ABSTRACT

HIV genotypic data help clinicians for better antiretroviral treatment of the infected individuals but genotypic resistance test in high HIV burden poor settings is not routinely performed for cost reason. Conventional population based Sanger genotyping method misses out low abundant clinically relevant viral variants. Ultra deep pyrosequencing (UDPS) is a massive parallel sequencing technology and is a promising alternative to the conventional genotyping method. But the shear volume of UDPS data necessitates a computational tool for analysis. We have developed Seq2Res computational pipeline for HIV drug resistance test from UDPS genotypic data.

Seq2Res uses QTrim to trim out low quality bases. QTrim comparison with other widely used tools showed that it is equivalent to the next best method at trimming good quality data but outperforms all methods at trimming poor quality data. Primer ID is novel technology to reduce the PCR and sequencing errors using a unique degenerate primer identifier. We have also included a module to analyze the data generated using primer ID technology.

We validated Seq2Res pipeline using two real biological datasets from Stanford HIV Database and five simulated datasets, each containing different known prevalence of drug resistant mutations (DRMs). The list of DRMs generated by standard HIV drug resistant interpretation algorithm in web sierra was compared the list generated by Seq2Res. In addition to all DRMs that were identified by the standard method, Seq2Res accurately identified a DRM that was incorrectly interpreted by the former method. Validity test on simulated datasets showed that Seq2Res is also capable of accurately identifying the presence of DRMs at the correct prevalence level

Finally, Seq2Res was applied on 471 samples, 630 samples and 349 samples genotyped using Roche/454 Junior, Roche/454 FLX and conventional Sanger genotyping methods respectively and were generated as part of CIPRA-SA. The date of Antiretroviral therapy (ART) and date of nevirapine (NVP) for those in PMTCT therapy were known. Drug resistance in a sample was predicted at prevalence cutoffs 20%, 15%, 10%, 5% and 1%, if a sample had drug resistant mutation to at least one drug at that prevalence cutoff. Both FLX and Junior data showed that the number of samples (baseline samples and virologic failure samples) that were predicted as resistant to at least one drug increased towards lower prevalence cutoff. The increment was higher in NVP exposed individuals than non-exposed individuals. Resistance call in samples sequenced both in UDPS and conventional Sanger genotyping method also showed that they were comparable. We further looked at virological resistance and the time of ART initiation since the data of NVP exposure. Our finding showed that the samples receiving ART at median time of 631 days were predicted as sensitive while the samples receiving ART at median time of 174 days were predicted as resistant. This showed that the prediction of NVP resistance significantly correlates (p < 0.05) with time since NVP exposure.