Chapter 5

**The application of Seq2Res to evaluate high-throughput sequencing as a large-scale, cost-effective alternative to conventional HIV resistance genotyping**

# Methods and Materials

The datasets used in this study had been generated as part of the CIPRA-SA study (Comprehensive International Program for Research in AIDS in South Africa) which was a prospective, unblinded, randomized controlled trial of comparing “doctor-initiative-doctor monitored” and “doctor-initiative-nurse-monitored” strategies for antiretroviral drug monitoring in resource poor setting (Sanne et al., 2010). The study population consisted of 831 HIV infected individuals with a CD4+ count less than 350-cells/mm3 or AIDS-defining illness were enrolled on the study. HIV positive mothers with previous exposure of single dose nevirapine (NVP) drug for prevention of viral transmission from mother to child (PMTCT) during their pregnancy were also included in the study.

562 patients were followed up with the remainder not included in the study for reasons like drug toxicity, death, withdrawn consent or lost to follow-up. Baseline blood samples were retrieved from all 562 patients (sampled from 2005 – 2006). In this instance, baseline describes samples obtained from individuals immediately before initiation of first line antiretroviral therapy (ART). 71% of these patients received the drug combination D4T-3TC-EFV, 20% received D4T-3TC-NVP, 8% received D4T-3TC-LPV/r and 1% received D4T-3TC-NLF.

Virologic failure to the treatment was defined as decline of viral load less than 1.5 log10 from baseline to 12 weeks of treatment or two consecutive samples from a patient taken four weeks apart have viral load greater than 1000 RNA copies/ml . Virologic failure to first line ART was identified in 79 patients, with 15 patients failing second-line therapy . Blood samples had been retrieved for all of these individuals upon failure detection.

From all the obtained samples, the entire pol and reverse transcriptase genes of HIV were amplified using HIV subtype C specific primers. Ten HTS sequencing runs using the Roche/454 Junior platform had been attempted for 471 samples using MID tags to pool. 48 samples per sequencing plate. Further, sequencing was attempted for 630 samples using the Roche/454 FLX platform. 12 FLX runs were undertaken, dividing each plate into 8 distinct sections with 8 MID tagged samples per section for each sequencing run.

Conventional Sanger-based genotyping results were also available for 349 of the samples. All of the sequence data had been generated by our collaborators in the laboratory of Prof Maria Papathanasopolous at the University of the Witwatersrand Medical School, South Africa.

In the preliminary assessment of the sample’s sequence data from FLX and Junior, the samples in which protease (PR) or reverse transcriptase (RT) or both were not amplified were not considered for analysis. A total of 599 samples from FLX and 468 samples from Junior had PR and RT sequences (**Table 5.1**) and were considered for analysis.

Out of the samples that were eligible for analysis, 464 samples were sequenced in FLX and Junior platforms, 327 samples in FLX and conventional Sanger’s consensus method and 257 samples in Junior and conventional Sanger’s consensus method (**Table 5.2**).

Sequence data for all samples were analyzed drug resistance using Seq2Res computational tool using default settings. A sample was termed as resistant (R) if at least of a drug in baseline regimen was resistant (not intermediate resistant) to the sample. If a patient’s baseline regimen was not known then resistance to at least one of the possible baseline drugs was taken.

# Results

### 3.2.1. Genotyping of baseline samples using the Roche/454 FLX platform

FLX sequencing was successful for baseline samples from a total of 526 patients of which 187 samples had previous ARV exposure as a result of PMTCT while the remaining 339 had no previous exposure to ARVs. The eventual clinical outcome of all of these individuals was known and showed that out of the 339 no-PMTCT patients, 50 had virologic failure and 289 had virologic success. On the other hand, out of 187 PMTCT exposed patients, 25 had virologic failure and 162 had virologic success.

The obtained baseline blood samples were sequenced using Roche/454 FLX technology and analyzed using Seq2Res. The observation showed that the number of patients with resistance call to at least one drug in baseline regimen increased from 1 to 5 as the prevalence cutoff decreased to 1% (**Figure 5.2**). Significant difference was observed at 1% and 15% prevalence cutoffs (**Figure 5.2**).

### 3.2.2. Genotyping of baseline samples using the Roche/454 Junior platform

407 patients were sampled at baseline and sequenced using Roche/454 Junior, 250 patients had no previous PMTCT therapy and 147 patients had previous PMTCT therapy. The clinical outcome showed that out of 250 non-PMTCT patients, 40 had virologic failure and 210 had virologic success. In the other hand, out of 147 PMTCT exposed patients, 21 had virologic failure and 136 had virologic success.

The obtained baseline blood samples were sequenced using Roche/454 Junior sequencing technology and again analyzed using Seq2Res. As expected, the observation on resistance call on the samples showed that the number of patients with resistance call to at least one drug in baseline regimen increased as the prevalence of resistant sequence reads for a sample decreased from 20 to 1 (**Figure 5.3**). Similar to FLX data, only one sample that had virologic failure from non PMTCT group had resistant call at 20% prevalence cutoff; the number of samples increased to 4 when the prevalence cutoff was decreased to 1%. Like in FLX data, the significant difference was observed at prevalence cutoff of 1% (**Figure 5.3**).

## 3.1 Comparison of genotyping results between the Roche/454 FLX and Junior platforms.

The Junior and FLX platforms do not differ in their chemistry but differ on the basis of number of reads output. Thus, it is plausible to suggest that the generation of more sequence reads for a sample may sequence the amplified viral population to a ‘deeper’ level, thereby detecting low-abundance resistant variants. Our initial analysis, therefore, involved comparison of the resistance calls of individuals for which sequencing had been successful on both the Junior and FLX platforms.

Sequencing had been successful on both HTS platforms for 464 samples (**TABLE**). Thus, initial analysis focused on comparing the number of sequence reads generated by each platform for each sample and identifying if ‘deeper’ sequencing coverage resulted in more sensitive prediction of resistance. We saw that FLX platform generated on average 6034 sequence reads per sample (stand deviation 2297) while the Junior platform generated an average 1532 sequence reads per sample (Standard deviation 595, **Figure 5.1**). Thus, it is clear thatthe FLX platform produced significantly (P-value < 2.2-16) more reads per sample than the Junior platform.

Baseline samples from 405 individuals had been sequenced using both FLX and Junior platforms. Of them, 249 had no previous PMTCT therapy while 146 had previous exposure to PMTCT therapy. Of the 249 patients, 40 had virologic failure and 209 had virologic success to the first line antiretroviral therapy regimen. Of the 146 PMTCT exposed patients, 21 had virologic failure and 135 had virologic success in first line antiretroviral therapy.

At all prevalence cutoffs, there was no significant difference observed between the numbers of resistant and non-resistant samples sequenced using FLX and Junior (**Figure 5.4**). Thus, FLX and Junior were comparable at HIV genotyping for drug resistance test. Although, the lower number of sequence reads were generated by Junior, it was sufficient for HIV drug resistance test. Because, the cost of sequencing per sample is relatively cheaper in Junior than in FLX, Junior could be choice for large-scale low cost HIV drug resistance genotyping.

## 3.4. Comparison of resistance prediction by High Throughput technology and conventional “gold standard” Sanger’s consensus sequencing technology

### Comparison high throughput and conventional method for resistance prediction using baseline samples

A total of 302 baseline samples were sequenced using both FLX high throughput sequencing and conventional Sanger’s consensus sequencing technology. 168 of them had no previous PMTCT therapy exposure and 134 had previous PMTCT exposure. Out of 168 non-PMTCT patients, 16 had virologic failure and 152 had virologic success in first line antiretroviral therapy. Similarly, out of 134 previously PMTCT exposed patients, 13 had virologic failure and 121 had virologic success (**Figure 5.5**).

The numbers of samples with resistance call to at least one baseline drug were counted. For the samples sequenced using FLX, the numbers of samples with resistance call was calculated at prevalence cutoffs 20%, 15%, 10%, 5% and 1%. On the other hand, resistance calls on the numbers of samples sequenced using conventional Sanger’s consensus method was calculated at only at 20% prevalence cutoff because the technology generates a single sequence that represents the whole viral population and has a limitation of inability to call a base below 20% frequency while sequencing (Hudelson et al., 2010; Larder et al., 1993; Leitner et al., 1993; Schuurman et al., 1999; Van Laethem et al., 1999).

At the prevalence cutoff 20%, there was no significant difference observed between the numbers of resistant and non-resistant samples sequenced using high throughput FLX technology and conventional Sanger’s consensus method (**Figure 5.5**). This showed that high throughput FLX was comparable to the conventional Sanger’s consensus method for HIV genotyping and drug resistance testing.

### Comparison of Roche/454 high throughput FLX and Junior using first line ART virologic failure samples

#### 3.4.2.1 Resistance genotyping of samples collected from individuals at virologic failure

51 of the first line ART virologic failure samples had been sequenced using Roche/454 high throughput FLX technology. 15 of these had previous ARV exposure through PMTCT while 36had no previous exposure through PMTCT.

Genotyping using the FLX platform predicted resistance to at least one of the first line drugs at all prevalence levels in 14 out of 15 PMTCT samples. On the other hand, in the no-PMTCT sample, 23 out of 36 had predicted resistance to at least one of the first line drugs at all prevalence levels while 13 had no resistance identified (**Figure 5.6**).

The observation showed that there was a significant difference (p-value < 0.05) at all prevalence cutoffs and that indicated that it was more likely to correctly predict resistance in virologic failure samples from patients previously exposed to PMTCT.

Out of the 36 1st line failure samples sequenced using the Junior platform, 23 had no previous PMTCT exposure while 13 had previous PMTCT exposure. The numbers of predicted resistant and non-resistant samples were calculated at all prevalence cutoffs (**Figure 5.7**).

We observed that all VF samples in PMTCT group had predicted resistance to at least one drug in the regimen. There were significant differences (p-value 0.05%) at all prevalence cutoffs, which showed that, like Roche/454 FLX, Junior was also more likely to correctly predict resistance in VF samples from patients previously exposed to PMTCT.

FLX versus Junior…no difference therefore just present the consensus vs junior results

#### 3.4.2.4: Comparison of the genotyping performance of the Roche/454 Junior platform and conventional Sanger sequence genotyping using virologic failure samples

Results from both genotyping approaches were available for 13 individuals’ virologic failure samples. Out of the 13 patients, 6 had no previous PMTCT therapy and 7 had PMTCT therapy. At all prevalence levels in the HTS there was 100% concordance between the resistance calls for both approaches across all clinical outcome categories (**Figure 5.9**)

## 3.5. Resistance to nevirapine is more likely to be present at baseline in PMTCT exposed individuals.

For each sequencing platform we compared the resistance predictions for PMTCT versus non-PMTCT exposed individuals and identified the percentage of individuals in whom resistance to nevirapine at baseline was predicted (Figure 5.10). In all comparisons we found that the percentage of individuals with predicted resistance to nevirapine was always significantly higher (p < 0.05) in the PMTCT group when compared with the non-PMTCT exposed group. This discordance became more evident the ‘deeper’ the prevalence cutoff )Figure 5.10), suggesting a large number of PMTCT-exposed individuals were harbouring low-abundance NVP resistant viruses.

To ascertain whether the prediction of NVP resistance in PMTCT exposed individuals correlates with the time since NVP exposure, we compared the time since NVP exposure in baseline PMTCT samples with predicted NVP resistance and those predicted as susceptible to NVP. At prevalence thresholds of 15% and below (for both FLX and Junior) we find that the prediction of NVP resistance significantlycorrelates (p < 0.05) with time since NVP exposure. The median number of days since PMTCT exposure was observed to be 674 days for those individuals predicted as susceptible to NVP and 172 days for those predicted as resistant

In the presence of an antiretroviral drugs, the virus that has drug resistant mutations to the drug can still replicate while the wild type virus cannot. In the course of time, the resistant virus explodes in the viral population leading to virologic failure and the wild type virus exists in low abundance (Delobel et al., 2011; Le et al., 2009; Li et al., 2011; Paredes et al., 2010; Simen et al., 2009). After the treatment interruption, the viral fitness of wild type virus is better than the virus with drug resistant mutations and thus the wild type virus replicates (Deeks et al., 2005; Paquet et al., 2011; Rosenbloom et al., 2012) and soon explodes where as resistant virus replication steadily decreases and persist as minor variants, even to undetectable level (Deeks et al., 2003; Metzner et al., 2011).