CoRSeq$_{V3}$: Novel Coreceptor Usage Prediction Algorithms for HIV-1 Subtypes B, C, D and CRF01_AE

Kieran Cashin[1,2*], Lachlan Gray[1,3], Katherine Harvey[1,2], Danielle Perez-Bercoff[4], Guinevere Q. Lee[5], Jasminka Sterjovski[1], Michael Roche[1], James F. Demarest[6], Fraser Drummond[7], P. Richard Harrigan[5], Melissa J. Churchill[1,3], and Paul R. Gorry[1,2,3] ***email: kieran@burnet.edu.au**

[1]Burnet Institute, Melbourne, Australia. [2]University of Melbourne, Melbourne, Australia. [3]Monash University, Melbourne, Australia. [4]Laboratory of Retrovirology, Centre Recherche Public de la Santé, Luxembourg. [5]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada. [6]ViiV Healthcare.

BACKGROUND

Entry into cells by HIV-1 is mediated viral envelope proteins (Env) binding CD4 and either CCR5 or CXCR4 on the host cell surface.

The only licensed HIV-1 entry inhibitor, maraviroc (MVC), blocks access to CCR5.

MVC is ineffective against CXCR4-using viruses and thus has the potential to select for these viruses and their associated rapid disease progression.

Hence a "tropism test" of a patient's circulating virus population is essential prior to administration of MVC.

The gold standard tropism test is Trofile®, a cell entry assay where virions pseudotyped with patient-derived Envs are exposed to target cells expressing either CCR5 or CXCR4.

Yet the cost, duration and specialised nature of Trofile® remains a major barrier to MVC accessibility, particularly within the developing world.

In silico programs that predict coreceptor specificity based on sequence alterations in the third variable (V3) loop of Env have been developed as rapid and inexpensive alternatives to Trofile®.

The currently available genotypic algorithms accurately predict coreceptor usage for subtype B HIV-1 (B-HIV), against which they were designed, but have limited success against non-B HIV-1 subtypes such as;

- Subtype C (C-HIV): >50% of worldwide infections and >95% of infections in southern Africa and central Asia.
- CRF01_AE (AE-HIV): endemic in Thailand, Indonesia and Vietnam.

AIMS

To develop a suite of highly accurate subtype-specific V3 sequence-based algorithms for B-HIV, C-HIV, D-HIV and AE-HIV.

To package as a free, user-friendly online platform (CoRSeq$_{V3}$).

METHODOLOGY

Every phenotypically characterised CCR5-specific (R5) and CXCR4-using (X4/R5X4) patient-derived B-HIV, C-HIV, D-HIV and AE-HIV V3 sequence was acquired from the Los Alamos HIV Database and our recent published studies.

One sequence per patient was randomly selected for the development of respective algorithms.

Sequences were randomly assigned to DESIGN or TEST sets for the development and validation of CoRSeq$_{V3}$ respectively.

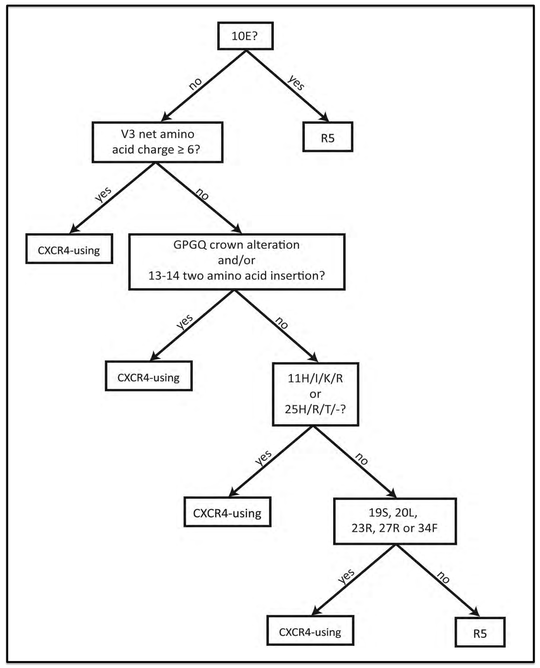
DESIGN PHASE

V3 sequence analysis was performed to highlight differences between CXCR4-using and R5 Envs. Below is the analysis of C-HIV DESIGN sequences. Identical analyses were performed for B-, D- and AE-HIV.

Distinguishing V3 characteristics include;

- Specific amino acid mutations (A)
- Net amino acid length (B)
- Net amino acid charge (C)
- Number of glycosylation sites (D)

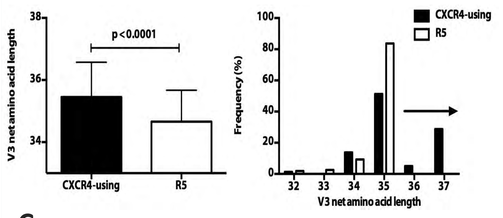
DESIGN V3 sequence analysis results informed CoRSeq$_{V3}$ prediction criteria. Below is the C-HIV CoRSeq$_{V3}$ decision tree.



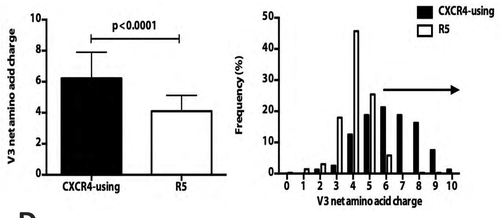
A

Mutation	Phenotype Frequency		Fisher's exact test (p value)	Mutation	Phenotype Frequency		Fisher's exact test (p value)
	CXCR4-using (%)	R5 (%)			CXCR4-using (%)	R5 (%)	
GPGQ alt	62.5	3.5	<0.0001	Phe20Val	7.5	0	0.0068
Asn6Tyr	5	0	ns	Val21His	12.5	0.9	0.0013
Asn7Lys	5	0	ns	Ile23Arg	6.25	0	0.0289
Thr8Ile	7.5	0.2	0.0068	Ile23Asn	11.25	2.6	0.0489
Thr8Lys	7.5	0.9	0.0349	Gly24Asp	11.25	2.6	0.0489
Arg9Ile	6.25	0	0.0289	Gly24Lys	11.25	1.6	0.0184
Lys10-	12.5	1.4	0.0013	Lys25Ala	20	8.4	0.0237
Lys10Glu	0	5.1	ns	Lys25Glu	10	21	0.0494
Arg11Arg	23.75	0.9	<0.0001	Lys25His	5	0	ns
Arg11Pro	5	0	ns	Lys25Lys	12.5	0.7	0.0013
13-14 insert	35	0	<0.0001	Lys25Asn	5	0	ns
Arg13Gly	11.25	2.1	0.0184	Lys25Arg	6.25	0.2	0.0289
Arg13Ser	12.5	0.7	0.0013	Lys25Thr	5	0	ns
Pro16Arg	25	0	<0.0001	Lys25-	12.5	0.7	0.0013
Gly17Arg	7.5	0.5	0.0349	Ile26Val	12.5	3.3	0.0165
Arg18Arg	23.75	1.2	<0.0001	-27Arg	6.25	0	0.0289
Arg18His	10	0.5	0.0050	Gln32Glu	1.25	7.9	0.0349
Ala19Ser	8.75	1.4	0.0185	His34Phe	5	0	ns

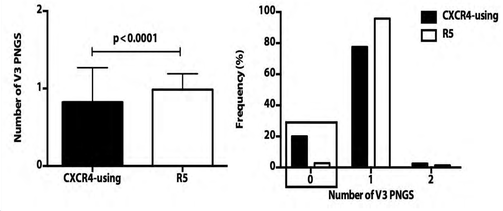
B



C



D



TEST PHASE

Performance of CoRSeq$_{V3}$ and currently available algorithms against TEST sequences.

Coreceptor usage prediction technique	B-HIV (185 CXCR4-using and 526 R5)		C-HIV (80 CXCR4-using and 429 R5)		D-HIV (57 CXCR4-using and 80 R5)		AE-HIV (50 CXCR4-using and 126 R5)	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
CoRSeq$_{V3-B}$	**78.9**	**81.9**	-	-	-	-	-	-
CoRSeq$_{V3-C}$	-	-	**91.3**	**90.2**	-	-	-	-
CoRSeq$_{V3-D}$	-	-	-	-	**89.5**	**83**	-	-
CoRSeq$_{V3-AE}$	-	-	-	-	-	-	**92**	**92.1**
g2p 1% FPR	34.6	99.3	52.5	99.8	49.1	97.5	60	99.2
g2p 2.5% FPR	58.4	97.2	67.5	99.8	70.2	87.5	78	92.9
g2p 5% FPR	68.1	93.7	72.5	97.2	86	75	82	82.5
g2p 10% FPR	74.1	85.3	80	93.7	89.5	53.8	88	57.1
g2p 15% FPR	74.6	80.1	83.8	91.4	89.5	52.5	88	51.6
g2p 20% FPR	77.8	69.8	85	85.8	89.5	48.8	90	40.5
WebPSSM$_{X4R5}$-B	54.6	96.1	67.5	95.1	89.5	55	78	80.2
WebPSSM$_{SINSI}$-B	61.6	94.2	56.3	98.6	86	67.5	80	82.5
WebPSSM$_{SINSI}$-C	-	-	81.3	89	-	-	-	-
11/25	51.9	93.7	81.3	89	78.9	65	42	99.2
11/24/25	54.6	93.1	47.5	98.4	80.7	63.8	44	99.2
HIVcoPRED (SAAC)	68.6	94.6	55	98.4	82.5	86.3	88	86.5
HIVcoPRED (SAAC+BLAST)	70.3	94.4	73.8	95.1	82.5	86.3	90	86.5
dsKernel	42.2	99.8	73.8	95.1	82.5	66.3	92	75.4

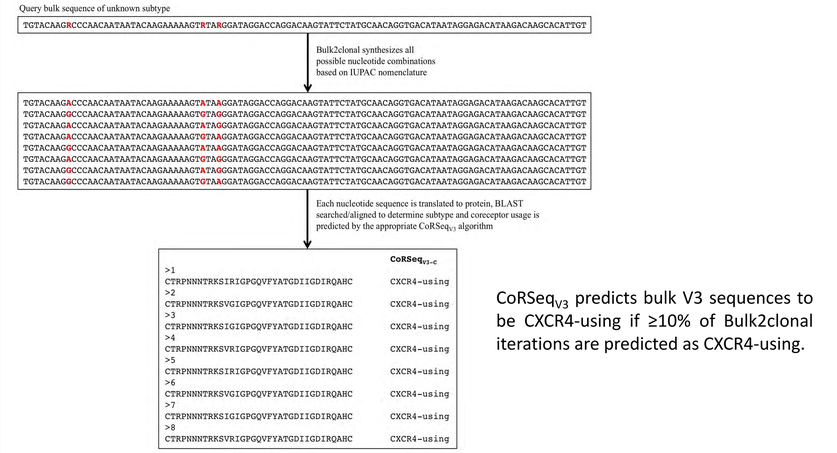
Sensitivity = no. of correctly predicted CXCR4-using sequences divided by no. of true CXCR4-using sequences
Specificity = no. of correctly predicted R5 sequences divided by no. of true R5 sequences

CoRSeq$_{V3}$ algorithms have the most favorable sensitivity and specificity profiles against TEST sequences.

CLINICAL VALIDATION

CoRSeq$_{V3}$ was developed against *clonal* V3 sequences whereas patient genotypic tests are typically performed using *bulk* V3 sequences.

To allow for CoRSeq$_{V3}$ bulk sequence predictions and thus clinical validation we developed Bulk2clonal.



CoRSeq$_{V3}$ predicts bulk V3 sequences to be CXCR4-using if ≥10% of Bulk2clonal iterations are predicted as CXCR4-using.

Patient-derived bulk V3 sequences were used to clinically validate CoRSeq$_{V3}$+Bulk2clonal.

A4001064 epidemiology study (Pfizer)
95 C-HIV and 87 D-HIV samples were phenotyped by Trofile® to test the efficacy of Trofile® against non-B-HIV.

MERIT maraviroc clinical trial (Pfizer)
21 C-HIV patient originally phenotyped by Trofile® as R5 failed MVC therapy. Re-testing of pre-treatment samples by the enhanced sensitivity Trofile® assay (ESTA) revealed low-frequency CXCR4-using viruses.

Performance of CoRSeq$_{V3}$ against A4001064 and MERIT patient sequences.

Coreceptor usage prediction technique		Epidemiology study A4001064 *Phenotyped by original Trofile assay*				MERIT *Phenotyped by ESTA*
		C-HIV 55 CXCR4-using and 40 R5		D-HIV 43 CXCR4-using and 44 R5		C-HIV 21 CXCR4-using
		Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)
CoRSeq$_{V3-C}$		83.6	92.5	-	-	76.2
CoRSeq$_{V3-D}$		-	-	80.5	73	-
g2p$_{[Narrowpool]}$ FRP	1%	32.7	97.5	12.2	95.7	0
	2.5%	56.4	97.5	48.8	84.8	4.8
	5%	75.6	95	78	69.6	9.5
	10%	81.8	95	87.8	58.7	47.6
	15%	83.6	90	90.2	56.5	57.1
	20%	85.5 [2]	85 [2]	90.2	52.2	57.1
g2p$_{[ALL]}$ FRP	1%	30.9 [2]	95 [2]	12.2	95.7	0
	2.5%	60 [2]	95 [2]	43.9	84.8	4.8
	3.75%	65.5 [2]	95 [2]	58.5	76.1	4.8
	5%	76.4 [2]	92.5 [2]	80.5	69.6	4.8
	10%	81.8 [2]	85 [2]	85.4	58.7	47.6
	15%	83.6 [2]	82.5 [2]	90.2	56.5	52.4
	20%	87.3 [2]	82.5 [2]	90.2	52.2	52.4

A4001064
CoRSeq$_{V3}$+Bulk2clonal has the most favorable sensitivity and specificity profile.

MERIT
CoRSeq$_{V3}$ was most sensitive at detecting low-frequency CXCR4-using viruses and would have forecast MVC therapy failure.

***N.B.** Specificities against MERIT could not be determined as all bulk V3 sequences were CXCR4-using.*

FINAL REMARKS

Extensive clinical validation of CoRSeq$_{V3}$ against MVC treatment outcomes of patients enrolled in other Pfizer trials such as MOTIVATE is ongoing.

CoRSeq$_{V3}$ may enhance access to MVC and future CCR5 inhibitors, particularly in regions where non-B-HIV predominates such as southern Africa, central Asia and southeast Asia where the HIV pandemic is worst.

C-HIV CoRSeq$_{V3}$ is available at www.burnet.edu.au/coreceptor. B-, D- and AE-HIV will be available soon.

Find C-HIV CoRSeq$_{V3}$ at: **www.burnet.edu.au/coreceptor**

We would like to thank ViiV Healthcare Australia for their support and collaboration throughout this study, specifically for supplying Pfizer trial V3 sequences and honoraria to attend and present at CROI 2014.