PhD

PROGRAM IN TRANSLATIONAL
AND MOLECULAR MEDICINE

DIMET

UNIVERSITY OF MILANO-BICOCCA
SCHOOL OF MEDICINE AND FACULTY OF SCIENCE

Single Nucleotide Polymorphisms and circulating microRNAs for monitoring HCV disease progression: an integrated approach

Coordinator: Prof. Andrea Biondi
Tutor: Dr. Sergio Abrignani

Dr. Francesco Marabita
Matr. No. 716450

XXIII CYCLE
ACADEMIC YEAR
2009-2010

Single Nucleotide Polymorphisms and circulating microRNAs for monitoring HCV disease progression: an integrated approach

Francesco Marabita

Table of Contents

Chapter 1. Hepatitis C: general introduction

Hepatitis C Virus (HCV) is a successful pathogen that establishes a persistent infection in about 70% of individuals who contract it. The chronic disease is associated with severe complications such as liver cirrhosis and hepatocellular carcinoma (HCC). Worldwide, an estimated 170 million people are chronically infected, with 3 to 4 million new cases each year.[1,2] HCV infection is currently considered one of the main indications for liver transplantation in many Western countries. Liver transplantation, however, is not a cure for HCV, as virus recurrence is rapid and universal.[3] Currently, no vaccination is available against HCV. A major challenge to vaccine development is the broad genetic variation between viral genotypes and strains and the substantial antigenic variation among different quasispecies within individual patients.[4] Currently, the most effective treatment is based on the combined use of Interferon-α and ribavirin, which however yields sustained response rate in about 50% of the patients, depending on host and viral factor. Furthermore, this treatment is expensive and has significant side effects and contraindications.[5]

Therefore, for HCV infection, there is a crucial need for novel, preferably non-invasive tests that could be used to improve the diagnosis and the management programs for patients at high risk of developing severe liver complications, such as cirrhosis and HCC. Classical biochemical markers to follow disease progression show limited potential mainly for invasiveness or little specificity and sensitivity, especially for HCC.

The goal of this work was to discover useful biomarkers of disease progression in the setting of chronic HCV infection. In the first part, a Genome Wide Association Study (GWAS) was carried out to study the genetic variability influencing the natural history of HCV infection. A cohort of patients was genotyped and disease stage was evaluated through a liver biopsy before any antiviral treatment. Although no hits reached a genome-wide significance level, candidate regions showing suggestive association were identified. In particular, the gene region containing the *GADD45G* gene resulted of particular interest, given its role in cell biology.

The second part focused on the use of circulating microRNA as blood-based biomarker of disease status. microRNAs were profiled in sera of healthy donors and patients infected with HCV at different stages of the disease. A list of differentially represented miRNAs was identified, containing miR-122 and miR-885-5p. These miRNAs could be used to monitor disease progression in chronic HCV infection, using small amounts of serum.

Hepatitis C Virus

The Hepatitis C virus is an enveloped, positive-strand RNA virus, classified as a distinct genus (*Hepacivirus*) within the *Flaviviridae* family. The HCV genome is 9.6 kb in length, containing 5' and 3' untranslated regions (UTRs) and a single open reading frame encoding a polyprotein that is processed by viral and cellular proteases into four structural (Core, E1, E2, p7) and six nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B). HCV infects primarly hepatocytes and its replication occurs entirely in the cytoplasm.[6]

The virus binds to its receptor at the cell membrane, resulting in an endocytotic process that leads to the viral uncoating ad release of the viral genome within the cytoplasm. Strong evidence exists for specific HCV entry factors: the tetraspanin CD81,[7] the scavenger receptor class B type I (SR-BI)[8] and the tight junction proteins claudin-1 and occludin.[9,10] After internalization, the genome is directly translated and viral proteins are produced. Protein translation depends on an Internal Ribosome Entry Site in the 5' UTR. Nonstructural proteins assemble to form a replication complex that drives RNA replication through a minus-strand RNA intermediate. Replication occurs within the so-called membranous web, which derives from intracellular membrane alterations induced by nonstructural proteins. Therefore, a nucleocapsid assembles and virions presumably form by budding into a subcompartment derived from the endoplasmic reticulum. Viral particles leave the cell through the classical secretory pathway.[6]

HCV was first identified in 1989,[11] but only in 2005 the production of infectious HCV particles was reported from a cell culture, using a genotype 2a isolate

(JFH1). This cell-culture derived viral particles were infectious *in vivo* in chimpanzees and in immunodeficient mice human liver xenografts.[12-14] Other model systems to study specific aspects of viral life cycle or immunobiology include the replicon system and the retroviral pseudoparticles.[15] Conversely, the chimpanzee is the only *in vivo* available model to study HCV, because small animal models that have been developed have only limited use.

HCV is classified into six major genotypes (designated 1-6), and many subtypes according to the heterogeneity distributed throughout the genome, except for the 5' UTR. The region coding for the envelope glycoproteins is the most variable. HCV lacks efficient proofreading activity during replication, hence the virus persists as a collection of quasispecies within an infected individual. This variability accounts for alterations of the antigenic properties of the proteins, thus allowing the virus to escape neutralizing antibodies.[16] Moreover, HCV diversity influences the therapy responsiveness and hampers the efforts towards the development of an efficacious vaccine. Genotypes 1-3 have a worldwide distribution. Genotype 1 is prevalent in Northern Europe and Northern America (subtype 1a), as well as in China, Japan, Southern and Eastern Europe (subtype 1b). Genotypes 2 and 3 are widely distributed throughout the world and are generally associated with higher rates of treatment success. Genotype 4 is prevalent in Middle East and North Africa, genotype 5 almost exclusively in South Africa and genotype 6 in South Asia and Australia.[17]

Natural history of HCV infection

The term natural history refers to the description of the course of the disease after infection, including clinical characteristics and factors that influence the progression of the disease. The accurate assessment of natural history has been very demanding, because generally the acute phase of the infection is silent, although the determination of the onset is critical to follow the full course of the disease. Furthermore, the disease progression is modified by factors and treatments that in principles may raise possible controversies on the definition of the long-term natural history of chronic HCV infection.

After initial exposure, the majority of cases are asymptomatic and rapid, while fulminant hepatitis is a rare event.[18] HCV infects the livers, and specifically the hepatocytes.[6] High viral titers are observed during the first weeks of infections, although the virus by itself is relatively noncytopathic and liver injury often occurs after 1-2 months.[19] About 80% of infected people fail to clear the virus and develop a persistent infection that is diagnosed by the detection of HCV RNA for at least 6 months. During chronic infection, HCV RNA generally ranges from 10^5 to 10^7 IU/ml, but the levels can fluctuate widely. The most important sequelae include progressive liver fibrosis, cirrhosis, liver failure and HCC. The chronic infection usually progresses to severe liver disease in about 20% of the persistently infected subjects, and once cirrhosis is established the risk of developing HCC is 1-4% per year.[18] The rate of progression is usually slow, with 20 or 30 years elapsing between the infection and the development of cirrhosis and HCC. However, the reported rates of cirrhosis development have been shown to vary among studies, from 2-8% in studies of young subjects to 20-30% in older patients.[20] These differences account for the fact that many host factors modulate the risk of liver disease progression, as further explained below.

The pathogenesis of the liver disease is mainly immune-mediated. Chronic infection is associated with portal inflammation, periportal necrosis, fibrosis and often steatosis.[21] Destruction of hepatocytes by the chronic inflammation is accompanied by liver regeneration. For mechanisms not completely understood, in a sizable fraction of cases, liver destruction is followed by scar formation and deposition of fibrotic tissue instead of the normal tissue. Hepatic fibrogenesis represents a wound-healing response characterized by a net accumulation of extracellular matrix (ECM) resulting from increased synthesis and decreased degradation.[22] Hepatic Stellate Cells (HSCs) represent the primary source of ECM.[23] In normal liver, HSCs are described as being in a quiescent state. A distinguished feature of quiescent HSC is the presence in their cytoplasm of multiple lipid droplets containing high amounts of vitamin A. In response to inflammatory stimuli and Reactive Oxygen Species (ROS), HSCs become activated, proliferate and transform into myofibroblasts expressing α-

smooth muscle actin (α-SMA), an actin isoform found in smooth muscle cells.[23] An increased production of collagen type I is characteristic of this phase, a process regulated both transcriptionally and post-transcriptionally.[24, 25] Hepatocytes, Kupffer cells, platelets and endothelial cells contribute to the activation of HSC. Kupffer cells are an important source of TGF-β1, which is a potent stimulus for the production of ECM. TGF-β1 also acts in an autocrine loop, because HSCs are themselves a source of this cytokine. In addition, TGF-β1 inhibits cell proliferation and promotes differentiation or apoptosis.[26, 27]

Nevertheless, a fraction of people with chronic HCV infection will never progress to cirrhosis, because the disease progression may be particularly slow, depending on both nonmodifiable and modifiable cofactors. The risk of progression is increased by many host factors, including older age at acquisition of the infection, male gender, alcohol consumption, coinfection with Hepatitis B Virus (HBV) and Human Immunodeficiency Virus (HIV), iron overload and other metabolic factors (insulin resistance, obesity). Their role in the natural history of HCV infection has been investigated through retrospective and prospective studies or a combination of both (for a review, see ref. [20] and [28]). In retrospective-prospective studies, the date of infection is dated back retrospectively but then the subjects are followed prospectively. Depending on the study design and the population being evaluated, different estimates of disease progression have been reported, underling the complexity of an accurate determination of the natural course of the chronic HCV infection.

Host-related factors. The age at infection represents one of the most important variables affecting the rate of progression of liver disease. In particular, the age at infection positively correlates with the rate of progression of liver fibrosis, even accounting for disease duration.[29-32] Furthermore, children who became chronically infected perinatally have been reported to have a very slow disease progression.[33, 34] An interesting question is whether the rate of fibrosis progression remains constant or increases over time during ageing. A study with paired biopsies taken at different times showed that the speed of fibrosis progression increases over time, independently of age at infection.[30] An explanation for the substantial role of the age in influencing the disease

outcome is still unclear. Possible hypotheses include a role of the modifications of the immune system during ageing or a reduced mitochondrial functionality.

Male gender[29-31] and alcohol consumption (>30 g/day in men and >20g/day in women)[2] have been repeatedly associated with faster fibrosis progression. Furthermore, concomitant infections with HIV or HBV have been shown to increase the risk of more rapid disease progression. Actually, these viruses have shared the same routes of infection as HCV, being common between injection drug users.

HBV/HCV coinfection is associated with more severe fibrosis and an increased risk of HCC.[35, 36] Commonly, during chronic hepatitis one of the two viruses is suppressed, with the most frequent scenario being active HCV replication in the setting of occult HBV infection, which is characterized by suppression of HBV surface antigens but persistence of HBV DNA. This interference appears to be mediated by indirect immune response rather than direct replicative suppression within the liver.[37]

In patients coinfected with HIV and HCV, reduced CD4+ and CD8+ T-cell responses are developed against HCV.[38] HIV causes a significant decline in CD4+ lymphocyte count, resulting in a reduced CD4+ T-cell help for viral-specific responses. Consequently, this immunosuppressed status negatively affects the course of chronic liver disease, leading to increased persistence rate, increased viral load and accelerated fibrosis progression.[39-41] These considerations underlie the importance of the immune system in mediating the liver damage and promoting the progression to cirrhosis. Indeed, within the liver of HIV/HCV coinfected patients, persistently replicating HCV is coupled to the continuous presence of T cell, which however are ineffective because of the inadequate CD4 help resulting from HIV infection. This setting represents a potential mechanism leading to increased progression to cirrhosis, because of the involvement of the immune system in the liver pathogenesis.

Other metabolic factors contribute to the modification of the natural course of HCV liver disease. These include: iron overload, steatosis, insulin resistance and obesity. All these factors have been shown to accelerate disease progression.[28, 42-44] With regard to steatosis, a complex relationship exists with the virus and,

additionally, the separation of the relative contributions of steatosis, obesity and insulin resistance is challenging, because of the underlying relationships between them. Moreover, accelerated rates of fibrosis progression associated with steatosis are observed in a genotype-specific way, as further explained below. Both *in vitro* and *in vivo*, HCV has been shown to modulate lipid metabolism, mainly through gene expression regulation or direct induction of intracellular lipid droplets accumulation.[45, 46] Moreover, HCV replicates within a compartment derived from intracellular membrane alterations induced by nonstructural proteins, thus providing another direct link between lipids and HCV life cycle.

Other host genetic factors may contribute to different disease outcomes during chronic HCV and numerous studies have been conducted so far to highlight the genetic of fibrosis progression.[47] The major limitations of these studies include a small sample size and lack of replication in independent populations. It is believed that the risk of advanced fibrosis is dependent upon the role of different genes, but with limited relative contribution. Each study assessing the role of host genetic factors must correct the reported associations for the covariates, which strongly influence disease progression. For instance, in order to eliminate possible confounding factors, the age at disease infection should always be considered in the association model, since it dramatically influences the risk of advanced liver disease.

Other unidentified host genetic factors may contribute to the risk of advanced disease, and numerous candidate-gene studies have been performed. For instance, Single Nucleotide Polymorphisms (SNPs) in the gene region of IFN-γ, TNF-α, or TGF-β have been associated with HCV-induced fibrosis/cirrhosis.[48-52] The rationale behind these studies is the involvement of the host immune system not only in developing antiviral responses, but also in primarily contributing to the pathogenesis of liver disease. Hence, polymorphisms in immune-related genes have been analyzed and significant associations have been reported. However, these studies neither were replicated nor resulted in relevant clinical implications. The main limitations that hampered the clear recognition of the host genetic contribute to the natural history included poor

study design, limited sample size, noninclusion of viral and host cofactors in the association model and small effect size of the causative variants.

One of the few studies providing both a large sample size and a replication phase is the work of Huang et al.[53] Using a custom panel of SNPs distributed throughout the genome, they performed a two-phase markers selection to identify SNPs associated to the risk of developing liver fibrosis. On the first cohort a thousand markers were selected for replication on the second group. Finally, a list of seven markers was used to identify high-risk patients. Compared to the clinical variables, the authors state an improvement in classifying patients with high-risk for cirrhosis in Caucasians, with a positive predictive value of 96% in the validation cohort. The authors proposed an algorithm to identify patients with an increased risk of cirrhosis, which would be useful in the process of decision making for therapy administration and individualization. However, the exact effect size of the identified genetic variant is hard to interpret from the study, because the age at the infection was not accounted for. Nevertheless, another independent study replicated the association using longitudinal observations.[54] The fibrosis progression was evaluated through serial biopsies, thus eliminating the uncertainty associated with the definition of the onset of the infection. The seven-marker signature was validated in males but not in females and resulted most evident when the initial biopsy showed no signs of fibrosis. Therefore, the authors propose the use of the algorithm to predict fibrosis progression in males with initial no/mild fibrosis.

Numerous reports highlighted the involvement of the human leukocyte antigen (HLA) locus on the outcome of HCV infection. The HLA is one of the most genetically variable loci in humans, and the effectiveness of the antiviral T-cell response is influenced by the ability of different HLA alleles to bind viral antigens. Actually, both for HLA class I and class II, specific alleles have been associated mainly to disease susceptibility and viral clearance but also to disease severity.[55-58] However, the effect of HLA allele as predictor of disease progression resulted small compared to other clinical covariates.

Viral factors. The influence of viral genotype in the natural history of HCV infection remains controversial. In particular, viral genotype has not been reported as a predictor of fibrosis progression,[30] although some investigators reported significant higher rates of progression for genotype 1b,[59] or more recently for genotype 3.[60] The latter case is noteworthy because a number of reports described an association with hepatic steatosis.[42-44] In principle, the enhanced fibrosis progression may be related to the increased steatosis and inflammation in patients with genotype 3. Thus, the accelerated rate of fibrosis progression in patients with genotype 3 is believed to be mediated by increased steatosis, although differences in the baseline characteristics of the studied population, inclusion criteria and other unknown genetic factors may account for the reported discordant results.

Conversely, the viral genotype has profound implications on treatment outcome: genotype 1 is generally associated with lower responses (40-50%), whereas genotypes 2 and 3 are associated with more favorable responses.[5, 61] Nevertheless, genotype 1 accounts globally for the majority of the infections, making the current standard of care unsatisfactory for many patients.

In summary, the genetic determinants of HCV pathogenesis and disease progression act in a polygenic setting. Thus, the natural history of HCV infection can be considered as a complex trait where the disease outcome is influenced at one side by the host genetic variability and at the other by environmental modifiable or non-modifiable factors. While the role of certain host characteristics, such as the gender and the age at acquisition of the infection, has been long established, the impact of genetic variants needs further investigation, with the ultimate goal of identifying patients at high risk of developing advanced liver complications.

Virus-host interactions: miR-122, IL28B and HCV infection

During the last five years two uppermost discoveries decisively contributed to understanding the complex interaction between the virus and the host biology. These findings were the identification of a human microRNA (miRNA) required

for HCV replication and the discovery that SNPs were strongly associated with the viral clearance.

miRNAs are small noncoding RNAs of 22-25 nucleotides that regulate gene expression by binding mostly the 3' UTR of target mRNAs. They are broadly conserved and act post-transcriptionally directing target gene repression. MiRNA genes are transcribed as capped and polyadenylated miRNA transcripts (pri-miRNA) that are processed into the nucleus to a ~70 nucleotides precursor form (pre-miRNA). Pre-miRNAs are exported into the cytoplasm by Exportin 5 and processed by the enzyme Dicer into an imperfect RNA duplex, with a 2 nt 3' overhangs.[62, 63] The strand of the duplex with the weakest base pairing at its 5' terminus is preferably loaded to Argonaute (Ago) proteins, the core of the RNA-induced silencing complex (RISC).[64] The RISC complex is directed by the miRNAs to the targets mRNAs by imperfect complementarity base pairing. The seed region (nucleotides 2-8) is the most important for target recognition and silencing.[65] About one thousand miRNAs have been identified in the human genome to date (miRBase release 16), which can influence expression of more than a half of the human genes.[66] MiRNAs are expressed in a tissue-specific manner and changes in miRNA expression within a tissue have been reported in several human diseases, including cancer, inflammation, and chronic viral infections.[67-70]

MicroRNA-122 (miR-122) is a liver-specific miRNA constituting the 70% of the total liver miRNA content.[71, 72] Replication of HCV is affected by miR-122 expression. Studies in hepatoma cells expressing a genotype 1 replicon demonstrated that miR-122 down-modulation resulted in a manifest decrease of HCV RNA replication.[70] This effect is mediated by the direct binding of miR-122 to HCV 5'-UTR and consequential stimulation of HCV RNA translation, through enhancement of the association with ribosomes.[70, 73] Moreover, a direct requirement of miR-122 in viral RNA replication has been established.[74] Further investigation confirmed that miR-122 was required for HCV genotype 2 replication and infectious virus production, using a full-length sequence (J6/JFH chimera). Interestingly, the same study demonstrated that a functionally operative miRNA biogenesis pathway is required for efficient HCV

infection.[75] *In vivo* studies demonstrated that patients poorly responding to IFN therapy had decreased pre-treatment miR-122 levels compared with responders, although the therapy did not influence the expression of miR-122. Furthermore, viral load did not correlate with miR-122 levels.[76, 77] However, in infected chimpanzees, inactivation of miR-122 resulted in suppression of HCV viremia, down-regulation of interferon-regulated genes, and improvement of liver pathology.[78] These findings need additional investigation to clarify the *in vivo* relationship between miR-122 and the virus.

MiRNAs play a pivotal role in regulating gene expression and because miR-122 is the most expressed miRNA in adult liver and, its role in tumorigenesis has been investigated. Indeed, miR-122 levels have been shown to vary in HCC, sometimes with opposite trend, probably reflecting different etiologies.[79-84] Generally, a decrease in miR-122 levels in hepatocarcinoma has been documented, although one study reported an up-regulation in HCV-associated HCC.[82] miR-122 is considered a tumor-suppressor within the liver. Actually, Cyclin G1 was identified as a target of miR-122[80] and this repression affected doxorubicin sensitivity of human hepatocarcinoma cells.[85] Cyclin G1 is involved in the regulation of cell cycle progression and it is overexpressed in a number of tumors, providing a direct link between miR-122 and tumorigenesis. In addition, the loss of miR-122 expression in HCC cells was associated with increased cell migration and invasiveness. The reversion of these tumorigenic properties was obtained through reintroduction of miR-122.[86]

Within the liver, miR-122 has a key role in regulating the metabolism of lipids. In mice treated with anti-sense oligonucleotides, silencing of miR-122 resulted in a decreased expression of genes involved in lipid metabolism and a substantial decrease of plasma cholesterol and fatty-acid levels.[87, 88] Moreover, a significant reduction in hepatic steatosis was observed in high-fat fed mice after miR-122 inhibition.[87] These results point out the crucial function of this miRNA in liver and make miR-122 an interesting therapeutic target. Indeed, a study in chimpanzees reported no adverse effects after prolonged treatment with an antisense oligonucleotide,[78] although extreme care should be used in translating these findings in humans. Actually, miR-122 is regarded as an onco-

suppressor within the hepatocyte and it is not known whether possible tumorigenic complications could arise after prolonged treatment.

Another outstanding advance in understanding the interplay between host immunogenetics and viral clearance was the discovery of the association between the IL28B locus on chromosome 19 and the spontaneous or treatment-induced viral clearance. IL28B encodes for intereferon-λ3 (IFN-λ3), a member of the type-3 interferon group. This cytokine is structurally related to the IL10 family, but is functionally related to type-1 IFN.[89] Actually, it is an antiviral cytokine involved in innate immune responses that functions via the JAK/STAT signaling pathway in regulating the expression of IFN-stimulated genes (ISGs) and thus suppressing viral infections. Indeed, recent evidence showed that IFN-λ suppresses HCV replication[90-93] and early clinical trials reported a successful antiviral effect of pegylated-IFN-λ1 plus ribavirin in treatment-naïve patients.[94]

The identification of IL28B as a significant predictor of treatment-induced viral clearance resulted from various GWASs performed in population of different ethnicity.[95-97] These findings were further validated and the association with spontaneous viral clearance was also demonstrated.[98-100] The first GWAS performed by Ge et al.[95] reported the association between rs12979860 a Sustained Virological Response (SVR) in patients with genotype 1 and treated with PegIFN-α_{2a} or PegIFN-α_{2b}. An approximate twofold change in response to treatment was observed in patients with the SNP rs12979860 CC genotype compared with the TT genotype. The reported Odds Ratio (OR) was 7.3 in European-Americans, 7.1 in African-Americans and 5.6 in Hispanics. The SNP rs12979860 is located ~3 kbp upstream the IL28B gene and resulted the strongest predictor of SVR, compared to baseline fibrosis or baseline viral load. The C allele is more common in East Asians, less common in African and intermediate in Europeans. This differential representation positively correlated with the rates of SVR in population of different ethnicity. Using different genotyping platforms, Tanaka et al. and Suppiah et al. found an association between rs8099917 and response to therapy in Japanese and Australians of European descent, respectively.[96, 97] rs8099917 is located ~8 kbp

upstream the IL28B gene and was reported to be in linkage disequilibrium with rs12979860.[95] The effect size varied between different populations, reflecting both distinct genetic backgrounds but also different study designs. However, when adjusting for rs12979860, the association remained significant for rs8099917, suggesting a possible independent effect for this variant.[95] Both rs12979860 and rs8099917 have also been associated with spontaneous viral clearance using a focused approach or a GWAS,[98, 99] further indicating the importance of innate immunity in HCV infection.

Indeed, the actual causative variants have not been yet identified, although targeted sequencing of the IL28B region revealed two possible candidates: a non-synonymous substitution (rs8103142, K70A) and a variant in the promoter region of IL28B (rs28416813). Recently, gene expression has been linked to IL28B genotypes, thus providing a direct link between genotype and function. Indeed, lower pre-treatment intrahepatic expression levels of ISGs have been reported for the genotypes of rs8099917 and rs12979860 associated with SVR.[77, 101] Conversely, no association was reported between the IL28B mRNA expression and different genotypes.[77] Hence, the emerging picture is the following: a) subjects with a IL28B genotype associated to SVR have a lower pre-treatment activation of ISGs; b) this reduced expression of ISGs in treatment-naïve patients may account for the higher viral loads observed in those carrying the IL28 genotype associated with SVR;[95] c) the IL28B variants associated with a poor response correspond to higher pre-treatment levels of ISGs, a condition that in principle might impair a strong response to exogenous IFN during therapy; d) the therapy outcome seems not to be influenced by differential expression of IL28B gene in groups with different IL28B genotype. Finally, the exact interplay between type-1 and type-3 IFNs remains to be fully elucidated and a mechanistic model to accurately explain this scenario is currently unavailable.

A short-term translational value of these finding could be the individualization of the therapy in accordance to the host genotype. A significant number of patients fail to respond to PegIFN plus ribavirin or have adverse effects, including influenza-like symptoms, haematologic abnormalities and

neuropsychiatric symptoms. Therefore, it is of outstanding interest to predict the therapeutic outcome. A laboratory test for IL28B genotype is already available in the United States and this could in principle be used together with viral load and genotype plus other host factor to predict SVR, since a 100% predictive power is currently not achievable with the genotype data alone.

Biomarkers of fibrosis and HCC

Currently, percutaneous liver biopsy is considered the "gold standard" for assessing hepatic disease status, since it provides a unique source of information on fibrosis and assessment of histology.[2] Although life-threatening complications are rare, it is an invasive and painful procedure.[102, 103] Moreover, the results could be affected by sampling and inter-observer variability.[104] Indeed, the specimen represents a really small part of the whole liver, which may not reflect the actual lesions that are heterogeneously distributed all through the organ. Nevertheless, liver biopsies have greatly contributed to the existing knowledge about fibrosis progression and natural history of HCV infection. Various noninvasive tests of hepatic fibrosis have been examined for monitoring patients with chronic HCV infection, and currently their effective value is under debate, although recent evidence confirms the interest in non-invasive methods for the staging of hepatic fibrosis or fibrogenic activity within the liver.[105] Current non-invasive methods to evaluate liver fibrosis include either a physical approach based on the measurement of liver stiffness using transient elastography (TE, FibroScan), or a biological approach based on serum biomarkers of fibrosis.

TE using Fibroscan is based on the measurement of liver stiffness using a low-frequency wave that produces an elastic shear vibration propagating through the liver with a velocity that is proportional to liver stiffness. An ultrasound acquisition determines the propagation speed that enables the estimation of liver elasticity, expressed in kPa: the stiffer the tissue, the faster the shear wave propagates.[106] TE measures liver stiffness in a volume that is at least 100 times bigger than a biopsy specimen, and is therefore more representative of the hepatic lesions. A number of clinical studies reported the agreement between

the observed hepatic elasticity and the histological fibrosis stage and a recent meta-analysis demonstrated that TE can be performed with diagnostic accuracy, which is better for cirrhosis but reduced for fibrosis.[107-109] This test has yet some limitation, including reduced inter-observer agreement in patients with low fibrosis, steatosis or obesity.[110] Recently, TE has been reported to produce results that could not be interpreted in nearly one in five cases.[111]

While TE directly assesses liver stiffness in relation to elasticity, corresponding to an intrinsic physical property of the liver, serum biomarkers are a combination of indirect serum surrogate of liver fibrosis. They are not totally liver-specific, but were optimized to reflect the fibrosis stage as assessed by liver biopsy. The two most widely studied are the Aspartate-to-Platelet Ratio Index (APRI)[110] and the FibroTest (FibroSure in the US).[112, 113] The first measures the ratio between the aspartate transaminase and the platelet count, while the latter combines the age and gender of the patient with six serum markers: Alpha-2-macroglobulin, Haptoglobin, Apolipoprotein A1, Gamma-glutamyl transpeptidase and total bilirubin. Actually, in a large recent study comparing TE with several biomarkers of liver fibrosis, the accuracy of TE did not differ significantly from those of serum biomarkers, using liver biopsy as a reference.[109] Limitations include a critical interpretation of the results, in order to avoid false positive related to other pathological conditions.[114] The combination of TE and biomarkers has been proposed in order to increase the diagnostic accuracy, although for the diagnosis of cirrhosis TE alone appears to be the most accurate method.[115]

Once cirrhosis is established the risk of developing Hepatocellular Carcinoma HCC is 1-4% per year.[18] Indeed, HCC is the leading cause of death among cirrhotic patients and the third cause of cancer-related mortality.[116] The poor survival rate is, in part, linked to the diagnosis of HCC at advanced stages, when effective therapies are lacking. Effective surveillance of patients with high risk for developing HCC (i.e., patients with cirrhosis) could potentially decrease HCC-related mortality rate. α-fetoprotein (AFP) is the most commonly used serum biomarker of HCC. However, AFP has low specificity and sensitivity.

Indeed, its levels may be normal in up to 40% of patients with HCC, particularly during the early stages. Furthermore, elevated AFP levels may be seen in patients with cirrhosis or flares of active hepatitis.[117, 118] Thus, there is an impellent need for novel, preferably non-invasive biomarkers that could be used for the detection of HCC in order to improve the efficacy of surveillance programs for patients at high risk.

In summary, the development of non-invasive and reliable serum markers that accurately reflect the spectrum of hepatic fibrotic and cirrhotic disease is considered a clinical and research priority. Routine assessment of fibrosis every 3-5 years through liver biopsy gives invaluable information about disease progression, but it is associated with poor patient acceptance. Thus consistent non-invasive *diagnostic* biomarkers are useful in improving the clinical management of patients with chronic HCV. However, since the rate of fibrosis progression significantly differs among individuals, *prognostic* predictors would also have great value. The importance of general host factors such as age at infection, gender, immune status and alcohol consumption has long been recognized. In addition, the identification of host genetic determinants of disease progression will likely improve the accurate prediction of the hazard of progression to cirrhosis and development of HCC.

Circulating microRNA as biomarkers

Over the last two years, accumulating evidence supported the use of miRNAs circulating in blood as non-invasive biomarkers of disease conditions, particularly cancer (see Table 1). These miRNAs were stably detected in the serum or plasma of healthy and diseased subjects and their levels did not appear to be dramatically influenced by multiple freeze-thaw cycles, long term storage or treatment with RNase before extraction.[119-121] Hence, circulating miRNAs are retained within a compartment that is non accessible to RNase activity. They have been shown to reside within exosomes, small (50 – 100 nm) vesicles of endosomal origin derived from multivesicular bodies that are released from various cell types.[122] However, miRNAs have also been found in other secreted microvesicles, apoptotic/senescent bodies or protected by non-

vesicle-associated protein complexes.[123-126] Currently, the function of miRNAs in the blood circulation is not known. Given their crucial regulatory role in cell biology, a plausible role for maintaining homeostasis in the blood is proposed. miR-223, one of the most abundant miRNA in serum, is highly specific for hematopoietic cells, and precisely for the myeloid compartment.[127-129] This miRNA plays important roles in the development and function of the myeloid lineage and innate immunity, since its deletion leads to an increase in the numbers of neutrophil progenitors and mature neutrophils.[130]

In one of the first works reporting the detection of circulating miRNAs, Lawrie et al. suggested that miR-155, miR-21 and miR-210 have potential as non-invasive diagnostic markers for Diffuse Large B-Cell Lymphoma.[131] In another key study, Mitchell et al. observed that serum levels of miR-141, a miRNA expressed in prostate cancer, discriminated between patients with prostate cancers and healthy controls.[120] Additionally, Chen et al. identified over 100 commonly expressed miRNAs in the serum of healthy individuals, using RNA sequencing. They also identified specific expression patterns of serum miRNAs for lung cancer, colorectal cancer and diabetes, providing evidence that serum miRNAs contain signatures for a range of diseases.[119] The importance of circulating miRNAs was further documented by a number of successive works, showing the potential of cell-free circulating miRNAs as disease biomarkers not only for cancer (lung,[132] colorectal,[133, 134], gastric,[135] ovarian,[136] pancreatic,[137] breast[138] and skeletal muscle[139] cancer), but also for acute myocardial infartion,[140] and liver pathologies including hepatotoxicity, chronic hepatitis and HCC.[141-145]

A common pitfall of some of these works is the lack of a solid normalization strategy to account for interindividual or intergroup variability. Most of the results are obtained by Real-time qPCR, using different reference genes. When quantifying cellular miRNAs, stable small RNA controls are currently used as reference RNAs. These include small non-coding RNAs (ncRNA), and specifically small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) such as RNU44, RNU48 and U6. For serum miRNAs, there is growing evidence that the above-mentioned small RNAs are highly variable or not stably

detectable,[146] thus leading to the search of suitable stable control miRNAs that are firmly detectable in human serum. The most common choice has been the use of a miRNA that does not vary considerably between individuals. However, this selection may be arbitrary when based on the Ct data only, without the assessment of the actual stability by suitable *wet* and *in silico* analyses. The choice of the normalization strategy has direct consequences in the quantification results, making more difficult to compare directly different studies. Moreover, a direct comparison may be further complicated by the fact that the starting material (whole blood, serum, plasma or microvesicles) differs, as well as the RNA extraction methods and quantification protocols. Because miRNA content in blood circulation is very low,[147] all these variables may significantly affect the final results. In our laboratory experience, the yield of RNA from small volume plasma or serum samples (i.e., 100–400 μl) is below the limit of accurate quantification by spectrophotometry, a finding that is confirmed by other groups.[148] Hence, a normalization based on a fixed amount of starting RNA is routinely impracticable. Nevertheless, the use of a fixed amount of starting material may be even incorrect, because little is known about the relationship between the total RNA amount in serum/plasma and the corresponding miRNA fraction.

On the topic of liver disease, recent literature highlight the potential use circulating miRNAs as sensitive and informative biomarkers of liver pathological states. The differential plasma expression of miRNAs was detected after acetaminophen-induced liver injury using a mouse model.[149] Several miRNAs originating from diverse tissues, including brain, heart, lung, liver and spleen were detected in the plasma of untreated animals. Most notably, specific microRNA species, such as miR-122 and miR-192, both enriched in liver tissue, showed exposure-dependent increase in the plasma that paralleled serum aminotransferase levels and the histopathology of liver degeneration, but their changes could be detected earlier and in a very sensitive fashion. Similar results were obtained using a rat model of tissue injury.[142] These works point out the possibility that a considerable fraction of the miRNA blood content may be a reflection of a passive release from tissues, after stress or injury.

Table 1. Circulating miRNAs associated with cancer

Cancer	*Source*	*Increased*	*Decreased*	*Ref.*
Diffuse Large B Cell Lymphoma	Serum	miR-155, miR-210, miR-21		131
Non-Small Cell Lung Carcinoma	Plasma & serum	miR-25, miR-233		119
Non-Small Cell Lung Carcinoma	Serum	miR-486, miR-30d	miR-1, miR-499	132
Colorectal cancer	Plasma	miR-17-3p, miR-92		133
Colorectal cancer	Plasma	miR-29a, miR-92a		134
Ovarian Cancer	Serum	miR-21, miR-29a, miR-126, miR-92, miR-93	miR-127 miR-99b miR-155	136
Squamous Cell Carcinoma	Plasma	miR-184		150
Squamous Cell Carcinoma	Plasma	miR-24		151
Squamous Cell Carcinoma	Plasma	miR-31		152
Pancreatic cancer	Plasma	miR-210		137
Breast cancer	Whole blood	miR-195, let-7a		138
Prostate cancer	Plasma	miR-141		120

Cancer	*Source*	*Increased*	*Decreased*	*Ref.*
Prostate cancer	Serum	miR-16, miR-92a, miR-92b, miR-103, miR-107, miR-197, miR-34b, miR-328, miR-485-3p, miR-486-5p, miR-574-3p, miR-636, miR-640, miR-766, miR-885-5p		[153]
Gastric cancer	Plasma	miR-17-5p, miR-21, miR-106a, miR-106b	let-7a	[135]
Hepatocellular Carcinoma	Serum	miR-500		[154]
Hepatocellular Carcinoma	Serum	miR-21, miR-122, miR-223		[144]
Hepatocellular Carcinoma	Serum	miR-885-5p		[141]
Glioblastoma	Exosomes	miR-21		[155]

Scope of the thesis

The goal of this work was to discover effective biomarkers of disease progression in the setting of chronic HCV infection.

- Chapter 1. The contribution of the host genetics to the fibrosis progression rate was investigated through a genome-wide association study, in a well-characterized cohort with known date of infection.

- Chapter 2. Circulating microRNAs were evaluated as biomarkers for monitoring HCV disease progression, defining serum miRNA signatures associated with chronic hepatitis, liver cirrhosis and hepatocellular carcinoma.

References

1. World Health Organization. Hepatitis C Fact Sheet No. 164 (Revised Oct 2000).
2. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002--June 10-12, 2002. Hepatology 2002;36:S3-20.
3. Brown RS. Hepatitis C and liver transplantation. Nature 2005;436:973-978.
4. Houghton M. Prospects for prophylactic and therapeutic vaccines against the hepatitis C viruses. Immunol Rev 2011;239:99-108.
5. Poynard T, Yuen M-F, Ratziu V, Lai CL. Viral hepatitis C. Lancet 2003;362:2095-100.
6. Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nat Rev Micro 2007;5:453-63.
7. Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, Weiner AJ, Houghton M, Rosa D, Grandi G, Abrignani S. Binding of hepatitis C virus to CD81. Science 1998;282:938-41.
8. Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, Filocamo G, Traboni C, Nicosia A, Cortese R, Vitelli A. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. EMBO J 2002;21:5017-25.
9. Evans MJ, von Hahn T, Tscherne DM, Syder AJ, Panis M, Wolk B, Hatziioannou T, McKeating JA, Bieniasz PD, Rice CM. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature 2007;446:801-5.
10. Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, Rice CM. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature 2009;457:882-6.
11. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 1989;244:359-62.
12. Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Krausslich HG, Mizokami M, Bartenschlager R, Liang TJ. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat Med 2005;11:791-6.
13. Zhong J, Gastaminza P, Cheng G, Kapadia S, Kato T, Burton DR, Wieland SF, Uprichard SL, Wakita T, Chisari FV. Robust hepatitis C virus infection in vitro. Proc Natl Acad Sci U S A 2005;102:9294-9.
14. Lindenbach BD, Meuleman P, Ploss A, Vanwolleghem T, Syder AJ, McKeating JA, Lanford RE, Feinstone SM, Major ME, Leroux-Roels G, Rice CM. Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. Proc Natl Acad Sci U S A 2006;103:3805-9.

15. Boonstra A, van der Laan LJ, Vanwolleghem T, Janssen HL. Experimental models for hepatitis C viral infection. Hepatology 2009;50:1646-55.
16. von Hahn T, Yoon JC, Alter H, Rice CM, Rehermann B, Balfe P, McKeating JA. Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. Gastroenterology 2007;132:667-78.
17. Simmonds P. Genetic diversity and evolution of hepatitis C virus--15 years on. J Gen Virol 2004;85:3173-88.
18. Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med 2001;345:41-52.
19. Hoofnagle JH. Course and outcome of hepatitis C. Hepatology 2002;36:S21-9.
20. Seeff LB. Natural history of chronic hepatitis C. Hepatology 2002;36:S35-46.
21. Lefkowitch JH. Liver biopsy assessment in chronic hepatitis. Archives of Medical Research 2007;38:634-43.
22. Friedman SL. Hepatic fibrosis -- overview. Toxicology 2008;254:120-9.
23. Friedman SL. Mechanisms of hepatic fibrogenesis. Gastroenterology 2008;134:1655-69.
24. Mann J, Mann DA. Transcriptional regulation of hepatic stellate cells. Advanced Drug Delivery Reviews 2009;61:497-512.
25. Tsukada S, Parsons CJ, Rippe RA. Mechanisms of liver fibrosis. Clin Chim Acta 2006;364:33-60.
26. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med 2000;342:1350-8.
27. Gordon KJ, Blobe GC. Role of transforming growth factor-beta superfamily signaling pathways in human disease. Biochim Biophys Acta 2008;1782:197-228.
28. Missiha SB, Ostrowski M, Heathcote EJ. Disease progression in chronic hepatitis C: modifiable and nonmodifiable factors. Gastroenterology 2008;134:1699-714.
29. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;349:825-32.
30. Poynard T, Ratziu V, Charlotte F, Goodman Z, McHutchison J, Albrecht J. Rates and risk factors of liver fibrosis progression in patients with chronic hepatitis c. JOURNAL OF HEPATOLOGY 2001;34:730-9.
31. Wright M, Goldin R, Fabre A, Lloyd J, Thomas H, Trepo C, Pradat P, Thursz M. Measurement and determinants of the natural history of liver fibrosis in hepatitis C virus infection: a cross sectional and longitudinal study. Gut 2003;52:574-9.
32. Minola E, Prati D, Suter F, Maggiolo F, Caprioli F, Sonzogni A, Fraquelli M, Paggi S, Conte D. Age at infection affects the long-term outcome of transfusion-associated chronic hepatitis C. Blood 2002;99:4588-91.
33. Casiraghi MA, De Paschale M, Romanò L, Biffi R, Assi A, Binelli G, Zanetti AR. Long-term outcome (35 years) of hepatitis C after

acquisition of infection through mini transfusions of blood given at birth. Hepatology 2004;39:90-6.
34. Tovo PA, Pembrey LJ, Newell ML. Persistence rate and progression of vertically acquired hepatitis C infection. European Paediatric Hepatitis C Virus Infection. J Infect Dis 2000;181:419-24.
35. Zarski JP, Bohn B, Bastie A, Pawlotsky JM, Baud M, Bost-Bezeaux F, Tran van Nhieu J, Seigneurin JM, Buffet C, Dhumeaux D. Characteristics of patients with dual infection by hepatitis B and C viruses. J Hepatol 1998;28:27-33.
36. Chu CJ, Lee SD. Hepatitis B virus/hepatitis C virus coinfection: epidemiology, clinical features, viral interactions and treatment. J Gastroenterol Hepatol 2008;23:512-20.
37. Bellecave P, Gouttenoire J, Gajer M, Brass V, Koutsoudakis G, Blum HE, Bartenschlager R, Nassal M, Moradpour D. Hepatitis B and C virus coinfection: a novel model system reveals the absence of direct viral interference. Hepatology 2009;50:46-55.
38. Lauer GM, Nguyen TN, Day CL, Robbins GK, Flynn T, McGowan K, Rosenberg ES, Lucas M, Klenerman P, Chung RT, Walker BD. Human immunodeficiency virus type 1-hepatitis C virus coinfection: intraindividual comparison of cellular immune responses against two persistent viruses. J Virol 2002;76:2817-26.
39. Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N, Nolt K, Nelson KE, Strathdee SA, Johnson L, Laeyendecker O, Boitnott J, Wilson LE, Vlahov D. The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA 2000;284:450-6.
40. Macias J, Berenguer J, Japon MA, Giron JA, Rivero A, Lopez-Cortes LF, Moreno A, Gonzalez-Serrano M, Iribarren JA, Ortega E, Miralles P, Mira JA, Pineda JA. Fast fibrosis progression between repeated liver biopsies in patients coinfected with human immunodeficiency virus/hepatitis C virus. Hepatology 2009;50:1056-63.
41. Thomas DL. Hepatitis C and human immunodeficiency virus infection. Hepatology 2002;36:S201-9.
42. Cross TJ, Quaglia A, Hughes S, Joshi D, Harrison PM. The impact of hepatic steatosis on the natural history of chronic hepatitis C infection. J Viral Hepat 2009;16:492-9.
43. Leandro G, Mangia A, Hui J, Fabris P, Rubbia-Brandt L, Colloredo G, Adinolfi LE, Asselah T, Jonsson JR, Smedile A, Terrault N, Pazienza V, Giordani MT, Giostra E, Sonzogni A, Ruggiero G, Marcellin P, Powell EE, George J, Negro F, Group HM-AoIPDS. Relationship between steatosis, inflammation, and fibrosis in chronic hepatitis C: a meta-analysis of individual patient data. Gastroenterology 2006;130:1636-42.
44. Rubbia-Brandt L, Fabris P, Paganin S, Leandro G, Male P-J, Giostra E, Carlotto A, Bozzola L, Smedile A, Negro F. Steatosis affects chronic hepatitis C progression in a genotype specific way. Gut 2004;53:406-12.
45. Su AI, Pezacki JP, Wodicka L, Brideau AD, Supekova L, Thimme R, Wieland S, Bukh J, Purcell RH, Schultz PG, Chisari FV. Genomic analysis

of the host response to hepatitis C virus infection. Proc Natl Acad Sci USA 2002;99:15669-74.
46. Patel K, Zekry A, McHutchison JG. Steatosis and chronic hepatitis C virus infection: mechanisms and significance. Clinics in Liver Disease 2005;9:399-410, vi.
47. Osterreicher CH, Stickel F, Brenner DA. Genomics of liver fibrosis and cirrhosis. Semin Liver Dis 2007;27:28-43.
48. Huang Y, Yang H, Borg B, Su X, Rhodes S, Yang K, Tong X, Tang G, Howell C, Rosen H, Thio C, Thomas D, Alter H, Sapp R, Liang T. A functional SNP of interferon-gamma gene is important for interferon-alpha-induced and spontaneous recovery from hepatitis C virus infection. Proc Natl Acad Sci U S A 2007;104:985-90.
49. Dai C, Chuang W, Lee L, Chen S, Hou N, Lin Z, Hsieh M, Hsieh M, Wang L, Chang W, Yu M. Associations of tumour necrosis factor alpha promoter polymorphisms at position -308 and -238 with clinical characteristics of chronic hepatitis C. J Viral Hepat 2006;13:770-4.
50. Yee LJ, Tang J, Herrera J, Kaslow RA, van Leeuwen DJ. Tumor necrosis factor gene polymorphisms in patients with cirrhosis from chronic hepatitis C virus infection. Genes Immun 2000;1:386-90.
51. Falleti E, Fabris C, Toniutto P, Fontanini E, Cussigh A, Bitetto D, Fornasiere E, Avellini C, Minisini R, Pirisi M. TGF-beta1 genotypes in cirrhosis: relationship with the occurrence of liver cancer. Cytokine 2008;44:256-61.
52. Powell EE, Edwards-Smith CJ, Hay JL, Clouston AD, Crawford DH, Shorthouse C, Purdie DM, Jonsson JR. Host genetic factors influence disease progression in chronic hepatitis C. Hepatology 2000;31:828-33.
53. Huang H, Shiffman ML, Friedman S, Venkatesh R, Bzowej N, Abar OT, Rowland CM, Catanese JJ, Leong DU, Sninsky JJ, Layden TJ, Wright TL, White T, Cheung RC. A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C. Hepatology 2007;46:297-306.
54. Marcolongo M, Young B, Dal Pero F, Fattovich G, Peraro L, Guido M, Sebastiani G, Palù G, Alberti A. A seven-gene signature (cirrhosis risk score) predicts liver fibrosis progression in patients with initially mild chronic hepatitis C. Hepatology 2009;50:1038-44.
55. Patel K, Norris S, Lebeck L, Feng A, Clare M, Pianko S, Portmann B, Blatt LM, Koziol J, Conrad A, McHutchison JG. HLA class I allelic diversity and progression of fibrosis in patients with chronic hepatitis C. Hepatology 2006;43:241-9.
56. Yee L. Host genetic determinants in hepatitis C virus infection. Genes Immun 2004;5:237-45.
57. Hue S, Cacoub P, Renou C, Halfon P, Thibault V, Charlotte F, Picon M, Rifflet H, Piette JC, Pol S, Caillat-Zucman S. Human leukocyte antigen class II alleles may contribute to the severity of hepatitis C virus-related liver disease. J Infect Dis 2002;186:106-9.

58. Neumann-Haefelin C, Thimme R. Impact of the genetic restriction of virus-specific T-cell responses in hepatitis C virus infection. Genes Immun 2007;8:181-92.
59. Kobayashi M, Tanaka E, Sodeyama T, Urushihara A, Matsumoto A, Kiyosawa K. The natural course of chronic hepatitis C: a comparison between patients with genotypes 1 and 2 hepatitis C viruses. Hepatology 1996;23:695-9.
60. Bochud PY, Cai T, Overbeck K, Bochud M, Dufour JF, Mullhaupt B, Borovicka J, Heim M, Moradpour D, Cerny A, Malinverni R, Francioli P, Negro F. Genotype 3 is associated with accelerated fibrosis progression in chronic hepatitis C. J Hepatol 2009;51:655-66.
61. Zein NN. Clinical significance of hepatitis C virus genotypes. Clin Microbiol Rev 2000;13:223-35.
62. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281-97.
63. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 2005;6:376-85.
64. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 2008;9:22-32.
65. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215-33.
66. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009;19:92-105.
67. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature 2005;435:834-8.
68. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA 2006;103:2257-61.
69. O'connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. Nature Reviews Immunology 2010;10:111-22.
70. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science 2005;309:1577-81.
71. Chang J, Nicolas E, Marks D, Sander C, Lerro A, Buendia MA, Xu C, Mason WS, Moloshok T, Bort R, Zaret KS, Taylor JM. miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. RNA Biol 2004;1:106-13.
72. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Curr Biol 2002;12:735-9.

73. Henke JI, Goergen D, Zheng J, Song Y, Schüttler CG, Fehr C, Jünemann C, Niepmann M. microRNA-122 stimulates translation of hepatitis C virus RNA. EMBO J 2008;27:3300-10.
74. Chang J, Guo JT, Jiang D, Guo H, Taylor JM, Block TM. Liver-specific microRNA miR-122 enhances the replication of hepatitis C virus in nonhepatic cells. J Virol 2008;82:8215-23.
75. Randall G, Panis M, Cooper JD, Tellinghuisen TL, Sukhodolets KE, Pfeffer S, Landthaler M, Landgraf P, Kan S, Lindenbach BD, Chien M, Weir DB, Russo JJ, Ju J, Brownstein MJ, Sheridan R, Sander C, Zavolan M, Tuschl T, Rice CM. Cellular cofactors affecting hepatitis C virus infection and replication. Proc Natl Acad Sci USA 2007;104:12884-9.
76. Sarasin-Filipowicz M, Krol J, Markiewicz I, Heim MH, Filipowicz W. Decreased levels of microRNA miR-122 in individuals with hepatitis C responding poorly to interferon therapy. Nat Med 2009;15:31-3.
77. Urban TJ, Thompson AJ, Bradrick SS, Fellay J, Schuppan D, Cronin KD, Hong L, McKenzie A, Patel K, Shianna KV, McHutchison JG, Goldstein DB, Afdhal N. IL28B genotype is associated with differential expression of intrahepatic interferon-stimulated genes in patients with chronic hepatitis C. Hepatology 2010.
78. Lanford RE, Hildebrandt-Eriksen ES, Petri A, Persson R, Lindow M, Munk ME, Kauppinen S, Orum H. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. Science 2010;327:198-201.
79. Kutay H, Bai S, Datta J, Motiwala T, Pogribny I, Frankel W, Jacob ST, Ghoshal K. Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. J Cell Biochem 2006;99:671-8.
80. Gramantieri L, Ferracin M, Fornari F, Veronese A, Sabbioni S, Liu C-G, Calin GA, Giovannini C, Ferrazzi E, Grazi GL, Croce CM, Bolondi L, Negrini M. Cyclin G1 is a target of miR-122a, a microRNA frequently down-regulated in human hepatocellular carcinoma. Cancer Research 2007;67:6092-9.
81. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. Gastroenterology 2007;133:647-58.
82. Varnholt H, Drebber U, Schulze F, Wedemeyer I, Schirmacher P, Dienes H-P, Odenthal M. MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. Hepatology 2008;47:1223-32.
83. Ura S, Honda M, Yamashita T, Ueda T, Takatori H, Nishino R, Sunakozaka H, Sakai Y, Horimoto K, Kaneko S. Differential microRNA expression between hepatitis B and hepatitis C leading disease progression to hepatocellular carcinoma. Hepatology 2009;49:1098-112.
84. Chen X-M. MicroRNA signatures in liver diseases. World J Gastroenterol 2009;15:1665-72.
85. Fornari F, Gramantieri L, Giovannini C, Veronese A, Ferracin M, Sabbioni S, Calin GA, Grazi GL, Croce CM, Tavolari S, Chieco P, Negrini M, Bolondi L. MiR-122/cyclin G1 interaction modulates p53 activity

and affects doxorubicin sensitivity of human hepatocarcinoma cells. Cancer Research 2009;69:5761-7.
86. Coulouarn C, Factor VM, Andersen JB, Durkin ME, Thorgeirsson SS. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. Oncogene 2009;28:3526-36.
87. Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. Silencing of microRNAs in vivo with 'antagomirs'. Nature 2005;438:685-9.
88. Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, Watts L, Booten SL, Graham M, McKay R, Subramaniam A, Propp S, Lollo BA, Freier S, Bennett CF, Bhanot S, Monia BP. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. Cell Metab 2006;3:87-98.
89. Gad HH, Dellgren C, Hamming OJ, Vends S, Paludan SR, Hartmann R. Interferon-lambda is functionally an interferon but structurally related to the interleukin-10 family. Journal of Biological Chemistry 2009;284:20869-75.
90. Robek MD, Boyd BS, Chisari FV. Lambda interferon inhibits hepatitis B and C virus replication. J Virol 2005;79:3851-4.
91. Marcello T, Grakoui A, Barba-Spaeth G, Machlin ES, Kotenko SV, MacDonald MR, Rice CM. Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. Gastroenterology 2006;131:1887-98.
92. Pagliaccetti NE, Robek MD. Interferon-lambda in the immune response to hepatitis B virus and hepatitis C virus. J Interferon Cytokine Res 2010;30:585-90.
93. Diegelmann J, Beigel F, Zitzmann K, Kaul A, Goke B, Auernhammer CJ, Bartenschlager R, Diepolder HM, Brand S. Comparative analysis of the lambda-interferons IL-28A and IL-29 regarding their transcriptome and their antiviral properties against hepatitis C virus. PLoS ONE 2010;5:e15200.
94. Muir AJ, Shiffman ML, Zaman A, Yoffe B, de la Torre A, Flamm S, Gordon SC, Marotta P, Vierling JM, Lopez-Talavera JC, Byrnes-Blake K, Fontana D, Freeman J, Gray T, Hausman D, Hunder NN, Lawitz E. Phase 1b study of pegylated interferon lambda 1 with or without ribavirin in patients with chronic genotype 1 hepatitis C virus infection. Hepatology 2010;52:822-32.
95. Ge D, Fellay J, Thompson A, Simon J, Shianna K, Urban T, Heinzen E, Qiu P, Bertelsen A, Muir A, Sulkowski M, McHutchison J, Goldstein D. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009.
96. Study tHC, Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate M, Bassendine M, Spengler U, Dore G, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Müller T, Bahlo M, Stewart G, Booth D, George J. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nature Genetics 2009.

97. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nature Genetics 2009.
98. Thomas D, Thio C, Martin M, Qi Y, Ge D, O'huigin C, Kidd J, Kidd K, Khakoo S, Alexander G, Goedert J, Kirk G, Donfield S, Rosen H, Tobler L, Busch M, McHutchison J, Goldstein D, Carrington M. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009.
99. Rauch A, Kutalik Z, Descombes P, Cai T, di Iulio J, Mueller T, Bochud M, Battegay M, Bernasconi E, Borovicka J, Colombo S, Cerny A, Dufour J-F, Furrer H, Günthard HF, Heim M, Hirschel B, Malinverni R, Moradpour D, Müllhaupt B, Witteck A, Beckmann JS, Berg T, Bergmann S, Negro F, Telenti A, Bochud P-Y, Studies SHCaHC. Genetic Variation in IL28B Is Associated With Chronic Hepatitis C and Treatment Failure: A Genome-wide Association Study. Gastroenterology 2010.
100. McCarthy JJ, Li JH, Thompson A, Suchindran S, Lao XQ, Patel K, Tillmann HL, Muir AJ, McHutchison JG. Replicated association between an IL28B gene variant and a sustained response to pegylated interferon and ribavirin. Gastroenterology 2010;138:2307-14.
101. Honda M, Sakai A, Yamashita T, Nakamoto Y, Mizukoshi E, Sakai Y, Nakamura M, Shirasaki T, Horimoto K, Tanaka Y, Tokunaga K, Mizokami M, Kaneko S. Hepatic ISG expression is associated with genetic variation in interleukin 28B and the outcome of IFN therapy for chronic hepatitis C. Gastroenterology 2010;139:499-509.
102. Piccinino F, Sagnelli E, Pasquale G, Giusti G. Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. J Hepatol 1986;2:165-73.
103. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med 2001;344:495-500.
104. Skripenova S, Trainer TD, Krawitt EL, Blaszyk H. Variability of grade and stage in simultaneous paired liver biopsies in patients with hepatitis C. J Clin Pathol 2007;60:321-4.
105. Castera L, Pinzani M. Non-invasive assessment of liver fibrosis: are we ready? Lancet 2010;375:1419-20.
106. Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, Christidis C, Ziol M, Poulet B, Kazemi F, Beaugrand M, Palau R. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound Med Biol 2003;29:1705-13.
107. Castera L. Transient elastography and other noninvasive tests to assess hepatic fibrosis in patients with viral hepatitis. J Viral Hepat 2009;16:300-14.
108. Friedrich-Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, Zeuzem S, Herrmann E. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology 2008;134:960-74.

109. Degos F, Perez P, Roche B, Mahmoudi A, Asselineau J, Voitot H, Bedossa P. Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study). J Hepatol 2010;53:1013-21.
110. Fraquelli M, Rigamonti C, Casazza G, Conte D, Donato MF, Ronchi G, Colombo M. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. Gut 2007;56:968-73.
111. Castera L, Bedossa P. How to assess liver fibrosis in chronic hepatitis C: serum markers or transient elastography vs. liver biopsy? Liver Int 2011;31 Suppl 1:13-7.
112. Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, Lok AS. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38:518-26.
113. Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poynard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study. Lancet 2001;357:1069-75.
114. Poynard T, Munteanu M, Imbert-Bismut F, Charlotte F, Thabut D, Le Calvez S, Messous D, Thibault V, Benhamou Y, Moussalli J, Ratziu V. Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C. Clin Chem 2004;50:1344-55.
115. Castera L, Sebastiani G, Le Bail B, de Ledinghen V, Couzigou P, Alberti A. Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. J Hepatol 2010;52:191-8.
116. Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr
117. Daniele B, Bencivenga A, Megna AS, Tinessa V. Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma. Gastroenterology 2004;127:S108-12.
118. Gebo KA, Chander G, Jenckes MW, Ghanem KG, Herlong HF, Torbenson MS, El-Kamary SS, Bass EB. Screening tests for hepatocellular carcinoma in patients with chronic hepatitis C: a systematic review. Hepatology 2002;36:S84-92.
119. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, Zhang Y, Wang J, Jiang X, Xiang Y, Xu C, Zheng P, Zhang J, Li R, Zhang H, Shang X, Gong T, Ning G, Wang J, Zen K, Zhang J, Zhang C-Y. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997-1006.
120. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R,

Vessella RL, Nelson PS, Martin DB, Tewari M. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA 2008;105:10513-8.
121. Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, Chajut A. Serum microRNAs are promising novel biomarkers. PLoS ONE 2008;3:e3148.
122. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol 2007;9:654-9.
123. Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, Xiao T, Schafer J, Lee M-LT, Schmittgen TD, Nana-Sinkam SP, Jarjoura D, Marsh CB. Detection of microRNA expression in human peripheral blood microvesicles. PLoS ONE 2008;3:e3694.
124. Lehmann BD, Paine MS, Brooks AM, McCubrey JA, Renegar RH, Wang R, Terrian DM. Senescence-associated exosome release from human prostate cancer cells. Cancer Res 2008;68:7864-71.
125. Skog J, Würdinger T, Van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, Curry WT, Carter BS, Krichevsky AM, Breakefield XO. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol 2008;10:1470-6.
126. Zernecke A, Bidzhekov K, Noels H, Shagdarsuren E, Gan L, Denecke B, Hristov M, Koppel T, Jahantigh MN, Lutgens E, Wang S, Olson EN, Schober A, Weber C. Delivery of microRNA-126 by apoptotic bodies induces CXCL12-dependent vascular protection. Sci Signal 2009;2:ra81.
127. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foà R, Schliwka J, Fuchs U, Novosel A, Müller R-U, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter H-I, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007;129:1401-14.
128. Chen C-Z, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004;303:83-6.
129. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, Wei L, Laurence A, Yasuda T, Peng S, Hu-Li J, Lu K, Dubois W, Kitamura Y, Charles N, Sun H-w, Muljo S, Schwartzberg PL, Paul WE, O'shea J, Rajewsky K, Casellas R. Regulation of microRNA expression and abundance during lymphopoiesis. Immunity 2010;32:828-39.
130. Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, Brummelkamp TR, Fleming MD, Camargo FD. Regulation of progenitor

cell proliferation and granulocyte function by microRNA-223. Nature 2008;451:1125-9.
131. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, Banham AH, Pezzella F, Boultwood J, Wainscoat JS, Hatton CSR, Harris AL. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol 2008;141:672-5.
132. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, Shen H. Serum MicroRNA Signatures Identified in a Genome-Wide Serum MicroRNA Expression Profiling Predict Survival of Non-Small-Cell Lung Cancer. Journal of Clinical Oncology 2010.
133. Ng EKO, Chong WWS, Jin H, Lam EKY, Shin VY, Yu J, Poon TCW, Ng SSM, Sung JJY. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut 2009;58:1375-81.
134. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. International journal of cancer Journal international du cancer 2009.
135. Tsujiura M, Ichikawa D, Komatsu S, Shiozaki A, Takeshita H, Kosuga T, Konishi H, Morimura R, Deguchi K, Fujiwara H, Okamoto K, Otsuji E. Circulating microRNAs in plasma of patients with gastric cancers. Br J Cancer 2010;102:1174-9.
136. Resnick KE, Alder H, Hagan JP, Richardson DL, Croce CM, Cohn DE. The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. Gynecologic Oncology 2009;112:55-9.
137. Ho AS, Huang X, Cao H, Christman-Skieller C, Bennewith K, Le QT, Koong AC. Circulating miR-210 as a Novel Hypoxia Marker in Pancreatic Cancer. Transl Oncol 2010;3:109-13.
138. Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin MJ. Circulating microRNAs as Novel Minimally Invasive Biomarkers for Breast Cancer. Annals of Surgery 2010;251:499-505.
139. Miyachi M, Tsuchiya K, Yoshida H, Yagyu S, Kikuchi K, Misawa A, Iehara T, Hosoi H. Circulating muscle-specific microRNA, miR-206, as a potential diagnostic marker for rhabdomyosarcoma. Biochemical and Biophysical Research Communications 2010.
140. Dimmeler S, Zeiher AM. Circulating microRNAs: novel biomarkers for cardiovascular diseases? European Heart Journal 2010.
141. Gui J, Tian Y, Wen X, Zhang W, Zhang P, Gao J, Run W, Tian L, Jia X, Gao Y. Serum microRNA characterization identifies miR-885-5p as a potential marker for detecting liver pathologies. Clin. Sci. 2011;120:183-93.
142. Laterza OF, Lim L, Garrett-Engele PW, Vlasakova K, Muniappa N, Tanaka WK, Johnson JM, Sina JF, Fare TL, Sistare FD, Glaab WE. Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clinical Chemistry 2009;55:1977-83.
143. Qu KZ, Zhang K, Li H, Afdhal NH, Albitar M. Circulating MicroRNAs as Biomarkers for Hepatocellular Carcinoma. J Clin Gastroenterol 2011.

144. Xu J, Wu C, Che X, Wang L, Yu D, Zhang T, Huang L, Li H, Tan W, Wang C, Lin D. Circulating MicroRNAs, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. Mol. Carcinog. 2010.
145. Zhang Y, Jia Y, Zheng R, Guo Y, Wang Y, Guo H, Fei M, Sun S. Plasma MicroRNA-122 as a Biomarker for Viral-, Alcohol-, and Chemical-Related Hepatic Diseases. Clinical Chemistry 2010.
146. Reid G, Kirschner MB, van Zandwijk N. Circulating microRNAs: Association with disease and potential use as biomarkers. Critical reviews in oncology/hematology 2010:In press.
147. Kroh EM, Parkin RK, Mitchell PS, Tewari M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). Methods 2010;50:298-301.
148. Kroh EM, Parkin RK, Mitchell PS, Tewari M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). Methods (San Diego, Calif) 2010;50:298-301.
149. Wang K, Zhang S, Marzolf B, Troisch P, Brightman A, Hu Z, Hood LE, Galas DJ. Circulating microRNAs, potential biomarkers for drug-induced liver injury. Proc Natl Acad Sci USA 2009;106:4402-7.
150. Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP, Wei WI. Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue. Clin Cancer Res 2008;14:2588-92.
151. Lin SC, Liu CJ, Lin JA, Chiang WF, Hung PS, Chang KW. miR-24 up-regulation in oral carcinoma: positive association from clinical and in vitro analysis. Oral Oncol 2010;46:204-8.
152. Liu CJ, Kao SY, Tu HF, Tsai MM, Chang KW, Lin SC. Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. Oral Dis 2010;16:360-4.
153. Lodes MJ, Caraballo M, Suciu D, Munro S, Kumar A, Anderson B. Detection of cancer with serum miRNAs on an oligonucleotide microarray. PLoS ONE 2009;4:e6229.
154. Yamamoto Y, Kosaka N, Tanaka M, Koizumi F, Kanai Y, Mizutani T, Murakami Y, Kuroda M, Miyajima A, Kato T, Ochiya T. MicroRNA-500 as a potential diagnostic marker for hepatocellular carcinoma. Biomarkers 2009;14:529-38.
155. Skog J, Wurdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, Curry WT, Jr., Carter BS, Krichevsky AM, Breakefield XO. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol 2008;10:1470-6.

Chapter 2. A Genome-Wide Association Study identifies variants showing suggestive evidence for association with liver fibrosis progression in patients with chronic hepatitis C

The data presented in this chapter were the starting material for a manuscript now submitted for publication to Hepatology:

Marabita F, Aghemo A, De Nicola S, Rumi MG, Cheroni C, Scavelli R, Crimi M, Soffredini R, Abrignani S, De Francesco R and Colombo M. *Genetic variation in IL28B gene is not associated with fibrosis progression in patients with chronic hepatitis C and known date of infection*. Submitted 02/18/2011.

Abstract

Hepatitis C Virus (HCV) is a pathogen that establishes a persistent infection in more than 70% of the cases. The chronic infection usually is associated with development of liver cirrhosis and hepatocellular carcinoma (HCC), with individual variable speed of progression from low fibrosis to cirrhosis. A number of host and viral factors have been previously associated to an increased speed of liver fibrosis progression, although the contribute of the host genetic variability remains to be fully elucidated. Therefore, this project aimed at the identification of SNPs associated to the progression of liver disease in chronic HCV infection.

Methods. 247 patients with an estimated date of HCV infection were included in this study. The fibrosis stage was evaluated through a percutaneous liver biopsy performed before any antiviral treatment. No patient had a history of alcohol abuse or co-infection with other viruses. The phenotypic classification was obtained by calculating the Fibrosis Progression Rate (ISHAK score/duration of disease). Clinical covariates previously associated with the natural history of HCV infection were analyzed. Genotyping was performed using Illumina 610-Quad BeadChip or 660W-Quad BeadChip. After accurate quality control of the genotype data, the association with fibrosis progression was tested for each of the 536,611 SNPs, using a linear model, a logistic regression or a survival analysis. All associations were adjusted for the age at infection, the gender and the viral genotype. The absence of population substructure was evaluated with Principal Component Analysis. In order to assess genetic variation at untyped SNP loci, we also performed genotype imputation.

Results. None of the tested SNPs reached a genome-wide level of significance ($p<10^{-7}$) to claim a definitive association. However, a list of candidate SNPs with suggestive evidence was obtained for upcoming validation. Above all, indicative evidence exists for a region containing the *GADD45G* gene (rs7871395, p=1.57E-06, OR=3.8, 95%CI=2.2 – 6.7). Conversely, the analysis of covariates revealed that they had a profound impact on the progression of the liver

disease. Indeed, the age at infection had a marked effect on fibrosis progression using both a linear model and Cox-proportional hazard regression ($p < 2E-16$). An 11.8% increase in the hazard of advanced fibrosis was estimated for each additional year at infection, suggesting that this is the major explanatory variable in our cohort. Male gender ($p < 0.05$) and HCV genotype 3 ($p < 0.01$) were also associated to faster fibrosis progression.

Conclusion. Suggestive evidence exists for the association of candidate SNPs with fibrosis progression. The relative limited sample size of this study does not allow claiming any definitive association, especially for SNPs with low minor allele frequency. Hence, these findings need to be confirmed either by testing them on an independent population or performing a meta-analysis.

Introduction

Hepatitis C virus (HCV) infection is a global health care burden with roughly 1-2% of the European and United States populations being chronic carriers of the virus.[1,2] Although HCV can lead to extrahepatic manifestations in a minority of patients, the prognosis of the disease is directly linked to the continuous accumulation of fibrotic tissue in the liver, which ultimately alters the organ architecture and its vascularisation, leading to the development of cirrhosis and its sequelae. Notably, once cirrhosis is established the risk of developing Hepatocellular Carcinoma (HCC) is 1-4% per year.[3] HCC is the leading cause of death among cirrhotic patients and the third cause of cancer-related mortality.[4] Standard therapy with pegylated Interferon-γ and ribavirin fails to clear HCV in ~ 50% of chronically infected individuals.[5,6] Moreover, no vaccination is currently available.

Hepatic fibrogenesis is characterized by a net accumulation of extracellular matrix (ECM) resulting from increased synthesis and decreased degradation. Hepatic Stellate Cells (HSCs) represent the primary source of ECM and in normal liver they are described as being in a quiescent state. When HSCs become activated, mainly in response to inflammatory stimuli, they proliferate and transform into myofibroblasts, actively producing type I collagen.[7, 8]

The risk of progression is variable among chronic patients and it is increased by many host factors, including older age at acquisition of infection, male gender, alcohol consumption, coinfection with Hepatitis B Virus (HBV) and Human Immunodeficiency Virus (HIV), iron overload and other metabolic factors (insulin resistance, obesity). Their role in the natural history of HCV infection has been investigated through retrospective and prospective studies or a combination of both (for a review, see reff. [9] and [10]).

Several studies have also investigated the role of candidate polymorphisms in the natural history of chronic HCV infection.[11] For instance, Single Nucleotide Polymorphisms (SNPs) in the gene region of IFN-γ, TNF-α, or TGF-β have been associated with HCV-induced fibrosis/cirrhosis.[12-16] However, these studies neither provided replication nor resulted in relevant clinical implications.

In contrast to candidate gene studies, Genome-Wide Association Studies (GWASs) represent a hypothesis-free approach, investigating the association with disease of a high number of variants distributed throughout the genome (500k – 1000k SNPs). Recently, GWASs have been reported also in the field of infectious diseases, including HCV.[17] In particular, the successful identification of variants in the IL28B region as significant predictors of the viral clearance resulted from various GWASs performed in population of different ethnicity.[18-21] Hence, an effort towards the exploration of the host genetic background in the setting of progression of liver fibrosis is really noteworthy. In fact, for HCV infection, there is a crucial need for novel, preferably non-invasive predictors that could be used to improve the diagnosis and the management programs for patients at high risk of developing severe liver complications, such as cirrhosis and HCC.

In this work, a GWAS was carried out to study the genetic variability influencing the natural history of HCV infection. A cohort of patients was genotyped and disease stage was evaluated through a liver biopsy before any antiviral treatment. Although no hits reached a genome-wide significant level, candidate regions showing suggestive association were identified. In particular, the gene region containing the *GADD45G* gene resulted of particular interest,

given its role in cell biology, including response to environmental stress, DNA repair, cell cycle arrest and apoptosis.

Materials and methods

Patients

From September 2008 to October 2009, all consecutive HCV patients followed at the Center for Liver Disease at Maggiore Hospital (Milan, Italy) that met the inclusion criteria were included in the study. Inclusion criteria comprised: a) HCV RNA serum positivity, b) an European descent, c) an estimated date of HCV infection based on the first reported parenteral risk factor, d) a diagnostic liver biopsy performed before any antiviral treatment at least four years after the infection, e) no history of alcohol abuse. Moreover, patients were excluded if they had any other cause of liver disease including: HBV infection, HIV infection, Wilson's disease, haemochromatosis or α1-antitrypsin deficiency. The duration of the disease was calculated considering the time elapsed between the year of infection and the liver biopsy. Liver biopsies were evaluated by a single expert pathologist and scored using the Ishak system in separate reports for grading and staging. The score for staging ranged from 0 (no fibrosis) to 6 (cirrhosis). The study was approved by the Institutional Review Board of the Department Of Internal Medicine. All patients gave their written informed consent to receive therapy and gave permission for use of their medical records. This group of patients will be referred as INGM cohort hereafter.

DNA extraction

Genomic DNA (gDNA) was extracted from 1 ml of frozen EDTA-whole blood. Briefly, magnetic beads (GeneCatcher™ gDNA 0.3-1 ml Blood Kit, Invitrogen) were used in a semi-automatic procedure on a FreedomEVO platform (Tecan), following manufacturer's instructions, with minor modifications. Blood samples were lysed manually in GeneCatcher™ Lysis Buffer containing DNA-binding magnetic beads. Then, samples were loaded on the robotic platform for

protease treatment, precipitation with isopropanol, washing and elution. After final elution, gDNA concentration and purity were evaluated on a NanoDrop1000 instrument. The sample was further processed if $A_{260}/A_{280} \geq 1.8$ and $A_{260}/A_{230} \geq 1.5$. On average, this method yielded 10-15 μg of gDNA.

Genotyping

Genotyping was performed using the Illumina 610-Quad BeadChip or 660W-Quad BeadChip starting from 250 ng gDNA. Illumina GenomeStudio software was used to obtain LogR Ratio (LRR), B Allele Frequency (BAF) values and genotype calls, using default settings. A common dataset of 561,332 biallelic SNPs was extracted from the two different genotyping arrays for further analysis.

For additional markers that were not included in the genotyping arrays, TaqMan genotyping was performed. Briefly, SNP alleles were detected using TaqMan SNP Genotyping Assays on a 7900HT real time PCR instrument (Applied Biosystems), following manufacturer's instruction.

Quality control

The following quality control steps were performed to ensure high-quality data for association analysis. The PLINK (v1.06) and R softwares were used.[22,23]

GenomeStudio QC. Genotype calls were obtained from two separate GenomeStudio files, one for each array class (610 or 660W), using Illumina-provided cluster files. Any sample with call rate < 99% was deleted. The LRR standard deviation of autosomal polymorphic probes represented another quality measure of genotyping procedures for each sample (low quality if ≥ 0.3).

Gender check. The specified gender was compared with the estimated sex from X chromosome genotypes. A male call was made if the homozygosity estimate of the X chromosome was > 80%, while a female call was made if X homozygosity was less than 20%. If the estimated sex and the sex from the phenotype database did not match, patients were individually inspected against original medical data records.

Filters on SNPs. All SNPs with Minor Allele Frequency (MAF) < 1% or missing rate (GENO) > 5% were removed from subsequent analyses.

Heterozygosity and inbreeding check. In order to obtain inbreeding coefficients (F) a pruned dataset was generated (using the "--indep-pairwise 100 25 0.5" option in PLINK) such that the SNPs in the subset were in approximate linkage equilibrium. The inbreeding coefficient was estimated for each individual. At the same time, heterozygosity rate was checked. Individuals were removed if F > 0.07.

Relatedness among patients. The Identity By State (IBS) distance was calculated from all markers that passed the quality control. Then Identity By Descent (IBD) proportion was estimated and this measure was used to check for cryptic relatedness between samples. Two individuals were considered as closely related if their pairwise IBD > 0.05.

Population stratification analysis

The population substructure was checked through both Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) as implemented in EIGENSTRAT and PLINK respectively. For PCA, the LD-pruned dataset was used (see above), since it has been shown that local patterns of LD can influence population stratification analysis.[24] Hapmap Rel23 (phase 2) CEU, YRI and JPT+CHB population were used as references. A merged dataset was generated containing 52,158 markers shared by the genotyped population and the reference groups. Visual inspection of eigenvector plots was employed to find unusual clustering patterns. Similarly, MDS plots were generated to check for population substructure, using calculated genome-wide IBS distance.

Statistical models for association tests

In order to quantitatively describe the disease outcome, a Fibrosis Progression Rate (FPR) was calculated by taking the ratio between the Ishak score (staging value) and the disease duration (in years). All association tests were performed using PLINK and R.

Continuous outcome: linear regression. For each SNP to be examined, the main association test was a model of the $\log_{10}$(FPR) as a linear combination of

the additive effect of allele dosage and a number of covariates, namely the age at the infection, the gender and the HCV genotype. The $\log_{10}$-transformation of the FPR was employed to obtain linearity. For cases having an Ishak = 0, the $\log_{10}$(FPR) was substituted with the lowest value observed. The model considered an additive effect of allele dosage, counting the minor allele. After association testing, Q-Q plots were generated to check for systematic sources of spurious association. Observed P-values were ranked and plotted against those obtained under the null hypothesis.

Case-control phenotype: logistic regression. On the basis of the FPR value, the whole population was split in two groups based on the median. Hence, patients were defined as cases if FPR ≥ median(FPR), conversely they were considered as controls if FPR < median(FPR). Here, the term "cases" refers to patients that progress more rapidly to cirrhosis, while the term "controls" denotes individuals with a slower disease progression. A logistic regression was implemented, including covariates (the age at the infection, the gender and the HCV genotype). The model considered an additive effect of allele dosage, counting the minor allele.

Cox-proportional hazard regression. Survival analysis was performed to model the time it takes for advanced fibrosis to occur. Here, advanced fibrosis was defined if Ishak ≥ 4. A Cox proportional-hazards regression model was fitted specifying the SNP genotype and two covariates, the age at the infection and the gender. The model considered an additive effect of allele dosage, counting the minor allele. The `survival` library was used in R.[25]

Genotype imputation

Genotype imputation was performed with MACH (v 1.0)[26] for autosomal SNPs not directly assayed through genotyping. The HapMap Rel22 CEU phased haplotypes were used as reference panel. Imputed SNP with low-quality score were filtered-out (Rsq_hat ≤ 0.3). Additionally, SNPs with MAF < 1% were also removed.

Results

Quality control overview

A total of 561,332 SNPS were initially included in this study. After missingness check and MAF filtering, 536,611 markers were eventually included. Of 251 participants, three were removed because of low call rate (<99%), while one was excluded because of unusual high homozygosity rate (F > 0.07). The reported and the estimated sex matched in all cases. Moreover, IBD analysis revealed that all individuals were unrelated. The total genotyping rate in remaining 247 individuals was 99.96%

Patients characteristics and population substructure

Patient characteristics are outlined in Table 1. The majority of the patients were infected with HCV genotype 1 (52%), although the present study includes subjects with HCV genotypes 1, 2, 3 and 4. Both males and females were well represented (52% males, 48% females). The median age at infection was 21 years, the median disease duration was 25 years, while the median age at biopsy was 47 years. The FPR distribution resulted right-skewed, but approached a normal distribution after $\log_{10}$-transformation. The majority of the patients (87%) had minimal to mild histological activity (grading<9), while a minor fraction (13%) showed moderate to severe activity (grading ≥9). The main reported risk factors for HCV infection were blood transfusions (75%) and the use of intravenous drugs (23%), with mother-to-child, needlestick or sexual transmission as the other reported risks. The mean Body Mass Index was 25.3 kg/m^2.

Population substructure was inspected using both MDS and PC analysis. Both approaches revealed that the population is genetically homogeneous, and only one cluster is evidenced. Figure 1 shows the MDS plots and the PC plots. In particular, no population admixture is present, as shown when plotting the three reference HapMap populations along with the INGM cohort.

The age at infection, the gender and the HCV genotype are associated with fibrosis progression

We first assessed the contribute of the covariates to the disease progression, in order to specifically consider all the confounding factors in the statistical model to evaluate genotype association. In a first analysis framework, we modeled the dependence of log10(FPR) on a linear combination of different explanatory variables already reported to influence fibrosis progression, i.e. the age at the infection, the gender, the HCV genotype, the BMI and the grading. The final model was chosen on the basis of the minimum Akaike Information Criterion. Table 2 shows that the age at infection resulted the variable most strongly and positively associated with fibrosis progression, ($p<2E-16$), while the gender and the HCV genotype resulted less but significantly associated with progression rate ($p<0.01$). In particular, HCV genotype 2 and 3 were respectively associated with slower and faster rates of fibrosis progression, while higher progression rates were observed in males. Overall, these variables explain 30.4% of the observed phenotypic variability, as resulted from the analysis of the deviance table.

The significant effect of the age at infection is manifested by the fact that patients infected vertically or perinatally have very slow progression of liver fibrosis, For those patients, the mean progression is 0.049 FPR units, corresponding approximately to an increase of 2 Ishak points in 40 years, similarly to other reports.[27, 28]

We also performed a survival analysis to examine the relationship of the above-mentioned explanatory factors with advanced fibrosis. In agreement with the linear model, the significant variables were the age at infection ($p<2E-16$), the gender ($p<0.05$) and the HCV genotype ($p<0.01$) (Table 3 and Figure 8). From the estimates of the coefficients we can derive the effect on the hazard. Thus, holding the other covariates constant, each additional year at infection produces an highly significant increase of the hazard of advanced fibrosis by a factor of $e^{0.112}=1.118$ (95% CI =1.087 - 1.151), or 11.8%. This effect is soundly evident when plotting the estimated survival functions for three representative infection times (0, 20 or 40 years of age respectively, Figure 8). Indeed, the

three survival functions are well separated and their corresponding confidence envelopes do not overlap. In summary, on the basis of these preliminary analyses, three critical covariates were included in all following statistical analyses, namely the age at the infection, the gender and the HCV genotype.

SNPs associated with fibrosis progression

Initially, we tested the association of each SNP with the progression of fibrosis by means of a linear regression model with the covariates indicated above and the phenotype specified by the FPR. No single marker showed a genome-wide significant association, resulting from a standard Bonferroni correction ($p<10^{-7}$, see Figure 2). The top associated SNPs are shown in Table 4. The two most strongly associated SNP were rs13378149 (p=3.07E-07) on chromosome 13 and rs17064865 (p= 4.60E-07) on chromosome 8. The first SNP lies within a gene-desert region, while rs17064865 is ~ 38 kb downstream the gene *MYOM2*. The minor allele (G) was associated with an increased fibrosis progression, as revealed from a positive regression coefficient. An enlarged view of this region is shown in Figure 4

The Q-Q plot shows the observed vs. the expected P-values. The inspection of the plot revealed that there were no systematic sources of spurious association. However, the close adherence of P-values to the diagonal line suggests that there is very scarce evidence for any SNP reaching a definitive level of genome-wide significance.

Additionally, the population was split in a case-control dataset, on the basis of the median FPR value, as indicated in materials and methods. After correcting for the covariates, the top associated SNPs are those shown in Table 5. There was no SNP reaching a genome-wide level of significance ($p<10^{-7}$, see Figure 2). The strongest association signal came from a region on chromosome 9 containing the *GADD45G* gene. This region features a number of closely linked SNPs ($r^2>0.8$) with a $10^{-7}<p<10^{-4}$. An enlarged view is given on Figure 5. The lowest P-value was observed for rs7871395 (p=1.57E-06). The minor allele (A) was associated with an increased fibrosis progression as shown in Figure 7

(from logistic regression with additive model, OR = 3.8, 95%CI = 2.2 – 6.7). The other top associated SNPs are shown in Table 5.

Alternatively, we tested the association of each SNP with the progression of fibrosis with a survival analysis, using a Cox proportional-hazard regression and correcting for the age at the infection and the gender. The most distinct association resulted from rs2233236 (p=5.29E-08). The minor allele (A) was associated with an increased hazard of advanced fibrosis. rs2233236 lies in an intron of the gene *TRAPPC9*. An enlarged view of this region is shown in Figure 6. The other top associated SNPs are shown in Table 6.

Genotype imputation was performed to predict the genotype of untyped SNPs, as described in Materials and Methods. These additional genotypes were tested for association with linear or logistic regression, in order to fine-map the candidate regions. However, after imputation, the markers most strongly associated with the disease resulted those directly assayed.

SNPs on IL28B have no effect on fibrosis progression

Here, we specifically checked if the genotype of rs8099917 and rs12979860 polymorphisms could influence fibrosis progression in the liver of HCV-infected patients. These SNPs were previously associated with both spontaneous and treatment-induced viral clearance in recent studies.[18-21, 29, 30] No evident effect can be evinced by looking at the host IL28B genotype (see Figure 8). Inclusion of these predictors in the linear regression model did not resulted in any significant association. Likewise, no impact on FPR was found even supposing a dominant model of inheritance for the alleles rs8099917 G or rs12979860 T, previously associated with treatment failure (data not shown). On the contrary, when these two SNPs were tested for association with the outcome of the therapy, using data from 91 patients with HCV genotype 1 and available treatment information, the effect was readily detectable. Actually, rs12979860 genotypes CT and TT vs. CC resulted associated with treatment failure ($p<0.01$, OR = 3.6, 95% CI = 1.3 – 10.2) in agreement with previous reports.

In order to further validate these findings, we assessed the effect of IL28B polymorphisms and covariates on advanced fibrosis using a selected subset of

patients, whose date of infection was certain, as a result of their medical history. Specifically, we selected those patients that manifested a documented spike in serum transaminase levels or a positive HCV test within six months from the exposure to the risk factor. Thus, using data from 105 patients, we confirmed the lack of any effect of IL28B SNPs, both with linear regression and survival analysis (data not shown). Remarkably, the estimated effects of covariates on the hazard of advanced fibrosis were similar to those obtained using the whole population. For instance, the hazard for each additional year at infection was calculated to increase by 1.118 (95% CI = 1.072 - 1.167) using the smaller population, comparably to the estimate obtained with the whole group, namely 1.118 (95% CI =1.087 - 1.151).

Discussion

In this work we described a GWAS on fibrosis progression in the setting of chronic HCV infection. In order to evaluate the contribute of genetic and non-genetic factors in the natural history of chronic HCV infection, we performed multiple analyses aimed at the definition of the predictors of fibrosis progression in a cohort of European patients, well selected for this purpose. The host genetic variation was previously reported as an important factor associated to treatment success. Indeed, the identification of some SNPs in the IL28B region as significant predictors of the treatment-induced viral clearance resulted from various GWASs performed in population of different ethnicity[18-20]. These findings were further validated and the association with spontaneous viral clearance was also demonstrated.[21, 29, 30] While the effect size varied between different populations, reflecting both distinct genetic backgrounds but also different study designs, the associations were extensively replicated and provided strong evidence on the role of host genetics in the field of infection diseases. Conversely, the role of common genetic variation on the natural history of chronic HCV infection is still under debate. The risk of progression to cirrhosis is increased by many host factors, including older age at acquisition of the infection, male gender, alcohol consumption, coinfection with Hepatitis B

Virus (HBV) and Human Immunodeficiency Virus (HIV), iron overload and other metabolic factors, like insulin resistance and obesity.[9,10] Other host genetic factors may contribute to the risk of advanced disease, and numerous candidate-gene studies have been performed. For instance, SNPs in the gene or promoter region of IFN-γ, TNF-α, or TGF-β have been associated with HCV-induced fibrosis/cirrhosis.[12-16] However, these studies neither were replicated nor resulted in relevant clinical implications. The main limitations that hampered the clear recognition of the host genetic contribute to the natural history included poor study design, limited sample size, noninclusion of viral and host cofactors in the association model and small effect size of the causative variants. Furthermore, no published GWAS exists in the field of fibrosis progression in chronic HCV infection. One work reported the use of ~24k SNPs to investigate the role of human variation, developing a significant genomic signature to predict cirrhosis.[31] However, the age at the infection was not accounted for in the study, although it is largely considered one of the main predictors of fibrosis progression.[2, 9, 10] Moreover, the number of markers used in the study is considered insufficient to cover the underlying common genetic variation in the population, because more than ~500,000 SNPs are required. Actually, it has been show that the vast majority of SNPs with a MAF of at least 5% could be reduced to ~550,000 LD bins for individuals of non-African ancestry.[32-34]

Hence, the aim of this study was to perform a hypotesis-free GWAS in a cohort of well-characterized patients with chronic HCV infection. In a first effort to correlate the disease progression with host and external variables, we modeled the fibrosis progression rate as a continuous outcome, considering the ratio between the fibrosis level and the disease duration. This approach assumes intrinsically that the rate of progression to cirrhosis remains constant over time. Although this assumption might actually be incorrect,[35] this method represents a way to consider the duration of the chronic disease within the model, instead of a simple split of the population in two groups on the basis of the fibrosis stage alone. This study design can potentially reveal a more accurate effect of the explanatory variables because the disease outcome is

represented by the rate of fibrosis progression that is, by definition, the change per unit time, at the expense of converting a qualitative assessment of fibrosis into a quantitative variable. One possible disadvantage is the fact that the regression coefficients are difficult to interpret, because they measure the effect on the $\log_{10}$(FPR), a measure not utilized in clinical practice. In summary, taking the duration of the disease directly into account, the well-established role of external factor is confirmed in our study. As expected, we detected a strong effect of the age at infection on the rate of disease progression, namely a 2.9% increase in the speed of disease progression for each additional year. Conversely, we observed no genome-wide significant effect of SNPs on fibrosis progression in this cohort ($p<10^{-7}$ after Bonferroni correction). Hence, the SNPs identified here show suggestive rather than definitive evidence for association. They can be considered as candidate hits to be further validated in an independent population.

The second approach involved the definition of a case-control dataset, obtained classifying the samples as "high progressors" or "slow progressors", on the basis of the median FPR value. This study design has the advantage of calculating an effect size that is directly interpretable, namely the Odds Ratio obtained from the logistic regression. Using these settings, we highlighted the association of a candidate region containing the *GADD45G* gene with the progression of fibrosis, although a direct functional relationship cannot be inferred from the assayed or imputed SNP. Notably, some SNPs in the same region are among the top hits using the linear regression analysis. rs7871395 is a common variant in the studied population (MAF=0.21) and in HapMap samples (MAF is 0.22 in CEU and 0.19 in TSI). The Growth Arrest and DNA Damage (Gadd45) family comprises three highly conserved genes (*GADD45A, GADD45B, GADD45G*) that contribute to the cellular homeostasis in response to a number of stresses. GADD45 proteins have been implicated in cell cycle arrest, DNA repair and induction of apoptosis. Their action is mediated by physical interactions with other cellular proteins that are implicated in the cell cycle regulation and the response of cells to stress. These include PCNA, p21, CDK1/CyclinB1, and the p38 and JNK kinases.[36, 37] Notably, GADD45-α was

found to be downregulated at the mRNA level in HCC, compared to matched cirrhosis.[38] More recently, another work showed that GADD45-γ was downregulated at the protein level in HCC vs. non-neoplastic tissue and it activated p38 and JNK kinase pathways in hepatoma Hep-G2 cell line, inducing G2/M arrest.[39] In addition, the expression of GADD45-β was reported to be inhibited by HCV infection in tumoral and non-tumoral tissues, supporting the evidence that GADD45-β was undetectable in HCCs with HCV etiology.[40] In the same paper the authors reported no significant effect on GADD45-γ by HCV, although a trend towards a reduced expression in HCC was shown also for this protein.

A contribution of GADD45 proteins during fibrosis development has also been proposed. In a rat model of liver fibrosis, a PPARγ agonist induced a decrease in HSC activation, revealed by diminished α-smooth muscle actin (α-SMA) and collagen α1 levels, two well-kwon markers of HSC activation. *In vitro*, the α-SMA reduction was accompanied by increased GADD45 expression, after treatment with the same agonist.[41] In another report, a decreased expression of GADD45-α and p53 was observed after stimulation with IL-1β of fibroblasts isolated from hypertrophic scar, but not from normal tissue. This reduced expression of a well-kwon cell cycle regulator was linked to the excessive fibroblast hyperproliferation and contraction associated with hypertrophic scars.[42] Likewise, liver fibrosis is a scarring response characterized by excessive extracellular matrix accumulation that ultimately leads to cirrhosis, associated with nodule formation and organ contraction. $TGF\beta_1$ is the most potent stimuli for the production of extracellular matrix from activated HSC, mainly resulting in transcriptional upregulation of type 1 collagen, a process mediated by Smad proteins.[43] Very interestingly, all GADD45 members are induced upon TGFβ treatment in a Smad4-dependent manner.[44] Hence, GADD45 genes reasonably play a function in influencing the progression of liver fibrosis through their modulation of HSC proliferation.

Taken together, these findings point out the relevance of GADD45 gene family members in the fibrogenesis and in particular in the pathogenesis of liver cirrhosis and HCC. In this work, we identified a region containing the gene

GADD45G, as being associated with faster fibrosis progression. A working hypothesis is that lower expression levels of GADD45-γ could be one of the predisposing factors for accelerated fibrosis progression. Actually, its role in the cell cycle arrest, induction of DNA repair and apoptosis has long been demonstrated. SNPs in the promoter region could be associated to differential expression of *GADD45G*. Alternatively, variants in the coding region could results in a non-functional protein. Either a decreased expression or a non-functional variant of this negative regulator of the cell cycle might predispose to increased proliferation of HSCs, the main players of fibrogenesis, and eventually contribute also to the neoplastic transformation of hepatocytes. Thus, a fine-map effort is required to gain further insight into the possible biological mechanisms underlying this association. More importantly, these findings need to be further validated in a replication phase or alternatively, through a meta-analysis with other genotyped cohorts, in order to provide genome-wide significant association.

In a third approach, we used a Cox-proportional hazard regression to directly estimate the hazard of developing advanced fibrosis as a function of host genotype, adjusting for covariates. This analysis conveniently outputs the effect of the variables on the hazard, providing useful insight on the natural history of HCV infection. Because survival analysis examines the time it takes for an event to occur, it directly considers the duration of the disease, avoiding the bias of splitting the population into groups with different periods of infection. Remarkably, the age at infection provides the major contribute to the hazard of advanced fibrosis, in agreement with other reports.[35, 45, 46] Using this model, the strongest genetic predictor of fast progression to advanced fibrosis/cirrhosis was rs2233236, reaching a genome-wide significant p-value. Although the signal is genome-wide significant, this SNP is a low-frequency variant (MAF=0.03) and no direct causal relationship can be inferred. This SNP lies in one of the introns on the gene *TRAPPC9* (previously known as *NIBP)*. This gene is associated with autosomal recessive mental retardation and involved in NF-κB signaling pathway, through direct interaction with IKK-β and MAP3K14 (NIK).[47-49] Within the liver, NF-κB has a role in the transcriptional regulation of

inflammation and fibrosis and it is considered a tumor-promoter.[179] Moreover, NF-κB expression increases substantially during HSC activation.[50, 51] Hence, although the ability of TRAPPC9/NIBP in modulating NF-κB activity is of potential interest, its expression is reported mainly in the nervous system and its role in liver physiology, if any, remains to be investigated.

A major limitation of our study is the small sample size that does not allow claiming a definitive association, especially for low frequency variants. The development of liver cirrhosis can be considered as a complex trait where the disease outcome is influenced at on side by the host genetic variability in a polygenic manner and at the other side by environmental modifiable or non-modifiable factors. Other GWAS in HCV infected patients have been performed with a first discovery cohort comparable in size to the one studied here, reporting very low, genome-wide significant p-values.[19, 20] One important difference with the present study is the effect size underlying the associated variant. Because the sample size required to obtain enough statistical power for detecting an association increases with smaller effect size,[52] only effects with sizeable OR may be detected with hundreds of samples (i.e. OR>5). Conversely, the majority of reported associations after GWAS has small OR (i.e. OR<2),[33] requiring large sample sizes. The IL28B gene variants have a large effect size on the therapy outcome, hence the association has been demonstrated in relatively small populations.

To our knowledge, this is the first GWAS in the field of the natural history of chronic HCV infection performed with enough genome coverage to catch the genetic variability in a European population. Although the sample size may be considered limited to gain enough statistical power on less common variants, we obtained a list of suggestive associations with a plausible biological role in determining the speed of progression to cirrhosis. Therefore, the reported associations merit further investigation, in particular a second phase validation. Additionally, an assessment of differential expression of the candidate genes in patients with rapid fibrosis progression would provide valuable biological insight into the process of hepatic fibrogenesis.

	All	HCV1	HCV2	HCV3	HCV4
Patients	247	129	74	34	10
Males	129 (52.2%)	66 (51.2%)	34 (45.9%)	21 (61.8%)	8 (80.0%)
Females	118 (47.8%)	63 (48.8%)	40 (4.1%)	13 (38.2%)	2 (20.0%)
Median age at infection, yrs (range)	21 (0 – 59)	22 (0 – 59)	22 (0 – 56)	19 (0 – 42)	21 (15 – 33)
Median age at biopsy, yrs (range)	47 (11 – 72)	48 (13 – 72)	50 (11 – 66)	43.5 (24 – 64)	41 (33 – 53)
Median duration, yrs (range)	25 (4 – 53)	25 (6 – 53)	25.5 (5 – 52)	23 (4 – 43)	19.5 (9 – 28)
Fibrosis stage, Ishak					
0	2 (0.8%)	2 (1.6%)	-	-	-
1	75 (30.4%)	36 (27.9%)	29 (39.2%)	5 (14.7%)	5 (50.0%)
2	80 (32.4%)	40 (31.0%)	26 (35.1%)	13 (38.2%)	1 (10.0%)
3	30 (12.1%)	11 (8.5%)	9 (12.2%)	8 (23.5%)	2 (20.0%)
4	20 (8.1%)	12 (9.3%)	5 (6.8%)	3 (8.8%)	-
5	18 (7.3)%	14 (10.9%)	3 (4.1%)	1 (2.9)%	-
6	22 (8.9)%	14 (10.9)%	2 (2.7%)	4 (11.8%)	2 (20.0%)
Grading					
Minimal to mild, <9	216 (87.4%)	113 (87.6%)	64 (86.5%)	31 (91.2%)	8 (80.0%)
Moderate so severe, ≥9	31 (12.6%)	16 (12.4%)	10 (13.5%)	3 (8.8%)	2 (20.0%)
Reported risk					
Blood transfusion	185 (74.9%)	97 (75.2%)	58 (78.4%)	24 (70.6%)	6 (60.0%)
Intravenous drug use	58 (23.5%)	31 (24.0%)	14 (18.9%)	9 (26.5%)	4 (40.0%)
Other	4 (1.6%)	1 (0.8%)	2 (2.7%)	1 (2.9%)	-
BMI, kg/m^2					
Mean (SD)	25.3 (2.8)	25.3 (2.8)	25.5 (2.8)	25.0 (2.5)	25.9 (2.2)

Table 1. Patient characteristics, stratified by HCV genotype.

	Estimate	95% CI	*P* value (contrast)	P value (predictor)
Age at infection	0.012	0.010 – 0.015	< 2E-16	< 2E-16
Gender (male)	0.091	0.023 – 0.159	0.0097	0.0098
HCV1	Reference	–	–	0.0013
HCV2	-0.083	-0.160 – -0.005	0.0370	
HCV3	0.144	0.041 – 0.247	0.0067	
HCV4	0.019	0.157 – 0.194	0.8353	

Table 2. Covariates significantly affecting the fibrosis progression (linear model). Estimate = effect size on $\log_{10}$(FPR); 95% CI = 95% Confidence interval for the estimate; P value (contrast) = P value for the estimate; P value (predictor) = P value for the linear predictor.

	Estimate	Effect = $e^{estimate}$	95% CI	*P* value (contrast)	P value (predictor)
Age at infection	0.112	1.118	1.087 – 1.151	1.97E-14	< 2E-16
Gender (male)	0.648	1.912	1.106 – 3.304	0.0203	0.0194
HCV1	Reference	–	–	–	0.0062
HCV2	-0.874	0.417	0.208 – 0.838	0.0141	
HCV3	0.657	1.929	0.865 – 4.304	0.1084	
HCV4	0.795	2.215	0.505 – 9.719	0.2918	

Table 3. Covariates significantly associated to the hazard of advanced fibrosis (Cox-proportional hazard model). Estimate = effect size on the log-hazard; Effect = $e^{estimate}$, or multiplicative effect on the hazard; 95% Confidence interval for the effect; P value (contrast) = P value for the estimate; P value (predictor) = P value for the linear predictor.

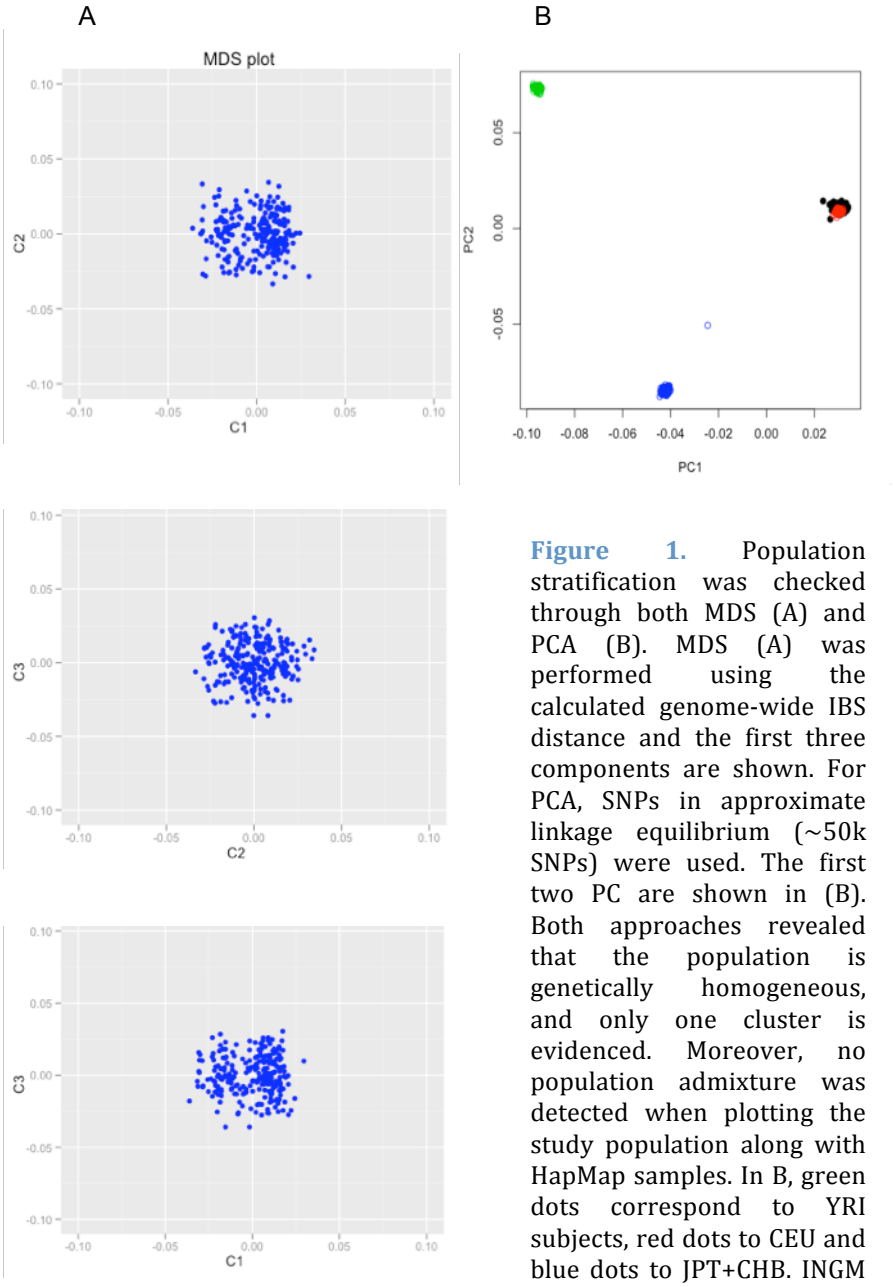


Figure 1. Population stratification was checked through both MDS (A) and PCA (B). MDS (A) was performed using the calculated genome-wide IBS distance and the first three components are shown. For PCA, SNPs in approximate linkage equilibrium (~50k SNPs) were used. The first two PC are shown in (B). Both approaches revealed that the population is genetically homogeneous, and only one cluster is evidenced. Moreover, no population admixture was detected when plotting the study population along with HapMap samples. In B, green dots correspond to YRI subjects, red dots to CEU and blue dots to JPT+CHB. INGM subjects are shown with black dots.

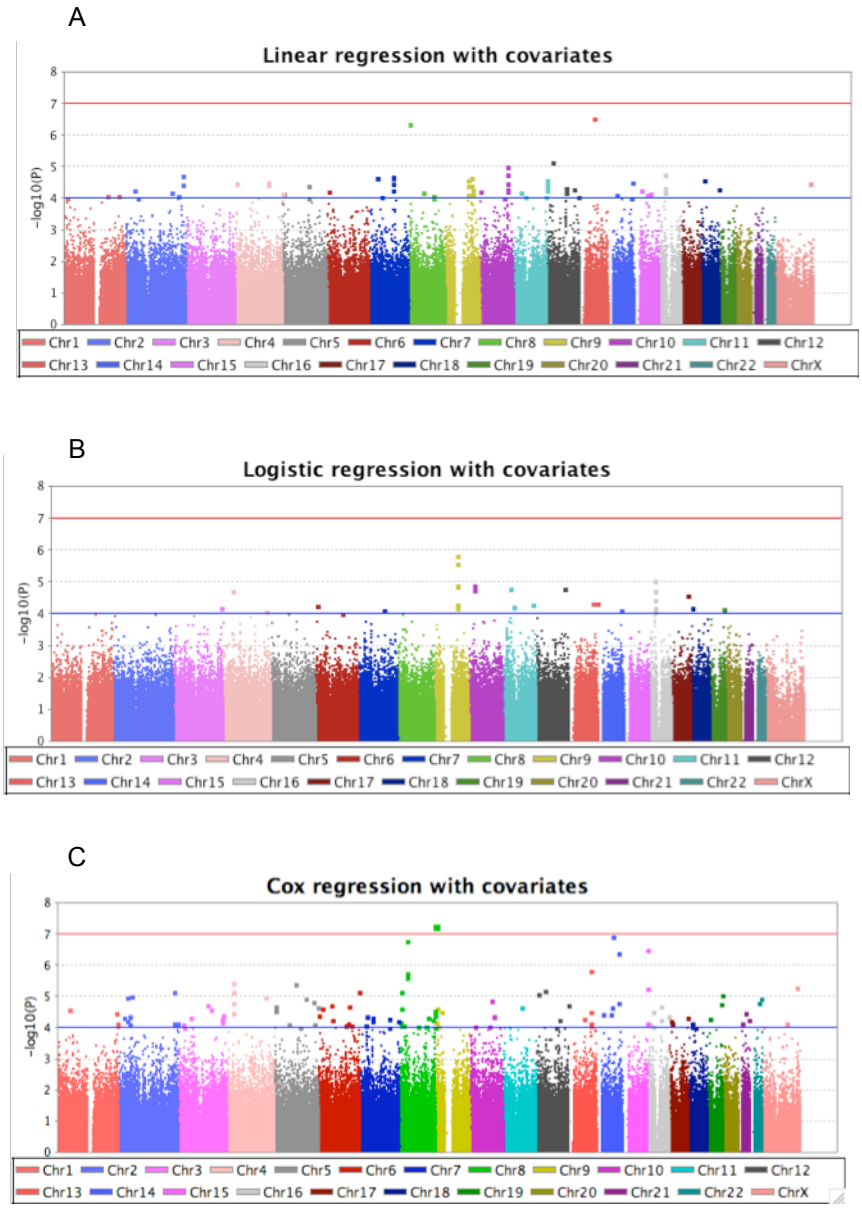


Figure 2. Manhattan plot of GWAS with fibrosis progression in chronic HCV infection. Each point corresponds to a SNP that passed the quality control. The vertical axis show the $-\log_{10}$(P-value), in tree different statistical model: the linear regression (A), the logistic regression (B) and the Cox proportional-hazard regression (C). The horizontal red line corresponds to the genome-wide level of significance ($p<10^{-7}$), while the blue line corresponds to suggestive evidence for association ($10^{-7}<p<10^{-4}$).

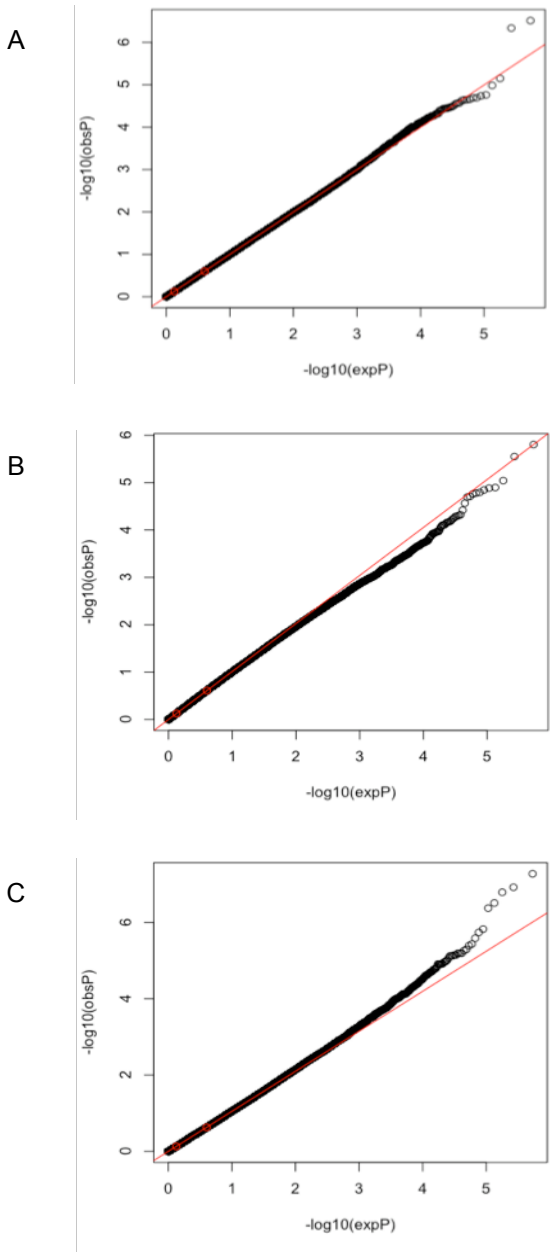


Figure 3. Q-Q plot obtained for the three association models, namely the linear (A), logistic (B) and Cox (C) regressions. The red line denotes expectation under the null hypothesis and black circles indicate the $-\log_{10}$(P-value) from the observed data (y-axis) plotted against the distribution that would be expected under the null hypothesis of no association (x-axis). The observed data closely adhere to the expectation indicating both the lack of systematic sources of spurious association and little evidence for definitive true associations.

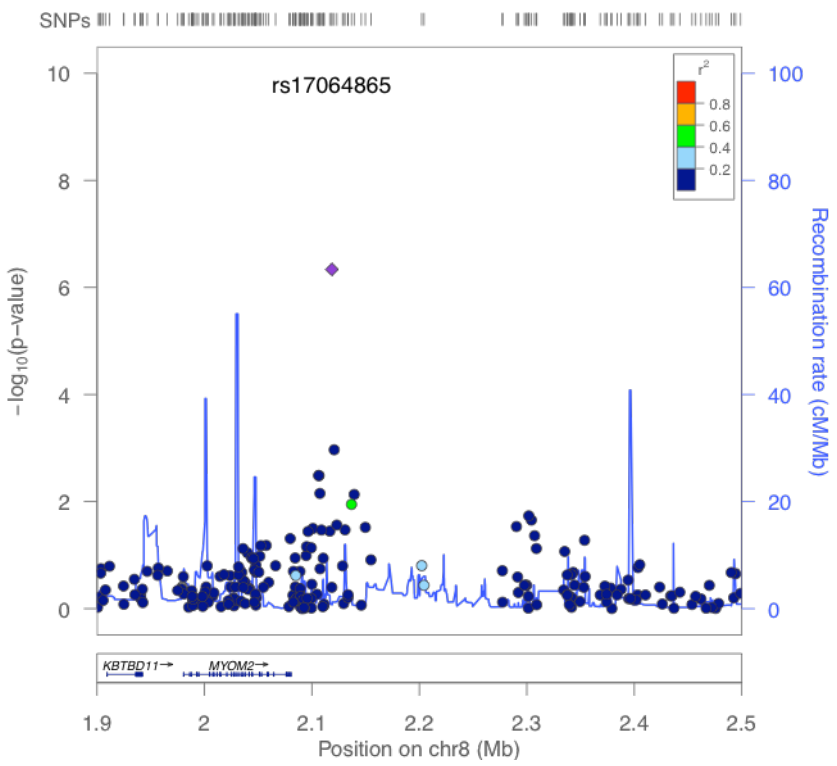
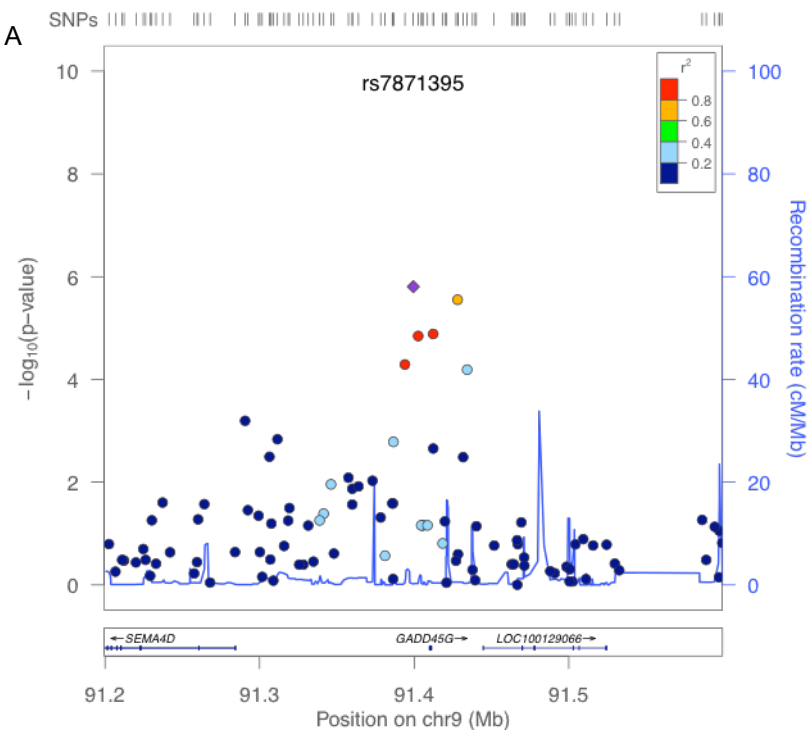


Figure 4. An enlarged view of the region containing the top hit identified through linear regression analysis. A diamond-shaped sign marks the top associated SNP, rs17064865. Each filled circle corresponds to a measured SNP, colored on the basis of the pairwise LD with rs17064865 (r^2), calculated from HapMap II samples. The –log10(P-value) is reported on the left vertical axis, while the recombination rate is reported on the right axis and plotted as a blue continuous line. This image was generated with LocusZoom.


A
SNPs
rs7871395
r²
0.8
0.6
0.4
0.2
10
8
6
4
2
0
100
80
60
40
20
0
−log10(p−value)
Recombination rate (cM/Mb)
SEMA4D
GADD45G
LOC100129066
91.2
91.3
91.4
91.5
Position on chr9 (Mb)

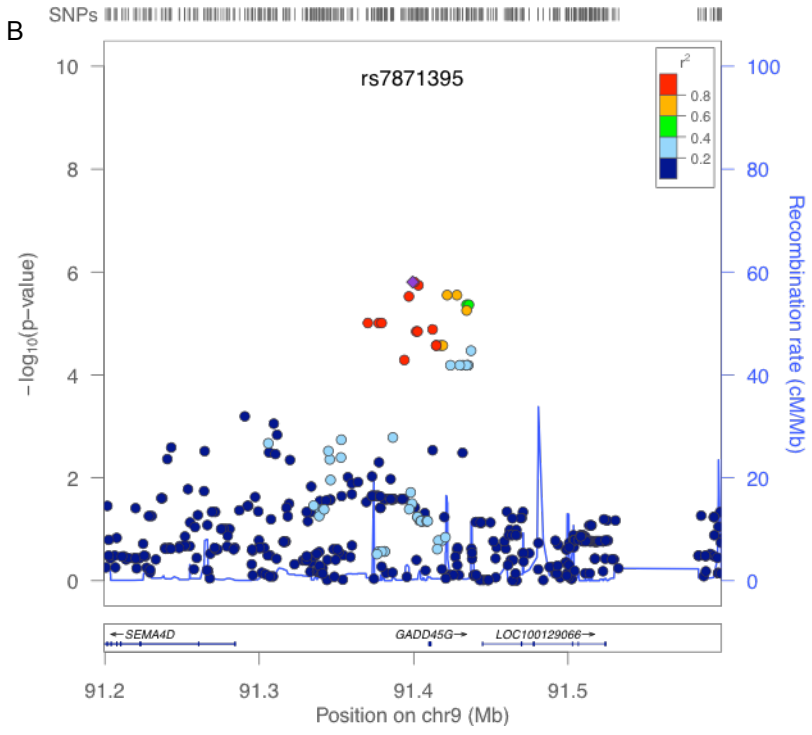

B
SNPs
rs7871395
r²
0.8
0.6
0.4
0.2
10
8
6
4
2
0
100
80
60
40
20
0
−log10(p−value)
Recombination rate (cM/Mb)
SEMA4D
GADD45G
LOC100129066
91.2
91.3
91.4
91.5
Position on chr9 (Mb)

Figure 5 (previous page). An enlarged view of the region containing the top hit identified through logistic regression analysis reveals that closely-linked SNPs in the region of the *GADD45G* gene are associated with a faster progression of fibrosis. A diamond-shaped sign marks the top associated SNP, rs7871395. Each filled circle corresponds to a SNP, colored on the basis of the pairwise LD with rs7871395 (r^2), calculated from HapMap II samples. The $-\log_{10}$(P-value) is reported on the left vertical axis, while the recombination rate is reported on the right axis and plotted as a blue continuous line. The upper panel (A) shows only the SNPs directly assayed, while the lower panel (B) depicts both measured and imputed SNP. This image was generated with LocusZoom.

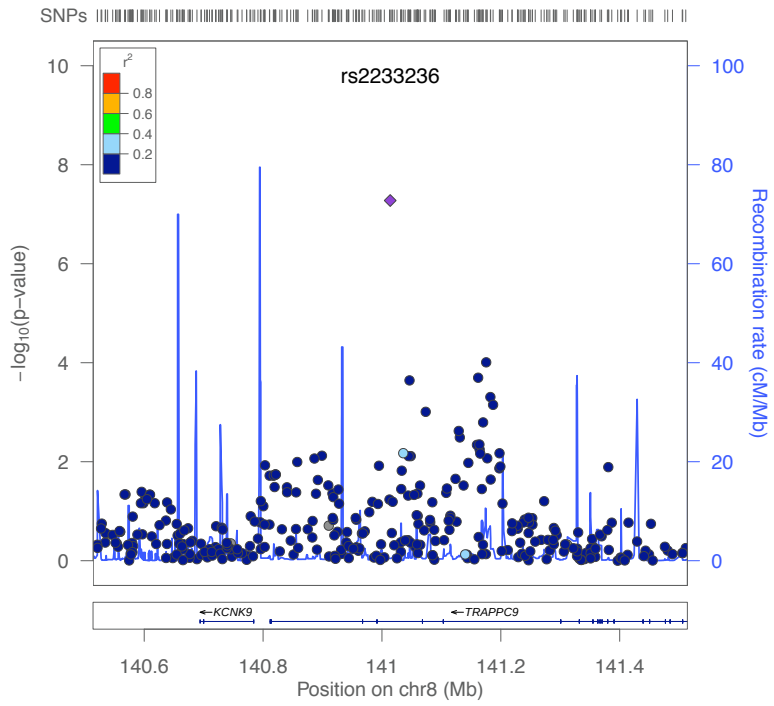


Figure 6. An enlarged view of the region containing the top hit identified through Cox proportional hazard regression analysis. A diamond-shaped sign marks the top associated SNP, rs2233236. Each filled circle corresponds to a measured SNP, colored on the basis of the pairwise LD with rs2233236 (r^2), calculated from HapMap II samples. The $-\log_{10}$(P-value) is reported on the left vertical axis, while the recombination rate is reported on the right axis and plotted as a blue continuous line. This image was generated with LocusZoom.

	rs7871395		
	GG	**GA**	**AA**
Fast fibrosis progression (n=122)	61	50	11
Slow fibrosis progression (n=125)	95	27	3

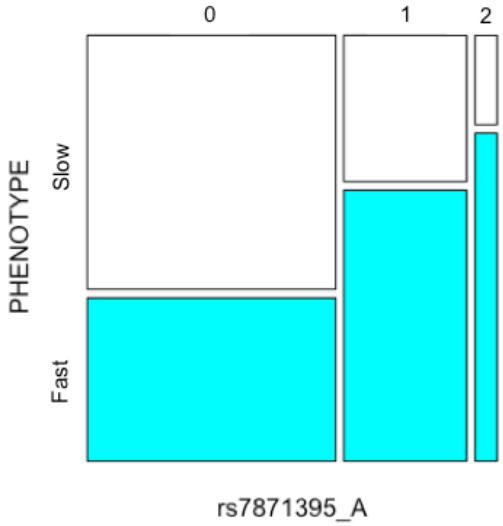


Figure 7. Mosaic plot showing the association of rs7871395 A allele with faster fibrosis progression. Genotypes are coded counting the minor allele, hence GG=0, GA=1 and AA=2. The width of each bar is proportional to the frequency of the genotype, while the height is proportional to the percentage of individuals showing a fast or slow fibrosis progression phenotype for each rs7871395 genotype. The additive effect of the allele dosage is evident. Genotype counts in the two categories are given in the top table.

SNP	*CHR*	*Position*	*P-value*	*A1*	*Genes within ± 100 kb*
rs13378149	13	58050022	3.07E-07	A	
rs17064865	8	2118593	4.60E-07	G	*MYOM2*
rs10842194	12	23672139	7.10E-06	G	*SOX5*
rs7082958	10	109784983	1.04E-05	A	
rs9927848	16	23740572	1.74E-05	C	*PRKCB1, LOC63928*
rs11193831	10	109739164	1.83E-05	A	
rs6436859	2	229905506	2.02E-05	A	*PID1, DNER*
rs2285694	7	94750340	2.03E-05	A	*PPP1R9A, PON1, PON3*
rs303543	9	105213031	2.21E-05	A	
rs41348	7	28683373	2.26E-05	A	*CREB5*
rs10156050	7	94743726	2.28E-05	A	*PPP1R9A, PON1, PON3*
rs303635	9	105188036	2.36E-05	C	
rs10908905	9	91428064	2.67E-05	A	*GADD45D*
rs7112210	11	132537509	2.72E-05	A	*OPCML*
rs7229921	18	12273547	2.81E-05	G	*CIDEA, TUBB6, AFG3L2*
rs7149001	14	96851587	3.09E-05	G	
rs628527	4	130742729	3.28E-05	G	
rs3735590	7	94765431	3.36E-05	A	*PPP1R9A, PON1, PON3*
rs1931600	10	109851845	3.43E-05	A	
rs4689316	4	4398406	3.54E-05	G	*TMEM128, LYAR, ZNF509, D4S234E, STX18*
rs4145397	X	138132502	3.60E-05	A	
rs2163561	11	132531564	3.64E-05	A	*OPCML*
rs6715968	2	229884476	3.64E-05	A	*PID1, DNER*
rs640314	4	130757562	3.76E-05	G	
rs303596	9	105162386	3.95E-05	A	
rs11741503	5	105663724	4.02E-05	A	
rs7871395	9	91399407	4.10E-05	A	*GADD45D*
rs11602730	11	132531363	4.74E-05	C	
rs15513	12	74174518	4.79E-05	A	*OPCML*
rs3741495	12	74183755	4.79E-05	A	*KRR1, LOC63928*

Table 4. The top 30 SNPs resulting from the linear model are shown. A1 is the minor allele, for which an additive effect was tested.

SNP	*CHR*	*Position*	*P-value*	*A1*	*Genes within ± 100 kb*
rs7871395	9	91399407	1.57E-06	A	*GADD45G*
rs10908905	9	91428064	2.81E-06	A	*GADD45G*
rs732798	16	23811451	9.05E-06	C	*PRKCB1*
rs7094334	10	17710038	1.26E-05	C	*PTPLA, STAM*
rs3138512	9	91412273	1.30E-05	A	*GADD45G*
rs1475537	9	91402570	1.43E-05	A	*GADD45G*
rs2779154	11	29405675	1.61E-05	G	
rs933741	12	113346102	1.67E-05	G	*TBX5*
rs7082590	10	17719624	1.73E-05	A	*PTPLA, STAM*
rs2188355	16	23775277	1.96E-05	A	*PRKCB1*
rs10010602	4	36971373	2.01E-05	G	
rs9890694	17	63827063	2.72E-05	A	*AMZ2, SLC16A6, ARSG*
rs12931116	16	23788374	3.81E-05	G	*PRKCB1*
rs571564	13	111773792	4.75E-05	G	*SOX1*
rs4773787	13	93821771	4.91E-05	G	*GPC6, DCT*
rs2031970	9	91393992	5.12E-05	A	*GADD45G*
rs17124120	11	120010838	5.22E-05	A	*GRIK4*
rs9503657	6	3589400	5.64E-05	A	*C6orf145*
rs12274692	11	43952769	5.96E-05	G	*ALKBH3, LOC390110, PHACS*
rs4877110	9	91434227	6.48E-05	G	*GADD45G*
rs12696601	3	191548366	6.65E-05	G	*CLDN1, CLDN16, UNQ846*
rs1395064	18	5135389	6.80E-05	A	
rs3785396	16	23810657	6.93E-05	A	*PRKCB1*
rs3764617	19	54188935	7.32E-05	G	*DHDH, BAX, FTL, GYS1, RUVBL2, LHB, CGB, CGB1, CGB2, CGB5*
rs1355274	14	91187081	7.76E-05	G	*C14ORF161*
rs1858782	7	103131063	7.79E-05	A	*RELN*
rs17154438	7	103127334	7.89E-05	G	*RELN*
rs7665426	4	175562261	8.30E-05	A	*KIAA1712, HPGD*
rs195990	16	23853886	8.85E-05	C	*PRKCB1*
rs6929649	6	107087881	9.98E-05	G	*AIM1, RTN4IP1, QRSL1*

Table 5. The top 30 SNPs resulting from the logistic model are shown. A1 is the minor allele, for which an additive effect was tested.

SNP	*CHR*	*Position*	*P-value*	*A1*	*Genes within ± 100 kb*
rs2233236	8	141013881	5.29E-08	A	*TRAPPC9 (NIBP)*
rs3742883	14	67304292	1.20E-07	A	*VTI1B, RDH11, RDH12, ZFYVE26, RAD51L1*
rs13279037	8	29360277	1.62E-07	G	*DUSP4*
rs3825926	15	99262964	3.10E-07	A	*ALDH1A3, LRRK1*
rs1959287	14	86182374	4.22E-07	A	
rs1929045	13	89986067	1.50E-06	G	
rs13278369	8	29351631	1.82E-06	G	*DUSP4*
rs11136067	8	29361227	2.60E-06	G	*DUSP4*
rs938070	4	23854677	3.66E-06	G	
rs12521803	5	85267203	4.04E-06	G	
rs4145397	X	138132502	5.08E-06	A	*FGF13*
rs3803426	15	99272709	5.46E-06	A	*ALDH1A3, LRRK1*
rs10743811	12	33158390	6.43E-06	G	
rs1846630	12	33159356	6.43E-06	C	
rs7303471	12	33162179	6.43E-06	G	
rs4458444	4	23881377	7.12E-06	A	
rs1868294	2	223735088	7.14E-06	A	*KCNE4*
rs10181857	2	223724035	7.48E-06	G	
rs2849576	6	162506056	7.50E-06	A	*PARK2*
rs17365733	8	5805820	7.56E-06	C	
rs10505741	12	9970794	8.53E-06	A	*KLRF1, CLEC2B, KLRF2, CLEC2A, CLEC12A, CLEC1B, CLEC12B*
rs1469335	19	59734440	9.66E-06	G	*TTYH1, LENG8,LENG9, CDC42EP5, LAIR2, KIR3DX1, LILRA1, LILRB1*
rs11896279	2	54220084	1.04E-05	G	*ACYP2*
rs604788	2	33517854	1.06E-05	A	*LTBP1, RASGRP3*
rs1531170	4	156133128	1.11E-05	A	
rs16996345	22	45868217	1.20E-05	A	*TBC1D22A*
rs17154260	5	125617360	1.23E-05	G	
rs11241880	5	125620169	1.23E-05	A	
rs964186	5	125631525	1.23E-05	G	*GRAMD3*
rs2408613	5	125633005	1.23E-05	G	*GRAMD3*

Table 6. The top 30 SNPs resulting from the Cox proportional-hazard regression are shown. A1 is the minor allele, for which an additive effect was tested.

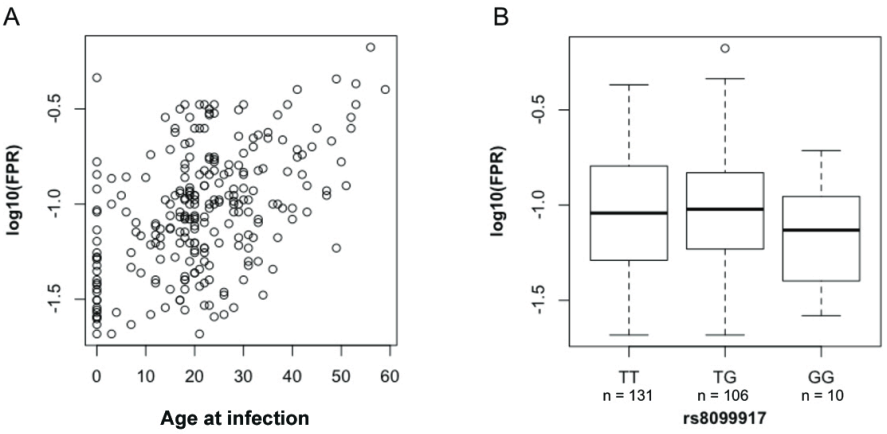


Figure 8. Effect of the age at infection and the IL28B genotype on the fibrosis progression. (A) A scatterplot between the age at infection (years) and the fibrosis progression (log_{10}FPR) shows that patients with younger age at infection have slower disease progression. (B) No association was found between IL28B rs8099917 genotype and disease progression. Boxplots depict the median value (bold lines), the interquartile range (boxes), the minimum and the maximum values (whiskers) and the outliers (circles).

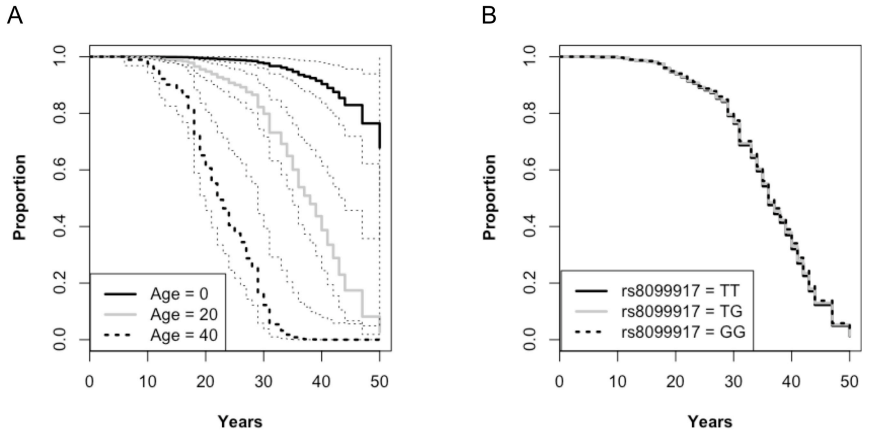


Figure 9. A younger age at infection reduces the hazard of developing advanced fibrosis, while the IL28B genotype has no effect. (A) Estimate survival functions for the Cox regression of the time to advanced fibrosis, on the age at infection, the gender and the HCV genotype. The proportion of patients not having advanced fibrosis is shown as a function of the disease duration in years. Three survival functions are estimated for three representative infection times (0, 20 or 40 years of age respectively), fixing other covariates (HCV genotype = 1 and gender = proportion of males). Thin dashed lines represent 95% confidence envelopes. (B) Estimated survival functions by rs8099917 genotype. IL28B has no effect on the hazard of advanced fibrosis. Three survival functions are estimated for the three rs8099917 genotypes, fixing other covariates (HCV genotype = 1, Age at infection = average value and Gender = proportion of males).

References

1. World Health Organization. Hepatitis C Fact Sheet No. 164 (Revised Oct 2000).
2. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002--June 10-12, 2002. Hepatology 2002;36:S3-20.
3. Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med 2001;345:41-52.
4. Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr
5. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 2001;358:958-65.
6. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Jr., Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;347:975-82.
7. Friedman SL. Mechanisms of hepatic fibrogenesis. Gastroenterology 2008;134:1655-69.
8. Friedman SL. Hepatic fibrosis -- overview. Toxicology 2008;254:120-9.
9. Seeff LB. Natural history of chronic hepatitis C. Hepatology 2002;36:S35-46.
10. Missiha SB, Ostrowski M, Heathcote EJ. Disease progression in chronic hepatitis C: modifiable and nonmodifiable factors. Gastroenterology 2008;134:1699-714.
11. Osterreicher CH, Stickel F, Brenner DA. Genomics of liver fibrosis and cirrhosis. Semin Liver Dis 2007;27:28-43.
12. Huang Y, Yang H, Borg B, Su X, Rhodes S, Yang K, Tong X, Tang G, Howell C, Rosen H, Thio C, Thomas D, Alter H, Sapp R, Liang T. A functional SNP of interferon-gamma gene is important for interferon-alpha-induced and spontaneous recovery from hepatitis C virus infection. Proc Natl Acad Sci U S A 2007;104:985-90.
13. Dai C, Chuang W, Lee L, Chen S, Hou N, Lin Z, Hsieh M, Hsieh M, Wang L, Chang W, Yu M. Associations of tumour necrosis factor alpha promoter polymorphisms at position -308 and -238 with clinical characteristics of chronic hepatitis C. J Viral Hepat 2006;13:770-4.
14. Yee LJ, Tang J, Herrera J, Kaslow RA, van Leeuwen DJ. Tumor necrosis factor gene polymorphisms in patients with cirrhosis from chronic hepatitis C virus infection. Genes Immun 2000;1:386-90.

15. Falleti E, Fabris C, Toniutto P, Fontanini E, Cussigh A, Bitetto D, Fornasiere E, Avellini C, Minisini R, Pirisi M. TGF-beta1 genotypes in cirrhosis: relationship with the occurrence of liver cancer. Cytokine 2008;44:256-61.
16. Powell EE, Edwards-Smith CJ, Hay JL, Clouston AD, Crawford DH, Shorthouse C, Purdie DM, Jonsson JR. Host genetic factors influence disease progression in chronic hepatitis C. Hepatology 2000;31:828-33.
17. Bakker PIWd, Telenti A. Infectious diseases not immune to genome-wide association. Nat Genet 2010;42:731-732.
18. Ge D, Fellay J, Thompson A, Simon J, Shianna K, Urban T, Heinzen E, Qiu P, Bertelsen A, Muir A, Sulkowski M, McHutchison J, Goldstein D. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009.
19. Study tHC, Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate M, Bassendine M, Spengler U, Dore G, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Müller T, Bahlo M, Stewart G, Booth D, George J. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nature Genetics 2009.
20. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nature Genetics 2009.
21. Rauch A, Kutalik Z, Descombes P, Cai T, di Iulio J, Mueller T, Bochud M, Battegay M, Bernasconi E, Borovicka J, Colombo S, Cerny A, Dufour J-F, Furrer H, Günthard HF, Heim M, Hirschel B, Malinverni R, Moradpour D, Müllhaupt B, Witteck A, Beckmann JS, Berg T, Bergmann S, Negro F, Telenti A, Bochud P-Y, Studies SHCaHC. Genetic Variation in IL28B Is Associated With Chronic Hepatitis C and Treatment Failure: A Genome-wide Association Study. Gastroenterology 2010.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
23. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.
24. Hinrichs AL, Larkin EK, Suarez BK. Population stratification and patterns of linkage disequilibrium. Genet Epidemiol 2009;33 Suppl 1:S88-92.
25. Terry Therneau and original R port by Thomas Lumley (2009). survival: Survival analysis, including penalised likelihood. R package version 2.35-7. http://CRAN.R-project.org/package=survival

26. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816-34.
27. Casiraghi MA, De Paschale M, Romanò L, Biffi R, Assi A, Binelli G, Zanetti AR. Long-term outcome (35 years) of hepatitis C after acquisition of infection through mini transfusions of blood given at birth. Hepatology 2004;39:90-6.
28. Tovo PA, Pembrey LJ, Newell ML. Persistence rate and progression of vertically acquired hepatitis C infection. European Paediatric Hepatitis C Virus Infection. J Infect Dis 2000;181:419-24.
29. Thomas D, Thio C, Martin M, Qi Y, Ge D, O'huigin C, Kidd J, Kidd K, Khakoo S, Alexander G, Goedert J, Kirk G, Donfield S, Rosen H, Tobler L, Busch M, McHutchison J, Goldstein D, Carrington M. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009.
30. McCarthy JJ, Li JH, Thompson A, Suchindran S, Lao XQ, Patel K, Tillmann HL, Muir AJ, McHutchison JG. Replicated association between an IL28B gene variant and a sustained response to pegylated interferon and ribavirin. Gastroenterology 2010;138:2307-14.
31. Huang H, Shiffman ML, Friedman S, Venkatesh R, Bzowej N, Abar OT, Rowland CM, Catanese JJ, Leong DU, Sninsky JJ, Layden TJ, Wright TL, White T, Cheung RC. A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C. Hepatology 2007;46:297-306.
32. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet 2006;38:659-62.
33. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009;10:241-51.
34. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-61.
35. Poynard T, Ratziu V, Charlotte F, Goodman Z, McHutchison J, Albrecht J. Rates and risk factors of liver fibrosis progression in patients with chronic hepatitis c. JOURNAL OF HEPATOLOGY 2001;34:730-9.
36. Liebermann DA, Hoffman B. Gadd45 in the response of hematopoietic cells to genotoxic stress. Blood Cells Mol Dis 2007;39:329-35.

37. Cretu A, Sha X, Tront J, Hoffman B, Liebermann DA. Stress sensor Gadd45 genes as therapeutic targets in cancer. Cancer Ther 2009;7:268-276.
38. Gramantieri L, Chieco P, Giovannini C, Lacchini M, Treré D, Grazi GL, Venturi A, Bolondi L. GADD45-alpha expression in cirrhosis and hepatocellular carcinoma: relationship with DNA repair and proliferation. Hum Pathol 2005;36:1154-62.
39. Zhu N, Shao Y, Xu L, Yu L, Sun L. Gadd45-alpha and Gadd45-gamma utilize p38 and JNK signaling pathways to induce cell cycle G2/M arrest in Hep-G2 hepatoma cells. Mol Biol Rep 2009;36:2075-85.
40. Higgs MR, Lerat H, Pawlotsky J-M. Downregulation of Gadd45beta expression by hepatitis C virus leads to defective cell cycle arrest. Cancer Research 2010;70:4901-11.
41. Bae M-A, Rhee SD, Jung WH, Ahn JH, Song B-J, Cheon HG. Selective inhibition of activated stellate cells and protection from carbon tetrachloride-induced liver injury in rats by a new PPARgamma agonist KR62776. Arch. Pharm. Res. 2010;33:433-42.
42. Barrow RE, Dasu MRK. Oxidative and heat stress gene changes in hypertrophic scar fibroblasts stimulated with interleukin-1beta. J Surg Res 2005;126:59-65.
43. Inagaki Y, Okazaki I. Emerging insights into Transforming growth factor beta Smad signal in hepatic fibrogenesis. Gut 2007;56:284-92.
44. van der Heide LP, van Dinther M, Moustakas A, Ten Dijke P. TGF{beta} activates mitogen- and stress-activated protein kinase-1 (MSK1) to attenuate cell death. J Biol Chem 2010.
45. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 1997;349:825-32.
46. Wright M, Goldin R, Fabre A, Lloyd J, Thomas H, Trepo C, Pradat P, Thursz M. Measurement and determinants of the natural history of liver fibrosis in hepatitis C virus infection: a cross sectional and longitudinal study. Gut 2003;52:574-9.
47. Mir A, Kaufman L, Noor A, Motazacker MM, Jamil T, Azam M, Kahrizi K, Rafiq MA, Weksberg R, Nasr T, Naeem F, Tzschach A, Kuss AW, Ishak GE, Doherty D, Ropers HH, Barkovich AJ, Najmabadi H, Ayub M, Vincent JB. Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am J Hum Genet 2009;85:909-15.
48. Philippe O, Rio M, Carioux A, Plaza JM, Guigue P, Molinari F, Boddaert N, Bole-Feysot C, Nitschke P, Smahi A, Munnich A, Colleaux L. Combination of linkage mapping and microarray-expression analysis identifies NF-kappaB signaling defect as a cause of autosomal-recessive mental retardation. Am J Hum Genet 2009;85:903-8.
49. Hu WH, Pendergast JS, Mo XM, Brambilla R, Bracchi-Ricard V, Li F, Walters WM, Blits B, He L, Schaal SM, Bethea JR. NIBP, a novel NIK and IKK(beta)-binding protein that enhances NF-(kappa)B activation. J Biol Chem 2005;280:29233-41.

50. Mann J, Mann DA. Transcriptional regulation of hepatic stellate cells. Advanced Drug Delivery Reviews 2009;61:497-512.
51. Elsharkawy AM, Mann DA. Nuclear factor-kappaB and the hepatic inflammation-fibrosis-cancer axis. Hepatology 2007;46:590-7.
52. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008;322:881-8.

Chapter 3. Circulating microRNAs miR-122 and miR-885-5p are promising markers for detecting disease progression in HCV-associated liver pathologies

The data presented in this chapter are part of an ongoing project.
The manuscript is in preparation and the contributors are listed below.

Francesco Marabita

Raoul J.P. Bonnal
Serena Curti
Paola de Candia
Lorena Donnici
Riccardo L. Rossi
Anna Torri

Ferruccio Bonino
Flavio Caprioli
Mario Comelli

Sergio Abrignani
Raffaele De Francesco
Massimiliano Pagani

Abstract

Circulating microRNAs (miRNAs) are emerging as a novel class of biomarkers for non-invasive diagnosis of several diseases. Therefore, in this study we investigated whether specific serum miRNA signatures may be detected in the serum of patients with chronic hepatitis C (HCV) infection at different stages of liver disease. Individual sera from healthy controls, patients with chronic hepatitis (CH), liver cirrhosis (LC) or hepatocellular carcinoma (HCC) were tested by high-throughput qPCR to profile the expression of the whole miRNome. A solid normalization strategy was applied to reduce variability, and 22 miRNAs were considered differentially represented among the four classes ($p<0.05$ after multiple testing correction). miR-122, miR-885-5p, miR-509-3p, miR-610 and miR-760 were selected and further characterized. These miRNAs showed increasing concentration as the disease progressed to LC and HCC. The levels of miR-122 and miR-885-5p specifically and consistently increased in patients with HCV infection compared to healthy controls or patients with Cronh's disease ($p<0.001$). miR-122 concentration changed significantly among groups with different alanine aminotransferase (ALT) activity ($p<0.001$), showing a positive trend. In summary, serum levels of miR-122 and miR-885-5p were significantly elevated in patients with HCV-associated liver pathologies. Our data show the potential use of serum miRNA as novel biomarker in the setting of chronic HCV infection.

Introduction

MicroRNAs (miRNAs) are small noncoding RNAs of 22-25 nucleotides that regulate gene expression by binding mostly the 3' untranslated region (3'UTR) of target mRNAs. They are broadly conserved and act post-transcriptionally directing target gene repression.[1, 2, 3] Their impaired expression has been associated with a variety of diseases, including cancers, inflammation and chronic viral infections.[4-7] Accumulating evidences support the use of miRNAs circulating in blood as non-invasive biomarkers of disease conditions,

particularly cancer.[8] miRNAs are stably detected in serum or plasma of healthy and diseased subjects and their levels do not appear to be dramatically influenced by multiple freeze-thaw cycles, long term storage or treatment with RNase before extraction.[9-11] Moreover, in some tissues particular miRNAs are expressed in specific manner, correlating to disease condition[4, 5, 12]. Hence, circulating miRNAs represent a promising alternative for the diagnosis of several pathological conditions and the monitoring of disease progression.

Hepatitis C Virus (HCV) is a positive-strand RNA virus that establishes a persistent infection in more than 70% of people who contract it. The chronic infection progresses to severe liver disease in about 20% of the persistently infected subjects, with development of liver cirrhosis and hepatocellular carcinoma (HCC).[13] HCC is the leading cause of death among cirrhotic patients and the third cause of cancer-related mortality.[14] Standard therapy with pegylated Interferon-γ and ribavirin fails to clear HCV in ~ 50% of chronically infected individuals.[15,16] Moreover, no vaccination is currently available. Classical biochemical markers to follow disease progression show limited potential mainly due to their invasiveness or little specificity and sensitivity, especially for HCC. Alanine aminotransferase (ALT) activity in the blood is the most widely used biochemical marker for measuring liver injury. However, ALT levels may fluctuate during HCV infection and values may even fall into the normal range.[17] Moreover, ALT has been shown to be increased in clinical disorders other than liver diseases.[18, 19] α-Fetoprotein (α-AFP) has long been used in clinic for the diagnosis of primary HCC, showing nonetheless unsatisfying sensitivity and specificity.[20] Furthermore, elevated AFP levels may be observed in patients with cirrhosis or flares of active hepatitis.[21, 22] Thus, there is a crucial need for novel, preferably non-invasive biomarkers that could be used for the detection of HCC in order to improve the diagnosis and the management programs for patients at high risk.

Here we investigate the hypothesis that specific serum miRNA signatures may be found in patients with different stages of HCV-associated liver diseases, given that the differential expression and the regulatory roles of miRNAs in liver pathologies, including chronic HCV and HCC, have been observed in the

hepatic tissue.[23-29] To prove this hypothesis, human sera from healthy donors and HCV-infected patients were tested by high-throughput qPCR and data were processed with specific bioinformatic tools, in order to profile the expression of the whole miRNome. This study shows that miR-122 and miR-885-5p are specifically increased in the serum of patients with chronic liver inflammation, highlighting their potential as blood biomarker of liver disease.

Materials and methods

Patients characteristics

A discovery panel of frozen human sera from patients infected with HCV was obtained from U.O. Epatologia, Azienda Ospedaliera Universitaria Pisana (Pisa). These sera were originally collected from 40 patients with chronic HCV infection and included: 10 patients with chronic hepatitis but without liver cirrhosis or HCC (hereafter CH), 15 patients with liver cirrhosis (hereafter LC) and 15 patients with liver cirrhosis and HCC (hereafter HCC). Details about the clinical characteristics are given in Table 1. Additionally, a group of sera from 10 healthy donors (HD) and 12 patients with Crohn's disease during active phase (CD) were included in this study as controls. These sera were obtained from U.O. Centro Trasfusionale (for HD) and U.O. Gastroenterologia (for CD), Ospedale Maggiore di Milano.

RNA extraction

Total RNA was extracted from human serum using miRVana miRNA isolation kit (Ambion), as specified in the provided protocol, with some modifications. Briefly, 70 µl of thawed serum were mixed with 530 µl of lysis solution composed of RNA Lysis Buffer and synthetic ath-miR-159a (final concentration 5 pM). This miRNA will be used as a process control, for technical normalization. To reduce sample-to-sample variation in quantification of ath-miR-159a, a mix of RNA Lysis Buffer and ath-miR-159a was prepared at the moment and then mixed to the sera being processed. RNA was extracted with

acid phenol-chloroform, then further purified with glass-fiber columns and finally eluted in 50 µl of nuclease-free water. RNA was then stored at -80 °C.

microRNA profiling

3 µl of total RNA were processed for Reverse Transcription and Preamplification with Megaplex Primer Pools A v2.1 and B v2.0 (Applied Biosystems), according manufacturer instruction. TaqMan Low Density Arrays (Applied Biosystems) were run on a 7900HT Fast Real-Time PCR System. A total of 664 human miRNAs, 6 human small RNA and 1 control miRNA from A. Thaliana were profiled in parallel. Ct values were extracted using RQ Manager, setting a manual threshold of 0.06. Multiple samples were assembled with StatMiner software.

Analysis workflow and definitions

Figure 1 shows the analytical workflow used throughout this study. Quantities are defined in this section. Relative Quantities (RQs) are those resulting after scaling Ct values with the external reference ath-miR-159a and moving to the linear scale:

$$RQ = 2^{-\Delta Ct}$$

$$\Delta Ct = Ct_{miRNA} - Ct_{ath-miR-159a}$$

The Normalization Factor (NF) is calculated as the geometric mean of the selected *k* normalizers, for each sample *j*. The NF is used to obtain the matrix of Normalized Relative Quantities (NRQ) for each miRNA *i* and sample *j*, starting from RQ:

$$NRQ_{ij} = \frac{RQ_{ij}}{NF_j}$$

Alternatively, NRQ may be obtained using a NF resulting from the arithmetic or geometric mean of all expressed miRNAs per sample, i.e. the mean obtained omitting detectors whose Ct is undetermined (Ct = 40).

Data normalization

Three different algorithms were used to reveal the best internal reference miRNAs, namely geNorm,[30] Normfinder[31] and a scoring on the basis of the

Coefficient of Variation (CV score). miRNAs detectable in all samples were selected. Moreover, we took ten samples per group, in order to provide equal group sizes. Outliers resulting from artificial, non-exponential amplifications were identified and removed after visual inspection of the amplification curves. geNorm was run using the `SLqPCR` library in R.[32] The CV score algorithm is described as follows: from the matrix of RQ (*m* detectors, *n* samples), calculate a new matrix **X** for miRNA *i* and sample *j* where:

$$X_{ij} = \frac{RQ_{ij}}{\sum_{i=1}^{n} RQ_{ij}}$$

Then, for each miRNA *i*, calculate the CV of X_{i*}. This parameter is useful to evaluate stability across samples. Actually the sum of all RQ for each sample *j* can be considered a surrogate of the total miRNA amount. As a consequence, the lesser the CV of X_{i*}, the better the miRNA *i* may be considered as normalizer. By definition, a normalizer must be correlated to the total miRNA amount.

The three algorithms output a score that represents the stability of the candidate miRNA in the expression dataset, in such a way that a smaller score corresponds to higher expression stability. The top candidate miRNAs were selected among those performing well in all three methods. To summarize the results, the distance from the origin in 3-dimensional space was used as a metric:

$$Distance = \sqrt{\left(Score_{geNorm}\right)^2 + \left(Score_{Normfincer}\right)^2 + \left(Score_{CV}\right)^2}$$

Ranking distances in ascending order, the final miRNA reference list was obtained.

Principal Component Analysis (PCA) was exploited as an independent method to look at variation in RQ. The input list of miRNAs detectable in all samples miRNAs was submitted to PCA, after $\log_{10}$-transformation. Both unscaled and autoscaled RQs were used. Autoscaled RQs for miRNA *i* and sample *j* are defined as follows:

$$ARQ_{ij} = \frac{\log_{10}(RQ_{ij}) - \overline{\log_{10}(RQ_{i*})}}{st.dev(\log_{10}(RQ_{i*}))}$$

where $\overline{\log_{10}(RQ_{i^*})}$ and $st.dev(\log_{10}(RQ_{i^*}))$ are the mean and the standard deviation of $\log_{10}$(RQ), respectively. Autoscaled data have mean zero and standard deviation of one.

Differential expression analysis

Data were loaded on MeV (v4.5) for testing differential expression and clustering. For two-class data, unpaired T-test was used to assess differential expression. Alternatively, ANOVA was used for multi-class data. Data were considered significant if Bonferroni-corrected $p<0.05$. Hierarchical clustering (HCL) was performed on samples and significant miRNAs, using Euclidean distance and complete linkage. HCL was run both with MeV and R, with the package cluster. Partition Around Medoids (PAM) clustering was performed in R using the package cluster, specifying *k*.

Exosome purification

Exosomes were purified by differential centrifugation as described previously.[33] Huh7.5 cells were cultured in exosome-free medium (DMEM high glucose, 10% FBS). Supernatants were collected after 48h from 3×10^8 cells. Conditioned medium was centrifuged for 10 min at 2000 × g and 30 min at 10,000 × g. The exosome-containing pellet was obtained centrifuging for 70 min at 100,000 × g and washing once in PBS for 70 min at 100,000 × g. For exosome purification from serum, 9 ml of serum were diluted with PBS and centrifuged for 30 min at 2,000 × g, for 45 min at 12,000 × g and for two hours at 110,000 × g. The pellet was resuspended in PBS, filtered through a 0.22 μm filter, centrifuged for 70 min at 110,000 × g and washed once in PBS for 70 min at 110,000 × g. RNA was extracted using the miRVana miRNA isolation kit (Ambion) as specified, adding ath-miR-159a at final concentration of 5 pM within RNA Lysis Buffer. For miRNA profiling, 200 ng of total RNA from Huh7.5 were used, while for exosome preparation a fixed amount of 3 μl was used, since the amount of RNA was below the limit for reliable quantification.

qPCR

5 µl of total RNA were processed for Reverse Transcription using the TaqMan miRNA Reverse Transcription Kit and miRNA-specific stem-loop primers (Applied Biosystems) according to manufacturer instructions. For real-time PCR, 1 µl of RT product were combined with TaqMan Universal PCR Master Mix, No AmpErase and TaqMan miRNA Assay (Applied Biosystem) in 15 µl of final volume. RQ were calculated using ath-miR-159a as reference, as specified above. To profile the expression of miR-122, miR-885-5p, mir-509-3p, miR-610 and miR-760 in a panel on human tissues, 10 ng of commercial RNA were processed for RT (FirstChoice Human Total RNA Survey Panel, Ambion).

Results

Reference miRNAs

The ultimate goal of data normalization was ensuring an adequate reduction of the experimentally-induced variability. The process workflow and the explanation of the variables are shown in Figure 1. We used a two-level normalization strategy. Firstly, we used a synthetic miRNA (ath-miR-159a) as spike-in control, within the RNA lysis buffer. With the expression of the miRNA levels relative to this external reference, we provided a quality control for differences in extraction or reverse transcription efficiencies. To further reduce experimental variability, we also used an endogenous control to eliminate variability linked to different quality and quantity of the starting RNA amount. This endogenous reference was either a global parameter, such as the mean of all expressed miRNAs, or a normalization factor calculated from the levels of the reference miRNAs.

Three different algorithms were used to reveal the best internal reference miRNAs (geNorm, Normfinder, CV score). A list containing the miRNAs detected in all samples and the arithmetic mean, the geometric mean and the median value of RQs was used as input. The final ranking was obtained summarizing the three scores, i.e. calculating the distance in a three-

dimensional space (Figure 2). There is a good agreement between the three methods (multiple $R^2 = 0.74$), suggesting that in principle an accurate selection of low-scored miRNAs can be used to calculate the NF. Global parameters, such as the arithmetic mean of all detected miRNA RQs, are ranked among the top items, indicating it is a good normalization factor. miRNAs showing a normalization performance similar to the mean expression value have been proposed as candidate normalizers.[34] From this analysis miR-17, miR-484 and miR-126 resulted the three most trusted internal references in this dataset, being stably expressed among all groups. NRQs were obtained using the geometric mean of these three miRNAs. Alternatively, NRQs were obtained using the global arithmetic or geometric mean of all expressed miRNAs per sample. miRNAs showing a normalization performance similar to the mean expression value have been proposed as candidate normalizers.[34] Hence, a combination of reference miRNAs may be suitable in reducing data variation. To test this hypothesis, the effective reduction in data variability was evaluated looking at the cumulative distribution of the coefficient of variation of NRQ and not-normalized data (Figure 4). Finally, two alternative normalization strategies were selected: the global geometric mean and the NF resulting from the three top miRNAs. The distribution of miRNA expression is shown in Figure 4.

Another powerful and independent approach in classifying miRNAs on the basis of their expression profile is PCA. A subset of patients with LC or HCC was selected for this analysis. By PCA candidate normalizers were tested and further evaluated. In Figure 3, the top panel shows a plot of the PCs for autoscaled data. Analysis of unscaled data takes into account the extent of the changes in a relative expression dataset, which is reported by the first PC. Therefore, the most informative components are usually PC2 and PC3. Whereas, for autoscaled data PC1 vs PC2 plot is helpful in identifying and evaluating reference miRNAs.[35] Interestingly, miRNAs scored as the stablest ones were clustered together at the narrow end of a funnel-like structure, showing that they are less variable with respect to the others. On the contrary, more variable miRNAs, such as those differentially expressed, should be located

away from the cluster of unvarying elements. miR-122, miR-192, and miR-885-5p appeared well separated in the plot from the group of candidate reference miRNAs, suggesting they carry a valuable information content in discriminating LC and HCC patients.

Differentially represented miRNAs

After normalization, miRNAs were filtered on the basis of their detectability, considering as expressed those miRNAs being detected in at least two third of samples per group. Hence, 91 miRNAs were further tested for differential expression and then clustered. For multi-class data, ANOVA was used. Results were considered significant if Bonferroni-corrected $p<0.05$. Hierarchical clustering was performed on samples and miRNAs, using significant results only. The list of significant miRNAs is shown in Table 3. This list was obtained using data normalized with a global parameter, namely the mean of RQs of all expressed miRNA, or alternatively using the selected reference miRNAs (miR-484, miR-17 and miR-126), which resulted the most stably expressed. Remarkably, 21 miRNAs were common between these two strategies and also the corresponding clustered heatmap appeared comparable (see Figure 5 and Table 3). This first evaluation revealed some interesting miRNAs. Notably, the liver-specific miR-122 was among them. Using these differentially represented miRNAs for hierarchical clustering, the two higher-order clusters mirrored a biological separation, showing a first group enriched in HD and CH patients and a second group populated mostly with patients with LC and HCC (Figure 5).

miR-122 and miR-885-5p specifically increase in the sera of patients with liver diseases

The cluster of miRNAs containing miR-122 and miR-885-5p was selected for further investigation. Actually, biological support exists for the association of miR-122 in hepatocarcinogenesis: miR-122 is the most expressed miRNA in normal liver and its level has been shown to vary in HCC.[23-28] Moreover, miR-122 is specifically required for HCV replication and infectious virus production.[7, 36] Here, a pattern may be recognized, mainly involving the coordinate expression of miR-122, miR-885-5p, miR-509-3p, miR-610 and

miR-760 (Figure 6). The representation of these miRNAs was lower in healthy individuals but increased progressively in patients with chronic HCV infection, liver cirrhosis and HCC, mirroring the progression of the disease (Figure 7). A closer look revealed that their levels were similar between patients with liver cirrhosis alone and patients with liver cirrhosis and cancerous lesions. Applying an unsupervised hierarchical clustering approach we demonstrated that this five-miRNA signature might differentiate healthy donors, chronic patients and cases with severe liver diseases (LC and HCC). This clustering revealed a remarkable classification, where a HD-enriched cluster (67% HD) is separated from a diseased-enriched cluster. Moreover, from this latest group a class containing almost only LC and HCC patients may be further separated from a class containing CH patients (Figure 6). As a further control, we included a group of sera from patients with Cronh's disease (CD) that were collected during the active phase. Interestingly, only miR-122 and miR-885-5p were specifically and consistently increased in subjects with chronic liver disease, but not in healthy controls or patients with CD. On the contrary, miR-509-3p, miR-610 and miR-760 were not homogeneously and significantly different between CD and all the three groups of HCV-infected patients (CH, LC and HCC). To gain further insight into the cellular origin of miRNAs belonging to the cluster described above, a panel of RNA from different human tissues was profiled (Figure 8). The well-know liver-specific miR-122 was highly abundant in liver tissue and miR-885-5p showed a moderate expression in liver, brain and skeletal muscle, suggesting that the liver may be responsible for the observed increase in sera of patients with HCV. Conversely, miR-760 was ubiquitous in the analyzed tissues, mir-509-3p was expressed in kidney, ovary and testes and miR-610 was overall poorly expressed.

The miRNA profiling was performed also in Huh7.5 cells, a clone from Huh7 cells that supports higher levels of HCV RNA replication as compared to parental Huh7 cells.[37] These cells were originally derived from a well-differentiated HCC, and are reported to contain higher levels of miR-122 than other hepatoma cell lines (i.e. HepG2).[7, 36] miR-122 and miR-885-5p resulted among the first ten most abundant miRNAs. Conversely, miR-509-3p and

miR-610 were detectable but showed extremely low expression and miR-760 was approximately thirty times less abundant than miR-122.

We also isolated the exosomes released *in vitro* from Huh7.5, using ultracentrifugation. Interestingly, miR-885-5p, but not miR-122, was abundant also in the final exoxome-containing pellet (Figure 8). We then evaluated the *in vivo* representation of miRNAs in matched samples of whole serum and serum-derived exosomes obtained from fresh serum specimen from four healthy donors. We detected 185 miRNAs in at least three out of four samples in the total serum (Ct<35), whereas 86 miRNAs were detected in the exosome preparations, in at least three out of four samples. The vast majority (98%) of miRNAs detected in exosomes were also present in the total serum (Figure 8). However, despite this large overlap between the whole serum and the exosomes, miR-122 and miR-885-5p were always detected in the serum but never in the exosome preparation, while miR-509-3p, miR-610 and miR-760 were detectable in both.

miR-122 correlates with ALT levels

In order to explain the accumulation of liver-related miRNAs in the blood circulation, we further analyzed a panel of serum samples from patients with different levels of alanine aminotransferase (ALT) activity, a marker commonly used for the assessment of liver disease and damage. To correlate ALT activity and serum concentration of miR-122, we divided the patients into 4 groups, on the basis of their ALT levels. (group I, ALT ≤ 80 U/l; group II, 80 U/l < ALT ≤100 U/l; group III, 100 U/l < ALT ≤200 U/l; group IV, ALT>200 U/l). We calculated the miR-122 RQs using the ath-miR-159a as reference. miR-122 concentration changed significantly among groups ($p = 0.0003$, Kruskal-Wallis test), showing a positive trend with increasing ALT activity (Figure 9).

Discussion

In this work we described a miRNA profiling in the serum of patients with chronic HCV infection. We observed that the level of organ-specific miRNAs is associated with the disease status. Specifically, mir-122 and miR-885-5p were

over-represented in the sera of patients with chronic HCV infection, with increasing concentration as the disease progressed to advanced phases, such as liver cirrhosis and HCC. Moreover, this increased representation was specific to chronic liver inflammation.

A solid experimental workflow for the evaluation of serum miRNA expression identified miRNAs that could be used as endogenous controls.

Over the last two years, accumulating evidence supported the use of miRNAs circulating in blood as non-invasive biomarkers of disease conditions, particularly cancer.[8] These miRNAs are stably detected in the serum or plasma of healthy and diseased subjects and their levels did not appear to be dramatically influenced by multiple freeze-thaw cycles, long term storage or treatment with RNase before extraction.[9-11] Lawrie et al. suggested that miR-155, miR-21 and miR-210 have potential as non-invasive diagnostic markers for Diffuse Large B-Cell Lymphoma.[38] In another key study, Mitchell et al. reported that serum levels of miR-141 discriminated between patients with prostate cancers and healthy controls.[10] The importance of circulating miRNAs was further documented by a number of successive works, showing the potential of cell-free circulating miRNAs as disease biomarkers not only for cancer (lung,[39] colorectal,[40, 41], gastric,[42] ovarian,[43] pancreatic,[44] breast[45] and skeletal muscle[46] cancer), but also for acute myocardial infartion,[47] and liver pathologies including hepatotoxicity, chronic hepatitis and HCC.[48-52]

A common pitfall of some of these works is the lack of a solid normalization strategy to account for interindividual or intergroup variability. Most of the results were obtained by Real-time qPCR, using different reference genes. When quantifying cellular miRNA, stable small RNA controls are currently used as reference RNAs. These include small non-coding RNAs (ncRNA), and specifically small nuclear RNA and small nucleolar RNA such as RNU44, RNU48 and U6. For serum miRNAs, there is growing evidence that the above-mentioned small RNAs are highly variable or not stably detectable,[8] thus leading to the investigation of suitable stable control miRNAs that are firmly detectable in human serum. The most common choice has been the use of a

miRNA that does not vary considerably between individuals. However, this selection may be arbitrary when based on the Ct data only, without the assessment of the actual stability by suitable *wet* and *in silico* analyses. In other words, "*the absence of evidence is not evidence of absence*" of differential expression. In our laboratory experience, the yield of RNA from small volume plasma or serum samples (i.e., 100–400 μl) is below the limit of accurate quantification by spectrophotometry, a finding that is confirmed by other groups[53]. Therefore, a normalization based on a fixed amount of starting RNA is routinely impracticable. Here, we used a fixed volume of starting RNA, although we were aware that little is known about the relationship between the total RNA amount in serum and the corresponding miRNA fraction. We developed a solid experimental workflow for the evaluation of serum miRNA expression starting from clinical samples, in a retrospective manner, using limited amount (70 μl) of frozen serum. The first challenge we met was to provide a robust normalization strategy, with the goal of minimizing data variation. The effect of a normalization process is highlighting true biological changes and eliminating, or at least reducing, the variability introduced during the whole experimental process. For array-based approaches, a normalization strategy based on the mean value of miRNA expression has been proposed for qPCR data.[34] This method is only applicable when high-throughput miRNA profiling is performed. Obtaining a number of solid normalizers is extremely advantageous, considering the prerequisite of reducing the number of reference miRNA and increasing the throughput and the feasibility of single or focused qPCR assays. Therefore, we focused on the search of candidate reference miRNA, using a combination of validated bioinformatic approaches. We successfully obtained candidate reference miRNAs that were valuable in reducing variability in the data. When data were normalized with miR-17, miR-126 and mir-484 as endogenous controls, those miRNAs identified as being differentially expressed remained significant as shown by even lower p-values. miR-17 lies in a well-known cluster of six miRNAs located in the third intron of a primary transcript (C13orf25) with unknown function. miR-17-92 cluster is involved both in development and in disease, having crucial functions in cancer, organ

development and immune system.[6, 54] miR-484 was firstly identified as human fetal liver miRNA but no well-established functions are attributed to this miRNA, except for its association with adrenocortical diseases.[55, 56] miR-126 is abundantly expressed in endothelial cells, playing important roles in vascular physiopathology.[196] Moreover, its role in the immune system has also been established.[6, 57, 58] Very interestingly, these miRNAs were consistently detected in the exosome preparation from this work and in another report.[59]

Circulating microRNAs are promising markers for detecting disease progression in HCV-associated liver pathologies. Here, we investigated the differential expression of serum miRNAs in the setting of chronic HCV infection, showing that the levels of organ-specific miRNAs are associated with the disease status. After multiple testing correction, we obtained a signature of significant miRNAs that were differentially represented between healthy donors and patients with chronic HCV infection, liver cirrhosis or HCC. Although the Bonferroni correction can be too conservative, we were particularly interested in reducing the false positive rate. Using the list of differentially represented miRNAs, we obtained an unsupervised hierarchical classification that correctly classified the majority of healthy subjects and diseased patients. The Partitioning Around Medoid (PAM) algorithm, a partitioning method that is considered to outperform hierarchical clustering,[60] confirmed this classification. Actually, 84% of samples were correctly classified, specifying three groups, corresponding to HD, CH and LC+HCC (data not shown). Notably, the HD group was homogeneously separated from the patients and only three HCC subjects were misclassified.

Starting from the list of significant miRNAs, we focused on the cluster of miRNAs containing miR-122 and miR-885-5p. miR-122 is the most expressed microRNA in normal liver and its levels have been shown to vary in HCC.[23-28] It is specifically required for HCV replication and infectious virus production.[7, 36] Moreover, in a mouse model of experimental drug-induced liver injury, the levels of miR-122 increased in plasma and decreased in liver at the same time, upon acetaminophen exposure.[61] Notably, the mouse does not have a homolog

for miR-885-5p. More recently, other findings suggested that circulating miR-122 levels are correlated to liver injury also in man.[52] Furthermore, from the analysis on normalizers, the liver-expressed miRNAs miR-122 and miR-885-5p emerged among the least stable and the PCA clustered them together, suggesting that they could have a common biological origin, reflected in correlated levels in the serum. For all these reasons, the cluster of genes containing miR-122 and miR-885-5p was extracted and cluster analysis was performed. We demonstrated that a five-miRNA signature might differentiate healthy donors, chronic patients and cases with severe liver diseases (LC and HCC). Moreover, miR-122 and miR-885-5p were specifically increased in patients with liver diseases compared to HD and CD. Indeed, CD is a form of inflammatory bowel disease characterized by a marked infiltration into the intestinal lamina propria of immune cells of the innate and adaptive compartment, mediated by an overactive Th17 and Th1 proinflammatory cytokine response.[62] Hence, this condition represented an ideal control to check for the presence of circulating miRNAs in another disease characterized by immune-mediated tissue injury, similarly to chronic HCV infection. Therefore, miR-122 and miR-885-5p were significantly and consistently increased in patients with chronic HCV, coherently with the hepatic origin of these miRNAs.

Because miR-122 is the most expressed miRNA in adult liver, its role in tumorigenesis has been long investigated. Indeed, miR-122 levels have been shown to vary in HCC, sometimes with opposite trend, probably reflecting different etiologies.[23-28] Generally, a decrease in miR-122 levels in HCC has been documented, although one study reported an up-regulation in HCV-associated HCC.[26] miR-122 is considered a tumor-suppressor within the liver. Actually, Cyclin G1 was identified as a target of miR-122,[24,63] and the loss of miR-122 expression in HCC cells was associated with increased cell migration and invasiveness.[64] How to correlate the common decrease in expression of miR-122 observed during tumorigenesis and the increased representation in the serum observed in this study? The changes in the concentration of a specific miRNA in the serum may reflect both variation in their expression within the originating tissue and differential release within the blood circulation. The

latter mechanism accounts for a combination of passive release, mainly related to cell death, and controlled release, mainly through exosomes and microvesicles containing miRNAs.[65-67] Changes in the release from the originating organ are a reasonable option during a disease characterized by chronic and progressive liver injury, as observed during HCV infection. Hence, it is conceivable that the reported differences with healthy individuals in the levels of miR-122 and miR-885-5p are attributable to increased release from hepatocytes of patients with chronic HCV, liver cirrhosis and HCC. While there was a very large overlap between the miRNAs detectable in the whole serum and in the exosomes, the preferential exclusion of miR-122 and miR-885 in the exosome compartment observed in this study may be suggestive of a passive release of these miRNA from the liver, after necrosis or apoptosis and accumulation of these nucleic acids within cell debris, protein complexes or apoptotic bodies, also in healthy individuals. Consistently with this hypothesis, we detected an association of miR-122 with ALT levels in the serum, the most widely used biochemical marker for measuring liver injury. This observation is in line with previous observations in animal models and humans showing a correlation between serum levels of miR-122 and hepatic damage.[52, 61]

Taken together, these results demonstrate that in patients with chronic HCV infection, miR-122 and miR-885-5p may be considered hallmarks of the presence of hepatic miRNAs in the blood circulation, although the precise mechanism of release from the hepatocytes remains to be better elucidated. These miRNA are specifically increased in the serum of patients with chronic liver inflammation, showing their potential as blood biomarker of liver disease.

During the preparation of this manuscript, Gui et al. also reported that increased levels of miR-885-5p were detected in the serum of patients with liver cirrhosis and HCC.[48] The authors did not report any association with other biochemical marker of liver disease (ALT, α-fetoprotein or aspartate transaminase), although they observe a specific increase in the serum of patients with liver pathologies. Their normalization strategy involved the use of U6 small nuclear RNA, even though this strategy have been controversial for circulating miRNAs,[8] also reporting very high Ct values for this reference RNA –

a possible source of less precise determinations. Moreover, their cohort included advanced liver disease mainly with HBV etiology, while we focused on chronic HCV infection.

In conclusion, we established the promising role of circulating miRNAs as biomarkers for liver diseases. However, our work represented a small-scale study, designed to screen for potential miRNAs to be further validated. In a second-phase validation study, we will assess the strength of the reported association, the sensitivity and specificity, the effect size and the usefulness of circulating miRNAs as biomarker of liver diseases. Moreover, a large-scale *prospective* study would be extremely advantageous to investigate the role of circulating miRNAs in the progression from liver cirrhosis to HCC. In future work, we anticipate the exploration of the physiological role of significant miRNAs in liver biology, by looking at their cellular targets and regulatory networks.

Table 1. Clinical characteristics of the patients

	Healthy Donors (HD) (n=10)	***Chronic Hepatitis (CH) (n=10)***	***Liver Cirrhosis (LC) (n=15)***	***Hepatocellular Carcinoma (HCC) (n=15)***
Sex – No. (%)				
Male	9 (90%)	5 (50%)	10 (67%)	9 (60%)
Female	1 (10%)	5 (50%)	5 (33%)	6 (40%)
Age - years				
Mean (SD)	42.2 (8.2)	52 (17)	60.7 (7.4)	71.6 (9.8)
Median (range)	37.5 (33 - 58)	48 (25 - 78)	63 (47 - 70)	77 (62 - 87)
ALT - No. (%)				
High, > 50 IU/l	-	8 (80%)	11 (73%)	13 (87%)
Normal, < 50 IU/l	10 (100%)	2 (20%)	4 (27%)	2 (13%)
Tumor size – No. (%)				
≤ 3 cm	-	-	-	10 (66%)
> 3 cm	-	-	-	5 (33%)
Multinodularity – No. (%)				
No	-	-	-	11 (73%)
Yes	-	-	-	4 (27%)
HCV genotype – No. (%)				
1	-	4 (40%)	9 (60%)	10 (66%)
2	-	5 (50%)	2 (13%)	-
3	-	1 (10%)	2 (13%)	-
Unknown	-	-	2 (13%)	5 (33%)
HBV exposure – No. (%)				
Yes	-	1 (10%)	6 (40%)	6 (40%)
No	10 (100%)	6 (60%)	7 (47%)	6 (40%)
Unknown	-	3 (30%)	2 (13%)	3 (20%)

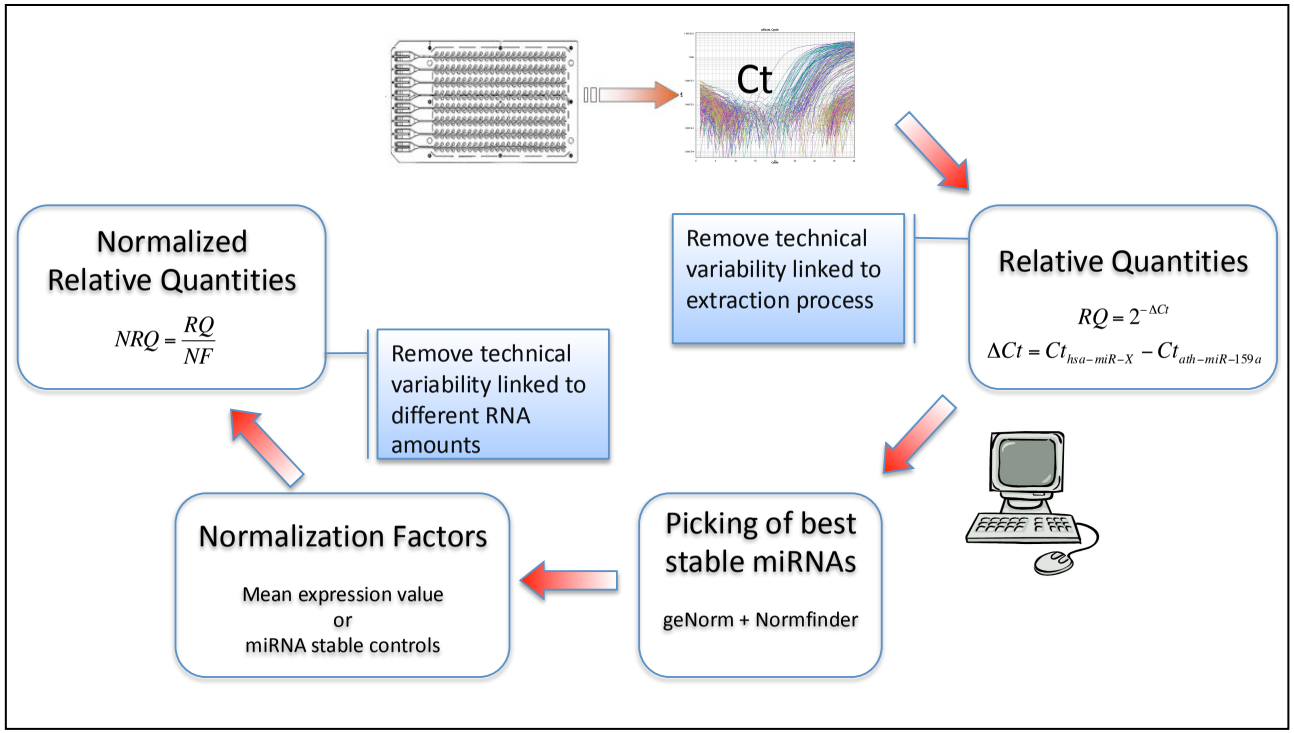


Figure 1. The workflow of the normalization process. Definition are given within the text

miRNA	*geNorm*	*Normfinder*	*CV*	*Distance*
hsa-miR-17	0.375	0.304	0.335	0.588
hsa-miR-484	0.375	0.321	0.368	0.616
hsa-miR-126	0.453	0.191	0.418	0.645
hsa-miR-146a	0.495	0.212	0.400	0.671
hsa-miR-20a	0.558	0.323	0.363	0.740

Table 2. The top five candidate reference miRNAs are shown ranked by the Distance column, which summarizes the three scoring algorithms. The lesser this value, the stablest the miRNA as normalization factor.

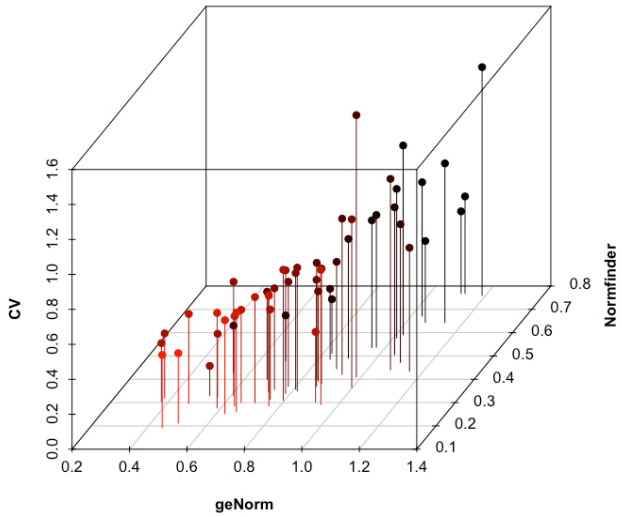


Figure 2. Three-dimensional scatter plot showing the scores resulting from the three algorithms for the selection of reference miRNAs. A good correlation between the methods is reported (multiple R^2=0.74)

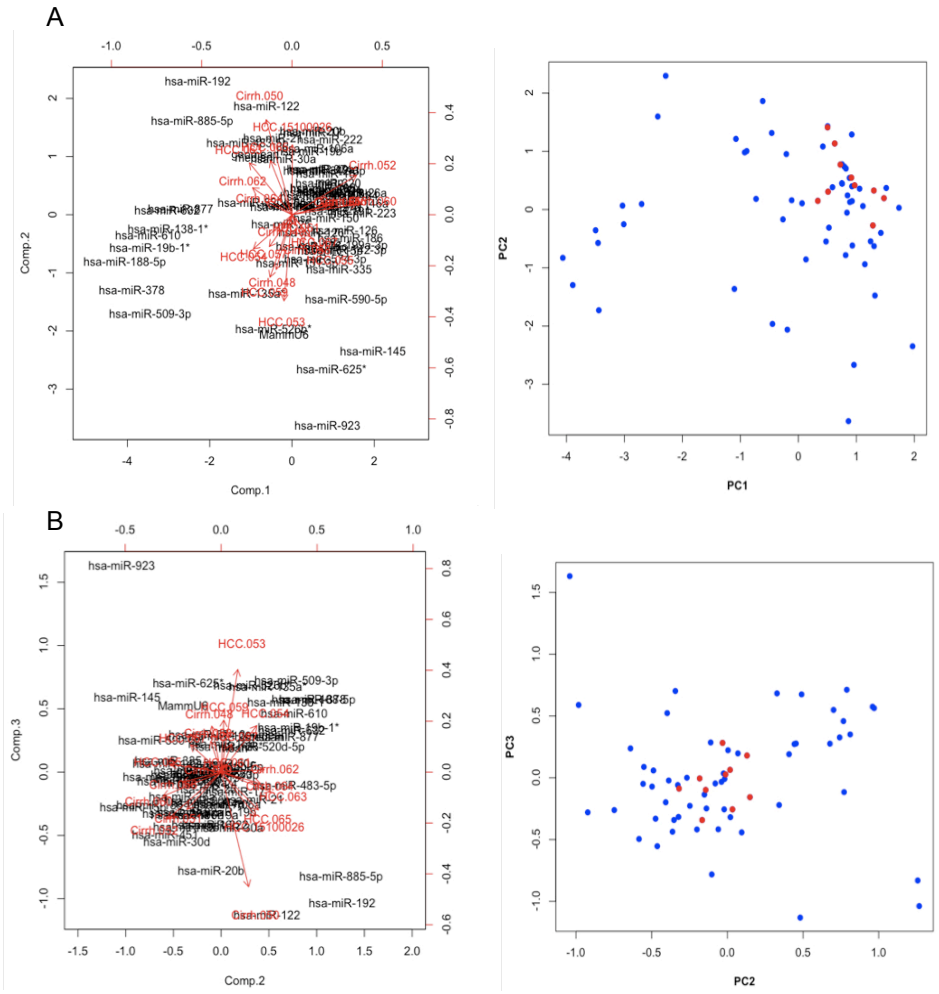


Figure 3. PCA reveals that candidate reference miRNAs are clustered together. The top panels (A) show a plot of the two first principal components of autoscaled data, while the bottom panels (B) show the second and the third principal components of unscaled data. On each plot on the left, component scores are shown by the corresponding miRNA name (bottom and left axis), while the loadings are shown by red arrows (top and right axis). The plots on the right depict the same data, but only miRNAs are shown. A red circle marks the first ten miRNAs ranked as the best endogenous controls by the combination of the three algorithms described in the text.

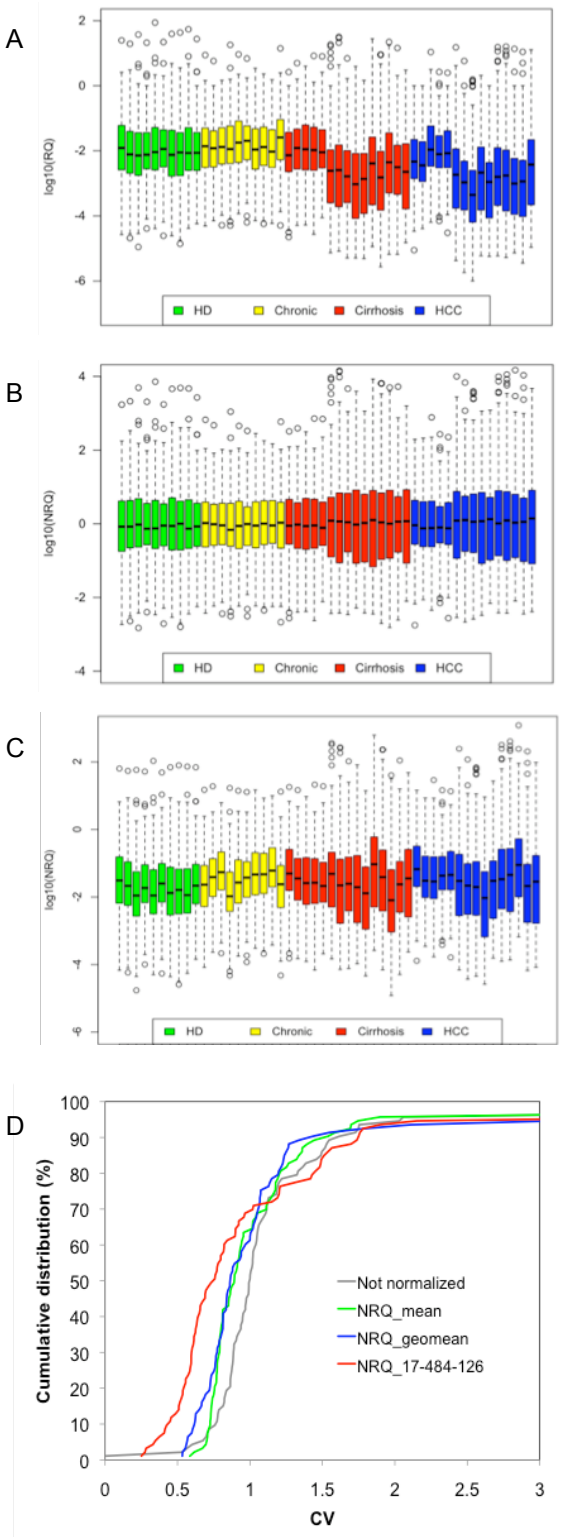


Figure 4. Normalized Relative Quantities (NRQs) were obtained using the global geometric or arithmetic mean of all expressed miRNAs per sample or alternatively using the geometric mean of the three top reference miRNAs (miR-17, miR-484, miR-126). Panels A, B and C show the distributions of miRNA levels in RQ and NRQ units. Data in B were normalized with the global geometric mean in B and with the reference miRNAs in C. The effective reduction in data variability was evaluated looking at the cumulative distribution of the coefficient of variation (CV) when data were not normalized (grey line) or normalized (yellow, blue and red lines).

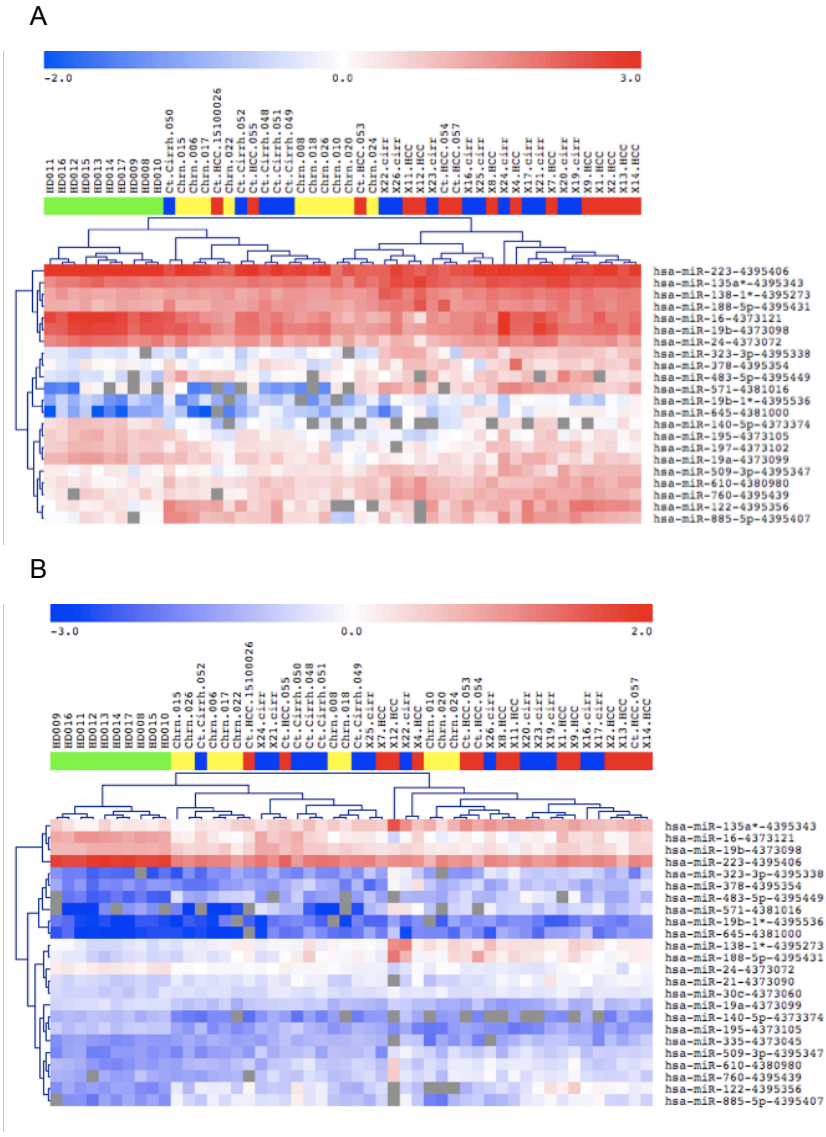


Figure 5. Differentially represented miRNAs after ANOVA on miRNAs detectable in at least two thirds of samples per group ($p<0.05$ after Bonferroni correction). Blue denotes low expression, white intermediate expression and red high expression, in logarithmic scale. All values are relative quantities normalized with the global sample geometric mean (A) or the geometric mean of mir-17, miR-126 and miR-484. Samples are coded as follows: green for healthy Donors (HD), yellow for chronic hepatitis (CH), blue for liver cirrhosis (LC) and red for hepatocellular carcinoma (HCC). Hierarchical clustering with Euclidean distance and complete linkage was used.

Table 3. miRNAs differentially represented among patients.

miRNA	*p-value*[*] *(global mean as reference)*	*p-value*[*] *(miRNAs as reference)*
hsa-miR-122	2.35E-03	6.38E-06
hsa-miR-135a*	1.77E-04	1.48E-03
hsa-miR-138-1*	1.59E-03	8.93E-03
hsa-miR-140-5p	3.35E-03	1.83E-03
hsa-miR-16	3.79E-05	1.18E-07
hsa-miR-188-5p	2.80E-07	8.10E-04
hsa-miR-195	4.55E-05	7.75E-06
hsa-miR-197	2.22E-02	n.s.[†]
hsa-miR-19a	1.48E-03	8.20E-04
hsa-miR-19b	7.02E-03	2.84E-05
hsa-miR-19b-1*	1.62E-02	7.17E-03
hsa-miR-21	n.s.[†]	1.03E-05
hsa-miR-223	2.68E-08	9.88E-15
hsa-miR-24	1.50E-02	7.33E-05
hsa-miR-30c	n.s.[†]	5.51E-05
hsa-miR-323-3p	9.25E-03	1.57E-03
hsa-miR-335	n.s.[†]	3.74E-02
hsa-miR-378	2.90E-03	5.01E-03
hsa-miR-483-5p	4.51E-04	1.36E-06
hsa-miR-509-3p	1.21E-02	1.28E-03
hsa-miR-571	6.56E-03	3.23E-03
hsa-miR-610	3.32E-08	1.37E-06
hsa-miR-645	1.19E-05	1.03E-05
hsa-miR-760	9.91E-06	1.47E-04
hsa-miR-885-5p	2.20E-02	6.49E-05

[*] after Bonferroni correction.

[†] n.s. = not significant ($p < 0.05$)

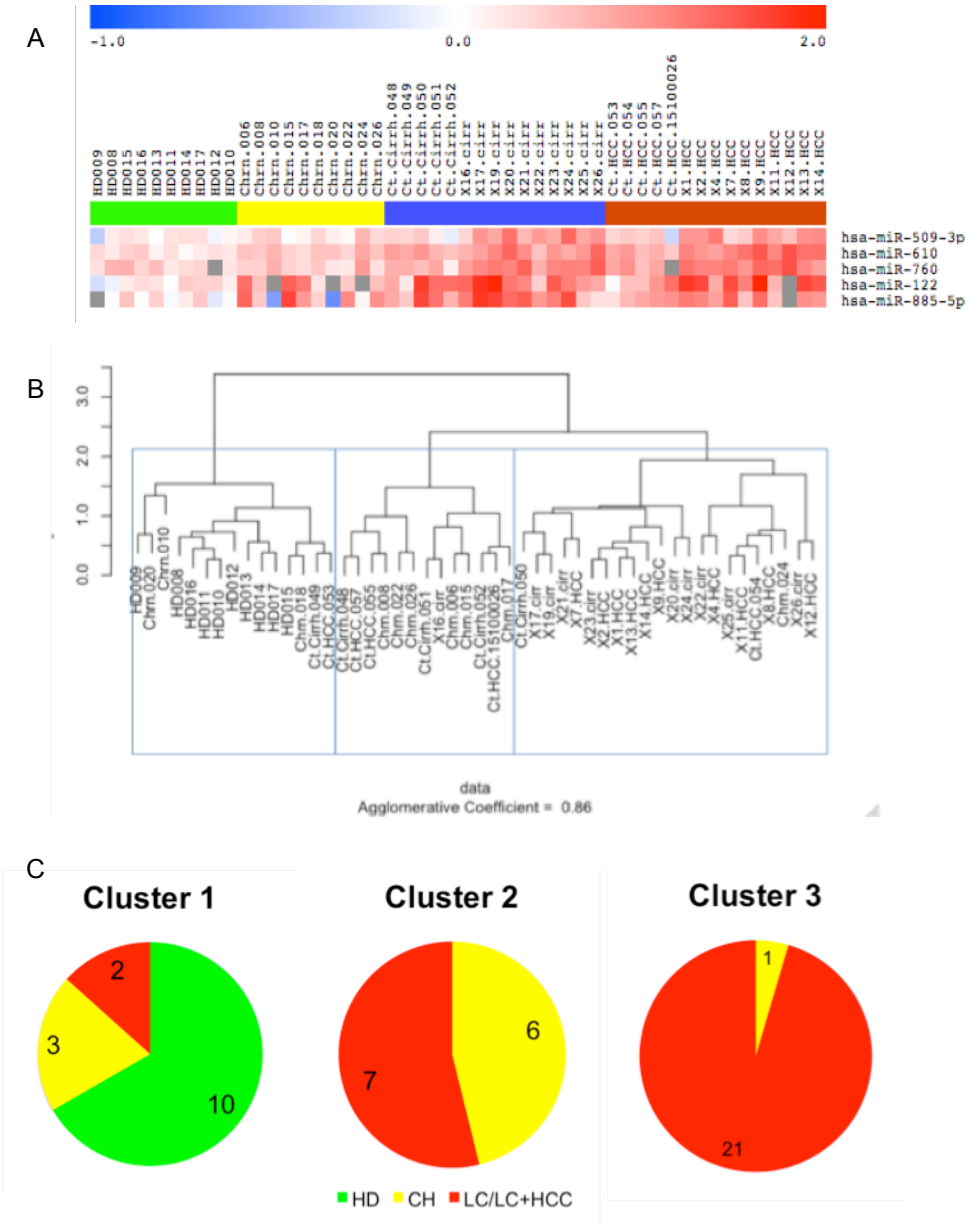


Figure 6. A five-miRNA signature clusters healthy donors and diseased subjects, mirroring disease progression. Panel A shows a heatmap of miR-509-3p, miR-610, miR-760, miR-122 and miR-885-5p, with the same color code indicated in **Figure 5**. In B, a hierarchical clustering is shown, obtained using Euclidean distance and complete linkage. In Panel C, each pie chart shows the portion of the subjects for each main cluster, stratified in three groups: healthy donors (HD, green), chronic hepatitis (CH, yellow) and advanced liver diseases (LC or HCC, red).

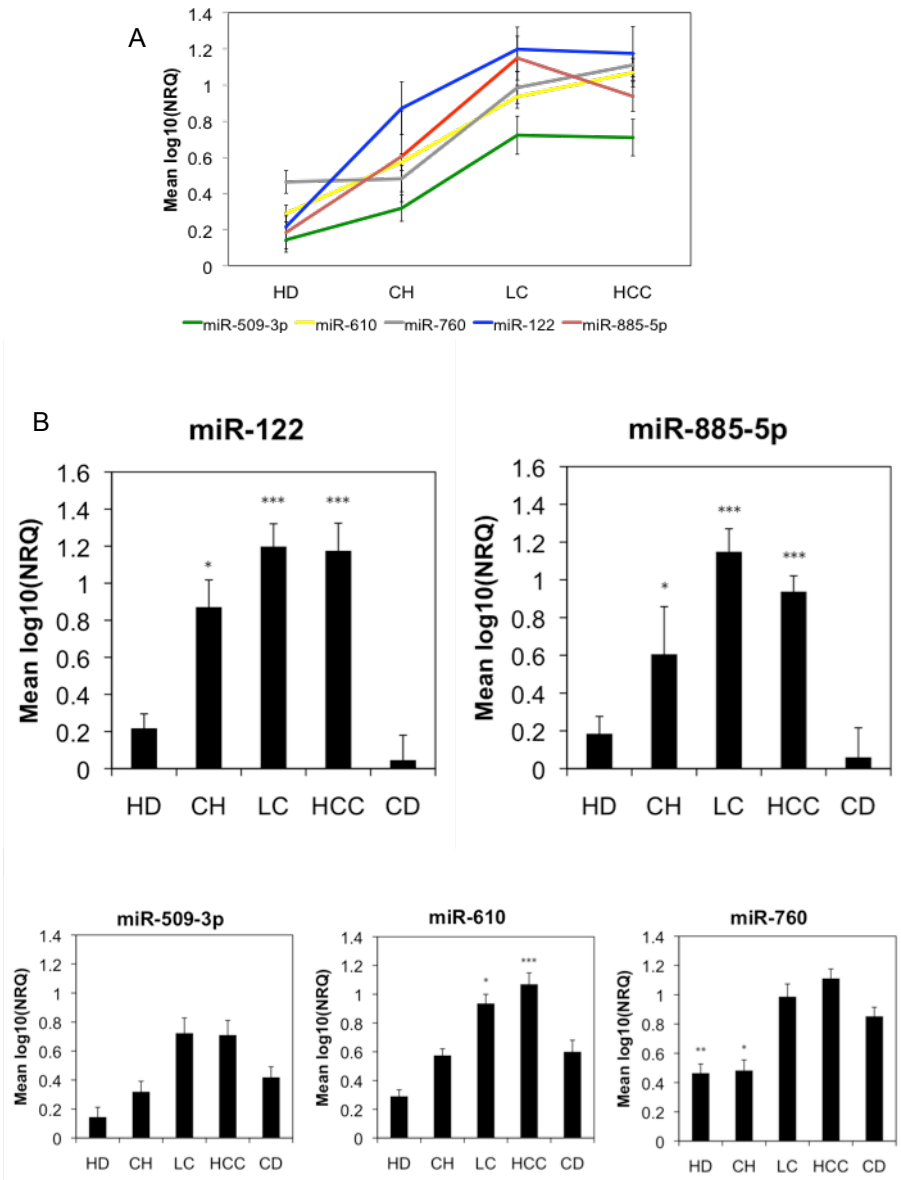


Figure 7. Circulating miRNAs are specifically and consistently increased in the serum of patients with chronic HCV infection, correlating to disease progression. As liver disease progresses to more advanced forms, the concentration of mirR-122, mir-885-5p, miR-509-3p, mir-610 and miR-760 increases in the serum (A). A closer look reveal that miR-122 and miR-885-5p are specific hallmark of chronic liver inflammation, compared to the control groups. Abbreviations: HD, healthy donors; CH, chronic hepatitis; LC, liver cirrhosis; HCC, hepatocellular carcinoma; CD, Chron's disease. Significant results are compared to CD and are indicated by: *, $p<0.05$; **, $p<0.01$, ***, $p<0.001$.

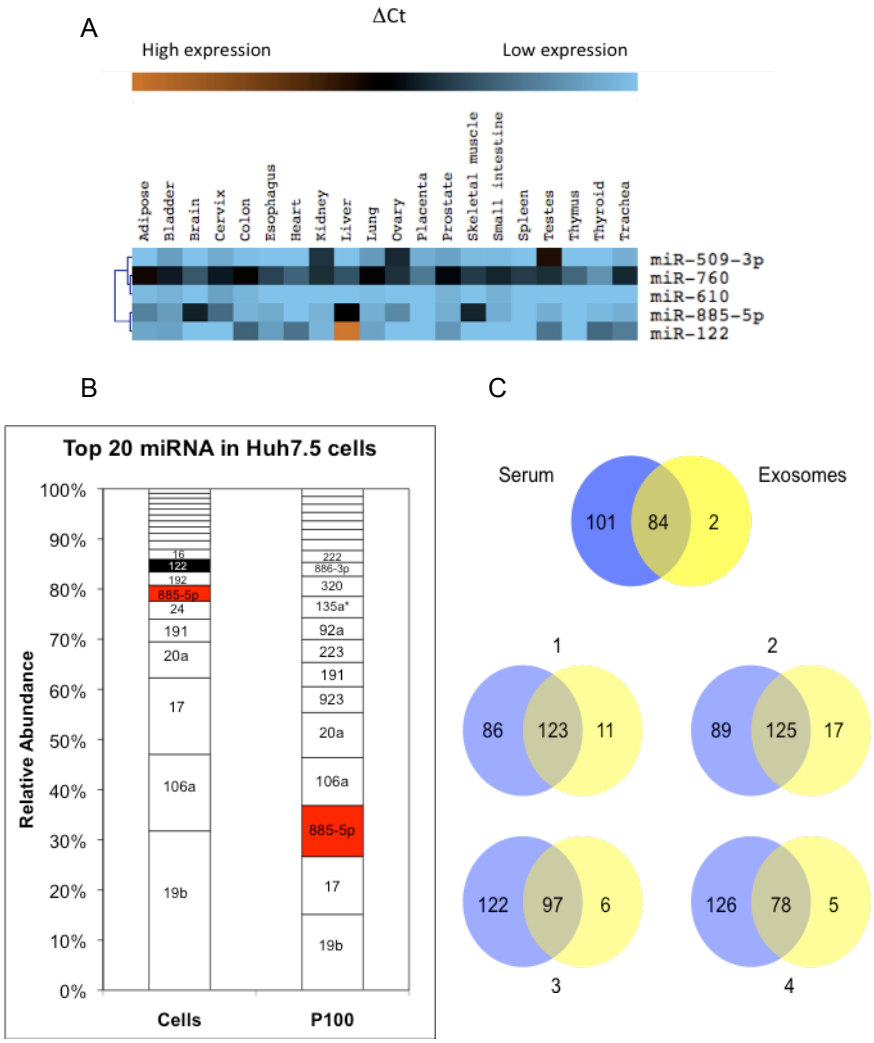


Figure 8. Tissue origin of selected miRNAs and expression in liver cells and exosomes. In panel A, the expression of the indicated miRNAs in adult normal tissues is shown on a ΔCt scale, after normalization with U6 small RNA. Panel B shows the relative abundance of the top miRNA in HuH7.5 cells and in the corresponding exosome preparation. Relative quantities were calculated with respect to the synthetic spike-in control (ath-miR-159a). In panel C, the Venn diagrams show the number of detected miRNAs (Ct<35) in whole serum vs. the corresponding exosome preparations, in four healthy controls. In summary, 185 miRNAs were detected in at least three out of for samples in the total serum, while 86 miRNAs were detected in at least three out of four corresponding exosome preparations.

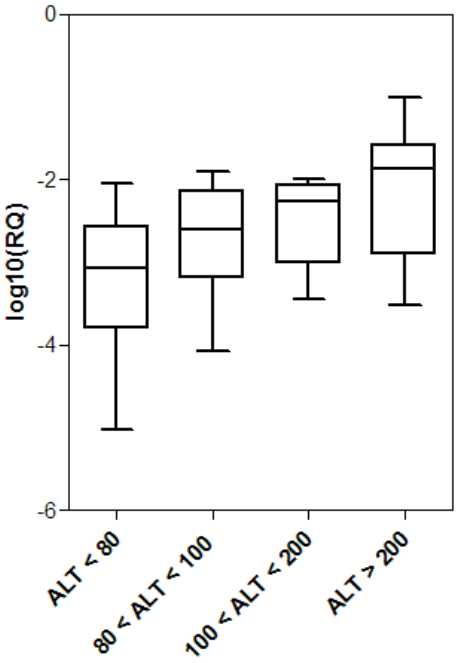


Figure 9. miR-122 levels correlated with alanine aminotransferase (ALT) activity. miR-122 Relative Quantities (RQs) were calculated by single TaqMan qPCR assays, using the ath-miR-159a as reference. miR-122 concentration changed significantly among groups (p=0.0003, Kruskal-Wallis test), showing a positive trend with increasing ALT activity, given in U/l.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281-97.
2. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215-33.
3. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010;466:835-40.
4. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature 2005;435:834-8.
5. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA 2006;103:2257-61.
6. O'connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. Nature Reviews Immunology 2010;10:111-22.
7. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science 2005;309:1577-81.
8. Reid G, Kirschner MB, van Zandwijk N. Circulating microRNAs: Association with disease and potential use as biomarkers. Critical reviews in oncology/hematology 2010:In press.
9. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, Zhang Y, Wang J, Jiang X, Xiang Y, Xu C, Zheng P, Zhang J, Li R, Zhang H, Shang X, Gong T, Ning G, Wang J, Zen K, Zhang J, Zhang C-Y. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997-1006.
10. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA 2008;105:10513-8.
11. Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, Chajut A. Serum microRNAs are promising novel biomarkers. PLoS ONE 2008;3:e3148.
12. Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 2007;8:166.

13. Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med 2001;345:41-52.
14. Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr
15. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 2001;358:958-65.
16. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Jr., Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002;347:975-82.
17. Kim WR, Flamm SL, Di Bisceglie AM, Bodenheimer HC. Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease. Hepatology 2008;47:1363-70.
18. Nathwani RA, Pais S, Reynolds TB, Kaplowitz N. Serum alanine aminotransferase in skeletal muscle diseases. Hepatology 2005;41:380-2.
19. Korones DN, Brown MR, Palis J. "Liver function tests" are not always tests of liver function. Am J Hematol 2001;66:46-8.
20. Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, Christensen E, Pagliaro L, Colombo M, Rodes J. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol 2001;35:421-30.
21. Daniele B, Bencivenga A, Megna AS, Tinessa V. Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma. Gastroenterology 2004;127:S108-12.
22. Gebo KA, Chander G, Jenckes MW, Ghanem KG, Herlong HF, Torbenson MS, El-Kamary SS, Bass EB. Screening tests for hepatocellular carcinoma in patients with chronic hepatitis C: a systematic review. Hepatology 2002;36:S84-92.
23. Kutay H, Bai S, Datta J, Motiwala T, Pogribny I, Frankel W, Jacob ST, Ghoshal K. Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. J Cell Biochem 2006;99:671-8.
24. Gramantieri L, Ferracin M, Fornari F, Veronese A, Sabbioni S, Liu C-G, Calin GA, Giovannini C, Ferrazzi E, Grazi GL, Croce CM, Bolondi L, Negrini M. Cyclin G1 is a target of miR-122a, a microRNA frequently down-regulated in human hepatocellular carcinoma. Cancer Research 2007;67:6092-9.
25. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. Gastroenterology 2007;133:647-58.
26. Varnholt H, Drebber U, Schulze F, Wedemeyer I, Schirmacher P, Dienes H-P, Odenthal M. MicroRNA gene expression profile of hepatitis C

virus-associated hepatocellular carcinoma. Hepatology 2008;47:1223-32.
27. Ura S, Honda M, Yamashita T, Ueda T, Takatori H, Nishino R, Sunakozaka H, Sakai Y, Horimoto K, Kaneko S. Differential microRNA expression between hepatitis B and hepatitis C leading disease progression to hepatocellular carcinoma. Hepatology 2009;49:1098-112.
28. Chen X-M. MicroRNA signatures in liver diseases. World J Gastroenterol 2009;15:1665-72.
29. Lewis AP, Jopling CL. Regulation and biological function of the liver-specific miR-122. Biochem. Soc. Trans 2010;38:1553-7.
30. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
31. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004;64:5245-50.
32. Kohl, M., (2007). SLqPCR: Functions for analysis of real-time quantitative PCR data at SIRS-Lab GmbH. R package, SIRS-Lab GmbH. Jena.
33. Théry C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol 2006;Chapter 3:Unit 3.22.
34. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol 2009;10:R64.
35. J. Vandesompele, M. Kubista and M. W. Pfaffl. Reference Gene Validation Software for Improved Normalization. Chapter 4 in Real-Time PCR: Current Technology and Applications, Publisher: Caister Academic Press. ISBN: 978-1-904455-39-4 (2009)
36. Randall G, Panis M, Cooper JD, Tellinghuisen TL, Sukhodolets KE, Pfeffer S, Landthaler M, Landgraf P, Kan S, Lindenbach BD, Chien M, Weir DB, Russo JJ, Ju J, Brownstein MJ, Sheridan R, Sander C, Zavolan M, Tuschl T, Rice CM. Cellular cofactors affecting hepatitis C virus infection and replication. Proc Natl Acad Sci USA 2007;104:12884-9.
37. Blight KJ, McKeating JA, Rice CM. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. J Virol 2002;76:13001-14.
38. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, Banham AH, Pezzella F, Boultwood J, Wainscoat JS, Hatton CSR, Harris AL. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol 2008;141:672-5.
39. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, Shen H. Serum MicroRNA Signatures Identified in a Genome-Wide

Serum MicroRNA Expression Profiling Predict Survival of Non-Small-Cell Lung Cancer. Journal of Clinical Oncology 2010.
40. Ng EKO, Chong WWS, Jin H, Lam EKY, Shin VY, Yu J, Poon TCW, Ng SSM, Sung JJY. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut 2009;58:1375-81.
41. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. International journal of cancer Journal international du cancer 2009.
42. Tsujiura M, Ichikawa D, Komatsu S, Shiozaki A, Takeshita H, Kosuga T, Konishi H, Morimura R, Deguchi K, Fujiwara H, Okamoto K, Otsuji E. Circulating microRNAs in plasma of patients with gastric cancers. Br J Cancer 2010;102:1174-9.
43. Resnick KE, Alder H, Hagan JP, Richardson DL, Croce CM, Cohn DE. The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. Gynecologic Oncology 2009;112:55-9.
44. Ho AS, Huang X, Cao H, Christman-Skieller C, Bennewith K, Le QT, Koong AC. Circulating miR-210 as a Novel Hypoxia Marker in Pancreatic Cancer. Transl Oncol 2010;3:109-13.
45. Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin MJ. Circulating microRNAs as Novel Minimally Invasive Biomarkers for Breast Cancer. Annals of Surgery 2010;251:499-505.
46. Miyachi M, Tsuchiya K, Yoshida H, Yagyu S, Kikuchi K, Misawa A, Iehara T, Hosoi H. Circulating muscle-specific microRNA, miR-206, as a potential diagnostic marker for rhabdomyosarcoma. Biochemical and Biophysical Research Communications 2010.
47. Dimmeler S, Zeiher AM. Circulating microRNAs: novel biomarkers for cardiovascular diseases? European Heart Journal 2010.
48. Gui J, Tian Y, Wen X, Zhang W, Zhang P, Gao J, Run W, Tian L, Jia X, Gao Y. Serum microRNA characterization identifies miR-885-5p as a potential marker for detecting liver pathologies. Clin. Sci. 2011;120:183-93.
49. Laterza OF, Lim L, Garrett-Engele PW, Vlasakova K, Muniappa N, Tanaka WK, Johnson JM, Sina JF, Fare TL, Sistare FD, Glaab WE. Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clinical Chemistry 2009;55:1977-83.
50. Qu KZ, Zhang K, Li H, Afdhal NH, Albitar M. Circulating MicroRNAs as Biomarkers for Hepatocellular Carcinoma. J Clin Gastroenterol 2011.
51. Xu J, Wu C, Che X, Wang L, Yu D, Zhang T, Huang L, Li H, Tan W, Wang C, Lin D. Circulating MicroRNAs, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. Mol. Carcinog. 2010.
52. Zhang Y, Jia Y, Zheng R, Guo Y, Wang Y, Guo H, Fei M, Sun S. Plasma MicroRNA-122 as a Biomarker for Viral-, Alcohol-, and Chemical-Related Hepatic Diseases. Clinical Chemistry 2010.
53. Kroh EM, Parkin RK, Mitchell PS, Tewari M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse

transcription-PCR (qRT-PCR). Methods (San Diego, Calif) 2010;50:298-301.
54. Mendell JT. miRiad roles for the miR-17-92 cluster in development and disease. Cell 2008;133:217-22.
55. Fu H, Tie Y, Xu C, Zhang Z, Zhu J, Shi Y, Jiang H, Sun Z, Zheng X. Identification of human fetal liver miRNAs by a novel method. FEBS Lett 2005;579:3849-54.
56. Bimpaki EI, Iliopoulos D, Moraitis A, Stratakis CA. MicroRNA signature in massive macronodular adrenocortical disease and implications for adrenocortical tumourigenesis. Clinical Endocrinology 2010;72:744-51.
57. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foà R, Schliwka J, Fuchs U, Novosel A, Müller R-U, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter H-I, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007;129:1401-14.
58. Urbich C, Kuehbacher A, Dimmeler S. Role of microRNAs in vascular diseases, inflammation, and angiogenesis. Cardiovasc Res 2008;79:581-8.
59. Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, Xiao T, Schafer J, Lee M-LT, Schmittgen TD, Nana-Sinkam SP, Jarjoura D, Marsh CB. Detection of microRNA expression in human peripheral blood microvesicles. PLoS ONE 2008;3:e3694.
60. D'haeseleer P. How does gene expression clustering work? Nat Biotechnol 2005;23:1499-501.
61. Wang K, Zhang S, Marzolf B, Troisch P, Brightman A, Hu Z, Hood LE, Galas DJ. Circulating microRNAs, potential biomarkers for drug-induced liver injury. Proc Natl Acad Sci USA 2009;106:4402-7.
62. Abraham C, Cho JH. Inflammatory bowel disease. N Engl J Med 2009;361:2066-78.
63. Fornari F, Gramantieri L, Giovannini C, Veronese A, Ferracin M, Sabbioni S, Calin GA, Grazi GL, Croce CM, Tavolari S, Chieco P, Negrini M, Bolondi L. MiR-122/cyclin G1 interaction modulates p53 activity and affects doxorubicin sensitivity of human hepatocarcinoma cells. Cancer Research 2009;69:5761-7.
64. Coulouarn C, Factor VM, Andersen JB, Durkin ME, Thorgeirsson SS. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. Oncogene 2009;28:3526-36.
65. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol 2007;9:654-9.

66. Gibbings DJ, Ciaudo C, Erhardt M, Voinnet O. Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. Nat Cell Biol 2009;11:1143-9.
67. Gibbings D, Voinnet O. Control of RNA silencing and localization by endolysosomes. Trends Cell Biol 2010;20:491-501.

Chapter 4. Summary, conclusions and future work

Hepatitis C Virus (HCV) infection is a major cause of chronic hepatitis, liver cirrhosis and HCC. Worldwide, an estimated 170 million people are chronically infected with HCV, translating into 3% of the population.[1,2] A principal attribute of chronic HCV infection is the characteristic progression from almost asymptomatic forms to severe liver disease in about 20% of the persistently infected subjects.[3] Indeed, HCV infection is currently considered one of the main causes of chronic liver disease, including cirrhosis and hepatocellular carcinoma (HCC) in Western countries. As a consequence, HCV infection is the most common indication for liver transplantation, even though viral recurrence is rapid and universal. Destruction of hepatocytes by the chronic inflammation is up to a certain point balanced by liver regeneration. For mechanisms not completely understood, in a considerable fraction of cases, liver destruction is followed by scar formation and deposition of fibrotic tissue. Fibrosis progresses with variable speed toward a massive alteration of the liver structure represented by liver cirrhosis. This is the real problem to avoid in chronic HCV infection, because cirrhosis is the door to two major complications such as liver failure and HCC.

Our ultimate aim was to identify significant predictors of disease progression in chronic HCV infection and elucidate their potential use as biomarkers for a non-invasive prognosis and/or monitoring of the natural history of HCV infection. To reach this goal, we designed an integrated approach based on a combination of state-of-the-art genomic technologies including:

a) a Genome-Wide Association Study (GWAS) performed on a well characterized cohort of patients with an accurate estimate of the date of infection, aimed at the identification of the genetic determinants of fibrosis progression
b) a high-throughput profiling of the microRNAs (miRNAs) circulating in the blood of patients with chronic HCV infection at different stages of the disease, including liver cirrhosis and HCC.

These two studies had intrinsically a translational perspective, given their connection to the clinical practice. Indeed, they meet the valued needs of both allowing a better understanding of the disease pathogenesis and guiding an

improved patient selection process for eligibility to antiviral therapy, identifying those patients with higher risk of developing hepatic complications and are therefore more likely to benefit from early treatment. Our studies involved a close collaboration with the physicians directly involved in the diagnosis and treatment of the patients. We successfully established a fruitful teamwork resulting in the synergy between experts with competences in different subject areas, as required by modern, well-designed translational studies.

The GWAS was a hypothesis-free approach based on the parallel genotyping of more than 500k SNPs on genomic DNA of patients with chronic HCV infection passing strict inclusion criteria. Most importantly, only patients with a known or accurately estimated date of infection were included in this study. The correct identification of the time of infection is a critical point in determining the role of any predictor of disease progression in an acquired disease such as chronic HCV. Our study, from this point of view, is particularly solid not only because in most of the patients the infection was acquired during a datable event such as multi blood transfusions or intravenous-drug abuse, but also because the magnitude of the effects was confirmed when we analyzed only those patients whose date of infection was certain, as they had either a positive anti-HCV test or an abrupt rise in serum transaminase levels following a known risk factor for HCV. Nevertheless, none of the tested SNPs was associated definitively with increased fibrosis progression, as we obtained only suggestive evidence ($10^{-7}<p<10^{-4}$). On the contrary, we detected an outstanding role of the older age at the infection as a predictor of worsening fibrosis. Moreover, male gender and HCV genotype, were significantly but less strongly associated to liver disease progression, as previously reported.[4, 5] Therefore, we strongly believe that from the clinical point of view our study provides important insights into the natural history of HCV infection and specifically into fibrosis progression and their relationship with host and external factors. While we absolutely characterized the role on non-genetic factors, we could not however dissect the heritability components underlying the differential disease progression rates observed during the progression to cirrhosis. We are aware

that a major limitation is represented by the limited sample size of this study. This feature resulted mainly from the strict inclusion criteria that we established for the selection of patients, limiting the number of available subjects in a monocentric planning at one side, but at the same time leading to more accurate estimates of the effect size of predictors of disease severity.

The second part of this framework involved the identification of circulating miRNAs that were specifically detected in sera of patients with chronic HCV infection and which correlated to disease progression. We successfully identified miR-122 and miR-885-5p as potential biomarkers of chronic HCV infection, being specifically and consistently increased in the sera of patients with chronic hepatitis, liver cirrhosis and HCC. We were particularly interested in achieving solid results, supported by proper normalization strategies to account for the variability introduced by the experimental process and successfully eliminate it. For this reason, we performed a scrupulous analysis of the variability in the expression data and effectively obtained a list of valuable reference miRNAs that reduced technical and experimental-induced differences. Next, we identified *ex vivo* disease-specific serum miRNA signatures associated specifically with different stages of liver disease in HCV-infected patients. Very importantly, the methods described here required a low amount of starting material (70 μl of serum), allowing a retrospective study design. This first phase, or "discovery" phase, will be followed by a second-phase validation study, when we will assess the strength of the reported associations, the sensitivity and specificity, the effect size and the usefulness of circulating miRNAs as biomarker of liver diseases. Moreover, a large-scale *prospective* study would be extremely advantageous to investigate the role of circulating miRNAs in the progression from liver cirrhosis to HCC. In future work, we anticipate the characterization of the physiological role of significant miRNAs in liver biology, by looking at their cellular targets and regulatory networks. The identification of candidate target genes will be based on a twofold approach that will combine the *in silico* prediction of miRNA targets and the information of the actually transcribed mRNAs in the cells of interest. In this regard, we will generate reference maps of cellular miRNA and mRNA

expression from cell-culture models of hepatic diseases including cultured primary and transformed hepatocytes infected with HCV or hepatic stellate cells (HSCs) undergoing in vitro activation. Actually, HSCs represent one of the main players during hepatic fibrogenesis, given their ability to produce extracellular matrix components. Because we are most interested in the use of miRNAs as potential serum-based biomarkers, the miRNA expression profiles will be studied both in the cells and in the exosomes released extracellularly. We are particularly fascinated by the mechanism responsible for the accumulation of miRNAs within the exosomes. It is really worth understanding if the exosomal miRNA content merely reflects the intracellular miRNA expression or an active and controlled release could be responsible of specific exosomal miRNA signatures. Actually, the functional export of miRNAs within exosomes might represent either a communication system between cells or a mechanism to "regulate the regulators", by preventing miRNA action through their exclusion from the cytoplasm.

References

1. World Health Organization. Hepatitis C Fact Sheet No. 164 (Revised Oct 2000).
2. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002--June 10-12, 2002. Hepatology 2002;36:S3-20.
3. Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med 2001;345:41-52.
4. Seeff LB. Natural history of chronic hepatitis C. Hepatology 2002;36:S35-46.
5. Missiha SB, Ostrowski M, Heathcote EJ. Disease progression in chronic hepatitis C: modifiable and nonmodifiable factors. Gastroenterology 2008;134:1699-714.

Concluding words

Esprimermi con il cuore non è mai stato il mio forte. Per questo mi perdonerete se l'ultima parte di questo lavoro la scrivo in italiano: non avrei potuto fare altrimenti. Nel corso di questi ultimi anni trascorsi dopo la laurea specialistica, una frase mi è stata particolarmente cara. Bertrand Russel (filosofo, logico e matematico inglese) scrisse: "*Science may set limits to knowledge, but should not set limits to imagination*",[*] sottolineando il profondo bisogno dell'uomo per la passione, l'arte e la religione nelle antiche civiltà. Credo fortemente che questa spinta verso attività al di fuori della stretta codifica empirica della scienza sia insita in ognuno di noi e si sia conservata intatta nella storia. Ad un osservatore esterno, potrebbe sfuggire la componente umana nell'attività di chi si occupa di scienza, immaginando il ricercatore immerso nei propri esperimenti e alieno alle pulsioni e ai tormenti. Ma l'attività del *fare* scienza comporta in sé un'intrecciarsi di relazioni umane che inevitabilmente partecipano al processo di raggiungimento della conoscenza, traducendo una attività rigorosamente oggettiva ed empirica, quale la conoscenza scientifica, in un'impresa con connotazioni di carattere emotivo, sia piacevolmente che dolorosamente. Allora non bisogna dimenticare che dietro esperimenti e risultati ci sono storie, vite ed eventi.

Dal mio personale punto di vista, dovendo fare un bilancio di questi anni di dottorato, credo che si sia sempre di più accresciuta in me la consapevolezza che uno tra gli elementi che rendono considerevole un percorso sia la passione che accompagna il processo stesso della ricerca. Ma intendo la ricerca in senso lato, come personale attività di miglioramento del proprio io, di arricchimento. Sono la forza che ti spinge al cambiamento, la meraviglia di ricercare e la propensione al divenire che hanno in fondo reso interessante il tracciato che mi ha portato fin qui oggi. Ammetto che i momenti di sfiducia sono stati parecchi, e spesso mi sono trovato altalenante tra l'abisso e l'esaltazione. Per cui il mio più

[*] Bertrand Russel. A History of Western Philosophy. 1945, Simon & Schuster, New York

grande e profondo ringraziamento va alla mia famiglia: a Raffaella, ai miei genitori Maria ed Enzo e alla mia sorellina Claudia che mi hanno aiutato a costruire radici e ali della mia persona, che mi hanno sostenuto, incoraggiato e a volte anche sopportato.

Non posso dimenticare poi tutti i miei colleghi con i quali ho condiviso tanto, persino la (poca) aria da respirare in quell'accogliente seminterrato dalle pareti blu/violetto dove quotidianamente ci si incontra. Non è fantastico?

In particolare devo ringraziare:

- Riccardo e Raoul che mi hanno iniziato all'analisi dei dati: avete creato un mostro…
- Raffaele e Massimiliano: lo yin e lo yang dei nostri laboratori. Grazie per il supporto e l'aver creduto in me.
- Sergio, da cui ho imparato la tenacia e la passione per il proprio lavoro. Grazie per la smisurata fiducia che mi hai dimostrato.
- Mauro: ne abbiamo passate davvero tante, durante questi anni, ne cito solo una: incidente!
- Paola: come avrei fatto in questi ultimi mesi senza di te in laboratorio? Abbiamo condiviso ansia e allegria. Grazie anche per la preziosa revisione della tesi.