UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di Dottorato in Scienze Farmacologiche

Curriculum in Farmacologia Molecolare e Cellulare

XXXII Ciclo

Tesi di Dottorato

Development of diagnostic assays for Hepatitis C Virus genotyping and identification of drug-resistance mutations

Tema vincolato: borsa Biofield Innovation S.r.l. - Tema: Ricerca e diagnostica in vitro connessa al trattamento farmacologico, sviluppo e validazione di tecniche analitiche e di dispositivi medico diagnostici in vitro (CE-IVD).

Supervisore esterno
Dott. Dino Paladin

Coordinatore
Prof. Nicola Ferri

Dottoranda
Serena Tiozzo

Anno Accademico 2018/2019

Ai miei genitori

Summary

1. ABSTRACT

Hepatitis C Virus (HCV) is one of the most widespread viruses in the world. World Health Organisation (WHO) planned to eliminate the disease in 2030. In order to do so, many efforts have to be made, especially in the countries where HCV is still endemic.

HCV is an extremely variable virus and it is divided into genetic groups called genotypes. In 2011 Direct Antiviral Agents (DAAs) have been introduced for the HCV therapy. In comparison with the previous therapy, these drugs have a higher Sustained Virological Response (SVR). However, they are genotype-dependent, and the identification of the viral genotype was considered the priority for the choice of a correct therapeutic regimen. Since genotyping analysis was considered mandatory during the care pathway, many *in vitro* diagnostic assays were developed by different companies to fill this clinical need. Biofield Innovation decided to redesign an HCV genotyping assay based on Reverse Line Blot technology to improve the performance of it. In this study, development and validation process of the new assay are discussed.

In 2016 pangenotypic antiviral drugs were commercialized. Their introduction makes the choice of the therapy regimen less and less influenced by the viral genotype. However, even with pangenotypic drugs, a low percentage of therapeutic failure is still registered. One of the main reasons of the treatment failure is the presence of Resistance Associated Substitutions (RASs) that are point mutations that confer drug-resistance to the virus. To avoid the viral relapse and to prevent the therapeutic failure, the choose of the most suitable treatment for the therapy course must involve the consideration of the presence of RASs.

Nowadays, the most powerfull technology that can be used to investigate the presence of RASs is Next Generation Sequencing (NGS). In this study the development of an in vitro diagnostic device for the identification of the genotype and the presence of RASs based on the NGS is described. In particular, prototype pre-validation and validation processes results of the prototype will be reported and analysed.

2. INTRODUCTION

2.1 Hepatitis C disease and its transmission

Hepatitis C is a liver disease caused by the *Hepatitis C virus* (HCV). The transmission of this infection occurs through the exposure to infected blood. World Health Organization (WHO) estimated that 71 million people in the world are infected by HCV. In 2017 there were about 2,85 million new cases worldwide. In 2016 WHO estimated that 399.000 people died because of the complication of the infection such as liver cirrhosis and hepatocellular carcinoma.

HCV could be found in all places of human habitation. The highest prevalence of HCV is reported to be in WHO Eastern Mediterranean Region (62,5 per 100.000) and the European Region (61,8 per 100.000) (Fig.1).

		Incidence of HCV infection			
		Incidence rate (per 100 000)		Total number (000)	
WHO region	Map key	Best estimate	Uncertainty interval	Best estimate	Uncertainty interval
African Region		31.0	22.5–54.4	309	222–544
Region of the Americas		6.4	5.9–7.0	63	59–69
Eastern Mediterranean Region		62.5	55.6–65.2	409	363–426
European Region		61.8	50.3–66.0	565	460–603
South-East Asia Region		14.8	12.5–26.9	287	243–524
Western Pacific Region		6.0	5.6–6.6	111	104–124
Global		23.7	21.3–28.7	1 751	1 572–2 120

Source: WHO, work conducted by the Center for Disease Analysis. See Annex 2.
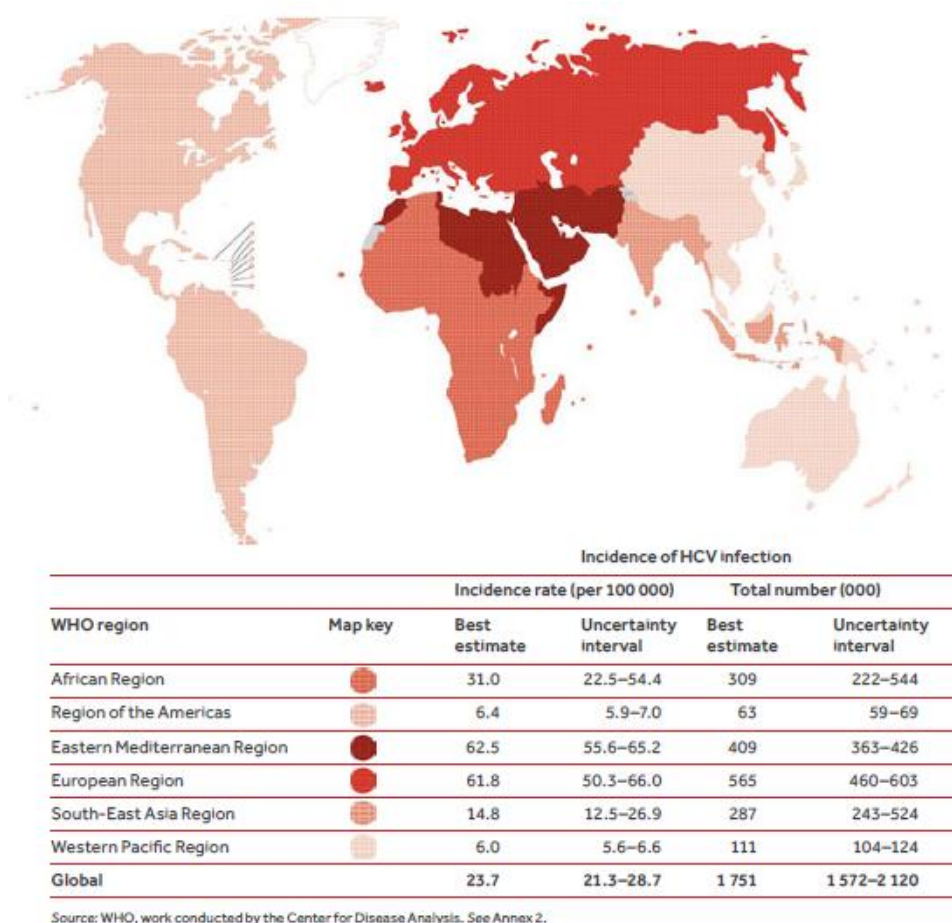
Fig. 1 Incidence of HCV infection in the general population, by WHO region, 2015 ("GLOBAL HEPATITIS REPORT,2017 WHO," n.d.)

Moreover, among the 3.7 million persons living with Human Immunodeficiency Virus (HIV), an estimated 2.3 million have serological evidence of past or present HCV infection (www.who.int/news-room/fact-sheets/detail/hepatitis-c, 20th July 2019). In the United States after a decrease, the incidence of HCV infection doubled between 2010 and 2014 (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019).

HCV is one of the viruses with the higher prevalence in the world, this is the reason why it is considered clinically important (Chigbu et al. 2019).

The most frequent ways of infection are: unsafe injection practices among Injection Drug Users (IDU) subjects, unsafe health care, the transfusion of unscreened blood and blood products, and promiscuous sexual behaviour (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019).

It seems that even vertical transmission is possible in three distinct moments of the pregnancy: intrauterine, peripartum and postpartum periods. HCV was found in new-born 28 days after birth and that demonstrates that the direct exposure to maternal blood is the easiest way for vertical transmission. Even amniocentesis exam or fetal monitoring could expose the fetus to the maternal blood. However, even if HCV was found in human colostrum, breast feeding seems not to be a way of vertical transmission (Prasad and Honegger 2013).

Moreover, in the 10% of cases, patients have not identifiable cause of infection. In these cases, the ways of infection could be disparate such as intranasal cocaine use, tattooing or body piercing and sharing of items that may have been contaminated with infected blood such as razors or toothbrushes. Beyond the cause, HCV ways of infection are all associated with a not adequate sterilization practice (https://www.hepatitis.va.gov/hcv/background/transmission-modes.asp, 26th May 2019).

Once inside the host, the virus begins to replicate, and the infection starts as an acute form. In the 80% of cases it's asymptomatic but patient with symptomatic acute phase could exhibit fever, fatigue, decreased appetite, nausea, vomiting, abdominal pain, dark urine, grey-coloured faeces, joint pain and jaundice (yellowing of skin and the white of the eyes) (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019).

Rarely, HCV infection is associated to extrahepatic disease such as secondary Sjögren syndrome, lichen planus, diabetes mellitus and other

lymphoproliferative disorders (Zignego et al. 2007).

In most of the cases HCV infection is asymptomatic, if it is not correctly diagnosticated and cured, it could become chronic and then it can develop in cirrhosis and Hepatocellular Carcinoma (HCC) (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019). Usually HCV-related hepatitis stays asymptomatic in the first ten years after the infection. It was observed that cirrhosis could be manifested even only after 30 years in 10-30% of cases, especially for people infected also with HBV, HIV or alcoholics. People with cirrhosis have been shown to increase the risk for developing HCC of 20-fold (Rosen 2011).

HCV is responsible of the world's 27% hepatic cirrhosis and of the 25% HCC (Mohamed et al. 2015).

In 2016, WHO published the first global health sector strategy on viral hepatitis 2016-2021 ("GLOBAL HEALTH SECTOR STRATEGY ON VIRAL HEPATITIS 2016-2021," n.d.). This plan has the purpose of eliminating viral hepatitis as a major public health threat by 2030. This would be possible through different approaches described in the five parts of the plan, such as: information, high-impact interventions in the local's public health, promoting equity in the cure, minimizing the costs, promoting the progress in scientific research.

2.2 Virus phylogenesis and structure

Hepatitis C virus was discovered in 1975 when most cases of transfusion-associated hepatitis were found to be not associated with Hepatitis A Virus (HAV) or Hepatitis B Virus (HBV) infections. This disease was defined as Non-A, Non-B Hepatitis (NANBH).

Further transmission studies in chimpanzees showed that NANBH seemed to be caused by a small agent. In 1989, Houghton and colleagues cloned and sequenced the genome of HCV (strain HCV-1) using a large samples size of NANBH-infected chimpanzee livers and plasmas samples and developed the first-generation blood diagnostic tests in 1990 (Bukh 2016; Houghton 2009).

HCV belongs to the *Flaviviridae* family and to the *Hapacivirus* genus (Simmonds et al. 2005), like Dengue virus and Zika virus (Chigbu

et al. 2019).

There were initially identified 6 different genetic groups of this virus called genotypes (GTs). The different GTs show a difference of the 30% in their genomic sequence. GTs are also divided into different subtypes. The subtypes of the same GT display a difference of 20% of the entire sequence between each other (Simmonds et al. 2005).

Genotypes are denoted with number and subtypes are denoted with lowercase letters (Ye et al. 2019).

In 2015 GT7 was identified in Central Africa (Murphy et al. 2015). Moreover, in 2018 was discovered in India a new genotype that was called 8 (Borgia et al. 2018). This novel HCV genotype, GT8, is genetically distinct from previously identified HCV GT1-7 with >30% nucleotide (nt) sequence divergence to the established HCV subtypes. The number of confirmed subtypes has increased from the 18 listed in 2005 to 67 in 2013, 86 in 2017 and 90 in May 2019 (https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification, 18th May 2019). HCV is still very present in some countries of the Middle East and the North Africa: 14,7% in Egypt and 4,8% in Pakistan (Chemaitelly, Chaabna, and Abu-Raddad 2015). The most widespread genotype in the world is the GT1 (46% of the total infections) (Gower et al. 2014), the most aggressive and more related to HCC and cirrhosis (Chigbu et al. 2019). Moreover, subtype 1b cause the 22% of the infections followed by GT3 (22%) and GT2 and 4 (13% each). GT1 is widespread in Australia-Asia, Europe and America. GT3 is mostly widespread in Asia, and it is usually related to steatosis and fibrosis (Chigbu et al. 2019), GT4 is very common in North Africa (71%) and Middle East (Gower et al. 2014) (Fig. 2). In addition to these typical local patterns, some "local epidemic" subtypes are observed like 4a in Egypt, 5a in Belgium (Jackowiak et al. 2014) and 3h in Campania, a region of southern Italy (Sodano et al. 2014) with a prevalence of 15,4% among GT3 infected patients (Minichini et al. 2018).

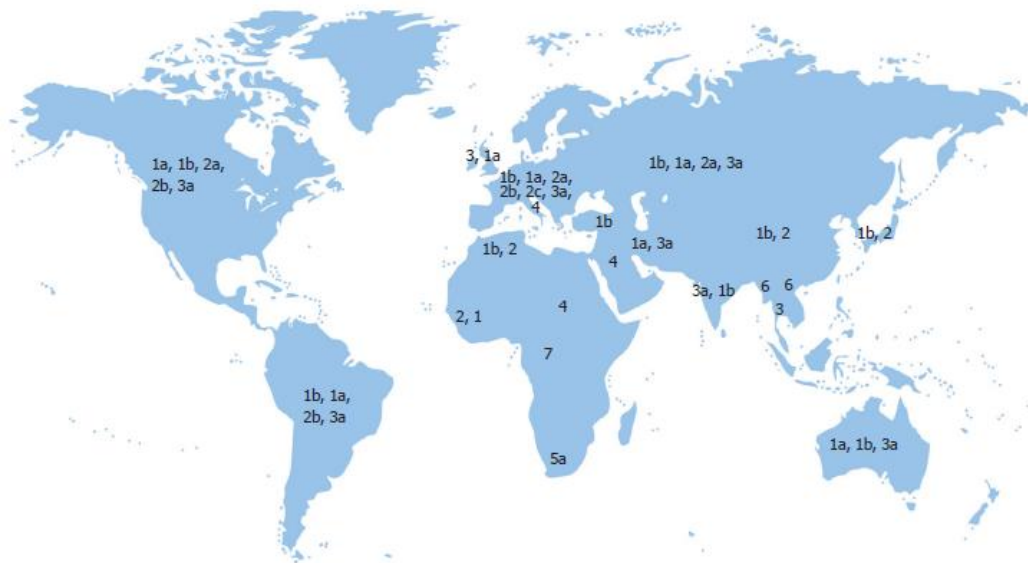**Fig. 2:** Geographical distribution of hepatitis C virus genotypes (Taherkhani and Farshadpour 2017)

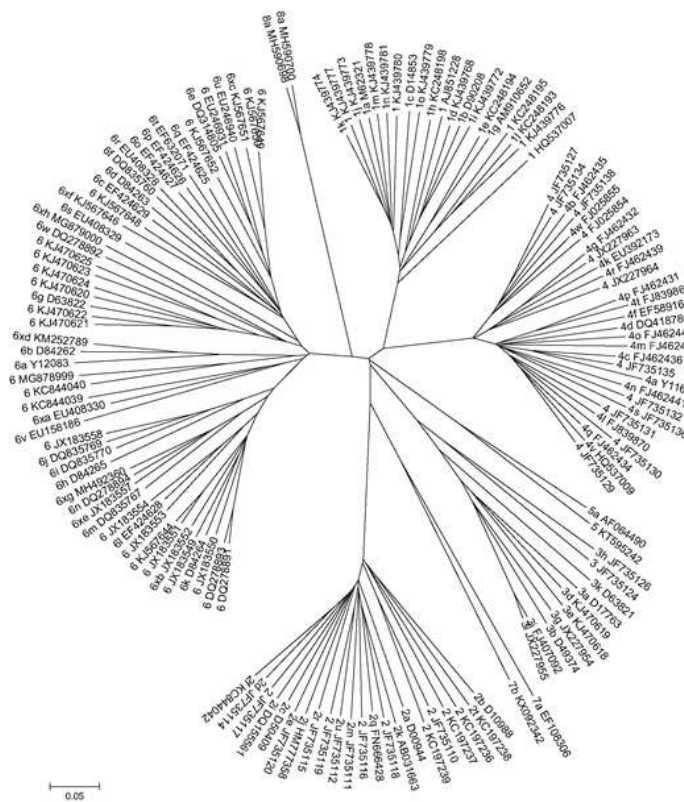The actual phylogenetic knowledge about HCV is described in the following phylogenetic tree (Fig. 3):

Fig. 3: A neighbour-joining phylogenetic tree of maximum composite likelihood nucleotide distances (https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification).

HCV viral particle has a diameter of 55-65 nm. It is formed by a bilayer pericapsid called envelope with a composition mostly lipidic, inside there is an icosaedric nucleocapsid formed by many units of CORE protein. This nucleocapsid contains a single strand RNA with positive polarity (Rosen 2011). HCV genome is 9600 nt long and it codify a single Open Reading Frame (ORF) of 3010 aminoacids (aa) (Tsukiyama-Kohara and Kohara 2017).

The viral RNA is divided in two parts (Fig. 4): the N-term, where there are the coding part for the structural proteins such as core, E1, E2, and also the channel protein p7 (Argentini et al. 2009), and the C-term with the non-structural (NS) proteins that are called NS2, NS3, NS4A, NS4B, NS5A e NS5B (Chevaliez and Pawlotsky 2006). The non-structural proteins have an important function in the replication of the virus. NS3 protein, for example, contains an RNA helicase, a nucleoside triphosphate (NTPase) and a NS serine protease (Chigbu et al. 2019). NS5B is an RNA-dependent RNA polymerase that is the main cause of the presence of genetically different HCV strains called genotypes. NS2 and p7 proteins are essential for the viral assembly and for its release (Chigbu et al. 2019). Both parts of the genome are flanked with two non-coding parts called 5'UTR and 3'UTR (untranslated region), 95-555 nt and 114-624 nt long respectively. They regulate the translation of the proteins (Chevaliez and Pawlotsky 2006).
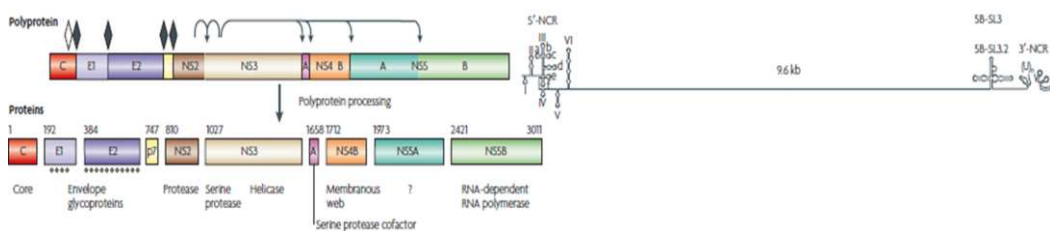
Fig. 4: HCV Genome structure (Moradpour, Penin, and Rice 2007)

2.3 Pathogenesis mechanism

Once the virus reaches the blood flow, it forms a complex with host's Low-Density Lipoproteins (LDL), Very-Low Density Lipoproteins (VLDL) and various Apolipoproteins (APOB, APOC and APOE) (Manns et al. 2017) that are called Lipoviroparticles (LVP). HCV virus starts its life cycle tiding to the glycosaminoglycans that are in the cytoplasm of the host's cells, from this moment the virus could be destroyed by the host's immune system.

HCV can be included by endocytosis when the interaction between the E1 and E2 glycoproteins expressed on the envelope and the receptors expressed by hepatocytes and lymphocytes B (such as SRB1, CD81, claudin-1, claudin-6, claudin-9, occludin, ephrin receptor type A2 and epidermal growth factors) occurs. This multi-receptor complex mediates the viral uptake inside the cell and defines organ and species-specificity (Manns et al. 2017). The endocytosis is mediated by Clathrin. Once inside the cell, the nucleocapsid is removed (uncoating) and the viral RNA is released into the cytoplasm (Chigbu et al. 2019) (Fig. 5). The formation of new virions does not depend on the active replication of viral RNA but only on the expression of non-structural proteins (Popescu et al. 2014). Human ribosomes produce directly from the positive polarity viral RNA a 3000 aa polyprotein which is divided into functional viral proteins by cellular and viral proteases. New-formed NS5B and NS3 proteins start the active viral RNA replication. During this process host factors such as microRNA-122 and Cyclophilin A helps viral replication (Chigbu et al. 2019).

Fig. 5: Entrance mechanism of HCV in the hepatocyte (Chigbu et al. 2019)
The origin of the inflammation process is due to the induction of

Reactive Oxygen Species (ROS) production. ROS reduce the intracellular concentration of glutathione and increase the production of peroxide-proteins. CORE and NS5A p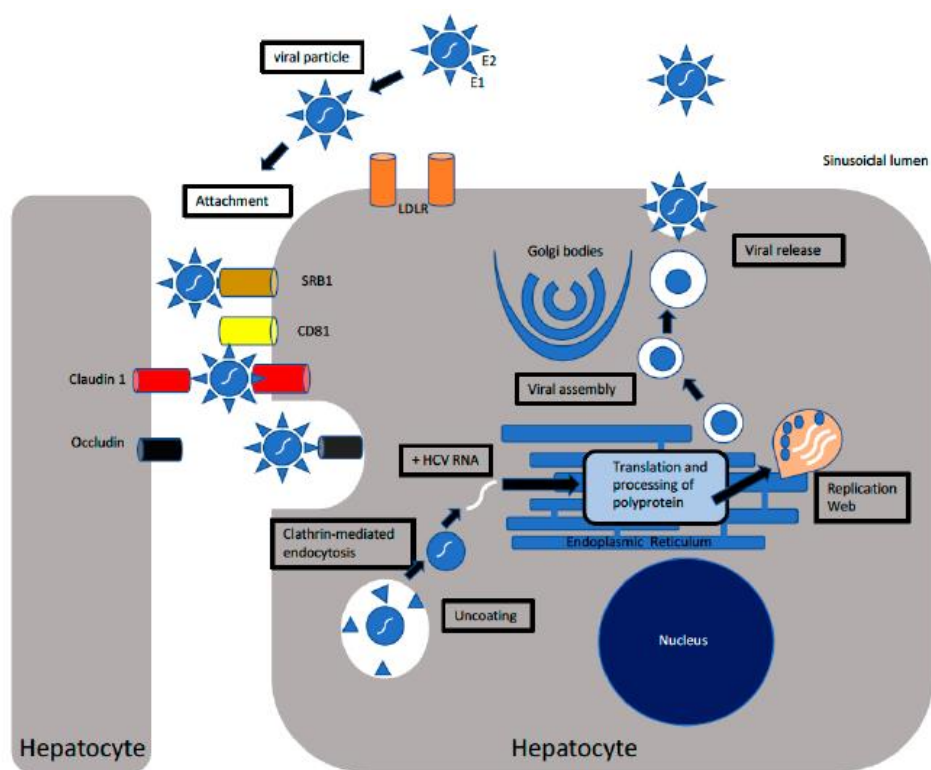roteins have a main role in the oxidative stress. In particular, the CORE protein could induce the production of ROS and the release of cytochrome C in mitochondria. On the other hand, the NS5A protein, induce an intracellular calcium release that leads to spontaneous production of ROS by the cell. The oxidative stress induced by HCV proteins leads to a pro-inflammatory cytokines production and to a chromosomal damage. Mutations could occur in the apoptotic chain such as β-globin, BCL, p53, and β-catenin genes in hypermutational areas, leading to carcinogenic process (Liang and Heller 2004).

2.4 Diagnosis

HCV infection is often asymptomatic, and it is difficult to diagnose during the acute phase. Usually, the HCV infection is detected when the hepatic damages are serious, and it becomes chronic. Hepatitis C is defined as chronic when the viral RNA persists more than six months. During the first six months from the contact with the virus, its elimination could happen in a spontaneous way, through the immune system's action (Wilkins et al. 2010).

Chronic infections of HCV are usually discovered after the analysis of the levels of hepatic enzymes or during a screening program of high-risks subjects (Wilkins et al. 2010).

The first diagnosis step is a serologic exam for the evaluation of the presence of Anti-HCV antibody. This first diagnosis test is made through ELISA technique. Even in a negative result, there is the possibility that the subject has been infected during the previous six months. It is recommended to repeat the test every 4-8 weeks until 6 months after the first result in all the high-risk subjects such as: HIV seropositive, drug addicts and people with risky sexual behaviour (Wilkins et al. 2010).

Since 2011, if the serological test results positive, a confirmation test is requested using the Recombinant Immuno Blot Assay (RIBA) (Chiron RIBA HCV 3.0 SIA). This test uses four recombinant HCV anti-antibody (c100p, c33c, p22p and NS-5) that are immobilized on the surface of the strip. If the antibodies in the patient's serum hybridize to anti-antibody probes, this results in different reactions such as: non-reactive, intermediate or reactive (Khudur Al-Nassary and Mahdi 2018). This assay was introduced in 2003 and discontinued in 2011 by US

Centers for Disease Control and Prevention (CDC) guidelines. Gong, Schmotzer, and Zhou, 2016 demonstrate that the omission of RIBA in the diagnosis algorithm does not affect the evaluation of the presence of the virus.

RIBA test was replaced with a qPCR assay for the quantification of the viral load. The COBAS®AmpliPrep/COBAS® TaqMan® HCV Quantitative Test, v2.0 (Roche Molecular Systems, Branchburg, NJ) is the most used Food and Drug Administration (FDA)-approved kit for HCV quantification via qPCR. The lower detection limit of the test is approximately 50 IU/ml in plasma and 60 IU/ml in serum. Thanks to the very low detection limit of this assay, HCV RNA could be found in the patient's plasma/serum from one or two weeks after the infection (Ozaras and Tahan 2009).

For the evaluation of the correct therapy, the virus genotyping is necessary, (Wilkins et al. 2010) in fact the viral genotype can influence the response to antiviral therapy (Manns et al. 2017). Nowadays, several genotyping methods are present on the market and they are all based on PCR technique. Especially, the most used viral region to determine the genotype are NS5B, CORE, E1 and 5' UTR regions (Petruzziello et al. 2016). Real Time PCR and Reverse Line Blot (RLB) techniques are the most used assay for virus genotyping. For example, VERSANT HCV Genotype 2.0 (Siemens Healthcare, Erlangen, Germany) is based on RLB technology and the assay targets simultaneously, the 5'UTR and CORE regions. While, the Abbott RealTime HCV Genotype II (Abbott Park, Illinois, U.S.A.) is based on Real Time PCR technology and the simultaneous analysis of 5'UTR and NS5B regions.

To elaborate a correct therapy for a naïve patient it is also necessary to evaluate the hepatic damage. The evaluation of the hepatic damage is required for the monitoring of the development due to therapy too (Wilkins et al. 2010). A chronic inflammation process in the liver leads to the fibrogenesis that is the substitution of hepatic parenchima with of extracellular matrix. For at least 60 years the gold standard method to evaluate the stadium of the progression of the fibrogenesis process was the biopsy of the liver. However, this method presents various limits due to the variability of the sampling process and the interpretation of the

results. Biopsy underrates cirrhosis presence in 10-30% of cases, on the bases of sample size (Rousselet et al. 2005). To avoid the interpretation variability of the biopsy results, different scoring systems were introduced during the year, such as Histological Activity Index, KNODELL HAI modified by Desmet et al. 1994, Histological grading and staging of chronic hepatitis C, Ishak et al. 1995 and finally Histological grading and staging of chronic hepatitis C METAVIR by Bedossa and Poynard 1996. All the models consider both necro-inflammatory activity and fibrosis, giving to these parameters a score on the basis of the severity of the damage.

Elastography is a new technique that could evaluate the stiffness of the liver expressed in kilopascal (kPa) through impulses. The result is highly reproducible and not invasive. This test is important to discriminate the level of the fibrosis severity during a hepatitis C infection (Foucher et al. 2006).

If fibrogenesis and necroinflammation are not stopped on time, cirrhosis pathogenic mechanism will start. It is characterized by nodular generation and creation of fibrotic septa that cause the liver structure collapse and its functionality (Tsochatzis, Bosch, and Burroughs 2014). Cirrhosis produces hepatocellular dysfunction and increases intrahepatic resistance to blood flow, which result in hepatic insufficiency and portal hypertension, respectively (Bataller and Brenner 2005). Major clinical complications of cirrhosis include ascites, renal failure, hepatic encephalopathy, and variceal bleeding. Patients with cirrhosis can remain free of major complications for several years. This asymptomatic state is called compensated cirrhosis. When symptoms appear, it is called decompensated cirrhosis and it is associated with short-term survival. Liver transplantation is often indicated as the only effective therapy (Davis et al. 2003). When liver cirrhosis is established, liver cancer may occur in the 3.5% of the cases (Conti et al. 2016). Hepatitis C- caused fibrosis, even in the cirrhotic range, could regress with specific antiviral therapy (Tsochatzis, Bosch, and Burroughs 2014).

HCC is an aggressive cancer that occurs in the advanced stage of cirrhosis. It is also the most common primary liver malignancy and the leading cause of cancer-related death worldwide (Balogh et al. 2016).

2.5 Therapy evolution

A vaccine to prevent the HCV infection has not been developed yet (https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019). It was observed that the virus tends to prevent the formation of antibody, using the tissue movement through the tight cells junctions (Brimacombe et al. 2011) avoiding immune system's surveillance. In fact, even immunocompromised people seem to be able to eliminate the virus. As a consequence, the vaccine against HCV does not necessarily have to induce the formation of antibody anti-HCV. Different experiments were made with recombinant proteins, peptides, plasmids and other vector-based approaches, but the costs for the production, the clinical assessment and the difficulty to obtain a good result in the animal experimentation are slowing down the project development (Swadling, Klenerman, and Barnes 2013).

Until 2011 the treatment was based on a mix of interferon apha-2b pegylated and ribavirin for a period of 24 weeks for GT2 and 3, and for 48 weeks for GT1 and 4 (McHutchison et al. 2009). Interferon alpha-2b pegylated is an immunomodulator that simulate the action of physiological interferon produced by immune system to stop viral replication. Ribavirin is an antiviral drug that acts on the transcription process of viral RNA. This therapy is characterized by numerous collateral effects and it succeeds in the 70-80% of cases with a GT2 or 3 infection and in the 45-60% of cases with GT1 or 4 (Zhang 2016).

The healing in the case of HCV infection is called Sustained Virologic Response (SVR). It is defined as the negative result of the virus presence after 4, 12 and 24 weeks, post therapy (Pearlman and Traub 2011). The therapy follow-up until the SVR reaching, is made by the evaluation of the viral load. The test, usually used for this purpose, is based on the Real Time PCR technology which have a high sensitivity to quantify the viral copy number in the patient's serum or plasma.

In May 2011 FDA approved the commercialization of new antiviral drugs called Direct Acting Agents (DAAs). During the experimental phase DAAs are proved to be more efficient than the dual therapy (interferon apha-2b pegylated and ribavirin). The first generation of DAAs includes Boceprevir and Telaprevir that act as inhibitors of the NS3 protease. The

use of these drugs lead to a major improvement of the SVR, from 45% to 70% together with Ribavirin in GT1 infections (Zhang 2016). These antiviral drugs were given only in naïve patients or in previously treated with interferon apha-2b pegylated. However, side effects of the treatment were still frequent (Chae, Park, and Youn 2013).

There is also a drug category that do not act on the virus directly, but on the endocellular structures of the host that interact with the pathogen itself. As previously discussed, Cyclophilin A (or peptidyl isomerase A) is a protein in the cytosol, involved in the viral replication that interact with cyclosporin. It is usually involved in the block of the calcium-calmodulin dependant phosphatase (calcineurine) and the rejection reaction of the organs after a transplantation surgery, stopping the production of TNF-α and interleuchin-2. The mechanism of action of the antiviral inhibitors of Cyclophilin A is not completely understood yet but from the *in vitro* studies it seems that its isomerasic activity inhibits the viral replication (Hopkins and Gallay 2012).

In 2013, the second generation DAAs were commercialized for the first time: Sofosbuvir and Simeprevir. With these new drugs the SVR increased to more than 90% without using interferon and reducing the patient's side effects (Zhang 2016).

Nevertheless, the eradication of the virus is still very difficult to achieve due to several reason, for example: the high cost of drugs, the absence of no-tested patients and the rapid evolution of viral forms resistant to DAAs (Tsukiyama-Kohara and Kohara 2017). In 2015 only the 20% of the 71 million infected were tested and among them, only 7,4% were cured due to the high cost of the therapy ("WHO | Global Hepatitis Report, 2017" 2018)

Currently the DAAs are divided into 3 different class based on the site of action (Fig 6):

- Viral protease NS3 inhibitors: Boceprevir, Telaprevir e Simeprevir
- Viral polymerase NS5B inhibitors, divided into:
  - Nucleosidic inhibitors (NI): Sofosbuvir
  - Non-nucleosidic inhibitors (NNI): Dasabuvir
- Viral polymerase complex NS5A inhibitors: Daclatasvir, Ledipasvir. (http://www.simit.org/IT/.xhtml. 31st May 2019).
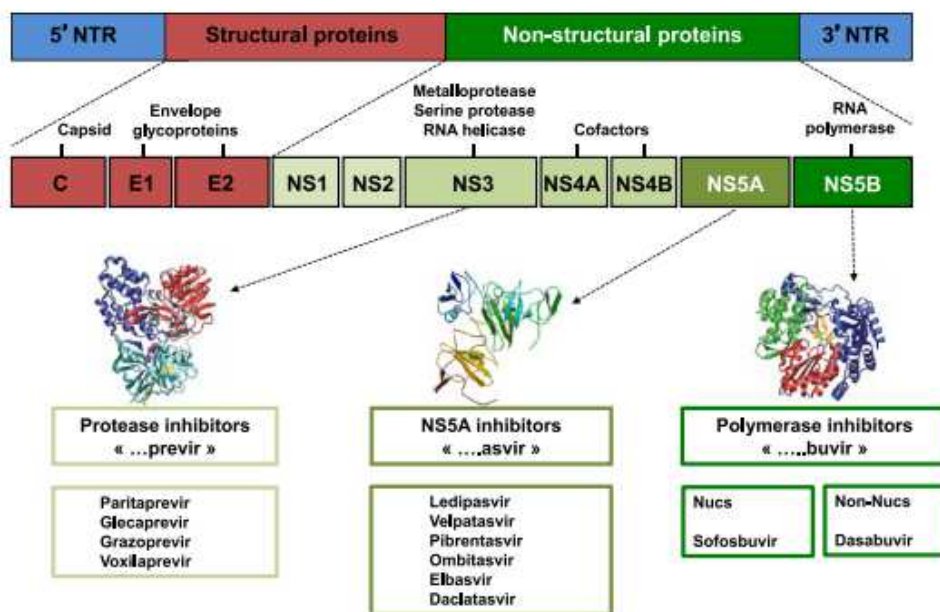
Fig. 6: Schematic description of HCV genome target of DAA's (Asselah, Marcellin, and Schinazi 2018)

In 2016 started the new drug era of pangenotypic antiviral therapy. In June 2016 the U.S. FDA approved Epclusa. It is a fixed-dose combination tablet containing Sofosbuvir and Velpatasvir and it is the first drug to treat all six major forms of HCV (pangenotypic treatment). Sofosbuvir is an inhibitor of NS5B protein and it has nanomolar *in vitro* activity against all HCV genotypes. Velpatasvir is a second-generation HCV NS5A inhibitor with antiviral activity against HCV replicons in genotypes 1 (Asselah, Marcellin, and Schinazi 2018). The Epclusa could be used for both patients with and without cirrhosis. For patients with moderate to severe cirrhosis (decompensated cirrhosis), the Epclusa is approved for use in combination with Ribavirin. (https://www.fda.gov/news-events/press-announcements/fda-approves-epclusa-treatment-chronic-hepatitis-c-virus-infection, 8th June 2019).

In July 2017 FDA approved Vosevi. It is a fixed-dose, combination tablet containing two previously approved drugs (Sofosbuvir and Velpatasvir) together with a new drug, Voxilaprevir. Voxilaprevir is a macrocyclic inhibitor of the NS3 and NS4A protease with picomolar *in vitro* antiviral activity against the 6 major HCV genotypes and an improved resistance profile compared to earlier protease inhibitors (Asselah, Marcellin, and Schinazi 2018). Vosevi could be used to treat adults with

chronic hepatitis C caused by genotypes 1-6 without cirrhosis or with mild cirrhosis. Moreover, it is the first treatment approved for patients who have been previously treated with the direct-acting antiviral drug sofosbuvir or other NS5A inhibitors such as Daclatasvir or Ledipasvir (https://www.fda.gov/news-events/press-announcements/fda-approves-vosevi-hepatitis-c, 8th June 2019).

In August 2017 FDA approved Mavyret. It is a fixed-dose combination of Glecaprevir and Pibrentasvir. Glecaprevir is a NS3/4A protease inhibitor with an *in vitro* nanomolar antiviral activity against HCV GT 1-6 and most known NS3 RASs. Pibrentasvir is an NS5A inhibitor with an *in vitro* picomolar antiviral activity against HCV GT 1-6 and most NS5A RASs (Asselah, Marcellin, and Schinazi 2018). It could be used in patients without cirrhosis or with mild cirrhosis, including patients with moderate to severe kidney disease and those who are on dialysis. Mavyret is also approved for adult patients with GT1 infection who have been previously treated with a regimen either containing, an NS5A inhibitor such as Daclatasvir or Ledipasvir, or an NS3/4A protease inhibitor such as Boceprevir, Telaprevir e Simeprevir but not both. Mavyret is the first treatment of 8-weeks duration (instead of 12 weeks) approved for GT 1-6 in adult patients without cirrhosis who have not been previously treated (https://www.fda.gov/news-events/press-announcements/fda-approves-mavyret-hepatitis-c, 8th June 2019).

At the moment only two pangenotypic DAAs regimens (Mavyret and Vosevi) are able to cure both treatment-naïve and treatment-experienced HCV patients with an SVR rate over 97%. Although the overall success of the new DAAs therapies, a small proportion of treated patients cannot achieve SVR. This is mainly due to several factors: cirrhosis, presence of other simultaneous infections, GT1a and 3 and pre-existing RASs selected by a previous DAAs regimen (Bourlière and Pietri 2019). A recent study (Chhatwal et al. 2018) claims that it is expected a failure to achieve the SVR in 124.000 (8.3%) people among the total of 1.50 million people that received a treatment between 2014 and 2020. These data demonstrate that despite the introduction of new pangenotypic therapies, the evaluation of genotype is still very important.

2.6 Genotyping assays

As mentioned earlier, HCV genotype is a substantial predictor of the response to the antiviral therapy. Therefore, many scientists have developed HCV genotyping assays. There are several molecular biology techniques to detect HCV RNA such as non-isothermal and isothermal nucleic acid amplification, lateral flow assays, nanotechnologies and sequencing (Warkad et al. 2018).

Commercial HCV genotyping assays are currently based on different techniques such as DNA sequencing, RLB, Real-Time PCR. They usually analyse 5'-UTR, CORE, NS5B regions. The most used *in vitro* devices are VERSANT HCV Genotype 2.0 assay (Siemens Healtcare) based on the analysis of 5'UTR and CORE regions by RLB technology, and Real-Time HCV genotype II (Abbott) based on the analysis of 5'UTR and CORE regions by Real Time PCR technology.

These assays could detect the most frequent HCV genotypes and subtypes, with some limitation that vary on the technique used. The most valuable system for the genotype/subtype evaluation is the sequencing. In particular, NS5B sequencing and phylogenetic analysis is considered the gold standard technique for HCV genotyping (Guettouche and Hnatyszyn 2005).

2.7 Genotyping techniques

2.7.1 Reverse Line Blot technology

The Reverse Line Blot (RLB) technology is based on the hybridization of the PCR products (amplicons) on specific probes, crosslinked on a nylon or nitrocellulose support. The visualization process is through a colorimetric reaction between alkaline phosphatase and nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate (NBT/BCIP).

Since HCV is a RNA virus a reverse transcription is needed before the amplification of the specific regions. The reverse transcription and the amplification have to be performed with biotinylated primers. The biotinylated amplicons are hybridized on spotted-specific probes in nitrocellulose strips and visualised through the RLB visualization protocol. After the hybridation of the amplicons on the probes, streptavidin alkaline

phosphatase conjugate is added. Streptavidin create a bond with biotinylated amplicons already hybridized with the specific probes. In the last step NBT/BCIP is added. The combination of NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) yields an intense, insoluble black-purple precipitate when reacted with alkaline phosphatase. It results in an intense purple-blue coloured spot where the probe was spotted, and only where biotinylated amplicons were hybridized.

The HCV genotyping assay analysed during this study is based on RLB technology. The intend use of this device is the detection of Hepatitis C virus (HCV) genotypes 1 to 7 and subtypes *a* and *b* of genotype 1 by one-step reverse transcription-PCR and Reverse Line Blot of the 5'UTR and CORE region. For each genotype and subtype different amplicons with different sequences will be formed. For each different amplicon there is the correspondent probe on the strip. This means that each different genotype and subtype reacts with different probes leading to a specific pattern of reacting probes for each subtype. The relationship between the probes pattern and the subtypes is resumed in an interpretation table that is periodically updated based on the new patterns found. All patterns are confirmed with Sanger sequencing and phylogenetic analysis.

Specifically, for this assay the identification of genotype and subtype is due to 5'UTR. The only exception is for subtype 1a and 1b and genotype 6 c-v that are recognized by specific probe in which only CORE region amplicons hybridize. In absence of bands for the CORE region HCV subtypes 1a and 1b cannot be discriminated. Genotyping is based solely on analysis of the 5'UTR. In addition, presence of the HCV 6 subtypes c-v cannot be excluded.

2.7.2   Next Generation Sequecing technology

DNA sequencing technique was developed in 1977 by Frederick Sanger. It is based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication. This revolutionary discover that allows to start to know the DNA structure in detail, however has a low throughput capacity because it permits to sequence one amplicon per reaction not longer than 500-600 bp.

At the beginning of new century, the development of a new technology was started. It was called Next Generation Sequencing (NGS) and since the beginning, it was characterized by higher throughput capacity. During the last

years different sequencing platforms were developed with different throughput capacity and different sequencing chemistries.

Since 2005, NGS technologies have changed high-throughput genomic research and have opened up many new research areas and novel applications. Many platforms were developed by companies such as, Roche Applied Science (454 Genome Sequencer FLX [GS FLX] System; CT, USA), Illumina (Genome Analyzer [GA] II; CA, USA), Life Technologies (Sequencing by Oligonucleotide Ligation and Detection [SOLiD™]; CA, USA) and Helicos BioSciences (HeliScope™ Single Molecule Sequencer; MA, USA). These sequencing platforms differ from each other in sequencing chemistries and technical details.

The final aim of the NGS technology is to overcome the limited scalability and sensitivity of traditional Sanger sequencing and have a high data throughput.

During the last few years, this technology starts to be applied in diagnostic. The most used sequencing platform in diagnostic field are Illumina, in particular MiSeq and iSeq100™ (Illumina) platforms. Illumina technology is based on Sequencing by Synthesis (SBS). This technique is similar to the Sanger one and it provided labelled nt. In particular, MiSeq (Illumina) platform requires four different labelled nt, and iSeq100™ (Illumina) platform has a one colour chemistry.

Sequencing by Synthesis

For all the NGS protocol, the DNA or RNA target to be investigated need to be isolated and prepared for sequencing. The selection and preparation process of the fragments is called Library Preparation.

The library preparation process for the Illumina platforms consists in 4 steps:

- Target enrichment. There are different strategies to select the target. All the fragment that will be sequenced have to be bind to different oligonucleotides, such as:

Illumina Adapters: complementary sequences to the flow cell probes, called P5 and P7 or PE1 and PE2.

Index: sequencing indexes are oligonucleotides sequence used for

the multiplexing of different samples in the same sequencing run. Each sample is associated to its own index, and during the data analysis each sample will be recognized through the index labelling.

All the libraries prepared have the following structure (Fig.7):
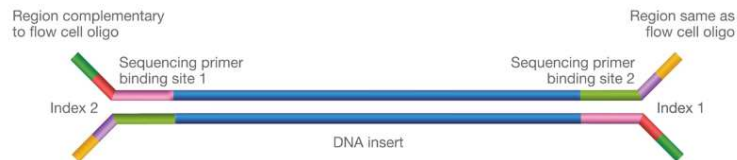
Fig. 7: Composition scheme of a library typical fragment

- Flow cell hybridization. All the fragment are denatured and pooled together in single tube. The samples pool is loaded in the sequencing cartridge were all the reagent for the sequencing are prefilled. The fragment pool is automatically poured in the flow cell surface. Flow cell is a slide in which two different probes are hybridized. Those probes are complementary to the Illumina Adapters of each fragment. Fragments in turn hybridized to the probes and a polymerase create the complementary strand of each fragment. At the end of the process, the original fragment will be washed away (Fig. 8).
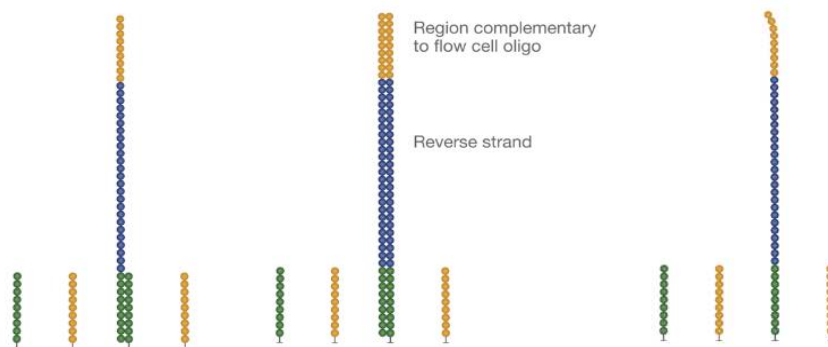
Fig. 8: Hybridization process of a fragment in the flow cell surface.

Bridge amplification and cluster creation. The free end of the fragment will bend to the nearest complementary flow cell's probe. A polymerase will create the second complementary strand of the bended fragment. The strands are denaturated and both parts are maintained, and they will continue the bridge amplification process (Fig. 9).
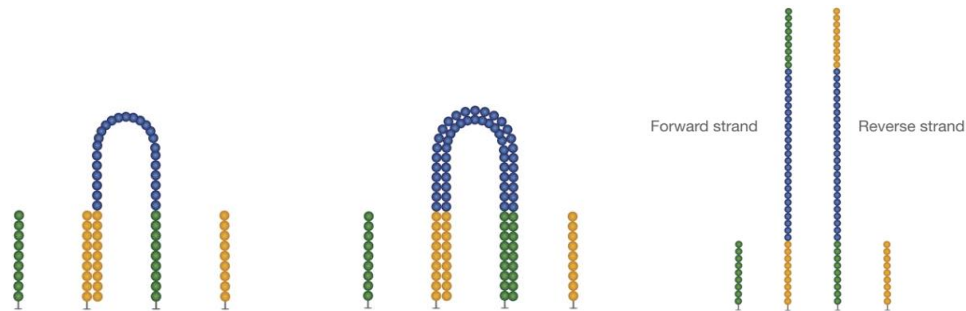
Fig. 9: Bridge amplification scheme

At the end of the bridge amplification process, groups of clones will be generated. That spot of clones generated by singles different fragments are called clusters (Fig. 10).
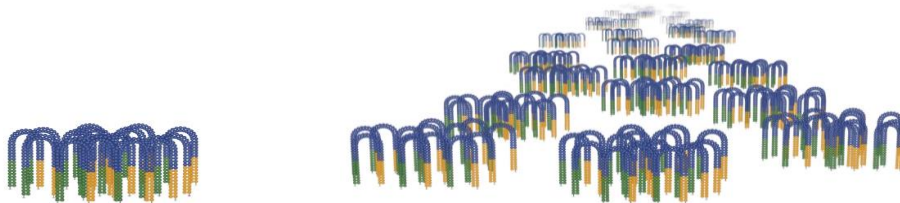
Fig. 10: Cluster creation scheme.

- Sequencing. Once the clusters are created, the double strands are denaturated and the reverse strands are washed away. Sequencing is made by cycles: at each cycle, nt are released. The nucleotides' labels are fluorescents and the fluorescence signal is read by a detector. Clusters allowed to amplify the fluorescence signal, because they are all made by the clones of the same fragment. The cycles are different in relation to the chemistry that the platform uses.

MiSeq (Illumina) Chemistry

MiSeq (Illumina) platform is based on four channels SBS. At each sequencing cycle all four nucleotides are released on the flow cell. Each

nucleotide is labelled with a different fluorophore. A polymerase adds, to the newly polymerized complement strand of each fragment, the most suitable nucleotide on the bases of the complementary rules of the DNA synthesis. Once the nucleotide is added, the fluorophore is released and detected. For each cluster, the same fluorophore during a cycle is released, amplifying the signal coming from the original fragment that originated the cluster.

For each cycle, an image is created of the signals coming from each cluster. The sum of all the image for each cycle, determined the final sequence of each fragment. The sequence of each cluster is called *read*.

iSeq™ 100 (Illumina) Chemistry

iSeq™ 100 (Illumina) platform is based on a one channel SBS. One channel sequencing is an evolution of the previous four (MiSeq and HiSeq) and the two channels (NextSeq) chemistry. It simplifies the nucleotide detection and requires two images to define the added nucleotide. One-channel SBS is based on the usage of a complementary metal-oxide semiconductor (CMOS) chip.

The system uses a patterned flow cell with nanowells fabricated over a CMOS chip. The clusters are already defined inside a nanowell and they are no more randomly distributed on the flowcell.

Unlike four-channel SBS chemistry, where sequencers use a different dye for each nucleotide, the iSeq100™ (Illumina) System uses one dye, two chemistry steps, and two imaging steps per sequencing cycle. In one-channel chemistry, adenine has a removable label and is labelled in the first image only. Cytosine has a linker group that can bind a label and is labelled in the second image only. Thymine has a permanent fluorescent label and is therefore labelled in both images, and guanine is permanently dark. Thanks to the ON/OFF pattern, the combination of the two images identifies the nucleotide.

- Data Analysis. At the end of the process, FastQ files with all reads will be analysed. Algorithms of the instrument used, will assign the reads to each specific sample using index, the algorithms of the bioinformatic pipeline will eliminated the low-quality reads and compared the complementarity between forward and reverse reads in paired-end sequencing analysis.

The first application area of NGS in human diagnostic was oncology, in particular for solid tumours. In fact, many diagnostic companies developed *in vitro* kits for point mutation, insertion or deletion in driver genes in the most common solid tumours. Other NGS kits can reveal fusion, Copy Number Variation and translocation.

More recently, NGS technology was applied in the microbiology field. Typical NGS methods are useful to discover new microorganisms and viruses and their drug-resistance mutations. The most used techniques are metagenomic, whole genome sequencing (WGS) for the investigation of microbial communities both in the environment and in human body niches. Moreover, thanks to the sequencing depth of NGS it is possible to analyse the variability of the virus within the host (i.e., *quasispecie*) and detect the low-abundance antiviral drug-resistance mutations in patients with HIV infection or viral hepatitis (Barzon et al. 2014).

Nowadays, NGS application in microbiology and virology is still rarely used in diagnostic due to technical challenges of the procedure, also the extreme variability of the organism increases the difficulties to perform the NGS test.

As a consequence of this poor application of the technology in the viral diagnostic field, there are not troubleshooting guides yet. One of the most studied viruses via NGS is HIV but the percentage of the paper regarding the detection of the HCV with NGS is still very low.

NGS technology could be used both for HCV genotyping and detection of drug resistance RASs.

2.8 Therapeutic failure causes

The new antiviral therapies cannot achieve the 100% of SVR for three main reasons:
- the presence of Resistance Associated Substitutions (the most common cause),
- recombinant genotypes
- dual infections.

2.8.1   Resistance Associated Substitutions - RAS's

HCV polymerase is defined as "error-prone" because it does not have a proof-reading activity. It was calculated that there is an error introduction rate of 10^{-3} errors/per site, especially for G:U-U:G mismatch (Tsukiyama-Kohara and Kohara 2017).

The mutations in the HCV genome are the result of external factors and polymerase characteristics. From a theorical point of view, NS5B polymerase can introduce mutation in random position of the genome. However, it was observed that there are regions in the viral genome in which mutations are more frequent such as the hypervariable region of E2 protein or the V3 domain of NS5A protein.

HCV variability is characterized by substitutions, with a higher frequency of transition than transversion. The mutations are mostly synonym or missense. Rarely they form STOP codons and insertion or deletion have a low frequency (Powdrill et al. 2011). This natural predisposition to variability in replication process, makes the viral population highly heterogenous: inside the host there is a huge viral populations complexity, strictly connected to each other. These populations are called quasispecie. Inside the patient, after the infection there will be a master variant and with the passing of time and replication cycles, a great number of variants will be formed with different distribution probability (Jackowiak et al. 2014).

Quasispecie are always submitted to a genetic selection inside the host, especially when antiviral drugs are introduced. Drugs represents a selection factor based on resistances creates by casual mutations that differentiate a quasispecie from the other. This casual mutation could have a drug-resistance effects (Cento, Chevaliez, and Perno 2015) and lead to a selection of the carrying quasispecie, and consequently a new increasing of the viral load.

Initially both the mutation and the quasispecie were called Resistance-Associated Variants (RAVs). This definition was considered not accurate for mutations and now they are called Resistance-Associated Substitutions (RASs) (Pawlotsky 2016). Drug-resistance mutations to be diagnosed are the ones in the viral genomic region hit by drugs such as NS3, NS5A and NS5B, because these genomic viral regions are the target of the DAAs. It was observed that the viral region most affected by RAS is NS5A. The mutations in that region cause a 100-fold decrease of the efficiency of the drugs in the *in vitro* experiments (Wyles 2017).

Before the antiviral drugs assumption, the mutations can be poorly expressed for two main reasons. Firstly, the meaningless selective pression before the therapy. In fact, all the quasispecie formed during the viral replication could exist without the selection caused by the therapy that eliminate the population without the specific RAS, that after therapy will become overexpressed. Secondly, random acquired mutations can lead to a damage in the virus fitness. Disadvantageous mutations cannot survive in the viral population. (Cento, Chevaliez, and Perno 2015). Before the therapy is started, it is difficult to detect the presence of quasispecie with critical mutation, in fact the sensitivity of the molecular biology techniques used nowadays is not enough.

The introduction, in the diagnostic routine, of a system to detect poorly expressed mutations to prevent the therapeutic failure seem to be necessary. Nowadays, the only technology that could have a sufficient sensitivity is NGS (Cento, Chevaliez, and Perno 2015). The new generation sequencing, differently from Sanger, allows to obtain thousands or millions of sequences for each sample. In the case of HCV drug-resistance mutations it could simultaneously sequence all the quasispecie in the sample with a sensitivity that vary from the sequencing depth. The introduction of such a powerful instrument in the diagnostic routine can lead to a more accurate selection of the antiviral drugs, avoiding the risk of therapeutic failure. An updated list of all the clinically relevant mutations is found in Geno2Pheno website (https://ngs.geno2pheno.org/hcvrules).

For Boceprevir, Telaprevir e Simeprevir that inhibit the NS3 protein, some prevalent RASs were found such as: V36A/M, T54A/S, V55A, Q80R/K, R155K/T (mostly in GT1a), I/V170A, D168A/E/K/T/V/Y and A156S/T/V that gives to the virus the highest level of resistance (mainly in GT1b). Post therapeutic failure analysis revealed that RASs are present in 80% of the patient that underwent to triple therapy with Boceprevir or Teleprevir. A special mention has to be made for Q80K. It is associated with significantly lower SVR rates for treatment with the triple therapy with Simeprevir, and it exists as a natural polymorphism mainly in HCV GT1a. In fact, the prevalence of this mutation in GT1a is 20–52%. The presence of Q80K is problematic in cirrhotic patients. In fact, in case

of cirrhosis, the presence of Q80K lowers SVR rates to the 74% versus the 92% in its absence. Therefore, monitoring of Q80K prior the therapy is recommended in all HCV GT1a infected patients starting the triple therapy with Simeprevir. While for therapy with SMV+SOF, testing is needed only for cirrhotic patients (Cuypers et al. 2016).

For inhibitors of NS5A protein such as Daclatasvir and Ledipasvir the most common RASs that were found are M/L28T/V (GT2), Q/L30E/H/R/S, L31M/V, H58D, Q54H and Y93C/H/ N that has a natural prevalence of 10% and confer resistance both for Daclatasvir and Ledipasvir in GT1b (Cuypers et al. 2016).

Finally, for NS5B inhibitors such as Sofosbuvir and Dasabuvir, L159F and V321A were found in several infected subjects who did not achieve SVR, with the highest proportion of failures detected in HCV GT1a, GT2 and GT3 infected patients. Also, C316N/H/F was found before the treatment in six case of therapeutic failure GT1b infected subjects and in one GT1a relapsing patient. For non-nucleosidic inhibitors such as Dasabuvir, commonly observed NS5B substitutions are M414T and S556G, or A421V and P495L/S (Cuypers et al. 2016).

The speed in the selection of the variants suggest that these drug-resistance mutations exist before the treatment. Recent studies that use NGS technology confirm that these quasispecie mutations are present in the host with a low frequency before the therapy. Data suggest that Early Virological Response (EVR - lowering of the viral load higher than 2 log), is correlated to heterogeneity of quasispecie population before the therapy intake (Jackowiak et al. 2014). The majority of the therapeutic failure cases are due to the presence of mutations that are not revealed by routine diagnosis, it is clear that there is a urgency to implement a sensitive system for their detection.

2.8.2 Recombinant genotypes

As a consequence of the viral genome variability, in the recent years a few natural inter-genotype and intra-genotype recombinants of HCV have been identified. Recombination is classified by the structure of the crossover junction: if recombination occurs at the same site in both parental templates it will be called homologous and it will produce molecules with the parental structure. If recombination occurs between different sites of the involved molecules it will be

called non-homologous, and genomes with duplications, deletions, or insertions will be formed. Consequently, homologous recombination occurs between related viruses that share similar genetic structure, whereas non-homologous recombinants can also be generated from molecules of very different origin. Recombination between different HCV strains it's always homologous (Galli and Bukh 2014).

Two mechanisms of recombination were identified for RNA viruses: the replicative one with a copy–choice model, where the viral polymerase changes template during synthesis of the nascent strand, produce a chimeric genome. In the second mechanism, the non-replicative, genetic fragments of different origins can be joined together generating hybrid genomes. Moreover, the breakage–rejoining model supposes that the nascent strand can dissociate from the RNA template (donor) and interact with a different template (acceptor) or with a different region of the same template. Nowadays we don't know exactly which is the recombination mechanism of HCV (Galli and Bukh 2014).

Since 2002, nine different inter-genotypic recombinant forms of HCV have been described (Table 1):

Recombinant form (RF)	Breakpoint*	Accession	Isolates	Reference
RF2k/1b	3186	AY587845	33	(Kalinina et al., 2002, 2004; Kurbanov et al., 2010)
RF2i/6p	3405-3464	DQ155560	1	(Noppornpanth et al., 2006)
RF2b/1b_1	3456	DQ364460	1	(Kageyama et al., 2006)
RF2/5	3366-3389	AM408911	1	(Legrand-Abravanel et al., 2007)
RF2b/6w	3429	EU643835	1	(Lee et al., 2010)
RF2b/1b_2	3432	AB622121	1	(Yokoyama et al., 2011)
RF2b/1a	3429-3440	JF779679	1	(Bhattacharya et al., 2011)
RF2b/1b_3	3286-3293	AB677530	1	(Hoshino et al., 2012)
RF2b/1b_4	3286-3293	AB677527	1	(Hoshino et al., 2012)

*in comparison with H77 isolate (AF009606)

Table 1: Inter-genotypic recombinant list of ICTV database (https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/38/table-4-recombinant-rf-hcv-genomes**)**

Until now, no multiple recombinants have been reported. The first recombinant was observed in Saint Petersburg (Kalinina et al. 2002) characterized as a Recombinant Form (RF) 2k/1b. Since then many other isolated were identify. In Georgia for example, 76% of GT2 patients were infected by RF2k/1b. RF2k/1b was found in different Middle east and Western Europe countries such as Russia, Georgia, Ukraine, Azerbaijan,

Chechnya, Kazakhstan, Armenia, Israel, Germany, Romania, Greece and Tadzhikistan (Susser et al. 2017).

Kalinina, Norder, and Magnius 2004, describe that in the parent strain of RF2k/1b there are two stable hairpins structure called HS1 and HS2, upstream and downstream of the breaking point. HS1 structure involves 58 to 88 nt depending on the genotype 1 strain and HS2 is formed by 60 nt in the sequence of subtype 2k strain. It seems that the RNA-dependent RNA polymerase of HCV usually pauses in the regions with strong secondary structure situated in the template. The preserved genomic structure of the recombinant shows that the chimera structure is formed from a homologous recombination, which occurs when two related RNA molecules recombine at corresponding sites. Intra genotypic recombinant forms of HCV were found such as: 1a/1b from Peru (Colina et al. 2004). In the following image, a resume of the HCV RFs found until 2014 (Fig. 11):
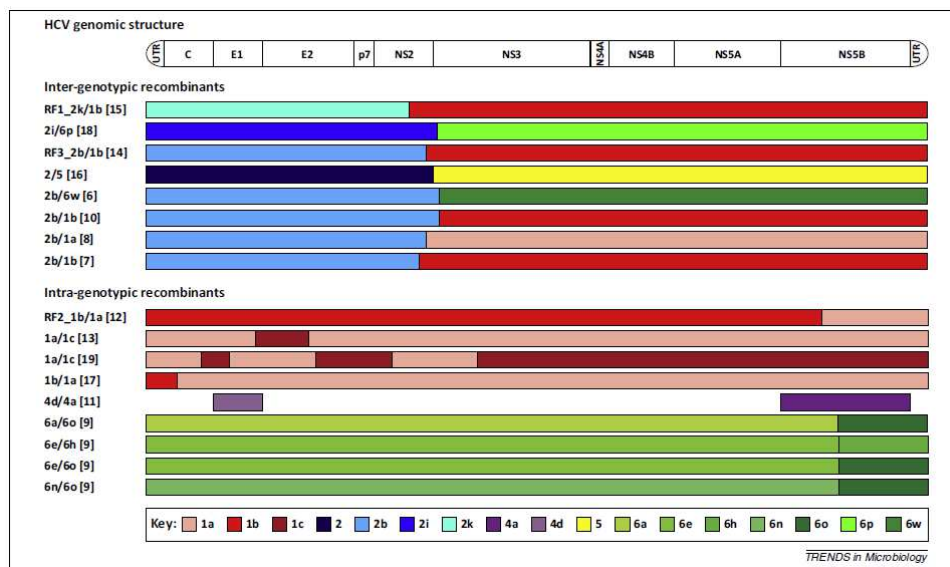
Fig. 11: Description of the major inter- and intra-genotypic recombinant found (Galli and Bukh 2014)

In diagnostic practice, HCV is usually typed with one of the several commercial kits available, their aim to detect genotype or subtype-specific mutations in conserved regions of the genome such as 5'UTR and CORE regions. Consequently, those devices are not able to detect recombinant strains and RF2k/1b will always results as a GT2.

De Keukeleire, Descheemaeker, and Reynders 2015 insist on the importance of determining the HCV genotype also by NS5B sequencing. The

detection of the real genotype and subtype is a useful guide for clinical decision-making.

Due to the limited information on susceptibility of the recombinant 2k/1b HCV genotype to antiviral treatment, no recommendations exist on the duration of therapy or the best treatment to use. As reported by Zakalashvili et al. 2017 a combination of Sofosbuvir with interferon apha-2b pegylated and Ribavirin for 12 weeks was used in Georgian patients and it seemed to be effective with an SVR12 of 96% but collateral effects for the interferon apha-2b pegylated usage have to be considered.

Recommended indications by AASLD ("HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C Welcome and Methods" 2014) for GT2 therapy are: daily fixed-dose combination of Glecaprevir (300 mg) / Pibrentasvir (120 mg) for 8 weeks or daily fixed-dose combination of Sofosbuvir (400 mg) / Velpatasvir (100 mg) for 12 weeks or daily Daclatasvir (60 mg) plus Sofosbuvir (400 mg) for 12 weeks in naïve patients without cirrhosis.

The RF2k/1b genome structure shows that all the DAA's target regions (i.e. Non-Structural proteins) belong to the GT1b part of the recombinant form.

Recommended indications by AASLD ("HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C Welcome and Methods" 2014) for GT1b therapy are: daily fixed-dose combination of Elbasvir (50 mg) / Grazoprevir (100 mg) for 12 weeks, daily fixed-dose combination of Glecaprevir (300 mg) / Pibrentasvir (120 mg) for 8 weeks, daily fixed-dose combination of Ledipasvir (90 mg) / Sofosbuvir (400 mg) for 12 weeks or daily fixed-dose combination of Ledipasvir (90 mg) / Sofosbuvir (400 mg) for patients who are non-Caucasian, HIV-uninfected, and whose HCV RNA level is <6 million IU/mL for 8 weeks, daily fixed-dose combination of Sofosbuvir (400 mg) / Velpatasvir (100 mg) for 12 weeks for treatment-naïve patients GT1b without cirrhosis.

If the chosen therapy for the detected GT2 isn't effective for GT1b target proteins the results could be a therapeutically failure.

Recombination can act as a catalyst for antiviral resistance because chimera forms can inherit the mutations of both parental strains.

Consequently, a more intense surveillance of recombination in HCV is necessary to prevent the further spread of resistance mutations (González-Candelas, Xavier López-Labrador, and Bracho 2011).

2.8.3 Dual infections

Coinfection with 2 or more distinct HCVs genotypes has been documented. The possibility of a double infection in highly exposed people is due to the lack of protective immunity generated by HCV infections and the possibility of re-exposure through in IDUs. The high prevalence of multiple infection in IDUs (Blackard and Sherman 2007) reported a rate of reinfection as a result of continued injection drug use estimated in 0–4.1 cases/ 100 person-years and the association with high-risk behaviour indicates that ongoing injection and needle sharing following primary infection can lead to subsequent acquisition of new HCV strains (Pham et al. 2010). Nevertheless, Mehta et al. 2002 observed that HCV infections were more frequent in naïve IDU patients (21%) than reinfection in the same group (12%) suggesting that the first exposure could partially protect against reinfection.

The *reinfection* term, as it's used for HIV infection, refers to a primary infection that is completely cleared before the secondary infection occurs, with either a homologous or a heterologous HCV.

Dual infections occur when an individual is infected by two different HCV genotypes. Dual infections could also be classified in *coinfections* and *superinfections*. Coinfection is an infection with 2 or more different HCV genotype simultaneously. Superinfection occurs when a second HCV strain infected the patient after the establishment persistent HCV infection and the development of an immunologic response to the first virus (Blackard and Sherman 2007).

The devices in use for HCV genotyping in clinical practice, either based on Real Time PCR or on Reverse Line Blot technologies and even Sanger sequencing, could not differentiate two or more genotypes in the same patient. So, nowadays it's impossible to determine the presence of both a coinfection or a superinfection because of this technical limitation.

3. AIM OF THE STUDY

Biofield Innovation is a Start Up company that, since 2015, offers specialized human resources and know-how for Research and Development, and Engineering.

Developments included *in vitro* diagnostic assays for molecular biology application such as Real Time PCR and Reverse Line Blot for the detection of the most widespread viral and bacterial infections and for the detection of genetic disease such as thrombophilia, hemochromatosis and celiac disease.

The company Research and Development department is involved in the study of a diagnostic system for the HCV genotyping.

A Reverse Line Blot technology-based protocol for HCV genotyping was already present in laboratory. It consists on a one-step multiplex retrotranscription and amplification of 5'UTR and CORE regions and a visualization part of the amplicons by Reverse Line Blot technique.

During internal evaluation, it was observed a lowering of performance in the genotype 1 subtyping. Preliminary analysis results demonstrated that to have a better performance in CORE region amplification, it was necessary to redesigned the mastermix protocol. For this assay, Reverse Line Blot technology was chosen because of the low-cost applicability to the majority of the diagnostic laboratories in Italy and in the world, and because only an end-point thermocycler and a Dubnoff bath are needed for the protocol. But, this kind of technology has some limitations due to the probes and the limited number of genotype and subtype detected.

In the last years the attention in HCV detection is no longer focus only on genotyping, but also in all the causes of therapeutically failure such as Resistance Associated Substitutions (RASs). The current method for the RASs diagnosis is the Sanger sequencing. However, the sensitivity of that method cannot detect low frequency RASs. A new step forward in the diagnostic technology is required to satisfy this new clinical need.

Next Generation Sequencing (NGS) technology is already widely used in oncological field for the detection of high and low expressed mutation in solid tumours or in liquid biopsy. Since it's invention, this technology has been applied to oncology diagnosis. Initially results were

given only by experimental research. After some years from the first appearance, guidelines for diagnosis using NGS technology has been written and CE-IVD kits were commercialized. Unfortunately, virology diagnosis field is not so developed yet, and fewer CE-IVD kits have been commercialized.

Biofield Innovation started in the last two year the development of a genotyping device based on the NGS technology. The aim of the device is HCV genotyping and subtyping using 5'UTR region and part of CORE region together with NS5B region, but also detecting the main RASs in NS3, NS5A and NS5B regions for GT1 samples.

With the regions selected, it will be possible to detect recombinant genotypes. Also, the higher depthof NGS technology can detect poorly represented RASs as well as the presence of mixed HCV infections and recombinant form cases.

The aim of this study is the restyling of HCV genotyping assay based on Reverse Line blot technology and the development of the new HCV NGS kit for genotyping and RASs detection of HCV.

4. MATERIAL AND METHODS

4.1 HCV genotyping assay based on Reverse Line Blot technology

The developing of a prototype with improved performance was based on the study of the previous assay. The entire protocol and the design of the assay is not an aim of this work.

The project for the performance improving of the device is all based on alignments made on Los Alamos (https://hcv.lanl.gov/content/index) and GenBank (https://www.ncbi.nlm.nih.gov/genbank/) databases. The performance improvement was obtained redesigning the primer for the amplification of the CORE region.

4.1.1 Clinical samples

For this study, some not-subtyped GT1 samples were collected.

Viral RNA from all samples was extracted with EZ1 Advanced XL (QIAGEN) with EZ1 DSP Virus Kit (QIAGEN) starting from 400 µL of serum or plasma and eluted in 60 µL. Carrier-RNA was used according to manufacturer user manual and Internal Control (IC RNA) was added. RT-PCR were performed on SimpliAmp Thermal Cycler (Thermo Fisher Scientific) and visualization protocol was performed on AUTOBLOT 3000H (MedTec) following the protocol of previous assay.

4.1.2 Synthetic samples

For the prototype under development, synthetic samples were made starting from DNA fragments.

Custom DNA were synthesised by GeneArt Strings DNA fragment service (Invitrogen - Thermo Scientific) for genotype HCV 1a, 1b and 6a starting from sequence with accession number KC844049.1 for HCV 1a and AB049088.1 for HCV 1b and KJ678756.1 for HCV 6a.

The DNAs were converted to RNA with MEGAshortscript T7 kit (Applied Biosystem) and purified with MEGAclear TM kit (Ambion) according to manufacturer indications. Synthetic RNA was quantified and

diluted to a final concentration of 10^3 copies/µL. RT-PCR were performed on SimpliAmp Thermal Cycler (Thermo Fisher Scientific) and visualization protocol was performed on AUTOBLOT 3000H (MedTec) following the instruction reported on AMPLIQUALITY HCV TYPE PLUS (AB ANALITICA) user manual.

4.1.3 Sanger sequencing

For all clinical samples, Sanger sequencing of the NS5B region was performed. Primers from Sandres-Sauné et al. 2003 were used. Thermal profile is described in the following table (Table 2):

Temperature	Time	Cycles
48°C	30 min	1X
95°C	10 min	1X
95°C	30 sec	
52°C	60 sec	40X
60°C	60 sec	

Table 2: Thermal profile used for NS5B region amplification

Sanger sequencing was performed on ABI PRISM 3130 (Applied Biosystem) using DeepCheck single round RT-PCR and sequencing (Advanced Biological Laboratories).

The phylogenetic analysis for the identification of the viral sequences were performed using the Basic Local Alignment Search Tool (BLAST - https://blast.ncbi.nlm.nih.gov) and the alignment with available HCV sequence on GenBank database.

4.1.4 Primer design

The tool used for the primer design was Primer Express™ Software v3.0.1 (Thermo Scientific) and dnaMATE v 1.0 (https://www.mybiosoftware.com/dnamate-1-0-consensus-melting-temperature-prediction-server-short-dna-sequences.html) for the melting temperature checking. Also, RealTime Design qPCR Assay Design Software (Biosearch technologies) was used for deepest evaluation on multiplex primer mix created with the new primers.

Specificity of the new primers were checked with BLAST. The specificity of the new primer was tested also *in vitro* using the new designed primers with DNA samples positive for the following target: *Epstein-Barr virus* (EBV), *Herpes simplex virus 1* (HSV-1), *Herpes simplex virus 2* (HSV-2), *Cytomegalovirus*

(CMV), *Varicella-zoster virus* (VZV), *Parvovirus B19*, Adenovirus (ADV), *BK virus* (BKV), *Enterovirus*, *Human herpes virus type 6* (HHV 6), *Human herpes virus type 8* (HHV 8), *JC virus* (JCV).

These pathogens were selected as they can be found in the same sample matrices of HCV such as serum or plasma.

4.1.5 CORE region amplification

To define the GT1 samples subtype in undetermined cases, the analysis of CORE region using the same primers of the original multiplex mastermix was set. The thermal profile used for the reverse transcription and the amplification is the same used for the multiplex one-step single tube PCR and it is described in the following table (Table 3):

Temperature	Time	Cycles
48°C	30 min	1X
95°C	10 min	1X
95°C	30 sec	45X
60°C	90 sec	

Table 3: Thermal profile for simplex and multiplex amplification of 5'UTR and CORE regions

Once the prototype conditions were assessed, a validation process was required.

4.1.6 Samples

For the prototype validation process a total of 111 positives plasma, 91 positive serum samples, 108 negative plasma and 47 negative serum samples were collected.

The RNA samples were extracted with different manual and automatic systems such as Invisorb Spin universal Kit (Stratec Molecular), QIAamp MinElute Virus Spin kit (QIAGEN), both starting from 200µL and eluted in 60µL, and GENEQUALITY X120 with GQ X120 Pathogen kit (AB ANALITICA) and MagCore HF16 Plus with MagCore Viral Nucleic Acid Extraction Kit (RCB Bioscience) starting from 400µL and eluted in 90µL and 60µL respectively. All the systems were used according to manufacturer instructions (Table 4).

Extraction kit	Automatic extraction platform	Extraction Volume (µL)	Elution Volume (µL)	Negative samples number	Positive samples number	Total
Invisorb® Spin universal Kit (STRATEC MOLECULAR)	-	200	60	12	27	39
QIAamp MinElute Virus Spin kit (QIAGEN)	-	200	50	20	57	77
GENEQUALITY X120 Pathogen kit (AB ANALITICA)	GENEQUALITY X120 (AB ANALITICA)	400	90	66	99	165
MagCore® Viral Nucleic Acid Extraction Kit (Low PCR Inhibition) 202 (RCB BIOSCIENCE)	MagCore HF16 Plus MagCore (DIATECH LABLINE)	400	60	10	19	29

Table 4: Summary of the extraction methods used for validation process

44 negative samples were confirmed with IVD Artus HCV RG RT-PCR Kit (QIAGEN) and 64 negative samples with Abbott RealTime HCV assay (ABBOTT).

The viral load of positive samples was quantified with Bosphore HCV Genotyping Kit v1 (ANATOLIA GENEWORKS) and the results are resumed in the following table (Table 5):

Viral Load Range	Samples number	%
≤ 2.00E+03 IU/mL	2	0.99
2.00E+03 IU/mL< viral load ≤ 1.00E+04 IU/mL	5	2.47
1.00E+04 IU/mL < viral load ≤ 1.00E+06 IU/mL	106	52.47
1.00E+06 IU/mL < viral load ≤ 2.00E+09 IU/mL	51	25.24
Data not available	38	18.81
Total	**202**	**100**

Table 5: Summary of viral load distribution among the samples.

For genotype 6 a spike sample was created using a synthetic RNA diluted in a negative sample.

183 positive sample were genotyped with AMPLIQUALITY HCV TYPE PLUS (AB ANALITICA) according to manufacturer instructions. RT-PCR were performed on SimpliAmp Thermal Cycler (Thermo Fisher Scientific) and visualization protocol was performed on AUTOBLOT 3000H (MedTec).

12 positive samples were Sanger sequenced in the NS5B region and the results are resume in the following table (Table 6):

Genotype	Samples	%
GT1	133	65.84
GT2	40	19.80
GT3	15	7.43
GT4	11	5.45
GT5	2	0.99
GT6	1	0.50
Total	**202**	

Table 6: Summary of genotypes distribution among the samples.

For the kit repeatability and reproducibility, evaluation of 6 known certified plasma samples INSTAND-EQAS positive for HCV 1a, HCV 1b, HCV 2, HCV 3a, HCV 4a, HCV 5a was performed. These plasmas were extracted with GENEQUALITY X120 extraction platform (AB ANALITICA) with GENEQUALITY X120 Pathogen kit (AB ANALITICA). Extracted samples were tested in duplicates 20 times. A total of 240 replicates were obtained. Each run was conducted in different days with different operators.

4.2 HCV NGS assay

For the development of an HCV genotyping device based on the NGS technology, two different library preparation approaches were used: Whole Genome Sequencing and Amplicon Based Sequencing approach.

Below the description of both methods.

4.2.1 Whole Genome Sequencing Approach

4.2.1.1 Samples, RNA extraction, rRNA Depletion

Seven samples were chosen with known high viral load and genotype to test WGS approach.

Viral load was quantified with Bosphore HCV Quantification Kit (ANATOLIA GENEWORKS) following the producer instruction, using an Applied Biosystem 7500 Fast Dx thermocycler (Thermofischer).

Genotyping was performed using AMPLIQUALITY HCV TYPE PLUS kit (AB ANALITICA) and with Abbott RealTime HCV Genotype II (Abbott Park, Illinois, U.S.A.) following the manufacturer instructions. Applied Biosystem SimpliAmp thermocycler, Applied Biosystem 7500 Fast Dx thermocycler (Thermofischer) and AUTOBLOT 3000H (MedTec) were used to perform the protocol.

For Viral RNA extraction QIAamp MinElute virus spin kit (QIAGEN) and Genequality X120 Pathogen kit for GENEQUALITY X120 platform (AB ANALITICA) were used. Clinsamp01704_S6, Clinsamp01792-2_S4, Clinsamp01907_S8, Clinsamp01980-2_S2, Clinsamp02007_S5, Clinsamp02022_S7 and Clinsamp02320 were extracted with QIAamp MinElute virus spin kit (QIAGEN). During the extraction phase 10µL of RNase-Free DNase I (QIAGEN) and 20 µL di DNase Booster Buffer (QIAGEN) were add to the solution after binding the RNA to the spin column. Enzyme mixture was incubated for 15 minutes at room temperature.

Sample RQ04309_S9 was extracted with Genequality X120 Pathogen kit for GENEQUALITY X120 platform (AB ANALITICA) following the manufacturer instructions without DNAse treatment.

Extraction efficiency was evaluated with High Sensitivity D5000 ScreenTape on 2200 TapeStation Nucleic Acid system (AGILENT).

Clinsamp01704_S6, Clinsamp01792dep_S3, Clinsamp01907_S8,

Clinsamp01980dep_S1, Clinsamp02007_S5, Clinsamp02022_S7 and Clin-Samp02320 were treated with NEBNext® rRNA Depletion Kit (NEW ENGLAND BIOLABS) for ribosomal RNA depletion. Ribosomal RNA (rRNA) is the most frequent RNA in Eukaryotes organisms' samples. In order to enrich the sample in viral RNA, the rRNA depletion is a fundamental step before the library preparation. The kit used in the WGS approach, for rRNA depletion eliminate cytoplasmatic rRNA such as 5S, 5,8S, 18S e 28S and the mitochondrial such as 12S and 16S.

The kit was used following manufacturer instructions. In the next table, samples and treatments were summarized (Table 7):

Sample ID	Extraction kit	DNAse treatment	rRNA depletion	Extraction volume (µL)	Elution Volume (µL)	Viral Load (IU/mL)	Expected Genotype
ClinSamp 01704	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	5.75E+05	1a
ClinSamp 01792-2	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	No	200	60	7.75E+05	1b
ClinSamp 01792	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	7.75E+05	1b
ClinSamp 01907	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	4.70E+06	1a
ClinSamp 01980-2	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	No	200	60	3.08E+05	1b
ClinSamp 01980	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	3.08E+05	1b
ClinSamp 02007	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	9.56E+05	1b
ClinSamp 02022	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	4.51E+06	1a
RQ04309	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	No	No	400	90	6.01E+08	1b
ClinSamp 02320	QIAmp MinElute Virus Spin Kit (QIAGEN)	Yes	Yes	200	60	2.00E+06	2k/1b

Table 7: Summary of the used samples for the Whole Genome Sequencing approach experiment.

4.2.1.2 Library preparation

The first step of the library preparation protocol is the retro transcription of the viral RNA in cDNA. For the reverse transcription was used the SuperScript™ IV First-Strand Synthesis System (THERMOFISHER SCIENTIFIC) kit in combination with the random hexamers.

The second strand cDNA was synthesized using NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module (NEW ENGLAND BIOLABS) kit in order to obtain a double strand DNA molecule.

The double strand cDNA was treated with NEBNext Ultra II End Repair/dA-Tailing Module (NEW ENGLAND BIOLABS). With this treatment, blunt ends are created on the 5' end of the cDNA fragment and poli-A were added at 3' end.

Adapter ligation was performed with NEBNext® Ultra™ II Ligation Module (NEW ENGLAND BIOLABS) kit. Illumina Adapters were provided in NEBNext® Multiplex Oligos for Illumina® (NEW ENGLAND BIOLABS) kit.

Sequencing libraries were amplified with KAPA HiFi Hot Start Ready Mix PCR Kit (KAPA BIOSYSTEM) kit. In this step, Illumina indexes are associated to each library fragments to allow the hybridization on the flow cell during the sequencing phase.

4.2.1.3 Purification

During the library preparation protocol several purification steps were performed. The off-target fragments are eliminated with the addition of magnetic beads to the samples, in fact magnetic beads are coated with positive charges that ligate covalently DNA's negative charges and allows the selection of the desired DNA fragments.

On the bases of the rate between amplicons and beads quantity it is possible to eliminate different sizes fragments. Lower is the rate, longer are the fragments ligated by the beads, and vice versa

Purifications steps were introduced after:
- rRNA depletion step
- Double cDNA synthesis
- Ligations step
- Library amplification step

In all the steps, Agencourt AMPure XP PCR Purification kit (BECKMAN COULTER, Inc) were used, following the manufacturer instructions.

The post-ligation steps need a size selection protocol in order to select right size library fragment. The expected fragment size for this protocol was 280 bp.

In the following table, purifying passages are described (Table 8):

Purification step	Agencourt AMPure XP PCR Purification concentration
Post rRNA depletion	2,2X
Post ds-cDNA synthesis	1,8X
Post adapters ligation (>280 bp)	1X
Post adapters ligation (<280 bp)	1X

Table 8: Purification passeges for Whole Genome sequencing approach

4.2.1.4 Library quantification

Library quantification was performed using KAPA Library Quantification Kit Illumina® Platforms (KAPA BIOSYSTEM) Applied Biosystem 7500 Fast Dx Real Time PCR thermocycler (Thermofischer).

Each library, after quantification, is diluted and pooled together to obtain an equimolar final concentration of 12pM to load on the Illumina instrument.

4.2.1.5 Sequencing

Sequencing was performed on Illumina MiSeq platform using MiSeq Reagent Micro Kit v2. This cartridge performs 2x150 cycles (paired end reads), with a throughput of 1,2 Gb (8 million of reads).

4.2.1.6 Bioinformatic Pipeline

Bioinformatic pipeline is a set of bioinformatic tools for NGS data analysis. Bioinformatic analysis is necessary for the interpretation of the sequencing results expressed as raw data from the sequencer. This analysis process was set to obtain information on viral genotype and on drug resistance mutations of the analysed samples. The pipeline was developed on SCI: Luigi. The chosen programming language was Python.

It is an interpreted, high-level, general-purpose programming language.

Once the sequencing run was finished, two files were created for all the paired-end experiments: R1.fastq for the forward sequences and R2.fastq for the reverse sequences.

The information contained in the files are (Fig. 12):

- Read name (label)

- Sequence Raw data (Sequence)

- Single base quality score of all the reads in ASCII language (Q score).

- Reads orientation (Forward or Reverse)

Fig. 12 Read structure in a FastQ file.

During the pre-analytical phase reads are checked for their quality control (QC) parameters with a bioinformatics tool called FastQC v. 0.11.8.

FastQC can show the overall quality of the runs and provide a Quality score for read selection during the next step. The next tool used on the WGS pipeline was CutAdapt that allows the removal of the adapter/primers from the reads.

Base calling accuracy, measured by the Phred quality score (Q score), is the most common metric used to assess the accuracy of a sequencing platform. It indicates the probability that a given base is called incorrectly by the sequencer. On the bases of Illumina quality parameters, reads Q-score has not to be lower than 30.

A Q score of 30 is equal to a probability of incorrect base call of 1 in 1000 bases, that means a base call accuracy of 99.9%. When sequencing quality reaches Q30, all the reads will be almost perfect, having a value near to zero of errors and ambiguities. This is why Q30 is considered a benchmark for quality in NGS technology.

The genotype assignment is made by a comparison between the unknown sample to reference sequences. References sequences are available on the official database (such as GenBank - https://www.ncbi.nlm.nih.gov/genbank/) where are listed known, publicly available nucleotide sequences and certificated genotypes.

The comparison between sequences is made by a process called alignment that is made by dedicated software called aligner.

The aligner used for the data analysis of this experiment is: Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), used for mapping reads on the reference HCV genomes.

Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters.

The Reference sequences used are: NC_004102.1 for genotype HCV 1a, AB016785.1 for genotype HCV 1b, NC_009827.1 for genotype HCV 6, NC_009823.1 for genotype HCV 2, NC_009824.1 for genotype HCV 3 and NC_009825.1 for genotype HCV 4, NC_009826.1 for genotype 5 and NC_030791.1 for genotype 7.

The reads of each analysed sample are compared with every reference sequence to find the best similarity match between the unknown sample and the known genotype sequences.

The similarity parameters considered are the depth and the coverage.

The overall depth value of a reconstructed sequence of each sample, is the medium number of times that a single base of the sample has been sequenced. It depends on the number of reads generated for the samples and it is calculated as the rate between medium reads length and the reference genome length, expressed as nX.

The coverage is the percentage of nucleotide, on the reference genome, covered by reads with a specific depth. The depth value is decided by the operator and depend on the study purpose. For this protocol, coverage was calculated with a decided value of 1X depth.

The aim of the experiment was to genotype samples and also to reveal the presence of RASs. Mutation analysis was performed with VarScan tool (http://varscan.sourceforge.net/). This tool can find the presence of mutation in the sample referring to reference genome sequence, based on parameters decided by the operator. The possibility to define a threshold permits to distinguish PCR artefacts from real mutations.

4.2.2   Amplicon-Based Sequencing approach

4.2.2.1 Primer design
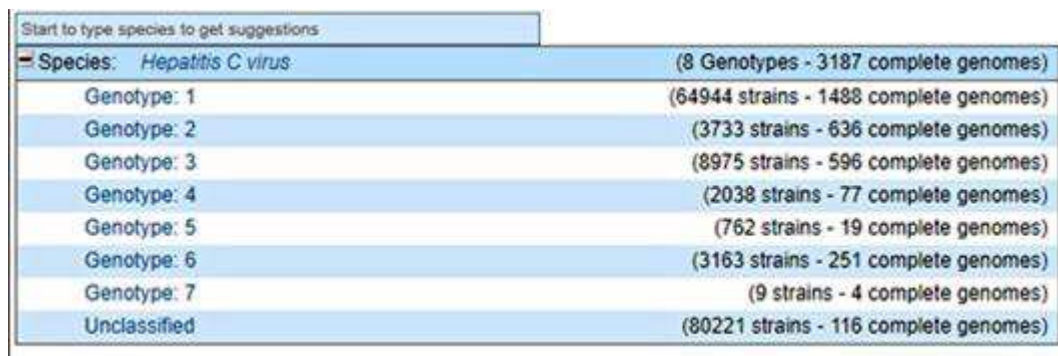
Two different approaches were used for the primers design. Both methods design primers for amplicons with the following characteristics:

- Overlapping pattern to avoid the off-target amplicons creation
- Less than 250 bp length because the chosen cartridge allows a maximum sequencing of 300 bp paired-end
- Target regions: 5'UTR – CORE and NS5B for all genotypes, NS3 and NS5A for GT1 RASs evaluation

Moreover, both methods followed the same step such as:

- Sequences database creation
- Sequences alignment
- Consensus sequence creation
- Annotation based on H77 strain (NC_004102.1)
- Primers design
- Primers validation

The first approach consists in downloading the HCV genotype and subtype reference sequences from ViPR database (https://www.viprbrc.org/brc/home.spg?decorator=flavi_hcv). In Fig. 13 is shown a screenshot of the ViPR HCV sequence list, up to date:

Start to type species to get suggestions	
⊟ Species: *Hepatitis C virus*	(8 Genotypes - 3187 complete genomes)
Genotype: 1	(64944 strains - 1488 complete genomes)
Genotype: 2	(3733 strains - 636 complete genomes)
Genotype: 3	(8975 strains - 596 complete genomes)
Genotype: 4	(2038 strains - 77 complete genomes)
Genotype: 5	(762 strains - 19 complete genomes)
Genotype: 6	(3163 strains - 251 complete genomes)
Genotype: 7	(9 strains - 4 complete genomes)
Unclassified	(80221 strains - 116 complete genomes)

Fig. 13: List of the ViPR HCV sequence list up to date (27th July 2019)

The sequences alignment was performed with BioEdit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). This tool allowed also the creation of consensus sequences (the sequence of most frequent nucleotide of all the sequences considered). From all the genotype and subtype selected was possible to identify conserved region for the primer design. The *in silico* validation

of the software was performed with a on line tool: RealTime Design qPCR Assay Design Software (Biosearch technologies - https://www.biosearchtech.com/support/tools/design-software/realtimedesign-software).

For primer designing the following requirements needed to be met:
- Annealing temp between 60°C and 62°C
- GC% around 55%
- Self-dimer and secondary structure creation avoiding
- Maximum of 3 degenerate bases

180 primers were designed with this approach. The first primer test showed a lack of amplification for some HCV region type probably due to the extreme variability inter-, intra-genotypes and subtypes.

A different approach was developed in order to amplify all the HCV region target. ICTV (https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification) is a database of HCV sequences regularly updated, ICTV provides a list of confirmed HCV genotype and subtype assignments and alignments (in FASTA and SSE formats). Its proposal was supported by the establishment of well-finished databases such as Los Alamos HCV Sequence Database (https://hcv.lanl.gov/content/index), a central website that provides annotated sequences and analysis tools for HCV.

All the reference sequences downloaded were used for a nucleotide sequence alignment with a tool called MAFFT (https://mafft.cbrc.jp/alignment/server/). The alignment obtained was the input file for Unipro UGENE tool (http://ugene.net/) that, with set criteria can build the consensus sequence. This step was reiterating several times for all the genotype/subtype considered.

UGENE allowed to set a threshold in percentage for the consensus generation; for the primer designed in this project, the threshold was set between 85% and 95%. This value was useful to exclude low expressed mutation in the aligned sequences group. For example: if threshold is set at 90%, mutations in a position are considered if the 90% of the sequence aligned in the group have that mutation.

The consensus sequence was used for the primer design with Gemi – primer design 1.3.2 tool. The criteria used were the same used

with the first approach:

- Amplicons length around 270 bp

- Annealing temperature between 62°C and 65°C.

- GC % around 55%

- Maximum of 3 degenerate bases

350 primers were designed with the second approach.

All the primers from the two approaches were tested both *in silico* and *in vitro*, and the most performing ones were selected and divided into 4 different pools. For the *in silico* validation were used FastPCR (https://primerdigital.com/fastpcr.html) and SimulatePCR (https://sourceforge.net/projects/simulatepcr/) tools. FastPCR is an integrated tool for PCR primers or probe design, performing *in silico* PCR; it tests oligonucleotide assembly, alignment and repeat searching. Simulate PCR can assess primer specificity and predict both desired and off-target amplification products using an HCV reference database provided by the user. Similar primers in sequence or interacting primers (G quadruplex, primer dimers) *in silico*, were exclude from the pools. The final pools contain 69, 47, 45 and 76 primers.

All the processes, from the sequence alignment to the primer validation were made for all genotype and subtype considered.

4.2.2.2 HCV NGS software

Sequencing raw data has to be analysed by a bioinformatics pipeline to obtain the final result.

HCV NGS kit pipeline is developed in Python together with Sci-Luigi, a framework for building complex pipelines. The project is running on Ubuntu, since most of the bioinformatics tools are Linux based.

In the first quality control passage, FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) check the quality of the raw reads on the fastq files and filters them. All the following analysis passages are made only on the quality filtered reads.

The reads that pass the quality control are trimmed with CutAdapt (https://cutadapt.readthedocs.io/en/stable/guide.html) to remove the low-quality bases, typically at the extremity of the read. Also, primers and artifacts are eliminated.

Reads alignment is performed with Bowtie 2 (http://bowtie-

bio.sourceforge.net/bowtie2/index.shtml). Alignments were performed using all the genome sequences of the ICTV database (https://talk.ictvonline.org/).

The reads alignment produces a SAM file. SAM files are converted in BAM files using SAMtools (http://samtools.sourceforge.net/). Moreover, indexing convert BAM into a BAI file, which gives quick access to any region of the reference.

Only primary reads and high-quality alignment reads are used for the variant calling.

Since analysis pipeline is not optimized for variant calling analysis yet and there are not reference methods for RASs evaluation, sequencing results are analysed with Geno2Pheno online tool (https://www.geno2pheno.org/). The RASs considered were the ones indicated in EASL guidelines only for GT1 samples (Pawlotsky et al. 2018).

4.2.3   Prototype pre-validation and Validation

4.2.3.1   Viral RNA extraction, quantification and genotyping

A diagnostic assay needs to be tested with real samples to evaluate performances and compliance with the declared intended use, before selling it. Prototype pre-validation process is an internal procedure useful to decide, with a small number of samples, if the prototype needs more adjustments or if it is ready for the commercialization. If the results of the prototype pre-validation process comply all the expected ones, verified prototype passes on the validation process.

For the kit prototype pre-validation process, 16 positive plasma samples were extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) with GENEQUALITY X120 workstation (AB ANALITICA). 10 samples were extracted with QIAamp MinElute virus spin kit (QIAGEN). Moreover, 9 samples from External Quality Control panel (QCMD, www.qcmd.org) for HCV genotyping assay were extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) with GENEQUALITY X120 workstation (AB ANALITICA). QCMD is an independent International External Quality Assessment Proficiency Testing organisation, providing a quality assessment service primarily focused on molecular infectious diseases (www.qcmd.org). QCMD evaluates the ability of an assay to detect the correct genotype / RASs for a given set of samples. In addition, two simulations of co-infection were created by mixing different genotypes from two different extracted RNA (sample RQ05942 and sample RQ05952) with different proportions of the virus (70% and 30% and viceversa). All the samples were sequenced with Illumina MiSeq platform and a subset of them were also sequenced with Illumina iSeq100™ (Illumina) platform for a preliminary evaluation of the consistency of data obtained with both platforms.

In the following table is described the resume of the samples results tested for the prototype pre-validation process (Table 9):

Sample ID	Clinical Sample ID	Extraction system	Extraction volume	Elution Volume	Viral load (IU/mL)	Expected Genotype	Sequencing platform
RQ05836	Clin-Samp 02631	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	1.28E+05	1a	Illumina MiSeq
RQ05830	Clin-Samp 02624	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.75E+05	4d	Illumina MiSeq
RQ05816	Clin-Samp 02636	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.02E+05	2	Illumina MiSeq
RQ05838	Clin-Samp 02633	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.86E+04	1a	Illumina MiSeq
RQ05788	Clin-Samp 02575	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.85E+05	1b	Illumina MiSeq
QCMD 2017 - 8	QCMD 2017 - 8	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.84E+03	5a	Illumina MiSeq
QCMD 2017 - 3	QCMD 2017 - 3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.00E+03	6a	Illumina MiSeq
RQ05819	Clin-Samp 02725	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.60E+05	3a	Illumina MiSeq
RQ05792	Clin-Samp 02579	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.02E+05	1b	Illumina MiSeq
RQ05832	Clin-Samp 02626	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	6.33E+05	4d	Illumina MiSeq
RQ05840	Clin-Samp 02635	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	9.71E+05	1a	Illumina MiSeq
RQ05787	Clin-Samp 02574	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	5.35E+03	1b	Illumina MiSeq
RQ05824	Clin-Samp 02618	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	5.82E+05	2	Illumina MiSeq
RQ05844	Clin-Samp 02639	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	1.60E+05	3a	Illumina MiSeq
QCMDDR2019-1	QCMDDR2019-1	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.32E+03	NA	Illumina iSeq™100 / Illumina MiSeq
QCMDDR2019-2	QCMDDR2019-2	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.90E+03	NA	Illumina iSeq™100 / Illumina MiSeq
QCMDDR2019-3	QCMDDR2019-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.17E+02	NA	Illumina iSeq™100 / Illumina MiSeq
QCMDDR2019-4	QCMDDR2019-4	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.70E+03	NA	Illumina iSeq™100 / Illumina MiSeq
QCMDDR2019-5	QCMDDR2019-5	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.60E+03	NA	Illumina iSeq™100 / Illumina MiSeq
RQ05942	Clin-Samp 02730	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.86E+06	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05943	Clin-Samp 02731	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.61E+05	4	Illumina iSeq™100 / Illumina MiSeq
RQ05944	Clin-Samp 02732	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.64E+05	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05945	Clin-Samp 02733	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.07E+05	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05946	Clin-Samp 02734	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.82E+05	3a	Illumina iSeq™100 / Illumina MiSeq
RQ05947	Clin-Samp 02735	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.37E+04	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05948	Clin-Samp 02736	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.32E+04	1a	Illumina iSeq™100 / Illumina MiSeq

Sample ID	Clinical Sample ID	Extraction system	Extraction volume	Elution Volume	Viral load (IU/mL)	Expected Genotype	Sequencing platform
RQ05949	Clin-Samp 02737	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.49E+06	2	Illumina iSeq™100 / Illumina MiSeq
RQ05950	Clin-Samp 02738	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.04E+07	2	Illumina iSeq™100 / Illumina MiSeq
RQ05951	Clin-Samp 02739	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.70E+04	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05952	Clin-Samp 02740	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.05E+06	4	Illumina iSeq™100 / Illumina MiSeq
RQ05953	Clin-Samp 02741	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	8.68E+05	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05955	Clin-Samp 02744	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	5.08E+06	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05959	Clin-Samp 02755	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.41E+03	1b	Illumina iSeq™100 / Illumina MiSeq
QCMD2019-5	QCMD2019-5	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.03E+03	5a	Illumina iSeq™100 / Illumina MiSeq
QCMD2019-8	QCMD2019-8	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.40E+05	6	Illumina iSeq™100 / Illumina MiSeq
RQ05942+RQ05952 (Coinf 70%+30%)	Clin-Samp02730 + Clin-Samp02740	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	70%+30%	1a-4	Illumina iSeq™100 / Illumina MiSeq
RQ05952+ RQ05942 (Coinf 70%+30%)	Clin-Samp02730 + Clin-Samp02740	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	70%+30%	1a-4	Illumina iSeq™100 / Illumina MiSeq

Table 9: Summary of samples used for the Prototype pre-validation process of HCV NGS kit

Validation process is necessary to define the kit performance on sensitivity and specificity. To define the kit's performances on sensitivity for genotype evaluation, 4 positive clinical samples were extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) with GENEQUALITY X120 workstation (AB ANALITICA). 7 clinical samples were extracted with QIAamp MinElute Virus Spin kit (QIAGEN). In order to evaluate differences between extractions kits, 3 plasma samples, were extracted with both kits. Details are explained on Table 10. Moreover, 6 samples from External Quality Control panel (QCMD, www.qcmd.org and Instand-EQAS www.instand-ev.de/en/eqas.html) for HCV genotyping assay were extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) with GENEQUALITY X120 workstation (AB ANALITICA).

In order to calculate the Limit of Detection (LoD) of the assay samples with different genotype were chosen. Extracted RNA of each sample were diluted in negative RNA to reach final concentration of 10^4 IU/mL, 10^3 IU/mL and 10^2

IU/mL for each genotype. Together with the evaluation of limit of detection of the genotype, for GT1 samples also RASs LoD were evaluated. For this purpose, QCMDDR 2019 panel GT1 samples already analysed in prototype pre-validation run, and QCMDDR 2018 panel GT1 samples were chosen.

Also, a preliminary reproducibility study was performed both for genotype and for RASs evaluation. From diagnostic sensitivity run, 3 GT1 samples were chosen and repeat in the Limit of Detection evaluation run. Chosen samples were RQ05954, RQ05972 and RQ05973. They were processed with the same protocol used for the diagnostic sensitivity library preparation to evaluate the presence of inter-run differences.

To define the diagnostic specificity 10 negative samples were processed. 3 of them were extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) and 7 with QIAamp MinElute Virus Spin kit (QIAGEN). 4 out of 10 are samples from External Quality Control panel (QCMD, www.qcmd.org). As reference positive control was used the same Instand-EQAS sample (/www.instand-ev.de/en/eqas.html) analysed in the diagnostic sensitivity run.

The viral load of the positive samples was evaluated with Bosphore HCV Quantification Kit (Anatolia Genework) using the manufacturer instruction on Applied Biosystem 7500 Fast Dx thermocycler (Thermofischer). Genotyping analysis was performed on the unknown clinical samples extracted with GENEQUALITY X120 Pathogen kit (AB ANALITICA) with GENEQUALITY X120 workstation (AB ANALITICA) and with QIAamp MinElute Virus Spin Kit (QIAGEN) were genotyped using AMPLIQUALITY HCV TYPE PLUS (AB ANALITICA) according to the user manual.

All the samples were sequenced with Illumina iSeq100™ (Illumina). Moreover, diagnostic sensitivity library was re-run on MiSeq platform to verify sequencing performances of library preparation protocol on both platforms.

In the following table is described the resume of the samples results tested for validation process (Table 10):

Diagnostic sensitivity

Sample ID	Clinical Sample ID	Extraction system	Extraction volume	Elution Volume	Viral load (IU/mL)	Expected Genotype	Sequencing platform
RQ05954	Clin-Samp 02743	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.22E+06	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05956	Clin-Samp 02745	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.87E+07	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05957	Clin-Samp 02746	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	9.50E+07	2	Illumina iSeq™100 / Illumina MiSeq
RQ05958	Clin-Samp 02749	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.76E+04	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05960	Clin-Samp 02761	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	8.70E+04	3a	Illumina iSeq™100 / Illumina MiSeq
RQ05961	Clin-Samp 02767	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.05E+04	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05968	Clin-Samp 02736	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	3.32E+04	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05969	Clin-Samp 02737	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	3.49E+06	2	Illumina iSeq™100 / Illumina MiSeq
RQ05970	Clin-Samp 02739	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.70E+04	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05971	Clin-Samp 02740	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.05E+06	4	Illumina iSeq™100 / Illumina MiSeq
RQ05972	Clin-Samp 02741	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	8.68E+05	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05973	Clin-Samp 02744	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	5.08E+06	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05974	Clin-Samp 02745	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	1.87E+07	1a	Illumina iSeq™100 / Illumina MiSeq
RQ05975	Clin-Samp 02746	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	9.50E+07	2	Illumina iSeq™100 / Illumina MiSeq
RQ05976	Clin-Samp 02755	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	4.41E+03	1b	Illumina iSeq™100 / Illumina MiSeq
RQ05977	Clin-Samp 02761	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	8.77E+03	3a	Illumina iSeq™100 / Illumina MiSeq
QCMD2016-5	QCMD 2016-5	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	N/A	5a	Illumina iSeq™100 / Illumina MiSeq
QCMD2016-7	QCMD 2016-7	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	N/A	5a	Illumina iSeq™100 / Illumina MiSeq
QCMD2018-3	QCMD 2018-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	5.09E+03	6a	Illumina iSeq™100 / Illumina MiSeq
QCMD2017-2	QCMD 2017-2	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.37E+02	2b	Illumina iSeq™100 / Illumina MiSeq
QCMD2017-4	QCMD 2017-4	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.26E+03	3a	Illumina iSeq™100 / Illumina MiSeq
RQ05982	Clin- Samp 02320	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	2.00E+06	2k/1b	Illumina iSeq™100 / Illumina MiSeq
Instand CTRL	Instand CTRL	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	N/A	1b	Illumina iSeq™100 / Illumina MiSeq

Sample ID	Clinical Sample ID	Extraction system	Extraction volume	Elution Volume	Viral load (IU/mL)	Expected Genotype	Sequencing platform
Diagnostic Specificity							
RQ04501	GLPN328	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
RQ04500	GLPN329	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
RQ04502	GLPN330	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
RQ04598	GLPN331	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
RQ04599	GLPN332	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
RQ04605	GLPN317	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
QCMD2019-4	QCMD 2019-4	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	undet	Neg	-
QCMD2018-6	QCMD 2018-6	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	undet	Neg	-
QCMD2017-6	QCMD 2017-6	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	undet	Neg	-
QCMD2016-2	QCMD 2016-2	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	undet	Neg	-
Instand CTRL	Instand CTRL	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	N/A	1b	-
Limit of Detection and repeatability							
QCMDDR 2019-1	QCMDDR 2019-1	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.32E+04	1a	Illumina iSeq™100
QCMDDR 2019-1 1:10	QCMDDR 2019-1	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.32E+03	1a	Illumina iSeq™100
QCMDDR 2019-1 1:100	QCMDDR 2019-1	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.32E+02	1a	Illumina iSeq™100
QCMDDR 2019-3	QCMDDR 2019-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.18E+02	1b	Illumina iSeq™100
QCMDDR 2019-5	QCMDDR 2019-5	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.66E+03	1a	Illumina iSeq™100
QCMDDR 2019-5 1:10	QCMDDR 2019-5	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	3.66E+02	1a	Illumina iSeq™100
QCMDDR 2018-3	QCMDDR 2018-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.30E+04	1a	Illumina iSeq™100
QCMDDR 2018-3 1:10	QCMDDR 2018-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.30E+03	1a	Illumina iSeq™100
QCMDDR 2018-3 1:100	QCMDDR 2018-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	1.30E+02	1a	Illumina iSeq™100
RQ05954	Clin-Samp 02743	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	4.22E+06	1a	Illumina iSeq™100
RQ05972	Clin-Samp 02741	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	8.68E+05	1a	Illumina iSeq™100

Sample ID	Clinical Sample ID	Extraction system	Extraction volume	Elution Volume	Viral load (IU/mL)	Expected Genotype	Sequencing platform
RQ05973	Clin-Samp 02744	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	5.08E+06	1b	Illumina iSeq™100
RQ05969 1:100	Clin-Samp 02737	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	3.49E+04	2	Illumina iSeq™100
RQ05969 1:1000	Clin-Samp 02737	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	3.49E+03	2	Illumina iSeq™100
RQ05960 1:10	Clin-Samp 02761	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	8.78E+03	3a	Illumina iSeq™100
RQ05960 1:100	Clin-Samp 02761	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	8.77E+02	3a	Illumina iSeq™100
RQ05971 1:1000	Clin-Samp 02740	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.05E+03	4	Illumina iSeq™100
RQ05971 1:10000	Clin-Samp 02740	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	2.05E+02	4	Illumina iSeq™100
QCMD2016-7 1:10	QCMD 2016-7	QIAmp MinElute Virus Spin Kit (QIAGEN)	200	60	5.00E+02	5a	Illumina iSeq™100
QCMD2018-3 1:10	QCMD 2018-3	GENEQUALITY X120 PATHOGEN Kit (AB ANALITICA)	400	90	5.09E+02	6a	Illumina iSeq™100

Table 10: Summary of samples used for the validation process of HCV NGS kit.

4.2.3.2 Library preparation

Library preparation for amplicon-based approach is divided into different passages described in the figure below (Fig. 14):
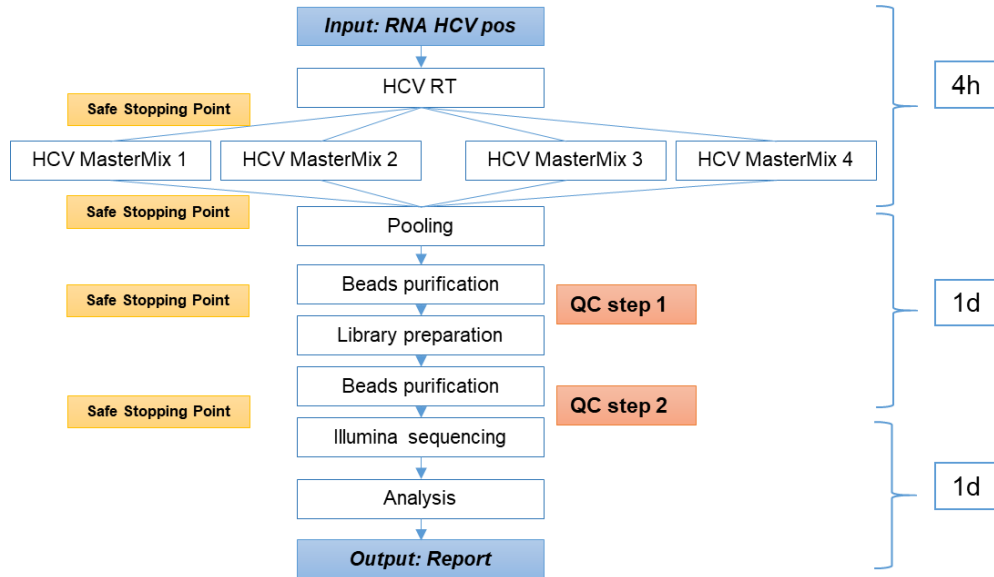
Fig. 14: Library prep workflow

Reverse transcription process was made with a ready-to-use mix containing random hexamers and oligo-dT primers, dNTPs, Murine RNase Inhibitor, and Reverse Transcriptase. 7 µL of sample's RNA was added.

The following thermal protocol was applied using SimpliAmp thermocycler (Applied Biosystem) (Table 11).

Step	Repeats	Time	°C
1	1	02:00	25,0
2	1	20:00	55,0
3	1	01:00	95,0
4	1	∞	4,0

Table 11: Thermal protocol for reverse transcription.

For the amplification of the target regions of the genome, an amplification with specific primers was set up. In order to limit the formation of off-target PCR products, four different ready-to-use mastermixes with four different sets of primers were created. Primers design and pools creation were previously discussed. The chosen mastermix for the amplification contains a High-Fidelity DNA polymerase.

Each cDNA synthesised in the reverse transcription reaction, was amplified with the 4 different mastermixes using 2µL. All 4 amplification reactions used the same thermal profile, so they can be performed in a single amplification run. The thermal profile used is described in the following table (Table 12):

Step	Repeats	Time	°C
1	1	02:00	98,0
2		00:10	98,0
3	40	00:15	62,0
4		00:30	72,0
5	1	02:00	72,0
6	1	∞	4,0

Table 12: Thermal protocol for the target amplification.

At the end of the amplification process, for each sample, amplicons derived from the 4 different PCR were pooled together in a single tube. The pooled amplicons were purified to eliminate dimers, primer residual and off-target amplification products.

The purification process was performed with MagBeads kit, using a 1.35X proportion of magnetic beads following the manufacturer instructions. Purification step allows to eliminate all the primer dimers.

After purification the first Quality Control (QC) step was performed. The purified pooled amplification products were visualized with a 2,5% agarose gel, using 5µL of each purified product. The expected length of the amplification products was 250-300 bp.

End-repair reaction was needed to transform overhang ends into blunt ends.

It was performed using the following thermal profile (Table 13).

Step	Repeats	Time	°C
1	1	30:00	20,0
2	1	30:00	65,0
3	1	∞	4,0

Table 13: thermal protocol for End Repair reaction.

For each reaction, 20µL of purified product was added to the End Repair Reaction.

The next step was the ligation of the Illumina adaptors to the end repaired amplicons. The following incubation was used (Table 14):

Step	Repeats	Time	°C
1	1	15:00	20,0

Table 14: thermal protocol for adaptor ligation.

The adaptors are hairpin-shaped. At the end of the incubation, an enzyme was added in order to cleave the adaptors hairpin with the following incubation step (Table 15):

Step	Repeats	Time	°C
1	1	15:00	37,0

Table 15: thermal protocol for adaptors' cleavage.

After adaptors ligations, size selection purification passages were performed in order to eliminate all the off-target products of the adapters' ligation reaction. MagBeads kit was used following the manufacturer instructions. Two purification steps were performed: 0.60X and 0.35X volume of beads.

To the purified final product, Illumina Unique Dual Indexes (UDI) were added using the following thermal profile (Table 16):

Step	Repeats	Time	°C
1	1	03:00	98,0
2	12	00:15	98,0
3		01:15	65,0
4	1	05:00	65,0
5	1	∞	4,0

Table 16: Thermal profile for library amplification.

UDI are oligonucleotides made by different (unique) combinations of 8 bases that identify unequivocally all the different amplicons or fragments derived from a single sample. Amplification reaction was set by adding to each library sample at 5' and 3' of the amplicons a specific UDI and allowing, at the end, the multiplexing of all the libraries.

The libraries were purified with magnetic beads kit using a 1X beads volume.

Before the sample's library pooling, each library was quantified with Qubit dsDNA HS (High Sensitivity) Assay Kit (ThermoFisher Scientific).

4.2.3.3 Sequencing

Sample sequencing was performed in two different Illumina platforms: MiSeq and iSeq100™ (Illumina).

For the MiSeq was used the MiSeq Reagent Kit v2 Micro (300-cycles), for iSeq100™ (Illumina) was used iSeq100™ (Illumina) i1

Reagent, with the paired end sequencing option.

For MiSeq platforms, 10pM final library concentration was load, for the iSeq 100 50pM was loaded. Libraries multiplexing and loading were conducted following Illumina instructions for both platforms.

5. RESULTS

5.1 HCV genotyping assay based on Reverse Line Blot technology

5.1.1 Development

The Agenzia Italiana del Farmaco (AIFA) in 2011 adapted its guidelines for HCV diagnosis to European Medicine Analysis Guideline on clinical evaluation of medicinal products for the treatment of chronic hepatitis C in which is reported that "*Techniques based on the analysis of the 5' non coding region are not recommended, as a too high incidence of erroneous determination of the subtype has been reported.*"

Chevaliez et al. 2009 reported in fact that a single region cannot distinguish different GT1 subtype and GT6, for the presence of polymorphism in common in the 5'UTR region sequence. The same paper highlights the ability of CORE region in subtyping GT1 and distinguish them from GT6 samples.

Biofield Innovation already had a HCV genotyping assay based on Reverse Line Blot technology which used 5'UTR and CORE regions analysis to better defined all GT1 subtype.

Nevertheless, an internal study revealed a limitation in subtyping GT1 samples. The undeterminable results for subtype was due to the missing or weak amplification of CORE region in some GT1 samples. For this lack of subtyping performances, it was decided to improve the device with the addition of new CORE region amplification primers.

In the prototype assessment study, subject of this study, an increase of performance was observed.

5.1.2 Validation

In this work diagnostic specificity and sensitivity, reproducibility and repeatability tests are described. Cross reactivity test was repeated with the redesigned amplification mastermix. Interfering substances tests and clinical sensitivity weren't repeated in respect to the previous assay.

5.1.2.1 Diagnostic sensitivity and specificity

202 positive samples and 108 negative samples were analysed

with new HCV genotyping assay prototype for diagnostic sensitivity test. Results are described in the following table (Table 17):

Genotype	Expected results (NS5B sequencing)	Obtained results (New HCV genotyping assay)
1	2	6
2	34	34
4	1	1
1a	48	47
1a+3	1	1
1b	83	80
2a/2c	6	6
3a	14	14
3h	1	1
4a	2	2
4a/b/c/d/f	6	6
4d	2	2
5a	2	2
6a	1	1
neg	108	108

Table 17: Resume of the obtain results with new HCV genotyping assay for Diagnostic sensitivity and specificity.

In positive samples group, 6 non-subtyped GT1 samples analysed during internal evaluation, were added to the study. These sample were subtyped with Sanger sequencing of the NS5B region. The sequencing results and the phylogenetic analysis identify the subtypes as GT1a (1 sample) and GT 1b (3 samples). 2 samples were not-subtyped because the viral load was below the LoD of the device (<2000 IU/mL). Further tests with previous HCV genotyping assay of the 2 samples shows the same results of non-subtyping. But, a weak amplification of 5'UTR region permitted to identify the genotypes (but not the subtyping). The same results were confirmed with the new version of the kit.

Regarding the GT2 and GT4 samples, the subtype was not indicated because of a device limitation. In fact, the assay was develop to detect only the subtyping of the most spread subtypes. However, since the subtypes for non-GT1 samples are not critical for the therapy decision, no further analysis were conducted on these samples. All the other samples analysis confirmed the same result of the previous version. To calculate the new HCV genotyping assay specificity the following formula was applied:

$$\text{DIAGNOSTIC SPECIFICITY: } \frac{\text{observed negative samples}}{\text{true negative+false positive}} \times 100$$

$$\text{DIAGNOSTIC SPECIFICITY: } \frac{108}{108+0} \times 100 = 100\%$$

Diagnostic specificity, or true negative rate, results as 100%.

Diagnostic specificity, or true negative rate, results as 100%.

Diagnostic sensitivity was divided into genotype level and subtype level. To calculate the new HCV genotyping assay sensitivity at the genotype level the following formula was applied:

$$\text{DIAGNOSTIC SENSITIVITY (GT level):} \frac{\text{positive and genotyped samples observed}}{\text{true positive with the right genotype+false negative or false genotype}} \times 100$$

$$\text{DIAGNOSTIC SENSITIVITY (GT level):} \frac{201}{201+0} \times 100 = 100\%$$

Diagnostic sensitivity at genotype level resulted as 100%.

As previously discussed, for 4 of the 201 samples the genotype was correctly evaluated but the subtype was not. For GT1 samples the subtype identification is made by the CORE region. Amplification of the CORE region is performed in the same reaction tube of the 5'UTR region.

For GT1 and non-GT6a and 6b samples 3 specific probe for CORE region amplicons were design. In particular, for subtype 1a and subtype 1b two specific probes were designed to allow the distinction between the different subtypes. For the not-subtyped GT1 samples a Sanger sequencing of the CORE region was performed. It resulted that there were mismatches between amplicons and specific GT1 subtype probe, that didn't allow the hybridization of the amplicons.

Diagnostic sensitivity at subtype level was calculate applying the following formula, considering as "not subtyped" only GT1 samples, because of the diagnostic significance of the GT1 samples subtype.

$$\text{DIAGNOSTIC SENSITIVITY (subtype level):} \frac{\text{positive and subtyped samples observed}}{\text{true positive with the right subtype+false negative or false subtype}} \times 100$$

$$\text{DIAGNOSTIC SENSITIVITY (subtype level):} \frac{197}{197+4} \times 100 = 98\%$$

5.1.2.2 Repeatability and reproducibility

For the kit repeatability and reproducibility evaluation, 6 known certified plasma samples INSTAND-EQAS positive for GT1a, GT1b, GT2, GT3a, GT4a, GT5a were used. Each sample was repeated in double in each run for a 20 consecutive runs. In the following table (table number)

are summarized the obtained results (Table 18):

Expected result	Total number of replicates	Suitable replicate observed	Percentage of suitable replicates (%)
GT1a	40	40	100 (40/40)
GT1b	40	40	100 (40/40)
GT2	40	40	100 (40/40)
GT3a	40	40	100 (40/40)
GT4a	40	40	100 (40/40)
GT5a	40	40	100 (40/40)
Total	240	240	100 (240/240)

Table 18: Summary of repeatability and reproducibility test

For each sample a total of 40 repetitions were performed. Each repetition, intra and inter run gave the expected result. Prototype repeatability (intra-run test) and reproducibility (inter-run test) are 100%.

5.1.2.3 Analytical specificity

The specificity of the new primers designed for CORE region was *in silico*-tested by using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and *in vitro*-tested by using samples positive for other viruses.

The *in silico* analytical specificity of the new HCV genotyping assay is guaranteed by accurate and specific selection of primers and probes, and by the use of stringent amplification conditions. Alignment of the primer and probe sequences with data sets of the most important databases did not reveal any non-specific pairing.

In vitro analytical specificity was tested with *Epstein-Barr virus* (EBV), *Herpes simplex virus 1* (HSV-1), *Herpes simplex virus 2* (HSV-2), *Cytomegalovirus* (CMV), *Varicella-zoster virus* (VZV), *Parvovirus B19*, Adenovirus (ADV), *BK virus* (BKV), *Enterovirus*, *Human herpes virus type 6* (HHV 6), *Human herpes virus type 8* (HHV 8) and *JC virus* (JCV) positive samples. All the samples resulted negative.

Both tests results highlight a low possibility of cross reaction with other pathogens.

5.2 HCV NGS assay

5.2.1   Whole Genome Sequencing approach

In the following table are summarized the obtained results from the experiment made using the WGS approach (Table 19):

Reference Sequence		Clinsamp01704_S6	Clinsamp01792-2_S4	Clinsamp01792dep_S3	Clinsamp01907_S8	Clinsamp01980-2_S2	Clinsamp01980dep_S1	Clinsamp02007_S5	Clinsamp02022_S7	RQ04309_S9	Clin-Samp02320
	Known genotype	1a	1b	1b	1a	1b	1b	1b	1a	1b	2k/1b
	Aligned reads	81123	6633	5973	82737	9124	45154	13360	228887	10663	N/A
	Aligned reads %	9,33	0,89	0,48	7,04	1,39	18,98	1,29	22,67	0,98	N/A
	Mapped\Filtered bases	2332	226	1012	15072	365	1378	1030	16367	533	8508
	Mapped\Filtered bases %	0,27	0,03	0,08	1,28	0,06	0,58	0,31	1,62	0,05	0.46
NC_004102.1 Genotype 1a	Depth	33,05	0,55	0,12	224,24	0,75	0,17	0,15	235,61	1,16	-
	Coverage	95,20	10,95	5,71	97,89	9,90	9,18	6,01	98,14	9,02	-
AB016785.1 Genotype 1b	Depth	1,07	0,69	14,67	0,27	0,37	19,96	15,02	0,68	1,26	2662
	Coverage	15,62	38,29	92,15	7,22	22,59	92,89	91,18	11,84	53,12	100
NC_009827.1 Genotype 6	Depth	0,01	0,08	0,05	0,05	0,14	0,03	-	0,07	0,14	-
	Coverage	0,31	0,77	3,56	2,05	0,78	1,04	-	2,25	0,38	-
NC_009823.1 Genotype 2	Depth	-	0,04	105,78	-	0,13	0,01	0,02	0.09	0,12	666
	Coverage	-	0,65	1,70	-	0,92	0,25	0,65	5,03	0,79	84.63
NC_009824.1 Genotype 3	Depth	0,07	0,02	0,14	0,06	0,05	0,07	0,03	0,10	0,05	-
	Coverage	1,34	1,16	2,13	3,16	0,33	3,22	0,53	2,91	0,34	-
NC_009825.1 Genotype 4	Depth	-	-	0,04	0,37	-	-	0,10	-	-	-
	Coverage	-	-	2,41	6,85	-	-	3,92	-	-	-
NC_009826.1 Genotype 5	Depth	0,02	-	-	0,11	-	0,02	-	0,24	-	-
	Coverage	1,51	-	-	4,92	-	1,16	-	6,45	-	-
NC_030791.1 Genotype 7	Depth	-	0,02	0,04	0,07	0,05	0,16	-	0,04	0,06	-
	Coverage	-	0,36	2,41	3,73	0,34	2,01	-	1,19	0,42	-

Table 19: Summary of the results of the WGS approach experiment. In the top row is reported the clinical sample's name, in the first column on the left, are listed the accession number of the used reference genome.

From the obtained results it's clear that rRNA depleted samples (Clinsamp01704_S6, Clinsamp01792dep_S3, Clinsamp01907_S8, Clinsamp01980dep_S1, Clinsamp02007_S5, Clinsamp02022_S7, Clin-Samp02320) have a higher mapped reads quantity than the other that were only treated with DNase (Clinsamp01792-2_S4, Clinsamp01980-2_S2).

Sample RQ04309_S9, that was no treated at all, had the same mapped reads results of the ones treated with DNase only. With both alignment tools samples with higher viral load (ClinSamp01907_S8 and ClinSamp02022_S7, Clin-Samp02320) have a higher number of mapped reads.

It's possible to assign the correct genotype comparing depth and coverage values obtained with all the reference genomes. Samples Clinsamp01792-2_S4 and Clinsamp01980-2_S2 in respect with other ones, have low coverage and depth values and the genotype assignment is uncertain.

ClinSampl02320 was previously analysed with AMPLIQUALITY HCV TYPE PLUS (AB ANALITICA) and with Abbott RealTime HCV Genotype II (Abbott Park, Illinois, U.S.A.). The analysis with Abbott RealTime HCV Genotype II (Abbott Park, Illinois, U.S.A.) that uses 5'UTR and NS5B regions to define the genotype highlights the possibility of the presence of a recombinant form, since 5'UTR region for both devices confirms the presence of GT2, but NS5B the presence of GT1b.

NGS WGS analysis confirmed the sample to be a Recombinant Form 2k/1b. The phylogenetic analysis shows that the genotype of the is closest to the reference sequence of 2k/1b recombinant genotype. The sequence coverage obtained was 99,3% with a mean depth of 103X.

The newly sequenced genome of the sample was submitted in GenBank database (reference number: MK039720)

Due to a not enough high depth values of all the samples it wasn't possible to make the variant calling analysis for the evaluation of the presence of mutations in the samples' genome.

The depth value that this approach can reach is too low for the intended use that the final commercial product has to obtain. In fact, the RASs detection in GT1 samples is one of the two main purpose of the

assay.

However, even if it could be possible to increase the efficiency of this approach-based assay, a second commercial reason turn the WGS approach less appealing. In fact, as it was described the library preparation protocol is time consuming for the operator and contamination-prone.

It was decided to develop the second approach for the library preparation that is the amplicon based one.

5.2.2   Amplicon Based Sequencing approach

5.2.2.1 Prototype pre-validation

For the prototype pre-validation process of HCV NGS kit, 35 positives samples and 2 artificial coinfections were analysed.

A total of two MiSeq (Illumina) and one iSeq100™ (Illumina) platforms runs were performed. In the first MiSeq sequencing run (samples from RQ05836 to RQ05844 in Table 20) a cluster density of 893 K/mm^2, 98.1% reads with >Q30 quality and a 93.4% cluster passing filter were obtained.

For the iSeq100™ (Illumina) the run parameters obtained were: 40% Occupancy, 93.5% reads1 with a >Q30 quality and 90% of reads2 with a >Q30 quality and 28.9% clusters passing filter.

In the second MiSeq sequencing run, a cluster density of 957 K/mm^2, 97,8% reads with a >Q30 quality and a 92,2% cluster passing filter were obtained.

Cluster density is an important factor in optimizing data quality and optimal reads number. The recommended raw cluster densities for balanced libraries in a MiSeq run is 850-1200 K/mm^2. The obtained values for both MiSeq sequencing runs were inside the expected optimum range. For iSeq100™ (Illumina) platform the cluster density value is not informative because of technology used in the flow cell. However, poor or high cluster forming, can be monitored during the sequencing run using the "% of occupancy" parameter.

The percentage of clusters passing filter (%PF) is an indication of signal purity from each cluster. Overclustered flow cells typically have higher numbers of overlapping clusters. This leads to poor template generation, which then causes a decrease in the %PF metric.

As previously discussed, the first analysis step is the quality filtering of the reads. To compare the sequencing efficiency of both platform, data on filtered reads yield of samples sequenced in double are shown in the following graphs (Fig. 15 and Fig. 16):
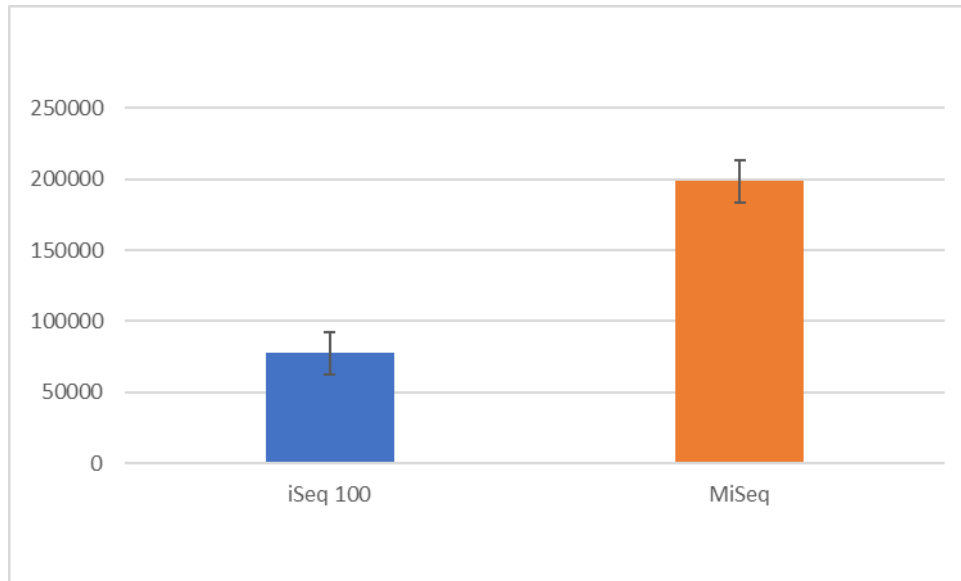
Fig. 15: Medium value of filtered reads obtained with the analysis on the same library in different sequencing platform (iSeq 100: 77619.21 average reads number, error: 11009.10, Miseq: 198753.56 average reads number, error: 14848.05)
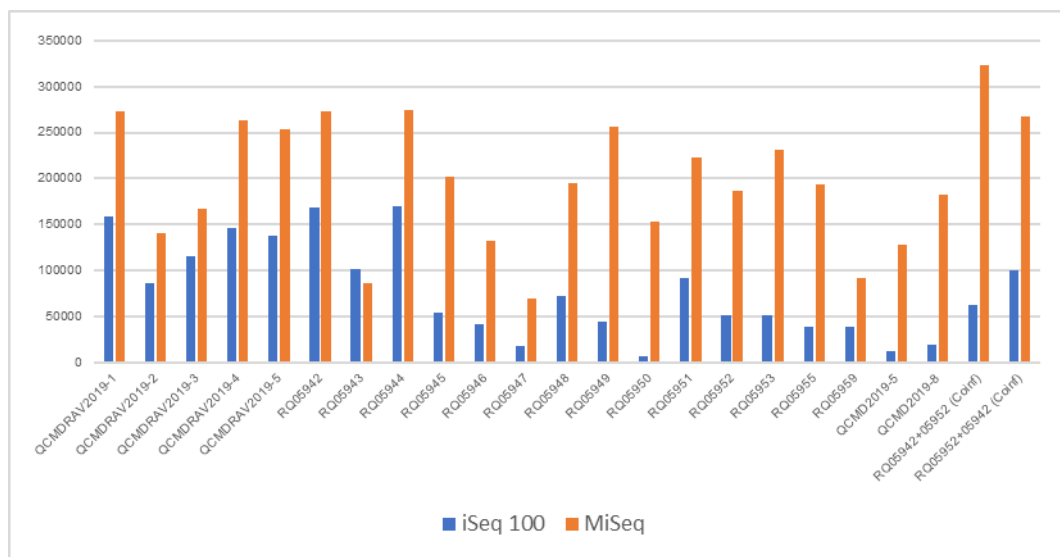
Fig. 16: Comparison between filtered reads number obtained in each sample in both sequencing platforms.

Data confirm that in this case iSeq™100 (Illumina) run has a lower efficiency performance than MiSeq one with the same samples' libraries.

In the following table (Table 20) is reported the summary of the obtained results with HCV NGS Software and Geno2Pheno program for genotyping and RASs evaluation.

In the result table the number of the filtered reads are reported as

indication of the final performance of the library preparation protocol.

Moreover, are reported the detected genotype, the covered amino acid positions for each analysed viral region (5'UTR, CORE, NS5B for each genotype, NS3 and NS5A for GT1 sample), the medium coverage and the RASs detected with Geno2Pheno tool (Table 20):

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
RQ05836	1.28E+05	1a	MiSeq (Illumina)	1a	131874(74.93%)	1-190 (4629X)	191-331 (312X)	7812-9172 (6082X)	3299-3928 (4082X)	6136-6822 (6274X)	none
RQ05830	2.75E+05	4d	MiSeq (Illumina)	4d	17838(12.32%)	1-160 (4032X)	201-275 (77X)	7812-8893 (690X)	3567-3928 (906X)	6332-6482 (132X)	N/A
RQ05816	4.02E+05	2	MiSeq (Illumina)	2c	80416(47.81%)	14-255 (9735X)	256-407 (413X)	8083-9211 (3986X)	-	-	N/A
RQ05838	2.86E+04	1a	MiSeq (Illumina)	1a	82406(53.00%)	1-160 (4511X)	200-275 (79X)	7812-8893 (2904X)	3320-3928 (5382X)	6140-6822 (3736X)	Y448H
RQ05788	2.85E+05	1b	MiSeq (Illumina)	1b	56336(49.54%)	66-315 (2980X)	316-465 (117X)	7920-9185 (3246X)	3424-4053 (1201X)	6265-6663 (3113X)	R30Q
QCMD 2017 - 8	3.84E+03	5a	MiSeq (Illumina)	5a	95884(50.77%)	1-216 (13263X)	failed	8010-8418 (11667X)	-	-	N/A
QCMD 2017 - 3	4.00E+03	6a	MiSeq (Illumina)	6a	73436(47.56%)	12-253 (11126X)	failed	8057-9270 (21260X)	-	-	N/A
RQ05819	1.60E+05	3a	MiSeq (Illumina)	3a	102268(72.45%)	90-285 (8787X)	356-424 (52X)	8112-9379 (3987X)	-	-	N/A
RQ05792	2.02E+05	1b	MiSeq (Illumina)	1b	72946(61.99%)	66-315 (3028X)	316-465 (97X)	7919-9015 (3329X)	3405-4479 (576X)	6265-6944 (5490X)	Y56F. C316N
RQ05832	6.33E+05	4d	MiSeq (Illumina)	4d	104522(72.16%)	13-262 (3637X)	263-412 (159X)	8019-9229 (4700X)	-	-	N/A
RQ05840	9.71E+05	1a	MiSeq (Illumina)	1a	147776(77.83%)	1-190 (3781X)	191-331 (392X)	7812-9172 (6547X)	3299-3928 (5418X)	6136-6822 (6653X)	S122G. M28V
RQ05787	5.35E+03	1b	MiSeq (Illumina)	1b	48860(30.41%)	66-276 (5352X)	failed	7971-8936 (1971X)	3820-4053 (82X)	6325-6944 (2503X)	C316N
RQ05824	5.82E+05	2	MiSeq (Illumina)	2c	103532(64.09%)	67-340 (18401X)	341-494 (1778X)	8147-9345 (5867X)	-	-	N/A
RQ05844	1.60E+05	3a	MiSeq (Illumina)	3a	78278(60.74%)	49-244 (6622X)	failed	8071-8740 (4678X)			N/A
QCMDDR2019-1	1.32E+04	NA	iSeq100™ (Illumina)	1a	159106(58.79%)	92-341 (15478X)	342-491 (4512X)	7963-9044 (3601X)	3518-4078 (3508X)	6291-6950 (11103X)	R155K. M28V
QCMDDR2019-2	4.98E+03	NA	iSeq100™ (Illumina)	3a	85850(35.49%)	49-298 (20908X)	299-452 (392X)	8105-9149 (983X)	-	-	NA
QCMDDR2019-3	4.17E+02	NA	iSeq100™ (Illumina)	1b	116094(37.00%)	80-329 (19829X)	330-479 (1328X)	8066-9029 (4189X)	3834-4057 (187X)	6286-6938 (12848X)	R30Q. C316N
QCMDDR2019-4	4.73E+03	NA	iSeq100™ (Illumina)	3a	146306(39.69%)	90-339 (20599X)	340-485 (3307X)	8112-9379 (7671X)	-	-	Y93H
QCMDDR2019-5	3.65E+03	NA	iSeq100™ (Illumina)	1a	137618(44.25%)	103-145 (4383X)	146-295 (6344X)	7767-8848 (9755X)	3302-3882 (4308X)	6095-6777 (7639X)	I170V
RQ05942	2.86E+06	1a	iSeq100™ (Illumina)	1a	169052(70.52%)	92-341 (4743X)	342-491 (5534X)	7963-9323 (7356X)	3440-4078 (4268X)	6291-6973 (10074X)	Q80K
RQ05943	2.61E+05	4	iSeq100™ (Illumina)	4d	101168(34.18%)	13-262 (6758X)	263-412 (7674X)	8002-9175 (5523X)	-	-	NA
RQ05944	3.64E+05	1a	iSeq100™ (Illumina)	1a	169668(68.08%)	92-341 (6312X)	342-491 (7131X)	7963-9323 (5844X)	3471-4079 (8452X)	6291-6973 (9029X)	Q80K

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
RQ05945	2.07E+05	1a	iSeq100™ (Illumina)	1a	54230(76.25%)	92-341 (2140X)	342-491 (1988X)	7963-9322 (2536X)	3518-4078 (2004X)	6291-6970 (3799X)	Q80L. S556G
RQ05946	1.82E+05	3a	iSeq100™ (Illumina)	3a	41620(44.21%)	90-339 (1950X)	340-485 (2875X)	8112-9298 (885X)	-	-	NA
RQ05947	23785.6	1b	iSeq100™ (Illumina)	1b	17964(30.38%)	1-190 (3720X)	191-340 (3077X)	7814-9172 (1402X)	failed	failed	none
RQ05948	33272.6	1a	iSeq100™ (Illumina)	1a	72864(60.73%)	92-341 (3294X)	342-491 (2738X)	7963-9323 (2667X)	3474-4078 (1058X)	6291-6938 (5471X)	Q80K
RQ05949	3.49E+06	2	iSeq100™ (Illumina)	2c	44754(57.52%)	14-255 (6648X)	256-407 (7161X)	8098-9293 (2054X)	-	-	NA
RQ05950	1.04E+07	2	iSeq100™ (Illumina)	2c	6010(45.86%)	14-255 (777X)	256-399 (1043X)	8098-9209 (303X)	-	-	NA
RQ05951	27048.1	1a	iSeq100™ (Illumina)	1a	91328(57.58%)	103-145 (3392X)	146-295 (4893X)	7767-9126 (3773X)	3302-3882 (4715X)	6095-6777 (5323X)	I170V
RQ05952	2.05E+06	4	iSeq100™ (Illumina)	4d	50724(58.11%)	13-262 (9579X)	263-412 (12956X)	7871-9260 (2616X)	-	-	NA
RQ05953	8.68E+05	1a	iSeq100™ (Illumina)	1a	50724(58.11%)	92-341 (1899X)	342-491 (2302X)	7963-9323 (1822X)	3471-4078 (1502X)	6291-6949 (3730X)	Q80K
RQ05955	5.08E+06	1b	iSeq100™ (Illumina)	1b	38426(33.69%)	66-315 (2724X)	316-465 (3133X)	7936-9324 (2088X)	3424-4201 (145X)	6265-6930 (3465X)	Y56F. R30Q. Y93H. S556G
RQ05959	4.41E+03	1b	iSeq100™ (Illumina)	1b	38426(33.69%)	66-315 (8457X)	316-465 (3906X)	7993-8936 (1154X)	3818-4053 (878X)	6456-6826 (3550X)	none
QCMD2019-5	4.03E+03	5a	iSeq100™ (Illumina)	5a	12228 (23.58%)	1-246 (3096X)	247-387 (384X)	8042-8609 (152X)	-	-	NA
QCMD2019-8	1.4E+05	6	iSeq100™ (Illumina)	6a	18894(42.86%)	31-283 (2713X)	284-424 (453X)	8076-8250 (5461X)	-	-	NA
RQ05942+05952 (Coinf)	70%+30%	1a-4	iSeq100™ (Illumina)	1a-4d	62660(68.16%)	92-341 (2064X)	342-491 (2778X)	7963-9323 (2580X)	3465-4078 (1483X)	6291-6923 (4034X)	Q80K
RQ05952+05942 (Coinf)	70%+30%	4-1a	iSeq100™ (Illumina)	4d+1a	99528 (66.75%)	92-341 (4067X)	342-491 (5752X)	7963-9323 (3923X)	3453-4078 (2340X)	6291-6973 (6149X)	Q80K
QCMDDR2019-1	1.32E+04	NA	MiSeq (Illumina)	1a	273826(61.88%)	70-341 (25660X)	342-491 (8069X)	7963-9045 (6415X)	3471-4078 (3947X)	6287-6956 (25627X)	R155K. M28V
QCMDDR2019-2	4.98E+03	NA	MiSeq (Illumina)	3a	141192(36.70%)	28-298 (31825X)	299-452 (649X)	8104-9184 (1580X)	-	-	NA
QCMDDR2019-3	4.17E+02	NA	MiSeq (Illumina)	1b	166986(38.26%)	56-329 (26774X)	330-479 (1960X)	8052-9029 (5704X)	3834-4057 (358X)	6286-6958 (16091X)	R30Q. C316N
QCMDDR2019-4	4.73E+03	NA	MiSeq (Illumina)	3a	263668(41.88%)	69-339 (34856X)	340-485 (6005X)	8112-9379 (17153X)	-	-	Y93H
QCMDDR2019-5	3.65E+03	NA	MiSeq (Illumina)	1a	254274(45.15%)	95-145 (7532X)	146-295 (12869X)	7767-8848 (17913X)	3302-3882 (8214X)	6095-6777 (10349X)	I170V
RQ05942	2.86E+06	1a	MiSeq (Illumina)	1a	273068(75.03%)	92-341 (7358X)	342-491 (8794X)	7963-9323 (12361X)	3440-4078 (7220X)	6291-6973 (15176X)	Q80K

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
RQ05943	2.61E+05	4	MiSeq (Illumina)	4d	85548(36.72%)	13-262 (5535X)	263-412 (6251X)	8004-8934 (7112X)	-	-	NA
RQ05944	3.64E+05	1a	MiSeq (Illumina)	1a	273904(73.35%)	92-341 (10397X)	342-491 (12163X)	7963-9323 (9991X)	3471-4079 (13372X)	6287-6923 (13030X)	Q80K
RQ05945	2.07E+05	1a	MiSeq (Illumina)	1a	202620(64.82%)	68-341 (9951X)	342-496 (7796X)	7963-9322 (6906X)	3498-4078 (6225X)	6291-6973 (13965X)	Q80L. S556G
RQ05946	1.82E+05	3a	MiSeq (Illumina)	3a	132482(52.09%)	90-339 (12926X)	340-485 (9311X)	8112-9379 (4283X)	-	-	NA
RQ05947	23785.6	1b	MiSeq (Illumina)	1b	69096(28.32%)	1-190 (17116X)	191-342 (11313X)	7812-9172 (5156X)	failed	failed	none
RQ05948	33272.6	1a	MiSeq (Illumina)	1a	195372(62.42%)	92-341 (5522X)	342-491 (7067X)	7963-9323 (7424X)	3474-4078 (2916X)	6291-6957 (14013X)	Q80K
RQ05949	3.49E+06	2	MiSeq (Illumina)	2c	256180(69.53%)	1-255 (35222X)	256-416 (39210X)	8094-9293 (11299X)	-	-	NA
RQ05950	1.04E+07	2	MiSeq (Illumina)	2c	153608(43.90%)	2-255 (19390X)	256-415 (21985X)	8083-9293 (6172X)	-	-	NA
RQ05951	27048.1	1a	MiSeq (Illumina)	1a	223200(68.25%)	92-341 (14241X)	342-491 (11666X)	7963-9323 (7945X)	3471-4079 (5416X)	6291-6973 (10670X)	I170V
RQ05952	2.05E+06	4	MiSeq (Illumina)	4d	187032(52.96%)	13-262 (19627X)	263-424 (22735X)	7871-9260 (4760X)	-	-	NA
RQ05953	8.68E+05	1a	MiSeq (Illumina)	1a	231622(59.27%)	92-341 (9853X)	342-496 (11081X)	7963-9323 (8351X)	3453-4078 (5916X)	6287-6973 (15921X)	Q80K
RQ05955	5.08E+06	1b	MiSeq (Illumina)	1b	193644(74.49%)	66-315 (7515X)	316-465 (12454X)	7914-9324 (7624X)	3422-4201 (455X)	6265-6944 (6445X)	Y56F. R30Q. Y93H. S556G
RQ05959	4.41E+03	1b	MiSeq (Illumina)	1b	91876(33.88%)	44-315 (19564X)	316-465 (9110X)	7993-9014 (19654X)	3818-4053 (2055X)	6341-6924 (5606X)	none
QCMD2019-5	4.03E+03	5a	MiSeq (Illumina)	5a	127938(21.71%)	1-246 (47625X)	247-400 (3394X)	8031-8609 (2088X)	-	-	NA
QCMD2019-8	1.4E+05	6	MiSeq (Illumina)	6a	182762(43.18%)	24-295 (32651X)	296-4436 (2759X)	8072-8287 (40385X)	-	-	NA
RQ05942+05952 (Coinf)	70%+30%	1a-4	MiSeq (Illumina)	1a-4d	323650(69.07%)	92-341 (10682X)	342-495 (13777X)	7963-9323 (13323X)	3440-4078 (6987X)	6291-6973 (20184X)	Q80K
RQ05952+05942 (Coinf)	70%+30%	4-1a	MiSeq (Illumina)	4d-1a	267784(68.25%)	92-341 (10619X)	342-491 (14843X)	7963-9323 (10930X)	4353-4078 (6334X)	6291-6973 (15545X)	Q80K

Table 20: Results obtained for the Prototype pre-validation process. Legend: *NA*, not applicable; *none*, no RASs found; *Failed*, no reads found.

As observed from the results table, all the expected genotypes obtained with AMPLIQUALITY HCV TYPE PLUS v2.0, are confirmed with HCV NGS kit.

According to the algorithm used for genotyping analysis, a 100X coverage of 300 bp on NS5B region is sufficient to correctly assign a genotype to the sample. 5'UTR and Core regions are necessary to confirm the genotype and to identify possible recombinant forms.

RASs confirmation is generally carried out with Sanger sequencing. However, the method has a lower sensitivity and a low depth capacity than NGS technology and for this stage of the NGS developing kit, it was not considered.

The evaluation of the RASs presence was performed with Geno2Pheno tool with a 2% frequency cut off, since during prototype pre-validation process HCV NGS software is not optimised for variant calling analysis.

The presence of drug-resistance mutation was evaluated for GT1a and GT1b samples. In the following table are shown the details RASs founded (Table 21) and considered relevant in 2018 EASL guidelines.

Sample	Viral load (IU/mL)	Resulted genotype	Sequencing platform	RASs detected	Viral region interested	AminoAcid Prevalence	Depth
RQ05836	1.28E+05	1a	MiSeq (Illumina)	none	-	-	-
RQ05838	2.86E+04	1a	MiSeq (Illumina)	Y448H	NS5B	4.83%	373/7727
RQ05788	2.85E+05	1b	MiSeq (Illumina)	R30Q	NS5A	39.86%	623/1563
RQ05792	2.02E+05	1b	MiSeq (Illumina)	Y56F	NS3	73.93%	207/280
RQ05792	2.02E+05	1b	MiSeq (Illumina)	C316N	NS5B	100%	1479/1479
RQ05840	9.71E+05	1a	MiSeq (Illumina)	S122G	NS3	2.04%	157/7709
RQ05840	9.71E+05	1a	MiSeq (Illumina)	M28V	NS5A	3.59%	73/2034
RQ05787	5.35E+03	1b	MiSeq (Illumina)	C316N	NS5B	99.75%	804/806
QCMDDR2019-1	1.32E+04	1a	MiSeq (Illumina)	R155K	NS3	99.75%	8666/8688
QCMDDR2019-2	1.32E+04	1a	MiSeq (Illumina)	M28V	NS5A	2.75%	171/6209
QCMDDR2019-3	4.17E+02	1b	MiSeq (Illumina)	R30Q	NS5A	99.90%	2045/2047
QCMDDR2019-4	4.17E+02	1b	MiSeq (Illumina)	C316N	NS5B	98.31%	174/177
QCMDDR2019-5	3.65E+03	1a	MiSeq (Illumina)	I170V	NS3	99.55%	8533/8572
RQ05942	2.86E+06	1a	MiSeq (Illumina)	Q80K	NS3	99.30%	8276/8334
RQ05944	3.64E+05	1a	MiSeq (Illumina)	Q80K	NS3	99.51%	16176/16256
RQ05945	2.07E+05	1a	MiSeq (Illumina)	Q80L	NS3	61.52%	5199/8451
RQ05945	2.07E+05	1a	MiSeq (Illumina)	S556G	NS5B	72.68%	4518/6216
RQ05947	2.37E+04	1b	MiSeq (Illumina)	none	-	-	-
RQ05948	3.32E+04	1a	MiSeq (Illumina)	Q80K	NS3	98.23%	2200/2217
RQ05951	2.70E+04	1a	MiSeq (Illumina)	I170V	NS3	99.42%	1871/1882
RQ05953	8.68E+05	1a	MiSeq (Illumina)	Q80K	NS3	98.74%	8888/9001
RQ05955	5.08E+06	1b	MiSeq (Illumina)	Y56F	NS3	98.99%	492/497
RQ05955	5.08E+06	1b	MiSeq (Illumina)	R30Q	NS5A	15.57%	1519/9756
RQ05955	5.08E+06	1b	MiSeq (Illumina)	Y93H	NS5A	13.34%	3962/29700
RQ05955	5.08E+06	1b	MiSeq (Illumina)	S556G	NS5B	54.04%	635/1175
RQ05959	4.41E+03	1b	MiSeq (Illumina)	none	-	-	-
RQ05942+05952 (Coinf)	70%+30%	1a-4d	MiSeq (Illumina)	Q80K	NS3	99.19%	6975/7032
RQ05952+05942 (Coinf)	70%+30%	4d-1a	MiSeq (Illumina)	Q80K	NS3	98.50%	6030/6122
QCMDDR2019-1	1.32E+04	1a	iSeq100™ (Illumina)	R155K	NS3	99.43%	4895/4923
QCMDDR2019-2	1.32E+04	1a	iSeq100™ (Illumina)	M28V	NS5A	2.99%	100/3340
QCMDDR2019-3	4.17E+02	1b	iSeq100™ (Illumina)	R30Q	NS5A	99.60%	1498/1504
QCMDDR2019-4	4.17E+02	1b	iSeq100™ (Illumina)	C316N	NS5B	100%	119/119
QCMDDR2019-5	3.65E+03	1a	iSeq100™ (Illumina)	I170V	NS3	99.30%	3994/4022

Sample	Viral load (IU/mL)	Resulted genotype	Sequencing platform	RASs detected	Viral region interested	AminoAcid Prevalence	Depth
RQ05942	2.86E+06	1a	iSeq100™ (Illumina)	Q80K	NS3	99.94%	5162/5198
RQ05944	3.64E+05	1a	iSeq100™ (Illumina)	Q80K	NS3	99.75%	10472/10498
RQ05945	2.07E+05	1a	iSeq100™ (Illumina)	Q80L	NS3	60.63%	1366/2253
RQ05945	2.07E+05	1a	iSeq100™ (Illumina)	S556G	NS5B	61.71%	4518/6216
RQ05947	2.37E+04	1b	iSeq100™ (Illumina)	none	-	-	-
RQ05948	3.32E+04	1a	iSeq100™ (Illumina)	Q80K	NS3	98.95%	751/759
RQ05951	2.70E+04	1a	iSeq100™ (Illumina)	I170V	NS3	99.51%	605/608
RQ05953	8.68E+05	1a	iSeq100™ (Illumina)	Q80K	NS3	99.19%	2074/2091
RQ05955	5.08E+06	1b	iSeq100™ (Illumina)	Y56F	NS3	99.30%	141/142
RQ05955	5.08E+06	1b	iSeq100™ (Illumina)	R30Q	NS5A	16.92%	397/2346
RQ05955	5.08E+06	1b	iSeq100™ (Illumina)	Y93H	NS5A	13.51%	948/7016
RQ05955	5.08E+06	1b	iSeq100™ (Illumina)	S556G	NS5B	42.02%	137/326
RQ05959	4.41E+03	1b	iSeq100™ (Illumina)	none	-	-	-
RQ05942+05952 (Coinf)	70%+30%	1a-4d	iSeq100™ (Illumina)	Q80K	NS3	99.54%	1528/1535
RQ05952+05942 (Coinf)	70%+30%	4d-1a	iSeq100™ (Illumina)	Q80K	NS3	98.60%	2387/2421

Table 21: Summary of the RASs in the sequenced GT1 samples, analysed with Geno2Pheno tool with a 2% cutoff. AminoAcid Prevalence is the mutation frequency observed in the sample for the mutated aminoacid; Depth of coverage is the number of reads sequenced for the specific aminoacid found.

Data obtained with the prototype pre-validation experiments demonstrate that kit's prototype complies with the intended use defined during project design.

A good sequencing depth, enough for genotyping analysis, was obtained for all the samples in both platforms and all the obtained genotype results are comparable with the expected ones.

For RASs evaluation no reference data obtained with a gold standard method such as Sanger sequencing are available on analysed samples. However, all the mutations found are supported with a sufficient depth of coverage.

The only exceptions are represented by C316N mutation in QCMDDR2019-4 sample supported by only 119X depth and Y56F in RQ05955 with a depth of 142X. Both results were obtained on iSeq™100 (Illumina) platform. The same mutations analysed with MiSeq (Illumina) platform had a 177X and a 497X depth respectively. These data suggest

that the difference between the depth results are due to the different sequencing efficiency of the run in the two platforms. Moreover, the lower depth obtained in these two cases in both sequencing runs, is probably due to the extreme genetic variability of the HCV that makes its amplification very difficult to achieve. In fact, the primer design process on this type of genome can lead to a lower coverage of certain regions (for example the regions' edge and GC rich regions).

Regarding sample RQ05947, where NS5A and NS3 regions were not detected (see Table 20), no RASs were found. Further analysis has to be conduct on that sample to confirm the data.

Among mutations, some of them were found with an extremely low frequency such as Y448H in RQ05838 sample (4.83%), S122G (2.04%) and M28V (3.59%) in RQ05840 sample. Even if we have no reference data that could confirm these evidences, all these mutations are supported with a high coverage: 7727X, 7709X and 2034X respectively.

The experiment on artificial coinfections demonstrate that it is not possible to define a coinfection percentage between the different genotypes. In fact, the mutation frequency does not change on the basis of the percentage of the GT1a sample spiked in the artificial coinfection. In both cases of coinfection, one with 70% and one with 30% of GT1a sample (RQ05942), the frequency of the mutation Q80K is around 99%. Additionally, the number of reads for the mutation in all the coinfection cases (RQ05942 70% and RQ05942 30%) do not correspond to the sample percentage inside the coinfection in comparison with the original sample (RQ05942 100%). Samples RQ05952 is a 4d genotype and no information regarding RASs are available.

Data obtained with the prototype pre-validation experiment demonstrate that kit's prototype complies with the intended use defined during project design. Further experiment will be necessary to define kit's performances and limitations and to optimized RASs evaluation on the pipeline.

In fact, considering all these evidences, it was decided to perform the validation process.

5.2.2.2 Validation

Validation process is necessary to define the performances values of a diagnostic assay.

For HCV NGS kit, were defined: genotyping sensitivity and specificity and Limit of Detection for genotype and RASs evaluation. In the following paragraph a detailed description.

Diagnostic sensitivity

For the sensitivity associated with genotypes detection, 23 positive samples were processed. The final library was sequenced on iSeq100™ (Illumina) platform and the run parameters obtained were: 71.3% Occupancy, 95.8% reads1 with a >Q30 quality and 93.2% of reads2 with a >Q30 quality and 64.4% clusters passing filter.

Library was also sequenced with MiSeq (Illumina) platform and the run parameters obtained were: 1074 K/mm^2 cluster density, 93,69% of passing filter cluster and 92.62% reads with a >Q30 quality.

Genotype analysis was performed with the bioinformatics pipeline in development, while results on RASs for GT1 samples were obtained with Geno2Pheno tool.

In the following table are summarized (Table 22) the evaluation results on genotype sensitivity:

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	1a	230332 (69.22%)	1-190 (8190X)	191-340 (10934X)	7812-9172 (9294X)	3320-3927 (9069X)	6140-6822 (12146)	I170V
RQ05956	1.87E+07	1a	iSeq100™ (Illumina)	1a	230972 (61.15%)	1-190 (10999X)	191-340 (13714X)	7812-9172 (9133X)	3297-3862 (8637X)	6140-6822 (11456X)	I170V
RQ05957	9.50E+07	2	iSeq100™ (Illumina)	2c	234430 (59.34%)	91-340 (15057X)	341-511 (19313X)	8114-9400 (14552X)	-	-	N/A
RQ05958	2.76E+04	1b	iSeq100™ (Illumina)	1b	207884 (43.23%)	32-304 (26128X)	305-454 (12909X)	8024-9154 (12455X)	3994-4032 (53X)	6445-6913 (6268X)	none
RQ05960	8.70E+04	3a	iSeq100™ (Illumina)	3a	265016 (64.50%)	69-309 (29265X)	362-424 (71X)	8113-9379 (4637X)	-	-	N/A
RQ05961	1.05E+04	1b	iSeq100™ (Illumina)	1b	194172 (59.32%)	66-315 (13858X)	316-465 (9682X)	7938-9324 (7770X)	3711-3815 (57X)	6267-6924 (14085X)	none
RQ05968	3.32E+04	1a	iSeq100™ (Illumina)	1a	237120 (72.87%)	52-341 (11155X)	342-491 (14311X)	7963-9198 (9293X)	3474-4078 (3742X)	6291-6973 (17041X)	Q80K
RQ05969	3.49E+06	2	iSeq100™ (Illumina)	2c	204620 (60.64%)	14-255 (24731X)	256-422 (32818X)	8094-9293 (11120X)	-	-	N/A
RQ05970	2.70E+04	1a	iSeq100™ (Illumina)	1a	202334 (61.28%)	92-341 (10606X)	342-491 (13065X)	7963-9204 (6890X)	3471-4079 (8883X)	6287-6973 (10661X)	V55A, I170V, S122T
RQ05971	2.05E+06	4	iSeq100™ (Illumina)	4d	205100 (63.43%)	13-262 (18126X)	263-413 (25862X)	7871-9260 (5929X)	-	-	N/A
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	1a	221846 (65.23%)	92-341 (7595X)	342-491 (11393X)	7952-9323 (8696X)	3453-4078 (7650X)	6287-6973 (12072X)	Q80K
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	1b	251250 (64.25%)	66-315 (11465X)	316-465 (14750X)	7913-9324 (10332X)	3422-4479 (1166X)	6265-6924 (18221X)	Y56F, R30Q, Y93H, S556G
RQ05974	1.87E+07	1a	iSeq100™ (Illumina)	1a	250242 (58.59%)	1-190 (12259X)	191-340 (16517X)	7812-9172 (7613X)	3297-3822 (9507X)	6140-6822 (14672X)	I170V
RQ05975	9.50E+07	2	iSeq100™ (Illumina)	2c	212154 (73.74%)	91-340 (16652X)	341-508 (19139X)	8114-9384 (12776X)	-	-	N/A
RQ05976	4.41E+03	1b	iSeq100™ (Illumina)	1b	199260 (35.21%)	65-315 (18250X)	316-465 (23077X)	7938-8936 (9356X)	3818-4053 (749X)	6305-6924 (11162X)	none
RQ05977	8.70E+04	3a	iSeq100™ (Illumina)	3a	244172 (59.70%)	48-260 (10594X)	316-383 (54X)	8072-8847 (15917X)	-	-	N/A
QCMD2016-5	4.36E+03	5a	iSeq100™ (Illumina)	5a	67476 (25.41%)	1-246 (24801X)	247-400 (2947X)	8031-8228 (139X)	-	-	N/A
QCMD2016-7	5.00E+03	5a	iSeq100™ (Illumina)	5a	72728 (26.52%)	6-279 (29364X)	280-433 (1424X)	8064-8326 (1182X)	-	-	N/A
QCMD2018-3	5.09E+03	6a	iSeq100™ (Illumina)	6a	49054 (16.69%)	46-295 (18980X)	296-436 (2287X)	8091-8265 (1625X)	-	-	N/A
QCMD2017-2	3.37E+02	2b	iSeq100™ (Illumina)	2b	5486 (1.21%)	92-287 (2562X))	failed	8151-8320 (1134X)	-	-	N/A
QCMD2017-4	2.26E+03	3a	iSeq100™ (Illumina)	3a	100344 (37.45%)	28-298 (27894X)	299-444 (1431X)	8071-9150 (1181X)	-	-	N/A

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
RQ05982	2.00E+06	2k/1b	iSeq100™ (Illumina)	2k/1b	199496 (71.19%)	66-315 (5763X)	316 – 470 (13591X)	7934-9185 (5929X)	3602-4056 (6258X)	6266-6933 (16745X)	S122T
Instand CTRL	N/A	1b	iSeq100™ (Illumina)	1b	229794 (52.30%)	66-315 (17005X)	316-465 (19374X)	7936-9324 (4562X)	3621-3840 (103X)	6265-6944 (28460X)	Y93H, L28F
RQ05954	4.22E+06	1a	MiSeq (Illumina)	1a	273506 (70.75%)	1-190 (9565X)	191-340 (13495X)	7812-9172 (11015X)	3320-3928 (11008X)	6140-6822 (14258X)	170V
RQ05956	1.87E+07	1a	MiSeq (Illumina)	1a	288722 (62.69%)	1-190 (13373X)	191-340 (17276X)	7812-9172 (11574X)	3297-3862 (10805X)	6140-6822 (14264X)	I170V
RQ05957	9.50E+07	2	MiSeq (Illumina)	2c	298398 (61.28%)	91-340 (18622X)	341-511 (24530X)	8114-9400 (18693X)	-	-	N/A
RQ05958	2.76E+04	1b	MiSeq (Illumina)	1b	245846 (42.58%)	31-304 (31298X)	305-454 (15954X)	7982-9174 (14687X)	3994-4468 (72X)	6293-6913 (4045X)	none
RQ05960	8.70E+04	3a	MiSeq (Illumina)	3a	316526 (65.07%)	69-339 (30429X)	340-485 (90X)	8113-9379 (5504X)	-	-	N/A
RQ05961	1.05E+04	1b	MiSeq (Illumina)	1b	230632 (59.12%)	66-315 (16483X)	316-465 (12072X)	7938-9324 (4467X)	3692-3840 (95X)	6265-6924 (16686X)	none
RQ05968	3.32E+04	1a	MiSeq (Illumina)	1a	306928 (75.30%)	92-341 (13795X)	342-500 (17563X)	7963-9198 (11938X)	3471-4078 (4932X)	6291-9198 (22562X)	Q80K
RQ05969	3.49E+06	2	MiSeq (Illumina)	2c	233428 (59.10%)	71-340 (28638X)	341-513 (38812X)	8158-9399 (11730X)	-	-	N/A
RQ05970	2.70E+04	1a	MiSeq (Illumina)	1a	245734 (62.46%)	92-341 (12464X)	342-501 (14933X)	7963-9204 (8410X)	3471-4079 (10818X)	6285-6973 (12963X)	V55A, I170V, A122T
RQ05971	2.05E+06	4	MiSeq (Illumina)	4d	237440 (65.94%)	12-262 (20516X)	263-419 (28184X)	7871-9260 (6925X)	-	-	N/A
RQ05972	8.68E+05	1a	MiSeq (Illumina)	1a	272982 (67.14%)	92-341 (9144X)	342-491 (13882X)	7935-9323 (10828X)	3453-4078 (9296X)	6287-6973 (14662X)	Q80K
RQ05973	5.08E+06	1b	MiSeq (Illumina)	1b	320832 (64.55%)	66-315 (14600X)	316-465 (19161X)	7913-9324 (13427X)	3422-4479 (890X)	6265-6924 (22870X)	Y56F, R30Q, Y63H, S556G
RQ05974	1.87E+07	1a	MiSeq (Illumina)	1a	345886 (60.02%)	1-190 (15775X)	191-350 (21645X)	7812-9172 (10733X)	3297-3927 (15519X)	6140-6822 (19882X)	I170V
RQ05975	9.50E+07	2	MiSeq (Illumina)	2c	256580 (75.58%)	91-340 (19433X)	341-512 (22792X)	8114-9384 (15632X)	-	-	N/A
RQ05976	4.41E+03	1b	MiSeq (Illumina)	1b	405298 (64.04%)	82-332 (20470X)	333-482 (25803X)	7955-8974 (283112X)	failed	6322-6941 (23358X)	none
RQ05977	8.70E+04	3a	MiSeq (Illumina)	3a	265386 (57.55%)	77-309 (11057X)	356-423 (67X)	8113-8888 (11583X)	-	-	N/A
QCMD2016-5	4.36E+03	5a	MiSeq (Illumina)	5a	79636 (25.87%)	1-246 (29235X)	247-400 (3618X)	8031-8228 (170X)	-	-	N/A
QCMD2016-7	5.00E+03	5a	MiSeq (Illumina)	5a	81986 (27.02%)	6-279 (33085X)	280-433 (1590X)	8064-8346 (1397X)	-	-	N/A
QCMD2018-3	5.09E+03	6a	MiSeq (Illumina)	6a	60952 (17.91%)	46-295 (23229X)	296-436 (3168X)	8091-8265 (2162X)	-	-	N/A

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
QCMD2017-2	3.37E+02	2b	MiSeq (Illumina)	2b	7534 (1.24%)	92-287 (3364X)	failed	8151-8320 (1758x)	-	-	N/A
QCMD2017-4	2.26E+03	3a	MiSeq (Illumina)	3a	143170 (38.53%)	28-298 (38202X)	299-444 (2277X)	8071-9184 (1829X)	-	-	N/A
RQ05982	2.00E+06	2k/1b	MiSeq (Illumina)	2k/1b	276900 (70.61%)	66-315 (10427X)	316-482 (17991X)	7934-9185 (8164X)			S112T
Instand CTRL	N/A	1b	MiSeq (Illumina)	1b	256726 (51.66%)	66-315 (18949X)	316-465 (22717X)	7936-9324 (5265X)	3621-3840 (126X)	6265-6944 (31603X)	Y93H, L28F

Table 22: Results obtained for genotype sensitivity performance. Legend: *NA*, not applicable; *none*, no RASs found; *Failed*, no reads found.

Sequences analysis demonstrates that the concordance between expected and obtained results on samples genotype was 100%.

Moreover, samples with no expected subtype such as RQ05957, RQ05969, RQ05971 and RQ05975, have a defined subtype with NGS analysis.

Genotype definition was identified analysing NS5B, 5'UTR and CORE regions. In sample QCMD2017-2, CORE regions was not sequenced. Probably it was due to the low viral load. The result on this low viral load sample defines a possible limit of detection of the kit that will be defined in a specific experiment.

As previously discussed, 6 out of 23 samples were extracted from the same clinical samples. In particular, RQ05956 and RQ05974 RNA were extracted from sample Clin-Samp02745, RQ05957 and RQ05975 RNA were extracted from Clin-Samp02746 and RQ05960 and RQ05977 RNA were extracted from sample Clin-Samp02761. Different extracts were made with different extraction systems as it is showed in Table 10. Overall, considering two parameters of the sequencing run (total quantity of filtered reads and region's depth of coverage), the obtained data showed that the two extraction systems gave similar results.

The only exception was for sample Clin-Samp02761, identified as GT3a (RNA extracts: RQ05960 and RQ05977), where the depth of coverage in the CORE region is low but comparable between the two extracts. These findings confirmed that the two extraction systems did not affect the results. Also, similar results were previously obtained for other samples with the same genotype (RQ05819 and RQ05844, genotype 3a) that gave low depth of coverage or failed in the CORE region.

Sample RQ05982 was previously sequenced with the WGS approach (Clin-Samp02320, Table 19). This sample was already confirmed to be a recombinant form but in the previous experiment it was impossible to analyse the presence of RASs due to the low sequencing depth. With amplicon-based approach it was possible not only to identify the presence of the recombinant form, but also to detect the presence of RASs.

The comparison between the iSeq 100 and Miseq showed evaluating differences on depth of coverage and amplified region

overlapping showed no significant difference between the two platforms (chi-squared test, p=0.99).

Diagnostic Specificity

In order to evaluate the specificity of HCV NGS kit, 10 negative samples were analysed. The QC1 step returned the following results (Table 23, Fig. 17):

Sample	Agerose gel sample ID (Fig 18)	Viral load (IU/mL)	Expected Genotype	QUBIT quantification (ng/μL)
RQ04501	1	undet	Neg	12.7
RQ04500	2	undet	Neg	3.44
RQ04502	3	undet	Neg	3.48
RQ04598	4	undet	Neg	4.06
RQ04599	5	undet	Neg	2.32
RQ04605	6	undet	Neg	6.48
QCMD2019-4	7	undet	Neg	3.46
QCMD2018-6	8	undet	Neg	3.28
QCMD2017-6	9	undet	Neg	2.84
QCMD2016-2	10	undet	Neg	3.64
Instand 1b	11	N/A	1b	17.7
NTC	NTC			1.57

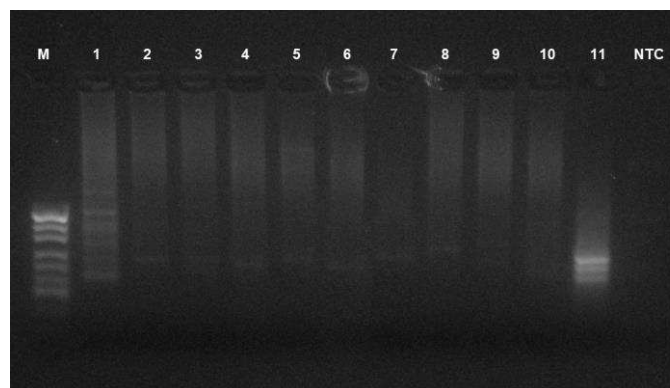**Table 23:** QC1 QUBIT quantification.

Fig. 17: QC1 electrophoresis 2.5% agarose gel result. Sample 11 is the reference positive control. Negative samples, 1 to 10. *NTC*: No Template Control.
Legend: **M**: marker puc19 (501/489, 404, 331, 242, 190, 147. 111/110, 67, 34, 25);

The first quality control step, that consists in a DNA quantification with QUBIT and in an electrophoresis gel (2.5% agarose), showed that negative samples have no specific bands, between 250 and 300 bp, compared with the reference positive control used (Fig. 16), and also a low amount of DNA (Table 23).

In silico and *in vitro* analysis revealed that the kit was specific on HCV amplification. However, the use of the kit will be intended only for HCV positive samples.

Limit of Detection

In order to evaluate the LoD of the device, 20 HCV positive samples were processed and sequenced on the iSeq 100™ (Illumina) platform.

The purpose of the experiment was to evaluate the lower viral load in which the device can detect both genotype and RASs.

For the genotype LoD were analysed 3 GT1a samples and 1 sample for GT1b GT2, GT3a, GT4, GT5 and GT6. Original samples were already analysed in diagnostic sensitivity run. The viral load of considered samples was between 10^2 and 10^3 IU/mL.

Sequencing values obtained were: 78.6% of occupancy, cluster passing filter 69.4%, %Q30 Read 1 94.9%, %Q30 Read 2 92.5%.

3 GT1a samples and 1 GT1b sample were used to calculate also LoD for RASs.

Each sample were diluted in HCV negative RNA to obtain useful dilution for LoD calculation. In the following table (Table 24) the summary of the obtained results:

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
QCMDDR 2019-1	1.32E+04	1a	iSeq100™ (Illumina)	1a	231842 (42.84%)	92-341 (5738X)	342-502 (23806X)	7963-9198 (7117X)	3518-4078 (4115X)	6291-6950 (21286X)	R155K
QCMDDR 2019-1 1:10	1.32E+03	1a	iSeq100™ (Illumina)	1a	159136 (29.84%)	72-341 (17195X)	342 – 491 (2765X)	7963-9044 (2607X)	3518-4078 (885X)	6291-6950 (17944X)	R155K
QCMDDR 2019-1 1:100	1.32E+02	1a	iSeq100™ (Illumina)	1a	131858 (34.73%)	96-145 (7254X)	146-305 (11255X)	7917-9126 (1361X)	3322-3642 (620X)	6095-6732 (16877X)	none
QCMDDR 2019-3	4.18E+02	1b	iSeq100™ (Illumina)	1b	107108 (56.25%)	82-332 (14638X)	333-476 (3031X)	8073-8956 (4459X)	failed	6284-6941 (13573X)	R30Q, C316N
QCMDDR 2019-5	3.66E+03	1a	iSeq100™ (Illumina)	1a	194920 (50.06%)	94-145 (12121X)	146-303 (20001X)	7767-8848 (8927X)	3322-3882 (8927X)	6095-6777 (24161X)	I170V
QCMDDR 2019-5 1:10	3.66E+02	1a	iSeq100™ (Illumina)	1a	70298 (15.05%)	94-341 (6933X)	342-496 (18617X)	7963-9044 (2710X)	3844-4075 (158X)	6293-6633 (179X)	none
QCMDDR 2018-3	1.30E+04	1a	iSeq100™ (Illumina)	1a	174860 (48.09%)	92-341 (20100X)	342-495 (15058X)	7963-9194 (2231X)	3518-4078 (1108X)	6291-6950 (16869X)	R155K, I170V, M28V
QCMDDR 2018-3 1:10	1.30E+03	1a	iSeq100™ (Illumina)	1a	74508 (23.04%)	1-190 (11405X)	191-340 (5258X)	7814-9172 (4023X)	failed	6140-6787 (3728X)	none
QCMDDR 2018-3 1:100	1.30E+02	1a	iSeq100™ (Illumina)	1a	33985 (16.20%)	failed	failed	8000-9062 (5021X)	failed	6291-6502 (2589X)	none
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	1a	294770 (75.02%)	1-190 ()11662X)	191-348 (17559X)	7812-9172 (11613X)	3320-4905 (4802X)	6140-6822 (15943X)	I170V
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	1a	225658 (64.99%)	92-341 (8819X)	342-491 (12746X)	7963-9323 (8165X)	3453-4078 (7910X)	6287-6950 (14404X)	Q80K
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	1b	219666 (63.27%)	66-315 (12249X)	316-465 (16094X)	7915-9324 (8329X)	3422-4479 (527X)	6265-6930 (16375X)	Y56F, R30Q, Y93H. S556G
RQ05969 1:100	3.49E+04	2	iSeq100™ (Illumina)	2c	219344 (40.11%)	1-255 (49367X)	256-416 (64163X)	8099-9293 (2361X)	-	-	N/A
RQ05969 1:1000	3.49E+03	2	iSeq100™ (Illumina)	2c	112908 (25.32%)	2-225 (30562X)	256-415 (32467X)	8141-9293 (590X)	-	-	N/A
RQ05960 1:10	8.78E+03	3a	iSeq100™ (Illumina)	3a	153192 (37.12%)	76-313 (18960X)	355-426 (65X)	8117-9353 (2966X)	-	-	N/A
RQ05960 1:100	8.77E+02	3a	iSeq100™ (Illumina)	3a	40812 (7.12%)	94-311 (7401X)	failed	8150-8283 (258X)	-	-	N/A
RQ05971 1:1000	2.05E+03	4	iSeq100™ (Illumina)	4d	108148 (31.40%)	1-262 (24938X)	263-416 (24315X)	8019-8851 (1048X)	-	-	N/A
RQ05971 1:10000	2.05E+02	4	iSeq100™ (Illumina)	4d	40194 (10.53%)	91-340 (9844X)	341-490 (10089X)	8097-8238 (1561X)	-	-	N/A
QCMD2016-7 1:10	5.00E+02	5a	iSeq100™ (Illumina)	5a	13676 (4.28%)	30-279 (4287X)	280-420 (1422X)	failed	-	-	N/A
QCMD2018-3 1:10	5.09E+02	6a	iSeq100™ (Illumina)	failled	-	-	-	-	-	-	N/A

Table 24: Summary of Limit of Detection (LoD) experiment. Obtained genotype results, depth of coverage and RASs are reported for each sample. Legend: *failed*: no reads found, *N/A*: Not Available, *none*: no RAS found.

For GT1 samples, the lowest viral load, in which regions for the genotype calling are sequenced, was 1.30×10^2 IU/mL

For GT2 samples, the viral load between 10^3 IU/mL and 10^4 IU/mL were evaluated. In the diagnostic sensitivity run, a sample with a viral load of 3.37×10^2 IU/mL sample (QCMD2017-2) was analysed and showed the failure of NS5B sequencing. The LoD run showed that the lowest viral load with significant results was: 3.49×10^3. Although further analysis are required for the final evaluation, the preliminary results on GT2 samples showed that the LoD can be set between 10^2 IU/mL and 10^3 IU/mL.

For the other genotypes the LoD was between 2.05×10^2 IU/mL (GT4) and 8.77×10^2 IU/mL (GT3)

Further experiments are needed to calculate the exact LoD for each genotype.

In the following table are shown the RASs details found in all the analysed GT1 samples for the kit's validation (Table 25):

Sample	Viral load (IU/mL)	Resulted genotype	Sequencing platform	RASs detected	Viral region interested	AminoAcid Prevalence	Depth
Diagnostic sensitivity							
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	I170V	NS3	45.57%	36/79
RQ05956	1.87E+07	1a	iSeq100™ (Illumina)	I170V	NS3	100%	138/138
RQ05958	2.76E+04	1b	iSeq100™ (Illumina)	none	-	-	-
RQ05961	1.05E+04	1b	iSeq100™ (Illumina)	none	-	-	-
RQ05968	3.32E+04	1a	iSeq100™ (Illumina)	Q80K	NS3	99.66%	2675/2684
RQ05970	2.70E+04	1a	iSeq100™ (Illumina)	V55A	NS3	99.58%	15644/15710
RQ05970	2.70E+04	1a	iSeq100™ (Illumina)	I170V	NS3	99.26%	2813/2834
RQ05970	2.70E+04	1a	iSeq100™ (Illumina)	S122T	NS3	97.61%	20214/20708
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	Q80K	NS3	98.97%	9959/10063
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	Y56F	NS3	98.82%	336/340
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	R30Q	NS5A	11.71%	842/7188
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	Y93H	NS5A	8.61%	2455/28515
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	S556G	NS5B	42.49%	362/852
RQ05974	1.87E+07	1a	iSeq100™ (Illumina)	I170V	NS3	96.88%	31/32
RQ05976	4.41E+03	1b	iSeq100™ (Illumina)	none	-	-	-
Instand CTRL	N/A	1b	iSeq100™ (Illumina)	Y93H	NS5A	51.81%	9853/19016
Instand CTRL	N/A	1b	iSeq100™ (Illumina)	L28F	NS5A	92.62%	20387/22012
RQ05982	2.00E+06	2k/1b	iSeq100™ (Illumina)	S122T	NS3	99.90%	2939/2942
RQ05954	4.22E+06	1a	MiSeq (Illumina)	I170V	NS3	44.35%	51/115
RQ05956	1.87E+07	1a	MiSeq (Illumina)	I170V	NS3	98.47%	258/262
RQ05958	2.76E+04	1b	MiSeq (Illumina)	none	-	-	-
RQ05961	1.05E+04	1b	MiSeq (Illumina)	none	-	-	-
RQ05968	3.32E+04	1a	MiSeq (Illumina)	Q80K	NS3	99.50%	3609/3627
RQ05970	2.70E+04	1a	MiSeq (Illumina)	V55A	NS3	99.53%	18956/19045
RQ05970	2.70E+04	1a	MiSeq (Illumina)	I170V	NS3	99.09%	3371/3402
RQ05970	2.70E+04	1a	MiSeq (Illumina)	S122T	NS3	97.75%	25064/25641
RQ05972	8.68E+05	1a	MiSeq (Illumina)	Q80K	NS3	99.04%	11676/11789
RQ05973	5.08E+06	1b	MiSeq (Illumina)	Y56F	NS3	99.04%	414/418
RQ05973	5.08E+06	1b	MiSeq (Illumina)	R30Q	NS5A	10.88%	1008/9263
RQ05973	5.08E+06	1b	MiSeq (Illumina)	Y93H	NS5A	9.08%	3029/33359
RQ05973	5.08E+06	1b	MiSeq (Illumina)	S556G	NS5B	46.13%	524/1136
RQ05974	1.87E+07	1a	MiSeq (Illumina)	I170V	NS3	98.11%	52/53

Sample	Viral load (IU/mL)	Resulted genotype	Sequencing platform	RASs detected	Viral region interested	AminoAcid Prevalence	Depth
RQ05976	4.41E+03	1b	MiSeq (Illumina)	none	-	-	-
Instand CTRL	N/A	1b	MiSeq (Illumina)	Y93H	NS5A	22.11%	2906/13144
Instand CTRL	N/A	1b	MiSeq (Illumina)	L28F	NS5A	89.92%	22249/24743
RQ05982	2.00E+06	2k/1b	MiSeq (Illumina)	S122T	NS3	99.48%	4397/4420

Limit of detection

Sample	Viral load (IU/mL)	Resulted genotype	Sequencing platform	RASs detected	Viral region interested	AminoAcid Prevalence	Depth
QCMDDR 2019-1	1.32E+04	1a	iSeq100™ (Illumina)	R155K	NS3	99.34%	6035/6075
QCMDDR 2019-1 1:10	1.32E+03	1a	iSeq100™ (Illumina)	R155K	NS3	99.53%	1469/1476
QCMDDR 2019-1 1:100	1.32E+02	1a	iSeq100™ (Illumina)	none	-	-	-
QCMDDR 2019-3	4.18E+02	1b	iSeq100™ (Illumina)	R30Q	NS5A	100%	166/166
QCMDDR 2019-3	4.18E+02	1b	iSeq100™ (Illumina)	C316N	NS5B	100%	142/142
QCMDDR 2019-5	3.66E+03	1a	iSeq100™ (Illumina)	I170V	NS3	99.67%	3281/3292
QCMDDR 2019-5 1:10	3.66E+02	1a	iSeq100™ (Illumina)	none	-	-	-
QCMDDR 2018-3	1.30E+04	1a	iSeq100™ (Illumina)	R155K	NS3	99.78%	2285/2290
QCMDDR 2018-3	1.30E+04	1a	iSeq100™ (Illumina)	I170V	NS3	3.79%	80/2112
QCMDDR 2018-3	1.30E+04	1a	iSeq100™ (Illumina)	M28V	NS5A	2.88%	106/3682
QCMDDR 2018-3 1:10	1.30E+03	1a	iSeq100™ (Illumina)	none	-	-	-
QCMDDR 2018-3 1:100	1.30E+02	1a	iSeq100™ (Illumina)	none	-	-	-
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	I170V	NS3	57.47%	50/87
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	Q80K	NS3	98.85%	10784/10909
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	Y56F	NS3	99.54%	217/218
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	R30Q	NS5A	10.96%	670/6113
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	Y93H	NS5A	9.16%	2494/27232
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	S556G	NS5B	46.48%	66/142

Table 25: Summary of the RASs in the sequenced GT1 samples, analysed with Geno2Pheno tool with a 2% cutoff. AminoAcid Prevalence is the mutation frequency observed in the sample for the mutated aminoacid; Depth of coverage is the number of reads sequenced for the specific aminoacid.

Regarding RASs LoD, QCMD panel samples for RAS evaluation was used. The preliminary results showed that the lowest viral load in which all the RASs are detected was around 10^4 UI/mL.

Further experiments are required for a better assessment of RASs LoD. Moreover a preliminary inter-run reproducibility test was performed analysing 3 samples in duplicate in two different runs. In Table 26 are reported the results.

Sample	Viral load (IU/mL)	Expected Genotype	Sequencing platform	Resulted genotype	Reads number (filtered)	5'UTR coverage	CORE coverage	NS5B coverage	NS3 coverage	NS5A coverage	RASs detected (2% cutoff)
									For GT1 only		
Diagnostic sensitivity											
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	1a	230332 (69.22%)	1-190 (8190X)	191-340 (10934X)	7812-9172 (9294X)	3320-3927 (9069X)	6140-6822 (12146)	I170V
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	1a	221846 (65.23%)	92-341 (7595X)	342-491 (11393X)	7952-9323 (8696X)	3453-4078 (7650X)	6287-6973 (12072X)	Q80K
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	1b	251250 (64.25%)	66-315 (11465X)	316-465 (14750X)	7913-9324 (10332X)	3422-4479 (1166X)	6265-6924 (18221X)	Y56F, R30Q, Y93H, S556G
Limit of Detection											
RQ05954	4.22E+06	1a	iSeq100™ (Illumina)	1a	294770 (75.02%)	1-190 ()11662X)	191-348 (17559X)	7812-9172 (11613X)	3320-4905 (4802X)	6140-6822 (15943X)	I170V
RQ05972	8.68E+05	1a	iSeq100™ (Illumina)	1a	225658 (64.99%)	92-341 (8819X)	342-491 (12746X)	7963-9323 (8165X)	3453-4078 (7910X)	6287-6950 (14404X)	Q80K
RQ05973	5.08E+06	1b	iSeq100™ (Illumina)	1b	219666 (63.27%)	66-315 (12249X)	316-465 (16094X)	7915-9324 (8329X)	3422-4479 (527X)	6265-6930 (16375X)	Y56F, R30Q, Y93H. S556G

Table 26: Preliminary results on HCV NGS kit reproducibility.

The results obtained on this preliminary reproducibility study showed that with high viral load samples, the device is extremely reproducible: region covered and depth of sequencing are comparable with an identical outcome of RASs detected.

Further studies are needed to evaluate reproducibility on lower viral load samples.

In the following table are listed the frequencies of each mutation found in all the clinical samples analysed in the prototype pre-validation and validation processes (Table 27):

Sample ID	Genotype	NS3								NS5A				NS5B		
		V55A	Y56F	Q80K	Q80L	S122G	S122T	R155K	I170V	M28V	L28F	R30Q	Y93H	Y448H	S556G	C316N
RQ05838 MiSeq	1a													4.83		
RQ05792 MiSeq	1b		73.93													100
RQ05788 MiSeq	1b											39.86				
RQ05840 MiSeq	1a					2.04				3.59						
RQ05787 MiSeq	1b															99.75
RQ05942 MiSeq	1a			99.30												
RQ05942 iSeq100				99.94												
RQ05944 MiSeq	1a			99.51												
RQ05944 iSeq100				99.75												
RQ05945 MiSeq	1a				61.52										72.68	
RQ05945 iSeq					60.63										61.71	
RQ05948 MiSeq	1a			98.23												
RQ05948 iSeq100				98.95												
RQ05951 MiSeq	1a								99.42							
RQ05951 iSeq100									99.51							
RQ05953 MiSeq	1a			98.74												
RQ05953 iSeq100				99.19												
RQ05955 MiSeq	1b		98.99									15.57	13.34		54.04	
RQ05955 iSeq100			99.30									16.92	13.51		42.02	
RQ05942 +05952 (Coinf) MiSeq	1a-4d			99.19												
RQ05942 +05952 (Coinf) iSeq100				99.54												
RQ05952 +05942 (Coinf) MiSeq	4d-1a			98.50												
RQ05952 +05942 (Coinf) iSeq100				98.60												

Sample ID	Genotype	NS3								NS5A				NS5B		
		V55A	Y56F	Q80K	Q80L	S122G	S122T	R155K	I170V	M28V	L28F	R30Q	Y93H	Y448H	S556G	C316N
RQ05954 MSeq	1a								44.35							
RQ05954 iSeq100	1a								45.57							
RQ05956 MiSeq	1a								98.47							
RQ05956 iSeq100	1a								100							
RQ05968 MiSeq	1a			99.50												
RQ05968 iSeq100	1a			99.66												
RQ05970 MiSeq	1a	99.53					97.75		99.09							
RQ05970 iSeq100	1a	99.58					97.61		99.26							
RQ05972 MiSeq	1a			99.04												
RQ05972 iSeq100	1a			98.97												
RQ05973 MiSeq	1b		99.04									10.88	9.08		46.13	
RQ05973 iSeq100	1b		98.82									11.71	8.61		42.49	
RQ05974 MiSeq	1a								98.11							
RQ05974 iSeq100	1a								96.88							
Instand CTRL MiSeq	1b										89.92		22.11			
Instand CTRL iSeq100	1b										92.62		51.81			
RQ05982 iSeq100	2k/1b						99.48									
RQ05982 iSeq100	2k/1b						99.90									
QCMDDR 2019-1 iSeq100	1a							99.34								
QCMDDR 2019-3 iSeq100	1b											100				100
QCMDDR 2019-5 iSeq100	1a								99.67							
QCMDDR 2018-3 iSeq100	1a							99.78	3.79							

Table 27: Details of RASs in GT1 clinical samples. Values are expressed in percentage. Percentage of each mutation is relative to the number of the reads that confirm the mutation (frequency).

Data obtained on RASs analysis revealed that the most frequent mutations were in the NS3 region. In particular Q80K and I170V are the most common among samples showing also high frequency in each sample.

As it was observed in prototype pre-validation, RAS identified in sequenced samples were generally well supported by a good sequencing depth, with some exception as I170V in RQ05954 and RQ05974 samples (both with a coverage below 50X).

Even low frequency mutations such as I170V in 'QCMDDR2018-3 sample' (3.79%) Y448H in RQ05838 sample (4.83%), S122G (2.04%) and M28V (3.59%) in RQ05840 sample were also well supported with good sequencing depth (2112X, 7727X, 7709X and 2034X respectively).

Double sequenced samples revealed a similar mutation frequency showing the strength of the library preparation protocol and its reproducibility. In conclusion both sequencing platforms are suitable for the HCV NGS kit's usage.

6. DISCUSSION

Hepatitis C virus is one of the most important pathogens worldwide and the World Health Organization (WHO) is making lot of efforts to eradicate this dangerous and expensive disease (GLOBAL HEALTH SECTOR STRATEGY ON VIRAL HEPATITIS 2016-2021). The best way to achieve the target is to decrease the costs of the diagnosis and the treatment for the poorest areas, especially where HCV is still an endemic disease. In fact, the overpopulation factor increases the contagion possibility and decrease the rate of the therapeutic failure. With the development of the DAAs, the achievement of SVR is possible for all the genotypes and all the patients with different disease stages progression.

However, a little percentage of cases still undergoes to therapeutic failure which depend on two main causes: a wrong genotyping result with the subsequent wrong genotype-dependent therapy or, the presence of drug resistance mutations.

Genotyping problem seems to be overcome by the recent introduction of pangenotypic antiviral drugs that demonstrated to be effective for the main HCV genotypes and subtypes. However, new pangenotypic drugs such as combination of Sofosbuvir/Velpatasvir/Voxilaprevir (Vosevi) and Sofosbuvir/Velpatasvir (Epclusa) have respectively an SVR12 rate of 97% and 90% (Pol and Parlati 2018). In fact, as it is reported in Cuypers et al. 2019, even though pangenotypic antiviral regimens demonstrate their efficacy to all HCV infected individuals, therapy success seems to depend on the HCV genotype background. Specific HCV subtypes such as GT1a or GT3a have a lower response rate to therapy than other genotypes and require the addition of Ribavirin or the extension of the therapy duration to reach SVR.

Genotype misclassification can result in wrong treatment management and in subsequent failure of the therapy. Commercial assays are mostly based on 5'UTR together with NS5B (Abbott RealTime HCV Genotype II) or CORE region analysis (INNO-LiPA-HCV-2.0 – Siemens Healthcare or AMPLIQUALITY HCV TYPE PLUS - AB ANALITICA). For probe-based method a sequencing-based assays is

required, in particular for undetermined GT1 subtypes. As previously discussed, undetermined genotype cases can depend on the presence of either a recombinant or coinfection. Given the importance of RASs detection in personalized medicine, the NGS technique is the best diagnostic tool able to gather both genotype and RASs information (Cuypers et al., 2016).

In particular, NGS technology can generate a higher data magnitude than Sanger sequencing. The cost reduction for samples analysis has increased the spread of NGS technique even in the diagnostic field. Clinical management of viral infections can greatly benefit from the ultra-wide and ultra-deep characteristics of NGS for genetically highly variable diseases such as HCV has previously discussed.

In the oncologic field, NGS technology is already widely used but for virology this technology is still not fully developed. A diagnostic kit based on NGS technology for HCV genotyping and RASs evaluations called Sentosa® SQ HCV Genotyping Assay was developed by Vela Diagnostic. This assay was developed for Ion Torrent platform.

Biofield Innovation developed an NGS based kit compatible with Illumina platforms. The prototype pre-validation process was performed with MiSeq and iSeq100™ (Illumina) platforms and the same instruments was used for validation. In fact, MiSeq platform is spread in the majority of diagnostic laboratories and the iSeq100™ (Illumina) platform has a very competitive price compared to the other instruments.

In two out of three sequencing runs an optimal cluster density was reached. The iSeq100™ (Illumina) run had a low cluster occupancy with a lower number of reads, but nevertheless it gave significant results.

In fact, obtained data showed that the library preparation protocol was strong enough to be not influenced by low efficiency sequencing runs. Moreover, the average depth and coverage values for each sample, showed that the target amplification worked accurately, and the primer design and pooling was correctly made.

Prototype pre-validation process was fundamental to verify the compliance of the prototype within the requirements defined in the project design. In particular, the results from the genotype identification were all consistent to the ones expected. Moreover, the NGS assay was able to identify the subtype for 7 clinical samples for which only the genotype was determined with the reference

method. The first results on RASs evaluation revealed a high percentage of clinical GT1 samples with RASs (12 out of 15). Also, a good sequencing depth was obtained, and all the detected mutation were supported by a consistent number of reads.

Since prototype pre-validation experiments demonstrated that the kit worked as expected, validation process was assessed.

The results from the first validation experiment demonstrated that the genotype sensitivity of the device was 100%. The device recognised all the genotypes and the subtypes of the analysed samples confirming that NGS assays can overcome the undefined subtypes issue that probe-based kits have. as it was observed in Dirani et al. 2018.

The results from prototype pre-validation process showed that the kit is able to identify coinfection and recombinant forms. Further studies are needed to assess the LoD for these particular cases.

Results on specificity performance revealed that the entire analysis process, cannot return false positive results and thus the observed device specificity can be considered 100%.

Results on genotype and RAS LoD reveals that kit's average limit of detection for genotype is between 10^2IU/mL and 10^3IU/mL and for RASs is around 10^3IU/mL. To better define the limit of detection for each genotype further studies with more samples are needed.

RASs analysis were performed with Geno2Pheno tool. The chosen cutoff was 2%. In fact, Perales et al. 2018 showed that RASs with a presence even lower than 10% (i.e. S122G substitution in NS3 and E62D in NS5A were present in 7.1% and 3.7% respectively, Q80L and R155K substitution in NS3 and Y93H in NS5A were present at 6.5%, 4.8%, and 3.5%, respectively) in naïve patient leads to treatment failure. The same patient after the treatment were found to have the 100% prevalence of the originally low frequency RASs, meaning that the detection of RASs at low frequency is recommended for the best treatment choice.

For this study it is not possible to speculate about the treatment efficacy or on the differences between naïve and treatment experienced patients since the clinical history of the samples analysed is unknown.

Regarding the population frequency of RASs, a study based on

Sanger sequencing from Wang et al. 2018, revealed that the prevalence of RASs in GT1 samples is completely different between GT1a and GT1b. In fact, regarding Q80K mutation, the most affected is GT1a (45%, 167 of 370 GT1a samples, versus 3% of GT1b samples, 3 out of 116). For NS5A region mutations, the most expressed is Y93H in GT1b (11%). R30Q has a prevalence of 4% in GT1a. In the same study, the most present mutation for NS5B is S556G in GT1b (14%).

A study from Gozlan et al. 2019, reported that Sanger sequencing can reveal variants present in at least the 15% of the quasispecie, meanwhile NGS sequencing can detect the variants with a prevalence of >1,5%. It means that RASs detected with NGS technology have higher prevalence than the ones reported in Wang et al, 2018 study, due to higher sensitivity of the method. It is challenging to make a comparison between the literature data and the ones presented here, given the limited population sample analysed. However, even with this small number of samples, the same most common RASs such as Q80K in GT1a, Y93H and S556G in GT1b have been found as in Wang et al, 2018 study.

On the other hand, other mutations such as I170V found at 1% in Wang et al, 2018 study is highly present in the analysed cohort. As reported in Kliemann et al. 2016 this mutation is found in 3.21% of GT1a samples, and 65.20% in GT1b.

Regarding recombinant form analysed the only mutation found is S122T. It is known that the patient was cured with Mavyret, one of the pangenotypic drugs that contains both Glecaprevir and Pibrentasvir. As previously discussed, Glecaprevir inhibits the NS3 protein, while Pibrentasvir the NS5A one. The only mutation found in the patient was S122T in NS3. In literature, for subtypes 1a and 1b (von Massow et al. 2019) RAS S122T, D168E and I170V were found to be the most spread mutations in NS3 region. S122T and D168E have been associated with resistance to antivirals. However, for RAS S122T it was never observed a drug resistance effect for Glecaprevir and Pibrentasvir (Sorbo et al. 2018). As a matter of fact, 12 weeks after the treatment, the patient reached the SVR. This finding demonstrated the importance of evaluating the presence of drug resistance mutations, to avoid treatments that could lead to a therapeutic failure.

7. CONCLUSION

Since the discovers of HCV in 1989, many efforts have been made for the eradication of the virus. The molecular biology techniques allowed the development of Direct Acting Antivirals (DAA) through the identification and characterization of HCV encoded proteins such as the NS3 protease, NS5A and the NS5B polymerase. The newer therapy regimens raised above 90% the healing among patients with chronic HCV infection.

Therapeutic failure still represents a challenge. In fact, neither the first generation DAAs nor the newest pangenotypic antiviral regimen can cure the virus in 100% of cases.

WHO is promoting an eradication plan within the 2030. Together with antiviral drugs, worldwide control of HCV would require the development of a prophylactic vaccine, and numerous candidates have been pursued without success so far.

The eradication of this virus could be realized through a prevention plan, promoting diagnosis and by a wider access to treatment to avoid the spreading of this disease.

The aim of this study was the development of diagnostic assays for the genotyping and for the evaluation of the presence of drug-resistance mutation. Validation results of both assays underlined the efficacy in genotype and mutations detection of HCV.

The increasing of diagnostic method based on NGS technology will reduce the cost of this analysis, leading to a broader access to the cure and a faster approach to the definitive viral eradication.

8. BIBLIOGRAPHY

Argentini, Claudio, Domenico Genovese, Stefano Dettori, and Maria Rapicetta. 2009. "HCV Genetic Variability: From Quasispecies Evolution to Genotype Classification." *Future Microbiology*. https://doi.org/10.2217/fmb.09.8.

Asselah, Tarik, Patrick Marcellin, and Raymond F. Schinazi. 2018. "Treatment of Hepatitis C Virus Infection with Direct-Acting Antiviral Agents: 100% Cure?" *Liver International*. Blackwell Publishing Ltd. https://doi.org/10.1111/liv.13673.

Balogh, Julius, David Victor, Emad H Asham, Sherilyn Gordon Burroughs, Maha Boktour, Ashish Saharia, Xian Li, R Mark Ghobrial, and Howard P Monsour. 2016. "Hepatocellular Carcinoma: A Review." *Journal of Hepatocellular Carcinoma* 3: 41–53. https://doi.org/10.2147/JHC.S61146.

Barzon, Luisa, Enrico Lavezzo, Valentina Militello, Stefano Toppo, and Giorgio Palù. 2014. "Applications of next Generation Sequencing Technologies to Diagnostic Virology." In *Omics in Clinical Practice: Genomics, Pharmacogenomics, Proteomics, and Transcriptomics in Clinical Research*, 351–80. Apple Academic Press. https://doi.org/10.1201/b17137.

Bataller, Ramón, and David A. Brenner. 2005. "Liver Fibrosis." *Journal of Clinical Investigation*. The American Society for Clinical Investigation. https://doi.org/10.1172/JCI24282.

Bedossa, P, and T Poynard. 1996. "An Algorithm for the Grading of Activity in Chronic Hepatitis C. The METAVIR Cooperative Study Group." *Hepatology (Baltimore, Md.)* 24 (2): 289–93. https://doi.org/10.1002/hep.510240201.

Blackard, Jason T., and Kenneth E. Sherman. 2007. "Hepatitis C Virus Coinfection and Superinfection." *The Journal of Infectious Diseases* 195 (4): 519–24. https://doi.org/10.1086/510858.

Borgia, Sergio M, Charlotte Hedskog, Bandita Parhy, Robert H Hyland, Luisa M Stamm, Diana M Brainard, Mani G Subramanian, et al. 2018. "Identification of a Novel Hepatitis C Virus Genotype From Punjab, India: Expanding Classification of Hepatitis C Virus Into 8 Genotypes." *The Journal of Infectious Diseases* 218 (11): 1722–29. https://doi.org/10.1093/infdis/jiy401.

Bourlière, Marc, and Olivia Pietri. 2019. "Hepatitis C Virus Therapy: No One Will Be Left Behind." *International Journal of Antimicrobial Agents*. Elsevier B.V. https://doi.org/10.1016/j.ijantimicag.2018.12.010.

Brimacombe, Claire L, Joe Grove, Luke W Meredith, Ke Hu, Andrew J Syder, Maria Victoria Flores, Jennifer M Timpe, et al. 2011. "Neutralizing Antibody-Resistant Hepatitis C Virus Cell-to-Cell Transmission." *Journal of Virology* 85 (1): 596–605. https://doi.org/10.1128/JVI.01592-10.

Bukh, Jens. 2016. "The History of Hepatitis C Virus (HCV): Basic Research Reveals Unique Features in Phylogeny, Evolution and the Viral Life Cycle with New Perspectives for Epidemic Control." *Journal of Hepatology* 65 (1 Suppl): S2–21. https://doi.org/10.1016/j.jhep.2016.07.035.

Cento, Valeria, Stephane Chevaliez, and Carlo Federico Perno. 2015. "Resistance to Direct-Acting Antiviral Agents: Clinical Utility and Significance." *Current Opinion in HIV and AIDS* 10 (5): 381–89. https://doi.org/10.1097/COH.0000000000000177.

Chae, Hee Bok, Seon Mee Park, and Sei Jin Youn. 2013. "Direct-Acting Antivirals for the Treatment of Chronic Hepatitis C: Open Issues and Future Perspectives." *The Scientific World Journal*. https://doi.org/10.1155/2013/704912.

Chemaitelly, Hiam, Karima Chaabna, and Laith J. Abu-Raddad. 2015. "The Epidemiology of Hepatitis C Virus in the Fertile Crescent: Systematic Review and Meta-Analysis." *PLoS ONE*. https://doi.org/10.1371/journal.pone.0135281.

Chevaliez, Stéphane, Magali Bouvier-Alias, Rozenn Brillet, and Jean Michel Pawlotsky. 2009. "Hepatitis C Virus (HCV) Genotype 1 Subtype Identification in New HCV Drug Development and Future Clinical Practice." *PLoS ONE* 4 (12): 1–9. https://doi.org/10.1371/journal.pone.0008209.

Chevaliez, Stéphane, and Jean-Michel Pawlotsky. 2006. *HCV Genome and Life Cycle*. *Hepatitis C Viruses: Genomes and Molecular Biology*.

Chhatwal, J., Q. Chen, T. Ayer, E. D. Bethea, F. Kanwal, K. V. Kowdley, X. Wang, M. S. Roberts, and S. C. Gordon. 2018. "Hepatitis C Virus Re-Treatment in the Era of Direct-Acting Antivirals: Projections in the USA." *Alimentary Pharmacology and Therapeutics* 47 (7): 1023–31. https://doi.org/10.1111/apt.14527.

Chigbu, Loonawat, Sehgal, Patel, and Jain. 2019. "Hepatitis C Virus Infection: Host–Virus Interaction and Mechanisms of Viral Persistence." *Cells* 8 (4): 376. https://doi.org/10.3390/cells8040376.

Colina, Rodney, Didier Casane, Silvia Vasquez, Laura García-Aguirre, Ausberto Chunga, Héctor Romero, Baldip Khan, and Juan Cristina. 2004. "Evidence of Intratypic Recombination in Natural Populations of Hepatitis C Virus." *Journal of General Virology*. https://doi.org/10.1099/vir.0.19472-0.

Conti, Fabio, Federica Buonfiglioli, Alessandra Scuteri, Cristina Crespi, Luigi Bolondi, Paolo Caraceni, Francesco Giuseppe Foschi, et al. 2016. "Early Occurrence and Recurrence of Hepatocellular Carcinoma in HCV-Related Cirrhosis Treated with Direct-Acting Antivirals." *Journal of Hepatology* 65 (4): 727–33. https://doi.org/10.1016/j.jhep.2016.06.015.

Cuypers, Lize, Francesca Ceccherini-Silberstein, Kristel Van Laethem, Guangdi Li, Anne Mieke Vandamme, and Jürgen Kurt Rockstroh. 2016. "Impact of HCV Genotype on Treatment Regimens and Drug Resistance: A Snapshot in Time." *Reviews in Medical Virology*. https://doi.org/10.1002/rmv.1895.

Cuypers, Lize, Marijn Thijssen, Arash Shakibzadeh, Farzaneh Sabahi, Mehrdad Ravanshad, and Mahmoud Reza Pourkarim. 2019. "Next-Generation Sequencing for the Clinical Management of Hepatitis C Virus Infections: Does One Test Fits All Purposes?" *Critical Reviews in Clinical Laboratory Sciences*, July, 1–15. https://doi.org/10.1080/10408363.2019.1637394.

Davis, Gary L., James E. Albright, Suzanne F. Cook, and Daniel M. Rosenberg. 2003. "Projecting Future Complications of Chronic Hepatitis C in the United States." *Liver Transplantation* 9 (4): 331–38. https://doi.org/10.1053/jlts.2003.50073.

Desmet, Valeer J., Michael Gerber, Jay H. Hoofnagle, Michael Manns, and Peter J. Scheuer. 1994. "Classification of Chronic Hepatitis: Diagnosis, Grading and Staging." *Hepatology*. https://doi.org/10.1002/hep.1840190629.

Dirani, G., E. Paesini, E. Mascetra, P. Farabegoli, B. Dalmo, B. Bartolini, A. R. Garbuglia, M. R. Capobianchi, and V. Sambri. 2018. "A Novel next Generation Sequencing Assay as an Alternative to Currently Available Methods for Hepatitis C Virus Genotyping." *Journal of Virological Methods*. https://doi.org/10.1016/j.jviromet.2017.10.005.

Foucher, Juliette, Laurent Casté Ra, Pierre-Henri Bernard, Xavier Adhoute, David Laharie, Julien Bertet, Patrice Couzigou, and Victor De Lé Dinghen. 2006. "Prevalence and Factors Associated with Failure of Liver Stiffness Measurement Using FibroScan in a Prospective Study of 2114 Examinations." *European Journal*

of Gastroenterology & Hepatology. Vol. 18. Lippincott Williams & Wilkins.

Galli, Andrea, and Jens Bukh. 2014. "Comparative Analysis of the Molecular Mechanisms of Recombination in Hepatitis C Virus." *Trends in Microbiology* 22 (6): 354–64. https://doi.org/10.1016/J.TIM.2014.02.005.

"GLOBAL HEALTH SECTOR STRATEGY ON VIRAL HEPATITIS 2016-2021." n.d.

"GLOBAL HEPATITIS REPORT,2017 WHO." n.d.

Gong, Shunyou, Christine L. Schmotzer, and Lan Zhou. 2016. "Evaluation of Quantitative Real-Time PCR as a Hepatitis C Virus Supplementary Test After RIBA Discontinuation." *Journal of Clinical Laboratory Analysis* 30 (5): 418–23. https://doi.org/10.1002/jcla.21873.

González-Candelas, Fernando, F. Xavier López-Labrador, and María Alma Bracho. 2011. "Recombination in Hepatitis C Virus." *Viruses*. https://doi.org/10.3390/v3102006.

Gower, Erin, Chris Estes, Sarah Blach, Kathryn Razavi-Shearer, and Homie Razavi. 2014. "Global Epidemiology and Genotype Distribution of the Hepatitis C Virus Infection." *Journal of Hepatology*. https://doi.org/10.1016/j.jhep.2014.07.027.

Gozlan, Yael, Efrat Bucris, Rachel Shirazi, Avia Rakovsky, Ziv Ben-Ari, Yana Davidov, Ella Veizman, et al. 2019. "High Frequency of Multiclass HCV Resistance-Associated Mutations in Patients Failing Direct-Acting Antivirals: Real-Life Data." *Antiviral Therapy*, March. https://doi.org/10.3851/IMP3301.

Guettouche, Toumy, and H James Hnatyszyn. 2005. "Chronic Hepatitis B and Viral Genotype: The Clinical Significance of Determining HBV Genotypes." *Antiviral Therapy* 10 (5): 593–604. http://www.ncbi.nlm.nih.gov/pubmed/16152753.

"HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C Welcome and Methods." 2014. www.HCVGuidance.org.

Hopkins, Sam, and Philippe Gallay. 2012. "Cyclophilin Inhibitors: An Emerging Class of Therapeutics for the Treatment of Chronic Hepatitis C Infection." *Viruses*. https://doi.org/10.3390/v4112558.

Houghton, Michael. 2009. "Discovery of the Hepatitis C Virus." *Liver International* 29 (January): 82–88. https://doi.org/10.1111/j.1478-3231.2008.01925.x.

Ishak, Kamal, Amelia Baptista, Leonardo Bianchi, Francesco Callea, Jan De Groote, Fred Gudat, Helmut Denk, et al. 1995. "Histological Grading and Staging of Chronic Hepatitis." *Journal of Hepatology*. https://doi.org/10.1016/0168-8278(95)80226-6.

Jackowiak, Paulina, Karolina Kuls, Lucyna Budzko, Anna Mania, Magdalena Figlerowicz, and Marek Figlerowicz. 2014. "Phylogeny and Molecular Evolution of the Hepatitis C Virus." *Infection, Genetics and Evolution*. Elsevier. https://doi.org/10.1016/j.meegid.2013.10.021.

Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius. 2002. "A Natural Intergenotypic Recombinant of Hepatitis C Virus Identified in St. Petersburg." *Journal of Virology* 76 (8): 4034–43. https://doi.org/10.1128/jvi.76.8.4034-4043.2002.

Kalinina, Olga, Helene Norder, and Lars O Magnius. 2004. "Full-Length Open Reading Frame of a Recombinant Hepatitis C Virus Strain from St Petersburg: Proposed Mechanism for Its Formation." *The Journal of General Virology* 85 (Pt 7): 1853–57. https://doi.org/10.1099/vir.0.79984-0.

Keukeleire, Steven De, Patrick Descheemaeker, and Marijke Reynders. 2015. "Diagnosis of Hepatitis C Virus Genotype 2k/1b Needs NS5B Sequencing." *International Journal of Infectious Diseases* 41 (December): 1–2.

https://doi.org/10.1016/J.IJID.2015.10.010.

Khudur Al-Nassary, Mithaq Sabeeh, and Batool Mutar Mahdi. 2018. "Study of Hepatitis C Virus Detection Assays." *Annals of Medicine and Surgery* 36 (December): 47–50. https://doi.org/10.1016/j.amsu.2018.10.002.

Kliemann, Dimas Alexandre, Cristiane Valle Tovo, Ana Beatriz Gorini Da Veiga, Angelo Alves De Mattos, and Charles Wood. 2016. "Polymorphisms and Resistance Mutations of Hepatitis C Virus on Sequences in the European Hepatitis C Virus Database." *World Journal of Gastroenterology* 22 (40): 8910–17. https://doi.org/10.3748/wjg.v22.i40.8910.

Liang, T. Jake, and Theo Heller. 2004. "Pathogenesis of Hepatitis C-Associated Hepatocellular Carcinoma." In *Gastroenterology*. Vol. 127. W.B. Saunders. https://doi.org/10.1053/j.gastro.2004.09.017.

Manns, Michael P, Maria Buti, Ed Gane, Jean-Michel Pawlotsky, Homie Razavi, Norah Terrault, and Zobair Younossi. 2017. "Hepatitis C Virus Infection." *Nature Reviews. Disease Primers* 3 (March): 17006. https://doi.org/10.1038/nrdp.2017.6.

Massow, Georg von, Damir Garcia-Cehic, Josep Gregori, Francisco Rodriguez-Frias, María Dolores Macià, Ana Escarda, Juan Ignacio Esteban, and Josep Quer. 2019. "Whole-Genome Characterization and Resistance-Associated Substitutions in a New HCV Genotype 1 Subtype." *Infection and Drug Resistance* 12: 947–55. https://doi.org/10.2147/IDR.S195441.

McHutchison, John G., Gregory T. Everson, Stuart C. Gordon, Ira M. Jacobson, Mark Sulkowski, Robert Kauffman, Lindsay McNair, John Alam, and Andrew J. Muir. 2009. "Telaprevir with Peginterferon and Ribavirin for Chronic HCV Genotype 1 Infection." *New England Journal of Medicine* 360 (18): 1827–38. https://doi.org/10.1056/NEJMoa0806104.

Mehta, Shruti H, Andrea Cox, Donald R Hoover, Xiao-Hong Wang, Qing Mao, Stuart Ray, Steffanie A Strathdee, David Vlahov, and David L Thomas. 2002. "Protection against Persistence of Hepatitis C." *Lancet (London, England)* 359 (9316): 1478–83. https://doi.org/10.1016/S0140-6736(02)08435-0.

Minichini, Carmine, Mario Starace, Stefania De Pascalis, Margherita Macera, Laura Occhiello, Mara Caroprese, Martina Vitrone, et al. 2018. "HCV-Genotype 3h, a Difficult-to-Diagnose Sub-Genotype in the DAA Era." *Antiviral Therapy* 23 (7): 605–9. https://doi.org/10.3851/IMP3228.

Mohamed, Amal Ahmed, Tamer A. Elbedewy, Magdy El-Serafy, Naglaa El-Toukhy, Wesam Ahmed, and Zaniab Ali El Din. 2015. "Hepatitis C Virus: A Global View." *World Journal of Hepatology*. https://doi.org/10.4254/wjh.v7.i26.2676.

Moradpour, Darius, François Penin, and Charles M. Rice. 2007. "Replication of Hepatitis C Virus." *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro1645.

Murphy, Donald G, Erwin Sablon, Jasmine Chamberland, Eric Fournier, Raymond Dandavino, and Cécile L Tremblay. 2015. "Hepatitis C Virus Genotype 7, a New Genotype Originating from Central Africa." *Journal of Clinical Microbiology* 53 (3): 967–72. https://doi.org/10.1128/JCM.02831-14.

Ozaras, Resat, and Veysel Tahan. 2009. "Acute Hepatitis C: Prevention and Treatment." *Expert Review of Anti-Infective Therapy* 7 (3): 351–61. https://doi.org/10.1586/eri.09.8.

Pawlotsky, Jean-Michel. 2016. "Hepatitis C Virus Resistance to Direct-Acting Antiviral Drugs in Interferon-Free Regimens." *Gastroenterology*. https://doi.org/10.1053/j.gastro.2016.04.003.

Pawlotsky, Jean-Michel, Francesco Negro, Alessio Aghemo, Marina Berenguer, Olav Dalgard, Geoffrey Dusheiko, Fiona Marra, Massimo Puoti, and Heiner Wedemeyer. 2018. "EASL Recommendations on Treatment of Hepatitis C 2018 Q." https://doi.org/10.1016/j.

Pearlman, Brian L, and Nomi Traub. 2011. "Sustained Virologic Response to Antiviral Therapy for Chronic Hepatitis C Virus Infection: A Cure and so Much More." *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 52 (7): 889–900. https://doi.org/10.1093/cid/cir076.

Perales, Celia, Qian Chen, Maria Eugenia Soria, Josep Gregori, Damir Garcia-Cehic, Leonardo Nieto-Aponte, Lluis Castells, et al. 2018. "Baseline Hepatitis C Virus Resistance-Associated Substitutions Present at Frequencies Lower than 15% May Be Clinically Significant." *Infection and Drug Resistance* 11: 2207–10. https://doi.org/10.2147/IDR.S172226.

Petruzziello, Arnolfo, Samantha Marigliano, Giovanna Loquercio, Anna Cozzolino, and Carmela Cacciapuoti. 2016. "Global Epidemiology of Hepatitis C Virus Infection: An up-Date of the Distribution and Circulation of Hepatitis C Virus Genotypes." *World Journal of Gastroenterology* 22 (34): 7824–40. https://doi.org/10.3748/wjg.v22.i34.7824.

Pham, Son T., Rowena A. Bull, James M. Bennett, William D. Rawlinson, Gregory J. Dore, Andrew R. Lloyd, and Peter A. White. 2010. "Frequent Multiple Hepatitis C Virus Infections among Injection Drug Users in a Prison Setting." *Hepatology* 52 (5): 1564–72. https://doi.org/10.1002/hep.23885.

Pol, Stanislas, and Lucia Parlati. 2018. "Treatment of Hepatitis C: The Use of the New Pangenotypic Direct-Acting Antivirals in 'Special Populations.'" *Liver International*. Blackwell Publishing Ltd. https://doi.org/10.1111/liv.13626.

Popescu, C I, L Riva, O Vlaicu, R Farhat, Y Rouille, and J Dubuisson. 2014. "Hepatitis C Virus Life Cycle and Lipid Metabolism." *Biology (Basel)*. https://doi.org/10.3390/biology3040892.

Powdrill, Megan H, Egor P Tchesnokov, Robert A Kozak, Rodney S Russell, Ross Martin, Evguenia S Svarovskaia, Hongmei Mo, Roger D Kouyos, and Matthias Götte. 2011. "Contribution of a Mutational Bias in Hepatitis C Virus Replication to the Genetic Barrier in the Development of Drug Resistance." *Proceedings of the National Academy of Sciences of the United States of America* 108 (51): 20509–13. https://doi.org/10.1073/pnas.1105797108.

Prasad, Mona R, and Jonathan R Honegger. 2013. "Hepatitis C Virus in Pregnancy." *American Journal of Perinatology* 30 (2): 149–59. https://doi.org/10.1055/s-0033-1334459.

Rosen, Hugo R. 2011. "Clinical Practice Chronic Hepatitis C Infection." *The New England Journal of Medicine Downloaded from Nejm.Org at VA LIBRARY NETWORK On*.

Rousselet, Marie Christine, Sophie Michalak, Florence Dupré, Anne Croué, Pierre Bedossa, Jean Paul Saint-André, and Paul Calès. 2005. "Sources of Variability in Histological Scoring of Chronic Viral Hepatitis." *Hepatology* 41 (2): 257–64. https://doi.org/10.1002/hep.20535.

Sandres-Sauné, K., P. Deny, C. Pasquier, V. Thibaut, G. Duverlie, and J. Izopet. 2003. "Determining Hepatitis C Genotype by Analyzing the Sequence of the NS5b Region." *Journal of Virological Methods*. https://doi.org/10.1016/S0166-0934(03)00070-3.

Simmonds, Peter, Jens Bukh, Christophe Combet, Gilbert Deléage, Nobuyuki Enomoto, Stephen Feinstone, Phillippe Halfon, et al. 2005. "Consensus Proposals for a

Unified System of Nomenclature of Hepatitis C Virus Genotypes." *Hepatology (Baltimore, Md.)* 42 (4): 962–73. https://doi.org/10.1002/hep.20819.

Sodano, Giuseppe, Erasmo Falco, Adriana Raddi, Maria Grimaldi, Francesco Labonia, Marcello Raffone, Mariano Bernardo, and Emanuele Durante Mangoni. 2014. "Acute Hepatitis HCV Genotype 3h: Virological Baseline Characterization and Monitoring 'On Therapy.'" *Microbiologia Medica*. https://doi.org/10.4081/mm.2010.2411.

Sorbo, Maria C., Valeria Cento, Velia C. Di Maio, Anita Y.M. Howe, Federico Garcia, Carlo F. Perno, and Francesca Ceccherini-Silberstein. 2018. "Hepatitis C Virus Drug Resistance Associated Substitutions and Their Clinical Relevance: Update 2018." *Drug Resistance Updates*. https://doi.org/10.1016/j.drup.2018.01.004.

Susser, Simone, Julia Dietz, Bernhard Schlevogt, Eli Zuckerman, Mira Barak, Valeria Piazzolla, Anita Howe, et al. 2017. "Origin, Prevalence and Response to Therapy of Hepatitis C Virus Genotype 2k/1b Chimeras." *Journal of Hepatology* 67 (4): 680–86. https://doi.org/10.1016/j.jhep.2017.05.027.

Swadling, Leo, Paul Klenerman, and Eleanor Barnes. 2013. "Ever Closer to a Prophylactic Vaccine for HCV." *Expert Opinion on Biological Therapy* 13 (8): 1109–24. https://doi.org/10.1517/14712598.2013.791277.

Taherkhani, Reza, and Fatemeh Farshadpour. 2017. "Global Elimination of Hepatitis C Virus Infection: Progresses and the Remaining Challenges." *World Journal of Hepatology*. https://doi.org/10.4254/wjh.v9.i33.1239.

Tsochatzis, Emmanuel A., Jaime Bosch, and Andrew K. Burroughs. 2014. "Liver Cirrhosis." In *The Lancet*, 383:1749–61. Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(14)60121-5.

Tsukiyama-Kohara, Kyoko, and Michinori Kohara. 2017. "Hepatitis C Virus: Viral Quasispecies and Genotypes." *International Journal of Molecular Sciences* 19 (1). https://doi.org/10.3390/ijms19010023.

Wang, Gary P, Norah Terrault, Jacqueline D Reeves, Lin Liu, Eric Li, Lisa Zhao, Joseph K Lim, et al. 2018. "Prevalence and Impact of Baseline Resistance-Associated Substitutions on the Efficacy of Ledipasvir/Sofosbuvir or Simeprevir/Sofosbuvir against GT1 HCV Infection." *Scientific Reports* 8 (1): 3199. https://doi.org/10.1038/s41598-018-21303-2.

Warkad, Shrikant Dashrath, Satish Balasaheb Nimse, Keum Soo Song, and Taisun Kim. 2018. "HCV Detection, Discrimination, and Genotyping Technologies." *Sensors (Switzerland)*. MDPI AG. https://doi.org/10.3390/s18103423.

"WHO | Global Hepatitis Report, 2017." 2018. *WHO*. World Health Organization.

Wilkins, Thad, Jennifer K Malcolm, Dimple Raina, and Robert R Schade. 2010. "Hepatitis C: Diagnosis and Treatment." *American Family Physician* 81 (11): 1351–57. http://www.ncbi.nlm.nih.gov/pubmed/20521755.

Wyles, David L. 2017. "Resistance to DAAs: When to Look and When It Matters." *Current HIV/AIDS Reports* 14 (6): 229–37. https://doi.org/10.1007/s11904-017-0369-5.

Ye, Mei, Xin Chen, Yu Wang, Lin Duo, Chiyu Zhang, and Yong-Tang Zheng. 2019. "Identification of a New HCV Subtype 6xg Among Injection Drug Users in Kachin, Myanmar." *Frontiers in Microbiology* 10 (April). https://doi.org/10.3389/fmicb.2019.00814.

Zakalashvili, M., J. Zarkua, M. Weizenegger, J. Bartel, M. Raabe, T. Telia, M. Zhamutashvili, et al. 2017. "Treatment Outcomes of Hepatitis C Virus Recombinant Form 2k/1b with Sofosbuvir Based Regimens in Georgia." *Journal of Hepatology* 66 (1): S302–3. https://doi.org/10.1016/S0168-8278(17)30922-4.

Zhang, Xingquan. 2016. "Direct Anti-HCV Agents." *Acta Pharmaceutica Sinica B.* Chinese Academy of Medical Sciences. https://doi.org/10.1016/j.apsb.2015.09.008.

Zignego, A. L., C. Ferri, S. A. Pileri, P. Caini, and F. B. Bianchi. 2007. "Extrahepatic Manifestations of Hepatitis C Virus Infection: A General Overview and Guidelines for a Clinical Approach." *Digestive and Liver Disease.* https://doi.org/10.1016/j.dld.2006.06.008.

http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

http://bowtie-bio.sourceforge.net/bowtie2/index.shtml).

http://samtools.sourceforge.net/

http://ugene.net/

http://varscan.sourceforge.net/

http://www.mbio.ncsu.edu/BioEdit/bioedit.html

http://www.simit.org/IT/.xhtml. 31st May 2019

https://blast.ncbi.nlm.nih.gov/Blast.cgi)

https://cutadapt.readthedocs.io/en/stable/guide.html

https://hcv.lanl.gov/content/index

https://hcv.lanl.gov/content/index

https://mafft.cbrc.jp/alignment/server/

https://ngs.geno2pheno.org/hcvrules

https://primerdigital.com/fastpcr.html)

https://sourceforge.net/projects/simulatepcr/

https://talk.ictvonline.org

https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification, 18th May 2019

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

https://www.biosearchtech.com/support/tools/design-software/realtimedesign-software

https://www.fda.gov/news-events/press-announcements/fda-approves-mavyret-hepatitis-c, 8th June 2019

https://www.geno2pheno.org/

https://www.hepatitis.va.gov/hcv/background/transmission-modes.asp, 26th May 2019

https://www.mybiosoftware.com/dnamate-1-0-consensus-melting-temperature-prediction-server-short-dna-sequences.html

https://www.ncbi.nlm.nih.gov/genbank/

https://www.viprbrc.org/brc/home.spg?decorator=flavi_hcv

https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 18th May 2019

https://www.who.int/news-room/fact-sheets/detail/hepatitis-c, 20th July 2019

ACKNOWLEDGMENT