Resolving the Problem of Multiple Reference Sequences in the UK HIV Drug Resistance Database

Esther Fearnhill, David Dunn (MRC Clinical Trials Un
for UK Collaborative Group on HIV Drug Resistance

Background (1)

- UK HIV Drug Resistance Database established in 2
  as central repository of resistance tests performed a
  part of routine clinical care (SQL Server)
- Nucleotide sequences now collected electronically
- When study started data were either
  - entered manually from copies of paper resistance
    report sent to clinicians
  - imported electronically (at amino acid level) from
    laboratories that had set-up local database

HIV resistance genotyping report

PHLS Antiviral Susceptibility Reference Unit, Birmingham Heartlands Hospital, Bordesley Green East, Birmingham B9 5ST, UK

Name - [redacted]

Our lab number: 02-111174

Number 1620313(D)

Request from: Churchill Hospital, Oxford

d.o.b. [redacted]

Clinician: [redacted]

Date of report:10.5.02

Your lab number/sample date: V13513; 17.4.02

Current treatment: RTV,SQV, 3TC, d4T

Previous treatment: AZT, ddI, NFV

N.B. Results from this test represent the majority viral population in plasma at the time of sample. Drug susceptibilities are estimated solely on these genotypic results with no account taken of the prior drug treatment history. It is essential that results are not interpreted in isolation from clinical information.

Protease (PR) codons analysed: 1-99

PR Resistance Mutations: K20M, M36I, M46LM, I54V, L63P, A71I, V82I, N88D, L90M, I93L

PR Other Mutations: I15V, L19V, L23I, E35G, R41K, D60E, Q61E, H69K, L89MV

Protease Inhibitors

APV	Intermediate resistance
IDV	Intermediate resistance
NFV	High-level resistance
RTV	Intermediate resistance
SQV	High-level resistance
LPV	Intermediate resistance

Reverse transcriptase codons analysed: 1-230

RT Resistance Mutations: M41L, D67N, K70R, V118I/V, V179I/V, M184V, L210W, T215Y, K219E

RT Other Mutations: V35T, T39E, V60IV, S68GS, K122E, D123DN, I135T, T139V, K173T, Q174K, T200A, E203EK, Q207E, R211K,

Background (2)

- Data variables entered included
  - laboratory conducting test
  - all reported amino acid changes (including mixture
  - range of codons sequenced in RT and PRO
  - whether “all” or “key” mutations only reported
- Database includes fields for reference sequence and sequencing system. This was completed through a batch update – based on information provided by laboratory

Data collected by March 2006

Type of test	Number
Nucleotide sequence	15,242
Amino acid changes (all)	3,056
Amino acid changes (key)	1,916
Total	**20,214**

Reference sequences (1)

- Amino acid sequences for different reference sequences were located from bioinformatic website
  - Consensus B (Stanford)
  - PNL4.3 (ABI)
  - LAV-1(VGI)
  - HXB2 (Virco and ABI)

Reference Sequence Differences

PR	3	37	57
Consensus B	I	N	R
PNL4.3	I	N	G/R
LAV-1	I	S	R
HXB2	V	S	R

RT	102	122	162	214	272	277	293	357	358	376	400
Consensus B	K	K	S	F	A	K	I	M	R	A	A
PNL4.3	Q	K	C	F	A	R	V	M	K	A	A
LAV-1	K	E	S	L	P	R	I	T	R	T	T
HXB2	K	E	S	L	P	R	I	M	R	T	T

Reference sequences (2)

- For some tests we were uncertain if we had inferred the correct reference sequence
- Ran a check based on the fact that an amino acid should be reported as a “change” or “mutation” only differs from the reference amino acid at that position

Reported mutations for test = 5240

Gene	codon	aa
PRO	3	V
PRO	13	V
PRO	35	D
PRO	37	S
PRO	57	R
PRO	63	P
PRO	64	V

Gene	codon	aa
RT	102	K
RT	122	E
RT	123	E
RT	184	V
RT	196	E
RT	211	K
RT	214	L
RT	245	E
RT	245	D
RT	249	R
RT	249	K
RT	272	P
RT	286	A
RT	293	I
RT	297	K

Limit to codons where reference sequences differ

Sequence	PRO			RT				
	3	37	57	102	122	214	272	29
5240	V	S	R	K	E	L	P	

Note: mixtures ignored for simplicity

Compare test sample with Consensus B

Sequence	PRO			RT				
	3	37	57	102	122	214	272	29
5240	V	S	R	K	E	L	P	
Consensus B	I	N	R	K	K	F	A	

Score = 3

Compare test sample with HXB2

Sequence	PRO			RT				
	3	37	57	102	122	214	272	29
5240	V	S	R	K	E	L	P	
HXB2	V	S	R	K	E	L	P	

Score = 8

Compare test sample with LAV-1

Sequence	PRO			RT				
	3	37	57	102	122	214	272	29
5240	V	S	R	K	E	L	P	
LAV-1	I	S	R	K	E	L	P	

Score = 7

Compare test sample with PNL4.3

Sequence	PRO			RT				
	3	37	57	102	122	214	272	29
5240	V	S	R	K	E	L	P	
PNL4.3	I	N	G/R	Q	K	F	A	V

Score = 0/1

Assignment Rule

- Can assign a resistance test to a reference sequenc unambiguously if score=0 and score for all other reference sequences is ≥ 1
- Otherwise resolve manually

Standardising reference sequence

- Having multiplicity of reference sequences can easi
  lead to analytical errors
- As nucleotide sequences are now processed throug
  Stanford program, natural to standardise database t
  Consensus B
- This was achieved through a STATA program (not a
  easy as it first seemed!)

Deriving full amino acid sequence

- Analytically easier to deal with observed amino acid sequence rather than file of mutations
- If no mutation reported impute the reference amino acid
- Excluded “key” mutation only tests from this progra
- Takes account of range of codons sequenced
- Can represent in “wide” or “long” format

Long format

- 1 record per amino acid per codon per gene per tes

TestID	Gene	Codon	aa
5240	1	1	P
5240	1	2	I
5240	1	3	S
5240	1	4	P
5240	1	4	F
5240	1	5	I
5240	1	6	E
.	.	.	.

Wide format

- 1 record per test

TestID	RT1	RT2	RT3	RT4	RT5	.	RT400	PR1	PR2
5240	P	I	S	X	I	.	~	P	Q
.	.	.	.	.	.	.	.	.	.

Acknowledgements

UK Collaborative Group on HIV Drug Resistance Steering Committee
David Dunn, Esther Fearnhill, Hannah Green, Kholoud Porter, (MRC Clinical Tr
Unit), Rob Gifford, Paul Kellam, Deenan Pillay, Andrew Phillips, Caroline Sabin
Free and University College Medical School), Sheila Burns (City Hospital, Edinb
Sheila Cameron (Gartnavel General Hospital, Glasgow), Pat Cane (Health Prot
Agency), Ian Chrystie (St. Thomas' Hospital, London), Duncan Churchill (Brigh
and Sussex University Hospitals NHS Trust), Valerie Delpeche (Health Protect
Agency, London), Philippa Easterbrook, Mark Zuckerman (King's College Hosp
London), Anna Maria Geretti (Royal Free NHS Trust, London), David Goldberg
(Scottish Centre For Infection and Environmental Health), Mark Gompels (Soutl
Hospital, Bristol), Tony Hale (PHLS, Leeds), Steve Kaye (St. Marys Hospital,
London), Andrew Leigh-Brown (University of Edinburgh), Anton Pozniak (Chels
Westminster Hospital, London), Gerry Robb (Department of Health, London), R
Smit (Health Protection Agency, Birmingham Heartlands Hospital), Peter Tilstor
(Manchester Royal Infirmary), Ian Williams (Mortimer Market Centre, RFUCMS)

www.hivrdb.org.uk