454-BASED RESEQUENCING USING THE VARIANT IDENTIFICATION AND INTERPRETATION PIPELINE (VI2P)

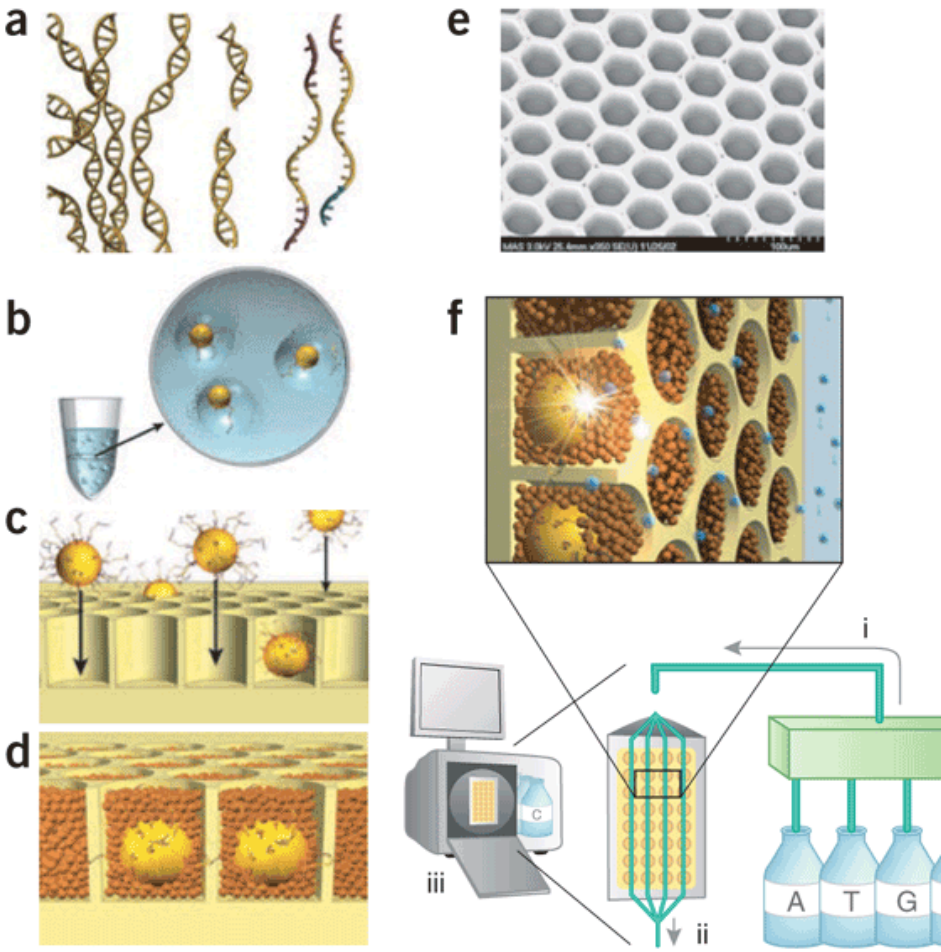
MoBi-KO: Jan 22 2010

Introduction to 454 sequencing

- Massive parallel (next-generation) pyrosequencing
- Specifications/run (2010)
  - 1,000,000 reads
  - 400bp avg. length
  - $0.025/kb
- Applications
  - De novo genome sequencing
  - Amplicon resequencing
  - Bisulphite sequencing

454 pyrosequencing technique

- A) Generate library sequences
  - Enrichment via chip/CHiP
  - PCR
- B) Emulsion PCR on beads
- C) Load plate: 1 bead per well
- D) Duplicate sequence on bead using ACTG cycles: 1 'read'/well =>1 image/cycle



Next-generation diagnostics

- Current state-of-the-art in diagnostics: Sanger sequencing
- Next-generation sequencing (NGS) in diagnostics:
  - Lower cost per screening
  - Faster results
  - Screen complex diseases involving multiple genes
- NGS diagnostics drawbacks :
  - Relative high error rate
  - Multiple reads needed (coverage >40)
  - Loads of data…
    - 454 GS-FLX: > 30GB/run, 2GB FASTA
    - Illumina GAII: > 3TB/run, 100GB FASTA
- Goal: Identify 'real' causal variants

Complex data analysis

Variant Identification/Interpretation Pipeline (1)

Sequencing

Image transfer

Athos server

Basecalling

FASTA files

Preprocess

Data Base

Map

BLAT

Data Base

Analysis

BioBiX Browser

<24h

- EnsEMBL
- dbSNP
- Exonic?
- Synon?
- Polyphen
- Intronic?
- Fruitfly.org
- …

Variant interpretation

Data Base

Variant Identification/Interpretation Pipeline (2)

- Modular & database pipeline
- Amplicon resequencing and targeted resequencing (Nimblegen)
- Current status:
  - 15 runs
  - 7,901,986 reads
  - 2,281,638,292 bp
  - Completely automated <24h
  - Validating putative variants
    - BRCA1/2
    - Genetic deafness
    - Heart aneurisma
    - …