Resolving HCV subtypes in a Belgian cohort by full genome next-generation sequencing

0561

Kasper T. Christensen[1], Florian Pierard[1], Kurt Beuselinck[2], David Bonsall[3], Rory Bowden[4], Katrien Lagrou[1,2], Frederik Nevens[1,2], Peter Simmonds[4], Anne-Mieke Vandamme[1], Eric Van Wijngaerden[1,2], Lize Cuypers[1,2], Kristel Van Laethem[1,2]

[1]KU Leuven, Leuven, Belgium, [2]University Hospitals Leuven, Leuven, Belgium, [3]Big Data Institute, University of Oxford, Oxford, UK, [4]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

BACKGROUND

In order to reach WHOs goal of HCV elimination it has been suggested to prioritize populations that are actively propagating infection, given e.g. several outbreaks of HCV among HIV-infected men who have sex with men in the last decade. Genotyping the infecting virus is part of routine care to guide antiviral treatment, but commercial assays have been shown to occasionally report inaccurate results. In this study, we use next-generation sequencing (NGS) to increase the discriminatory power and evaluate the accuracy of genotyping in routine HCV care.

METHODS

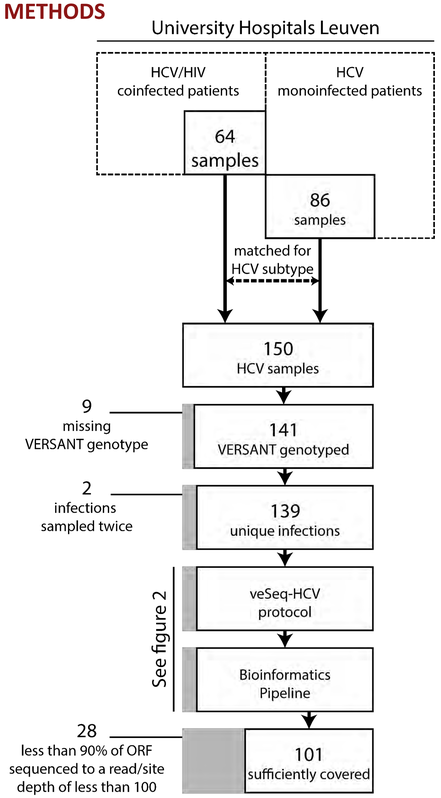


Figure 1: Sample selection, inclusion criteria and study flow. Upon final analysis, 101 samples passed all inclusion criteria and were examined for concordance between VERSANT and NGS.

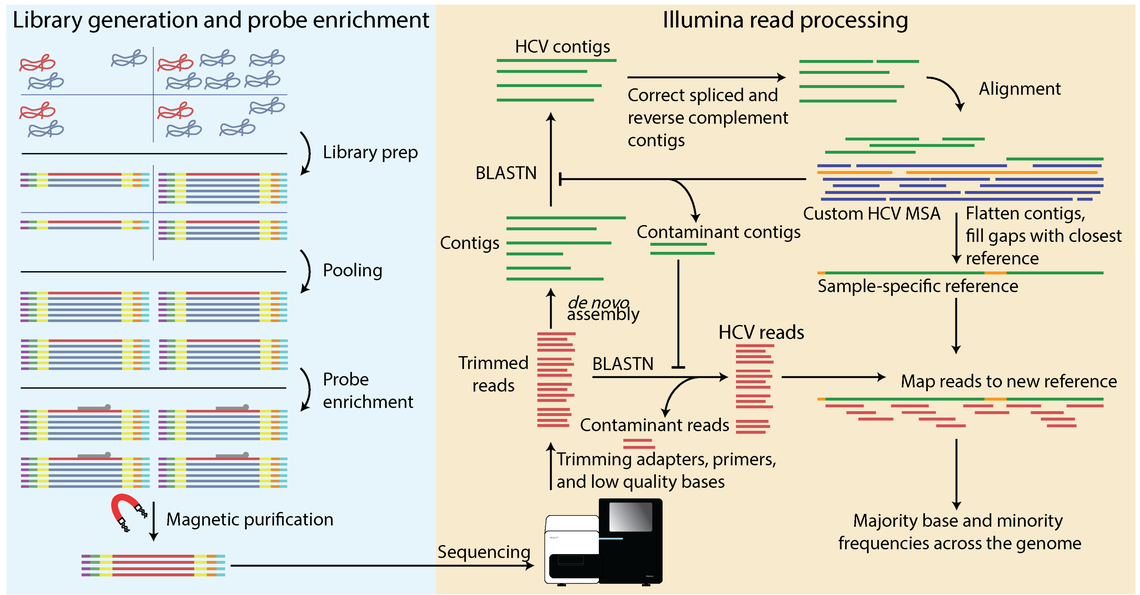


Figure 2: Overview of the veSeq-HCV protocol (Bonsall D. *et al*, 2015) and an *in-house* bioinformatics pipeline for HCV full NGS (Pierard F., poster presentation at VEME workshop, 2019, Wymant C. *et al*, 2018) that was used to obtain 101 HCV full genome sequences. NGS subtype classification was based on the coverage and similarity of each contig to known reference strains as well as read depth of coverage for separate analyses using a sample-specific reference generated with contigs of each subtype.

RESULTS

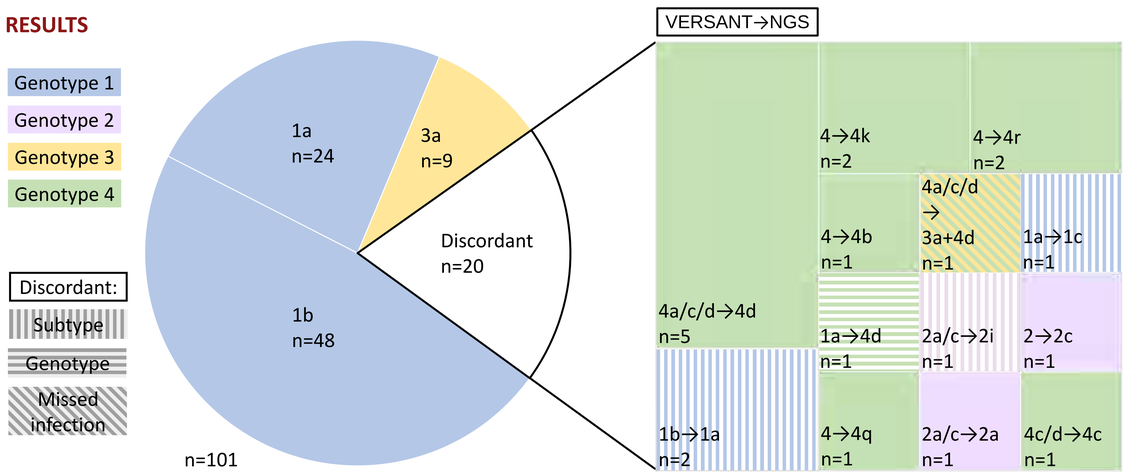


Figure 3: HCV subtype distribution in 101 samples for which >90% of HCV ORF was covered to a depth >100. The predominant genotype in our study was 1, followed by 4, 3 and 2. Distinct subtype classifications were mainly found for genotype 2 and 4. 81 samples, all either 1a, 1b or 3a, were called similarly by VERSANT and NGS. Of the remaining 20 samples, 14 samples consisted of refinements of an ambiguous classification of VERSANT into one of its constituent subtypes by NGS. Four samples had a different subtype called by NGS (vertical stripes), one sample had a different genotype called by NGS (horizontal stripes), and one sample had a dual infection of which only one was identified by VERSANT (diagonal, see Figure 4 for details).

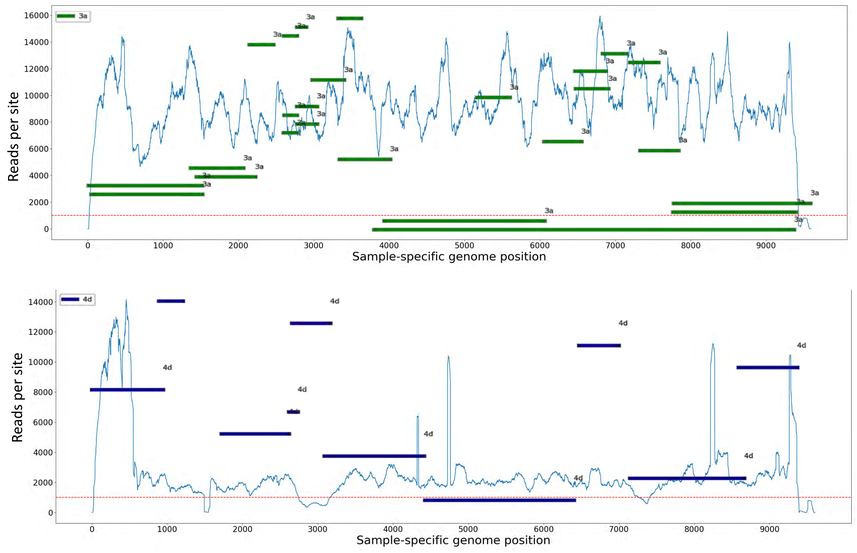


Figure 4: Evidence of a sample with a dual infection of subtypes 3a and 4d, that was missed by VERSANT (classified as 4a/c/d). Read and contigs for one sample were aligned to two different sample-specific references generated with the 3a (green – top) and 4d (blue – bottom) contigs. The red lines correspond to a coverage of 1,000.

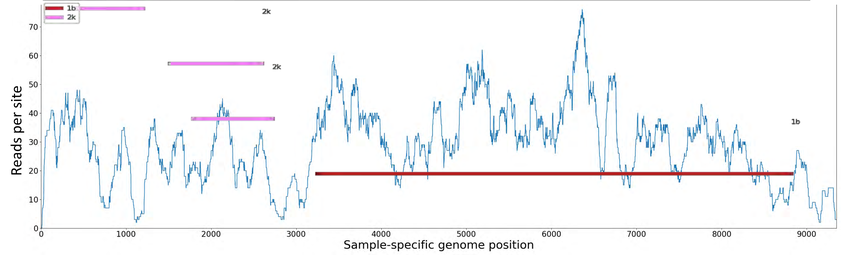


Figure 5: Evidence for a recombinant strain (2k/1b), that was missed by VERSANT. Read and contigs were aligned to a sample-specific reference with all generated contigs. Despite not meeting the coverage threshold to be included in the final 101 samples, a clear indication of an infection with a 2k/1b recombinant is observed. This sample was classified by VERSANT simply as a subtype 2a/c.

CONCLUSIONS

While the applied sequencing strategy requires further optimization to reliably classify all geno- and subtypes across a broad viral load range, a good overall concordance was found with the sub- and genotypes determined by VERSANT. The higher resolution of NGS proves capable of resolving specific subtypes and detecting cases of potential mixed infections and (inter-genotype) recombinants.

ADDITIONAL KEY INFORMATION

This work was funded in part by grants from the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO, G.0B23.17N and 1S38819N).

Contact author: *kasperthorhauge.christensen@kuleuven.be*