Genome Medicine

METHOD Open Access

HLA typing from RNA-Seq sequence reads

Sebastian Boegel[1,2], Martin Löwer[1], Michael Schäfer[1,2], Thomas Bukur[1], Jos de Graaf[1], Valesca Boisguérin[1], Özlem Türeci[3], Mustafa Diken[1], John C Castle[1*] and Ugur Sahin[1,2*]

Abstract

We present a method, seq2HLA, for obtaining an individual's human leukocyte antigen (HLA) class I and II type and expression using standard next generation sequencing RNA-Seq data. RNA-Seq reads are mapped against a reference database of HLA alleles, and HLA type, confidence score and locus-specific expression level are determined. We successfully applied seq2HLA to 50 individuals included in the HapMap project, yielding 100% specificity and 94% sensitivity at a *P*-value of 0.1 for two-digit HLA types. We determined HLA type and expression for previously un-typed Illumina Body Map tissues and a cohort of Korean patients with lung cancer. Because the algorithm uses standard RNA-Seq reads and requires no change to laboratory protocols, it can be used for both existing datasets and future studies, thus adding a new dimension for HLA typing and biomarker studies.

Background

The major histocompatibility complex (MHC) molecules display peptide antigens that are derived from intracellular (class I) and extracellular (class II) proteins on the surface of vertebrate nucleated cells. The human MHC, called the human leukocyte antigen (HLA), is highly polymorphic and comprises three major gene loci for class I (A, B, C) (Figure 1) and three major gene loci for class II (DP, DQ, DR), which are expressed co-dominantly. Each cell expresses three maternal and three paternal HLA class I and three maternal and three paternal class II alpha and beta alleles. Determining the sequence of these molecules, HLA typing, is essential for clinical work (for example, organ transplantation), immune system research, and biomarker and drug development. Current HLA typing techniques use labor- and time-intensive methods, such as sequence-specific oligonucleotide probe (SSOP) hybridization [1], PCR amplification with sequence-specific primers [2], Sanger sequencing [3] and sero-typing [4].

Next generation sequencing (NGS) is a novel platform that enables rapid generation of billions of short nucleic acid sequence reads. Several studies described the use of NGS in high-throughput HLA genotyping using genomic DNA (for examples, see [5,6]). Recently, Lank *et al.* described a method using RNA for high-throughput MHC class I genotyping [7,8], which was applied to assess genotype- and allele-specific expression of MHC class I in human and macaque leukocyte subsets [9]. This method involves the reverse transcription of RNA into cDNA, amplification using highly specific MHC I primers and subsequent bi-directional cDNA amplicon sequencing using Roche/454 GS FLX pyrosequencing, and is able to unambiguously resolve MHC alleles with high accuracy. However, all of these techniques use specialized NGS protocols including primer design to amplify only MHC class I alleles and amplicon sequencing with long reads (≥150 nucleotides) using Roche/454 GS FLX or Illumina GAIIx.

By contrast, gene expression profiling in patient samples using 'whole transcriptome' sequencing (RNA-Seq profiling) typically uses much shorter reads. The adoption of the RNA-Seq platform has been rapid: clinical and research laboratories worldwide have deposited over 14,600 'RNA-Seq' sample profiles into public repositories such as the National Center for Biotechnology Information Sequence Read Archive, including 4,304 human RNA-Seq samples as of 8 October 2012. As opposed to previous methods for determining gene expression, the RNA-Seq platform not only generates expression profiles but the data also contain nucleotide sequence information. Given the large number of RNA-Seq profiles in the public domain and our efforts to develop individualized T cell-mediated cancer vaccines, which require the knowledge of a patient HLA type and HLA expression to prioritize target epitopes [10-12], we sought to develop an algorithm to

* Correspondence: john.castle@tron-mainz.de; sahin@uni-mainz.de
[1]TRON - Translational Oncology at the University Medical Center of the Johannes Gutenberg University, Langenbeckstrasse 1, Building 708, 55131 Mainz, Germany
Full list of author information is available at the end of the article