The International HIV Resistance Response Database Initiative: A New Global Collaborative Approach to Relating Viral Genotype and Treatment to Clinical Outcome

B.A. Larder, V.DeGruttola, S. Hammer, R. Harrigan, S. Wegner, D. Winslow & M. Zazzi

On behalf of the HIV Resistance Response Database Initiative (RDI)

Abstract 102
XI International HIV Drug Resistance Workshop
Seville, Spain, 2-5 July 2002

Goal

- To develop a relational database to correlate HIV drug resistance-associated genotype data with virological response to antiretroviral agents

Aim

- Initial aim is to collect genotype, treatment and clinical outcome information (viral load) from substantial numbers of patients (many thousands)
- To organize data in an Oracle-based relational database
- To analyze data using a variety of methodologies to relate resistance mutation patterns to clinical response
- Provide wide access to enable the database to be queried via the Internet

Basic concept

- The development of the RDI database is a collaborative initiative. Several research groups are working together to make significant contributions and will subsequently have access to all of the data, enabling them to perform analyses with the consensus of the core group. It is the intention to share the outcome of these analyses with the FDA for purposes of their evaluation of interpretation algorithms submitted to them for review & approval. In addition, it is also the intention of the RDI to present significant findings at major HIV conferences and to submit the results for publication. The database will be made available publicly as a query only database. As such, the database will be available to address specific questions physicians or scientists have using a user-friendly, web-based query tool, although the raw data itself will not be available to download. The RDI "core group" will be instrumental in developing the database & acting as reviewers for any requests to perform analyses on the data set

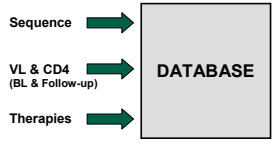
Current Core Team

- Victor DeGruttola (Harvard School of Public Health, Boston, MA, USA)
- Scott Hammer (Columbia University, New York, NY, USA)
- Richard Harrigan (BC Center for Excellence in HIV/AIDS, Vancouver BC, Canada)
- Brendan Larder (Visible Genetics, Cambridge, UK)
- Scott Wegner (US Military HIV Research Program, Rockville, MD, USA)
- Dean Winslow (Visible Genetics, Suwanee, GA, USA)
- Maurizio Zazzi (University of Siena, Siena, Italy)

Core Team

- *Mission of the Core Team:*
  - Contribute data and develop database
  - Ensure data meets appropriate QA standards
  - Develop initial data analysis plan
  - Review requests to analyse data from the database

Data Input



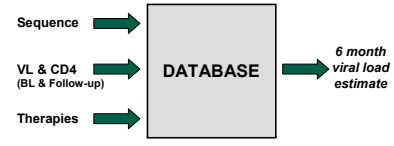
VL = Viral Load
BL = Baseline

Data Captured

Mandatory Fields:
- Viral load (including assay used)
- Viral load assay date
- Unique Patient Identifier
- Patient Sex
- Pediatric/Adult variable: patient >12 years YES/NO
- Genotype Test Date
- Reference Sequence
- Protease Start/End Index
- RT Start/End Index
- Generated Protease Sequence
- Generated RT Sequence
- Therapy during time period

Optional Fields:
- CD4 Count
- CD4 date
- Adherence indication
- Plasma drug levels

Data Output



Data QA

CINICAL DATA:
- To comply with the data protection privacy acts of all countries all data submitted to the HIV Resistance RDI database is stripped of identifiers that can trace the data to the patient source. All investigators contributing data are required to sign an agreement that confirms data submitted has been re-coded & the link between the original code & this code has been severed
- Cohort data & clinical trial data separately designated in the database as data quality is likely to be different

SEQUENCE DATA:
- All submitted data should follow a genotyping protocol or procedure to ensure consistency
- Laboratories submitting data shall be accredited from a local regulated body (i.e. CLIA/CAP), to ensure the laboratory is complying to Good Laboratory Practice (GLP) standards
- Minimal Length & standard of Sequence:
  - *Protease:* A bi-directional sequence of the protease gene from codons 10 to 99.
  - *RT:* A bi-directional sequence of the reverse transcriptase region from codons 41 to codon 235

Data Analysis

- To date, data identified from about 3500 patients
  - More when additional genotyping is performed
- Oracle database hardware & software in place (with dedicated support)
  - DB architecture constructed
- Power calculations have been performed to estimate approximate number of required data points *(see DiRienzo & DeGruttola)*
- Initial NN models constructed *(see Wang et al)*

Summary

- The RDI is focused on establishing relationships between baseline genotype and virological response via analysis of a large clinical dataset
- Significant progress has been made:
  - Sources of data identified
  - Database architecture constructed
  - Modeling work has begun
- This initiative is open for groups to join and aims to provide open access to query the database
- Utilization of large databases is likely to improve the accuracy of genotypic interpretation

References

DiRienzo G & DeGruttola V. Collaborative HIV resistance-response database initiatives: Sample size for detection of relationships between HIV-1 genotype and HIV-1 RNA response using a nonparametric approach. *XI International Workshop on HIV Drug Resistance.* Seville, Spain, 2002 [Abstract X].

Wang D, DeGruttola V, Hammer S, Harrigan R, Larder B, Wegner S, Winslow D & Zazzi, M. On Behalf of The HIV Resistance Response Database Initiative (RDI). A Collaborative HIV Resistance Response Database Initiative: predicting virological response using neural network models. *XI International Workshop on HIV Drug Resistance.* Seville, Spain, 2002 [Abstract 119].