Chapter 5

**The application of Seq2Res to facilitate large-scale, cost-effective HIV drug resistance genotyping using 454 pyrosequencing**

# Methods and Materials

The datasets used in this study were derived from a study called Comprehensive International Program for Research in AIDS in South Africa (CIPRA-SA), 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. The result of the study was published in Sanne et al 2010 (Sanne et al., 2010). The study population consisted 831 individuals confirmed with HIV infection and had CD4+ count less than 350-cells/mm3 or AIDS-defining illness. 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; the rest were not included in the study for reasons like – drug toxicity, death, withdrew consent or lost to follow-up. 660 baseline blood samples were obtained from 562 patients sampled from 2005 – 2006. The baseline blood samples were obtained from the patients before they underwent first line antiretroviral therapy (ART). 71% of the 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 is 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. 79 patients under the first line ART had virologic failure to the first line therapy. 51 blood samples from the first line ART virologic failure patients were obtained. 15 under the second line ART showed virologic failure.

From all the obtained samples, the entire protease (PR) and reverse transcriptase (RT) genes of HIV were amplified using HIV subtype C specific primers. 471 samples were attempted to be sequenced using Roche/454 Junior platform. 48 samples per sequencing plate, each tagged with unique multiplexed identifier (MID) were sequenced together in Junior. 630 samples were attempted to be sequenced using Roche/454 FLX platform. Each FLX run has 8 plates and each plate can sequence 8 samples, tagged with a unique MID. In total, 12 FLX runs and 10 Junior runs were obtained. Using conventional Sanger’s consensus method, the sequence data for 349 samples were obtained (**Table 5.1**).

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

## 3.1 Comparison of number of sequence reads per sample generated by FLX and Junior

Sequencing data from samples sequenced in Roche/454 FLX and Junior were used to compare and know if there was a significant difference in the number of sequence reads generated by both the platform. Of the 464 number of samples sequenced using both FLX and Junior, the number of sequence reads from Junior was compared to the number of sequence reads from FLX for the same sample. The comparison result showed that FLX generated on average 6034 sequence reads while Junior generated on average 1532 sequence reads (**Figure 5.1**). The observation showed that FLX produced significantly (P-value < 2.2-16) more reads per sample than Junior.

## 3.2. Comparison of resistant call in FLX and Junior at baseline samples

### 3.2.1. Prediction of resistance on baseline samples sequenced using FLX data

A total of 526 patients of which, 339 patients had no previous exposure to Prevention from Mother To Child Transmission (PMTCT) therapy and 187 samples had previous PMTCT therapy. The clinical outcome showed that out of 339 non-PMTCT patients, 50 had virologic failure and 289 had virologic success. In 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. Prediction of resistance on baseline samples sequenced using Junior data

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**).

## 3.3. Comparison of FLX and Junior sensitivity at resistance prediction using baseline samples

405 patients were sampled at baseline and sequenced using both FLX and Junior technology. 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 in the first line antiretroviral therapy. 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 Prediction of resistance on virologic failure samples sequenced using FLX data

In further analysis, blood samples were collected from the patients after virologic failure at first line ART. 51 of the first line ART virologic failure samples were sequenced using Roche/454 high throughput FLX technology. Out of them, 36 had no previous PMTCT therapy and 15 had previous PMTCT therapy. 14 out of 15 samples that had previous PMTCT therapy and virologic failure to the first line ART had resistance call. On the other hand, in samples that had no previous PMTCT therapy, 23 out of 36 had resistance to at least one drug while 13 had no resistance call to any drugs at all prevalence cutoffs (**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.

#### 3.4.2.2: Prediction of resistance on virologic failure samples sequenced using Junior data

50 virologic failure (VF) samples were sequenced using Roche/454 Junior technology but only 36 samples’ baseline ART was known. Out of the 36 VF samples, 23 had no previous PMTCT exposure while 13 had previous PMTCT exposure. The numbers of resistant and non-resistant samples were calculated at all prevalence cutoffs (**Figure 5.7**).

We observed that all VF samples in PMTCT group had 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.

#### 3.4.2.3: Comparison of Roche/454 FLX and conventional Sanger’s consensus method using virologic failure samples

Roche/454 FLX and conventional Sanger’s consensus method were observed to be comparable using the baseline samples. This was retested using the virologic failure samples. Of the 51 first line ART failed patients, sequenced using Roche/454 high throughput FLX technology, the sequence data from conventional Sanger’s consensus method was also available for 20 of them. Out of those 20 patients, 11 had no previous PMTCT therapy and 9 had previous PMTCT therapy. The number of resistant and non-resistant samples sequenced using conventional Sanger’s consensus method was calculated at prevalence cutoff 20% and for FLX at all prevalence cutoffs (**Figure 5.8**).

The observation showed that the number of resistant and non-resistant samples for both FLX and conventional Sanger’s consensus method were same at 20% prevalence cutoff in both non-PMTCT and PMTCT patients that VF at the first line ART therapy (**Figure 5.8**). Therefore, Roche/454 high throughput FLX technology was comparable to conventional Sanger’s consensus method for HIV genotyping and drug resistance testing using VF samples.

#### 3.4.2.4: Comparison of Roche/454 Junior and conventional Sanger’s consensus method using virologic failure samples

Roche/454 Junior and conventional Sanger’s consensus method were observed to be comparable using the baseline samples. This was retested with virologic failure samples. In another analysis, Roche/454 Junior high throughput technology and conventional Sanger’s consensus method were compared using the sequence data from the first line ART failure patients that were sequenced using both methods. Samples from 13 patients that had first line ART virologic failure were available that were sequenced using both Junior and conventional Sanger’s consensus method. Out of the 13 patients, 6 had no previous PMTCT therapy and 7 had PMTCT therapy. The number of resistant and non-resistant samples sequenced using Conventional Sanger’s consensus method was calculated at prevalence cutoff 20% and for Junior at all prevalence cutoffs (**Figure 5.9**).

The observation showed that the number of resistant and non-resistant samples for both Junior and conventional Sanger’s consensus method were same at 20% prevalence cutoff in both non-PMTCT and PMTCT patients that had the first line ART therapy (**Figure 5.9**). Therefore, Roche/454 Junior and conventional Sanger’s consensus method were comparable at HIV genotyping for drug resistance test using VF samples.

## 3.5. Pre-exposure of a drug drives viral evolution to the drug resistance

Patients that were sampled at baseline were grouped together as consensus, FLX and Junior based on the sequencing technologies used to sequence them. In each group – consensus, FLX and Junior, the number of patients were subgrouped as non-PMTCT and PMTCT based on their previous exposure to PMTCT therapy.

For baseline samples sequenced using conventional Sanger’s consensus method, a resistance calls to the drug NVP was made at 20% prevalence cutoff. The percentage of patients with previous PMTCT exposure and with NVP resistance was significantly different (p-value < 0.05) than the percentage of patients without previous PMTCT exposure and with NVP resistance (**Figure 5.10 A**).

In both non-PMTCT and PMTCT subgroups for Roche/454 FLX and Junior, the number of patients resistant to drug NVP was calculated at prevalence cutoffs 20, 15, 10, 5 and 1. At prevalence cutoff 20, resistant call prediction to drug NVP in non-PMTCT and PMTCT subgroups in both FLX and Junior were comparable to that of conventional Sanger’s consensus method (**Figure 5.10).**

However, as the prevalence cutoff was decreased to 1%, the percentage of patients with previous PMTCT exposure increased significantly than at prevalence 20% while there was no significant increment in percentage of patients without previous PMTCT exposure at 1% cutoff in both FLX and Junior (**Figure 5.10 B, C**). This indicated that resistance to NVP is significantly more likely (p < 0.05) to develop in PMTCT exposed patients than in PMTCT naive patients (regardless of prevalence level).

## 3.6. Correlation of resistant viral prevalence and time of sampling

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).

For previous PMTCT exposed patients that had virologic failure, the median days of the last nevirapine exposure at the time of sampling was observed to be 172 days in patients who were predicted as resistant to the drug NVP and 674 days in patients who were predicted as sensitive to the drug NVP. This showed that the prediction of nevirapine resistance significantly (p-value <0.05) correlates with the time since nevirapine exposure.