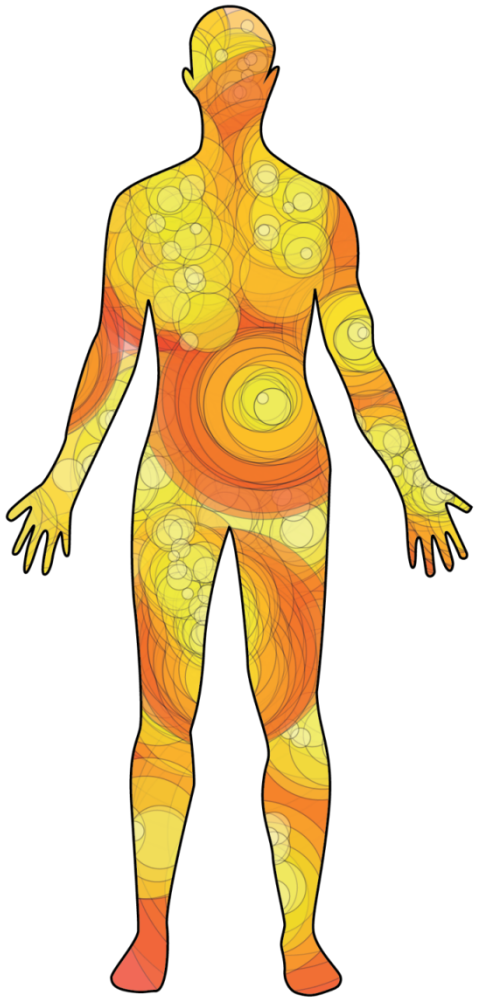
GLOBAL HAPLOTYPE RECONSTRUCTION OF HIV-1

Francesca Di Giallonardo

Armin Töpfer

SANGER-SEQUENCING VS. NEXT-GENERATION SEQUENCING (NGS)

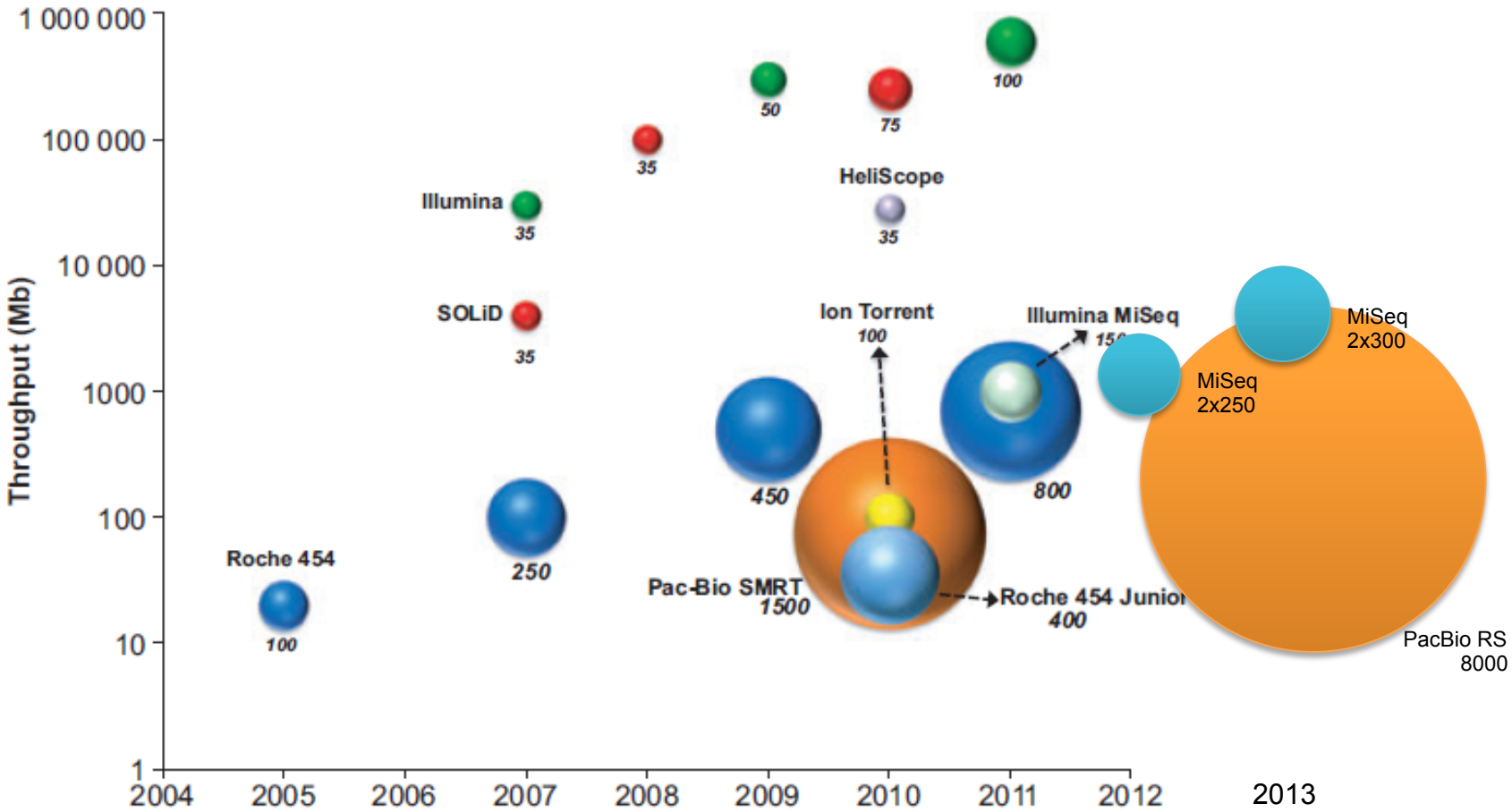


INFECTED WITH A POPULATION OF STRAINS

SANGER - CONSENSUS

NGS – INDIVIDUAL SEQUENCES

OVERVIEW OF NGS PLATFORMS



Shokralla S *et al*, Mol Ecol. 2012 Apr;21(8):1794-805

LOCAL AND GLOBAL HAPLOTYPE RECONSTRUCTION

```
      ↓     ↓     ↓     ↓     ↓       ↓     ↓       ↓
A:  CCTGAAATCACTCTATGGCAACGACCCATCGTCACAATAAAGATAGGG  60%
B:  CCTCAAATCACTCTTTGGCAACGACGCATCGTCACAATATAGATAGGA  30%
C:  CCTCAAATCTCTCTTTGGCACCGACCCATCGTCCCAATAAAGATAGGG  10%
```

NGS

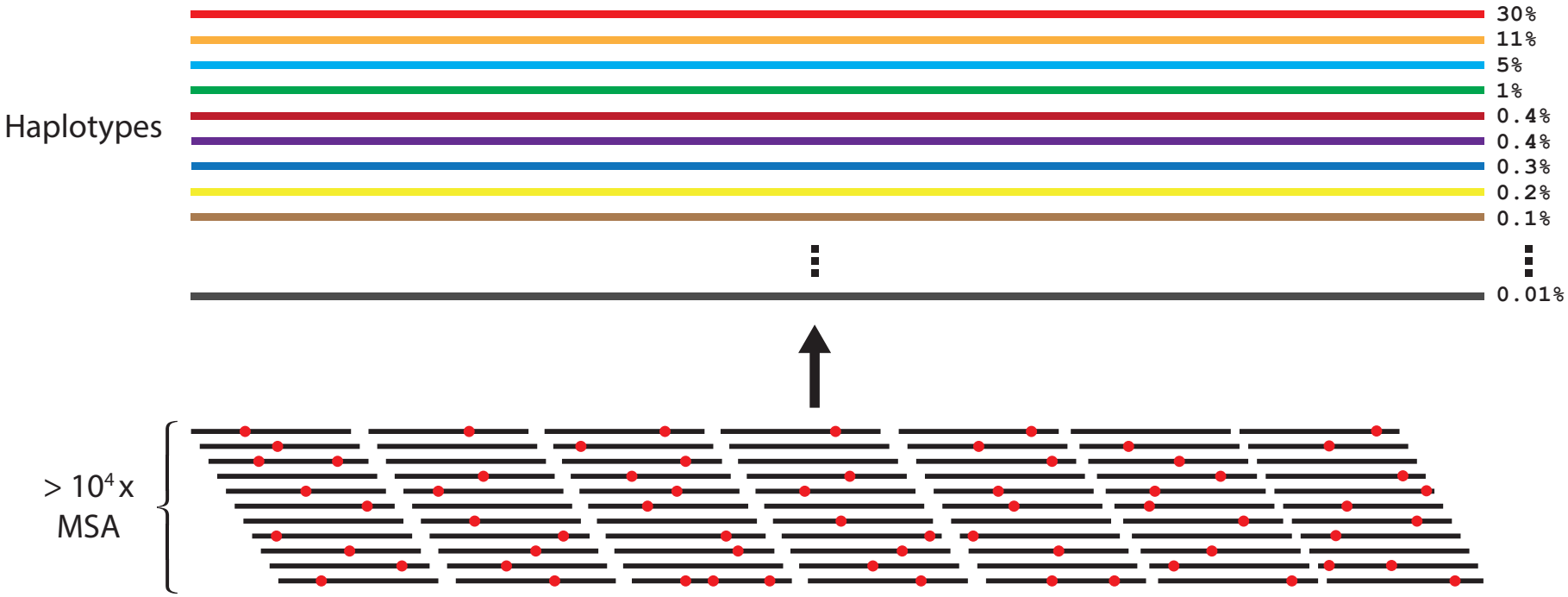
```
 1  CCTGAAATCACTCTATGGCA
 2     GAAAACACTCTATGGCAACG
 3        ATCACTCTTTGGCAAGGCCG
 4         TCACTCTATGGCAACGACCC
 5          CTCTTTTGGGCACCGACCCA
 6              CTATGGTAACGACCCATCGT
 7               TATGGCAACGAGCCATCGTC
 8                ATGGCACGGACCCATCCCCC
 9                 TGGCAACGACGCATCGTCAC
10                    CAACGACGCATCGTCACAAT
11                    CAACGACGCATCGTCACGAT
12                      AACGACCCTTCGTCACAATA
13                        CGACCCATCGTCTCAATAAA
14                            GCATCGTCACAATATAGAGA
15                             CATCGTCACAAAATAGATAG
16                               TCGTCACAATAAAGATAGGG
17                                  TCACAATAAAGATGGGG
18                                    CCAATAAAGATAGGG
19                                       AATAAGGATGGGG
20                                          ATAGATAGGA
```

errors

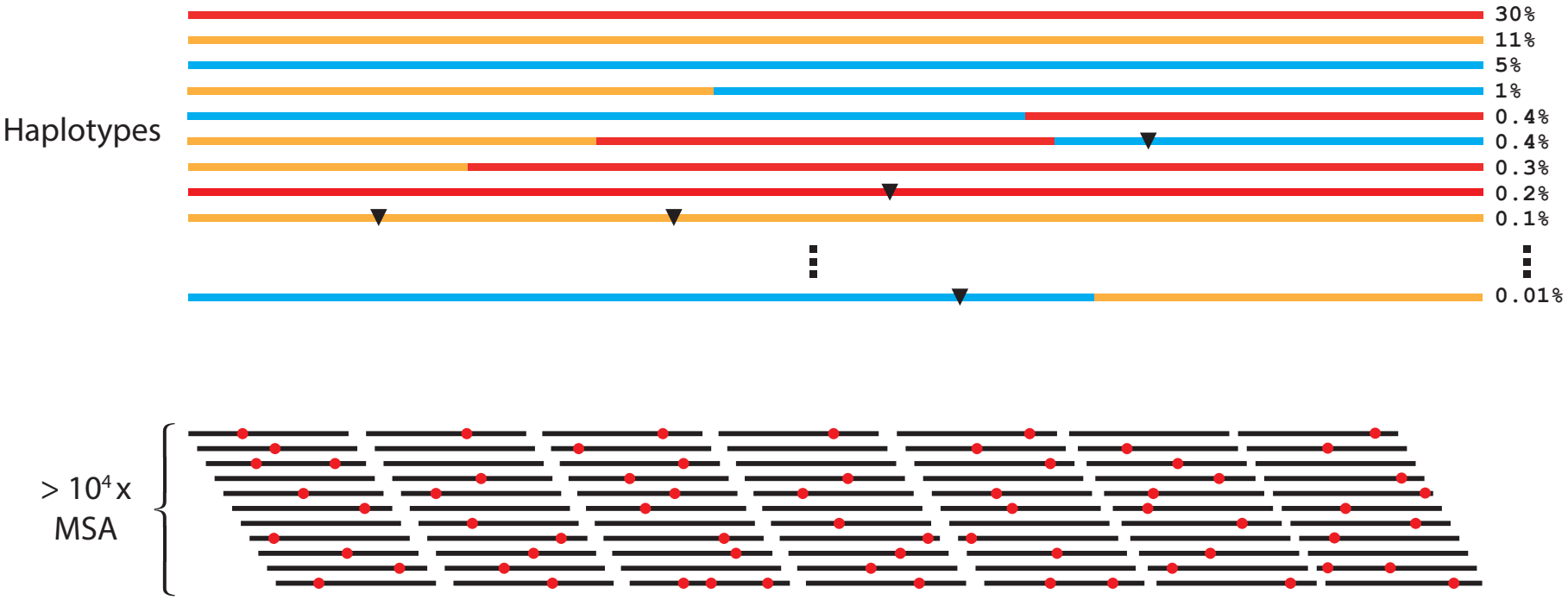
···· global
--- local
— SNV

Beerenwinkel et al. Front Microbiol. 2012;3:329

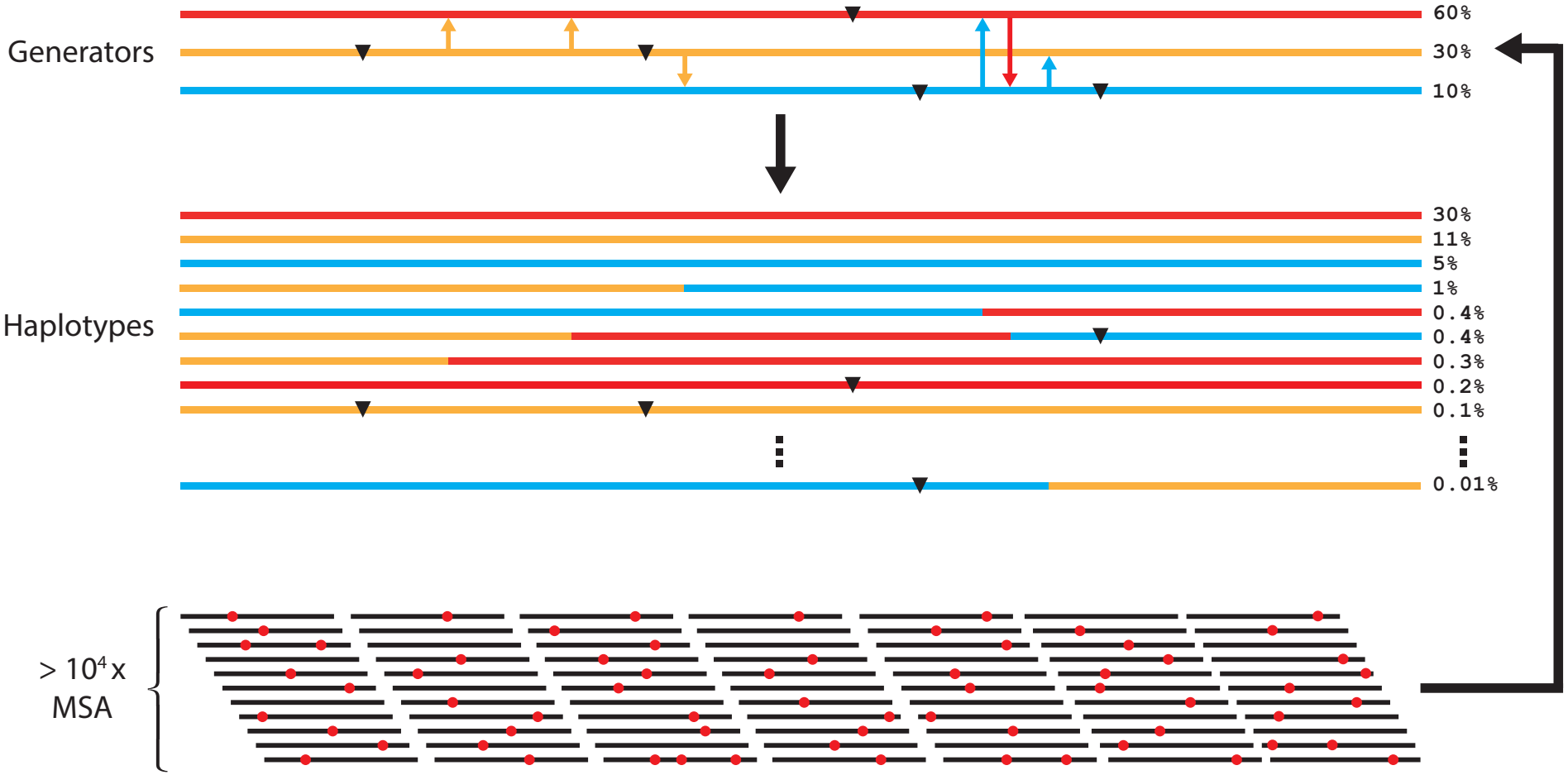
GLOBAL PHASING STRATEGY, OTHER APPROACHES



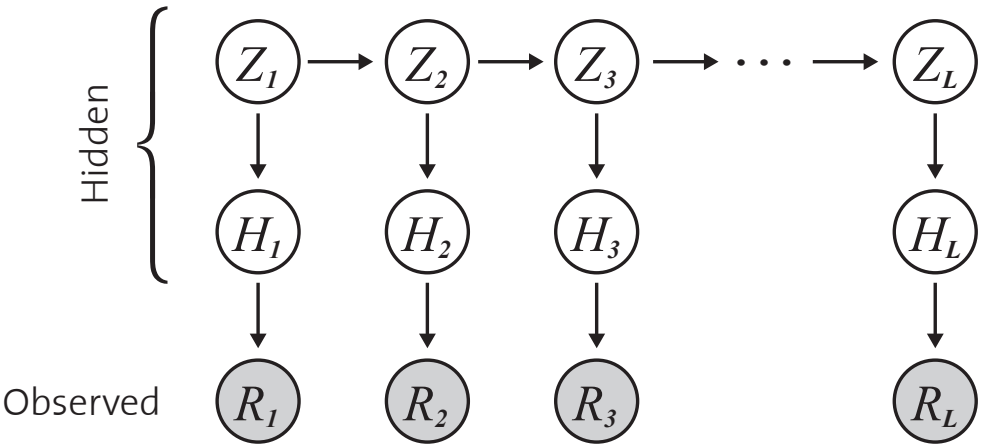
GLOBAL PHASING STRATEGY, NEW

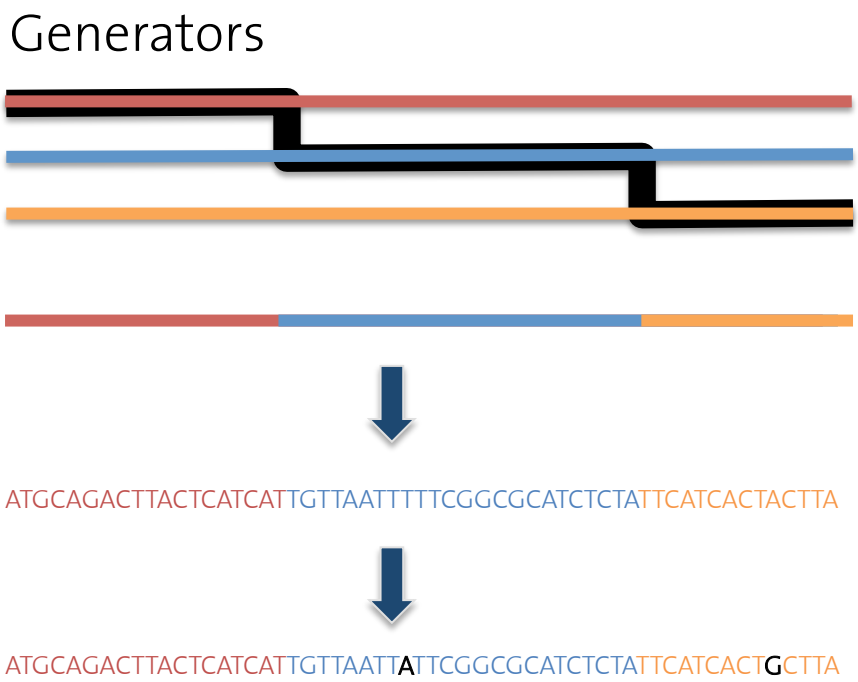


GLOBAL PHASING STRATEGY, NEW



MODEL

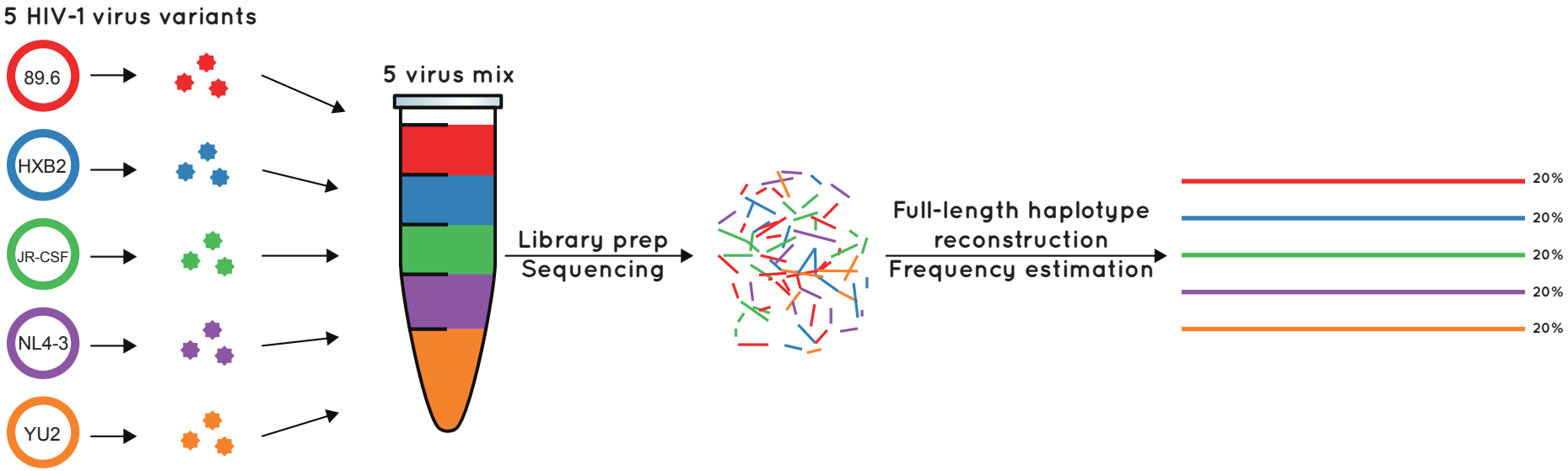




- Maximum a posteriori estimation:

$$P(\theta \mid R) \propto P(R \mid \theta) P(\theta)$$

5-VIRUS-MIX


5 HIV-1 virus variants
89.6
HXB2
JR-CSF
NL4-3
YU2
5 virus mix
Library prep
Sequencing
Full-length haplotype
reconstruction
Frequency estimation
20%
20%
20%
20%
20%

METHOD: SAMPLE PREPARATION FOR FULL-LENGTH SEQUENCING

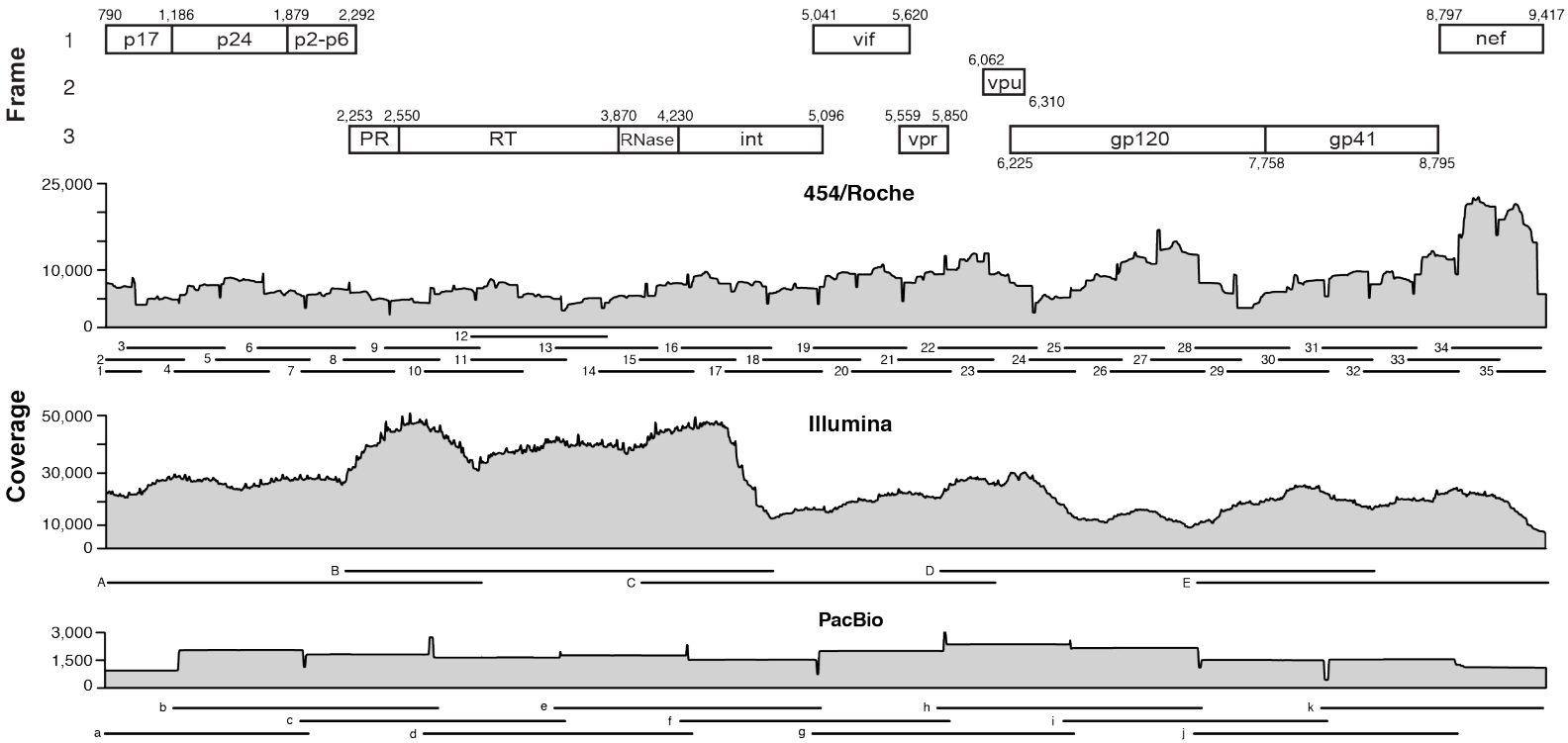
RNA isolation

cDNA synthesis of protease only	cDNA synthesis with 35 RT primers	cDNA synthesis with 5 RT primers	cDNA synthesis with 6 RT primers
End-point dilution	35 separate PCRs	5 separate PCRs	11 separate PCRs
qPCRs (one of five positive)		random fragmentation and adaptor ligation	
Sanger sequencing	454/Roche pyrosequencing	Illumina paired-end sequencing	Pacific Biosciences Single-molecule sequencing

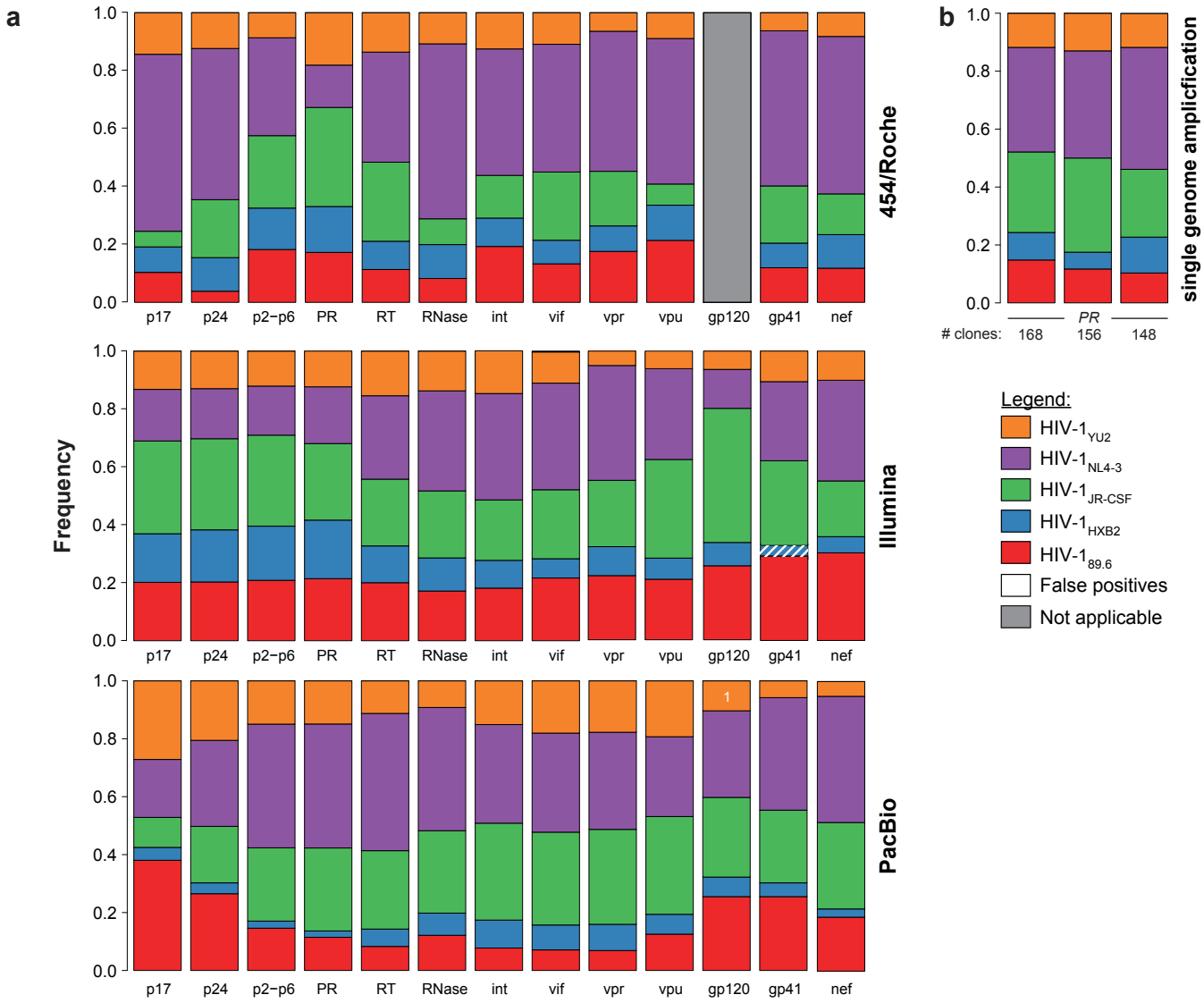
GOLD STANDARD

AMPLICON LAYOUT AND READ COVERAGE

Technology	Amplicons	~ Length (bp)	~ Overlap (bp)
454/Roche	35	510	250
Illumina MiSeq	5	2,500	750
PacBio RS	11	1,500	750

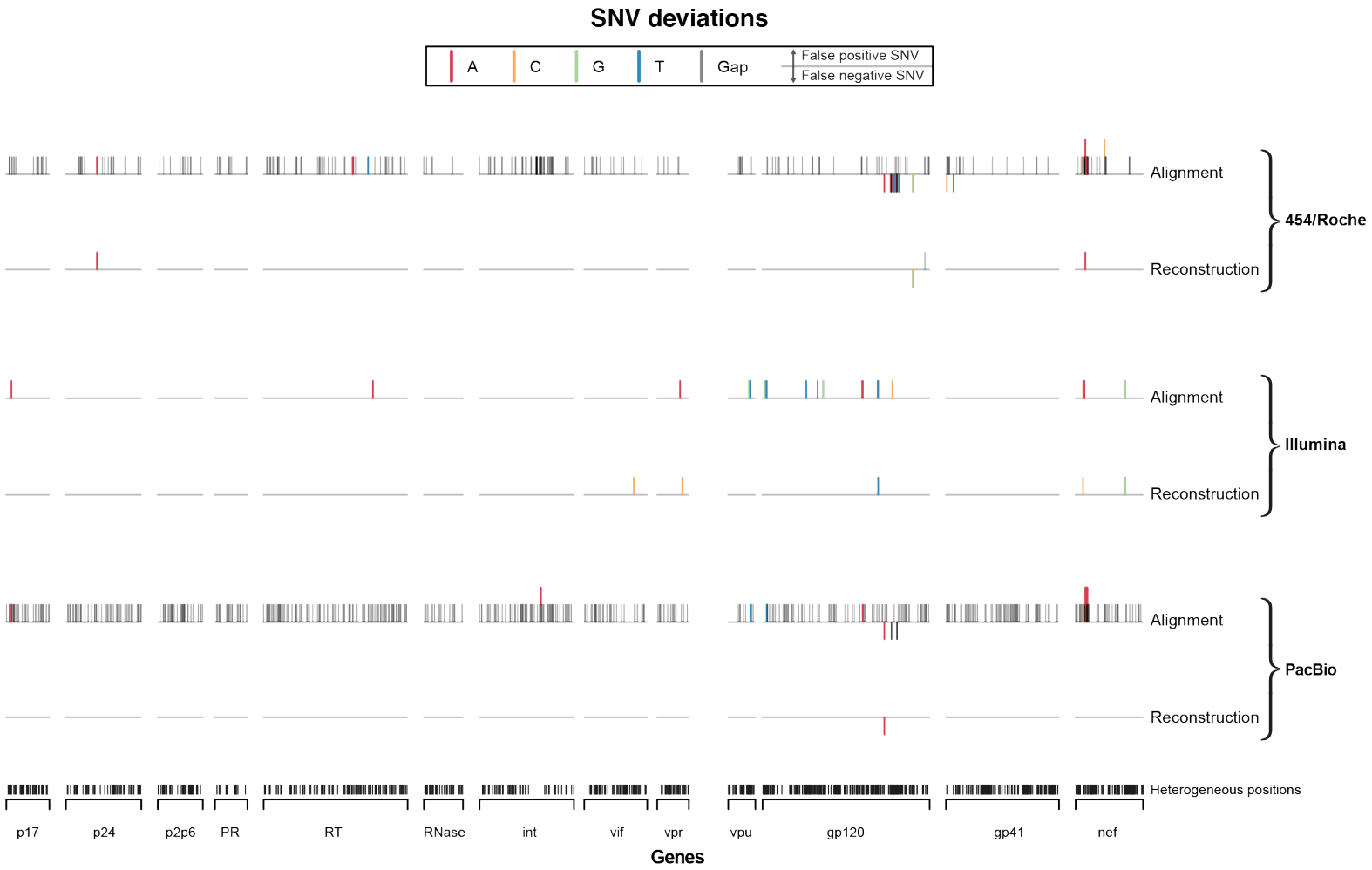


GENE-WISE RECONSTRUCTION



	p17	p24	p2−p6	PR	RT	RNase	int	vif	vpr	vpu	gp120	gp41	nef
Gene length (bp):	396	693	413	297	1,320	350	866	578	291	248	1,533	1,037	620
Min distance (%):	3.77	1.72	2.17	1.34	2.49	2.49	1.73	3.45	3.08	2.82	1.89	0.96	1.44
Max distance (%):	6.80	3.60	4.34	4.69	3.33	5.54	2.88	5.52	6.50	16.93	11.60	8.09	13.36
Mean distance (%):	5.23	2.82	3.30	2.75	2.88	4.62	2.29	4.81	4.96	12.82	9.50	6.29	8.06

NAÏVE SNV CALLING



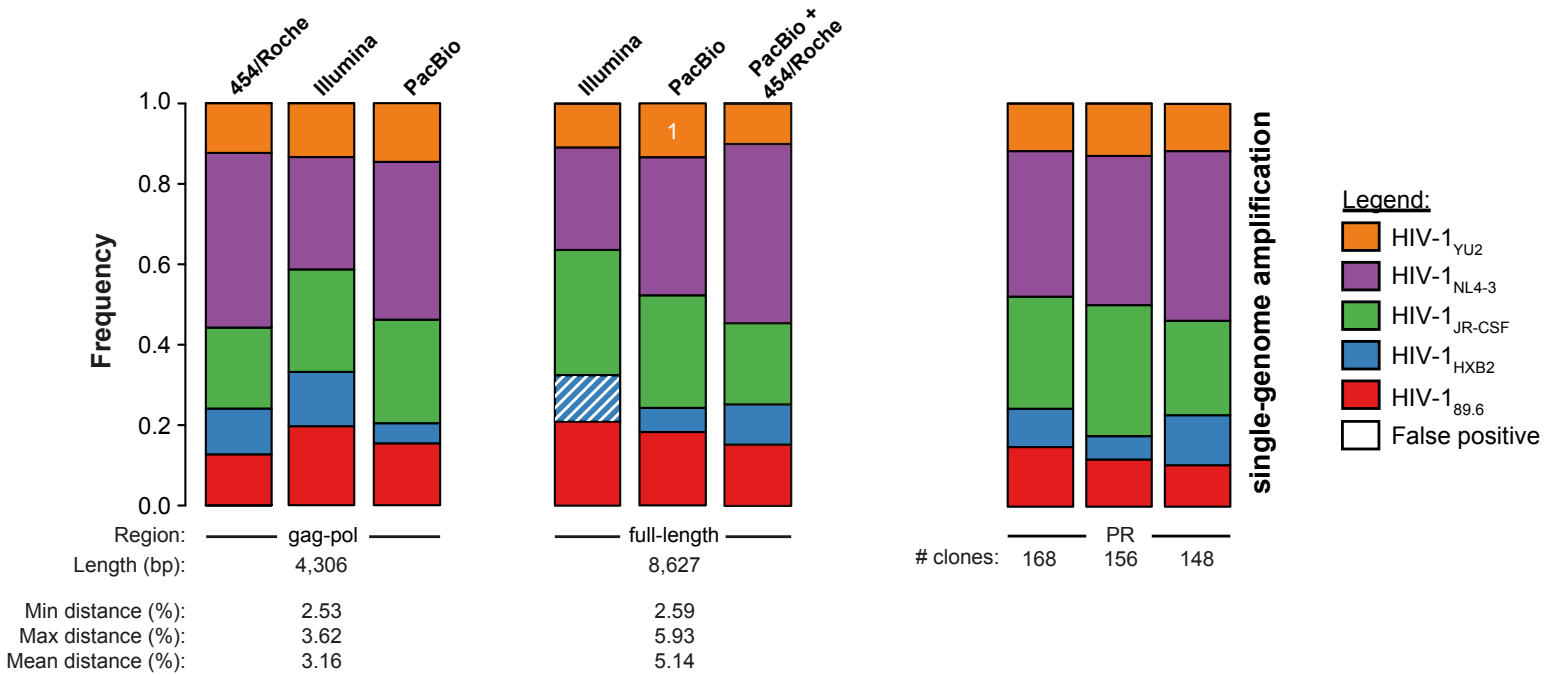
GENE-WISE RECONSTRUCTION

Estimated frequencies of reconstructed haplotypes across genes

	estimated frequencies % (mean±sd)				
Sequencing technique	HIV-$1_{89.6}$	HIV-1_{HXB2}	HIV-1_{JR-CSF}	HIV-1_{NL4-3}	HIV-1_{YU2}
454/Roche	13.7±5.1	11.0±2.4	18.2±8.6	46.2±12.9	10.9±3.5
Illumina	22.1±3.9	11.2±5.9	28.0±7.3	27.3±9.2	11.1±3.2
PacBio	16.5±9.7	5.8±2.6	27.2±6.4	35.9±7.9	14.6±6.2
SGA	12.3±2.3	9.2±3.3	27.9±4.6	38.4±3.2	12.2±0.7

NGS estimates **do not differ significantly** from SGA estimates

GLOBAL RECONSTRUCTION OF FUNCTIONAL REGIONS



Full-length variance of **0.4%** for 100 bootstrap samples

LESSONS LEARNED

- Naïve SNV calling is insufficient
- Experimental design:
  - Amplicon layout & overlap
  - Read length
  - Coverage
- Full-length reconstruction **is possible** with 454/Roche, Illumina, and PacBio
- Global frequency **estimates are reliable**

ACKNOWLEDGMENTS

University Hospital Zurich

Karin Metzner

Huldrych Günthard

Yannick Duport

Christine Leemann

Stefan Schmutz

Nottania Campbell

All other group members

ETH Zürich

Niko Beerenwinkel

Christian Beisel

Ina Niessen

University of Basel

Volker Roth

Melanie Rey

Sandhya Prabhakaran

Institute of Medical Virology

Alexandra Trkola

Michael Huber

Osvaldo Zagordi

Jürg Böni

NZR

Institute für Immunologie und Genetik (Kaiserslautern)

Martin Däumer

FGCZ

Rita Lecca

Andrea Patrignani

THANKS

bit.ly/quasirecomb