Detection and quantification of rare mutations with massively parallel sequencing

Isaac Kinde, Jian Wu, Nick Papadopoulos, Kenneth W. Kinzler and Bert Vogelstein

Johns Hopkins Kimmel Cancer Center
June 7, 2011

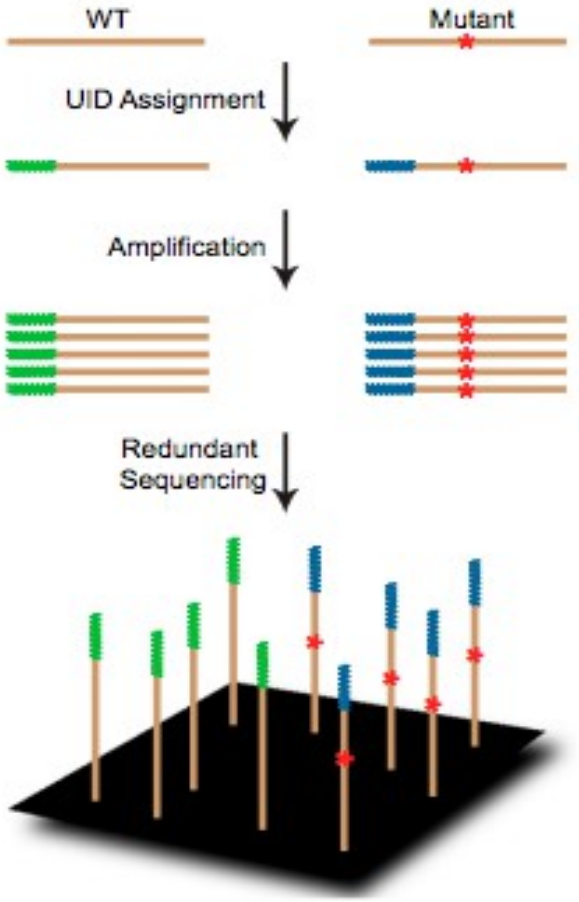
Tim Farrell
BF831: Translational Bioinfo Seminar
October 14, 2015

Motivation

- Low-prevalence mutations are difficult to identify with confidence
  - In clinical setting, prevalence can be <0.01%
- Illumina, and sequencing technologies more generally, still have error rates high enough to mask these rare mutations
  - estimates ranged from 1% to 0.05% in 2011
  - ~0.01% in 2015 (1)
- Many aspects of sequencing factor into this error:
  - Library prep
  - Choice of primers
  - Base-calling
  - Read length

Safe-Sequencing System ("Safe-SeqS") Methodology

1. Assign unique id (UID; aka barcodes or indices) to each DNA template
2. Amplify of each uniquely-tagged template:
   - Creates many daughter molecules = "UID family"
3. Sequence
4. If 95% of any UID family has identical mutation → "supermutant"



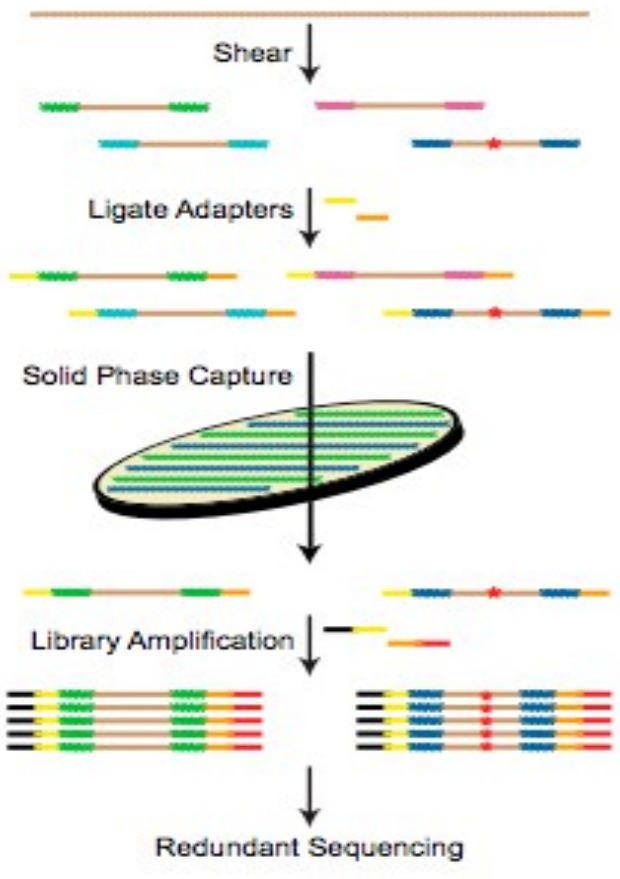
UID Assignment: Endogenous

Randomly sheared DNA inherently contains unique seq at either end → call UID

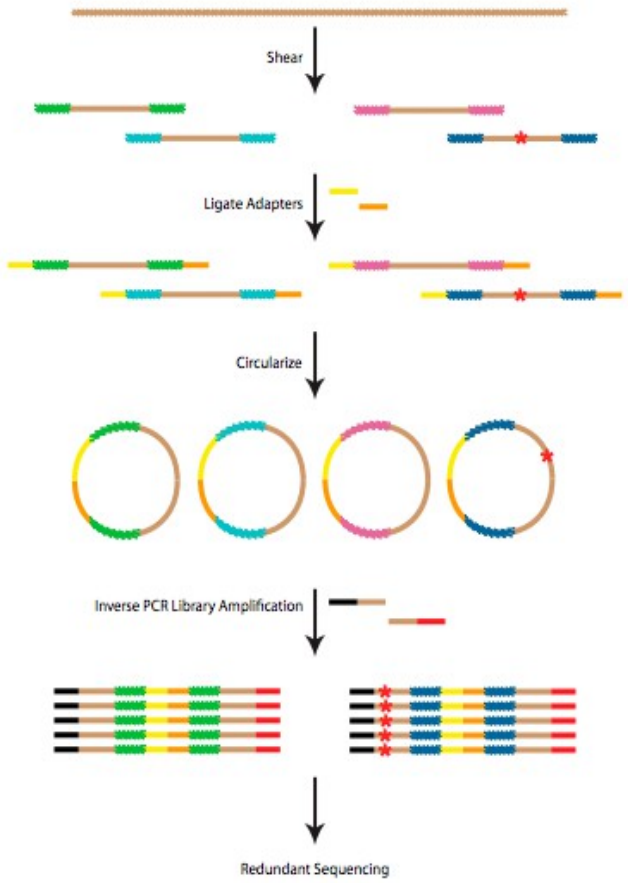
2 approaches:

- Capture: many genes simultaneously
- Inverse PCR: single gene in depth

Capture



Inverse PCR



Endogenous UID Experiments

Capture:

- 15000 normal cells
- 2.6kb from 6 genes
- SafeSeqS: error lowered 70 fold

Inverse PCR:

- 1750 normal cells
- 1 gene of interest
- 1057 unique molecules assessed
- SafeSeqS: conventional results likely all FP

	Capture	Inverse PCR
Conventional analysis		
High-quality base pairs	106,958,863	1,041,346,645
Mean high-quality base pairs read depth	38,620×	2,085,600×
Mutations identified	25,563	234,352
Mutations/bp	2.4*E*-04	2.3*E*-04
Safe-SeqS analysis		
High-quality base pairs	106,958,863	1,041,346,645
Mean high-quality base pairs read depth	38,620×	2,085,600×
UID families	69,505	1,057
Average no. of members/UID family	40	21,688
Median no. of members/UID family	19	4
Supermutants identified	8	0
Supermutants/bp	3.5*E*-06	0.0

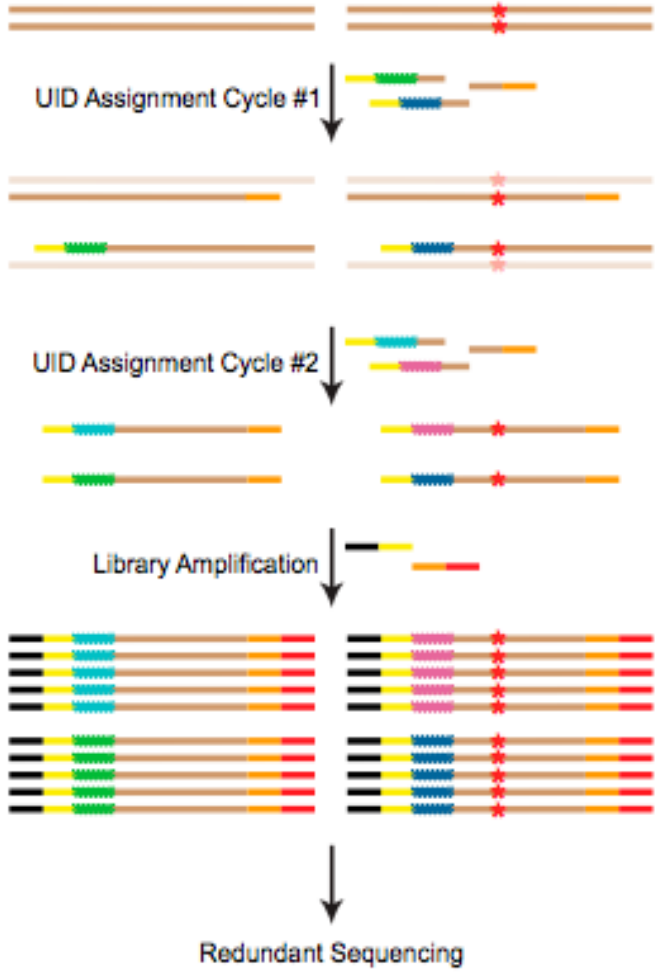
UID Assignment: Exogenous

2 cycles of UID assignment during PCR

12-14 random nucleotides b/t forward primer and seq serves as UID

Amplification: 25 PCR cycles

Gives a uniquely-tagged, dsDNA fragment for each of the original template strands



Mean SD

Exogenous UID Experiments

CTNNB1:

- 3 individuals
- SafeSeqS: reduced apparent error 24 fold

Mitochondrial DNA:

- 7 individuals
- SafeSeqS: 15 fold reduction

Discussion

How do we know conventional analyses represent artifacts rather than true mutations?

- consistency of results
- polymerase fidelity experiment results consistent with previous biological assays

Limitations:

- Exogenous UID strategy cannot be applied to multiple amplicons from sample containing a limited # of templates
- Still does not overcome error introduced in amplification

Future directions

References

(1) Schrimer et al. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nuc Acid Res*. doi: 10.1093/nar/gku1341.