Comparison of 454 (Roche) versus MiSeq (Illumina) deep sequencing technologies in clinical samples from the (ANRS) 139 TRIO trial

CAHR 2014
P022

Guinevere Q. Lee[1,2], Christopher Lachowski[1,2], Eric Martin[1], Charlotte Charpentier[3], Christophe Rodriguez[3], Benoit Visseaux[4], Catherine Fagard[5], Jean-Michel Molin[6], Christine Katlama[7], Yazdan Yazdanpanah[8], Diane Descamps[3], Art Poon[1], P. Richard Harrigan[1,2]

1BC. Centre for Excellence in HIV/AIDS, Vancouver, BC, 2University of British Columbia, Vancouver, BC, 3Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Bichat-Claude Bernard, Laboratoire de Virologie, EA 4409 Université Paris-Diderot, Paris, France, 4Laboratoire de Virologie, Hôpital Bichat-Claude Bernard, Groupe Hospitalier Paris Nord Val de Seine, APHP, 46 rue Henri Huchard, 75018 Paris, France. , 5INSERM U897, Bordeaux, France, 6AP-HP, Hôpital Saint-Louis, Service de Maladies Infectieuses, Université Paris-Diderot, Paris, France, 7Service de Maladies-Infectieuses et Tropicales, Hôpital de la Pitié Salpêtrière, Groupe hospitalier Pitié Salpêtrière – Charles Foix, APHP, 47 boulevard Hôpital, 75013 Paris, France
8Hôpital de Tourcoing, Service de Maladies Infectieuses et Tropicales, Tourcoing, France.

Background & Objective

- "454 deep sequencing" has been popular in HIV drug resistance testing, but is susceptible to sequencing errors in homopolymer-rich regions.
- MiSeq is a promising replacement due to its robustness to homopolymers resulting in a lower error rate.
- Here, we compared the prevalence of amino acid variants in HIV *reverse transcriptase* (RT) and *integrase* (INT) regions reported by 454 versus MiSeq from samples derived from the Agence nationale de recherches sur le SIDA et les hépatites virales (ANRS) 139 TRIO study in France.

Methods

- ANRS 139 TRIO evaluated efficacy of raltegravir, etravirine, and darunavir in heavily treatment-experienced HIV-infected patients naïve to all three drugs.
- Here, we sequenced TRIO baseline samples using standard Sanger, 454 and MiSeq Nextera XT methods. For each sample, we performed one reverse transcription with first-round PCR amplification, then three different "nested-PCRs," one specific for each technology.
- Nucleotides were translated into amino acids for this comparison.
- To reduce noise, we excluded positions with MiSeq depth-of-coverage below 10,000 and variants of <1% prevalence by 454.

Results

- Paired 454 and MiSeq results were obtained in n=79 RT and n=59 INT samples at positions RT96-194 and INT83-193. Overall concordance of the observed % prevalence of amino acid variants was good (RT r^2=0.886, 8014-pairs; INT r^2=0.940, 6149-pairs).
- In RT and INT respectively, 138/8014-pairs (2%) and 68/6149-pairs (1%) had discordance >=20% (Figure 1 a and 1b).
- The majority of these discordances, 125/138 (91%) and 60/68 (88%), were found in homopolymer-rich regions (defined as >=3 consecutive identical nucleotides in HXB2), with the most frequent homopolymer-related RT error 99/125 (79%) observed in the vicinity of a critical resistance-associated position K103; MiSeq and Sanger were concordant at these positions, implying 454 homopolymer errors.
- The remaining pairs with >=20% discordances were nucleotide-mixtures by Sanger-sequencing (Figure 2), suggesting sampling bias in the second-round PCR rather than sequencing error.
- Finally, we focused on prevalence range 1-20% because 20% is roughly Sanger method's sensitivity limit. Concordance was extremely poor (RT r^2=0.143, 205-pairs; INT r^2=0.200 131-pairs; Figure 3a and 3b).

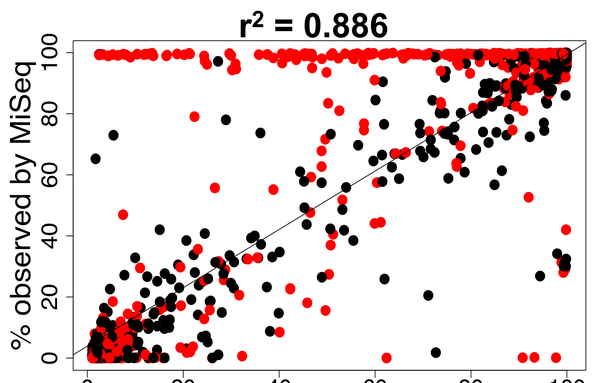
Conclusion

- Majority of discordances over 20% prevalence were found around homopolymer regions, where MiSeq outperformed 454 in predicting Sanger sequences.
- Even though overall concordance between 454 and MiSeq was good, concordance below Sanger-sequencing's sensitivity limit was poor.
- Future directions: Correlation with clinical outcome data will further assist in the interpretation of these observations in 454 and MiSeq.

Figure 1. Prevalence of amino acid variants observed by MiSeq versus 454 showed good overall concordance

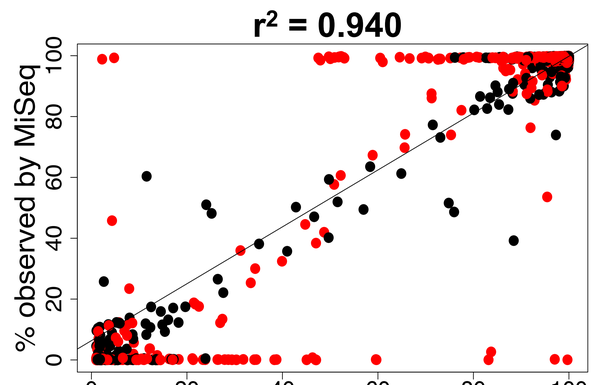
*Red dots correspond to homopolymer positions

(a) HIV reverse transcriptase 96-194



	Coverage	
	min	max
454	319	2247
MiSeq	10004	159930

(b) HIV integrase 83-193



	Coverage	
	min	max
454	309	2767
MiSeq	10005	146386

Figure 2. Most discordances of >=20% in amino acid prevalence were found in homopolymer-rich regions where MiSeq and Sanger agreed, implying MiSeq outperformed 454

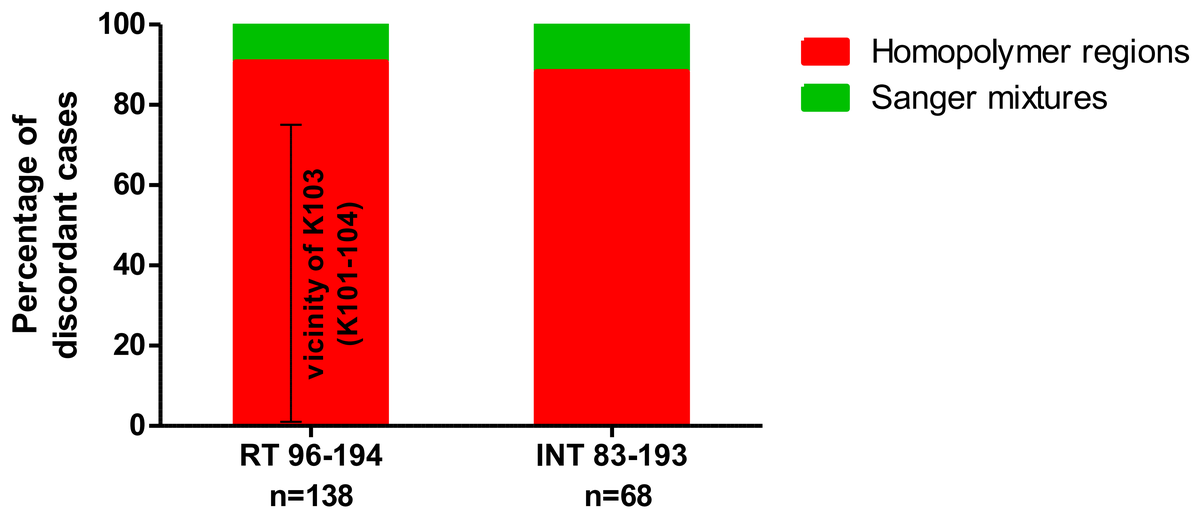
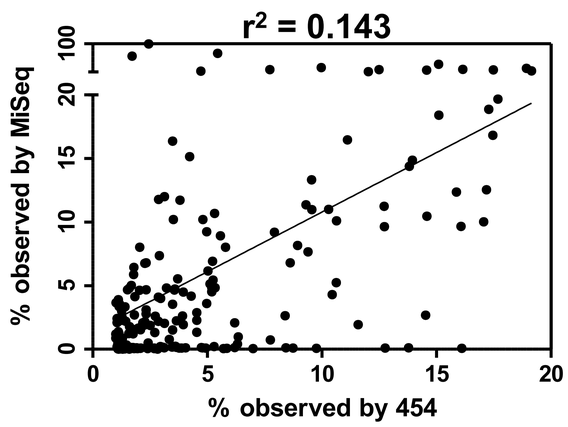
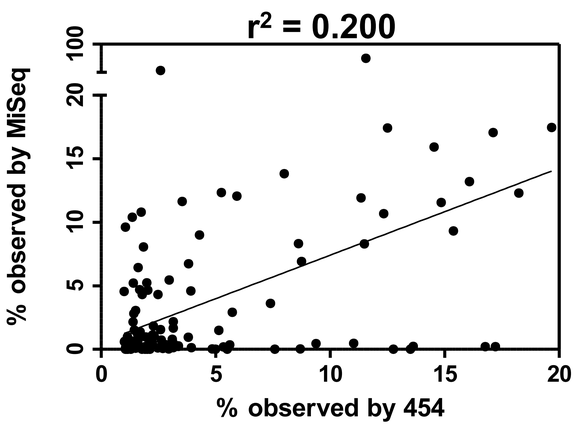


Figure 3. 454 and MiSeq correlated poorly below Sanger detection limit (1-20% by 454) despite the exclusion of homopolymer regions from analysis

(a) HIV reverse transcriptase 96-194



(b) HIV integrase 83-193



BRITISH COLUMBIA CENTRE *for* EXCELLENCE *in* HIV/AIDS

Providence HEALTH CARE
How you want to be treated

anrs France REcherche Nord & sud Sida-hiv Hépatites
Agence autonome de l'Inserm