Chapter Outline

Chapter 2: This chapter introduces and describes a novel algorithm, QTrim, designed for the quality trimming of UDPS sequence data with Phred-like quality scores (e.g. Roche/454, Illumina and Ion Torrent) and the evaluation of its performance in comparison to other widely used tools. The results show that QTrim is comparable to the next best tool while quality trimming a good quality data set and outperforms all the tools while trimming a poor quality data set. The tool has been published in BMC Bioinformatics (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).

Chapter 3: This chapter introduces and discusses the application of primer ID technology developed by Jabara and colleagues (Jabara et al., 2011) to reduce the PCR and sequencing errors. We developed a module to facilitate the analysis of HIV drug resistance genotyping data generated using the primer ID technology in the Seq2Res drug resistance testing pipeline. We describe the workflow of the primer ID module and discuss the limitations of the technology including primer ID collision and underrepresentation of the HIV variants in the viral population.

Chapter 4: This chapter introduces and describes the Seq2Res computational pipeline that facilitates low cost HIV drug resistance through easy analysis of HIV drug resistance genotyping data generated using UDPS sequencing technologies. The chapter describes a workflow of the pipeline, the requirements of the pipeline, HIV drug resistance output files and plots that summaries overall analysis. We evaluated and validated drug resistant mutation calls in the Seq2Res pipeline in comparison with the Stanford HIV drug resistance interpretation algorithm using two biological datasets downloaded from the Stanford Database. We validated the prevalence calls of drug resistant mutations in the Seq2Res pipeline using five simulation datasets with known prevalence of known drug resistant mutations. We observed that regardless of the prevalence level of the drug resistant mutations in the dataset, Seq2Res is capable of accurately identifying their presence at the correct prevalence level.

Chapter 5: In this chapter we present the application of Seq2Res to analyze HIV drug resistance genotyping sequence data generated using the Roche/454 Junior platform and the Roche/454 FLX platform from HIV infected individuals sampled as part of CIPRA-SA (Comprehensive International Program for Research in AIDS in South Africa) study. We compared the results with those obtained using Standard population based consensus sequencing and observed that:

1. The sensitivity of the Roche/454 Junior and Roche/454 FLX platforms are comparable for HIV drug resistance genotyping.
2. Ultra deep pyrosequencing is, at least, comparable to conventional population based Sanger method at HIV drug resistance genotyping
3. Resistance to nevirapine is significantly more likely to be observed in individuals previously exposed to ARVs through PMTCT than in drug naïve individuals.
4. At 15% and below (both FLX and Junior), the prediction of NVP resistance significantly correlates (p < 0.05) with time since NVP exposure

Chapter 6: This chapter summary the significance of the work, development of Seq2Res pipeline, validation of the pipeline and the application of the pipeline