

RDI

HIV Resistance Response Database Initiative

Seville 2002 Presentation



Goal

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

Aim

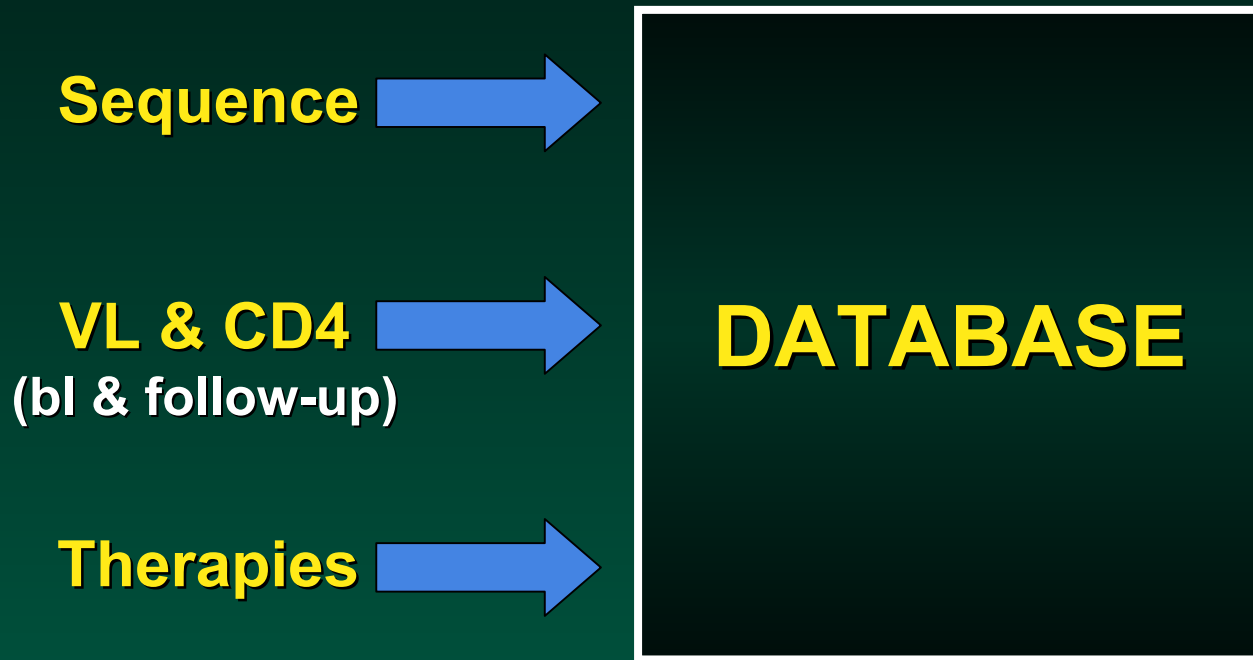
- **Initial aim is to collect genotype, treatment & clinical outcome information (VL & CD4) from substantial numbers of patients**
- **To organise data in an oracle-based relational database**
- **To analyse data using a number of approaches to relate resistance mutation patterns with clinical response**
- **Provide wide access to interrogation of data via the internet**

Core Team

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



Data Input



Data QA

- **Specific QA standards applied to data submitted to the database**
- **Developed for :**
 - **Clinical Data (Cohort, Clinical Trial)**
 - **Sequencing Data (RT, Protease regions)**



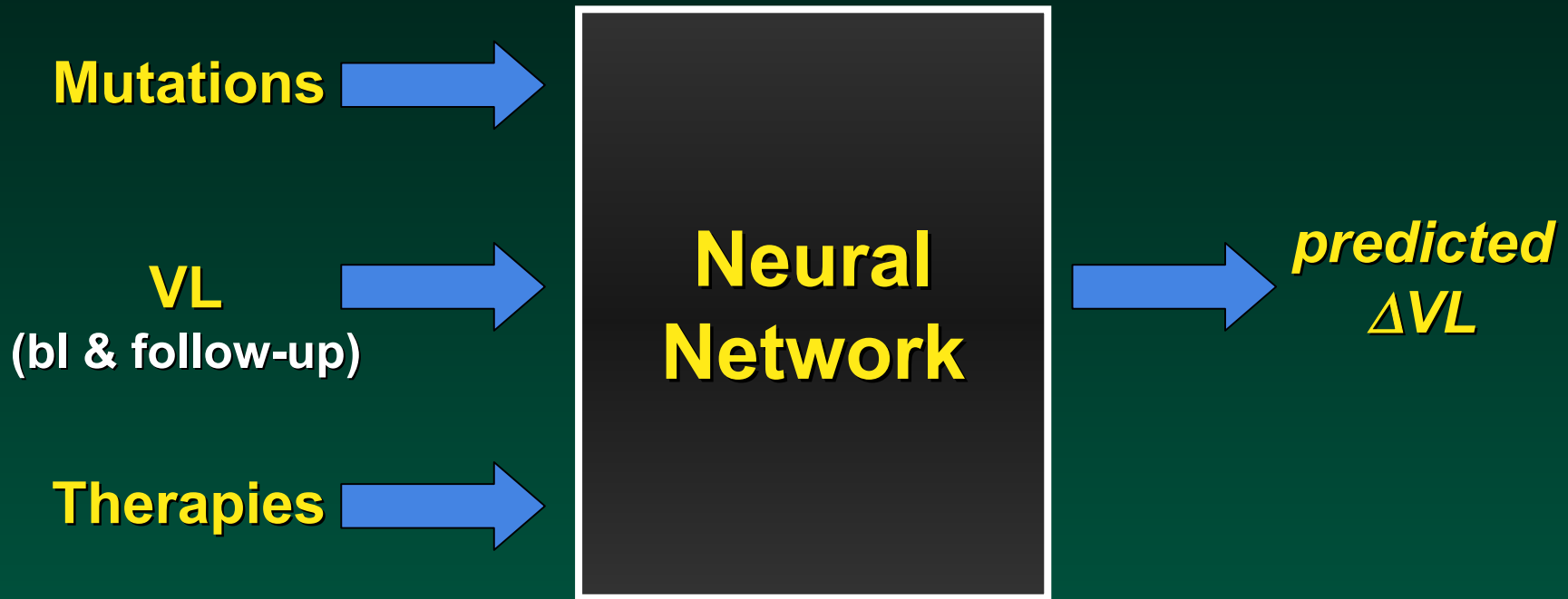
Data Analysis

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



Neural Network Model

(Dechao Wang et al)



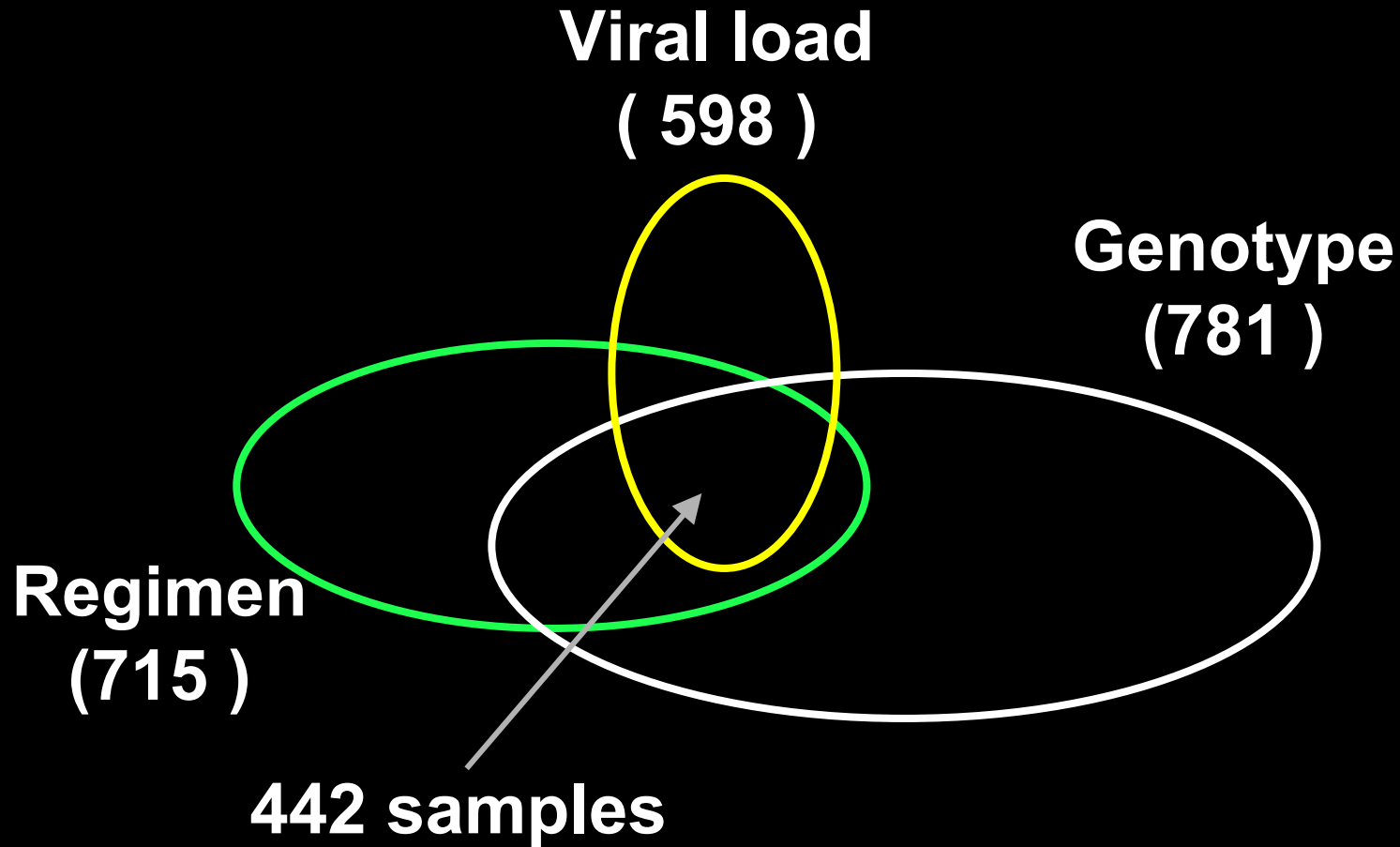
62-Parameter Neural Network model

(Dechao Wang et al)

- **Input variables**
 - 20 PI codon positions
 - 29 RT codon positions
 - 12 drugs (5 PIs, 5 NRTIs, 2 NNRTIs)
 - Duration of therapy (weeks)
- **Output variable**
 - Viral load change at on-therapy time points

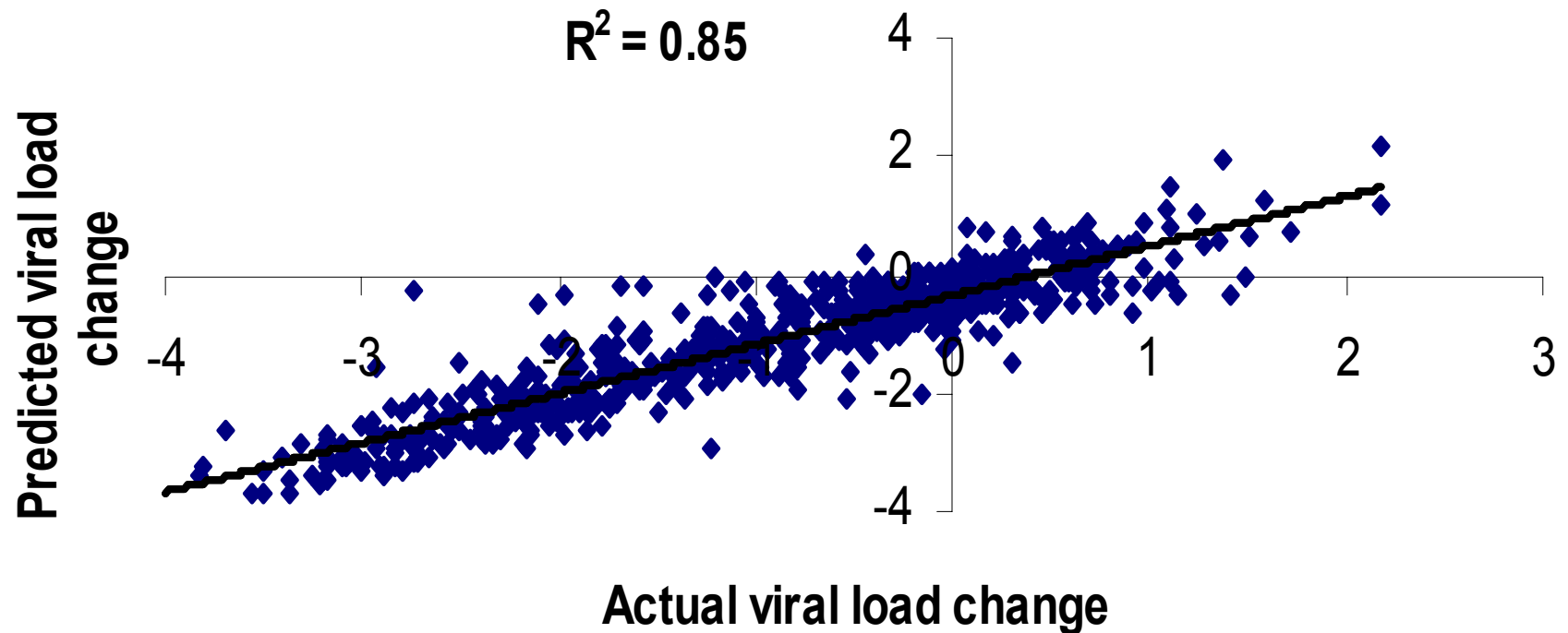


VGI Vigilance II Database

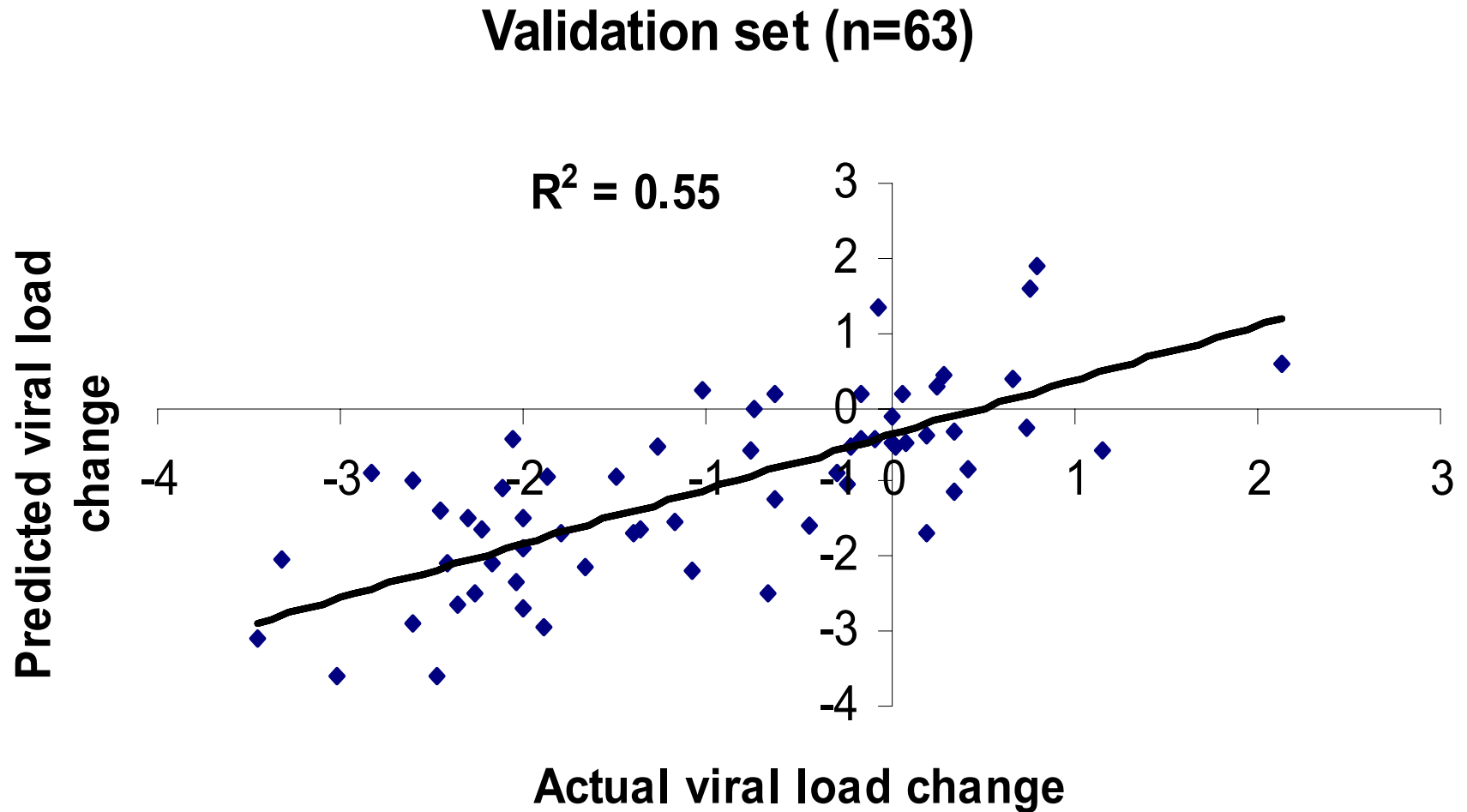


Linear Regression Analysis: *Training Set*

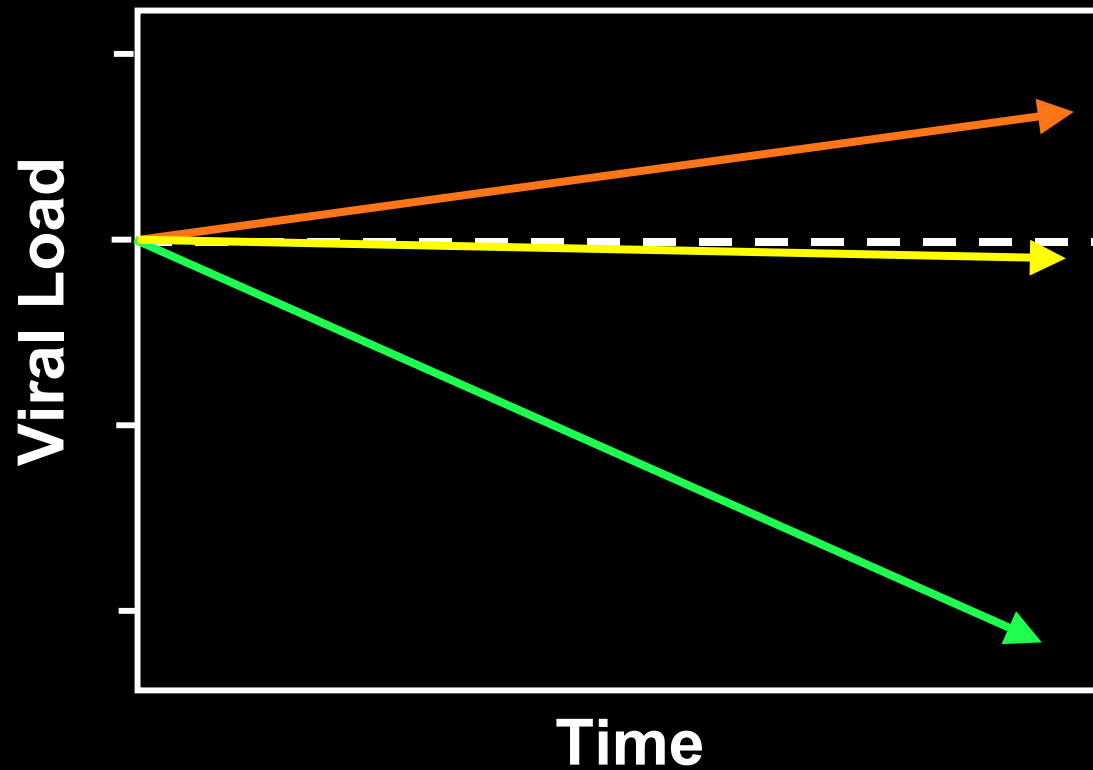
Training set (n=639)



Linear Regression Analysis: *Independent Set*



Predicting VL Trajectory: Test Data Set



NN Prediction:

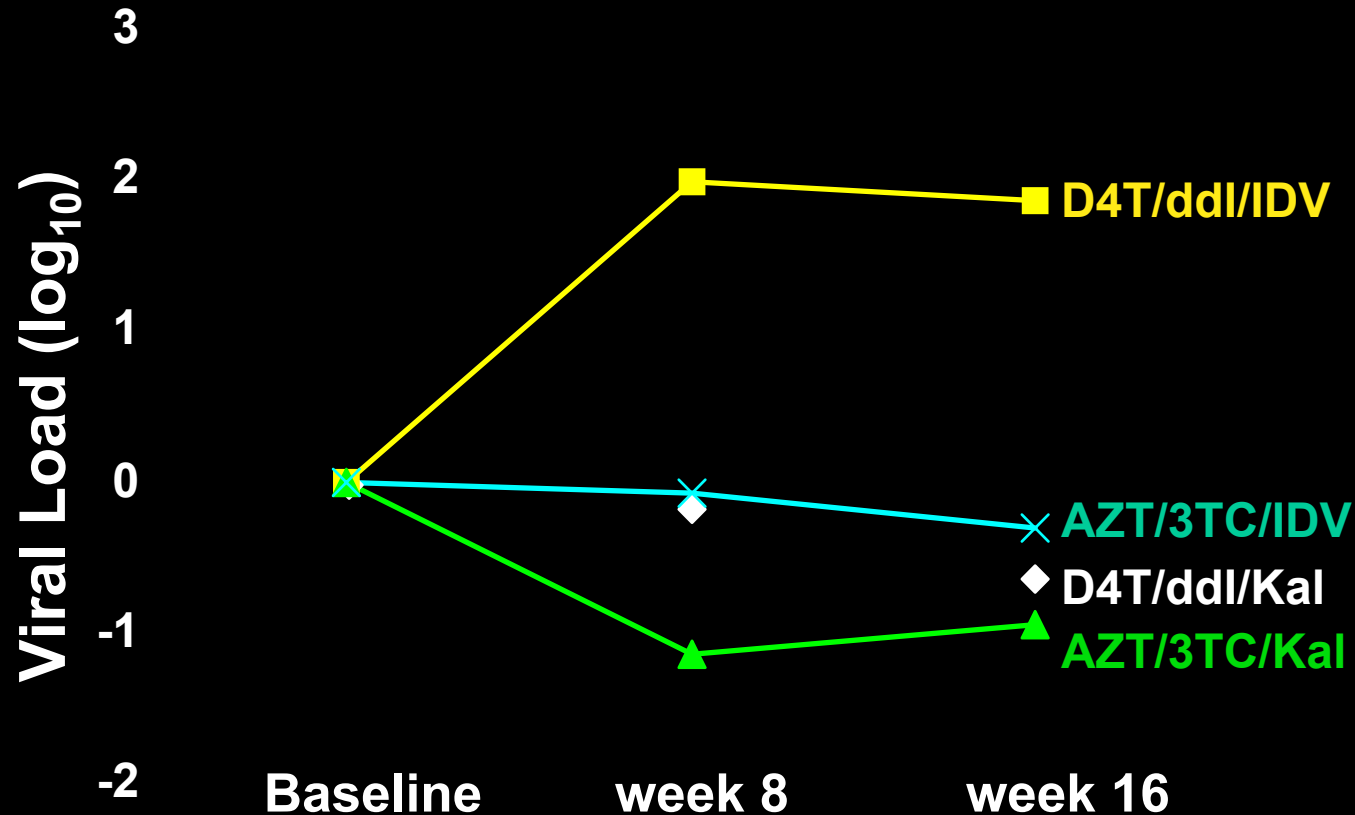
**75%±1.8%
correct
(~86/115)**

“In silico” Response Prediction

- **Baseline genotype for a virtual patient**
 - RT: 41, 67, 118, 210, 215
 - PI: 10, 46, 82, 90
- **Alternative therapy regimens**
 - a. D4T, ddl, Kaletra
 - b. D4T, ddl, indinavir
 - c. AZT, 3TC, Kaletra
 - d. AZT, 3TC, indinavir



“In silico” Response Prediction



Summary

- **The RDI is focused on establishing relationships between baseline genotype & 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 & aims to provide open access to query the database**
- **Utilization of large databases is likely to improve the accuracy of genotypic interpretation**

