UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Biologia

SCUOLA DI DOTTORATO DI RICERCA IN BIOSCIENZE E BIOTECNOLOGIE
INDIRIZZO GENETICA E BIOLOGIA MOLECOLARE DELLO SVILUPPO
CICLO XXVII

DEVELOPMENT OF NEW GENETIC TESTS: ATYPICAL CYSTIC FIBROSIS AND HCV GENOTYPING

Direttore della Scuola : Ch.mo Prof. Giuseppe Zanotti
Coordinatore d'indirizzo: Ch.mo Prof. Rodolfo Costa
Supervisore : Ch.ma Prof.ssa Maria Luisa Mostacciuolo

Dottorando : Alice Renesto

INDEX

INDEX 1
1. ABSTRACT 5
RIASSUNTO 7
2. FOREWORD 9
2. AIM OF THE STUDY 11
3. PART I: ATYPICAL CYSTIC FIBROSIS 12
3.1. INTRODUCTION 12
3.1.1. CFTR 12
3.1.2. Epidemiology of CFTR mutations 14
3.1.3. Type of CFTR mutations 14
3.1.4. Diagnosis 16
Clinical manifestation 17
Neonatal screening 17
Family history 17
Sweat test 17
CFTR mutation analysis 18
Nasal potential difference 18
3.1.5. Atypical CF 19
Disseminated bronchiectasis 19
Idiopathic chronic pancreatitis (ICP) 19
CBAVD 20
3.1.6. CFTR modifier genes 21
3.1.7. ENaC 22
Interaction between CFTR and ENaC 22
ENaC mutations and other diseases 23
ENaC mutations and atypical CF 24
3.2. MATERIALS AND METHODS 27
3.2.1. Design of ENac typing test prototype 27
3.2.2. PCR design 27
Design of primers 27
Design of PCR reaction 30
3.2.3. RLB design 31

Design of probes....31
Strip manufacturing....31
Design of RLB protocol....32
3.2.4. DNA sequencing....33
3.2.5. Statistical analysis....34
3.2.6. Samples....34
3.3. RESULTS AND DISCUSSION....36
3.3.1. Optimization of ENaC typing test prototype....36
3.3.2. Validation of ENaC typing test prototype....36
Healthy control group....36
Patients with respiratory manifestations....37
Patients with pancreatic manifestations....40
4. PART II: HCV GENOTYPING....43
4.1. INTRODUCTION....43
4.1.1. Genetic variability of HCV....43
4.1.2. Genotypes....43
Distribution of HCV genotypes....44
4.1.3. Viral structure....46
Virion structure....47
Structure and function of the viral proteins....47
4.1.4. Life cycle....50
Viral entry....50
Translations of viral proteins....50
Polyprotein processing....50
The viral replication complex....50
Packaging, assembly and particle release....51
4.1.5. Infections by HCV....51
Natural history of HCV....52
4.1.6. Therapy....53
Direct-acting antiviral agents....54
4.1.7. Diagnosis....57
Indirect diagnosis....57
Direct diagnosis....58

4.2. MATERIALS AND METHODS 61
4.2.1. Design of HCV genotyping test prototype 61
4.2.2. HCV sequences databases 61
4.2.3. Design of RT-PCR 62
Design of primers 62
RNA internal control (IC-RNA) 63
RT-PCR reaction design 64
4.2.4. RLB design 65
Design of probes 65
Strip manufacturing 67
Design of RLB protocol 67
4.2.5. Sequencing 68
4.2.6. Samples 68
4.2.7. RNA extraction 69
4.3. RESULTS 70
4.3.1. OPTIMIZATION OF HCV GENOTYPING TEST PROTOTYPE 70
Evaluation of enzyme mixture 70
Evaluation of RT-PCR efficiency 70
Primers evaluation and redesign 72
Starting materials evaluation 72
Evaluation of genotyping strip 72
Strip results interpretation 74
Evaluation of RLB protocol 77
4.3.2. VALIDATION OF HCV GENOTYPING TEST PROTOTYPE 77
Diagnostic specificity 77
Diagnostic sensitivity 78
5. DISCUSSION AND CONCLUSIONS 84
6. REFERENCES 89
7. AKNOWLEDGMENTS 102

1. ABSTRACT

Recent revolutionary progress in human genomics is reshaping the approach to therapy and diagnosis. Nucleic acid–based testing is the major diagnostic tool not only in the setting of inherited genetic disease but in a wide variety of neoplastic and infectious processes. The "Reverse Line Blot" (RLB) is a robust method for the molecular characterization of several disorders. It is a simple technique that can be used in diagnostic routine also by small laboratories, because the request instruments are few and not expensive.
The aim of our project is to exploit the potential of this method for assessment of two different disorders: a rare hereditary disease (atypical cystic fibrosis) and a hepatic pathology caused by the Hepatitis C Virus (HCV).

In the first part of the study we explore the potentiality of RLB in detecting known genetic mutations in the human genome and our interest was focused on atypical CF. This pathology is characterized by cystic fibrosis (CF) mild symptoms, involvement of few organs, and only one or none mutations in the CFTR gene. This phenotypic heterogeneity could be explained by modifier genes of the CFTR gene, such as the genes coding for the Epithelial Sodium Channel (ENaC). During this project, we developed a new RLB molecular assay for detection of nine already described variations in the ENaC genes, involved in CF-like disorders. The performance of the method were evaluated analyzing 208 patient samples, with respiratory or pancreatic manifestations, and 169 true healthy individuals as control group. Among the atypical CF patients, with one or none CFTR mutations and showing respiratory symptoms, we found seven ENaC different variants: the mutations p.S82C-SCNN1B, p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B, p.1670-2A>G-SCNN1B, and the polymorphisms p.W493R-SCNN1A and p.R181W-SCNN1A. In particular, p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B, p.1670-2A>G-SCNN1B showed a frequency significantly higher in patients rather in controls ($P<0.05$), while for the other variants there was no significantly differences between the two groups. Instead, in the group of patients with pancreatitis, we found only 3 variants (p.W493R-SCNN1A, p.R181W-SCNN1A, p.S82C-SCNN1B), but without significantly differences between patient group and controls.

In the second part of the study, we explore the potentiality of RLB in the identification of the genotypic strand of Hepatitis C virus. HCV is a single-strand positive RNA virus, which strains can be classified into 7 genotypes. Knowing HCV genotype is essential for the therapeutic management and it is related to antiviral therapy outcome. The 5' untranslated region (5' UTR) of the HCV genome is used as target region in most genotyping test. However it is not enough informative for subtyping genotype 1, distinguishing 1a and 1b, and for discrimination of genotypes 1 from 6 (subtypes c-l). During this project, we developed a molecular assay for HCV genotyping, which allows reliable distinction between genotypes and subtypes. The method is based on reverse-transcription of two viral regions (5'UTR and core), followed by RLB genotyping. For the performance evaluation of the developed assay, 283 serum samples were collected: 212 HCV positive, 71 HCV negative and 8 synthetic samples (representative of less common

genotypes 5, 6 and 7a). Another CE-IVD genotyping assay and the sequencing of NS5b viral region were used as reference methods. Our results showed that the use of two viral regions for genotyping allows a better discrimination of subtypes 1a and 1b. Among 145 genotype 1 samples, 22 samples (whose subtyping 1a or 1b was not solved by the use of the sole 5'UTR) were correctly typed as 1a or 1b. The genotyping of other 8 samples, previously incorrectly subtyped, were solved. Moreover, the new assay has successfully genotyped 5 samples with genotype 6 that were misdiagnosed as genotype 1 with the analysis of 5'UTR. Our assay showed a diagnostic specificity of 100% and a sensibility of 98.6% both at genotype and subtype levels, in a range of viral loads between 1×10^3 and 6×10^7 IU/mL.
Therefore, combined analysis of both 5'UTR and core regions could be very useful for a better discrimination between subtypes 1a and 1b, and between genotypes 1 and 6, leading to a more powerful instrument for the application of appropriate therapeutic regimen.

In conclusion, this work has demonstrated that RLB is a robust method for the management of several diseases, requiring a genetic characterization. However, genetic knowledge about target disorders must to be well defined.

RIASSUNTO

I recenti progressi nel campo della genomica umana hanno fortemente influenzato l'approccio diagnostico e terapeutico alle patologie. La biologia molecolare si è affermata come strumento di elezione per la diagnosi non solo di malattie ereditarie ma anche di una grande varietà di processi neoplastici ed infettivi. Tra le metodologie di questa disciplina, il "Reverse Line Blot" (RLB) occupa un posto di rilievo, grazie alla sua versatilità. E' una tecnologia molto semplice e può essere utilizzata per la routine diagnostica anche da piccoli laboratori, in quanto la strumentazione necessaria è limitata e non eccessivamente costosa. L'obiettivo del nostro progetto è stato quello di sfruttare le potenzialità di questa tecnologia applicandola allo studio di due diverse patologie: una malattia ereditaria (fibrosi cistica atipica) ed una patologia epatica causata dall'infezione del virus dell'epatite C (HCV).

Nella prima parte dello studio, abbiamo esplorato le potenzialità della tecnologia RLB per l'identificazione di mutazioni genetiche note, concentrandoci sullo studio della fibrosi cistica atipica. Si tratta di una patologia caratterizzata dagli stessi sintomi della fibrosi cistica (FC), che si presentano però in forma più lieve, con il coinvolgimento solamente di alcuni organi, ed una o nessuna mutazione nel gene CFTR. Questa eterogeneità fenotipica può essere dovuta all'azione di geni modificatori di CFTR, come, per esempio, quelli che codificano il canale epiteliale del sodio (ENaC). Nel corso di questo progetto è stato sviluppato un nuovo test molecolare basato sulla metodologia RLB, per la rilevazione di 9 varianti nei geni ENaC, coinvolte nella FC atipica e precedentemente descritte da altri gruppi. Le performance del metodo sono state valutate analizzando 208 campioni di pazienti con sintomi respiratori o manifestazioni pancreatiche, e 169 individui sani del gruppo di controllo. Tra i pazienti con sintomi respiratori e affetti da FC atipica, recanti una o nessuna mutazione CFTR, abbiamo trovato 7 diverse varianti ENaC: le mutazioni p.S82C-SCNN1B, p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B, p.1670-2A>G-SCNN1B, ed i polimorfismi p.W493R-SCNN1A e p.R181W-SCNN1A. In particolare, p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B, p.1670-2A>G-SCNN1B mostravano una frequenza significativamente più alta nei pazienti rispetto ai controlli ($p < 0,05$), mentre per le altre varianti non abbiamo rilevato differenze significative tra i 2 gruppi. Invece nel gruppo di pazienti con pancreatiti, abbiamo identificato 3 varianti ENaC (p.W493R-SCNN1A, p.R181W-SCNN1A, p.S82C-SCNN1B), ma le loro frequenze non erano significativamente diverse tra il gruppo degli affetti e quello dei controlli.

Nella seconda parte dello studio, abbiamo esplorato le potenzialità della tecnologia RLB per l'identificazione dei ceppi genotipici del virus dell'epatite C. HCV è un virus a RNA, a singolo filamento positivo, i cui ceppi possono essere classificati in 7 genotipi. L'identificazione del genotipo di HCV è essenziale per la gestione terapeutica ed è correlato all'esito della terapia antivirale. La regione non tradotta al 5' (5' UTR) del genoma di HCV viene utilizzata come target nella maggior parte dei test di genotipizzazione. Tuttavia tale regione, sebbene altamente conservata, si è rivelata non

sufficientemente informativa per la distinzione dei sottotipi HCV 1a e 1b, e per la discriminazione dei genotipi 1 dal genotipo 6 (sottotipi da c a l).
Nel corso di questo progetto è stato sviluppato un test molecolare per la genotipizzazione di HCV, che consente una buona distinzione tra genotipi e sottotipi. Il metodo è basato sulla retrotrascrizione di due regioni virali (5'UTR e core), seguita dalla genotipizzazione in RLB. Per la valutazione delle performance del nostro test sono stati analizzati 283 campioni di siero: 212 erano positivi per HCV, 71 erano negativi, mentre 8 erano campioni sintetici, e corrispondevano ai rari genotipi 5, 6 a 7a. Come metodi di riferimento sono stati utilizzati un dispositivo CE-IVD che analizza la sola regione 5'UTR ed il sequenziamento della regione virale NS5b.
I risultati ottenuti hanno dimostrato che l'utilizzo di due regioni virali per la genotipizzazione consente una migliore discriminazione dei sottotipi 1a e 1b. Su un totale di 145 isolati con genotipo 1, 22 campioni, non sottotipizzati mediante la sola analisi della 5'UTR, sono stati correttamente tipizzati come 1a o 1b grazie all'analisi combinata 5'UTR e core. Inoltre il nostro metodo ha permesso di effettuare la sottotipizzazione di altri 8 campioni, per i quali l'analisi della sola 5'UTR aveva assegnato una tipizzazione non corretta. Per quanto riguarda i campioni con genotipo 6, il test sviluppato è stato in grado di tipizzarne correttamente 5, che erano stati erroneamente definiti come genotipo 1 mediante l'analisi della sola regione 5'UTR. Il nostro test ha evidenziato una specificità diagnostica del 100% e una sensibilità del 98,6%, sia a livello di genotipizzazione che di sottotipizzazione, in un range di carica virale compreso tra 1 x 10^3 and 6 x 10^7 UI/mL. Pertanto, l'analisi combinata delle due regioni 5'UTR e core determina un'accurata discriminazione tra i sottotipi 1a e 1b, e tra i genotipi 1 e 6, e diventa uno strumento molto importante per l'applicazione del regime terapeutico più appropriato.

In conclusione, il nostro lavoro ha dimostrato che la tecnologia RLB è un metodo affidabile per la diagnosi di molte patologie, per le quali si renda necessaria una caratterizzazione genetica. Va però sottolineato che dispositivi basati sulla metodologia RLB possono diventare validi strumenti diagnostici solo per malattie ben caratterizzate da un punto di vista genetico.

2. FOREWORD

This project concerns about the work that I carried out at AB ANALITICA s.r.l. in the context of an advanced apprenticeship program for the acquisition of a Ph. D. qualification. AB ANALITICA is a certified company established in 1990 and specialized in the design, development, production and trade of *in vitro* medical diagnostic devices (IVD) of diagnostic molecular systems. Fields of interest are microbiology, virology, onco-hematology and genetics.

Recent revolutionary progress in human genomics is reshaping the approach to therapy and diagnosis. Nucleic acid–based testing is becoming a crucial diagnostic tool not only in the setting of inherited genetic disease, but in a wide variety of neoplastic and infectious processes. Following diagnosis, molecular testing can help guide to appropriate therapy by identifying specific therapeutic targets of several newly tailored drugs, thus playing an integral role in the application of pharmacogenomics. Molecular diagnostics provides the necessary support for any successful application of gene therapy or biologic response modifiers. It offers a great tool for assessing disease prognosis and therapy response and detecting minimal residual disease [1].

Molecular diagnostic techniques for viral testing have experimented a rapid development during last years, and have been introduced in the majority of laboratories as a new way for the diagnosis of human pathogens like viruses. The use of amplification techniques such as polymerase chain reaction (PCR), real-time PCR or reverse line blot (RLB) for virus detection, genotyping and quantification have some advantages like high sensitivity and reproducibility, as well as a broad dynamic range.

Molecular diagnostic methods for the inherited genetic disease can be divided in 2 main groups: direct techniques that test for known mutations (genotyping) and indirect techniques that scan for any mutation in a particular target region (mutation scanning). There are several well-known genotyping and scanning methods in routine diagnostic use.

A direct technique for genotyping is the Reverse Line (or Dot) Blot (RLB or RDB), a relatively simple and less expensive method that can be used in diagnostic routine also by small laboratories. Reverse allele specific oligonucleotide assays provide a robust method for the molecular characterization of several disorders, caused by genetic mutations or by infective agents. RLB rely upon highly multiplexed PCR reactions containing biotinylated primers and upon the hybridization between amplification products and oligonucleotides bound on a nylon or nitrocellulose membrane. The detection system is based on a colorimetric reaction, catalysed by strepatvidin alkaline phosphatase or by streptavidin horseradish peroxidase.

Briefly the protocol involves the extraction of DNA or RNA from the specimen and then the selectively amplification of the sample. A strip is coated with highly specific probes, which are complementary to the amplified nucleic acid sequences. The amplification products are chemically denatured, and they are specifically bound by the complementary probes during the hybridization step. Non-specifically bound products are removed in subsequent washing steps. In the following step, bound product is

labelled with the enzyme alkaline phosphatase and made visible by a colorimetric reaction. In this way, a specific band pattern appears on the strip (fig. 1-2).

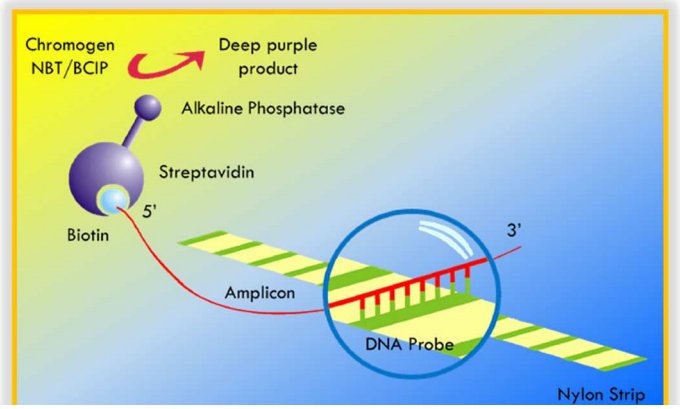


Fig. 1 Schematic representation of the reverse line blot technique.

The technique can be utilized in a number of ways. Multiple probes can be designed to detect sequence variation within a single amplified product, or multiple targets can be amplified simultaneously, with one (or more) probes used for subsequent detection. A combination of both approaches can also be used within a single assay. The ability to include multiple probes for a single target sequence makes the assay highly specific. The flexibility of the method allows its use for a wide range of applications.

The entire procedure from patient's sample to result usually requires less than 1 day. Large numbers of individuals can be rapidly screened for multiple targets using this technology. The method is simple, rapid, and generally shows high sensitivity and specificity. Commercial instruments are available, that automate the hybridization and colour development. In addition, scanning software can capture the probe reactivity pattern and interpret it in terms of a genotype.

RLB is rapidly becoming a standard molecular tool for diagnostic and epidemiological studies in an increasing number of laboratories all over the world, thanks to its peculiar features.

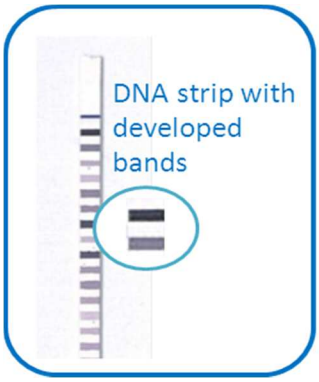


Fig. 2 Image of a developed strip.

RLB is a method that can be applied to several fields, such as human genetics, microbiology and parasitology.

2. AIM OF THE STUDY

The aim of our project was the development of RLB tests and the evaluation of their application in the management of diseases, requiring a genetic characterization.
We focused our attention on two different disorders, one regarding a hereditary disease and the other caused by an infectious agent.
In the first part of the study, we explored the potentiality of RLB in detecting known genetic mutations in the human genome and our interest was focused on atypical cystic fibrosis. This is a rare genetic disease, not well genetically characterized and probably related to the presence of mutations in the ENaC genes. Knowledge about causative mutations, in this case, is very limited.
In the second part of the study, we evaluated usefulness of RLB in the identification of the genotypic strand of Hepatitis C virus (HCV), an infectious agent that causes chronic hepatitis, cirrhosis and hepatocellular carcinoma. In this case, information about HCV genotypes is widely defined.
Our purpose was to evaluate potentiality of RLB in the application to clinical diagnosis and therapy management of both diseases, different not only in the transmission way but also in the genetic knowledge available for each one.

3. PART I: ATYPICAL CYSTIC FIBROSIS

3.1. INTRODUCTION

Cystic fibrosis (CF) is the most common autosomal disorder in Caucasians, with a frequency of about 1 in 2500 live births [2].
This pathology affects the chloride transport of polarized epithelia cells. Abnormally viscous secretions in the airways of the lungs cause obstructions that lead to inflammation and subsequent chronic bacterial infections by pathogens such as *Pseudomonas aeruginosa, Staphylococcus aureus* and *Burkholderia cepacia.* Finally, the chronic inflammation lead to tissue damage and destruction of the organ system; obstructive lung disease is currently the primary cause of morbidity and is responsible for about 80% of mortality in these patients.
Other organ systems containing epithelia, such as ducts of the pancreas, the sweat gland, biliary duct of the liver, the male reproductive tract and the intestine, are also affected. Loss of pancreatic exocrine function results in malnutrition and poor growth, which leads to death in the first decade of life for most untreated individuals [3].
98% of men with cystic fibrosis are infertile, with aspermia secondary to atretic or absent *vasa deferentia* and dilated or absent seminal vesicles. Female reproductive function is normal, although cervical mucus can be dehydrated, which might impair fertility [2].

3.1.1. CFTR

Cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane regulator (CFTR) gene, which has been identified in the 1989 by positional cloning. It is located on long arm of chromosome 7 and it is composed by 27 exons, spanning over 230 kb [4] (fig. 3).

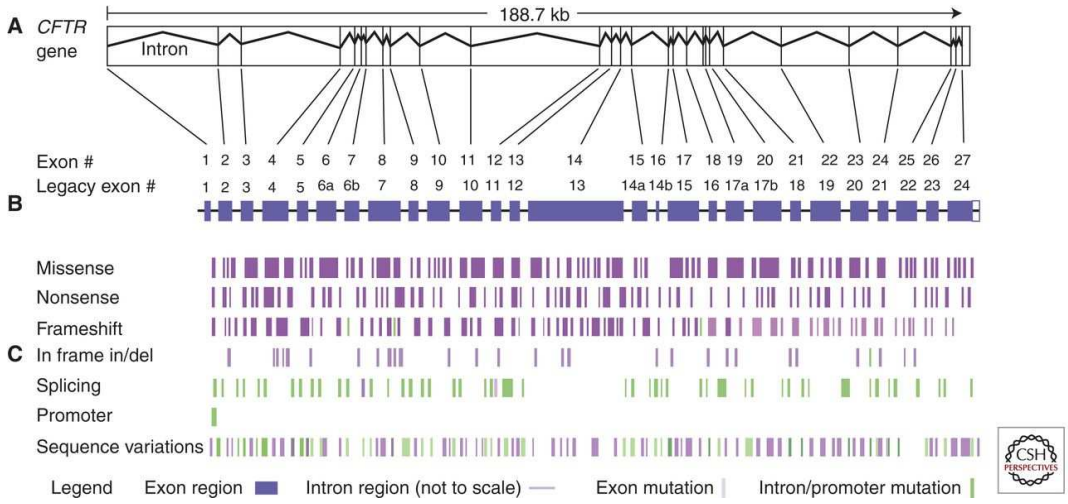


Fig. 3 Scheme of CFTR gene, transcript, and mutation distribution. (A) Scaled schematic of exons and introns followed by current and historical exon numbering. (B) A rescaled exon scheme with minimized introns, which are drawn not to scale. (C) Distribution of known mutations and polymorphisms as vertical bars [5].

The predicted structure of the CFTR, based on aminoacid sequence, is composed by twelve transmembrane domains, two nucleotide binding motifs and a regulatory R domain in the center [6]. These topological features associated the CFTR with the members of the ABC (ATP binding cassette) superfamily of transporters (fig. 4).
Protein purification and gene transfer studies have demonstrated that CFTR functions as a chloride channel regulated by cyclic AMP (cAMP)-dependent phosphorylation [7]. It is expressed in the apical membrane of the epithelial cells.

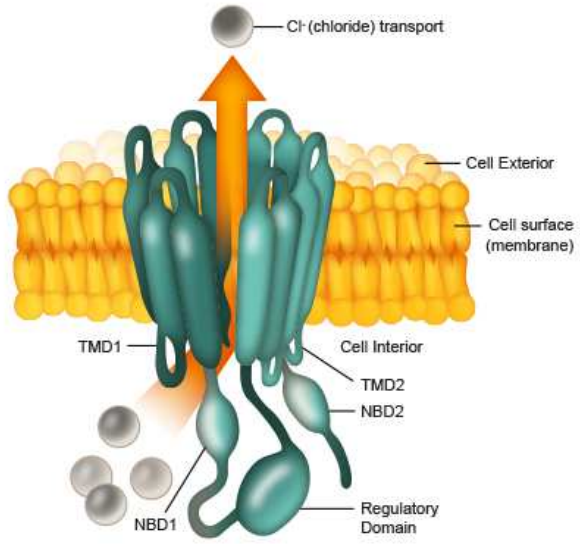


Fig. 4 Structure of CFTR protein

3.1.2. Epidemiology of CFTR mutations

Almost 2000 variants have been reported to the Cystic Fibrosis Mutation Database, one of the first locus-specific databases.

The most common mutation is F508del, which accounts for approximately two thirds of all CFTR alleles in patients with CF, with a decreasing prevalence from Northwest to Southeast Europe. The remaining third of alleles is substantially heterogeneous, with fewer than 20 mutations occurring at a worldwide frequency of more than 0.1% (tab. 1). In Europe, on average 1 in 2000-3000 newborns are affected with CF. Even where populations appear relatively homogeneous, there may be marked local and regional variations.

CFTR gene mutations have been well characterized in most European populations. F508del frequencies vary from a maximum of 100% in the isolated Faroe Islands of Denmark, to a minimum of about 20% in Turkey. In central, northern, western, and north-eastern Europe, F508del has a frequency of about 70%. Apart from F508del, 5 to 10 relatively frequent mutations contribute to 10%-15% of all CF-causing CFTR mutations, such as the G542X, N1303K, and G551D mutations. Ethnic specific mutations are observed in some populations such as the Nordic mutation 394delTT, the 3905insT mutation in Switzerland, the R1162X mutation in Northeast Italy, and the Eastern Slavic CFTR dele 2,3 (21kb) mutation. The remaining mutations are heterogeneous, private, or limited to a small number of individuals [8].

E60X	Southern European	S549N	Indian
CFTR	Slavic — Eastern European	G551D	United Kingdom, Central Europe
R75X	Southern European, US-Hispanic	Q552X	Southern European, Italian
394delTT	Nordic — Baltic sea region	R553X	Central European
G85E	Southern Europe	A559T	African-American
406-1G>A	US-Hispanic	R560T	Northern Irish
R117H	European-derived populations	1811+1.6kbA>G	Spanish, US-Hispanic
R117C	Northern European	1898+1G>A	United Kingdom, Central Europe
621+1G>T	Southern European	1898+5G>T	East Asian populations
711+1G>T	French, French Canadian	2143delT	Slavic — Eastern European
711+5G>A	US-Hispanic	2183delAA>G	Southern Europe, Middle Eastern, Iranian, Latin American
L206W	Spanish and US-Hispanic	2184delA	European-derived populations
V232D	Spanish and US-Hispanic	2789+5G>A	European-derived populations
1078delT	French Brittany	Q890X	Southern European
R334W	Southern European, Latin American	3120+1G>A	African, Arabian, African-American, Southern Europe
1161delC	Indian	3272–26A>G	European-derived populations
R347P	European-derived, Latin America	3659delC	Scandinavian
R347H	Turkish	3849+10kbC>T	Ashkenazi-Jewish, Southern European, Middle Eastern, Iranian, Indian
A455E	Dutch	R1066C	Southern European
1609delCA	Spanish, US-Hispanic	Y1092X (C>A)	Southern European
I506T	Southern European, Spanish	M1101K	US-Hutterite
I507del	European-derived populations	3905insT	Swiss
F508del	European-derived populations	D1152H	European-derived populations
1677delTA	Southern European, Middle Eastern	R1158X	Southern European
1717–G>A	European-derived populations	R1162X	Italian, Amerindian, Latin America
V520F	Irish	S1251N	European-derived populations
G542X	Southern European, Mediterranean	W1282X	Ashkenazi-Jewish, Middle Eastern
S549R(T>G)	Middle Eastern	N1303K	Southern European, Middle Eastern

Tab. 1 Geographical distribution of the most common CFTR mutations [9].

3.1.3. Type of CFTR mutations

Among the reported 2000 variants, 40% are predicted to cause substitution of a single aminoacid, 36% are expected to alter RNA processing (including nonsense, frameshift and mis-splicing variants), about 3% involve large rearrangements of CFTR, and 1% affects promoter regions; 14% seem to be neutral variants, and the effect of the

remaining 6% is unclear. Disease-causing variants can affect the quantity and/or function of CFTR at the cell membrane [3].

Historically, CFTR variants have been grouped into five functional classes.

1. The I class comprise nonsense, frameshift and splice junction mutations, that lead to nonfunctional proteins, rapidly degraded.
2. Mutations of II class are associated to defective processing and maturation of the CFTR protein, that is degraded before to reach plasmatic membrane. They are missense variants. This class includes the F508del.
3. The III class comprises missense mutations that lead to defective regulation of the channel. They are located in the ATP binding domain or in site of phosphorylation into the R domain and cause suppression of CFTR function.
4. Mutations of class IV are associated to defective conductance of the chloride. These are missense mutations that lead to functional CFTR protein, but with minor activity [10].
5. The V class includes splicing defects and missense mutations, leading to a reduced function and/or synthesis of CFTR channel (fig. 5).

In general, patients homozygous for class I–III mutations exhibit a phenotype associated with pancreatic insufficiency, higher frequency of meconium ileus, premature mortality, earlier and more severe deterioration of lung function, higher incidence of malnutrition and severe liver disease. Class IV–V mutations are usually associated with milder lung disease, older age at death, pancreatic sufficiency. Class IV–V mutations are phenotypically dominant when occurring in combination with class I–III mutations [11].

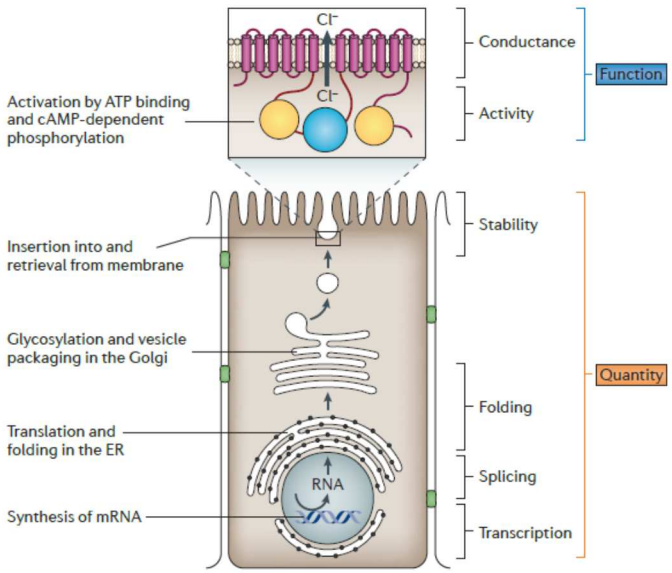


Fig. 5 Molecular consequences of variants in CFTR. The degree to which epithelial ion transport is altered in an individual with cystic fibrosis is determined by the effect of each disease-causing variant on the quantity and the function of cystic fibrosis transmembrane conductance regulator (CFTR) [3].

CFTR mutations can be clustered also according to their predicted clinical consequences [9]. They are divided into 4 groups:

a. mutations that cause CF disease;
b. mutations that results in a CFTR-related disorder;
c. mutations with no known clinical consequences;
d. mutations of unproven or uncertain clinical relevance.

There is some overlapping of groups A and B, as some mutations may sometimes be detected in association with pancreatic sufficient CF, some other times with CFTR-related, mono-symptomatic disorders (tab. 2). However, these relationships between certain classes of CFTR mutations and phenotypes are looser than expected and the phenotypic consequences of a significant number of CFTR gene changes remain unclear. Factors such as the age related progression of the disease, the environment, and modifier genes, all play a role in the clinical heterogeneity of patients carrying these "borderline" mutations [9].

Mutation group	Examples
A. CF-causing	F508del, R553X, R1162X, R1158X, 2184delA, 2184insA, 3120+1G>A, I507del, 1677delTA, G542X, G551D, W1282X, N1303K, 621+1G>T, 1717-1G>A, A455E, R560T, G85E, R334W, R347P, 711+1G>T, 711+3A>G*, 1898+1G>A, S549N, 3849+10kbC>T, E822X, 1078delT, 2789+5G>A, 3659delC, R117H-T5 (*),R117H-T7 (*), D1152H (*), L206W (*), TG13-T5 (*)
B. CFTR-related disorders associated	R117H-T7(*), TG12-T5(*), R117H-T5 (*), D1152H (*), TG13-T5 (*), S997F, R297Q*, L997F, M952I, D565G*, G576A*, TG11-T5**, R668C-G576A-D443Y, R74W-D1270N
C. No clinical consequences	I148T, R75Q, 875+40A/G, M470V, E528E, T854T, P1290P, 2752-15G/C, I807M, I521F, F508C, I506V, TG11-T5**
D. Unknown or uncertain clinical relevance	Mainly missense mutations***

Tab. 2 Examples of CFTR mutations with regard to their clinical consequences.
(*) mutations which may belong either to Group A or to Group B.
(**) mutations which may belong either to Group B or to Group C.
(***) certain common sequence (missense) variants with subclinical molecular consequences (e.g. M470V) may co-segregate on the same chromosome and exert more potent, cumulative phenotypic effect [9].

3.1.4. Diagnosis

The diagnosis of CF is suggested by one or more clinical signs and confirmed by an increased sweat chloride concentration.
In 1998, a first consensus statement listing criteria for the diagnosis of CF was issued by the US Cystic Fibrosis Foundation [12].
The fundamental criteria were: (i) one or more of the phenotypic features of the disease or (ii) a history of CF in a sibling or (iii) a positive immunoreactive trypsin (IRT, a neonatal screening test), in association with at least one other feature. The additional features

included a positive sweat test result on two occasions, a CF-causing mutation in each CFTR gene or an abnormal nasal potential difference (NPD) [12].
In 2006 a European working group proposed a modified consensus statement, by focusing on different diagnosis of classic CF and atypical manifestation of CF. They suggested two diagnostic alghoritms, starting from different fundamental criteria exhibited by patients [13].
The CF diagnostic cascade start with the finding of one of these three situations: (1) clinical manifestations; (2) neonatal screening; and (3) family history.

Clinical manifestation

Many clinical problems are compatible with a diagnosis of CF, because this multiorgan disease is very heterogeneous and it has, at times, an atypical clinical presentation.
The phenotypic features consistent with a diagnosis of CF are:

1. Chronic sinopulmonary disease, manifested by persistent colonization/infection with typical CF pathogens, chronic cough and sputum production, persistent chest radiograph abnormalities, airway obstruction and nasal polyps.
2. Gastrointestinal and nutritional abnormalities, including intestinal manifestations, pancreatic disease, hepatic symptoms.
3. Salt loss syndromes.
4. Male urogenital abnormalities resulting in obstructive azoospermia (CBAVD) [12].

Neonatal screening

CF neonatal screening is based on the immunoreactive trypsinogen (IRT) assay, which is relatively inexpensive and adaptable to large numbers. Increased IRT concentrations at birth are characteristic of newborns affected by CF, but can also be found in healthy infants [13].

Family history

It is strongly advised that siblings of affected children are investigated by a sweat test. Because of the remarkable clinical heterogeneity, even within families, lack of symptoms is insufficient to exclude the diagnosis of CF.

The two different diagnostic CF algorithms proposed by Boeck *et al.* could be applied in the presence of one of the above mentioned three situations. In settings where the sweat test is available, the approach starting with this analysis is the preferred route. In patients with atypical disease manifestations the reliability of the sweat test is much lower and additional diagnostic tests will be necessary to substantiate the diagnosis. Thus, the algorithm starting with DNA analysis could be more appropriate.

Sweat test

The sweat test retains its position as a key diagnostic test for CF: the only acceptable procedure is the quantitative pilocarpine iontophoresis sweat test.

A sweat chloride concentration of more than 60 mmol/L is consistent with the diagnosis of CF, but the result must be interpreted in the context of the patient's age and clinical picture by a physician knowledgeable about CF. On the other hand, levels below 40 mmol/L are considered normal, but with a cautious interpretation in infancy and prematurity. The diagnosis of CF should be made only if the elevated sweat chloride concentration (>60 mmol/L) is revealed on two separate occasions [12].

CFTR mutation analysis

The analysis for the identification of CFTR mutations could be divided into 3 categories:

1. The analysis of first level identify specific or most common mutations, through technologies such as Reverse Dot Blot (RDB), Oligo Ligation Assay (OLA) or Amplification Refractory Mutation System (ARMS). Several commercial assays are available for CFTR mutation screening. Most tests only screen for about 30 mutations, the majority of which are associated with classic CF.
2. The second level analysis examine large portions of the gene to search every type of eventually present variants: they includes Denaturing HPLC (D-HPLC), Denaturing Gradient Gel Electrophoresis (DGGE) analysis, sequencing of the exons and intron-exon junctions of CFTR gene. They can be used in case of borderline sweat test and negative first level analysis.
3. Finally, the third level analysis could identify rare rearrangements and intron variants not so close to exons and that cause splicing alterations. The technologies used are Multiplex Ligation-dependent Probe Amplification (MLPA), Quantitative Multiplex Polymerase chain reaction of Short Fluorescent fragments (QMPSF) and the study of nasal epithelia CFTR mRNA [14].

Only sequencing will approach 100% sensitivity. The other techniques, such as indirect mutation scanning assays, allow sensitivities varying from close to 100% to as low as 90%. Even if a mutation is found, its involvement in disease may not be clear. For many CFTR mutations the functional consequences are unknown; they may even be polymorphisms [13].

Nasal potential difference

Respiratory epithelia, including nasal epithelia, regulate the composition of fluids that wet airway surfaces by transport of ions such as sodium (Na+) and chloride (Cl–). This active transport of ions generates a transepithelial electrical PD, which can be measured in vivo. Abnormalities of ion transport in respiratory epithelia of patients with CF are associated with a different pattern of nasal PD compared with normal epithelia [12].

3.1.5. Atypical CF

There is great heterogeneity in the clinical manifestations of cystic fibrosis. Some patients may have all the classical manifestations of CF from infancy and have a relatively poor prognosis, while others have much milder or even atypical disease manifestations and still carry mutations on each of the CFTR genes. The CF phenotype is a *continuum* of symptoms and cannot be easily defined in two distinct disease categories (typical versus atypical CF).

It has been demonstrated that, apart from their involvement in classical CF, CFTR mutations also cause, or contribute to, the CFTR-Related Disorders and so-called atypical CF. Patients with non-classic or atypical CF have a CF phenotype in at least one organ system and a normal (40 mmol/l) or borderline (40–60 mmol/l) sweat chloride level. Most of these patients has exocrine pancreatic sufficiency and milder lung disease [13]. A CFTR-related Disorder (CFTR-RD) is defined as a clinical entity associated with CFTR dysfunction that does not fulfill the diagnostic criteria for CF. Three main clinical entities illustrate these phenotypes: disseminated bronchiectasis, acute recurrent or chronic pancreatitis and congenital bilateral absence of the vas deferens (CBAVD) [15].

Disseminated bronchiectasis

Bronchiectasis is a pathological description of lung damage characterized by an abnormal and irreversible dilatation of thick-walled bronchi. Affected areas are inflamed and easily collapsible, resulting in airflow obstruction and impaired clearance of secretions. Symptoms include recurrent lower respiratory tract infections, chronic cough and mucopurulent sputum production. In approximately 50% of cases, bronchiectasis is associated with underlying conditions such as CF, childhood infections, allergic broncho-pulmonary aspergillosis, immune defects, primary ciliary dyskinesia, aspiration of irritants, ulcerative colitis, rheumatoid arthritis and other connective tissue disorders. In the remainder of cases, causative factors cannot be identified [16]. An increased incidence of CFTR gene mutations has been found in bronchiectasis. At least one CFTR mutation was reported in 10–50% of a series of patients in different studies [17, 18]. Two mutations were found in 5–20% of cases, but not all studies specified whether a segregation analysis had been performed to establish if those subjects carried the two mutations in *cis* or in *trans*. Often, in these patients only one mutation is CF-causing. No specific CFTR mutation is associated directly with bronchiectasis. Instead, a wide spectrum of CFTR mutations have been identified, most being uncommon and likely to result in residual CFTR function [18, 19]. The variety of CFTR mutations associated with bronchiectasis reflects the heterogeneous nature of this condition [15].

Idiopathic chronic pancreatitis (ICP)

Chronic pancreatitis (CP) is a progressive inflammatory disease in which pancreatic secretory parenchyma is destroyed and replaced by fibrous tissue, eventually leading to impairment of the exocrine and endocrine functions of the organ [20].

About 10-30% of patients with chronic pancreatitis do not have an apparent underlying cause, and disease is classified as idiopathic pancreatitis [21].

Genetic studies led to identification of a number of potential defects in genes, involved into pancreatic activities, which may contribute to the development of pancreatitis.
In 1998 a research group reported an association between CFTR mutations and idiopathic chronic pancreatitis (ICP), by starting from three evidences: (i) both ICP and CF pancreatic disease show early ductal plugging resulting from inspissated secretions, (ii) CF patients occasionally suffer from pancreatitis and (iii) chronic pancreatitis is a known cause of false-sweat tests [22].
In the human exocrine pancreas, CFTR is predominantly expressed at the apical membrane of the ductal and centro acinar cells that line small pancreatic ducts, and controls cAMP-stimulated HCO^{-3} secretion into the duct lumen. The major role of CFTR in pancreatic ducts is to dilute and alkalinize the protein-rich acinar secretions, thereby preventing the formation of protein plugs that predispose to pancreatic injury [23].
About 30% of patients with ICP or recurrent acute pancreatitis are found to carry CFTR mutations. No specific CFTR mutations are associated with ICP, but rare or private class 4 or class 5 mutations are generally found in these patients. CF-causing mutations, F508del being the most common, that generally have < 2% of normal CFTR function, lead typically to pancreatic insufficiency in homozygotes. In contrast, CF patients with genotypes producing about 5% of normal CFTR function often have pancreatic sufficiency [24].
Recently, in a group of 33 patients with recurrent pancreatitis, Segal *et al.* [25] found that seven (21%) had an abnormal NPD even though their sweat chloride concentration and mutation profile did not differ from control subjects.

CBAVD

Congenital bilateral absence of the vas deferens (CBAVD) in otherwise healthy males accounts for approximately 3% of cases of infertility. The incidence of CBAVD, based on estimations, is approximately 1:1000 males [26]. While the prevalence of CF is very low in non-Caucasian countries, the prevalence of CBAVD does not seem to differ between populations.
In the majority of cases, isolated CBAVD is recognized as an autosomal recessive genetic disorder associated with anomalies of the CFTR gene: alterations that retain enough residual CFTR function might result in milder phenotypes such as CBAVD.
CBAVD is caused by mutations in the two copies of the CFTR gene in 70–90% of cases depending on ethnic/geographic populations [27]. Thereby, the distribution of CFTR mutations and genotypes in CBAVD differs substantially from classical CF. Among males with two identified CFTR mutations, CF patients have either two severe (88%) or one severe and one mild/variable CFTR mutations (12%), whereas CBAVD males have either a severe and a mild/variable (88%) or two mild/variable (12%) CFTR mutations [28]. The two most common compound heterozygous genotypes found in European males with CBAVD are F508del in *trans* with IVS8-5T (28%) and F508del in *trans* with R117H (6%).
CFTR gene defects in CBAVD are essentially point mutations. However, in a very small number of cases large rearrangements (deletions or duplications) within the CFTR locus are identified [29].

Several studies provide evidence for genetic heterogeneity in CBAVD: a proportion of extensively studied men with CBAVD, variable depending on country of origin (6–15% in Europe [29]), do not display any abnormalities in the CFTR gene.

3.1.6. CFTR modifier genes

In the vast majority of classical CF patients both copies of the CFTR gene are mutated. However, in a minority of CF patients a mutation cannot be identified on both CFTR genes, i.e., in 1 to 2% of the CF patients of Northern Europe and in up to 8 to 10% in CF patients of Southern Europe (www.who.int/genomics/publications/en). In an even higher proportion of patients with CBAVD, atypical CF, chronic pancreatitis, or disseminated bronchiectasis, a mutation cannot be identified on both CFTR genes [28, 30].

There is substantial evidence that genes other than CFTR may cause CF or CF-like disease. Indeed, a German family was described with a CF patient without CFTR mutations, and with an unaffected healthy sister who inherited the same CFTR genes from her parents [31]. In two American families, each with two affected sibs, no mutations could be found on both CFTR genes, the affected sibs inherited different parental CFTR genes [32].

Extensive understanding of CF pathophysiology presents an opportunity to interrogate candidate genes as potential modifiers (fig. 6). In the lungs, loss of CFTR leads to inflammation, neutrophil recruitment, tissue damage and replacement with fibrotic connective tissue. At least 50 genes encoding proteins that participate in these cellular and tissue functions have been investigated as candidate modifiers [33].

Moreover, apart from the defective chloride secretion, it has been demonstrated that loss of functional CFTR leads to an increased sodium absorption in the airways of CF patients. Sodium transport is mediated through the amiloride sensitive epithelial sodium channel (ENaC) [34]. Thus, ENaC could be considered a good candidate as CFTR modifier gene.

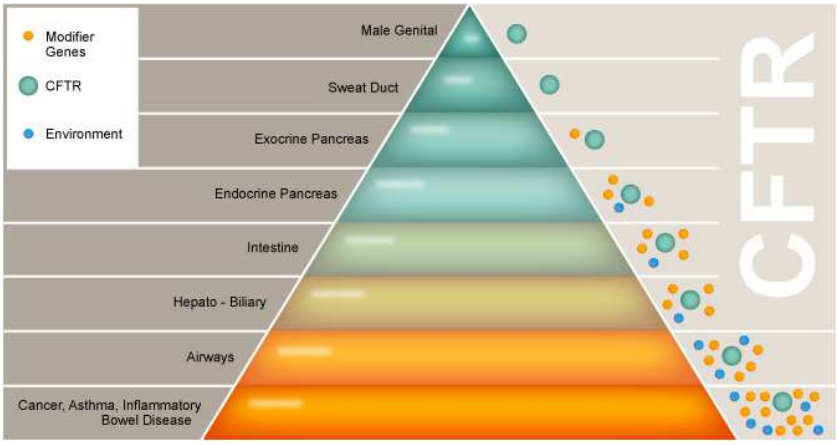


Fig. 6 The relative contribution of modifier genes, CFTR, and environment on phenotype [35].

3.1.7. ENaC

The ENaC proteins were identified by expression cloning in the early 1990s [36].
The ENaC subunits belong to the degenerin/ENaC family of ion channels, which fulfills a key role in Na^+ and water homeostasis.
ENaC forms a heteromeric channel composed of three homologous subunits α, β and γ. Each subunit has two trans-membrane domains, a large extracellular loop and cytosolic N- and C-termini. There are multiple ENaC proteins expressed in various epithelia, with the prototypical ENaC thought to consist of at least 1α, 1β, and 1γ ENaC subunit interacting to form a channel (fig. 7). Of these, the α-ENaC subunit is required for a functional channel, while β-ENaC or γ-ENaC alone do not appear to form a conducting protein [36].
The three subunits are encoded by three different genes: SCNN1A (sodium channel nonvoltage-gated 1, alpha, 12p13), SCNN1B (sodium channel nonvoltage-gated 1, beta, 16p12.2-p12.1) and SCNN1G (sodium channel nonvoltage-gated 1, gamma, 16p12).
ENaC proteins are found in the apical membrane of sodium-absorbing epithelial cells, e.g. in the respiratory tract, distal nephron, distal colon, sweat and salivary ducts. In these epithelia, ENaC is the rate-limiting step of sodium absorption [37, 38].

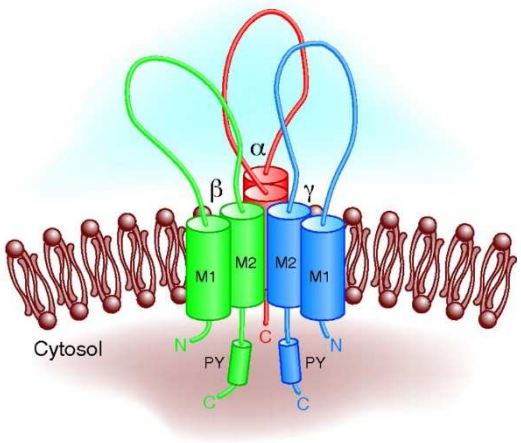


Fig. 7 Schematic representation of ENaC structure.

Interaction between CFTR and ENaC

At present, the molecular mechanisms of the regulatory relationship of the three subunits within ENaC, and between ENaC and CFTR, are incompletely understood and subject of considerable controversy.
In 1995 Stutts *et al* found that MDCK cells and 3T3 fibroblasts, when co-transfected with CFTR and αβγ-ENaC, exhibited reduced amiloride sensitive Na^+ current in a Cl^- free solution as compared to cells expressing αβγ-ENaC in the absence of CFTR [34]. Following studies observed the inhibitory effects of CFTR on ENaCs also in other cells [39, 40].
Although there is a general agreement that understanding of the CFTR–ENaC interaction can clarify the pathophysiology of CF, the exact mechanism of their relationship is not clear. A number of possible mechanisms have proposed, such as regulation by chloride

concentration [41], indirect effect through intermediary proteins [42, 43, 44] or direct protein–protein interaction between the two molecules [45].
Despite the poor knowledge about mechanism of interaction between CFTR and ENaC, it is generally accepted that the coordinated regulation and activity of CFTR and ENaC determine the composition of the airway surface liquid (ASL) in the lungs (fig. 8). CFTR-mediated Cl^- secretion enhances ASL volume while ENaC mediated Na^+ absorption reduces it. Thus, an appropriate balance between CFTR and ENaC activity is essential for maintaining an optimal ASL volume; if not, disease will occur [46, 47].
In classical CF patients with loss-of-function mutations in both copies of the CFTR gene, the ASL depletion is explained by the lack of CFTR-mediated Cl^- secretion and accelerated Na^+ absorption. Interestingly, transgenic mice that overexpress SCNN1B in the lower airways have increased airway epithelial sodium absorption and present CF-like lung disease symptoms [48, 49].

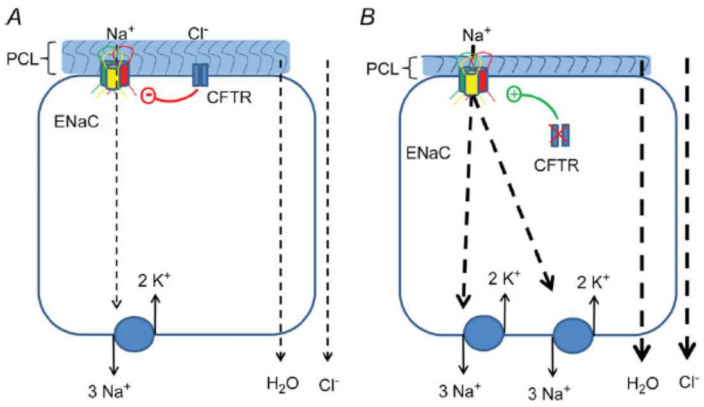


Fig. 8 Na^+ hyperabsorption leads to increased O_2 consumption and airway surface liquid (ASL) volume depletion. A and B, schematic diagrams of Na^+ absorption in normal and cystic fibrosis airway epithelia, respectively. Na^+ is transported down its electrochemical gradient through ENaC in the apical membrane and is pumped out across the basolateral membrane by the Na^+/K^+-ATPase. Because airway epithelia are highly water permeable and have NaCl-permeable paracellular pathways, Cl^- and H_2O follow Na^+ [50].

ENaC mutations and other diseases

Mutations in ENaC are also known to be involved in two different human genetic diseases. Either too low or too high amounts of ENaC will thus cause disease. Activation of ENaC by mutations in either SCNN1B or SCNN1G causes Liddle's syndrome, an inherited form of salt-sensitive arterial hypertension with enhanced renal sodium retention [51]. These dominant gain of function mutations lead to an enhanced channel activity by increasing the number of ENaC channels expressed at the cell surface and by increasing the channel's open probability [52, 53]. Loss of function mutations in SCNN1A, SCNN1B, or SCNN1G cause autosomal recessive pseudohypoaldosteronism type I, characterized by severe renal salt-wasting and arterial hypotension [54]. Moreover, in PHA-I patients, the reduced reabsorptive capacity of the lungs leads to an increased ASL volume, which often results in recurrent respiratory problems [55]. Indeed, PHA-I patients were found with chronic lung disease that resembles that of CF in the absence of common CFTR mutations [56, 57].

ENaC mutations and atypical CF

Given the observation that sodium hyperabsorption through ENaC is part of the basic CF pathology, and that mice that overexpress SCNN1B present CF-like disease, mutations in the genes encoding ENaC may potentially explain disease in patients with CF-like disease in whom a mutation cannot be identified on both CFTR genes.
Two different groups have analyzed patients with atypical CF and with one or none CFTR mutations, to find causative ENaC mutations [58, 59].
In 2005 Sheridan *et al.* have analyzed 20 non-classic CF patients without CFTR mutations, showing respiratory manifestations. They sequenced the exons and the flanking introns of the genes encoding the α, β and γ-subunits of ENaC and they identified six novel sequence changes. Five were predicted to cause amino acid changes: R181W in SCNN1A and S82C, P267L, G294S and E539K in SCNN1B, whereas the sixth changed the highly conserved penultimate A in the 3' splice site of SCNN1B intron 12 (1670–2 A to G). P267L, G294S and E539K variants were not found in ethnically matched control alleles; the three aminoacids mutated in patients were completely conserved in ENaC orthologues and paralogues in human, rabbit, mouse and rat.
RT–PCR analysis of nasal epithelial RNA revealed that the 1670–2 A to G mutation resulted in two stable SCNN1B transcripts. The first retains 83 nucleotides from the 3' end of intron 12 and is predicted to alter the amino acid sequence following codon 514 and terminate after the addition of 188 novel residues. The second transcript lacks the first 33 nucleotides of SCNN1B exon 13 leading to a deletion of 11 amino acids that precede the second transmembrane domain (fig. 9). These results confirmed that 1670-2 A to G causes aberrant splicing leading to mRNA transcripts with substantial alterations in sequence.

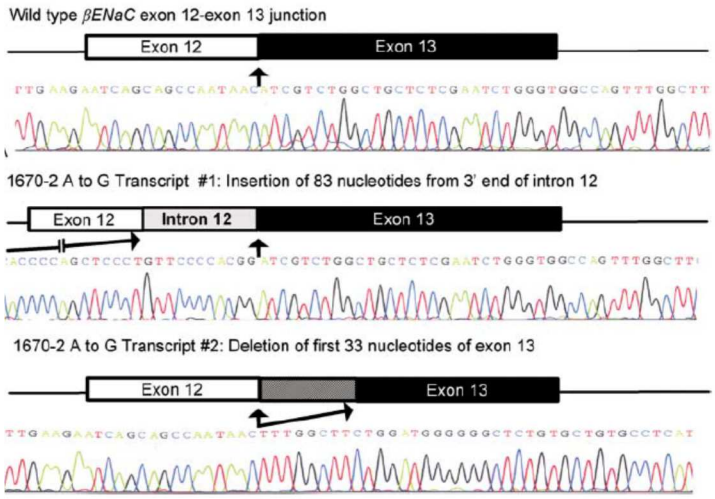


Fig. 9 RT–PCR analysis of SCNN1B 1670–2 A to G.
RT-PCR with primers placed in SCNN1B exons 12 and 13 to analyze the exon 12/exon 13 junction in transcripts was performed. RT–PCR products were sequenced. Two different mutant transcripts were present [58].

They also investigated the effect of the three missense mutations on ENaC function, by using *Xenopus laevis* oocytes; they found that oocytes injected with E539K βENaC or

P267L βENaC generated Na^+ currents significantly lower than those of wild-type. In contrast, the co-expression of G294S βENaC produced Na^+ currents that were significantly higher than those of wild-type. Thus, each of the found missense SCNN1B mutations was associated with abnormal function (fig. 10).
Together, these data indicate that the SCNN1B mutations discovered in these patients were likely to be deleterious [58].

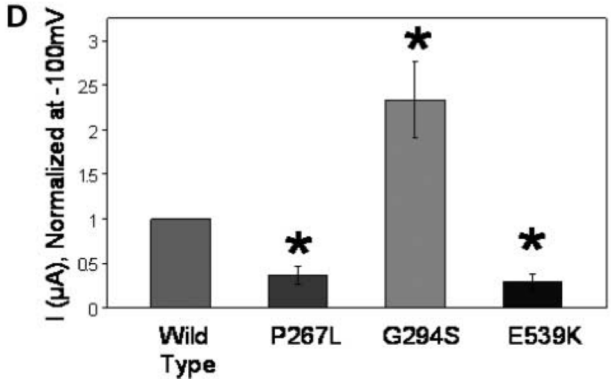


Fig. 10 Functional studies of SCNN1B mutants expressed in *Xenopus* oocytes. Histogram of normalized mean currents (+SEM) at a membrane potential of -100 mV for oocytes injected with RNA encoding wild-type α- and γ-ENaC plus wild-type βENaC, n= 40 from nine batches; βENaC P267L, n= 14 from four batches; βENaC G294S, n= 27 from six batches; or βENaC E539K, n=12 from six batches. Asterisks indicate P-values less than 0.005 when compared with wild-type [58].

In 2009 Azad *et al.* have analyzed 76 patients with CF or CF-like lung symptoms, carrying one or none CFTR mutation and/or with a positive or borderline sweat test. They sequenced the exons and the flanking introns of the genes encoding α, β and γ-subunits of ENaC. They identified a total of 30 sequence variants, some of which were previously described by Sheridan (fig. 11).
Among selected variants, only the hyperactive V114I and hypoactive F61L mutations in SCNN1A were observed once in their patients, but not in controls. They suggested that these mutations could potentially cause disease by a Mendelian mechanism. Interestingly, the hyperactive V114I-SCNN1A mutation was identified in the German patient that provided the first evidence that mutations in other genes than CFTR may cause CF-like disease [31].
The cumulative frequency of mutations that have a minor allele frequency of <2.5% in controls had a more than three-fold significantly increased cumulative frequency (30%) in the studied patient group. Among these, the most relevant functional variant W493R-SCNN1A was found at a more than two-fold significantly increased incidence in patients (8%).
From a genetic point of view, all these observations suggested an involvement of ENaC in disease in some of these patients [59].

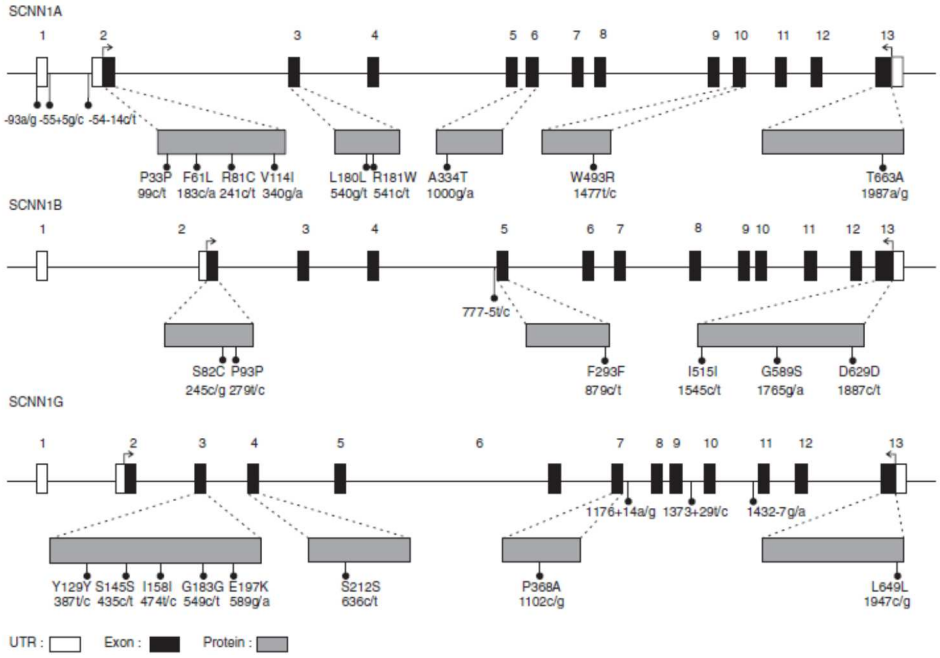


Fig. 11 Schematic diagram of the different SCNN1A, SCNN1B, and SCNN1G mutations found in the study of Azad *et al.* The coding exons and their flanking 30 nucleotides in the introns were analyzed. Translation initiation and termination are shown by arrows. Open boxes refer to the untranslated region (UTR), black boxes refer to coding exon, and gray boxes refer to protein [59].

3.2. MATERIALS AND METHODS

3.2.1. Design of ENac typing test prototype

A prototype for identification of ENaC variants, probable involved in pathogenesis of atypical cystic fibrosis (GENEQUALITY ENaC-TYPE), was planned. After an *in silico* study for primers and probes selection, the prototype was manufactured.
Our test allows detection of 9 ENaC variants through RLB technology. These mutations have been previously described, first by Sheridan *et al.* [58] and then by Azad *et al.* [59]. They demonstrated a correlation between these functional variants and atypical respiratory phenotypes. Mutations frequencies were significantly higher in patients rather than in controls.
Variants, their genomic position, their effect on protein and frequency in the general population are summarized in tab. 5.

3.2.2. PCR design

Design of primers

Design and analysis of primers can be divided into three phases:

1. Selection of the target region.
2. Design of primer pairs.
3. Analysis of primers.

The genomic target sequence was identified using GenBank and dbSNP. These are open access databases of the NCBI (National Center for Biotechnology Information) containing a collection of publicly available nucleic acid and protein sequences and their genetic variations within and across different species.
The UCSC Genome Browser is a web-based graphical viewer that allows locating DNA sequences in the whole genome. It provides a fast display of any requested portion of the genome at any scale, together with annotation tracks (known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologies, and more) to recognize other possibly relevant features present in such a region.
NCBI genomic sequences used are summarized in tab. 3.

Primers have been designed by using 2 bioinformatics tools: Primer Express® Software v3.0.1 (Life Technologies) and DNA Mate (http://melolab.org/dnaMATE/tm-pred.html). Primer Express® is a tool licensed from Life Technologies, that allows to design primers and probes using TaqMan® and SYBR® Green I dye chemistries for gene quantitation and allelic discrimination (SNP) real-time PCR applications. It is possible to project primers starting form an input sequence given by the operator or to draw manually oligonucleotides. Finally, for each primer, it gives information about temperature of melting, content of GC, amplicon length, probability of secondary structures formation.

DNA Mate is a program that calculates the annealing or melting temperature of any given short DNA sequence (in the range of 16-30 nt) using five different approximations. A merged or consensus temperature among all calculations is also given. In addition to this, the server will inform to the user about the expected variation of the melting temperature estimation, which depends on the specific oligonucleotide sequence.

All the selected primers satisfy some common features:

- primer length: 18-25bp
- amplicon length: 100-300bp
- no dinucleotide repeats or long stretches of guanosine (G)
- less than 3bp difference in length between primer pairs
- melting temperatures (Tm) between 60°C and 65°C
- GC content between 40% and 60%
- balanced distribution of GC-rich and AT-rich domains
- no intra-primer homology (more than 3 bases that complement within the primer) or inter-primer homology (forward and reverse primers having complementary sequences) to avoid self-dimers or primer-dimer formation.

After designing primers, a BLAST search was performed to determine their specificity (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The goal is to avoid no specific alignment with other regions different from the target sequence, not only in the human genome, but also for other microorganisms.

Sequences of primers are not reported, in protection of trade secret.

SNP	GENE	ALTERNATE NAMES	REFERENCE SEQUENCE *(RefSeq-GeneBank ID)*	GENOMIC POSITION*	GENIC FUNCTION	FREQUENCY **	AMPLIFIED PRODUCT
rs61758859	SCNN1A	c. C183A p. F61L	NT_009759.15 NM_001038.4	chr12: 6,483,717-6,483,817 band:12p13.31 exon 2	Loss of function mutation	N.A.	290 bp
rs61759861	SCNN1A	c. C340T p. V114I	NT_009759.15 NM_001038.4	chr12: 6,483,560-6,483,660 band: 12p13.31 exon 2	Gain of function mutation	MAF < 0.01 (T)	290 bp
rs61759925	SCNN1A	c.G541A p.R181W	NT_009759.15 NM_001038.4	chr12: 6,472,702-6,472,802 band: 12p13.31 exon 3	Gain of function polymorphism	MAF=0.02 (A)	70 bp
rs5742912	SCNN1A	c.A1477G p.W493R	NT_009759.15 NM_001038.4	chr12: 6,458,300-6,458,400 band: 12p13.31 exon 10	Gain of function polymorphism	MAF=0.02 (G)	150 bp
rs35731153	SCNN1B	c.C245G p.S82C	NT_010393.15 NM_000336.2	chr16: 23,360,115-23,360,215 band: 16p12.2 exon 2	Loss of function mutation	MAF < 0.01 (G)	200 bp
rs137852709	SCNN1B	c.C799T p.P267L	NT_010393.15 NM_000336.2	chr16: 23,379,150-23,379,250 band: 16p12.2 exon 5	Loss of function mutation	N.A.	300 bp
rs72654338	SCNN1B	c.G880A p.G294S	NT_010393.15 NM_000336.2	chr16: 23,379,230-23,379,330 band: 16p12.2 exon 5	Gain of function mutation	MAF < 0.01 (A)	300 bp
No registered	SCNN1B	p.1670-2A>G	NT_010393.15 NM_000336.2	chr16: 23,391,764-23,391,739 band: 16p12.2 Intron 12	Loss of function mutation	N.A.	260 bp
rs137852710	SCNN1B	c.G1615A p.E539K	NT_010393.15 NM_000336.2	chr16: 23,391,764-23,391,864 band: 16p12.2 exon 12	Loss of function mutation	N.A.	260 bp

Tab. 3. Description of ENaC variants detected with our test. * chromosomic localization of variants, as reported in UCSC (Human Genome Browser). ** frequency of each variant, reported in the 1000 Genomes database, for European population

Design of PCR reaction

The polymerase chain reaction (PCR) is a technology that allow the specific and exponential amplification of a specific sequence of DNA.
The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA.
Specificity is given by primers, which are complementary to the target region; they also be necessary for the initiation of DNA synthesis by DNA polymerase. This enzyme enzymatically assembles a new DNA strand from the nucleotides, by using single-stranded DNA as a template.
The reaction of PCR is composed by three phases: denaturation, annealing and extension, which are repeated for a number of times variable between 25 and 50.
For our amplification protocol, a particular Hot Start Taq Polymerase has been used; it is a recombinant enzyme, given in inactive state through link with antibodies, and requires a thermal activation. This enzyme avoid no specific amplifications during phases of reagents preparation, which could be performed at room temperature.
A specific reaction buffer, provided with Taq polymerase and containing the denaturing agent DMSO, has been used in order to raise efficiency of amplification for some sequences with a high percentage of guanine and cytosine.
The amplification protocol has been designed as a multiplex PCR; all the investigated ENaC variants are amplified in a unique reaction.
Commercial names of reaction Mix and Taq polymerase, qualitative and quantitative composition of oligomix are omitted, in protection of trade secret.
In the following table the thermal profiling of PCR is reported.

Hold	Temperature	Time	Number of
Hot start	95°C	5 minutes	1
Denaturation	95°C	30 seconds	40
Annealing	58°C	90 seconds	
Extension	72°C	30 seconds	
Final Extension	68°C	10 minutes	1
Storage	10°C	∞	1

Tab. 4. Thermal cycling used for the multiplex amplification of ENaC variants.

3.2.3. RLB design

Design of probes

For each investigated ENaC variant two probes have been designed: one specific for the wild-type sequence and one for the allelic variant. In total 18 probes have been designed. Genotyping probes have been elaborated on the same genes RefSeq, used for the primers.
3 bioinformatics tools have been used:

1. *DNA Mate* (http://melolab.org/dnaMATE/tm-pred.html), for the evaluation of melting temperature.
2. *DNA Folding Form* application (http://mfold.rna.albany.edu/?q=mfold), to analyze the amplicon folding and to evaluate the presence of possible secondary structures that can prevent annealing with probe.
3. *Two state melting* application (http://mfold.rna.albany.edu/?q=DINAMelt), to evaluate temperature of hybridization between probe and target sequence.

All the selected probes satisfy some common features:

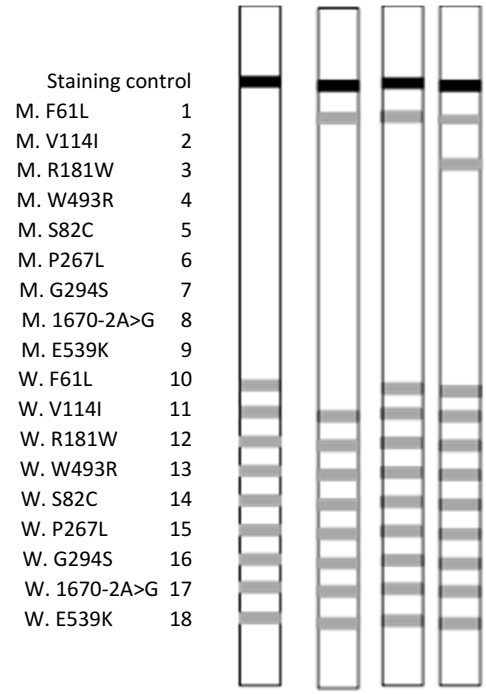
- central position, in the sequence, of mismatch that differentiate wild-type from variant;
- average length of 20 base pairs;
- designed in regions not involved in strong secondary structures;
- 53°C as temperature of hybridization with specific sequences;
- ΔT of almost 4°C between temperatures of hybridization of the probes that identify the two allelic variants;
- average temperature of melting between 65° and 70°C.

Sequences of probes are not reported, in protection of trade secret.

Strip manufacturing

The support of the strip is a nitrocellulose baked membrane. The probes are spotted on nitrocellulose through an automatic work-station and then are immobilized on support by exposure to UV light, which induces the covalent bounds between oligos and membrane.
Then, spotted membrane is incubated with a blocking solution containing BSA, for 30 minutes in moderate shaking. This step prevents non-specific background binding of the DNA to membrane.
Finally, membrane is cutted into single strips, through an automatic instrument.
Each strip contains 18 genotyping probes: in the upper part of the strip there are probes relatives to variant alleles, while in the bottom part there are probes for wild-type alleles. Moreover, there is a staining control that guarantee the correct execution of post-hybridization steps and a reference line for the interpretation of resulting pattern (tab. 5 – fig. 12).

Position	Description
	Staining control
1	M. F61L
2	M. V144I
3	M. R181W
4	M. W493R
5	M. S82C
6	M. P267L
7	M. G294S
8	M. 1670-2A>G
9	M. E539K
10	W. F61L
11	W. V144I
12	W. R181W
13	W. W493R
14	W. S82C
15	W. P267L
16	W. G294 S
17	W. 1670-2A>G
18	W. E539K



Tab. 5. Position of probes on strip. Probes M. type (pos. 1-9) identify variant alleles. Probes W. type (pos. 10-18) identify wild-type alleles.

Fig. 12 Example of probes pattern. Sample 1: individual omozygous wild-type for all analyzed variants. Sample 2: individual omozygous for F61L mutation and wild-type for all the other variants. Sample 3: individual heterozygous for F61L mutation and omozygous wild-type for all the other variants. Sample 4: individual heterozygous for F61L mutation and R181W polymorphism and omozygous wild-type for all the other variants.

Probes concentration, buffer compositions, commercial name of nitrocellulose membrane and other details are omitted, for respecting intellectual proprietary and trade secret.

Design of RLB protocol

A RLB assay is influenced by several interconnected parameters.

The stringency of the assay could be achieved by acting on two fundamental factors: temperature and ionic strenght of solutions.

Temperature is a key on which it is possible to operate in order to achieve more stringency; a higher temperature makes less stable nonspecific duplex DNA-oligo and limits their formation. Temperature of the assay depends essentially by the optimal temperature of hybridization between DNA and probe. This is given by the structure of

the probes: it is influenced by content of GC, length of oligos, number, position and types of mismatch between DNA target and oligos.
The other factor that influences stringency of the assay is the ionic strength of the hybridization and washing solutions. A minor ionic strength results in a higher stringency.
Finally, it is possible to act on incubation time of the several washing steps.
Table 6 summarizes the RLB protocol of our test.

Step	Reagents	Incubation	Time of incubation
1) Denaturation of amplicon	Amplicon + Denaturation Solution, containing Sodium hydroxide	Room temperature	5’
2) Hybridization	Hybridization solution, containing Sodium citrate, Sodium chloride, Sodium dodecyl sulfate and preservatives	52°C Moderate shaking	30’
3) Stringent wash	Stringent wash solution, containing Sodium citrate, Sodium chloride, Sodium dodecyl sulfate, and preservatives	52°C Moderate shaking	10’
5) Incubation with Streptavidine AP-Conjugated	Streptavidine AP-Conjugated diluted in stringent wash solution	52°C Moderate shaking	30’
6) Wash	Rinse solution, containing magnesium chloride, sodium chloride, Trizma and detergents.	Room temperature Moderate shaking	2’
7) Wash	Rinse solution, containing magnesium chloride, sodium chloride, Trizma and detergents.	Room temperature Moderate shaking	2’
8) Colorimetric reaction	NBT/BCIP solution, containing NBT, BCIP, Tris buffer and $MgSO_4$	Room temperature Moderate shaking AT DARK LIGHT	10’
9) Blocking of reaction	Blocking solution, containing Citric acid	Room temperature Moderate shaking	2’
10) Final wash	Distilled water	Room temperature Moderate shaking	2’

Tab. 6. Scheme of visualization on strip protocol.

3.2.4. DNA sequencing

Samples resulting carrier of ENaC variants by RLB assay were sequenced to confirm the outcome.
Amplicons have been purified by using ExoSap-IT® (Affymetrix), according to the manufacturer’s protocol. Reaction of sequencing was performed by BMR Genomics, with the BigDyes terminator cycle sequencing kit v3.1 (Applied Biosystems). The sequencing reactions were run on ABI 3730XL (Applied Biosystems) and ABI 3100 (Applied Biosystems) and were analyzed with FinchTV software v1.4 (Geospiza).

3.2.5. Statistical analysis

Differences in allelic frequencies between patients and controls were tested for each SNP with the use of a Fisher exact probability, because more than 25% of the cells had a count of <5. Genotypic odds ratios (ORs) and 95% confidence intervals (CIs) were estimated.

3.2.6. Samples

Clinical samples used in this study have been provided by different Italian institutes and analysis laboratories. Samples have been selected and afterwards they were made anonymous, randomized, classified and stored with a biobanking system.
Genomic DNA was extracted from peripheral whole blood with the systems routinely used by the providing laboratory (tab. 7).

Number	Laboratory of origin	Matrix	Extraction system	Geographical informations
226	Laboratory of Molecular Pathology, Department of Pathology and Diagnostics, University Hospital of Verona	Peripheral blood	Qiamp DNA Blood Mini Kit (QIAGEN)	Northern Italy
7	Laboratory of Molecular Pathology, Department of Pathology and Diagnostics, University Hospital of Verona	Peripheral blood	Qiamp DNA Blood Mini Kit (QIAGEN)	Central Europe
100	Laboratory of Molecular Pathology, Department of Pathology and Diagnostics, University Hospital of Verona	Peripheral blood	BioRobot EZ1 DSP system (QIAGEN)	Northern Italy
69	Analysis Laboratory, Hospital of Bentivoglio, USL Bologna Nord	Peripheral blood	MagNA Pure Compact System (ROCHE)	Northern Italy

Tab. 7. Providing laboratory, matrix, extraction system and geographical information of the clinical samples.

In the following table, features of subjects, as phenotype and number of CFTR mutations, are summarized.

	Phenotype	CFTR mutation number			Total of subjects
		2	1	0*	
Number of subjects	Respiratory manifestations	58	9	27	94
	Chronic pancreatitis	13	28	29	70
	Recurrent pancreatitis	12	25	32	69
	Healthy	-	-	169	169

Table 8. Number of subjects analyzed, divided on the basis of phenotype and number of CFTR mutations.*defined by second level analysis

3.3. RESULTS AND DISCUSSION

3.3.1. Optimization of ENaC typing test prototype

In this Ph. D. project, we have developed a prototype for identification of ENaC variants, named GENEQUALITY ENaC-TYPE.
Preliminary experimental tests have been made to set conditions of PCR and visualization on strip. We tested samples of known mutational status with different master mixes, different concentrations of primers and probes. The visualization of the PCR products by RLB was optimized testing different combinations of temperatures, washing solutions, and duration of incubation. Step by step, we compared the different conditions considering both intensity of signal on strip and absence of aspecific bands, to select a compromise between these features.

3.3.2. Validation of ENaC typing test prototype

After achieving the prototype standardization of our test, a phase of the project dedicated to the evaluation of the diagnostic performance was started.
The performances of this assay have been evaluated on a total of 233 samples from affected patients and 169 from healthy individuals, as control group.
The clinical criteria for patient recruitment were the presence of CF or CF-like lung or pancreatic symptoms, as well as of a positive or borderline sweat test and/or the presence of one CF-causing CFTR mutation. Only patients truly carrying one or no CF-causing CFTR mutations were included. The absence of two CFTR mutations was confirmed by second level molecular analysis, such as sequencing of the complete CFTR coding region and its exon/intron junctions, or mutation scanning of the complete CFTR coding region and its exon/intron junctions using DGGE or dHPLC. The sensitivity of these scanning assays may not reach 100%, so a CFTR mutation might have escaped detection.
Patients have been divided into 2 groups on the bases of their phenotype: 94 individuals with respiratory symptoms and 139 patients with pancreatic manifestations.

Healthy control group

The prevalence of ENaC variants was first determined in the group of 169 healthy individuals.
In this group, 17 individuals carried an ENaC variant, mainly the polymorphism p.W493R-SCNN1A. A smaller part showed the mutation p.S82C-SCNN1B or the polymorphism p.R181W-SCNN1A.

In the following table there is summarized the identified ENaC variants for the 17 individuals.

N. of individuals	SCNN1A VARIANT	SCNN1B VARIANT
10	W493R	-
3	R181W	-
4	-	S82C

Tab. 9. Distribution of ENaC variants in the group of healthy controls.

Patients with respiratory manifestations

Patients with CF-like lungs symptoms were divided into 3 groups based on the number of CFTR mutations:

A. 9 individuals with atypical CF, heterozygous for CFTR mutation
B. 27 patients with atypical CF, without any CFTR mutation
C. 58 patients with classic CF, homozygous or compound heterozygous for CFTR mutations

In the group A, 7 ENaC variants have been identified: two polymorphisms p.R181W-SCNN1A and p.W493R-SCNN1A and five mutations p.S82C-SCNN1B, p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B and p.1670-2A>G-SCNN1B. Three patients carried a single ENaC variant, while five were compound heterozygous for ENaC mutations. One individual did not carry any ENaC variant. Patients n. 5 and n. 6 belonged from the same family; the same was for individuals n. 7 and n. 8.
Table 10 summarizes the identified ENaC variants and CFTR mutation for each patient.

PATIENT AH*	SCNN1A VARIANT	SCNN1B VARIANT	CFTR MUTATION
1	R181W	-	G542X
2	W493R	-	ΔF508
3	-	S82C	3849+10 kbC>T
4	W493R	S82C	R347P
5	-	P267L 1670-2A>G	ΔF508
6	-	P267L 1670-2A>G	ΔF508
7	-	G294S E539K	ΔF508
8	-	G294S E539K	ΔF508
9	-	-	R1162X

Tab. 10. Atypical heterozygous patients, with relative ENaC variants and CFTR mutation (group A). *AH= Atypical Heterozygous

In the group B, only three patients carried ENaC variants: one showed the polymorphism p.R181W-SCNN1A, and two the p.W493R-SCNN1A polymorphism, as summarized in table 11.

PATIENT AN*	SCNN1A VARIANT	SCNN1B VARIANT
1	R181W	-
2	W493R	-
3	W493R	-

Tab. 11. Atypical patients without CFTR mutations and their ENaC variants (group B). *AN= Atypical No CFTR mutation

In the group C, only three individuals carried ENaC variants. Two of them showed the p.W493R-SCNN1A polymorphism, while the other carried the p.S82C-SCNN1B mutation, as summarized in table 12.

PATIENT CCF*	SCNN1A VARIANT	SCNN1B VARIANT	CFTR MUTATIONS
1	-	S82C	ΔF508/3849+10 kbC>T
2	W493R	-	ΔF508/ΔF508
3	W493R	-	ΔI507/ΔI507

Tab. 12. Classic CF patients with relative ENaC variants and CFTR mutations (group C). *CCF= Classic CF

The prevalence and distribution of each ENaC variant in patients and control groups have been analyzed and relative allelic frequencies have been calculated (tab. 13).

SNP REFERENCE	VARIANTS	OBSERVED ALLELIC FREQUENCY (mutated allele)			EXPECTED ALLELIC FREQUENCY*
		ATYPICAL CF PATIENTS (n=36)	CLASSIC CF PATIENTS (n=58)	TRUE CONTROLS (n=169)	
rs61758859 C>A	F61L	-	-	-	no data
rs61759861 C>T	V114I	-	-	-	0
rs61759925 G>A	R181W	0.03	-	0.009	0.02
rs5742912 A>G	W493R	0.05	0.02	0.03	0.02
rs35731153 C>G	S82C	0.03	0.009	0.01	<0.01
rs137852709 C>T	P267L	0.03	-	-	no data
rs72654338 G>A	G294S	0.03	-	-	<0.01
No registered A>G	1670-2A>G	0.03	-	-	no
rs137852710 G>A	E539K	0.03	-	-	no data

Tab. 13. Summary of the detected ENaC variants and their allelic frequencies. *Frequencies reported in the *1000 Genomes* Database, for European population.

The allelic frequencies have been used to perform a statistical analysis, in order to evaluate possible significative differences between atypical affected and healthy controls.

On the total of seven identified variants in this group of patients, only four had a frequency significantly higher in affected than in controls. These are the loss of function mutations p.E539K-SCNN1B, p.1670-2A>G-SCNN1B and p.P267L-SCNN1B and the gain of function mutation p.G294S-SCNN1B. They were found in two atypical patients (5.5%), at a more than five-fold increased incidence compared to controls (0%) (p-value=0.03), with an OR of 5.97 (95% CI, 4.39-8.11).

The four significative mutations are located on SCNN1B gene. These results are partial in agreement with the findings that transgenic mice, overexpressing SCNN1B in the lower airways, have increased airway epithelial sodium absorption and present CF-like lung disease symptoms. It is possible that the amount of different ENaC subunits that is present in the cell is likely not equal. Indeed, the amount of a given subunit depends on many factors at the transcriptional and translational level. One subunit will thus be the limiting factor for the generation of functional ENaC channels. SCNN1B might be the limiting factor in the formation of functional ENaC channels, such that overexpression results in more ENaC activity and disease in transgenic mice [48].

The polymorphism p.R181W-SCNN1A and the loss of function mutation p.S82C-SCNN1B were found in two affected individuals (5.5%), a three-fold increased incidence respect to control group of healthy individuals (1.7%), but the statistical analysis revealed that this difference was not significant (p-value=0.21).

The remaining polymorphism p.W493R-SCNN1A was identified in four patients (11.1%) and in ten healthy individuals (5.6%); also in this case the difference was not significantly different between the two groups (p-value=0.27).

The only difference between frequencies observed in our group of control and those reported in database was identified for p.R181W-SCNN1B polymorphism. This finding could be related to different geographical belonging of individuals. Our control group was entirely composed by Italian individuals, while frequencies reported by *1000 Genomes* Database were calculated on individuals from different European countries. Despite this difference, p.R181W-SCNN1B frequency in our group of atypicals was similar to value reported by database, although patients carrying the polymorphism were Italian.

The four significative variants, identified in this study, were found in patients heterozygous for CFTR. They carried ΔF508 mutation, the most common causative variant of CF. When a disease-causing mutation is found on both CFTR genes, CFTR network is almost completely nonfunctional. However, in patients in whom a mutation is only found on one CFTR gene, about 50% of the network is still functional. This part of CFTR network may be rendered nonfunctional by a mutation in at least one other member of the CFTR network, such as ENaC. In these cases, ΔF508 account for damage of CFTR network and, probably, p.E539K-SCNN1B, p.1670-2A>G-SCNN1B, p.P267L-

SCNN1B, p.G294S-SCNN1B affect functionality of remaining 50%. Thus, it is possible that ENaC mutation, in combination with a loss-of function mutation in one copy of the CFTR gene, may cause or predispose to CF or CF-like disease.
However it is very difficult to develop an explanation of the effect of the different mutations, especially in the light that either too low or too high amounts of ENaC can already result in two different diseases, which will be even further complicated if the CFTR protein is implicated.
Other ENaC variants were also detected in patients with explained CF and in healthy individuals, but their effect was null. In CF affects, disease is essentially given by the two CFTR disease-causing mutations, as expected. Therefore, it is possible to speculate that no disease will occur when these mutations are found alone in a heterozygous state.

Patients with pancreatic manifestations

Patients with CF-like pancreatic symptoms (chronic or recurrent) were divided in 3 groups based on the number of CFTR mutations:

A. 53 individuals with atypical CF, heterozygous for CFTR mutation
B. 61 patients with atypical CF, without any CFTR mutation
C. 25 patients with classic CF, homozygous or compound heterozygous for CFTR mutations

In the group A, we have identified four individuals carrying ENaC variants. In particular, two of them showed chronic pancreatitis, while the other two had a clinical picture of recurrent pancreatitis. They carried the polymorphisms p.R181W-SCNN1A and p.W493R-SCNN1A and the mutation p.S82C-SCNN1B. Patients were heterozygous for CF causing mutations, as summarized in the table 14.

PATIENT AH*	PHENOTYPE	SCNN1A VARIANT	SCNN1B VARIANT	CFTR MUTATION
1	Recurrent pancreatitis	R181W	-	W1282X
2	Recurrent pancreatitis	W493R	-	P5L
3	Chronic pancreatitis	R181W	-	ΔF508
4	Chronic pancreatitis	-	S82C	ΔF508

Tab. 14. Atypical heterozygous patients with pancreatic symptoms, relative ENaC variants and CFTR mutations (group A). *AH= Atypical Heterozygous

In the group B, only five patients carried ENaC variants: four of them showed the p.W493R-SCNN1A polymorphism, while one the p.R181W-SCNN1A.

Table 15 summarizes the identified ENaC variants for the five patients.

PATIENT AN*	PHENOTYPE	SCNN1A VARIANT	SCNN1B VARIANT
1	Recurrent pancreatitis	R181W	-
2	Recurrent pancreatitis	W493R	-
3	Chronic pancreatitis	W493R	-
4	Chronic pancreatitis	W493R	-
5	Chronic pancreatitis	W493R	-

Tab. 15. Atypical patients with pancreatic manifestations and no CFTR mutations and their ENaC variants (group B). *AN= Atypical No CFTR mutation

The prevalence of the ENaC variants was also determined in the group C (tab. 16).
In this group, only two individuals carried ENaC variants: one was a compound heterozygous for p.R181W-SCNN1A and p.S82C-SCNN1B, while the other carried the p.S82C-SCNN1B mutation. Interestingly, these two patients exhibited chronic pancreatitis, while individuals with recurrent pancreatitis did not carry any ENaC variants.

PATIENT CCF*	PHENOTYPE	SCNN1A VARIANT	SCNN1B VARIANT	CFTR MUTATIONS
1	Chronic pancreatitis	R181W	S82C	ΔF508/(TG)12T5
2	Chronic pancreatitis	-	S82C	ΔF508/D1152H

Tab. 16. Classic CF patients with relative ENaC variants and CFTR mutations (group C). *CCF= Classic CF patients

In addition, the prevalence and distribution of each ENaC variant, in the group of patients with pancreatic manifestations, have been analyzed and relative allelic frequencies have been calculated (tab. 17).

SNP REFERENCE	VARIANTS	OBSERVED ALLELIC FREQUENCY (mutated allele)			EXPECTED ALLELIC FREQUENCY*
		ATYPICAL CF PATIENTS (n=114)	CLASSIC CF PATIENTS (n=25)	TRUE CONTROLS (n=169)	
rs61758859 C>A	F61L	-	-	-	no data
rs61759861 C>T	V114I	-	-	-	0
rs61759925 G>A	R181W	0.01	0.01	0.009	0.02
rs5742912 A>G	W493R	0.02	-	0.03	0.02
rs35731153 C>G	S82C	0.004	0.009	0.01	<0.01
rs137852709 C>T	P267L	-	-	-	no data
rs72654338 G>A	G294S	-	-	-	<0.01
No registered A>G	1670-2A>G	-	-	-	no registered SNP
rs137852710 G>A	E539K	-	-	-	no data

Tab. 17. Summary of the detected mutations and their allelic frequencies. *Frequencies reported in the *1000 Genomes Database*, for European population.

Statistical analysis, performed on allelic frequencies calculated above, has underlined the absence of any significative differences between atypical affected and healthy controls.
The two polymorphisms p.R181W-SCNN1A, p.W493R-SCNN1A and loss of function mutation p.S82C-SCNN1B were identified both in the group of patients and in the controls, without any statistical difference between them.
The polymorphism p.R181W-SCNN1A was found in three affected individuals (2.6%) and in three healthy individuals (1.7%); the difference was not clearly statistical significant (p-value=0.39). The second polymorphism p.W493R-SCNN1A was identified in five atypicals (4.4%) and in ten controls (5.6%) (p-value=0.78); finally p.S82C-SCNN1B mutation was found in only one affected (0.87%) and in the control group with the value of 1.7% (p-value=0.65).
Interestingly, ENaC variants that appeared significative in the group of patients with respiratory symptoms (e.g. p.P267L-SCNN1B, p.G294S-SCNN1B, p.E539K-SCNN1B and p.1670-2A>G-SCNN1B), have not been identified in individuals with pancreatic manifestations.
Frequency of identified ENaC variants were not significantly different between patients with recurrent pancreatitis and individuals with chronic pancreatitis. Among the patients with explained CF, only individuals with chronic pancreatitis carried ENaC variants. This finding has no statistical significance, because ENaC variants were detected in affected with recurrent pancreatitis, carrying one or none CFTR mutations. Therefore, the observed situation could be due to the low number of individuals in the group.

All the variants identified by RLB assay were confirmed through sequencing (data not shown).

4. PART II: HCV GENOTYPING

4.1. INTRODUCTION

Hepatitis C virus (HCV) has been cloned in the 1989 [60, 61], and it is the principal cause of the so-called post-transfusion non-A, non-B hepatitis.
HCV is a single-strand RNA virus, member of Flaviviridae family, Hepacivirus genus, although it is different in many details of its genome organization from the original members of the family. HCV is additionally distinct and somewhat unusual for an RNA virus in being able to establish persistent infections in the majority of exposed individuals. This phenomenon has attracted the greatest interest in HCV research, not least because long-term, chronic infections underlie its disease manifestations and effective therapy must break this ongoing cycle of replication in the liver [62].

4.1.1. Genetic variability of HCV

Genetic variability of HCV exists at several different levels. Most obvious is the substantial genetic divergence of the main genotypes of HCV, which frequently show specific geographical ranges in the human population and associations with particular risk groups for infection. Below this, variability is observed between individual variants (or strains). It reflects processes of neutral sequence drift over time after the introduction of HCV into new risk groups in the 20th century. Some of the sequence divergence may represent phenotypically selected changes, associated with adaptation for replication in individuals with different immune responses. Finally, HCV diversifies within an infected individual overtime, forming what has been described as a *quasispecies*. This pre-existing genetic variability, combined with an extremely large replicating population size of HCV in a chronically infected individual, provides a large pool of genetic variants that can adapt to new selection pressures [62].

4.1.2. Genotypes

Comparison of nucleotide sequences of variants, recovered from infected individuals in different risk groups for infection and from different geographical regions, has revealed the existence of at least seven major genetic groups. On average over the complete genome, these differ in 30–35% of nucleotide sites, with more variability concentrated in regions such as the E1 and E2 glycoproteins. Whereas sequences of the core gene and some of the non-structural protein genes, such as NS3, are more conserved. The lowest sequence variability between genotypes is found in the 5' UTR, where specific sequences and RNA secondary structures are required for replication and translation functions.
Within each genotype, HCV is further classified into 67 confirmed and 20 provisional subtypes (fig. 13) that typically differ from each other by 20–25% in nucleotide sequences [63].

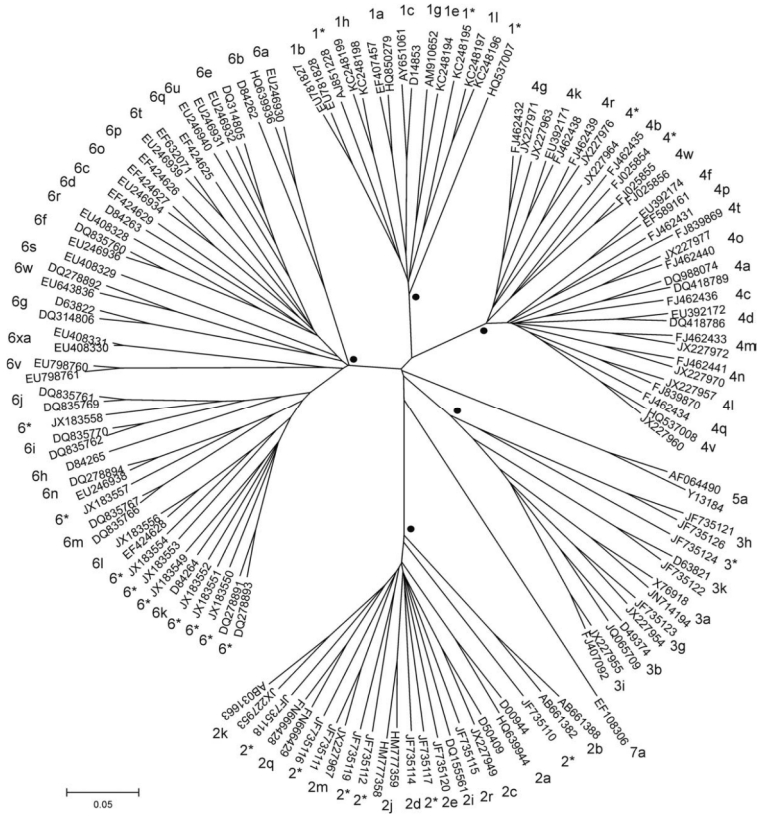


Fig. 12 Phylogenetic tree of 129 representative complete coding region sequences. Up to two representatives of each confirmed genotype/subtype were aligned and a neighbor joining tree constructed using maximum composite likelihood nucleotide distances between coding regions using MEGA5.Sequences were chosen to illustrate the maximum diversity within a subtype. Tips are labeled by accession number and subtype (*unassigned subtype). For genotypes 1, 2, 3, 4, and 6, the lowest common branch shared by all subtypes and supported by 100% of bootstrap replicates (n= 1,000) is indicated by● [63].

Distribution of HCV genotypes

The contemporary global geographic distribution of HCV genotypes is complex. It has already been established that a few subtypes—specifically 1a, 1b, 2a, and 3a—are widely distributed across the globe and account for a large proportion of HCV infections in high income countries. These so-called “epidemic subtypes” are thought to have spread rapidly in the decades prior to the discovery of HCV by way of infected blood, blood products, injecting drug use, and other routes [64, 65, 66].

The model suggested by these genotype distributions is that HCV has been endemic in sub-Saharan Africa and South-East Asia for a considerable time, and that the occurrence of infection in Western and other non-tropical countries represents a relatively recent emergence of infection in new risk groups [67, 68]. In the 20th century, parenteral exposure to blood-borne viruses became frequent through the widespread adoption of blood transfusion since the 1940s, the medical use of often unsterilized needles for injections and vaccinations (a practice that continues in many developing countries) and, most specifically, to industrialized countries, injecting drug use and the sharing of

injection equipment. These new routes for transmission plausibly account for the epidemiological and genetic evidence for recent epidemic spread of HCV over the past 50 years in Europe, Egypt and elsewhere [68, 69, 70].

Many other HCV subtypes are considered “endemic” strains; these are comparatively rare and have circulated for long periods in more restricted regions. Endemic strains from genotypes 1 and 2 are primarily in West Africa, 3 in south Asia, 4 in Central Africa and the Middle East, 5 in Southern Africa, and 6 in South East Asia (fig.14) [64, 67, 69].

To date, only one genotype 7 infection has been reported; it was isolated in Canada from a Central African immigrant [71].

The global distribution of HCV genetic variation has likely been influenced by historical and contemporary trends in human migration. For example, strains from West Africa appear to have been transferred to the America by way of the trans-Atlantic slave trade [72].

Globally, genotype 1 is estimated to account for more HCV cases than any other genotype at 83.4 million (46.2%), with over one-third of genotype 1 cases located in East Asia.

HCV genotype 3 is the next most common and is estimated to account for 54.3 million (30.1%) cases globally, approximately three-quarters of which occur in south Asia; the remaining are located in parts of Scandinavia.

Genotypes 2, 4, and 6 are responsible for the majority of the remaining cases of HCV worldwide, with an estimated 16.5 million (9.1%), 15.0 million (8.3%), and 9.8 million (5.4%) cases, respectively. East Asia accounts for the greatest numbers of genotype 2 and genotype 6 HCV cases, while North Africa and the Middle East have the largest number of genotype 4 cases.

Genotype 5 is responsible for the fewest HCV cases globally (1.4 million, <1% of all HCV cases), the great majority of which occur in Southern and Eastern sub-Saharan Africa.

Genotype 6 is present at the highest frequencies in East and Southeast Asia, but is the dominant genotype in only one country, Laos; it is also prevalent in neighboring Vietnam [73].

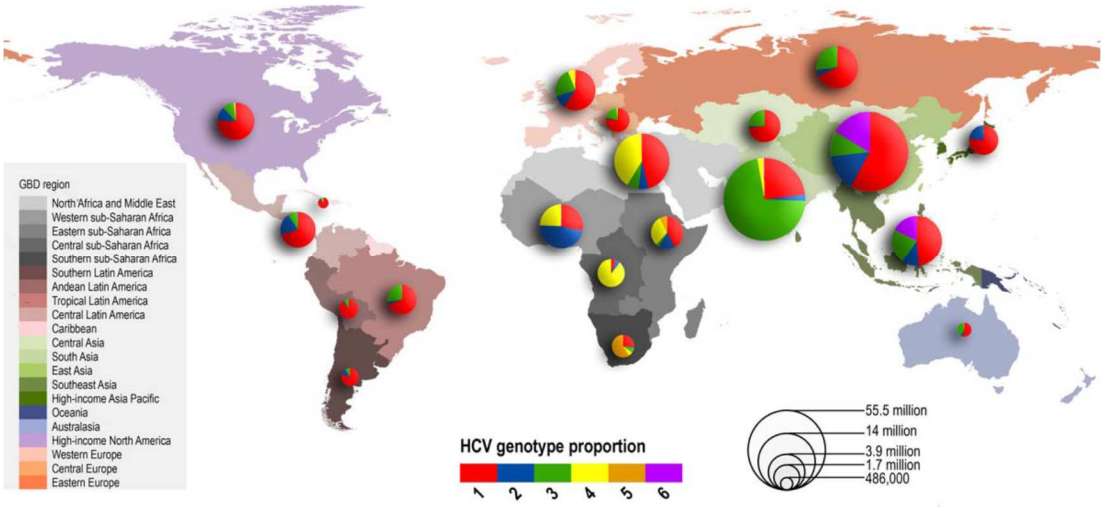


Fig. 14 Relative prevalence of each HCV genotype. Size of pie charts is proportional to the number of seroprevalent cases [74].

4.1.3. Viral structure

HCV is an enveloped virus with a positive single-stranded RNA genome, of approximately 9400 nucleotides in length. The genome contains a single ORF (Open Reading Frame), which is translated into polyprotein of 3000 aminoacids. This polyprotein is subsequently processed by viral and host proteases into 10 different proteins, three structural and seven nonstructural.

The three structural proteins, which constitute the viral particle, include the core protein and the envelope glycoproteins E1 and E2. Two regions in E2, known as hypervariable regions 1 and 2, are reported to have extreme sequence variability. The seven nonstructural components include the p7 polypeptide, the NS2-3 protease, the NS3 serine protease and RNA helicase, the NS4A polypeptide, the NA4B and NS5A proteins, and the NS5B RNA-dependent RNA polymerase.

At both ends of the open reading frame lie the 5'- and 3'-untranslated regions (5'-UTR and 3'-UTR). The 5'-UTR is highly conserved among different HCV isolates and it is composed by four highly ordered domains numbered I to IV. Domains II and III constitute the IRES that is essential for cap-independent translation of the viral RNA. Domains I and II are both essential for HCV RNA replication [74]. The 3'-UTR is composed of a short variable region, a poly (U/UC) tract with an average length of 80 nucleotides, and an almost invariant 98 nucleotide RNA element, designated the X-tail. The conserved elements in the 3'-UTR, including a minimal poly (U) tract of about 25 bases, are essential for replication [75].

Besides the 5'- and 3'-UTRs, an essential cis-acting replication element (CRE) was identified in the sequence that encodes the C-terminal region of NS5B [76].

Virion structure

Based on filtration and electron microscopic studies, HCV particles are 40–70 nm in diameter. It is thought that the core protein and the envelope glycoproteins E1 and E2 are the principal protein components of the virion. E1 and E2 are presumably anchored to a host cell-derived double-layer lipid envelope that surrounds a nucleocapsid composed of multiple copies of the core protein and the genomic RNA (fig. 15).
HCV circulates in various forms in the infected host. Virus can be associated with low-density lipoproteins (LDL) and very-low-density lipoproteins (VLDL), both of which seem to represent the infectious fraction. It also circulates as free virions or bound to immunoglobulins [77].

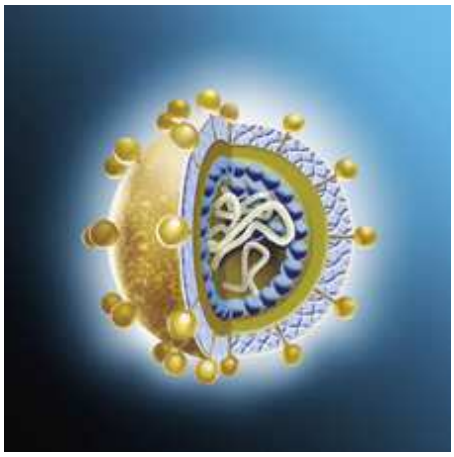


Fig. 15 Structure of HCV virion.

Structure and function of the viral proteins

Core

The first structural protein encoded by the HCV open reading frame is the core protein, which forms the viral nucleocapsid. An internal signal sequence located between the core and E1 sequences targets the nascent polypeptide to the endoplasmic reticulum (ER) membrane for translocation of the E1 ectodomain into the ER lumen. Cleavage of the signal sequence by signal peptidase yields an immature 191 aminoacids core protein. Further C-terminal processing yields the mature 21-kDa core protein of 173–179 aminoacids.
The N-terminal hydrophilic domain (D1) of the core protein contains a high proportion of basic aminoacid residues and has been implicated both in RNA binding and homo-oligomerization. The core protein is a α-helical protein that is found on membranes of the ER and on the surface of lipid droplets. The association with lipid droplets, which is mediated by the central, relatively hydrophobic domain (D2), may have a role during viral replication and/or virion morphogenesis [78].

Envelope glycoproteins

The envelope proteins E1 and E2 are glycosylated and form a non-covalent complex, which is the building block for the viral envelope. HCV glycoprotein maturation and folding is a complex process that involves the ER chaperone machinery and depends on disulphide bond formation as well as glycosylation. The transmembrane domains of E1 and E2, located at their C-termini, are involved in heterodimerization and have ER retention properties. Each of them contains a hydrophobic patch that functions as an internal signal peptide for the downstream E2 and p7 proteins [79].

p7

p7 is a 63 aminoacid polypeptide that is often incompletely cleaved from E2. HCV p7 is not required for RNA replication *in vitro* but is essential for productive infection *in vivo*. It has been suggested to belong to the viroporin family, and that could have an important role in viral particle maturation and release [80].

NS2–3 protease

The NS2–3 protease is also known as the autoprotease. As with all of the HCV proteins, NS2 is associated with intracellular membranes. It is dispensable for RNA replication *in vitro* but is essential for the complete replication cycle *in vitro* and *in vivo*, possibly by affecting a late step of the viral life cycle [81].

NS3–4A complex

NS3 is a multifunctional protein, with a serine protease located in the N-terminal one third and an RNA helicase/NTPase located in the C-terminal two thirds of the protein. The NS4A polypeptide functions as a cofactor for the NS3 serine protease. Its central portion is incorporated as an integral component into the enzyme core, and its N-terminal portion is responsible for membrane association of the NS3–4A complex [82].

NS4B

NS4B is a relatively poorly characterized 27-kDa protein. One of its functions is to induce the formation of the membranous network, the specific membrane alteration that serves as a scaffold for the HCV replication complex [83].

NS5A

NS5A is a phosphoprotein that can be found in basally phosphorylated (56 kDa) and hyperphosphorylated (58 kDa) forms. Phosphorylation of NS5A is a conserved feature among hepaciviruses and pestiviruses and is also found in flavivirus NS5 proteins. Thus, NS5A has an important role in the HCV lifecycle and probably modulates the efficiency of HCV RNA replication. NS5A would attach the viral RNA onto intracellular membranes and coordinate its different fates during HCV replication [84].

NS5B

NS5B RNA-dependent RNA polymerase (RdRp) is the key enzyme responsible for HCV replication. First, a complementary negative-strand RNA is synthesized using the genome as a template and then, a genomic positive-strand RNA is synthesized from this negative-strand RNA template. This enzyme lacks of a proof-reading function; this feature, together with the high replicative activity, accounts for genetic variability of HCV. NS5b RdRp has been extensively characterized, and has emerged as a major target for antiviral intervention. [85].

ARFP/F proteins

An alternative reading frame (ARF) was identified in the HCV core coding region, as a result of a –2/+1 ribosomal frameshift in genotype 1a. It has the potential to encode a protein of up to 160 amino acids, designated ARFP (alternative reading frame protein) or F (frameshift) protein. Amino-acid sequencing indicated that the frameshift probably occurs at, or near to, codon 11 of the core protein sequence [86].
However, the functions, if any, of the ARFP/F proteins in the life cycle and pathogenesis of HCV remain to be elucidated.

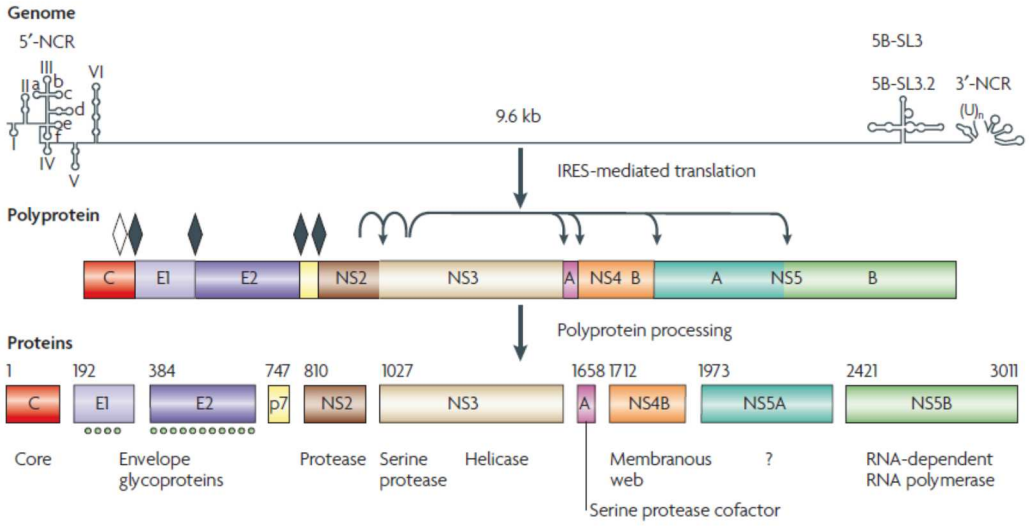


Fig. 16 Genetic organization and polyprotein processing of hepatitis C virus (HCV). The 9.6-kb positive-strand RNA genome is schematically depicted at the top. Simplified RNA secondary structures in the 5′- and 3′-UTR and the core gene are shown. Internal ribosome entry site (IRES)-mediated translation yields a polyprotein precursor that is processed into the mature structural and non-structural proteins. Amino-acid numbers are shown above each protein (HCV H strain; genotype 1a; GenBank accession number AF009606). Solid diamonds denote cleavage sites of the HCV polyprotein precursor by the endoplasmic reticulum signal peptidase. The open diamond indicates further C-terminal processing of the core protein by signal peptide peptidase. Arrows indicate cleavages by the HCV NS2–3 and NS3–4A proteases. Dots in E1 and E2 indicate the glycosylation of the envelope proteins [75].

4.1.4. Life cycle

Viral entry

HCV only infects humans and chimpanzees. Hepatocytes are the main target cells but infection of B cells, dendritic cells and other cell types has also been reported. CD81, a tetraspanin protein that is found on the surface of many cell types, including hepatocytes, the LDL receptor (LDLR), the scavenger receptor class B type I (SR-BI) and, most recently, claudin-1 (CLDN1) have, among others, been proposed as HCV receptors [87, 88, 89], but their role remains still not fully elucidated. CLDN1 functions at a late stage of cell entry, possibly at tight junctions of polarized hepatocytes.
Virus internalization depends on clathrin-mediated endocytosis and the acidification of the endosome induces HCV glycoprotein membrane fusion. Little is known about the uncoating process, which results in genome release into the cytosol.

Translations of viral proteins

Domains II, III and IV of the 5'-UTR, together with the first 24–40 nucleotides of the core coding region, constitute the IRES. HCV translation initiation occurs through the formation of a binary complex between the IRES and the 40S ribosomal subunit. Then, a 48S-like complex is assembled at the AUG initiation codon, after the association of eukaryotic translation initiation factor3 (eIF3) and ternary complex (eIF2•Met-tRNAi•GTP) [75].

Polyprotein processing

Translation of the HCV open reading frame yields a polyprotein precursor that is co- and post-translationally processed by cellular and viral proteases into the mature structural and non-structural proteins. The structural proteins and the p7 polypeptide are processed by the ER signal peptidase whereas the non-structural proteins are processed by two viral proteases, the NS2–3 protease and the NS3–4A serine protease [75].

The viral replication complex

The peculiar steps of HCV replication are the formation of a membrane-associated replication complex, the composition of viral proteins, replication of RNA and alteration of cellular membranes. Replication might occur on altered membranes derived from the ER, Golgi apparatus, mitochondria or even lysosomes.
The role of membranes in viral RNA synthesis is not well understood. Several functions have been suggested: physical support and organization of the RNA replication complex, compartmentalization and local concentration of viral products, attachment of the viral RNA during unwinding, provision of lipid constituents important for replication and protection of the viral RNA from double-strand RNA-mediated host defenses or RNA interference [75].

Recent studies have revealed a complex interaction between HCV RNA replication and cellular lipid metabolism, presumably through the trafficking and association of viral and host proteins with intracellular membranes [90].

Packaging, assembly and particle release

Little is known about the late steps of the viral lifecycle. NS2 and possibly other non-structural proteins, as well as undefined RNA structures, are involved in these processes. Virions presumably form by budding into the ER, or an ER-derived compartment, and exit the cell through the secretory pathway [75].

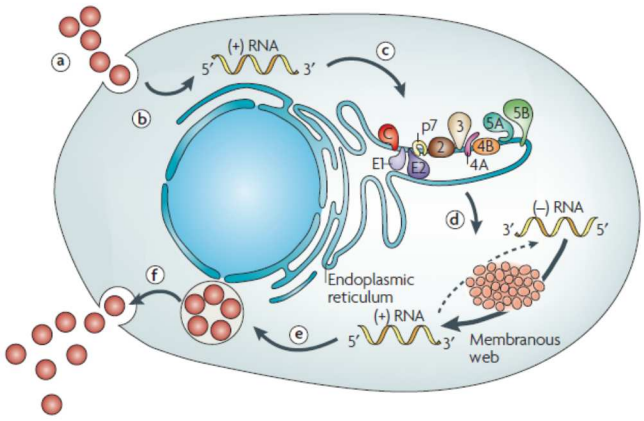


Fig. 17 Lifecycle of hepatitis C virus (HCV). Virus binding and internalization (a); cytoplasmic release and uncoating (b); IRES-mediated translation and polyprotein processing (c); RNA replication (d); packaging and assembly (e); virion maturation and release (f) [75].

4.1.5. Infections by HCV

The World Health Organization (WHO) estimates that 170–200 million people worldwide, i.e., 3% of the world's population, is infected with HCV.

HCV prevalence is characterized by a high variability between world's regions, individual countries and between age and risk groups. HCV prevalence is highest in Africa and the Middle East, where Egypt, Cameroon, Saudi Arabia, Iraq and Syria account for the majority of cases and prevalence ranges from 2% to 15%. North America, Australia, Japan and Northern and Western Europe report lower prevalence of HCV infection, with no country showing a rate > 2%. China, India, Egypt, Pakistan and Indonesia account for approximately half of the global HCV-infected subjects [91]. In general, developing countries present the major HCV-related burden but also the major limitations in surveillance: data from most African, Asian and South American countries are lacking.

HCV prevalence in the majority of developed countries is classified as low, but marked differences in the epidemiological picture exist among countries, principally related to

temporal and transmission factors and resulted in diverse age-specific distribution of HCV cases [92].

Natural history of HCV

Acute HCV infection is mostly asymptomatic and rarely recognized clinically.
In up to 45% of cases, acute HCV infection completely resolves, and this seems to be associated with a younger age at infection, female sex, and possibly certain major histocompatibility complex genes. However, about 55% to 86% of HCV infected patients develop chronic infection, manifested by the persistence of detectable HCV in the serum. This has been primarily attributed to the propensity of HCV to mutate and evade host defenses [93].
Chronic HCV is usually characterized by a lack of symptoms or only fatigue or vague abdominal pain. Extra hepatic manifestations of chronic HCV may be identified, and these are associated primarily with autoimmune or lymphoproliferative states. Increases in serum alanine aminotransferase (ALT) reflect hepatocyte injury, but these values typically fluctuate overtime and may be even normal on occasion.
The major complication of chronic HCV infection is progressive hepatic fibrosis leading to cirrhosis, which develops in about 20% of patients with chronic HCV [94]. The natural history of chronic HCV is variable, and progression of chronic liver disease is insidious in most patients. About one third of patients with chronic HCV develop hepatic cirrhosis 15 to 20 years after infection (‘‘rapid fibrotic progressors’’), one third develop cirrhosis 20 to 30 years after infection (‘‘intermediate fibrotic progressors’’). The remaining one third develop it only after 30 years of HCV infection (‘‘slow fibrotic progressors’’) [95].
However, a number of factors can accelerate progression to advanced liver disease, such alcohol consumption, co-infection with HIV or hepatitis B virus, and older age at the time of infection. Obesity and hepatic steatosis are also emerging independent predictors of more severe liver fibrosis.
Chronic HCV infection is associated with an increased risk of hepatocellular carcinoma, but this occurs primarily in patients with cirrhosis. Factors that increase the risk of hepatocellular carcinoma among HCV infected people include male sex, older age, chronic hepatitis B infection, and heavy alcohol ingestion (>50 grams/day). Once cirrhosis is established, the incidence of hepatocellular carcinoma is 1% to 4% per year. Death from chronic HCV typically occurs because of decompensated cirrhosis or hepatocellular carcinoma [94].

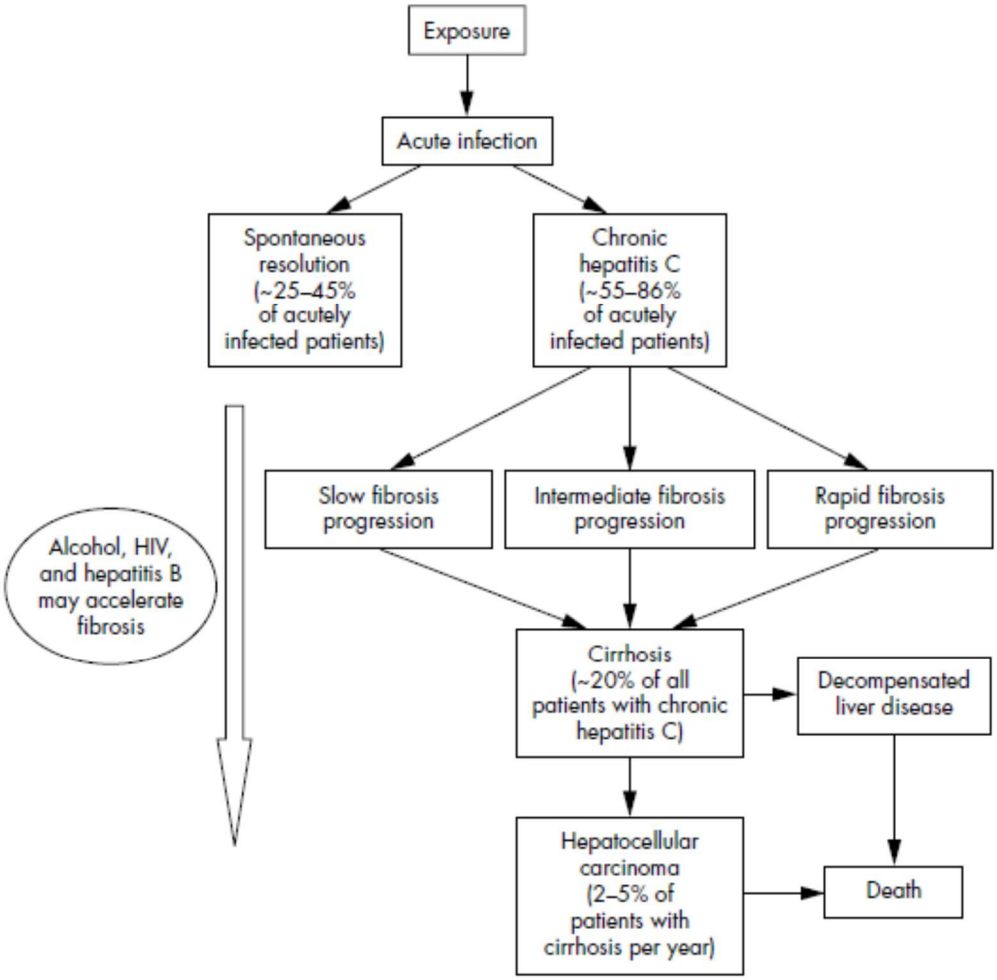


Fig. 18 Natural history of HCV infection [94].

4.1.6. Therapy

The treatment available for HCV has changed significantly over recent decades, with the standard of care shifting from conventional interferon (IFN) monotherapy to IFN and ribavirin combination therapy to PEGylated IFN (PEG-IFN) with ribavirin. IFNα has potent antiviral activity due to its ability to induce IFN-stimulated genes that encode proteins, inhibiting various stages of viral replication. In addition, IFNα has an immunomodulatory effect, interacting with both the adaptive and innate immune response of the host. IFNα promotes T-helper (Th) cell differentiation of T-lymphocytes over Th2 cells, leading to increased production of interleukin (IL)-2 and IFNγ. Moreover, IFNα exerts an anti-inflammatory effect by inhibiting the synthesis of various cytokines, including tumor necrosis factor-alpha and IL-1 [96].

The mechanisms of action of ribavirin are not fully understood. It has been postulated that ribavirin acts via direct inhibition of HCV replication, inhibition of the host inosine monophosphate dehydrogenase enzyme, induction of mutagenesis to drive a rapidly

replicating virus beyond the threshold to error catastrophe, and immunomodulation by inducing a Th1 immune response [97].

The primary goal of antiviral therapy in patients with chronic hepatitis C is achieving a Sustained Virological Response (SVR), defined as undetectable serum HCV-RNA by a sensitive molecular assay 24 weeks after completion of therapy. Although the standard of care improves SVR rates in HCV genotypes 2 and 3, the response is still suboptimal in genotypes 1 and 4 and in particular patient populations; therapy fails in 50%–60% of patients with HCV genotype 1 and approximately in 20% of those with HCV 2 and 3.

Management of relapsers and nonresponders remains a challenging and controversial issue. In addition, all of the IFN-based regimens have moderate to severe side effects, including hematologic adverse events (neutropenia, thrombocytopenia), fatigue, irritability, fever, myalgia, arthralgia, inflammation at the injection site, and cardiac dysrhythmia, that negatively influence the tolerability and adherence of patients with therapy [98].

Moreover, a number of host and viral factors influence SVR rates in patients with chronic HCV. A SVR is more likely in young individuals, females, patients infected with genotypes 2 or 3, and those with low pretreatment HCV-RNA levels, no or minimal liver fibrosis, and adequate adherence to therapy. Infection with HCV genotype 1 or 4, high baseline HCV RNA levels (>800,000 IU/mL), steatosis, insulin resistance, and co-infection with HIV are associated with low response rates [99].

Since genotype 1 is the most frequent genotype in chronically infected patients worldwide, the need for more efficacious therapies is becoming urgent.

All of the above factors have driven a need to develop new treatments that are safer and more effective. Recently, a number of direct-acting antiviral agents (DAAs) have been developed for use with PEG-IFN/ribavirin as triple therapies or IFN-free therapy. The efficacy of such therapeutic regimens varies according to genotype and host characteristics.

Direct-acting antiviral agents

DAAs were developed to increase SVR rates, reduce adverse events, and improve adherence to therapy in HCV patients.

Increased understanding of the HCV life cycle in recent years has supported the development of direct-acting antiviral (DAA) agents that specifically target post-translational processing and HCV replication.

The targets of currently approved or in development molecules are related with HCV replication, specifically translation and polyprotein processing (NS3/4A), HCV genome replication (NS5B polymerase and NS5A), and viral assembly (NS5A) [100].

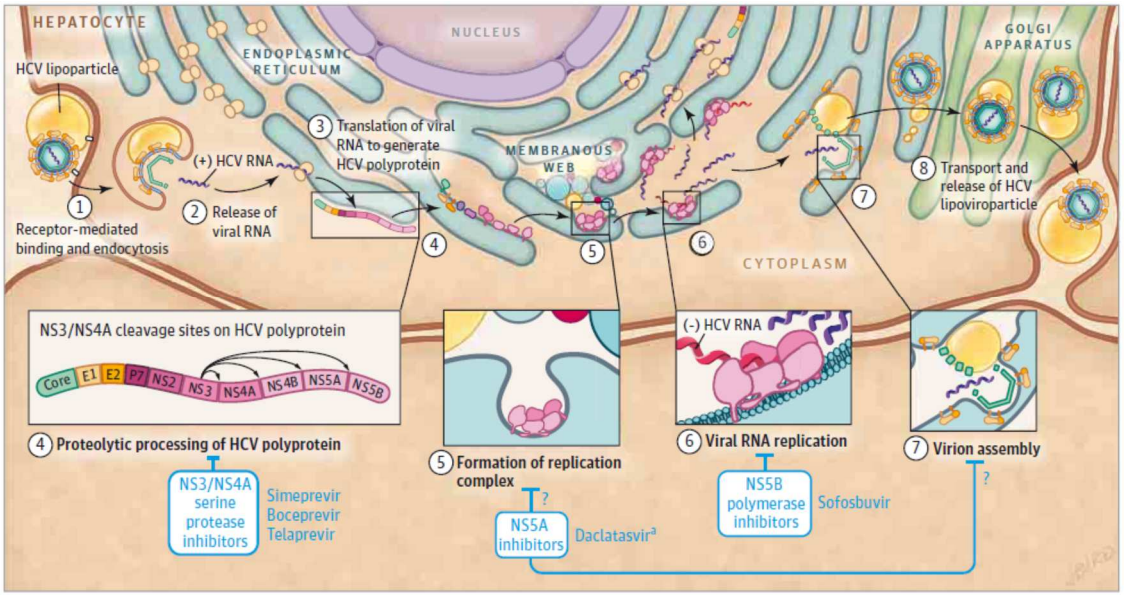


Fig. 19 Steps in the HCV life cycle targeted by DAAs (shown in cyan). Abbreviations: E, enevelope glycoprotein; NS, nonstructural protein; + and -, positive and negative HCV RNA strands. [a]Not approved by the US Food and Drug Administration [102].

NS3/4A Inhibitors

Inhibition of NS3 and its cofactor, NS4A, results in blocking proteolytic maturation of a large portion of the nonstructural region of the HCV polyprotein, NS3 to NS5B.

A number of other protease inhibitors, which have been developed and in phase II or III clinical trials, are classified as “first-generation” and “second-generation” according to degree of genetic barrier to resistant HCV and genotype coverage. The first-generation protease inhibitors include boceprevir, telaprevir, simeprevir (TMC-435), faldaprevir (BI201335), vaniprevir (MK-7009), and asunaprevir (BMS-650032).

The second-generation protease inhibitors, characterized by potent activity against pan-genotypes and high genetic barrier to resistance, include MK-5172 and ACH-2684 in phase II clinical trial [102].

Boceprevir and telaprevir represents the first-wave of the first-generation oral protease inhibitors and they have been assessed in large clinical trials. These agents have been approved by regulatory authorities and are currently used in clinical practice. Boceprevir acts as a noncovalent inhibitor of cytochrome P450 A4 and P-glycoprotein. Addition of boceprevir or telaprevir to PEG-IFN and ribavirin significantly increased SVR rates and shortened the treatment duration in naïve, relapsing, and partially responding patients. Current practice guidelines recommend a triple therapy regimen combining PEG-IFN, ribavirin, and telaprevir or boceprevir. However, triple therapy has some drawbacks, including drug–drug interactions and viral resistance. This regimen increases adverse events such as rash and moderate to severe anemia to an extent that might require reduction of the ribavirin dose. Boceprevir and telaprevir are only effective against genotype 1, with recent studies showing that these protease inhibitors have no antiviral

activity against genotype 2, 3, or 4. Further, triple therapy is ineffective in patients who have not responded to previous dual PEG-IFN/ribavirin therapy [98].

The second-wave of first-generation protease inhibitors comprises simeprevir, asunaprevir, and danoprevir, which are currently being evaluated in an effort to overcome the limited efficacy of the first-wave protease inhibitors in HCV genotypes 2, 3, and 4 and to minimize their adverse events [98].

NS5A Inhibitors

NS5A is a dimeric protein required for HCV RNA replication and virion assembly. NS5A inhibitors have potent antiviral activity, but the genetic barrier to resistance is low.

Daclatasvir (BMS-790052) is the first of DAA targeting against hepatitis C virus NS5A showing a very potent antiviral effect on several HCV genotypes. The overall adverse event profile is acceptable. Due to a relatively low genetic barrier, combination regimen including daclatasvir and other NS3/4A, PEG-IFN/RBV, or NS5B drugs is recommended for the treatment of hepatitis C [103].

NS5B Inhibitors

Polymerase inhibitors are another class of DAAs that have recently shown much potential. These drugs bind to NS5B polymerase to stop replication of the virus.

They can be divided into 2 groups: the nucleoside analog inhibitors, that are incorporated into the HCV RNA chain, leading to direct chain termination, and the non-nucleoside inhibitors that bind to several discrete sites outside of the polymerase active center, causing a conformational protein change.

The nucleoside analog inhibitors are potentially active against all HCV genotypes, and viral resistance to these agents is low and less frequent than with the non-nucleoside inhibitors.

Sofosbuvir is a nucleoside analog inhibitor and has recently been approved by the US Food and Drug Administration. It has a high barrier to viral resistance, and no virologic breakthrough has been recorded so far. One major feature of sofosbuvir is its pan-genotypic antiviral effect [104]. Actually, sofosbuvir has opened the first window for the era of “interferon-free” treatment of hepatitis C. Many clinical trials of sofosbuvir are now ongoing to optimize the regimen and treatment duration for each HCV genotype.

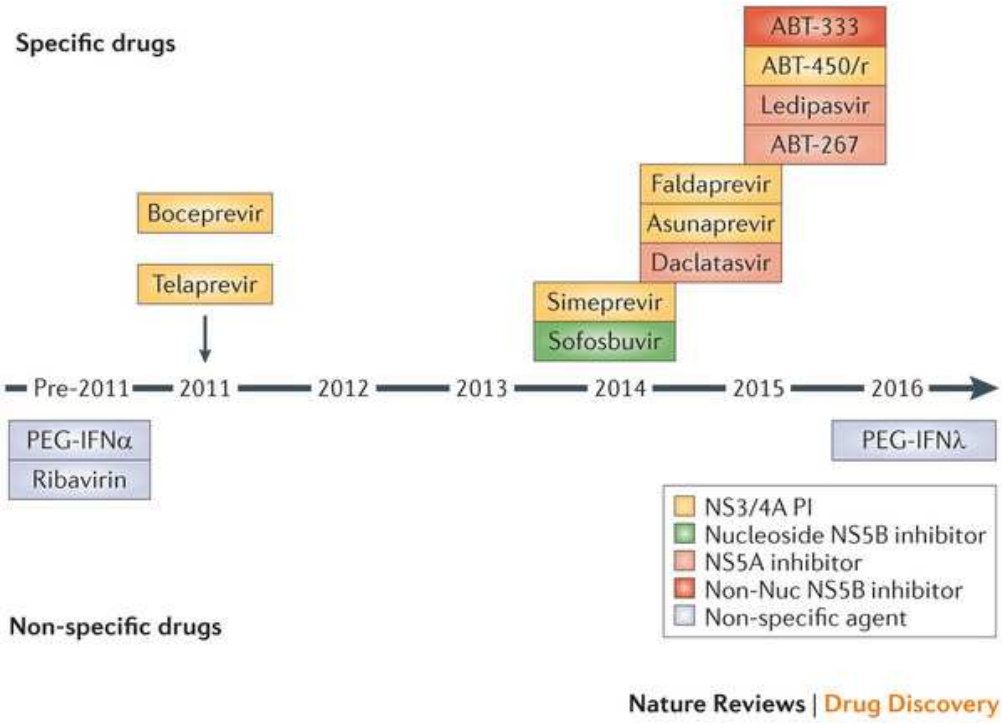


Fig. 20 A schematic of drugs approved for treating hepatitis C virus (HCV) infection as well as drugs in advanced development with tentative future launch dates. PEG-IFN, peglyated-interferon; r, ritonavir [106].

4.1.7. Diagnosis

Diagnosis and management of HCV infection can be achieved through two major types of assays: direct methods, that detect viral components (e.g., the core antigen or the viral genome) and indirect assays, which detect HCV specific antibodies.

Indirect diagnosis

Serologic assays detect HCV specific total antibodies (IgM and IgG) and they are used to screen and diagnose HCV exposure. However, these assays do not discriminate between active and resolved infections.
Serological methods can be divided into screening assays and confirmatory methods.

Screening assays

Since 1989, when HCV was discovered and its immunodominant epitopes were identified, HCV infection has been mainly diagnosed by detecting HCV antibodies from serum samples using enzyme immunoassays (EIA). Over time, serologic assays have evolved, and current third-generation assays now include multiple recombinant HCV antigens from the core, NS3, NS4 and NS5 regions. This has resulted in the reduction of the window period and in an overall improved detection of patients exposed to HCV.

Confirmatory assays

Recombinant immunoblot assays (RIBA) can be used to confirm the presence of HCV specific antibodies for individuals who have tested positive by EIA, especially in the screening populations with a low prevalence of HCV infection. This assay is highly specific, as the presence of antibodies against each of several HCV proteins is assessed as individual bands on a membrane strip [106].

Direct diagnosis

The detection of viral components is needed to diagnose an active HCV infection. Direct methods include molecular HCV assays and quantification of HCV Core antigen.

Molecular HCV assays

Molecular assays to detect the HCV genome are used for several purposes in the clinical setting. First, the presence of circulating HCV-RNA reflects viral replication, such that sensitive molecular assays (with a lower limit of detection < 50 IU/mL) are used to diagnose active HCV infection in patients with a positive antibody test. Second, molecular testing is required for an early diagnosis of acute HCV infection, as the HCV-RNA can be detected before specific antibodies become detectable (within 1-3 weeks after exposure). Finally, the diagnosis of a chronic HCV infection is confirmed by the presence of both HCV antibodies (with the exception of severely immunosuppressed patients) and HCV RNA over 6 months [107].

The highly conserved HCV 5'UTR region is the target of choice for HCV genome detection across different genotypes. Real-time reverse-transcription PCR (RT-PCR) is the method of reference for the quantification of HCVRNA levels in clinical practice according to European and American guidelines, given its high sensitivity and wide dynamic range of quantification.

Quantitative tests are also used to monitor antiviral therapy. In order to minimize side-effects, emergence of resistance and costs, HCV-RNA must be periodically quantified to strictly follow treatment stopping rules.

With the advent of new treatment regimens that include a protease inhibitor and the response-guided treatment algorithms, only assays with a lower limit of quantification of $\leq$ 25 IU/mL and a lower limit of detection of approximately 10-15 IU/mL, should be used. Additionally, the presence of detectable but not quantifiable HCV-RNA below those levels is clinically relevant, as it reflects true viremia [108].

HCV Core antigen detection and quantification

The HCV Core antigen can be detected in the serum of HCV-infected patients, and its levels are significantly related to those of HCV-RNA [109].

HCV Core assays could be used as an alternative to HCV-RNA assays for three different situations:

(1) to distinguish active from resolved HCV infections;

(2) to identify HCV infection in the antibody window period;
(3) to identify HCV infection in seronegative individuals at high risk for HCV infection, such as in hemodialysis patients.
When HCV infection is confirmed, the following step is the determination of HCV genotype, in order to apply the correct therapeutic regimen.

HCV genotyping assays

HCV genotype is considered to be the major baseline predictor of a sustained virological SVR to IFN-α-based therapy. Moreover, new DAAs (e.g. boceprevir and telaprevir), already available, have shown different efficacies on various HCV genotypes.
While subtyping is not considered to be clinically relevant for PegIFN-α and RBV treatment regimens, it may be relevant in the era of directly acting antivirals.
For instance, subtypes 1a and 1b have been described to have subtype-specific resistance profiles to linear protease inhibitors [105].
Since the HCV genotype is predictive of the response to, genotyping is mandatory to tailor dose and duration of treatment. Furthermore, it is necessary for deciding on triple therapy eligibility with currently approved protease inhibitors, which are effective against HCV genotype 1.
The gold standard for HCV genotyping is genome sequencing of the NS5B region and subsequent phylogenetic analysis [62]. However, this in-house method is restricted to reference centers.
Actually, many commercial assays for HCV genotyping are available and they are based on Real-Time technology or upon RLB technique.
The target of choice for commercially available genotyping assays has classically been the highly conserved 5'UTR. This region allows a well differentiation between several HCV genotypes (1 to 7a), except genotype 1 from 6, subtypes c to l. Moreover, discrimination among subtypes 1a and 1b is not always reliable using this region.
Pickett *et al.* demonstrated that exist two distinct clades within the 1a subtype (clade I and II) with each clade having a star-like tree topology and lacking definite correlation between time or place of isolation and phylogeny. Identification of significant phylogenetically-informative sites at the nucleotide level revealed positions not only contributing to clade differentiation, but which are located at or proximal to codons associated with resistance to: protease inhibitors (NS3 Q41) or polymerase inhibitors (NS5B S368). Synonymous/non synonymous substitution mutation analyses revealed that the majority of nucleotide mutations yielded synonymous amino acids, indicating the presence of purifying selection pressure across the polyprotein with pockets of positive selection also being detected [110].
Moreover, Chevaliez demonstrated that the two clades share some nucleotides with subtype 1b in the 5'UTR, in particular in positions 107 and 204 (fig. 21). At position 107 subtype 1a clade I exhibit a G, as subtype 1b, leading to a msidiagnosis.

At position 204, A is the most frequent nucleotide for subtype 1a clade I, whereas C is the most frequent nucleotide for subtype 1a clade II, and C or T are the most frequent nucleotides for subtype 1b. The usual presence of a C at position 204 in subtype 1a clade II explains why misclassifications were far more frequent with this clade than with subtype 1a clade I [111].

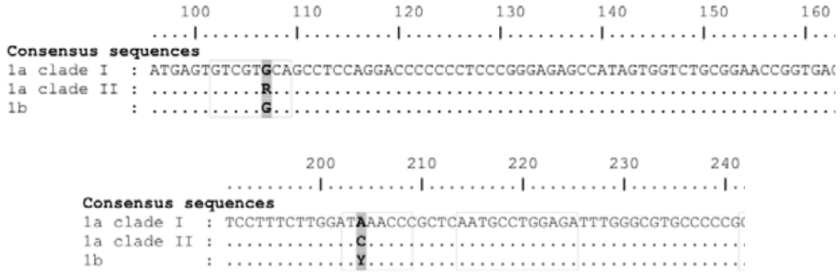


Fig. 21 Alignment of the 5' UTR consensus sequences of subtype 1a clade I (1a-I), subtype 1a clade II (1a-II) and subtype 1b. Positions 107 and 204, that differentiate subtypes 1a and 1b are in bold [111].

The analysis of the sole 5'UTR is not sufficient to allow a good discrimination between subtypes 1a and 1b, and a differentiation of genotype 1 and 6. Thus, assays targeting other regions in addition to the 5'UTR have been recently developed.

The two regions most frequently chosen for the combined analysis with the 5'UTR are NS5b and core regions. They are informative to discriminate subtypes and genotypes 1 to 6, but also sufficiently conserved to allow primers annealing.

Commercialized genotyping assays may result in <5% of indeterminate results due to the high genetic variability of HCV. Clinicians are forced to treat patients with an indeterminate result as if they were infected with genotypes 1 or 4 (resulting in longer treatment duration and higher ribavirin doses than for genotypes 2 and 3) and cannot decide on triple therapy eligibility. Therefore, patients with an indeterminate genotype result should be retested using either an alternative commercial assay or the reference method.

4.2. MATERIALS AND METHODS

4.2.1. Design of HCV genotyping test prototype

A prototype for HCV genotyping (AMPLIQUALITY HCV TYPE PLUS) was planned and manufactured, after an *in silico* study for primers and probes selection.
An approach of two viral regions combined analysis was chosen. The first selected region was 5'UTR, because it is the most conserved one and it allows a good discrimination between genotypes. The choice of second region was based on bibliographic and bioinformatic research. Targets as NS5b, E1 and core were evaluated. NS5b and E1 regions are too polymorphic for this type of assay. Their high variability could affected primers annealing and led to a failure of test. On the other hand, core region contains both polymorphic and conserved regions, ideal for primers and genotyping probes design.
The new prototype for HCV genotyping targets 5'UTR and core regions, to allow a good discrimination between genotypes 1 to 7a and subtypes 1a and 1b.

4.2.2. HCV sequences databases

In order to create a dataset of sequences for each HCV genotype and subtype, suitable for the design of specific primers and probes, different databases were used.
All available HCV partial and whole genome sequences were obtained from:

1. Virus Pathogen Database and Analysis Resources-ViPR (www.virpr.org)
2. BLAST tool (www.ncbi.nlm.nih.gov/blast/).

ViPR is a publicly available database and analysis resource. ViPR integrates data from external sources (GenBank, UniProt, Immune Epitope Database, Protein Data Bank), direct submissions, internal curation and analysis pipelines. It provides a suite of bioinformatics analysis and visualization tools and allows several types of studies on pathogenic viruses.
The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The software compares nucleotide or protein sequences to databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences, as well as help identify members of gene families.
Sequences of complete or partial genomes were obtained for all the 7 genotypes and for the main subtypes. Each HCV subtype is represented by a certain number of single isolates, and this number varies consistently among different subtypes, in particular for subtype 1a and 1b, being the most frequent ones. (tab. 18)

GENOTYPE/SUBTYPE	COMPLETE GENOME SEQUENCES	PARTIAL GENOME SEQUENCES	TOTAL OF SEQUENCES
1a	409	3805	4214
1b	204	4283	4487
2 (all subtypes)	113	345	458
3 (all subtypes)	30	1904	1934
4 (all subtypes)	30	675	705
5 (all subtypes)	2	72	74
6 (all subtypes)	17	553	570
7a	-	1	1

Tab. 18 Sequences obtained by ViPR Database, divided on the basis of HCV genotype or subtype.

4.2.3. Design of RT-PCR

Design of primers

Design and analysis of primers can be divided into three phases:

1. Alignment of all available sequences for 5'UTR and core region of the seven genotypes.
2. Design of primer pairs.
3. Analysis of primers.

The sequence alignment was performed by using two bioinformatics tools:

- MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm in the ViPR database (www.virpr.org), to align sequences belonging from the same genotype.
- ClustalW2 software (www.ebi.ac.uk/Tools/msa/clustalw2/), to align sequences belonging from the same subtype.

Because of high sequences variability, no one software was used for primer design. For each alignment, conserved part of sequences were identified as possible regions for primer design. Shared sequences between all genotypes were analyzed in order to evaluate consensus primers both for the 5'UTR and the core regions.
Since HCV variability is very high, some isolates cannot be amplified by consensus primers, because of high number of SNPs in the sequence. In this case, design of specific primers was preferred to degenerate ones.

Primers analysis was performed by using 3 bioinformatic tools:

- Primer Express® Software v3.0.1 (Life Technologies);
- DNA Mate (http://melolab.org/dnaMATE/tm-pred.html);
- BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Primer Express® and DNA Mate software were used to analyze melting temperature, content of GC, amplicon length and the probability of secondary structures formation.

Primers have been also analyzed by using BLAST tool:

- to avoid no specific alignment with other regions different from the target sequence, not only in the HCV genome, but also in other microorganisms and human genome;
- to check if all HCV isolates can be amplified by selected primers or if there was some SNP that could affect primer annealing.

All the selected primers satisfy some common features:

- specificity for all isolates of the genotype
- primer length: 20-28bp
- amplicon length: 200-300bp
- no dinucleotide repeats or long stretches of guanosine (G)
- less than 3bp difference in length between primer pairs
- melting temperatures (Tm) between 65°C and 70°C
- GC content between 50% and 60%
- balanced distribution of GC-rich and AT-rich domains
- no intra-primer homology (more than 3 bases that complement within the primer) or inter-primer homology (forward and reverse primers having complementary sequences) to avoid self-dimers or primer-dimer formation.

Sequences of designed primers are not reported, for respecting intellectual proprietary and trade secret.

RNA internal control (IC-RNA)

Diagnostic assays require an internal control (IC), in order to monitoring the presence of inhibitors or other factors which may cause false-negative results. Internal controls are used as indicator of good nucleic acid extraction, quality of samples and PCR. In case of clinical samples from human, the internal control will indicate the samples have been collected, transported and stored properly. An IC for diagnostic RT-PCR assays should be easy to produce and to standardize. Additionally, ICs should be stable, noninfectious, absent from clinical samples, and suitable for different assays.

An endogenous IC is a template that occurs naturally within the specimen being analyzed. In gene expression analysis and virus screenings, housekeeping genes are often used as ICs and references for transcript quantification, but they have to be proven for each experiment and target. Exogenous ICs are added before nucleic acid isolation (extraction control) or amplification (amplification control), where co-amplification is performed within the same reaction. Ideally, these ICs hybridize to the same primers, have identical amplification efficiencies, and contain discriminating features, such as length or sequence variations, targeted by hybridization probes. However, these competitive ICs can lower the amplification efficiency, which results in a lower detection limit. Therefore, noncompetitive IC templates are used, where the target and IC are

amplified with different primer sets. The disadvantage is that amplification of the IC may not accurately reflect amplification of the target.
For viral nucleic acid amplification tests (NAT), the detection of model viruses, such as *Escherichia coli* bacteriophages, has been described. In these approaches, clinical specimens were spiked with a known amount of an animal virus to monitor the efficiency of extraction, reverse transcription, and amplification. The advantage of such model viruses is the stability of RNA and the control of decapsulation of the viral RNA during the extraction procedure [112].
Since human serum is not a cellular matrix, the amplification of a housekeeping gene is not possible and therefore, the addition of an external amplifiable nucleic acid in the PCR assay serves as an internal control. We decide to design, for our assay, an IC able to monitoring all the analytical protocol: from viral RNA extraction to strip visualization. For this purpose, an *E. coli* phage was selected as IC. It is resistant to RNase degradation, even at high storage temperatures, sufficiently stable for routine use and it does not interfere with the multiplex RT-PCR.
The *E. coli* phage genome sequence is absent from the human specimens, cell cultures, and veterinary samples. The IC is reverse-transcribed and amplified in the same multiplex reaction of the target pathogen; primers and relative probe have been designed. Our IC did not affected efficiency of amplification.
Sequences of IC primer pairs and probe designed are not reported, for respecting intellectual proprietary and trade secret.

RT-PCR reaction design

Reverse transcription polymerase chain reaction (RT-PCR) is one of the variants of polymerase chain reaction (PCR). In RT-PCR, the RNA template is first converted into a complementary DNA (cDNA) using a reverse transcriptase. The cDNA is then used as a template for exponential amplification using PCR.
This process can be achieved as either a one-step or a two-step reaction. In the one-step approach, the entire reaction from cDNA synthesis to PCR amplification occurs in a single tube. The one-step approach minimizes experimental variation by containing all of the enzymatic reactions in a single environment. However, the starting RNA templates are prone to degradation in the one-step approach, and the use of this method is not recommended when repeated assays from the same sample is required.
On the other hand, the two-step reaction requires that the reverse transcriptase reaction and PCR amplification be performed in separate tubes. The disadvantage of the two-step approach is susceptibility to contamination due to more frequent sample handling.
A multiplex one-step approach was decided for our test. The entire reaction, from RNA to amplified products, is made in a single tube, without intermediated manipulations. Both reverse transcription and amplification of cDNA are performed by using specific

primers for 5'UTR and core HCV regions in a unique reaction. The run time of the reaction is 2 hours and 10 minutes; the detailed thermal profile is reported in table 19.

Hold	Temperature	Time	Number of cycles
Reverse transcription	48°C	30 minutes	1
HotStart	95°C	10 minutes	1
Denaturation	95°C	30 seconds	45
Annealing and Extension	60°C	90 seconds	
Storage	10°C	∞	1

Tab. 19 Thermal profile used for the multiplex reverse transcription and amplification of 5'UTR and core HCV regions.

Several mixtures of enzymes have been commercially available: these mixes contains both the Reverse Transcriptase and the Hot Start Taq polymerase.
Reverse Transcriptase are active in a range of temperatures between 42°C and 50°C, on the basis of their chemical features. Hot Start Taq polymerases are supplied in an inactive state and they have no enzymatic activity at room temperature. The enzyme remains completely inactive during the reverse-transcription reaction and does not interfere with it. This prevents formation of misprimed RT-PCR products and primer-dimers during reaction setup, reverse transcription and the first denaturation step. The enzyme is activated after the reverse-transcription by a denaturation step. The hot start also inactivates the Reverse Transcriptase, ensuring temporal separation of reverse transcription and PCR, allowing both steps to be performed sequentially in a single tube.
Several commercial mixtures of enzymes were tested in order to choose the best performing reagent that fitted our purpose. Commercial names of reaction mix and enzymes, qualitative and quantitative composition of oligomix are not reported, in protection of industrial secrecy.

4.2.4. RLB design

Design of probes

Probes have been designed using the alignments of 5'UTR and core regions, previously made for primers selection.
Conserved parts of sequences, delimitated by primers, have been identified both in the 5'UTR and in the core regions. Among these conserved segments, one sequence in the 5'UTR and one in the core region were selected for the design of the universal probes. These oligos ensure to identify the presence of HCV RNA, even in case of high polymorphic sequences.
Instead, discriminative parts of sequence have been identified to design genotyping probes. In the 5'UTR specific probes for all genotypes have been determined, while in the core region only probes for subtypes 1a and 1b and genotype 6 have been evaluated.

Once defined these regions, 30 nucleotides in length, genotyping probes have been designed, by using 3 bioinformatic tools:

4. *DNA Mate* (http://melolab.org/dnaMATE/tm-pred.html), to evaluate melting temperature.
5. *DNA Folding Form* application (http://mfold.rna.albany.edu/?q=mfold), to analyze the amplicon folding and to evaluate the presence of possible secondary structures that can prevent annealing with probe.
6. *Two state melting* application (http://mfold.rna.albany.edu/?q=DINAMelt), to evaluate temperature of hybridization between probe and target sequence.

All the selected probes satisfy some common features:

- average length of 20 base pairs;
- designed in regions not involved in strong secondary structures;
- 50°C as temperature of hybridization with specific sequences;
- average temperature of melting between 60° and 65°C.

Sequences of probes are not reported, in protection of industrial secrecy.

```
gi_254546399_4d   CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGA-TAAACCGCTCAAT 149
gi_239836688_4d   CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAT-AAACCGCTCAAT 149
L29625_4e         CCGGAATCGCCGAGATGACCGGGTCCTTTCTTGGA-TCAACCGCTCAAT 126
L29591_4e         CCGGAATCGCCGAGATGACCGGGTCCTTTCTTGGA-TCAACCGCTCAAT 126
EF115679_4e       CCGGAATCGCCGAGATGACCGGGTCCTTTCTTGGA-TTAACCGCTCAAT 121
Gi_239836676_4f   CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGA-TAAACCGCTCAAT 149
gi_254546397_4f   CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGA-TTAACCGCTCAAT 149
gb_JX227963.1_4g  CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 138
EF115673_4h       CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAACTAACCGCTCAAT 122
gi_239836678_4k   CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 150
gi_239836690_4k   CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAACTAACCGCTCAAT 150
gi_239836692_     CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATTAACCGCTCAATG 149
gb_JX227958.1_4L  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGA-TCAACCGCTCAAT 137
gb_JX227957.1_4L  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGA-TTAACCGCTCAAT 135
gi_239836680_4m   CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATCAAACCGCTCAAT 150
gb_JX227972.1_4m  CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATTAAACCGCTCAAT 138
gb_JX227961.1_4m  CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATCAACCGCTCAATG 138
gb_JX227970.1_4n  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGA-TTAACCGCTCAAT 137
gi_239836696_4n   CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGA-TCAACCGCTCAAT 148
gb_JX227979.1_4o  CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 97
gb_JX227978.1_4o  CCGGAATCGCCGGGACGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 97
gi_239836682_4q   CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 150
gb_JX227976.1_4r  CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATTAACCGCTCAATG 138
gb_JX227962.1_4r  CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGATTAACCGCTCAATG 138
gb_JX227960.1_4v  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAATAAACCGCTCAAT 101
gb_JX227959.1_4v  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAACAAACCGCTCAAT 148
SEQ_AB_BEP051038  CCGGAATCGCCGGGATGACCGGGTCCTTTCTTGGAT-AAACCGCTCAAT 112
Probe             CCGGAATCGCCAGGACGACCGGGTCCTTTCTTGGA---------- 35
                  ************  ** ******************
```

Fig. 22 Example of alignment between several genotype 4 isolates and selection of a discriminative region for probe design.

Strip manufacturing

The support of the strip is a nitrocellulose baked membrane. The probes are spotted on nitrocellulose through an automatic work-station and then are immobilized on support by exposure to UV light, which induces the covalent bounds between oligos and membrane.
Then, spotted membrane is incubated with a blocking solution containing BSA, for 30 minutes in moderate shaking. This step prevents non-specific background binding of the DNA to membrane.
Finally, membrane is cutted into single strips, through an automatic instrument.
Probes concentration, buffer compositions, commercial name of nitrocellulose membrane and other details are omitted, for respecting intellectual proprietary and trade secret.

Design of RLB protocol

A RLB assay is influenced by several interconnected parameters.
The stringency of the assay could be achieved by acting on two fundamental factors: temperature and ionic strenght of solutions. It is also possible to vary incubation time of the several washing steps.
Table 20 summarizes the RLB protocol of our genotyping test.

STEP	REAGENTS	INCUBATION	TIME OF INCUBATION
1) Denaturation of amplicon	Amplicon + Denaturation Solution, containing Sodium hydroxide	Room temperature	5’
2) Hybridization	Hybridization solution, containing Sodium citrate, Sodium chloride, Sodium dodecyl sulfate and preservatives	50°C Moderate shaking	1 h
3) Stringent wash	Stringent wash solution, containing Sodium citrate, Sodium chloride, Sodium dodecyl sulfate, and preservatives	50°C Moderate shaking	2’
5) Incubation with Streptavidine AP-Conjugated	Streptavidine AP-Conjugated diluted in stringent wash solution	50°C Moderate shaking	15’
6) Wash	Rinse solution, containing magnesium chloride, sodium chloride, Trizma and detergents.	Room temperature Moderate shaking	2’
7) Wash	Rinse solution, containing magnesium chloride, sodium chloride, Trizma and detergents.	Room temperature Moderate shaking	2’
8) Colorimetric reaction	NBT/BCIP solution, containing NBT, BCIP, Tris buffer and $MgSO_4$	Room temperature Moderate shaking AT DARK LIGHT	10’
9) Blocking of reaction	Blocking solution, containing Citric acid	Room temperature Moderate shaking	2’
10) Final wash	Distilled water	Room temperature Moderate shaking	2’

Tab. 20 Scheme of protocol for visualization on strip.

4.2.5. Sequencing

NS5b sequencing is still the gold standard method for HCV genotyping.
In order to confirm results of RLB subtyping, NS5b sequencing of genotype 1 samples was performed.
RT-PCR was made with specific primers for NS5b region, previously reported by Sandres-Sauné *et al*. [113]; the same enzyme mix utilized for our HCV genotyping test was used, but with a different thermal profile, that fitted to melting temperature of these primers. Table 21 summarizes thermal profile used for NS5b amplification.

Hold	Temperature	Time	Number of cycles
Reverse transcription	48°C	30 minutes	1
HotStart	95°C	10 minutes	1
Denaturation	95°C	30 seconds	40
Annealing	52°C	60 seconds	
Extension	60°C	60 seconds	
Storage	10°C	∞	1

Tab. 21 Thermal profile of NS5b sequencing.

Amplified products have been purified by using ExoSap-IT® (Affymetrix), according to the manufacturer's protocol. Reaction of sequencing was performed by BMR Genomics, with the BigDyes terminator cycle sequencing kit v3.1 (Applied Biosystems). The sequencing reactions were run on ABI 3730XL (Applied Biosystems) and ABI 3100 (Applied Biosystems) and were analyzed with FinchTV software v1.4 (Geospiza).

4.2.6. Samples

Clinical samples used in this study have been provided by the Molecular Hepatology Laboratory of VIMM (Venetian Institute of Molecular Medicine, Padova). Samples have been selected and afterwards they were made anonymous, randomized, classified and stored with a biobanking system.
All human serum samples were previously analyzed for detection of HCV RNA with a Real Time PCR system by the Molecular Hepatology Laboratory: 212 samples were positive for HCV RNA and 71 were negative (tab. 22).
The 212 HCV positive samples were representative of genotypes 1, 2, 3 and 4; they also reflected different stages of viral infection, characterized by different viral loads.
For the less common genotypes 5, 6 and 7a, commercial synthetic controls were used (tab. 22). The AcroMetrix® HCV Control (Life Tecnologies) has been carefully formulated to mimic naturally occurring human specimens containing HCV RNA. They are prepared by quantitatively diluting HCV positive human serum or plasma of a specific genotype into normal human plasma (NHP). The NHP was previously tested and found to be non-

reactive for antibodies to HIV-1, HIV-2, HTLV-I-II, and HCV, HIV1-RNA, HBsAg, HBV DNA, HCV RNA, WNV RNA and CMV DNA. Each control contains HCV RNA calibrated against the WHO International Standard for HCV RNA, which is genotype 1. The final viral load of these controls is about $2x10^4$ IU/mL.
In this study, we have used AcroMetrix® controls for genotypes 5a and 6a.
For the other rare subtypes of genotype 6 (6c, 6d, 6e, 6l and 6n) and for genotype 7a, plasmids pEX-A2, containing both 5'UTR and core sequences, were provided by Eurofins genomics.
They were quantified and diluted to a final concentration of 10^4 c/µL.

Number	Matrix	HCV
212	Human serum	positive
71	Human serum	negative
2	Synthetic serum added with HCV (Acrometrix® controls)	positive
6	Synthetic DNA controls (Eurofins genomics plasmids)	positive

Tab. 22 List of samples used in the evaluation study

4.2.7. RNA extraction

Total RNA was extracted from both clinical and synthetic human serum, by using the BioRobot EZ1 DSP system, with the EZ1 DSP Virus kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions.
RNA was extracted from 400 µL of human serum and it was eluted in 60 µL.
In order to monitoring the efficiency of extraction, 10 µL of internal control was also added during the elution step of RNA, according to the manufacturer's instructions.

4.3. RESULTS

4.3.1. OPTIMIZATION OF HCV GENOTYPING TEST PROTOTYPE

In this Ph. D. project, we have developed a prototype for a new HCV genotyping test, named AMPLIQUALITY HCV TYPE PLUS, with the 5'UTR and core as target regions. Preliminary experimental tests have been made to set conditions of RT-PCR and visualization on strip. During the optimization phase of the project, these conditions were tested with a subset of 50 samples selected for their positivity to different genotypes, with different viral load. This evaluation phase shows a randomly failure of the test. Therefore, a series of experiments was planned in order to investigated the cause.

Evaluation of enzyme mixture

Different commercial mix were tested to select the best performing enzyme mixture for RT-PCR. Testing experiments were planned in order to achieve the major intra-assay variability: samples with different viral loads and various genotypes were used. Tests were made in parallel, to lower variability of other factors, that can affect result.
We also evaluated the introduction of an enhancer solution (constituted by DMSO) to avoid the typical secondary structures of RNA, and to increase the RT-PCR efficiency. However, the enhancer solution did not led to the expected amplification improvement, but conversely it seemed to have a sort of inhibition effect. (fig. 23)

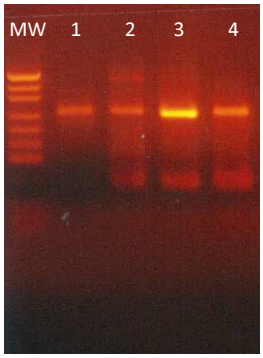


Fig. 23 Imagine of agarose gel visualization of core region simplex PCR, performed on two different RNA samples, without and with the adding of the enhancer. The amplified product has a length of 280 bp. Legend: MW Molecular Weight (501-489, 404, 331, 242, 190, 147, 111-110, 67, 34-34, 26 bp), 1= sample 1, mix without enhancer; 2= sample 1, mix with enhancer; 3= sample 2, mix without enhancer; 4= sample 2, mix with enhancer.

Evaluation of RT-PCR efficiency

We optimized the multiplex amplification reaction step by step, in order to obtain a good amplification of both the target regions: 5'UTR and Core.
Indeed, the two regions showed an impaired amplification efficiency, with the prevalence of 5'UTR on core. The 5'UTR region was always correctly amplified, instead in some samples, the core amplification failed.

One hypothesis was the absence of binding between probes and amplicon of core region, caused by the possible formation of secondary structures. The visualization of PCR products by gel electrophoresis showed that the band of core region (280 bp) was not present. This result demonstrated that there was a problem in the RT-PCR set up.
The relationship between this PCR failure and samples viral load or storage condition (i.e. number of freeze and thaw cycles) were excluded, because of the random way with which it occurs.
Once excluded these factors, we have focused our attention on the primers efficiency.
We decided to perform a simplex PCR reaction with synthetic DNA for genotype 1 and 6 to investigate the primers efficiency. Results showed that primers for 5'UTR worked well and had a good signal on strip, while those for core led to results of less intensity.
Therefore, we decided to test the effect of increasing the PCR annealing temperature, by experiments with gradient tool. The current annealing temperature of 52° was increased to 60°C (fig. 24).

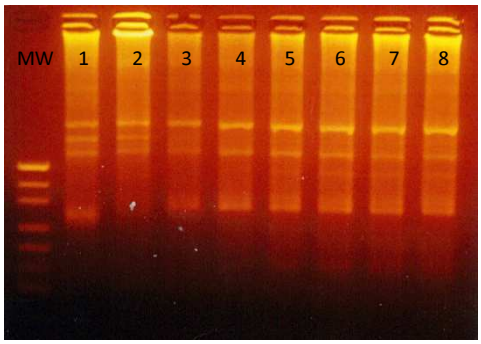


Fig. 134 Imagine of agarose gel visualization of core region simplex PCR, performed on synthetic DNA for genotype 1, with gradient tool. The amplification product has a length of 280 bp. Legend: MW Molecular Weight (501-489, 404, 331, 242, 190, 147, 111-110, 67, 34-34, 26 bp), 1= annealing temperature of 52.0°C, 2= 52.5°C, 3= 53.5°C, 4= 54.9°C, 5= 56.8°C, 6= 58.4°C, 7= 59.4°C, 8= 60.0°C.

As shown in fig. 24, the intensity of core amplification product was less with the annealing temperature of 52° and get better in combination with increasing of temperature.
A similar experiment with gradient tool, showed that also annealing temperature of 5'UTR primers could be increased at 60°C without impairment (data not shown). Therefore, the annealing temperature of 60°C was selected as a good compromise for the efficient amplification of both viral regions.
HCV RNA positive samples were tested with the new annealing condition of reverse-transcription. The reaction was yet not proficient for all the samples tested, and randomly the core amplification has continued to fail.

Primers evaluation and redesign

At the beginning of this study, primers have been newly designed on a dataset containing all the available sequences for HCV genotypes and major subtypes. This bioinformatics study has taken long time and led to a design of a primer set for HCV 5'UTR and core regions. Moreover, primers for amplification of the internal control, an *E. coli* phage, were evaluated.

The entire set is constituted by several primers, because of the high variability of HCV sequences. Specific primers, covering all SNP, were preferred to degenerate primers that can impair RT-PCR efficiency.

Because of the continues random failure of the multiplex reaction, even though efforts made for RT-PCR set up, we decided to modify the primers sequence. For the new design, the same position in the viral genome was more or less maintained, but certain number of nucleotides down or upstream were shifted respect the old localization.

Newly designed primers were tested with the selected amplification conditions (annealing temperature of 60°C) and they led to good results with all samples of the subset.

Starting materials evaluation

During the optimization phase, also the amount of RNA used in RT-PCR was standardized.

Initially we have tried to solve the random failure of amplification by increasing the quantity of RNA, added in RT-PCR, to a final volume of 20 µL. The results were not clear, because a major amount of RNA lead to a good results in some samples, but has an inhibitory effect on others.

At the end the use of 10 µL of RNA extracted from 400 µL of human serum was established as best condition.

Evaluation of genotyping strip

The layout of HCV genotyping strip, obtained at the end of the optimization phase of this project, is shown in fig 25. The first part of the strip is dedicated to 5'UTR region and it contains 16 probes, able to recognize genome sequence of genotypes 1-7a. A detailed description of these probes is shown in table 23.

The bottom part of the strip is reserved to core region, and has three specific probes respectively for HCV 1a and 1b subtypes and for genotype 6. This part of the strip must be considered only in presence of a genotype 1 pattern obtained in the 5'UTR. A detailed description of these probes is shown in table 24.

Every strip also contains 4 control lines:

- Internal control line, that hybridizes IC amplification product and allows the user to monitor the success of extraction and RT-PCR procedure;
- Staining control line, that guarantee the correct execution of post-hybridization steps;

- Universal 5'UTR line, that hybridizes amplification product of this region;
- Universal core line, that hybridizes amplification product of this region and must be evaluated only in presence of a genotype 1 pattern obtained in the 5'UTR region.

Probe Number	Description
1	Identification of HCV 1 genotype
2	Identification of HCV 1 genotype
3	Identification of HCV 1b subtype
4	Identification of HCV 1b subtype
5	Identification of HCV 1a subtype
6	Identification of HCV 2 genotype
7	Identification of HCV 2a/2c subtype
8	Identification of HCV 2b subtype
9	Identification of HCV 3 genotype
10	Identification of HCV 3a subtype
11	Identification of HCV 3h subtype
12	Identification of HCV 4 genotype
13	Identification of HCV 4a subtype
14	Identification of HCV 4r and 4m subtypes
15	Identification of HCV 5 genotype, subtype a
16	Identification of HCV 6 genotype, subtype a and b

Tab. 23 Probes for 5'UTR genotyping and their specificity.

Probe Number	Description
17	This probe is specific for HCV 1a subtype
18	This probe is specific for HCV 1b subtype
19	This probe identifies HCV 6 genotype, all subtypes

Tab. 24 Probes for core genotyping and their specificity.

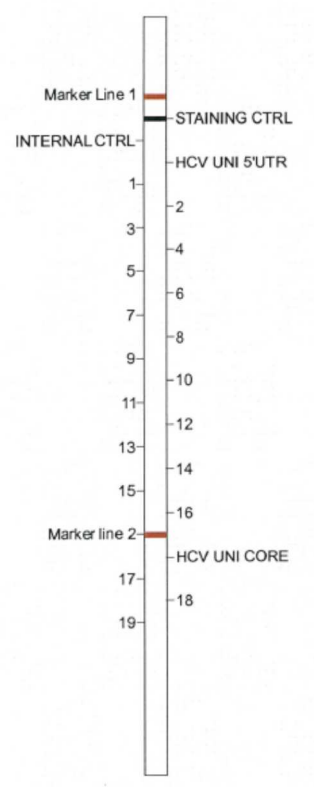


Fig. 25 Scheme of strip layout.

Strip results interpretation

The results of the genotyping test is not immediate to understand, because more probes give a signal for the same genotype. Therefore, every genotype can have more patterns of bands, due to the high sequence variability of HCV genome.

Moreover, even if each probe is specific for a certain genotype or subtype, it could led to a signal also for others, because of cross-reactivity of oligos. Some SNP differences did not affected annealing between probe and amplified products, because they are located at the extremity of the probe.

The bioinformatics analysis led us to create a series of putative hybridization patterns of HCV isolates, determined by this set of probes.

In order to simplify the interpretation procedure, we have decided that the analysis of core region must be considered only in presence of a genotype 1 pattern obtained in the 5'UTR. Probes designed in the 5'UTR, alone are not able to correctly subtype all genotype 1 samples, but there are not problems for subtyping genotype 2, 3, 4 and 5. Moreover, subtypes 6a and 6b can be discriminated in the 5'UTR, because the hybridization pattern is well defined, but subtypes 6c to 6n show the same profile

pattern of genotype 1. In this case, the analysis of 5'UTR alone is not sufficient and the core region information are needed to avoid a misdiagnosis.
This complicated situation makes necessary the use of a tool for strip interpretation. For this reason an interpretation chart were developed, as shown in fig. 26.
For each genotype, several probe patterns were identified, as reported in table 25.

Genotype	Number of patterns (5'UTR)
1	23
2	13
3	8
4	38
5	2
6	5
7a	4

Tab. 25 Number of band patterns reported in the interpretation chart for each genotype.

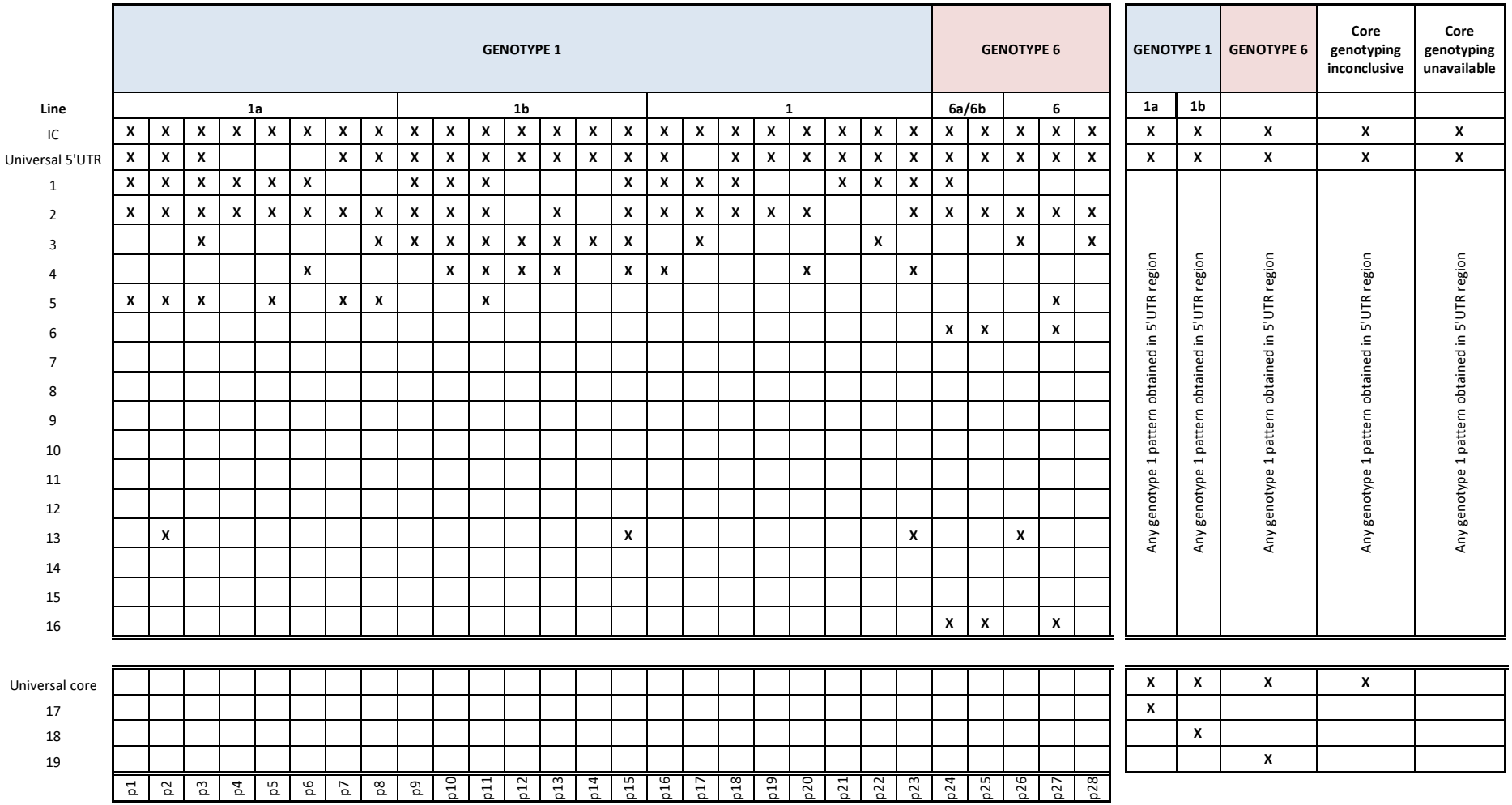


Fig. 26 Part of the interpretation chart, relative to patterns of genotypes 1 and 6. On the left genotype 1 and 6 patterns in the 5'UTR, on the right patterns relative to core region

Evaluation of RLB protocol

A series of tests were performed in order to achieve a triple objective: the standardization of RLB protocol, the presence of a good signal and the absence of background. Different combination of hybridization and washing solutions, by modulating level of stringency, were tried. The final formulation obtained allow results with well defined bands on strip and absence of aspecific signals.

4.3.2. VALIDATION OF HCV GENOTYPING TEST PROTOTYPE

After achieving the prototype standardization of HCV genotyping test, a phase of the project dedicated to the evaluation of the diagnostic performance was started. As requested by the *Common Technical Specifications for in vitro diagnostic medical devices (2009)* for the test validation, the following parameters were investigated: diagnostic specificity and diagnostic sensitivity.

Diagnostic specificity

Diagnostic specificity is the conditional probability that a person not having a disease will be correctly identified by a clinical test. It is calculated from the number of true negative results divided by the total number of those without the disease (which is the sum of the numbers of true negative plus false positive results).
Diagnostic specificity of our assay was assessed by the evaluation of 71 negative samples for HCV RNA. As reference method another CE *in vitro* dispositive, that target the 5'UTR alone, was used.
All the samples were negative with our genotyping test and results were concordant with reference method (fig. 27).
Therefore, our HCV genotyping test shows a diagnostic specificity of 100%.

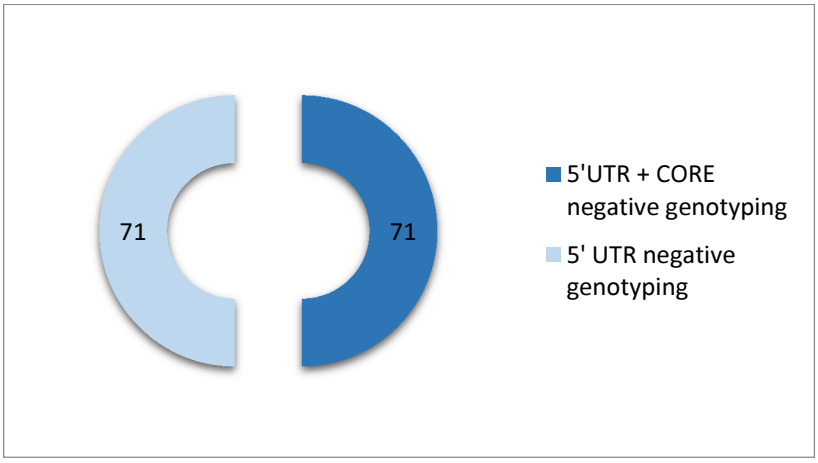


Fig. 27 Comparison of results obtained with AMPLIQUALITY HCV TYPE PLUS (5'UTR and core) and with reference method (5'UTR alone).

Diagnostic sensitivity

Diagnostic sensitivity is the conditional probability that a person having a disease will be correctly identified by a clinical test. It is calculated from the number of true positive results divided by the total number with the disease (which is the sum of the numbers of true positive plus false negative results).

Diagnostic sensitivity of our genotyping test was defined by the evaluation of 220 positive samples, considering both clinical and synthetic ones.

The major part of samples had HCV genotype 1 (n=145, 66%), reflecting the worldwide distribution of viral strains. In Italy, genotype 1 is largely diffused and it is not so difficult recruiting patients with this HCV genotype. Instead, genotypes 2, 3 and 4 are less common and samples with these strains are rare.

Genotype 5, 6 and 7a are the less represented, because of their epidemiological distribution. In Europe genotypes 5 and 6 are extremely rare, they are prevalently diffused in Southern Africa and East Asia. Thus, the availability of these particular samples is very limited.

On the other hand, genotype 7a was officially classified only in January 2014, by the publication of consensus classification of HCV variants by Smith *et al*. [63]; only one genotype 7 infection has been reported and it was isolated in Canada from a Central African immigrant [71].

Therefore, for genotypes 5 and 6a, we have used Acrometrix® synthetic controls (human serum added with HCV, at a final viral load of 2 x 10^4 IU/mL). While, for subtypes 6c, 6d, 6e, 6l, 6n and genotype 7a, we were forced to use synthetic customized DNA control, to check efficiency of primers and probes.

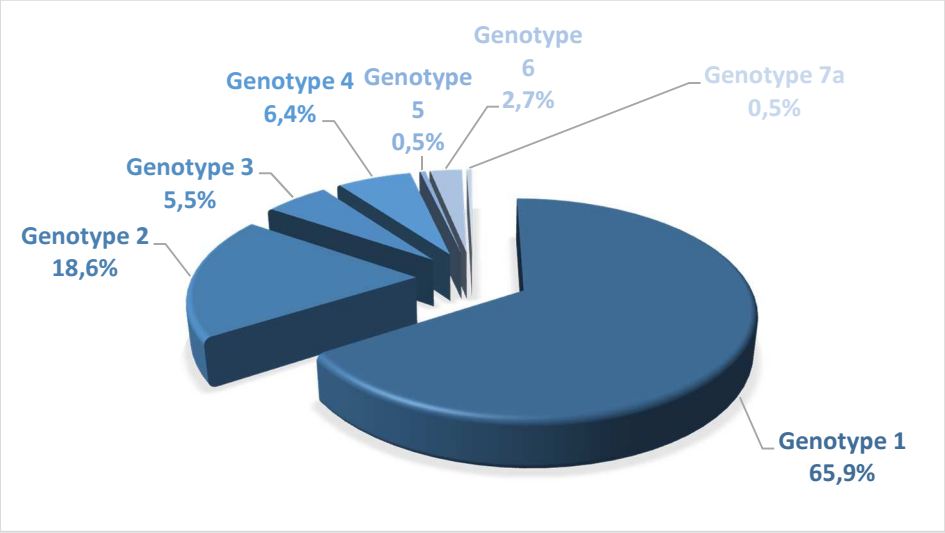


Fig. 28 Distribution of samples in relation to their genotype.

Samples used in this study showed different viral loads, included in a range between 1 x 10^3 and 6 x 10^7 IU/mL. The major part of samples fell into the range > 500.000 IU/mL e < 5.000.000 IU/mL.

We have also 37 samples with an elevated viral load, useful to determine possible impairment of reaction due to high amount of RNA. Instead, we had only 4 samples with a viral load lower of 5000 IU/mL. We had no data about viral loads of 14 samples.

Figure 29 summarizes the distribution of samples in relation to viral load.

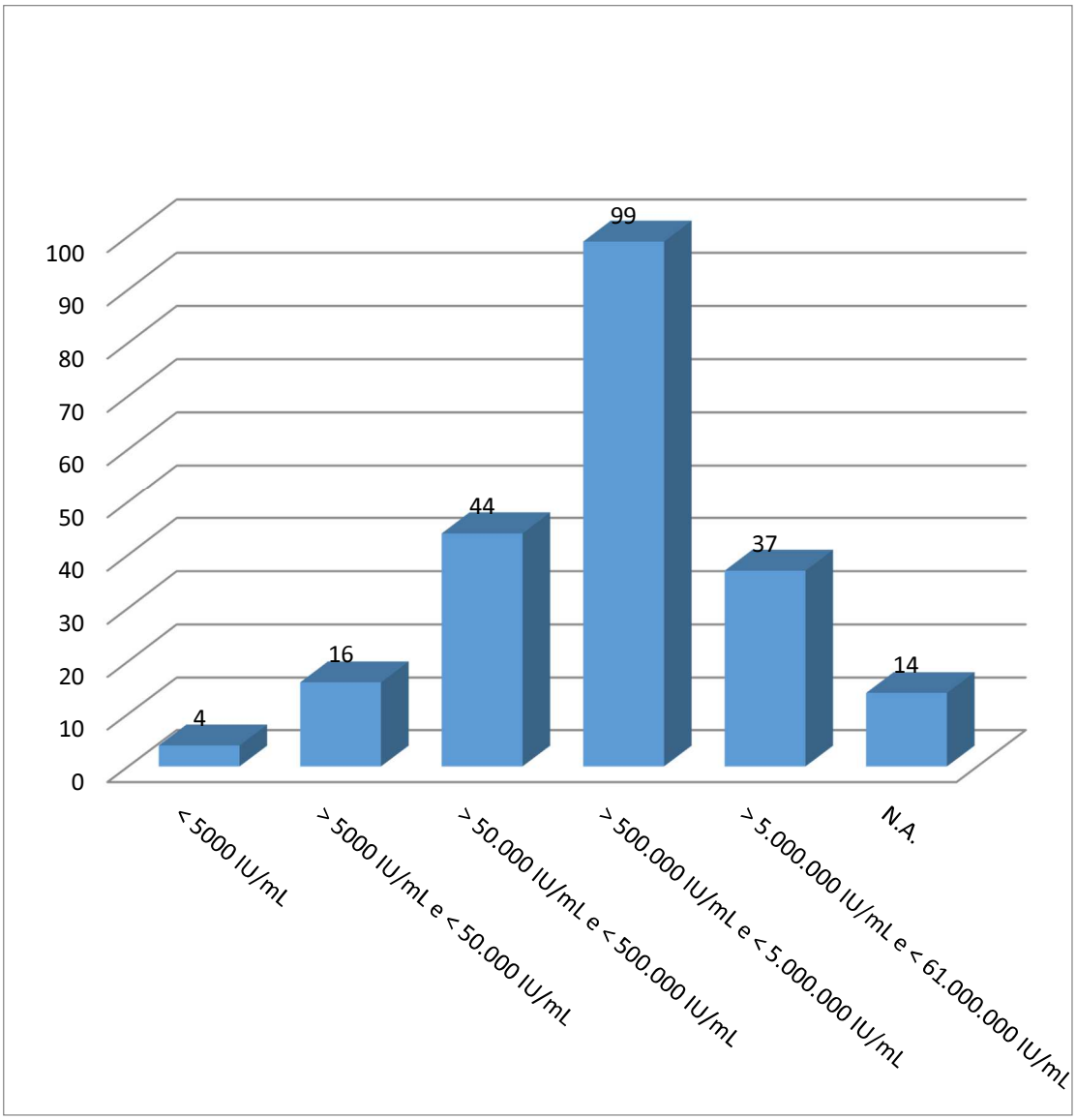


Fig. 29 Distribution of samples in relation to viral load.

First, diagnostic sensitivity was assessed at genotype level. Results obtained with our test were compared to those achieved with the reference method, the assay that target the 5'UTR alone. Discrepant results were confirmed by NS5b sequencing, as second reference method.

Genotype	**5'UTR + core** (AMPLIQUALITY HCV TYPE PLUS)	**Sole 5'UTR** (1st reference method)	**NS5b sequencing** (2nd reference method)
1	143	148	143
2	41	41	n.a.
3	12	12	n.a.
4	14	14	n.a.
5	1	1	n.a.
6	6	1	n.a.
7	1	1	n.a.

Tab. 26 Comparison of genotyping results obtained with different methods. Total of genotyped samples is 218, because of 2 samples RT-PCR failure.

The concordance between the two genotyping methods was 97.7%. Samples with genotype 2, 3, 4, 5 and 7 were identified in the same way, while for genotype 1 and 6 there were some discrepancy.

Sequencing of NS5b region of samples with discrepant results had confirmed the AMPLIQUALITY HCV TYPE PLUS genotyping. This allow to calculate for our assay a diagnostic sensitivity, at genotype level, of 100%.

Furthermore these data demonstrate that the 5'UTR alone is not able to discriminate between genotype 1 and 6, subtypes c-n; the resulting pattern of band on strip is the same of a genotype 1, leading to a misdiagnosis. Combining the analysis of 5'UTR and core regions, a discrimination between these genotypes is allowed, as shown in fig. 30.

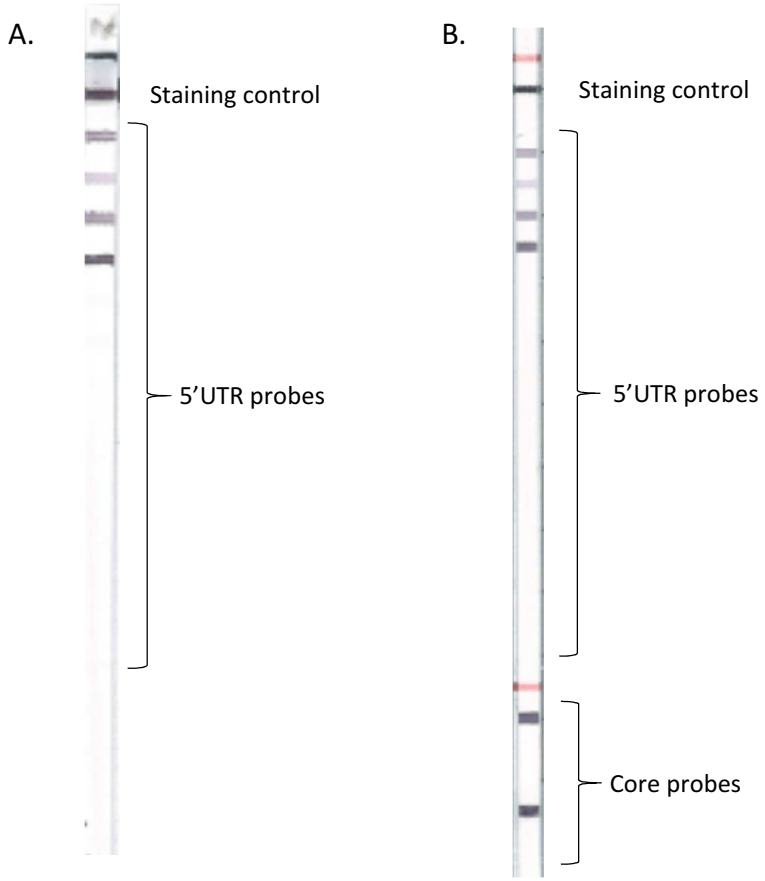


Fig. 30 Imagine of strip result for a genotype 6 sample.
A. Strip relative to the analysis of sole 5'UTR: band pattern identifies a 1b isolate, leading to a misdiagnosis.
B. Strip relative to combined analysis 5'UTR and core: band pattern of core region identifies a genotype 6 isolate. (see interpretation chart page 76)

Then we have calculated diagnostic sensitivity at subtype level, in particular for genotype 1.

Results obtained with our test were compared to those achieved both with the genotyping of the sole 5'UTR and NS5b sequencing.

Regard samples infected by subtype HCV 1a, results obtained with our assay are similar to those of NS5b sequencing. Two samples were not genotyped, because of failure of RT-PCR.

The analysis of 5'UTR alone led to a correct subtyping only for 20 samples (44.4%); other 19 samples (42.2%) were not subtyped and defined only as genotype 1. Four samples (8.9%) were wrongly identified as subtype 1b (tab. 27).

Regard samples infected by subtype HCV 1b, results obtained with our assay are similar to those of NS5b sequencing. One sample was not genotyped, because of failure of RT-PCR.

The analysis of 5'UTR alone led to a correct subtyping of 92 samples (92%); other 3 samples (3%) were not subtyped and defined only as genotype 1. Instead, 4 samples (4%) were wrongly identified as subtype 1a (tab. 28).

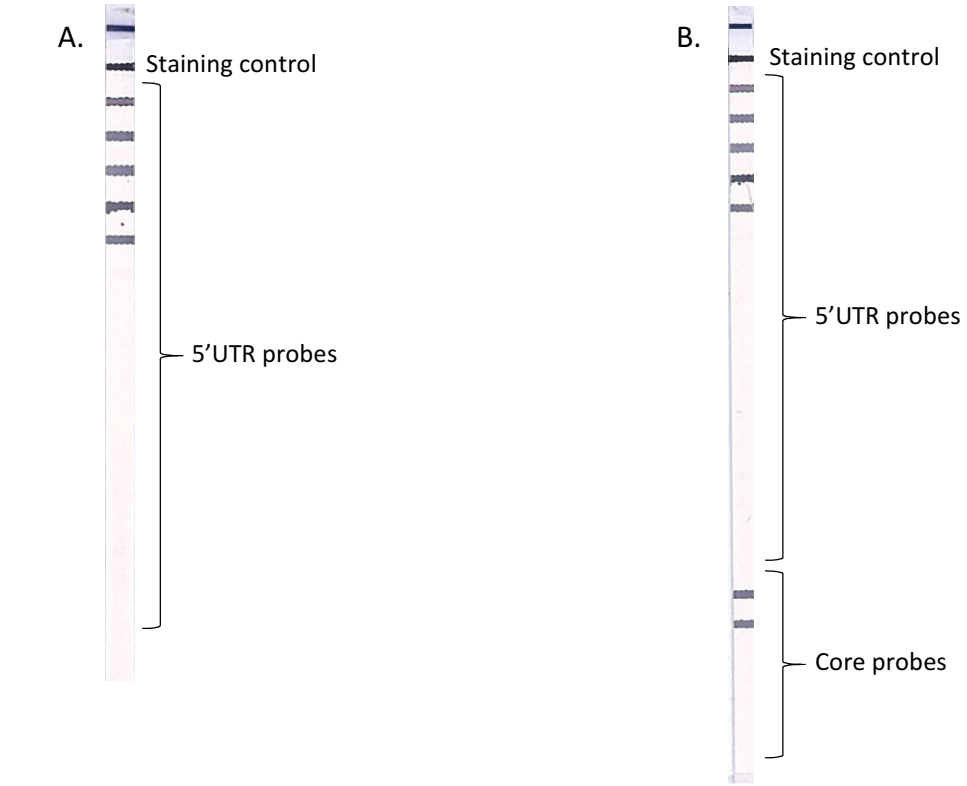
Combining the analysis of 5'UTR and core regions, a discrimination between these subtypes is allowed, as shown in fig. 31.
In conclusion, our HCV genotyping test shows a diagnostic sensitivity, at subtype level, of 100%.

Genotype	**5'UTR + core** (AMPLIQUALITY HCV TYPE PLUS)	**Sole 5'UTR** (1st reference method)	**NS5b sequencing** (2nd reference method)
1a	43	20	45
1b	0	4	0
1	0	19	0
No genotyping	2	2	0
Tot	45	45	45

Tab. 27 Comparison of subtyping results obtained with AMPLIQUALITY HCV TYPE PLUS assay and the two reference methods.

Genotype	**5'UTR + core** (AMPLIQUALITY HCV TYPE PLUS)	**Sole 5'UTR** (1st reference method)	**NS5b sequencing** (2nd reference method)
1b	99	92	100
1a	0	4	0
1	0	3	0
No genotyping	1	1	0
Tot	100	100	100

Tab. 28 Comparison of subtyping results obtained with AMPLIQUALITY HCV TYPE PLUS assay and the two reference methods.



D.

>gb|FJ024281.1| Hepatitis C virus subtype 1a isolate HCV-1a/US/BID-V1718/2007, complete genome
Length=9301

Score = 501 bits (271), Expect = 1e-138
Identities = 302/317 (95%), Gaps = 2/317 (0%)
Strand=Plus/Plus

Fig. 31 Comparison of results obtained with the 3 genotyping methods for a misdiagnosed sample.
A. Strip relative to the analysis of sole 5'UTR: band pattern identifies a 1b isolate, leading to a misdiagnosis.
B. Strip relative to combined analysis 5'UTR and core: band pattern of core region identifies a subtype 1a isolate. (see interpretation chart page 76)
C. Electropherogram relative to NS5b sequence of the sample.
D. BLAST result relative to NS5b sequence: sample is confirmed to be subtype 1a.

5. DISCUSSION AND CONCLUSIONS

Reverse Line Blot is rapidly becoming a standard molecular tool for diagnostic and epidemiological studies in an increasing number of laboratories all over the world, thanks to its high grade of usability. RLB is a simple technique that can be used in diagnostic routine also by small laboratories, because the request instruments are few and not expensive.
The technique can be utilized in a number of ways. Multiple probes can be designed to detect sequence variation within a single amplified product, or multiple targets can be amplified simultaneously, with one (or more) probes used for subsequent detection. A combination of both approaches can also be used within a single assay. The ability to include multiple probes for a single target sequence makes the assay highly specific. The flexibility of the method allows its use for a wide range of applications.
The aim of our project was the development of RLB tests and the evaluation of their application in the management of diseases, requiring a genetic characterization.

In the first part of the study we explore the potentiality of RLB in detecting known genetic mutations in the human genome and our interest was focused on atypical CF. This is a rare genetic disease, not well genetically characterized and probably related to the presence of mutation in the ENaC genes, as discussed in recent epidemiological studies [58, 59]. Until today, only sequencing approach was used for investigate the presence of specific ENaC mutations, and this limiting the diffusion of this type of investigation.
We developed GENEQUALITY ENaC-TYPE, which is a device based on RLB, able to detect 9 ENaC variants, previously identified by sequencing works of Sheridan and Azad.
Once defined and analytically validate the prototype, the assay has been evaluated on a series of patients with atypical CF, in collaboration with Laboratory of Molecular Pathology, Department of Pathology and Diagnostics, University Hospital of Verona, a reference center for cystic fibrosis.
Two groups of patients were selected on the basis of their different CF phenotypes: lungs disease vs. pancreatic manifestation and the results were separately analyzed. Until now, studies about correlation of ENaC mutations with atypical cystic fibrosis have been performed only on atypical CF patients with respiratory symptoms and not on individuals with different CF phenotypes.
In the group of lungs disease phenotypes, four significative variants were detected (p.E539K-SCNN1B, p.1670-2A>G-SCNN1B, p.P267L-SCNN1B and p.G294S-SCNN1B). Mutations were found at a more than five-fold significantly increased in patients (5.5%) respect to the controls (0%).
Patients that carried these variants were heterozygous for CFTR mutations. Thus, about 50% of the CFTR network is still functional. However, the remainder 50% may be

rendered nonfunctional by a mutation in at least one other member of the network, such as ENaC.

In the group of pancreatic disease phenotypes, none ENaC mutation showed a frequency significantly higher in patients rather than in controls.

Despite CFTR, which expression and activity in pancreatic cells have been well established, ENaC role in pancreas is still an argument of controversial opinions. Findings of several studies are in contrast about presence of functional ENaC in pancreatic cells and the possible physiologic function of ENaC, therefore, remains obscure. Thus, it is impossible to define if our results are due to lack of ENaC involvement in the pathogenesis of CF-like pancreatic disorders or to absence of functional channels in pancreatic cells.

RLB method is limited to the detection of specific ENaC mutations; functional analysis of each variants has not been performed. Thus, it is impossible to develop an explanation of the effect of the different mutations, especially in the light that either too low or too high amounts of ENaC can already result in two different diseases, which will be even further complicated if the CFTR protein is implicated.

Despite classic CF in which a large part of CFTR causative mutations has been defined, role of ENaC variants in atypical cystic fibrosis is still to be defined. Studies about prevalence of ENaC mutations in patients with CF-like disorders are very few and the causative relationship between ENaC and disorder is only supposed. Phenotypes of atypical CF are very wide, suggesting an implication of additional genes/proteins and environmental factors. Indeed, in a considerable number of these patients no involvement of ENaC was found.

Our test has demonstrated to be easily applied to a diagnostic routine, for his usability and simple results interpretation but it allows detection of only 9 specific variants. In a context of undefined correlation, this test is too limited. Further studies of sequencing of all exons and exon/intron junctions have to be performed on patients with different CF-like disorder, to collect more data. Then, functional studies will clarify effect of identified variants on phenotype.

Therefore, this test at the moment could not be used as a confirmatory assay for diagnosis and management of therapy. Probably our assay could find a collocation in the epidemiological field, to evaluate prevalence of the 9 variant in particular group of individuals. Moreover, we think that this study has given preliminary results that could be achieved for a future work based on a more advanced technology, such as Next Generation Sequencing. For example, we have seen that study could be performed on patients with CF-like lungs disorder and not on individuals with pancreatic pathology. Moreover, it could be interesting to evaluate prevalence of ENaC variants in patients with CBAVD, one of the major related CF pathologies.

We have defined that RLB is not the ideal approach for studying a genetic disease with so many obscure points. On the other hand, this technique is appropriate for diagnosis of disorder in which causative mutations have been well defined, e. g. classic CF. In this

case RLB is a very useful method that allow contemporary detection of principal mutations in each individual. Each strip can contain more diffuse mutations and regional variants. This could be a test of first analysis, to screen a part of patients.

In the second part of the study, we explore the potentiality of RLB in the identification of the genotypic strand of Hepatitis C virus. Genotypes differ at 31-34 % of nucleotide positions of complete genomic sequences with approximately 30% amino acid sequence divergence.

HCV genotype is one of the factor influencing the disease progression and the major predictor of a SVR to double therapy peg-IFN/ribavirin. SVR vary according to the HCV genotype involved. HCV genotype 1 and 4 were found to be the prevalent resistant genotypes. Moreover, new DAAs have shown different efficacy on subtypes 1a and 1b, due to different resistance profile of these subtypes. Thus, HCV genotyping is essential to apply the most appropriate therapeutic regimen.

The gold standard for HCV genotyping is NS5b sequencing and subsequent phylogenetic analysis. Generally, this type of assay is used by reference centers, which design home-made protocols, based on their specific required. Therefore, other type of assay, based on Real-Time PCR or RLB technologies, are commercially available. The major part of them target the viral 5'UTR, because it is the most conserved one. This region is able to well differentiate the virus at the genotype level, except for genotype 1 and 6 (subtypes c to l). Moreover, it cannot lead to the subtyping genotype 1, because the HCV subtype 1a segregates in two distinct clades, which are termed 1a clade I and 1a clade II. The 5'UTR sequence of 1a clade I is identical to that of subtype 1b. Recent study shows that using the sole 5'UTR as target for genotyping, there is a mistyping approximately of 25% and 10% of HCV subtype 1a and 1b strains, respectively [111].

Therefore, it is evident that the analysis of the sole 5'UTR is not sufficient to allow a good determination of HCV genotype/subtype and that evaluation of another region is necessary.

The AMPLIQUALITY HCV TYPE PLUS device designed and developed during the second part of this project is an RLB assay which can well differentiate genotype 1 to 6 and subtypes 1a and 1b, thanks to the presence of core as target region in addition to 5'UTR. We selected the core gene because it contains both conserved and variable regions and this condition is optimal for designing primers and probe, respectively.

The optimization of the assay, unlike of the experience of GENEQUALITY ENaC-TYPE development, has taken long time. The most critical issue was the randomly failure of HCV RNA identification in several samples. Step by step, we have analyzed all the possible reasons and focused our attention on RT-PCR efficiency, probably limited by the first set of primers designed.

A more accurate selection of the primers have led to better results, suitable for diagnostic purpose. For performance evaluation of our HCV genotyping assay a set of 220 samples were used.

Our assay showed a diagnostic specificity of 100% and a diagnostic sensitivity of 98.6% both at genotype and subtype levels.

Results were compared with two reference methods: a CE-IVD assay that target the sole 5'UTR and NS5b sequencing. As confirmed by NS5b sequencing, our test using a combined analysis of 5'UTR and core regions was able to correctly genotype all the samples. Conversely, the analysis of the sole 5'UTR has showed mistyping of genotype 6 samples. As reported in literature, this region is not able alone to recognize genotype 6, subtypes c to n, and to differentiate them from genotype 1 [114].

This kind of mistyping could affect the management of therapeutic regimen. Indeed, SVR rates for genotypes 5 and 6 with PEG-based therapy are minor than SVR for genotypes 2 and 3, but major than SVR for genotype 1. Dose of drugs is the same for genotypes 1 and 6, but duration is lesser for genotype 6. On the other hand, clinical studies about administration of new DAAs to patients infected with genotype 6 are very few. Knowledge about efficacy, adverse effects, resistant profiles of strains are very limited. The major part of available DAAs shows a great efficacy against only genotype 1 strains. Among new DAAs, only sofosbuvir (Gilead) has been tested on patients infected by genotype 6 HCV (NEUTRINO clinical study) and can be used in the clinical practice.

Therefore, misdiagnosis of a genotype 6 HCV in a genotype 1 could led to administration of a not adequate therapy, with risk for patient health.

At subtype level, NS5b sequencing confirmed that our assay had correctly assigned the subtype to all samples infected with genotype 1. Instead, the analysis of the sole 5'UTR showed discrepant results, mainly in the subtyping of 1a strains, confirming the findings reported in previous studies [111]. The 5'UTR was not able to subtype almost half of samples 1a and about 10% was misdiagnosed as 1b. Less critical was the mistyping in the 1b samples group, in which the discrepancy was 4%.

Until the introduction of new DAAs in the clinical practice, HCV subtyping was not mandatory to define correct therapeutic approach. However, antivirals molecules, such telaprevir and boceprevir, have shown variable efficacies on subtypes 1a and 1b. Probably the genetic barrier and resistance profiles substantially differ between the various genotype 1 subtypes. For this reason, the correct identification of HCV subtypes 1a and 1b become crucial in clinical trials assessing new HCV drugs in order to stratify and interpret efficacy and resistance data. Misdiagnosis between subtypes 1a and 1b could led to application of a not appropriate therapy, causing failure of SVR and selection of resistant strains.

The good subtyping results obtained by our test, confirm that RLB assay is an ideal tool for HCV genotyping. It is a simple and fast method and exhibits a high sensitivity and specificity. The same detailed genotyping information could be achieved with Real-Time technologies, but is necessary performing multiple amplification reactions for each sample. Instead, RLB allows the acquisition of a lot of sequence information in a single run. RLB is also the ideal technology for identification of co-infections by different HCV strains. NS5b sequencing can give a definite result for genotype but evaluation of more

HCV isolates presence in the same sample is very difficult. On the other hand, a critical issue of our test could be the results interpretation that required the careful use of an interpretation table, in which more than 90 different patterns are reported. For this reason, our study probably will continue with the implementation of an interpretative software tool.

In conclusion, this work has demonstrated that RLB is a robust method for the management of several diseases, requiring a genetic characterization. However, genetic knowledge about target disorders must to be well defined.

6. REFERENCES

[1] Netto G.J., Saad R.D., Dysert II P.A.
Diagnostic molecular pathology: current techniques and clinical applications.
BUMC Proceedings 2003;16:379–383

[2] Ratjen F., Döring G.
Cystic fibrosis
Lancet 2003;361: 681–89

[3] Cutting G.R.
Cystic fibrosis genetics: from molecular understanding to clinical application
Nat Rev Genet. 2015;16(1):45-56.

[4] Riordan J.R., Rommens J.M., Kerem B.S., *et al.*
Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.
Science 1989;245: 1066–73.

[5] Tsui L., and Dorfman R.
The cystic fibrosis gene: a molecular genetic perspective.
Cold Spring Harb Perspect Med. 2013;3(2):a009472

[6] Higgins C.F.
Cystic fibrosis transmembrane conductance regulator (CFTR).
Br Med Bull. 1992;48(4):754-65.

[7] Kartner N., Hanrahan J.W., Jensen T.J., Naismith A.L., Sun S.Z., Ackerley C.A., Reyes E.F., Tsui L.C., Rommens J.M., Bear C.E., *et al*.
Expression of the cystic fibrosis gene in non-epithelial invertebrate cells produces a regulated anion conductance.
Cell. 1991;64(4):681-91.

[8] World Health Organization (WHO)
The molecular genetic epidemiology of cystic fibrosis. Report of a joint meeting of WHO/ECFTN/ICF(M)A/ECFS.
WHO Geneva, Human Genetics Programme WHO/HGN/CF/WG/04.02.

[9] Castellani C., Cuppens H., Macek Jr.M., Cassiman J.J., Kerem E., Durie P., Tullis E., Assael B.M., Bombieri C., Brown A., Casals T., Claustres M., Cutting G.R., *et al*.
Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice
Journal of Cystic Fibrosis 2008;7:179–196

[10] Welsh M.J., Smith A.E.
Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis.
Cell 1993;73(7):1251–1254.

[11] McKone E.F., Emerson S.S., Edwards K.L., Aitken M.L.
Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study.
Lancet. 2003;361(9370):1671-6.

[12] Rosenstein B.J., Cutting G.R.
The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel.
J Pediatr. 1998;132(4):589-95.

[13] De Boeck K., Wilschanski M., Castellani C., Taylor C., Cuppens H., Dodge J., Sinaasappel M.
Cystic fibrosis: terminology and diagnostic algorithms.
Thorax. 2006;61(7):627-35.

[14] Bonizzato A., Castellani C.
La diagnostica molecolare della fibrosi cistica e delle malattie correlate a mutazioni del gene CFTR
Biochimica clinica, 2008, vol. 32, n. 2.

[15] Bombieri C., Claustres M., De Boeck K., Derichs N., Dodge J., Girodon E., Sermet I.,cSchwarz M., Tzetis M., Wilschanski M., Bareil C., Bilton D., Castellani C., Cuppens H., Cutting G.R., Drevinek P., *et al*.
Recommendations for the classification of diseases as CFTR-related Disorders
J Cyst Fibros. 2011;10 Suppl 2:S86-102.

[16] Pasteur M.C., Helliwell S.M., Houghton S.J., Webb S.C., Foweraker J.E., Coulden R.A., Flower C.D., Bilton D., Keogan M.T.
An investigation into causative factors in patients with bronchiectasis.
Am J Respir Crit Care Med 2000;162:1277–84.

[17] Bombieri C., Benetazzo M., Saccomani A., Belpinati F., Gilè L.S., Luisetti M., Pignatti P.F.
Complete mutational screening of the CFTR gene in 120 patients with pulmonary disease.
Hum Genet 1998;103:718–22.

[18] Casals T., De-Garcia J., Gallego M., Dorca J., Rodríguez-Sanchón B., Ramos M.D., Giménez J., Cisteró-Bahima A., Olveira C., Estivill X.
Bronchiectasis in adult patients: an expression of heterozyosity for CFTR gene mutations?
Clin Genet. 2004;65(6):490-5.

[19] Ziedalski T., Kao P., Heing N.R., Jacobs S.S., Ruoss S.J.
Prospective analysis of cystic fibrosis transmembrane regulator mutations in adults with bronchiectasis or pulmonary nontuberculous mycobacterial infections.
Chest 2006;130(4):995‑1002.

[20] Braganza J.M., Lee S.H., McCloy R.F., McMahon M.J.
Chronic pancreatitis.
Lancet. 2011;377(9772):1184-97.

[21] Grigorescu M., Grigorescu M.D.
Genetic factors in pancreatitis.
Rom J Gastroenterol. 2005;14(1):53-61.

[22] Cohn J.A., Friedman K.J., Noone P.G., Knowles M.R., Silverman L.M., Jowell P.S.
Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis.
N Engl J Med 1998;339(10):653‑8.

[23] Park H.W., Nam J.H., Kim J.Y., Namkung W., Yoon J.S., Lee J.S., Kim K.S., Venglovecz V., Gray M.A., Kim K.H., Lee M.G.
Dynamic regulation of CFTR bicarbonate permeability by [Cl‑]i and its role in pancreatic bicarbonate secretion.
Gastroenterology. 2010;139(2):620-31.

[24] Chen J.M., Ferec C.
Chronic pancreatitis: genetics and pathogenesis.
Annu Rev Genomics Hum Genet. 2009;10:63-87.

[25] Segal I., Yaakov Y., Adler S.N., Blau H., Broide E., Santo M., Yahav Y., Klar A., Lerner A., Aviram M., Ellis I., Mountford R., Shteyer E., Kerem E., Wilschanski M.
Cystic fibrosis transmembrane conductance regulator ion channel function testing in recurrent acute pancreatitis.
J Clin Gastroenterol. 2008;42(7):810-4.

[26] Mak V., Jarvi K.A.
The genetics of male infertility.
J Urol. 1996;156(4):1245-56; discussion 1256-7.

[27] Cuppens H., Cassiman J.J.
CFTR mutations and polymorphisms in male infertility.
Int J Androl. 2004;27(5):251-6.

[28] Claustres M., Guittard C., Bozon D., Chevalier F., Verlingue C., Ferec C., Girodon E., Cazeneuve C., Bienvenu T., Lalau G., Dumur V., Feldmann D., Bieth E., Blayau M., Clavel C., *et al*.
Spectrum of CFTR mutations in cystic fibrosis and in congenital absence of the vas deferens in France.
Hum Mutat. 2000;16(2):143-56.

[29] Ratbi I., Legendre M., Niel F., Martin J., Soufir J.C., Izard V., Costes B., Costa C., Goossens M., Girodon E.
Detection of cystic fibrosis transmembrane conductance regulator (CFTR) gene rearrangements enriches the mutation spectrum in congenital bilateral absence of the vas deferens and impacts on genetic counselling.
Hum Reprod. 2007;22(5):1285-91.

[30] Groman J.D., Karczeski B., Sheridan M., Robinson T.E., Fallin M.D., Cutting G.R.
Phenotypic and genetic characterization of patients with features of "nonclassic" forms of cystic fibrosis.
J Pediatr. 2005;146(5):675-80.

[31] Mekus F., Ballmann M., Bronsveld I., Dörk T., Bijman J., Tümmler B., Veeze H.J.
Cystic-fibrosis-like disease unrelated to the cystic fibrosis transmembrane conductance regulator.
Hum Genet. 1998;102(5):582-6.

[32] Groman J.D., Meyer M.E., Wilmott R.W., Zeitlin P.L., Cutting G.R.
Variant cystic fibrosis phenotypes in the absence of CFTR mutations.
N Engl J Med. 2002;347(6):401-7.

[33] Weiler C.A., Drumm M.L.
Genetic influences on cystic fibrosis lung disease severity.
Front. Pharmacol. 2013;4:40.

[34] Stutts M.J., Canessa C.M., Olsen J.C., Hamrick M., Cohn J.A., Rossier B.C., Boucher R.C.
CFTR as a cAMP-dependent regulator of sodium channels.
Science. 1995;269(5225):847-50.

[35] Borowitz D., Durie P.R., Clarke L.L., Werlin S.L., Taylor C.J., Semler J., De Lisle R.C., Lewindon P., Lichtman S.M., Sinaasappel M., Baker R.D., Baker S.S., Verkade H.J., Lowe M.E., Stallings V.A., Janghorbani M., Butler R., Heubi J.
Gastrointestinal outcomes and confounders in cystic fibrosis.
J Pediatr Gastroenterol Nutr. 2005;41(3):273-85.

[36] Canessa C.M., Merillat A.M., Rossier B.C.
Membrane topology of the epithelial sodium channel in intact cells.
Am J Physiol. 1994;267(6 Pt 1):C1682-90.

[37] Garty H., Palmer L.G.
Epithelial sodium channels: function, structure, and regulation.
Physiol Rev. 1997;77(2):359-96.

[38] Kellenberger S., Schild L.
Epithelial sodium channel/degenerin family of ion channels: a variety of functions for a shared structure.
Physiol Rev. 2002;82(3):735-67.

[39] Ling B.N., Zuckerman J.B., Lin C., Harte B.J., McNulty K.A., Smith P.R., Gomez L.M., Worrell R.T., Eaton D.C., Kleyman T.R.
Expression of the cystic fibrosis phenotype in a renal amphibian epithelial cell line.
J Biol Chem. 1997;272(1):594-600.

[40] Letz B., Korbmacher C.
cAMP stimulates CFTR-like Cl- channels and inhibits amiloride-sensitive Na+ channels in mouse CCD cells.
Am J Physiol. 1997;272(2 Pt 1):C657-66.

[41] Bachhuber T., König J., Voelcker T., Mürle B., Schreiber R., Kunzelmann K.
Cl- interference with the epithelial Na+ channel ENaC.
Biol Chem. 2005;280(36):31587-94.

[42] Naren A.P., Cormet-Boyaka E., Fu J., Villain M., Blalock J.E., Quick M.W., Kirk K.L.
CFTR chloride channel regulation by an interdomain interaction.
Science. 1999;286(5439):544-8.

[43] Cormet-Boyaka E., Di A., Chang S.Y., Naren A.P., Tousson A., Nelson D.J., Kirk K.L.
CFTR chloride channels are regulated by a SNAP-23/syntaxin 1A complex.
Proc Natl Acad Sci U S A. 2002;99(19):12477-82.

[44] Hallows K.R., Raghuram V., Kemp B.E., Witters L.A., Foskett J.K.
Inhibition of cystic fibrosis transmembrane conductance regulator by novel interaction with the metabolic sensor AMP-activated protein kinase.
J Clin Invest. 2000;105(12):1711-21.

[45] Berdiev B.K., Qadri Y.J., Benos D.J.
Assessment of the CFTR and ENaC association.
Mol Biosyst. 2009;5(2):123-7.

[46] Boucher R.C.
Cystic fibrosis: a disease of vulnerability to airway surface dehydration.
Trends Mol Med 2007;13:231–240.

[47] Knowles M.R., Boucher R.C.
Mucus clearance as a primary innate defense mechanism for mammalian airways.
J Clin Invest 2002;109:571–577.

[48] Mall M., Grubb B.R., Harkema J.R., O'Neal W.K., Boucher R.C.
Increased airway epithelial Na+ absorption produces cystic fibrosis-like lung disease in mice.
Nat Med. 2004;10(5):487-93.

[49] Johannesson B., Hirtz S., Schatterny J., Schultz C., Mall M.A.
CFTR regulates early pathogenesis of chronic obstructive lung disease in βENaC-overexpressing mice.
PLoS One. 2012;7(8):e44059.

[50] Hobbs C.A., Da Tan C., Tarran R.
Does epithelial sodium channel hyperactivity contribute to cystic fibrosis lung disease?
J Physiol. 2013;591(Pt 18):4377-87.

[51] Shimkets R.A., Warnock D.G., Bositis C.M., Nelson-Williams C., Hansson J.H., Schambelan M., Gill J.R. Jr, Ulick S., Milora R.V., Findling J.W., *et al.*
Liddle's syndrome: heritable human hypertension caused by mutations in the beta subunit of the epithelial sodium channel.
Cell. 1994;79(3):407-14.

[52] Firsov D., Schild L., Gautschi I., Mérillat A.M., Schneeberger E., Rossier B.C.
Cell surface expression of the epithelial Na channel and a mutant causing Liddle syndrome: a quantitative approach.
Proc Natl Acad Sci U S A. 1996;93(26):15370-5.

[53] Snyder P.M., Price M.P., Mcdonald F.J., Adams C.M., Volk K.A., Zeiher B.G., Stokes J.B., Welsh M.J.
Mechanism by which Liddle's syndrome mutations increase activity of a human epithelial Na+ channel.
Cell 1995;83:969–978.

[54] Chang S.S., Grunder S., Hanukoglu A., Roosler A., Mathew P.M., Hanukoglu I., Schild L., Lu Y., Shimkets R.A., Nelson-Williams C., Rossier B.C., Lifton R.P.
Mutations in subunits of the epithelial sodium channel cause salt wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1.
Nat Genet 1996;12:248–253.

[55] Kerem E., Bistritzer T., Hanukoglu A., Hofmann T., Zhou Z., Bennett W., MacLaughlin E., Barker P., Nash M., Quittell L., Boucher R., Knowles M.R.
Pulmonary epithelial sodium-channel dysfunction and excess airway liquid in pseudohypoaldosteronism.
N Engl J Med. 1999;341(3):156-62.

[56] Marthinsen L., Kornfält R., Aili M., Andersson D., Westgren U., Schaedel C.
Recurrent Pseudomonas bronchopneumonia and other symptoms as in cystic fibrosis in a child with type I pseudohypoaldosteronism.
Acta Paediatr. 1998;87(4):472-4.

[57] Schaedel C., Marthinsen L., Kristoffersson A.C., Kornfält R., Nilsson K.O., Orlenius B., Holmberg L.
Lung symptoms in pseudohypoaldosteronism type 1 are associated with deficiency of the alpha-subunit of the epithelial sodium channel.
J Pediatr. 1999;135(6):739-45.

[58] Sheridan M.B., Fong P., Groman J.D., Conrad C., Flume P., Diaz R., Harris C., Knowles M., Cutting G.R.
Mutations in the beta-subunit of the epithelial Na+ channel in patients with a cystic fibrosis-like syndrome.
Hum Mol Genet. 2005;14(22):3493-8.

[59] Azad A.K., Rauh R., Vermeulen F., Jaspers M., Korbmacher J., Boissier B., Bassinet L., Fichou Y., des Georges M., Stanke F., De Boeck K., Dupont L., Balascáková M., Hjelte L., Lebecque P., Radojkovic D., Castellani C., Schwartz M., Stuhrmann M., Schwarz M., Skalicka V., de Monestrol I., Girodon E., Férec C., Claustres M., Tümmler B., Cassiman J.J., Korbmacher C., Cuppens H.
Mutations in the amiloride-sensitive epithelial sodium channel in patients with cystic fibrosis-like disease.
Hum Mutat. 2009;30(7):1093-103.

[60] Choo Q.L., Weiner A.J., Overby L.R., Kuo G., Houghton M., Bradley D.W.
Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis.
Br Med Bull. 1990;46(2):423-41.

[61] Kuo G., Choo Q.L., Alter H.J., Gitnick G.L., Redeker A.G., Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E., *et al.*
An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis.
Science. 1989;244(4902):362-4.

[62] Simmonds P.
Genetic diversity and evolution of hepatitis C virus: 15 years on.
J Gen Virol. 2004;85(Pt 11):3173-88.

[63] Smith D.B., Bukh J., Kuiken C., Muerhoff A.S., Rice C.M., Stapleton J.T., Simmonds P.
Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource.
Hepatology. 2014;59(1):318-27.

[64] Smith D.B., Pathirana S., Davidson F., Lawlor E., Power J., Yap P.L., Simmonds P.
The origin of hepatitis C virus genotypes.
J Gen Virol. 1997;78 (Pt 2):321-8.

[65] Pybus O.G., Cochrane A., Holmes E.C., Simmonds P.
The hepatitis C virus epidemic among injecting drug users.
Infect Genet Evol. 2005;5(2):131-9.

[66] Magiorkinis G., Magiorkinis E., Paraskevis D., Ho S.Y., Shapiro B., Pybus O.G., Allain J.P., Hatzakis A.
The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis.
PLoS Med. 2009;6(12):e1000198.

[67] Simmonds P.
The origin and evolution of hepatitis viruses in humans.
J Gen Virol. 2001;82(Pt 4):693-712.

[68] Ndjomou J., Pybus O.G., Matz B.
Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon.
J Gen Virol. 2003;84(Pt 9):2333-41.

[69] Pybus O.G., Barnes E., Taggart R., Lemey P., Markov P.V., Rasachak B., Syhavong B., Phetsouvanah R., Sheridan I., Humphreys I.S., Lu L., Newton P.N., Klenerman P.
Genetic history of hepatitis C virus in East Asia.
J Virol. 2009;83(2):1071-82.

[70] Cochrane A., Searle B., Hardie A., Robertson R., Delahooke T., Cameron S., Tedder R.S., Dusheiko G.M., De Lamballerie X., Simmonds P.
A genetic analysis of hepatitis C virus transmission between injection drug users.
J Infect Dis. 2002;186(9):1212-21. Epub 2002 Oct 11.

[71] Murphy D.G., Willems B., Deschênes M., Hilzenrat N., Mousseau R., Sabbah S.
Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5' untranslated region sequences.
J Clin Microbiol. 2007;45(4):1102-12.

[72] Markov P.V., van de Laar T.J., Thomas X.V., Aronson S.J., Weegink C.J., van den Berk G.E., Prins M., Pybus O.G., Schinkel J.
Colonial history and contemporary transmission shape the genetic diversity of hepatitis C virus genotype 2 in Amsterdam.
J Virol. 2012;86(14):7677-87.

[73] Messina J.P., Humphreys I., Flaxman A., Brown A., Cooke G.S., Pybus O.G., Barnes E.
Global distribution and prevalence of hepatitis C virus genotypes.
Hepatology. 2015;61(1):77-87.

[74] Wang Y., Okamoto H., Tsuda F., Nagayama R., Tao Q.M., Mishiro S.
Prevalence, genotypes, and an isolate (HC-C2) of hepatitis C virus in Chinese patients with liver disease.
J Med Virol. 1993;40(3):254-60.

[75] Moradpour D., Penin F., Rice C.M.
Replication of hepatitis C virus.
Nat Rev Microbiol. 2007;5(6):453-63.

[76] You S., Stump D.D., Branch A.D., Rice C.M.
A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication.
J Virol. 2004;78(3):1352-66.

[77] André P., Perlemuter G., Budkowska A., Bréchot C., Lotteau V.
Hepatitis C virus particles and lipoprotein metabolism.
Semin Liver Dis. 2005;25(1):93-104.

[78] Boulant S., Montserret R., Hope R.G., Ratinier M., Targett-Adams P., Lavergne J.P., Penin F., McLauchlan J.
Structural determinants that target the hepatitis C virus core protein to lipid droplets.
J Biol Chem. 2006;281(31):22236-47.

[79] Dubuisson J., Penin F., Moradpour D.
Interaction of hepatitis C virus proteins with host cell membranes and lipids.
Trends Cell Biol. 2002;12(11):517-23.

[80] Sakai A., Claire M.S., Faulk K., Govindarajan S., Emerson S.U., Purcell R.H., Bukh J.
The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences.
Proc Natl Acad Sci U S A. 2003;100(20):11646-51.

[81] Pietschmann T., Kaul A., Koutsoudakis G., Shavinskaya A., Kallis S., Steinmann E., Abid K., Negro F., Dreux M., Cosset F.L., Bartenschlager R.
Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras.
Proc Natl Acad Sci U S A. 2006;103(19):7408-13.

[82] Wölk B., Sansonno D., Kräusslich H.G., Dammacco F., Rice C.M., Blum H.E., Moradpour D.
Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines.
J Virol. 2000;74(5):2293-304.

[83] Egger D., Wölk B., Gosert R., Bianchi L., Blum H.E., Moradpour D., Bienz K.
Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex.
J Virol. 2002;76(12):5974-84.

[84] Appel N., Pietschmann T., Bartenschlager R.
Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain.
J Virol. 2005;79(5):3187-94.

[85] Moradpour D., Brass V., Bieck E., Friebe P., Gosert R., Blum H.E., Bartenschlager R., Penin F., Lohmann V.
Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication.
J Virol. 2004;78(23):13278-84.

[86] Branch A.D., Stump D.D., Gutierrez J.A., Eng F., Walewski J.L.
The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others.
Semin Liver Dis. 2005;25(1):105-17.

[87] Agnello V., Abel G., Elfahal M., Knight G.B., Zhang Q.X.
Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor.
Proc Natl Acad Sci U S A. 1999;96(22):12766-71.

[88] Scarselli E., Ansuini H., Cerino R., Roccasecca R.M., Acali S., Filocamo G., Traboni C., Nicosia A., Cortese R., Vitelli A.
The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus.
EMBO J. 2002;21(19):5017-25.

[89] Evans M.J., von Hahn T., Tscherne D.M., Syder A.J., Panis M., Wölk B., Hatziioannou T., McKeating J.A., Bieniasz P.D., Rice C.M.
Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry.
Nature. 2007;446(7137):801-5.

[90] Kapadia S.B., Chisari F.V.
Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids.
Proc Natl Acad Sci U S A. 2005;102(7):2561-6.

[91] Lavanchy D.
Evolving epidemiology of hepatitis C virus.
Clin Microbiol Infect. 2011;17(2):107-15.

[92] Alter M.J.
Epidemiology of hepatitis C virus infection.
World J Gastroenterol. 2007;13(17):2436-41.

[93] Lauer G.M., Walker B.D.
Hepatitis C virus infection.
N Engl J Med. 2001;345(1):41-52.

[94] Lo Re III V., Kostman J.R.
Management of chronic hepatitis C
Postgrad Med J 2005;81:376–382.

[95] Alberti A., Chemello L., Benvegnù L.
Natural history of hepatitis C.
J Hepatol. 1999;31 Suppl 1:17-24.

[96] Tilg H.
New insights into the mechanisms of interferon alfa: an immunoregulatory and anti-inflammatory cytokine.
Gastroenterology. 1997;112(3):1017-21.

[97] Lau J.Y., Tam R.C., Liang T.J., Hong Z.
Mechanism of action of ribavirin in the combination treatment of chronic HCV infection.
Hepatology. 2002;35(5):1002-9.

[98] Kamal S.M.
Pharmacogenetics of hepatitis C: transition from interferon-based therapies to direct-acting antiviral agents.
Hepat Med. 2014;6:61-77

[99] Asselah T., Estrabaud E., Bieche I., Lapalus M., De Muynck S., Vidaud M., Saadoun D., Soumelis V., Marcellin P.
Hepatitis C: viral and host factors associated with non-response to pegylated interferon plus ribavirin.
Liver Int. 2010;30(9):1259-69.

[100] Scheel T.K., Rice C.M.
Understanding the hepatitis C virus life cycle paves the way for highly effective therapies.
Nat Med. 2013;19(7):837-49.

[101] Kohli A., Shaffer A., Sherman A., Kottilil S.
Treatment of hepatitis C: a systematic review
JAMA. 2014;312(6):631-40.

[102] Kim Do Y., Ahn S.H., Han K.H.
Emerging therapies for hepatitis C.
Gut Liver. 2014;8(5):471-9

[103] Gentile I., Borgia F., Coppola N., Buonomo A.R., Castaldo G., Borgia G.
Daclatasvir: the first of a new class of drugs targeted against hepatitis C virus NS5A.
Curr Med Chem. 2014;21(12):1391-404.

[104] Ilyas J.A., Vierling J.M.
An overview of emerging therapies for the treatment of chronic hepatitis C.
Med Clin North Am. 2014;98(1):17-38.

[105] McCown M.F., Rajyaguru S., Kular S., Cammack N., Nájera I.
GT-1a or GT-1b subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-796.
Antimicrob Agents Chemother. 2009;53(5):2129-32.

[106] Saludes V., González V., Planas R., Matas L., Ausina V., Martró E.
Tools for the diagnosis of hepatitis C virus infection and hepatic fibrosis staging
World J Gastroenterol 2014;20(13):3431-42.

[107] European Association for the Study of the Liver.
EASL Clinical Practice Guidelines: management of hepatitis C virus infection.
J Hepatol. 2011;55(2):245-64.

[108] Harrington P.R., Zeng W., Naeger L.K.
Clinical relevance of detectable but not quantifiable hepatitis C virus RNA during boceprevir or telaprevir treatment.
Hepatology. 2012;55(4):1048-57.

[109] Descamps V., Op de Beeck A., Plassart C., Brochot E., François C., Helle F., Adler M., Bourgeois N., Degré D., Duverlie G., Castelain S.
Strong correlation between liver and serum levels of hepatitis C virus core antigen and RNA in chronically infected patients.
J Clin Microbiol. 2012;50(2):465-8.

[110] Pickett B.E., Striker R., Lefkowitz E.J.
Evidence for separation of HCV subtype 1a into two distinct clades.
J Viral Hepat. 2011;18(9):608-18

[111] Chevaliez S., Bouvier-Alias M., Brillet R., Pawlotsky J.M.
Hepatitis C virus (HCV) genotype 1 subtype identification in new HCV drug development and future clinical practice.
PLoS One. 2009;4(12):e8209.

[112] Dreier J., Störmer M., Kleesiek K.
Use of bacteriophage MS2 as an internal control in viral reverse transcription-PCR assays.
J Clin Microbiol. 2005;43(9):4551-7.

[113] Sandres-Sauné K., Deny P., Pasquier C., Thibaut V., Duverlie G., Izopet J.
Determining hepatitis C genotype by analyzing the sequence of the NS5b region.
J Virol Methods. 2003;109(2):187-93

[114] Chinchai T., Labout J., Noppornpanth S., Theamboonlers A., Haagmans B.L., Osterhaus A.D., Poovorawan Y.
Comparative study of different methods to genotype hepatitis C virus type 6 variants.
J Virol Methods. 2003;109(2):195-201.

7. AKNOWLEDGMENTS

I am using this opportunity to express my gratitude to AB ANALITICA, in the person of Dr. Dino Paladin, for give me the possibility to carry out this Ph.D. project.
I express my warm thanks to Dr.ssa Katia Bortolozzo, Dr.ssa Silvia Mason and Dr.ssa Anna Gani for their support and guidance at AB ANALITICA.
I would also thank my project external supervisor Prof.ssa Maria Luisa Mostacciuolo for her critical guidance.
I thank the following experts who provided me clinical samples and scientific advice required for my project:

- Dr.ssa Maria Cristina Dechecchi, Dr. Giulio Cabrini (Laboratory of Molecular Pathology, Department of Pathology and Diagnostics, University Hospital of Verona), for providing samples for ENaC study.
- Prof. Luca Frulloni (Department of Medicine, Section of Gastroenterology, University of Verona), for his scientific support for ENaC study, relative to atypical CF patients with pancreatitis.
- Prof. Alfredo Alberti (Venetian Institute of Molecular Medicine, Padova), for providing samples for HCV study.

Finally, I would thank my colleagues of AB ANALITICA, my parents and Daniele, for his unconditionally support.