BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Algorithms for Bioinformatics
Title	RAAVioli: a bioinformatic tool for the characterization of AAV integration and recombination processes
All Authors	Cipriani C(1,2), Spinozzi G(2), Rudilosso L(2), Dalvadi D(3), Grompe M(4), Masseroli M (1), Montini E(2), Calabria A(2), Cesana D (2)

Affiliation

(1) Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
(2) San Raffaele Telethon Institute for Gene Therapy (Sr-Tiget), IRCCS San Raffaele Scientific Institute, Milan, Italy, Milan, Italy
(3) OHSU, Portland, OR,
(4) Yecuris Corporation, Tualatin, OR

Motivation

Recombinant Adeno Associated Viral (rAAV)-based gene therapy (GT) applications have been successfully exploited for the treatment of several disorders. rAAV genome mainly remains episomal in the nucleus of transduced cells; however, numerous studies demonstrated the integration of fragmented, rearranged, or full-length AAV DNA within the transduced cell genome, where double-strand DNA breaks or nicks have occurred. Furthermore, preclinical studies revealed the occurrence of hepatocellular carcinoma and clonal expansion events consequent to rAAV insertions, posing safety concerns for their clinical use.
Although several pipelines studying vector integration sites (IS) exist in the literature, none of them takes into consideration the presence of vector rearrangements (recombination of the viral DNA with other copies of itself), thus leading to a misidentification of the integration locus.
For this reason, we decided to build a pipeline, named RAAVioli (Recombinant Adeno-Associated Viral IntegratiOn anaLysIs), able not only to correctly identify IS but also to characterize vector rearrangements, using both long and short reads.

Methods

RAAVioli is composed of three major steps: 1) quality filtering; 2) alignment; 3) identification of IS and characterization of vector rearrangements.
After initial quality analysis and filtering steps (performed with FastqC and MultiQC), only reads containing AAV sequences are aligned (using bwa-mem) on a hybrid genome composed of both the AAV and target genome (such as human or mouse).
Next, primary and supplementary alignments are collected and analyzed with custom Python and R scripts to parse the corresponding CIGAR (Concise Idiosyncratic Gapped Alignment Report) strings in order to reassemble the read composition. Indeed, by looking at the soft and hard clipped portion of the CIGAR string, each aligned fragment is ordered within the read to precisely identify the integration locus on the target genome and characterize vector rearrangements. Reads having the same AAV/target genome breakpoint and the same number and type of AAV rearrangements are considered identical IS. The closest gene within the target region and surrounding the IS is identified with BedTools. The abundance of each IS is computed by counting, for the same vector/cell genome junction, the number of different DNA fragments that varied based on the shear site location and on the cell genome of the starting cell population.

Results

We collected data from a humanized liver mouse model, where human primary hepatocytes have been transduced ex-vivo or in-vivo with a tomato expressing AAV. PCR amplicons or DNA fragments containing AAV vector portions were sequenced by both short paired-end and long reads, and then analyzed by RAAVioli to characterize vector rearrangements and IS.
We identified 811 and 370 IS from short paired-end Illumina reads and long PacBio reads, respectively. The distribution of AAV IS was sparse in the human genome, similarly in both datasets, and ALB was the most targeted gene as described in the literature. Furthermore, 32 ISs were found in both datasets, demonstrating the consistency of the results independently from the sequencing platform adopted. Additionally, more than 25% of the identified IS in each dataset contained vector rearrangements, thus confirming the need for an approach able to combine both rearrangement characterization and IS identification.
Precision and accuracy of RAAVioli, in the identification of the exact integration locus and characterization of vector rearrangements (as start and end position), were assessed by producing in silico datasets that recapitulate the lengths and the number of rearrangements observed in the experimental dataset. Overall, scores higher than 0.95 were obtained for both long and short reads.
These results show that RAAVioli is a comprehensive and flexible tool that supports the characterization of AAV integration and recombination processes utilizing both long and short paired ends sequencing reads. These methods are essential in AAV-based gene therapy and gene editing applications since they enable the assessment of safety in AAV studies.

Info

-

filename	-

Figure	
-	
Availability	-

Corresponding Author

Name, Surname	Carlo, Cipriani
Email	carlo.cipriani@polimi.it
Submitted on	28.04.2023