www.geno2pheno.org

Inferring viral tropism from genotype with massively parallel sequencing: Qualitative and quantitative analysis

Martin P. Däumer[1], Rolf Kaiser[2], Rolf Klein[1], Thomas Lengauer[3], Bernhard Thiele[1], Alexander Thielen[3]

1 Institute of Immunology and Genetics, Kaiserslautern, Germany; 2 Institute of Virology, University of Cologne, Germany; 3 Max Planck Institute for Informatics, Saarbrücken, Germany

(1) Background

Genotypic prediction of HIV-1 tropism is an inexpensive and fast alternative to phenotypic approaches. However, using standard sequencing approaches, these advantages are accompanied by lower sensitivities of X4-detection, especially in clinical isolates.
Since prediction from genotype performs well on clonal data, it is commonly thought that "false" predictions are mainly attributed to the lack of detection of minor variants. In comparison to standard bulk-sequencing, the 454-technology detects single clones, thereby being much more sensitive.
The aim of this study was to address the question, whether this method can be successfully combined with bioinformatic approaches to generate a qualitative and quantitative prediction of coreceptor usage from V3-genotype.

(2) Methods

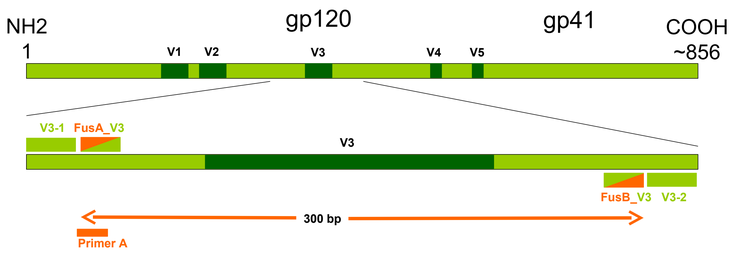
Patients

Plasma samples from 55 antiretroviral-treated patients with tropism documented by Monogram's **tro**file® assay were sequenced with standard population based approaches and their tropism predicted with geno2pheno$_{[coreceptor]}$ (default false-positive-rate of 10%). From these, 14 samples (7 R5, 7 X4) were selected for further analysis with massively parallel sequencing (454 Life Sciences/ROCHE).

Massively parallel sequencing (MPS)

300 bp amplicons including the entire V3 region of the *env* gene were generated by nested PCR (Figure 1). Amplicons were sequenced using the amplicon sequencing protocol with 250 base pair (bp) average read length to a depth of approximately 10,000 reads/V3-sequences per patient on Genome Sequencer FLX (454 Life Sciences/ROCHE).

Figure 1: Strategy for amplification and massively parallel sequencing of V3



(2) Methods cont.

Prediction of coreceptor usage from MPS data

MPS-generated sequences were directly extracted from the "sff"-files, analysed and processed for full-length V3. Viral tropism was predicted with the internal batch-version of geno2pheno$_{[coreceptor]}$. The prediction-score (not shown in the web-version) was used as a measure of coreceptor-affinity. For quantitative analysis of the R5/X4-distribution, the prediction-score of each variant containing the V3-loop was plotted against its frequency within the viral population (Figure 2).

(3) Results

Prediction of coreceptor usage from the 55 bulk-sequences showed a sensitivity of 59.1% and a specificity of 90.9% (data not shown). Among the samples used in ultra-deep analysis, one R5 and four X4 were falsely predicted (Table 1).

Using the 454-technology, 10,000 reads per V3-containing amplicon were generated on average. Minorities of sequences with high confidence in CXCR4-usage were found in all samples, irrespective of phenotypic determined tropism (Figure 2). For comparison with the Trofile-results, a minority-cutoff of 5%, reflecting the proposed sensitivity of the standard Trofile-assay, was applied. Using the default false-positive-rate of geno2pheno$_{[coreceptor]}$ results were concordant except for one sample (Table 1).

Table 1: Correlation between phenotype, prediction from bulk-sequence and MPS data

Patient	*phenotype* (trofile®)	*Bulk sequence, predicted as X4 at FPR* of:*	*MPS, prop. predicted as X4 at FPR* of 10%*
1	D/M	2.5%	85.3%
2	D/M	5%	86.7%
3	D/M	10%	14.2%
4	D/M	15%	5.4%
5	D/M	15%	48.0%
6	D/M	---	66.1%
7	D/M	---	1.0%
8	R5	10%	1.4%
9	R5	---	3.4%
10	R5	---	0.3%
11	R5	---	0.5%
12	R5	---	1.6%
13	R5	---	1.8%
14	R5	---	2.3%

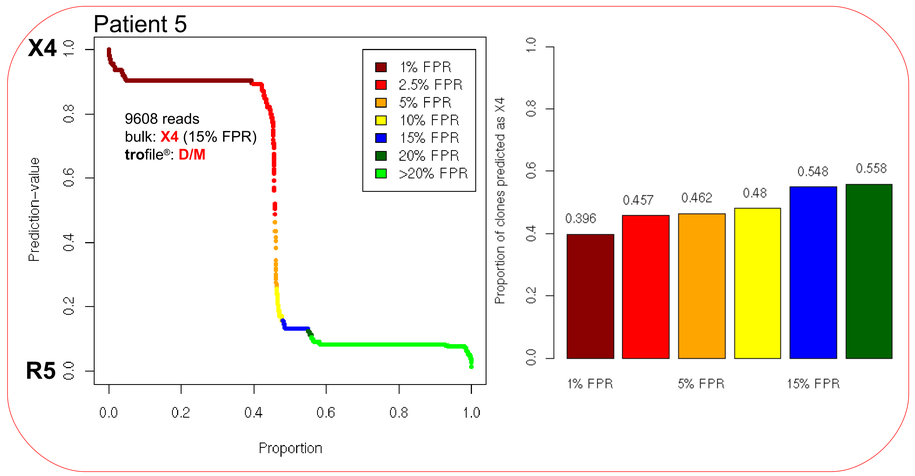
**FPR: false positive rate, --- = FPR > 20%*

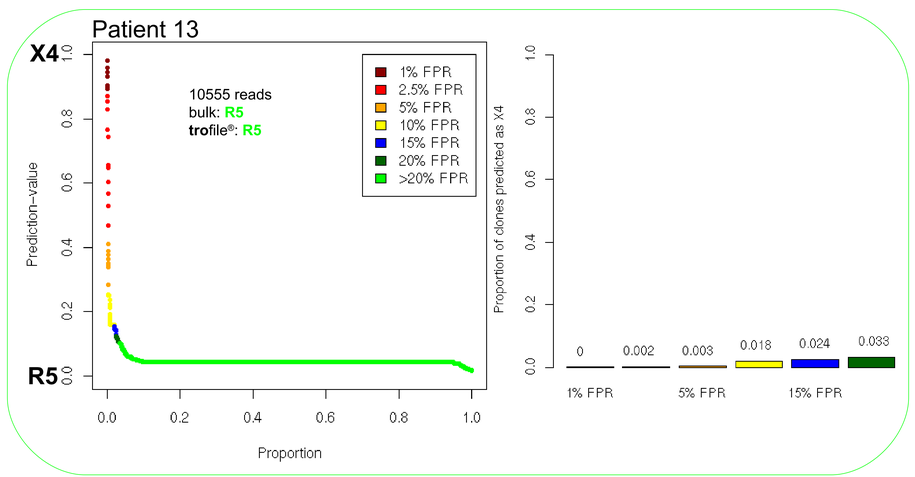
References

- Margulies M et al. Nature 2005;437:376-380.
- Sing T et al. Antivir Ther. 2007;12:1097-106.
- Low AJ et al. AIDS. 2007 Sep 12;21(14):F17-24..

(3) Results cont.

Figure 2: Proportion of predicted X4-viruses at different FPR in massively parallel sequenced V3 amplicons. Red frame: **tro**file® D/M-typed; green frame: **tro**file® R5-typed





Conclusions

Combining genotypic prediction with ultra-deep sequencing results in a fast and accurate alternative to phenotypic assays. The detection of X4-viruses in all isolates suggests that coreceptor usage as well as fitness of minorities is important for therapy outcome. The high sensitivity of this technology in combination with a quantitative description of the viral population may allow implementing meaningful cutoffs for predicting response to CCR5-antagonists in the presence of X4-minorities.