**Chapter 6**

**Final Summary**

In countries with a high burden of HIV, drug resistance testing is not routinely performed in a clinical setting due to the cost. The advent of Roche/454, and other, pyrosequencing platforms holds great promise in the development of a high-throughput, robust, reliable and affordable HIV drug resistance test. Further, the depth of coverage obtained using ultra deep pyrosequencing means that the presence of clinically relevant low abundance drug resistant viral variants within an individual’s viral population can also be explored.

The sheer volume of data generated by such an approach means that there is a need for powerful, sensitive and user-friendly bioinformatics applications for management and analysis of the data. We have developed a bioinformatics pipeline (Seq2Res) that takes sequence data directly from the Roche/454 sequencing platform and outputs drug resistance information for each sample.

As part of this pipeline we have developed a novel approach (QTrim) for quality trimming of the Roche/454 sequence reads. UDPS data outputs a quality score associated with every nucleotide and these must be accounted for prior to downstream analysis of the data, to ensure accurate drug resistance results and avoid false positives and false negatives. We compared the performance of QTrim with other widely used algorithms on both good and poor quality Roche/454 sequence data. We evaluated the methods based on mean read length, total number of reads and the percentage of poor quality bases in the resulting output. We found that our approach performed marginally better than the next best method for good quality data and significantly outperformed all methods when poor quality data is analyzed.

UDPS reads often contain PCR and sequencing induced errors which can artificially inflate viral diversity and may result in the generation of false positives in HIV resistance genotyping. These errors can potentially be corrected if all of the sequence reads originating from the same viral template. Recent studies have described the use of a primer ID approach that tags each viral template with a unique identifier during the cDNA generation step. Thus sequence reads deriving from the same viral template sequence can be identified and the generation of a consensus sequence from all of the reads originating from the same template can accurately reduce the presence of PCR and sequencing induced errors in the viral genotyping data. The analysis of this data is complex with no know algorithms currently published for analysis of such data. We have developed a module for the Seq2Res resistance testing platform that is capable of processing sample specific sequence data produced using the Primer ID approach. Application of this approach to primer ID data showed that while the primer ID approach can potentially reduce the presence of errors, it also has the potential for underrepresenting viral diversity by removing viral sequences from subsequent analysis because there are insufficient reads for a template to generate a consensus sequence.

We evaluated and verified the sensitivity of the resistance testing module of Seq2Res at accurately identifying drug resistant mutations using the two real consensus sequence datasets downloaded from Stanford resistance database. We found that Seq2Res is at least as sensitive as the Stanford database as it identified all the drug resistant mutations that were reported by the Web Sierra. In addition, Seq2Res was also able to identify a drug resistant mutation that was reported by Web Sierra incorrectly.

We also tested the ability of Seq2Res to report the prevalence of the drug resistant mutations correctly using five simulated datasets and showed that regardless of the prevalence level of the DRMs in the dataset, Seq2Res is capable of accurately identifying their presence at the correct prevalence level.

Finally, we applied Seq2Res for drug resistance genotyping on real biological datasets generated as part of the CIPRA-SA (Comprehensive International Program for Research in AIDS in South Africa) study. 471 samples were genotyped using Roche/454 Junior and 630 samples were genotyped using Roche/454 FLX. Both datasets were evaluated using the Seq2Res resistance testing platform. Consensus sequences for 349 samples from conventional genotyping were also available and were used for comparison with the UDPS data.

Roche/454 FLX and Roche/454 Junior genotypic data for baseline samples showed that the number of samples that were predicted as resistant increased when the prevalence of resistance was decreased to 1% cutoff. There was no significant difference in the number of samples predicted as resistant, showing that FLX and Junior were comparable at genotypic HIV drug resistance prediction. We further compared sensitivity of genotypic HIV drug resistance prediction of both UDPS and population based Sanger genotyping at prevalence cutoff 20%, which showed that UDPS and population based Sanger genotyping were comparable. However, in an earlier study by Wang et al (Wang et al., 2007) showed that UDPS detected 58 HIV variants per sample as compared to only 8 variants per sample detected by population based Sanger genotyping method. This showed that UDPS is more sensitive than population bases Sanger method.

In our analysis, the number of baseline samples predicted as resistant, using UDPS, was higher in PMTCT group than in no-PMTCT group at the prevalence cutoff 15% and below. It showed that NVP resistance was more likely to develop in NVP exposed patients in PMTCT group than in the drug naïve group. Further, we analyzed the correlation of time since NVP exposure and virologic clinical outcome to first line ART. We knew the date of ART and the date of PMTCT for the individuals. The time (in days) of NVP exposure before ART initiation was calculated from these known time points for the individuals in both PMTCT group and no-PMTCT group. The result showed that the prediction of NVP resistance significantly correlates (p < 0.05) with time since NVP exposure. The samples from patients that receive sdNVP at median time of 174 days before ART initiation were predicted resistant while the samples from patients that receive sdNPV at median time of 631 days were predicted sensitive. Our finding correlates with the previous studies as well. Chi et al (Chi et al., 2007) showed that ART initiation before six months of NVP exposure could be a risk factor for clinical poor outcome. Coovadia et al (Coovadia et al., 2009) observed that NVP based ART initiation after 18-36 months of NVP exposure had sustained likelihood of virologic suppression. Similarly, Stringer et al (Stringer et al., 2010) observed as the time since NVP exposure increased, the rate of VF in NVP containing ART was declined.

Roche/454 FLX and Roche/454 Junior genotypic data from first line virologic failure (VF1) samples were studied for resistance. Seq2Res reported that up to 100% of the VF1 samples from PMTCT and ~65% of the VF1 samples from no-PMTCT were resistant. It showed that drug exposed individuals are more likely to experience first line VF than the drug naive individuals.

In conclusion, UDPS is capable of massive parallel pyrosequencing to explore unprecedented range of HIV variants, including the clinically relevant low abundant drug resistant variants missed by conventional population based Sanger genotyping method. In addition, UDPS is much cost effective per sample meaning that it can be applied for massive resistance test in the poor settings like sub-Saharan African regions. However, PCR and sequencing errors in UDPS genotypic data can create artificial biodiversity. Low quality bases in the UDPS genotypic data need to be removed, thus, we have created a novel trimming algorithm – QTrim (Shrestha et al., 2014) that performs better than other widely used tools in poor quality data. Any remaining poor quality bases could still compromise the downstream drug resistance analysis. Therefore, the quality of the bases at the DRM positions should be considered. We have created a HIV drug resistance testing pipeline called Seq2Res that can take the UDPS genotypic data directly as input. We have incorporated primer ID module for analyzing the data generated using primer ID technology. Seq2Res has been validated for drug resistance call sensitivity and accuracy using the two real data sets and five simulated data set at five different prevalence cutoffs. Finally, Seq2Res was applied on the real biological datasets generated using Roche/454 FLX and Roche/454 Junior platforms from NVP exposed and non-exposed individuals. Our findings showed that at 15% and below (both Roche/454 FLX and Roche/454 Junior), the prediction of NVP resistance significantly correlates (p < 0.05) with time since NVP exposure. Our findings also correlated with other previous studies (Chi et al., 2007; Coovadia et al., 2009; Stringer et al., 2010).

Chi, BH, Sinkala, M, Stringer, EM, Cantrell, RA, Mtonga, V, Bulterys, M, Zulu, I, Kankasa, C, Wilfert, C, Weidle, PJ, Vermund, SH, Stringer, JS (2007) Early clinical and immune response to NNRTI-based antiretroviral therapy among women with prior exposure to single-dose nevirapine. *AIDS* **21**: 957-964.

Coovadia, A, Hunt, G, Abrams, EJ, Sherman, G, Meyers, T, Barry, G, Malan, E, Marais, B, Stehlau, R, Ledwaba, J, Hammer, SM, Morris, L, Kuhn, L (2009) Persistent minority K103N mutations among women exposed to single-dose nevirapine and virologic response to nonnucleoside reverse-transcriptase inhibitor-based therapy. *Clin Infect Dis* **48**: 462-472.

Shrestha, RK, Lubinsky, B, Bansode, VB, Moinz, MB, McCormack, GP, Travers, SA (2014) QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BMC Bioinformatics* **15**: 33.

Stringer, JS, McConnell, MS, Kiarie, J, Bolu, O, Anekthananon, T, Jariyasethpong, T, Potter, D, Mutsotso, W, Borkowf, CB, Mbori-Ngacha, D, Muiruri, P, Ong'ech, JO, Zulu, I, Njobvu, L, Jetsawang, B, Pathak, S, Bulterys, M, Shaffer, N, Weidle, PJ (2010) Effectiveness of non-nucleoside reverse-transcriptase inhibitor-based antiretroviral therapy in women previously exposed to a single intrapartum dose of nevirapine: a multi-country, prospective cohort study. *PLoS Med* **7**: e1000233.

Wang, C, Mitsuya, Y, Gharizadeh, B, Ronaghi, M, Shafer, RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res* **17**: 1195-1201.