Chapter Outline

Chapter 2: This chapter introduces and describes a novel algorithm - QTrim designed for quality trimming of Roche/454 ultra-deep high throughput sequence data and its performance comparison with other widely used tools. The comparison result shows 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 just been published in the journal “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 analyze ultra deep pyrosequence genotypic data generated using primer ID technology as a part of Seq2Res drug resistance testing pipeline. We talk through the workflow of the primer ID module to generation of consensus sequences from the sequence reads with same primer ID tag. We also discuss the limitations of the technology like primer ID collision and underrepresentation of the HIV variants in the viral population.

Chapter 4: This chapter introduces and describes Seq2Res computational pipeline that facilitates low cost HIV drug resistance test. 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 Seq2Res pipeline in comparison with the Stanford HIV drug resistance interpretation algorithm using the two biological datasets downloaded from Stanford Database. We validated the prevalence call of the drug resistant mutations in 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: After validation of Seq2Res pipeline, in this chapter, we discuss the application of Seq2Res to analyze 471 samples generated using Roche/454 Junior platform and 630 samples generated using Roche/454 FLX platform from HIV infected individuals generated as part of CIPRA-SA (Comprehensive International Program for Research in AIDS in South Africa) study. We observed the following results from the data analysis with Seq2Res:

1. Roche/454 Junior platform and Roche/454 FLX platform were comparable at 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 develop in the drug-exposed individuals than in the 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