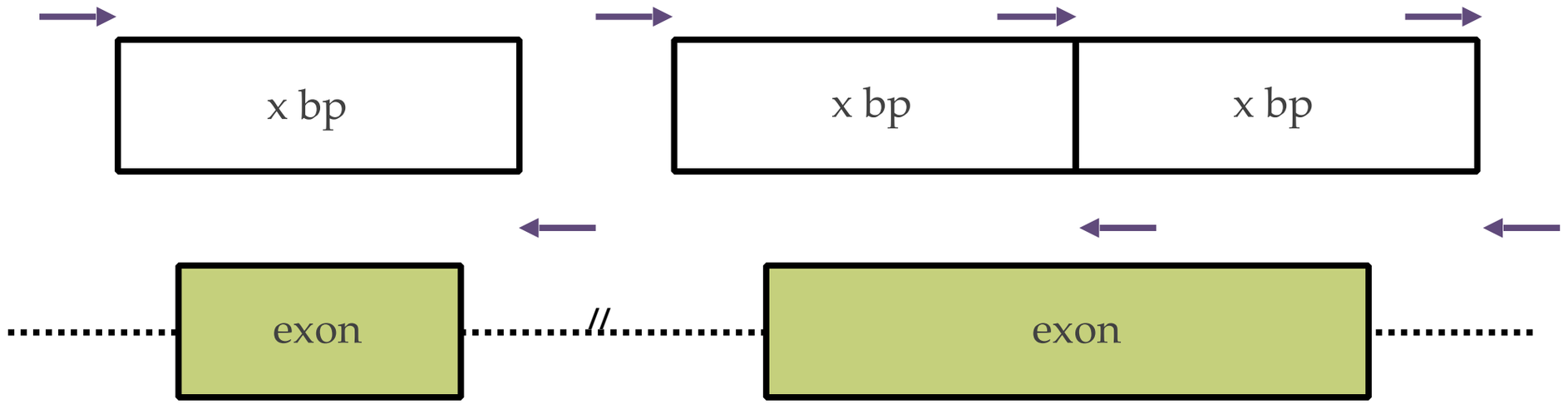
Commonly seen in pathology labs. DNA can be degraded.

Hi-Plex results

- Further adoption of the technology:
    - 4 gene breast cancer panel with Cancer Research Malaysia (formerly CARIF).
    - a number of large scale cancer screens (1000s of samples and 10s of genes) with various international collaborators

Multiplex primer design software

Target regions are covered by tiles



Each tile denotes an amplicon with a pair of primers

Multiplex primer design software

- Requirements:
    - bad primers should be avoided
    - entire target region must be covered by tiles
    - primers may vary in size (to a small degree)
    - tiles may vary in size (to a small degree)
    - tiles may overlap (to a small degree)
    - tiles may overhang at the ends of regions
    - sequencing resource should be used efficiently

Multiplex primer design software

- Up until now we have been working with an overly simplistic primer design tool:
    - Inaccurate melting temperature calculation for long primers
    - Sliding window search only considered a small portion of the primer/tile search space
    - Did not adequately account for primer-dimers and other off-target effects

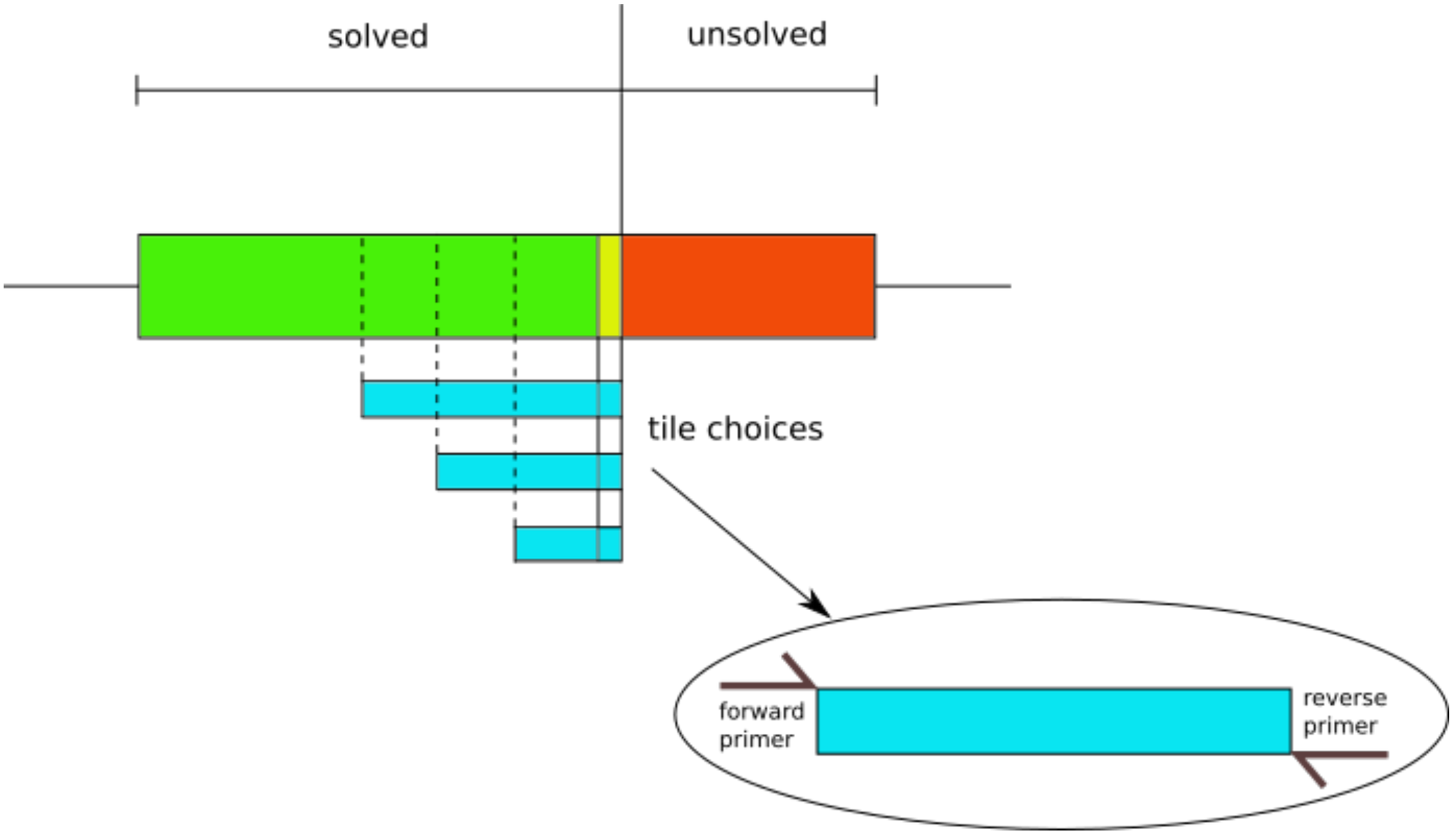
Multiplex primer design software

- Algorithmic challenges:
  - Number of possible tilings of a region grows exponentially
  - It is not feasible to consider all possible tilings in a naive brute-force way
  - Pairwise primer-dimer considerations are costly, so cannot be done frequently

Multiplex primer design software

- Multiplex primer design can be viewed as an optimisation problem
- Find the set of primers/tiles that maximises a fitness function
- A (simple) fitness function can be based on the sum of the primer scores for each chosen tile
- Individual primers can be scored on qualities such as melting temperature, entropy, length, off-target priming

Multiplex primer design software



New primer design tool

- Uses dynamic programming to search the tile space efficiently (in linear time)
- Primer-dimers are not considered in the scoring of fitness
- We apply a final post-processing stage to identify bad cases of primer-dimers (which occur infrequently)
- Off-target priming can be approximated by hashing k-mers of the reference sequence

New primer design tool

- In future work we plan to consider randomised approximation techniques such as simulated annealing

Future Potential Applications

- Currently Hi-Plex has been used to screen for SNPs and small INDELs
- We may adapt the system to screen for other effects such as structural variants and methylation

Acknowledgements

- Core Hi-Plex team:
    - Daniel Park
        - Head, Melbourne Bioinformatics Platform
        - Head, Genomic Technologies Group, University of Melbourne
    - Tú Nguyen-Dumont
        - Research Fellow, Genetic Epidemiology Laboratory, University of Melbourne

Acknowledgements

- Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne
    - Melissa Southey
    - Joyce Teo
    - Fleur Hammet
    - Maryam Mahmoodi
    - Helen Tsimiklis
- Victorian Life Sciences Computation Initiative (VLSCI)
- NHMRC project grant APP1025879

Acknowledgements

- Students:
    - Luke Shillabeer (Primer Design)
    - Sori Kang (Primer Design, Annokey)
    - Roger Li (ROVER, UNDR-ROVER)
    - Edmund Lau (Primer Design)

More information

- www.hiplex.org
- www.hiplex.org/papers.html
- www.hiplex.org/software.html