GuavaH: A Compendium of Host Genetic Data in HIV Biology and Disease

István Bartha[1,2,3], Paul J McLaren[1,2,3], Angela Ciuffi[1], Jacques Fellay[1,2,3] and Amalio Telenti[1]

[1]Institute of Microbiology, University Hospital Lausanne, Switzerland, [2]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland, [3]Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Figure 1

SUMMARY

There is an ever-increasing volume of data on host genes that are modulated during HIV infection, influence disease susceptibility or carry genetic variants that impact HIV infection. We created GuavaH (Genomic Utility for Association and Viral Analyses in HIV, http://www.guavah.org), a public resource that supports queries of genome-wide genetic variation and gene expression profile across multiple phenotypes relevant to HIV biology (Figure 1).

We included original data from 8 genome and transcriptome studies which cover phenotypes such as HIV acquisition, plasma viral load, disease progression, viral life cycle, latency and viral-host genomic interaction (Figure 2).

The GuavaH framework (http://www.guavah.org) supports queries on

- **single nucleotide polymorphism across different HIV related phenotypes,**
- **in vivo gene expression in the setting of human infection (CD4+ T cells),**
- **in vitro gene expression data in models of permissive infection and latency.**

Figure 2

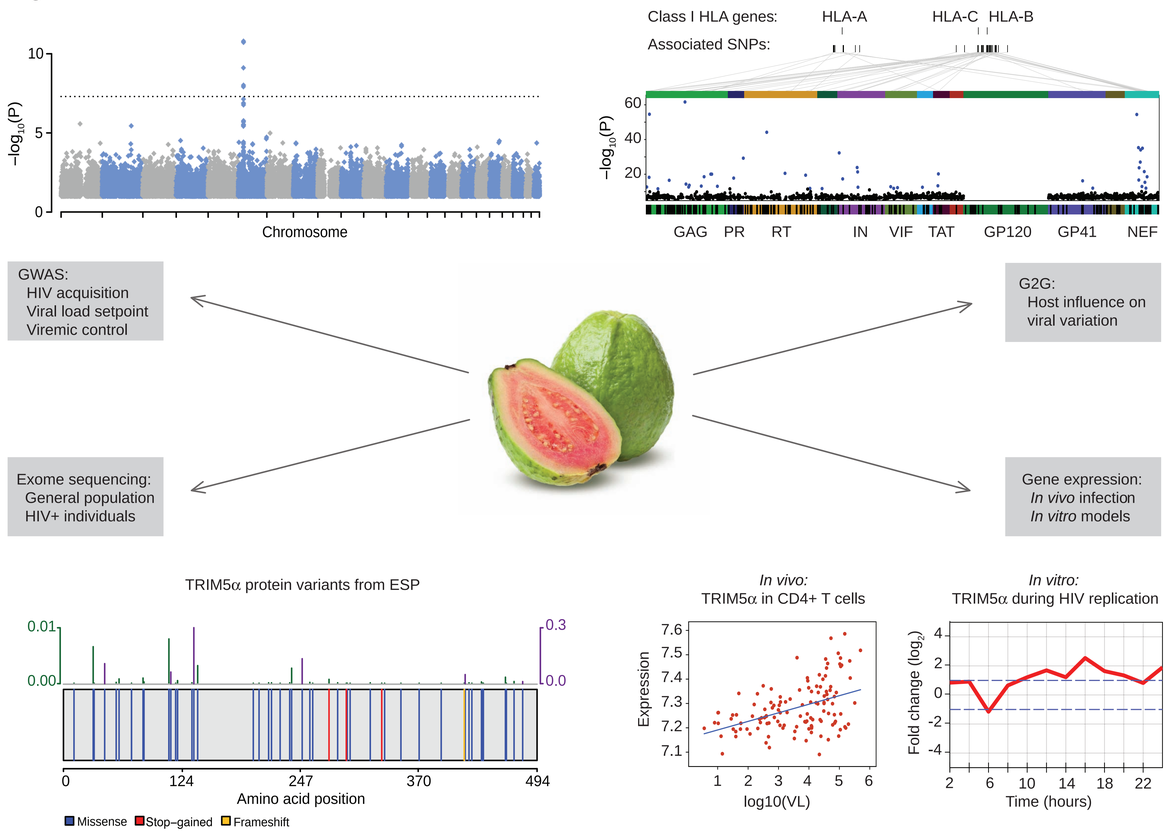
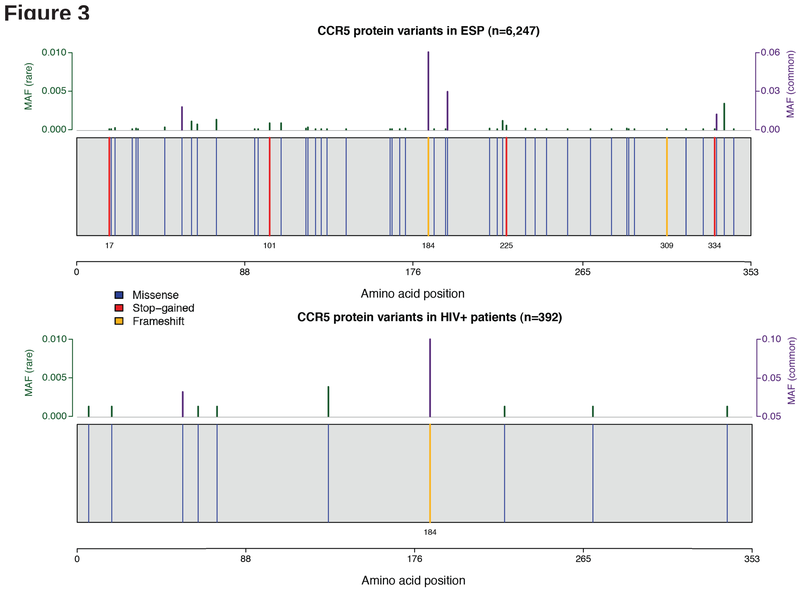


Figure 3



INCLUDED DATASETS

Common Genetic Variation of the Host

Genome-wide association results on HIV control (set point plasma viral load [1, 2] and elite control [3]) and on susceptibility to infection in a cohort of highly exposed seronegative individuals [4].

HIV Sequence Mutations Affected by Host Polymorphisms

In addition to traditional GWASs of clinically related outcomes we included data from a recent joint host-HIV genomic analysis [5]. The genome-to-genome approach identifies loci of host-pathogen conflict independently of clinical data, and it provides insights in the distribution of selective pressure exerted to the viral genome.

Rare Exonic Variation of the Host

Gene-level graphical representation of functional variants. We included sequence variation on a large sample taken from non-infected population (http://evs.gs.washington.edu/EVS/), and on 392 HIV infected individuals. This allows for first-pass estimation of the mutational burden of a given gene (Figure 3). Access to exome data in the HIV+ sample is restricted but gene-level queries are possible upon request (contact@guavah.org).

Gene Expression Variation

Results of functional transcriptome analyses. *In vivo* data were obtained by microarray studies of CD4+ T cells from 127 individuals chronically infected with HIV [7]. These are augmented with temporal *in vitro* analysis of the HIV replication cycle in a T cell line (Sup T1) [8].

REFERENCES

Fellay J, Ge D, Shianna KV et al: Common genetic variation and the control of HIV-1 in humans. PLoS Genet 2009, 5:e1000791.
Fellay J, Shianna KV et al: A whole-genome association study of major determinants for host control of HIV-1. Science 2007, 317:944–947.
Pereyra F, Jia X, McLaren PJ et al: The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010, 330:1551–1557.
Lane J, McLaren PJ et al: A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. Hum Mol Genet 2013, 22:1903-1910.
Bartha I, Carlson JM et al: A genome-to- genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. eLife 2013, 2:e01123.
Rotger M, Dang KK et al: Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. PLoS Pathog 2010, 6:e1000781.
Mohammadi P, Desfarges S et al: 24 hours in the life of HIV-1 in a T cell line. PLoS Pathog 2013, 9:e1003161.