EBioMedicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Transcriptomics and Targeted Proteomics Analysis to Gain Insights Into the Immune-control Mechanisms of HIV-1 Infected Elite Controllers

Wang Zhang [a,b,1], Anoop T. Ambikan [a,1], Maike Sperk [a,c,1], Robert van Domselaar [d], Piotr Nowak [a,d], Kajsa Noyan [a], Aman Russom [b], Anders Sönnerborg [a,d], Ujjwal Neogi [a,b,*]

[a] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden
[b] Science for Life Laboratory, Division of Proteomics and Nanobiotechnology, KTH Royal Institute of Technology, Solna, Stockholm, Sweden
[c] Faculty of Medicine, University of Tuebingen, Tuebingen, Germany
[d] Department of Medicine Huddinge, Unit of Infectious Diseases, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 26 October 2017
Received in revised form 22 November 2017
Accepted 30 November 2017
Available online xxxx

Keywords:
HIV-1 Elite Controllers
Transcriptome
Proteome

ABSTRACT

A small subset of HIV-1 infected individuals, the "Elite Controllers" (EC), can control viral replication and restrain progression to immunodeficiency without antiretroviral therapy (ART). In this study, a cross-sectional transcriptomics and targeted proteomics analysis were performed in a well-defined Swedish cohort of untreated EC (n = 19), treatment naïve patients with viremia (VP, n = 32) and HIV-1-negative healthy controls (HC, n = 23). The blood transcriptome identified 151 protein-coding genes that were differentially expressed (DE) in VP compared to EC. Genes like CXCR6 and SIGLEC1 were downregulated in EC compared to VP. A definite distinction in gene expression between males and females among all patient-groups were observed. The gene expression profile between female EC and the healthy females was similar but did differ between male EC and healthy males. At targeted proteomics analysis, 90% (29/32) of VPs clustered together while EC and HC clustered separately from VP. Among the soluble factors, 33 were distinctive to be statistically significant (False discovery rate = 0.02). Cell surface receptor signaling pathway, programmed cell death, response to cytokine and cytokine-mediated signaling seem to synergistically play an essential role in HIV-1 control in EC.

1. Introduction

A small subset of human immunodeficiency virus type 1 (HIV-1) infected individuals, the "Elite Controllers" (EC), can control the virus and restrain progression to immunodeficiency without antiretroviral therapy (ART). It has been suggested that EC can hold the key to how a functional HIV-cure can be reached. Strict criteria have been used for defining EC and only <0.5% of the HIV-infected population fulfill such criteria, e.g., known HIV-1 positivity for ≥10 year with none or very few episodes of viremia and stable $CD4^+$ T cell counts (Olson et al., 2014). However, we hypothesize that EC still is a heterogeneous group as a result of that several definitions exists (Olson et al., 2014) which is illustrated by the claim that EC also can develop acquired immunodeficiency syndrome (Deeks and Walker, 2007).Therefore, we suggest that an even more strict definition of EC should be applied to minimize the bias of false phenotypic classification (Zhang et al., 2017).

Genome-wide association studies (GWAS) have tried to identify the background for HIV control. Multiple independent polymorphisms have been defined within the *HLA* and *CCR5-CCR2* locus that together explains ~25% of the observed variability in viral load (McLaren et al., 2012; Pereyra et al., 2010; McLaren et al., 2015). Apart from the HLAs, the AIDS restriction genes (ARGs) like CCR5-Δ32 (rs333), CCR5 59029G (rs1799987), and SDF1-3′A (rs1801157) that have been reported to have a strong association with disease control (Poropatich and Sullivan, 2011; O'Brien and Nelson, 2004). Though understanding about HIV-1 disease progression was generated, the studies were not designed to assess the impact of the full spectrum of functional variants within coding regions. Recent research has shown that several GWAS hits have no specific biological relevance to disease (Boyle et al., 2017). Most of the earlier studies trying to elicit the mechanism of disease progression and control in HIV have focused on specific predefined molecules or pathways, thereby ignoring the systemic, interconnected, immunological programs that are associated with individual immune defense mechanisms.

Gene expression data from EC are limited, mainly derived from the $CD4^+$ T cells. Using low-throughput TaqMan® Low-Density Arrays (TILDA) assay, elevated levels of Schlafen-11 were identified as a signature in T cells of the EC, as one of only 34 genes tested (Abdel-Mohsen

* Corresponding author at: F68, Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm 14186, Sweden.
E-mail address: ujjwal.neogi@ki.se (U. Neogi).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ebiom.2017.11.031

et al., 2013). Another study using high throughput RNA sequencing on $CD4^+$ T cells from EC did not reveal an entirely distinctive mRNA expression pattern (Rotger et al., 2010). Further transcriptome analysis of EC in a third recent study reported that seven genes were differentially expressed in EC compared to healthy controls. Notably, the gene encoding for macrophage inflammatory protein 1α (MIP-1α), a natural ligand for CCR5, was found to be upregulated (Walker et al., 2015). However, as the immune system is a complex, adaptable system and specific immune cells are dependent upon each other for stimulations and inhibition (Kaczorowski et al., 2017), analysis of single cell population may not reveal the complete picture.

Therefore, to understand the underlying mechanisms, which ensure that disease progression is prevented in EC, a comprehensive analysis of clinical phenotypes coupled to genetics and biomolecular mechanisms is required. The rapidly increasing accessibility of genetic and biomolecular expression data from new high-throughput technologies is the foundation to shift the traditional phenotype-first approach to explorative genetic or molecular data-first approaches. In this study, we aimed to explore a comprehensive analysis of host transcriptomics and proteomics data coupled to clinical phenotypes in a well-defined Swedish EC cohort with up to 20 years of clinical follow-up data.

2. Materials and Methods

2.1. Patients

Whole peripheral blood was obtained between 2014 and 2016 from three categories of individuals; an unbiased cohort of untreated HIV-1-infected EC (n = 19), treatment naïve patients with viremia (VP, n = 8), and HIV-1-negative persons (HC, n = 14). Viral load was below detection limit (<20 copies/mL) in all EC at the time of sample collection. Peripheral blood mononuclear cells (PBMCs) were isolated using density gradient centrifugation. For the proteomics profiling, additional plasma samples were obtained from therapy naïve patients with viremia (n = 24) and HIV-1-negative persons (n = 9) from the InfCareHIV cohort at the Karolinska University Hospital, Huddinge collected between 2010 and 2016. The EC was defined as known HIV-positivity more than a year and ≥3 consecutive viral load <75 copies/ml over one year (and all previous VLs <1000 cp/ml) and known HIV-1 positivity ≥10 years and minimum two VL-measurements (≥90% of all VLs <400 copies/ml). Group characteristics of the patient-categories are presented in Table 1. The individual characteristics of ECs have been given elsewhere (Zhang et al., 2017).

The study was approved by regional ethics committees of Stockholm (2013/1944–31/4). All participants gave informed consent. The patient identity was anonymized and delinked before analysis. The complete study design is illustrated in Fig. 1.

2.2. RNA Sequencing (RNA-Seq)

Total RNA was extracted from PBMCs using RNeasy Mini (Qiagen, Hilden, Germany) according to the manufacturer's protocol and the RNA quality was determined by Agilent RNA 6000 p kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany). The mean and SD RIN value was 8.85 (0.94). The library preparation, and RNA-Seq, were performed at the National Genomics Infrastructure, Science for Life Laboratory, Stockholm, Sweden. The libraries were prepared from 100 ng of total RNA samples using Illumina TruSeq® Stranded mRNA Library Prep Kit with poly-A selection following manufacturer's guidelines. The clonal cluster amplification was carried out with Illumina cBOT System. The pooled libraries were sequenced on HiSeq2500 (HiSeq Control Software 2.2.58/RTA 1.18.64) with a 2 × 126 setup using 'HiSeq SBS Kit v4' chemistry. The Base call (Bcl) files to FastQ conversion was performed using bcl2fastq from the Illumina's Consensus Assessment of Sequence and Variation (CASAVA) software suite (Illumina, US).

Table 1
Cohort characteristics.

	Elite controllers	Viremic progressors	Healthy controls	p-Value
n	19	32	23	
Age in years, median (IQR)	46 (40–52)	39 (32–52)	45 (39–52)	0.36[a]
Male, n (%)	10 (53)	15 (47)	13 (57)	0.77[a]
Ethnicity, n (%)				
Caucasian	8 (42)	20 (63)	18 (78)	0.19
Black	10 (53)	11 (34)	4 (17)	
Other	1 (5)	1 (3)	1 (4)	
Mode of transmission, n (%)				
Heterosexual	10 (53)	23 (72)	NA	0.11
MSM	4 (22)	8 (25)		
PWID	2 (10)	0		
Blood product	2 (10)	0		
Unknown	1 (5)	1 (3)		
Time since diagnosis, years-median (IQR)	9.0 (5.0–14.0)	1.3 (0.1–3.6)	NA	<0.001[b]
CD4 count, cells/mm^3, median (IQR)	950 (695–1160)	410 (302–527)	NA	<0.001[b]
CD8 count, cells/mm^3, median (IQR)	670 (570–900)	975 (735–1215)	NA	0.005[b]
CD4/CD8 ratio	1.52 (0.82–1.74)	0.46 (0.29–0.57)	NA	<0.001[b]
Viral load, Log_{10} copies/mL, median (IQR)	<1.27	4.60 (3.85–5.03)	NA	<0.001[b]
HIV-1 subtype, n (%)				
A1	0	6 (19)	NA	0.001
B	3 (16)	11 (34)		
C	5 (26)	5 (16)		
Recombinant forms	2 (11)	10 (31)		
Unknown	9 (47)	0		

NA: Not available, IQR: Interquartile range, MSM: Men who sex with men, PWID: People who inject drugs.
[a] The *Kruskal–Wallis* test.
[b] Mann-Whitney *U* test.

2.3. Data Processing

The quality control of raw sequencing reads was performed with FastQC, (www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were then aligned to the human reference genome GRCh37 Ensembl release 75 using Tophat v2.0.4. Gene-level count data was generated using Htseq v2.0.4 to summarize read counts for each gene. Counts per million (CPM) for protein-coding genes and non-protein coding transcripts were generated using edgeR (Robinson et al., 2010), and the CPM were log transformed to reduce the sequencing depth differences among the libraries. Transcripts with CPM value <0.5 in the whole group of minimal sample size were discarded to remove transcripts that were not expressed at a biologically meaningful level in any condition. Fragments per kilobase of transcript per Million mapped reads (FPKMs) were generated using cufflinks/2.2.1 (http://cufflinks.cbcb.umd.edu/).

2.4. Clustering Analysis

To analyze the similarities and dissimilarities between the samples, unsupervised principal component analysis and clustering analysis were performed using PlotMDS function in the Limma (linear models for microarray data) R package (Ritchie et al., 2015) to generate Multi-Dimensional Scaling plot (MDS). Heatmap of hierarchical clustering of the samples was generated by calculating the matrix of Euclidean distances from the logCPM to examine the relationships between samples.

Please cite this article as: Zhang, W., et al., Transcriptomics and Targeted Proteomics Analysis to Gain Insights Into the Immune-control Mechanisms of HIV-1 Infected Elite Contr..., EBioMedicine (2017), https://doi.org/10.1016/j.ebiom.2017.11.031