GWAS reveals new genetic associations with liver fibrosis progression in HIV/HCV

642LB

Sigrid Le Clerc[1], Damien Ulveling[1], Taoufik Labib[1], Maria Winnock[2,3], Ioannis Théodorou[4,3], Laurent Abel[5,3], Yves Lévy[6,3], Dominique Salmon[7,3], Jean-François Zagury[1,3], Stéphanie Dominguez[6,3]

[1]Chaire de Bioinformatique; Laboratoire GBA (EA4627); Conservatoire National des Arts et Métiers, Paris, France[2]: INSERM U897/ISPED; Bordeaux Segalen University, Bordeaux, France, [3]ANRS Genomic Group (French Agency for Research on AIDS and Hepatitis), Paris, France, [4]INSERM U945; Pierre et Marie Curie University (UPMC), Paris, France, [5] *INSERM U980; Laboratory of Human Genetics of Infectious Diseases; University Paris Descartes, Imagine Institute, Paris, France,* [6]INSERM U955; Department of Clinical Immunology and infectious disease, Henri Mondor hospital, Creteil, France, [7]Department of Infectious and Tropical Diseases; Cochin Hospital, AP-HP, Paris, France

Abstract

Background: Genetic studies with common variants in HIV mono-infection have mainly revealed associations with HIV non-progression and elite control (notably in the HLA locus). In HCV infection, the main association results were found between IL28B variants and clearance of the virus either under treatment or spontaneously. We took the opportunity of the large French ANRS CO13 HEPAVIH cohort of co-infected HIV/HCV patients to study these signals and search for additional markers associated with various phenotypes related to the course of the co-infection in a genome-wide association study (GWAS)

Methods: We used the Illumina Omni2.5 BeadChip to genotype 439 Caucasian patients who had given their informed consent for genetic studies. We first performed the usual quality controls (SNP calling >98%, patient calling >95%, HWE >10^{-6}, IBD <0.125, MAF >1%) and then measured the stratification using the Eigenstrat® software. We report here the early findings of GWAS screening for host genetic determinants of progression of liver fibrosis. The fibrosis stage at baseline was assessed by elastography (Fibroscan®) scoring fibrosis in kilopascal. For each single nucleotide polymorphism (SNP), the genetic association was estimated by linear regression using as covariates the age at baseline, gender, the HCV genotype, the duration of both HIV and HCV infection, alcohol and drug consumption, effective cumulative time on antiretroviral treatment, and the first two principal components of the stratification analysis.

Results: After quality control and stratification analysis, there were 412 patients left to search for genetic associations on 922,971 SNPs. Fibrosis stage was available at baseline in 399 patients. Three SNPs in two regions passed the Bonferroni threshold for genome-wide significance (5x10^{-8}): the rs11790131 SNP located in the locus 9p21.3 in the gene *LOC392288* (*P*=1.8x10^{-8}), the rs73060964 and rs73060937 SNPs (r^2=1), located in the locus 3p22.1 in the gene *LOC101928135* (*P*=4x10^{-8}). The previously published genetic markers of HIV and HCV mono-infection did not reach this statistical significance.

Conclusion: We report new signals, reaching genome-wide significance, associated with liver fibrosis in HIV/HCV co-infected patients. Confirmation in other genetic studies is needed. These potential markers of fibrosis will be studied prospectively in the cohort as well as other phenotypes such as HIV and HCV treatment response and tolerance, HIV progression and occurrence of hepato-cellular carcinoma.

Background

The development of new DNA technologies has allowed the successful completion of genome-wide association studies (GWAS) and multiple genetic associations were identified in several diseases (1-2). In HCV infection, the main association results were found between *IL28B* variants and clearance of the virus either under treatment or spontaneously, while in HIV-1 mono-infection, main result are located in *HLA* region. In this study, we took the opportunity of the large French ANRS CO13 HEPAVIH cohort of co-infected HIV/HCV patients to study these signals and search for additional markers associated with liver fibrosis progression in a genome-wide association study.

Methods

Subjects: The patients analyzed in this study belonged to a the large French ANRS CO13 HEPAVIH cohort of co-infected HIV/HCV patients. The cohort is composed of 1075 patients. Among this 1075 patients, 653 have given their informed consent for genetic studies and among the 653, 471 patients have declared a European ancestry and were kept for the analysis.

Genotyping and quality control: The 471 patients extract from the French ANRS CO13 HEPAVIH cohort were genotyped using the Illumina Omni2.5 BeadChip containing 2 391 739 SNPs. We performed a usual quality control: SNP calling > 98%, patient calling > 95%, Hardy-Weinberg equilibrium > 10^{-6}, identity by descent < 0.125, minor allele frequency > 1%. After quality control 922,971 SNPs and 439 patients were available for the association analysis.

Stratification: To correct for possible population stratification, genotypes were analyzed using EIGENSTRAT utility of the EIGENSOFT package version 2.0 (3). This analysis with the Eigenstrat software pointed out 27 outliers, which were removed from further analysis. It thus remained 412 patients for the following step.

Phenotypes analyzed: The main phenotype studied was the progression of liver fibrosis. The fibrosis stage at baseline was assessed by elastography (Fibroscan®) scoring fibrosis in kilopascal. Among the 412 patients fibrosis stage was available at baseline for 386 patients and duration of infection was available for 329 patients. In the first step, we considered the liver fibrosis stage in kPa as a quantitative phenotype. In a second step, we used survival analysis techniques to take into account the individual duration of infection in the analysis; Failure (liver fibrosis > 9 kPa) and censored (liver fibrosis < 9 kPa) events were defined, and the duration of infection was estimated from the presumed year of HCV acquisition to the year in which elastography was carried out.

Statistical analysis: For each SNP, we performed a linear regression with PLINK (4) to test association between liver fibrosis stage in kPa and genotypes in the additive model. For duration phenotypes assessing the liver fibrosis progression, we used a Cox model using Probabel (5) considering the presumed year of HCV acquisition to the year in which elastography was carried out. To focus more specifically on the genetic factors affecting liver fibrosis, several characteristics known to affect it had to be taken into account as covariates in the Logistic regression and Cox model analyses: age at baseline, gender, VHC genotype, duration of VHC contamination (only in linear regression), Alcool consumption, cumulative duration of the antiretroviral treatment. Genetic stratification was also taken into account by adding as covariates the 2 principal components identified by the Eigenstrat method (3).

Results

Association with liver fibrosis stage (in kPa)

After quality control and stratification analysis, we disposed of 386 patients with clinical data to search for genetic associations on 922,971 SNPs. Figure 1 shows the distribution of the p-values along the chromosomes, and table 1 presents the most significant signals obtained for analysis of liver fibrosis in kPa in additive mode.

Three SNPs in two regions passed the Bonferroni threshold for genome-wide significance (5x10^{-8}):

- The rs11790131 SNP located in the locus 9p21.3 in the gene *MAP1LC3BP1* (*P*=1.8x10^{-8}). This SNP is ~ 20 kb apart from gene *ACER2*. The rs11790131-AA genotype favored severe fibrosis (Figure 2A).
- The rs73060964 and rs73060937 SNPs in total linkage disequilibrium (r^2=1), located in the locus 3p22.1 in the gene *LOC101928135* (*P*=4x10^{-8}). The rs73060937-CC genotype favored severe fibrosis (Figure 2B, the rs73060964 and rs73060937 are in high linkage disequilibrium, thus we present the plot for only one SNP).

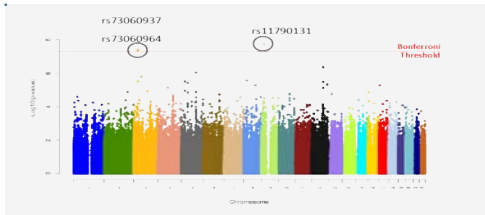

Figure 1: Manhattan plot of the distribution along the human autosomes of -$\log_{10}$(p-values) obtained for analysis of the liver fibrosis stage measured in kPa in the additive model.

Table 1: Significant signals obtained for analysis of the liver fibrosis stage measured in kPa. The rs11790131-A and rs73060937-C allelic frequency in population with extreme stage of liver fibrosis (<7.2 kPa, >12.5 kPa) are supplied.

SNP	chr	*P*	Location
rs11790131	9	1.8x10^{-8}	*MAP1LC3BP1* (intron)
rs73060937	3	4x10^{-8}	*LOC101928135* (intron)
rs73060964	3	4x10^{-8}	*LOC101928135* (intron)

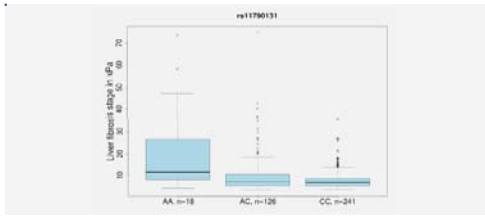

Figure 2A: Box plot representing the liver fibrosis stage in kPa of subjects in function of the rs11790131 genotypes.

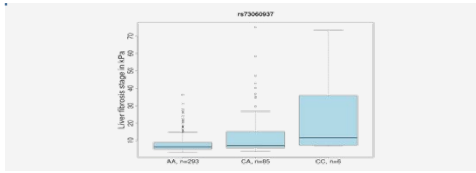

Figure 2B: Box plots representing the distribution of the liver fibrosis stage (in kPa) according to the rs7306937 genotypes.

Survival analysis with liver fibrosis event >9 kPa

In this second analysis we worked with the 330 patients with clinical data available all along the infection (Figure 3). This analysis highlighted a new signal on chromosome 16 passing the Bonferroni threshold:

- Two SNPs, rs16975170 and rs8056742, in total linkage disequilibrium (r^2=1) are located in region 16q24.1 in *KIAA0513* gene (*P*=1.87x10^{-8}). The rs8056742-G allele favored a more rapid progression towards a fibrosis stage > 9 kPa (Figure 4, the rs16975170 and rs8056742 are in high linkage disequilibrium, thus we present the plot for only one SNP).
- Interestingly, The two SNPs seem to modulate the *ZDHHC7* gene expression (*P*=6.25x10^{-8}) according to the eQTL database "GHS Express" (6). *ZDHHC7* is ~ 86 kb away from the two SNPs. Moreover there exist several SNPs in high LD with rs16975170 and rs8056742 SNPs, located in a region with particular chromatin state in lymphoblastoid cell line, that can influence nearby gene expression.

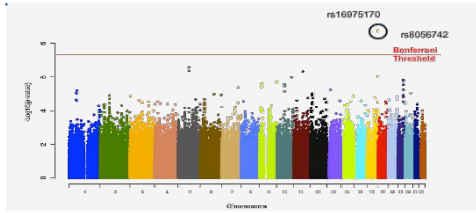

Figure 3: Manhattan plot of the distribution along the human autosomes of -$\log_{10}$(p-values) obtained for analysis of the liver fibrosis progression measured in kPa in the additive model.

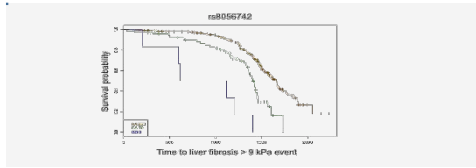

Figure 4: Survival curves of the 3 rs8056742 genotypes with the event "liver fibrosis > 9 kPa ".

Discussion

Biological interpretation

- The first approach of association analysis with the quantitative liver fibrosis stage at inclusion (in kPa) identified 3 SNPs (rs11790131, rs73060964 and rs73060937) on chromosome 9 and 3, all located within pseudo-genes. The rs11790131 SNP is 20 kb away from a relevant gene, *ACER2*, that encodes a protein promoting cell proliferation and survival. More precisely this protein hydrolyzes the sphingolipid ceramide into sphingosine and free fatty acid (Entrez gene, UniProtKB) and It may play a role in metabolic troubles according to MalaCards. This gene is expressed widely and notably in the liver.

- The second approach looked for associations with liver fibrosis progression using a survival analysis and highlighted 2 SNPs in LD (rs16975170 and rs8056742) on chromosome 16, located in the gene *KIAA0513*. *KIAA0513* codes for a signaling molecule that interacts with modulators of neuroplasticity, apoptosis, and the cytoskeleton. Interestingly, *KIAA0513* is link with the host cell responses induced by hepatitis C virus binding (7). The 2 SNPs are located in a specific chromatin region which can influence nearby gene expression according to the Encode project. Moreover, they seem to modulate *ZDHHC7* gene expression according to a eQTL database (6). The *ZDHHC7* protein may function as a specific palmitoylacyl-transferase that participates to the palmitoylation of sex steroid receptors.

Conclusion

- This preliminary GWA study on the HEPAVIH cohort has identified positive signals associated with liver fibrosis. Of course these signals need to be confirmed either by replication in other cohorts or by experiments assessing the impact of the identified genes/polymorphisms. We are dealing here with HIV/HCV co-infected patients which bring a new interest but also more complexity because of the role of dual infection and treatment regimens.

- Among the forecast next steps, we plan to increase the number of patients with more extreme phenotypes to enhance the power of the analysis, and to explore the other phenotypes such as HIV natural course, HIV natural course, occurrence of hepato-cellular carcinoma, tolerance and efficacy of both HIV and HCV treatments and HCV spontaneous clearance.

References

1. Cooper JD, Smyth DJ, Smiles AM *et al.* Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet.* 2008;40:1399-401.
2. International HIV Controllers Study1, Pereyra F, Jia X *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* 2010:1551-7.
3. Price AL, Patterson NJ, Plenge RM *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904-9.
4. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-75.
5. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics.* 2010;11:134.
6. Zeller T, Wild P, Szymczak S *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One.* 2010;5:e10693.
7. Fang X, Zeisel MB, Wilpert J *et al.* Host cell responses induced by hepatitis C virus binding. *Hepatology.* 2006;43:1326-36.