Roberta Marino

Role of rs16944 (IL1B) and rs5743899 (TOLLIP) polymorphisms in the year-long presence of rhinovirus in the nostrils of healthy young volunteers

Universidade Fernando Pessoa

Porto, 2018

Roberta Marino

Role of rs16944 (IL1B) and rs5743899 (TOLLIP) polymorphisms in the year-long detection of rhinovirus in the nostrils of healthy young volunteers

Universidade Fernando Pessoa

Porto, 2018

Roberta Marino

Role of rs16944 (IL1B) and rs5743899 (TOLLIP) polymorphisms in the year-long detection of rhinovirus in the nostrils of healthy young volunteers

Roberta Marino

Paper presented to Fernando Pessoa University as part of the requirements for obtaining a Master's Degree in Pharmaceutical Sciences, under the guidance of Professor José Manuel Cabeda

RESUMO

A infecção por rinovírus é uma das causas mais frequentes da constipação. Os mais afetados são crianças, idosos e portadores de doença respiratória crónica, que podem incorrer em doenças mais graves e complicações. Os rinovírus pertencem à família Picornaviridae sendo assim pequenos vírus de RNA de cadeia positiva com uma cápside proteica icosaédrica. Estes vírus são transmitidos muito facilmente atacando principalmente a mucosa nasal, mas a frequência e intensidade da infeção varia muito com os indivíduos e a época do ano.

No presente trabalho procuramos caracterizar a influência de 2 polimorfismos em genes do sistema imunológico inato na infeção pelo rinovírus. Para tal, 89 voluntários (estudantes universitários) doaram mensalmente amostras nasais para a deteção rinovírus por RT-qPCR. Alguns destes indivíduos doaram também uma amostra de sangue (n=33) ou de epitélio bucal (n=15) para obter DNA genómico e caracterizar SNP's dos genes TOLLIP e IL1B (rs5743899 e rs16944 respetivamente) por PCR-RFLP.

Os resultados não revelaram qualquer associação do polimorfismo rs16944 (IL1B) e a presença de rinovírus nasal. Contudo todos os indivíduos que nunca testaram positivo para o rinovírus apresentaram o alelo G do polimorfismo rs5743899 (gene TOLLIP), sugerido um possível efeito deste alelo na eficaz eliminação imediata do rinovírus antes de ocorrer infeção viral. Contudo, o mesmo alelo apresentava também os valores mais elevados de títulos nasais de rinovírus, sugerindo que uma vez estabelecida a infeção estes indivíduos têm maior dificuldade em a controlar. Estes resultados estão em linha com estudos anteriores indicando que este alelo diminui a expressão de TOLLIP, potenciando a sinalização via TLR2, uma via particularmente importante no reconhecimento da cápside viral. Outros estudos revelaram ainda que o mesmo alelo está associado à diminuição da expressão de genes antivirais importantes como o IFN-λ1, o que pode potenciar uma menor eficácia de eliminação viral assim que a infeção viral ocorre.

Palavras-chave: Rhinovirus, Picornaviridae, SNP, TOLLIP, IL1B, rs5743899, rs16944, qPCR, PCR-RFLP.

ABSTRACT

Rhinovirus infection is one of the most frequent causes of the common cold. Individuals most affected are children, elders and chronic respiratory patients who may present complications and a more severe, even life-threatening disease. Rhinoviruses belong to the Picornaviridae family thus being positive stranded RNA viruses with an icosahedral proteic capsid. These viruses are easily transmitted attacking mainly the nasal mucosa, but infection frequency varies considerably among individuals and with the season.

In the present work we characterized the influence of two polymorphisms in innate immune system genes on rhinovirus infection. For that purpose, 89 voluntaries (university students) donated monthly nasal swabs for rhinovirus detection by RT-qPCR. Additionally, some individuals donated a blood sample (n=33) or buccal swab (n=15) for DNA extraction and SNP characterization of rs5743899 (TOLLIP gene) and rs16944 (IL1B gene) by PCR-RFLP.

Results revealed no relation between rs16944 (IL1B) and rhinovirus nasal detection. However, every single individual that did never had rhinovirus detected in their nostril throughout the one year observation period showed the rs574388 G allele (TOLLIP gene), suggesting a possible effect of this allele in the efficacious prompt elimination of the virus before infection is attained. However, the same allele showed the highest viral titres, suggesting that once the infection is established, these individuals struggle to control it. These results are in agreement with previous studies indicating that the G allele decreases TOLLIP expression increasing TLR2 signalling, a particularly important route for RHV capsid detection. Other results also indicate that the same allele decreases antiviral gene expression (particularly IFN-λ1) which may potentiate a lower efficacy of viral clearance after infection is established.

Keywords: Rhinovirus, Picornaviridae, SNP, TOLLIP, IL1B, rs5743899, rs16944, qPCR, PCR-RFLP.

Dedicatoria

A chi c'è e ci sarà sempre;
a chi non c'è più ma ci sarà sempre;
a tutti voi ma soprattutto a me stessa!

Acknowledgments

I would like to thank the University Fernando Pessoa and the University Federico II, for having allowed me to do this important and wonderful experience.

But above all, I thank Professor José Manuel Cabeda, for all the teachings, the knowledge and the passion for this work that he has transmitted to me. For the patience, understanding, commitment and availability that has always shown towards me.

My parents who support me and realize my desires.

And Francesco who never stopped encouraging me.

INDEX OF CONTENT

List of Figures

List of Tables

List of abbreviations

A – Adenine

ACAS – Community-acquired acute sinusitis

AERD – Aspirin-exacerbated respiratory disease

AOM – Acute Otitis media

ATA – Aspirin-tolerant asthma

C – Cytosine

COPD – Chronic obstructive pulmonary disease

COX-2 – Cyclooxygenase 2

DNA - Deoxyribonucleic Acid

dNTP – Deoxynucleoside triphosphates

EDTA – Ethylenediamine tetraacetic acid

FRET – Fluorescent resonance energy

G – Guanine

GM-CSF – Granulocyte-macrophage colony-stimulating factor

HRV – Human Rhinovirus

ICAM-1 - Intercellular Adhesion Molecule 1

IFNγ - Interferon gamma

IL-1 – Interleukin 1

IL-6 – Interleukin 6

iNOS – Inducible nitric oxide synthase

LD – Linkage disequilibrium

LDL – Low-density lipoprotein

mRNA – Messenger RNA

n.s. – Not significant

PCR – Polymerase Chain Reaction

PGE2 – Prostaglandin E2

PRR – Pattern Recognition Receptors

qPCR – Real Time PCR (quantitative PCR)

RFLP – Restriction fragment length polymorphism

RFU – Relative fluorescence units

RHV – Rhinovirus

RHV^{+} – Positive Rhinovirus Sample(s) / Episode(s)

RHV^{-} – Negative Rhinovirus Sample(s) / Episode(s)

RNA – Ribonucleic Acid

RT – Reverse-transcription reaction

SNPs – Single Nucleotide Polymorphism

SPSS - Statistical Package for Social Science

T – Thymine

Ta – Annealing temperature

TLR – Toll-Like receptor

Tm – Melting temperature

TNF – Tumour necrosis factor

TRIS-HCL – Tris hydrochloride

UTR – Untranslated regions

1. INTRODUCTION

1.1. Rhinoviruses

Rhinovirus infection is one of the most frequent causes of the acute upper respiratory infection, otherwise known as common cold. Despite being usually a mild, self-resolving infection, complications are not rare. In fact, these viruses are also associated with more serious diseases, such as acute otitis media (AOM) and community-acquired acute sinusitis (ACAS) (Blomqvists, 2004). Most at risk are children, the elderly and the immunocompromised individuals (Lu, 2014). Also, people who already suffer from chronic respiratory diseases, such as asthma, chronic bronchitis, chronic obstructive pulmonary disease (COPD) and cystic fibrosis are more likely to have complications associated with rhinovirus infection (Semler, 2004).

The rhinoviruses (HRVs) belong to the Picornaviridae family, together with the human Enteroviruses, and are genetically distinguished in three groups A, B and C (Figure 1).

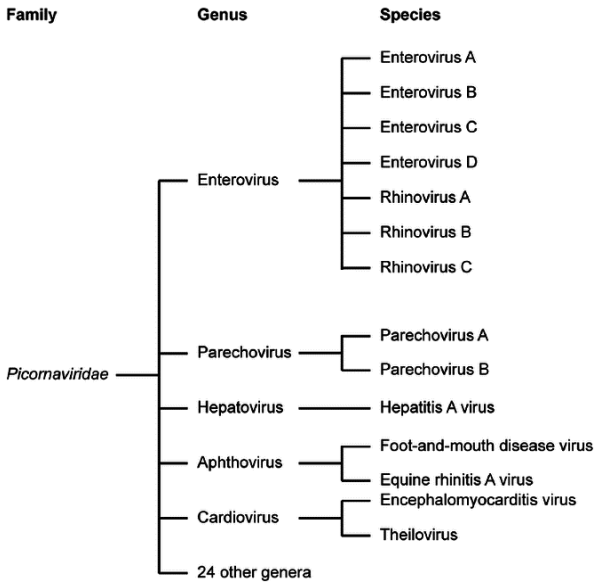

Figure 1. Classification of the clinically most important virus of the Picornaviridae family (van der Linden, 2015)

To date, more than 100 types of serologically distinct HRVs are known (Cordey, 2008). While leaving a type-specific permanent immunity, infection with one type of virus does not offer protection to other types (Monto, 2002). In fact, the cross-protection between the various serotypes is very low or absent, and this explains why young children experience on average 5 to 10 cold episodes a year, especially during the second year of life and in the autumn, winter and spring, right in the period when nurseries are more frequented. The adult, on the other hand, presents on average two to four cold respiratory infections per year, (Waris, 2008).

1.1.1. Structure

All picornaviruses (pico = very small; RNA = from the nucleic acid type) are small (the virion has a diameter of 27-30 nm), have a spherical protein capsid containing an RNA molecule and a molecular weight that is around the 8.25 x 10^6 Da (Semler, 2004). The capsid has an icosahedral symmetry and consists of 60 protomers, each of which is composed of 4 non-glycosylated proteins: VP1, VP2, VP3, on the surface of the capsid, and VP4, inside the capsid and in close contact with viral RNA (Figure 2) (Lewis-Rogers, 2009).

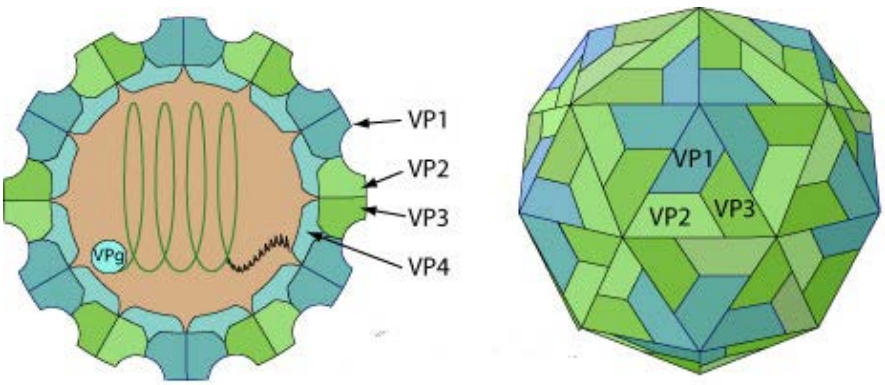

Figure 2. Picornaviridae virion structure (ViralZone 2008, Swiss institute of bioinformatic)

As said, it is composed of molecular RNA with a molecular weight of about $2.6x10^6$Da, a single chain and a positive polarity. It has only one "open reading frame" that encodes both structural and functional proteins (Figure 3). In the viral particle and in the initial phases of replication it is associated with the viral protein VPg (Figure 2 and 3) (Cordey, 2008).

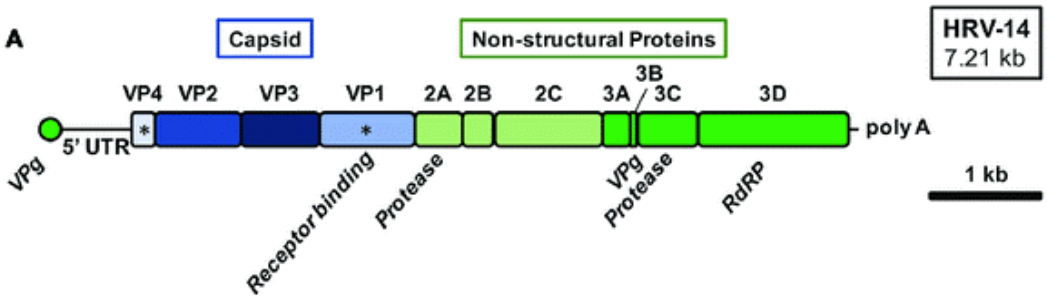

Figure 3. Human rhinovirus genomic organization (Stobart, 2017)

1.1.2. Replication

Viruses are obligatory endocellular parasites, they must therefore find, adhere and penetrate the host cell in which their replicative role can be played.

The attack occurs through the specific recognition between viral antireceptor and the cell receptor. Receptor molecules are surface molecules, whose physiological functions are subverted by viruses. They may be specific receptors such as proteins (often glycoproteins) or less specific receptors such as carbohydrate residues (glycoproteins or glycolipids) (Blomqvists, 2004).

In the case of HRVs, the binding occurs on the bottom of the so-called "canyons" (depressions on the surface of the virus), which are areas with a constant sequence that are not affected by the selective pressure of the immune system, given their difficult accessibility for antibody binding (figure 4). The binding occurs either with the

intercellular adhesion molecule 1 (ICAM-1) (main receptor group) or a member of the low-density lipoprotein (LDL) receptor family (Lewis-Rogers, 2009).

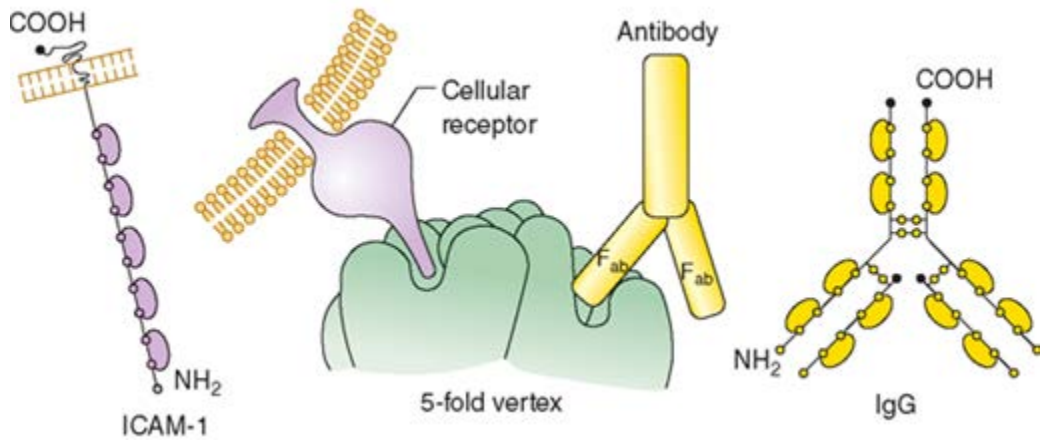

Figure 4. Binding of cellular receptor ICAM-1 and of the antibody to the floor of the canyon. ICAM-1 has a diameter roughly half that of an immunoglobulin G (IgG) antibody molecule (Jawetz, Melnick, & Adelberg's Medical Microbiology)

The attachment of the virus to the cell receptor triggers conformational changes in the capsid of the virus, eventually leading to a release of viral RNA into the cytoplasm (Figure 5-A).

In the cell cytoplasm, the viral genome is translated to provide viral proteins essential for the replication of the genome and the production of new viral particles. The long polyprotein precursor splits first in the intermediates P1, P2 and P3 (Figure 5-B) through two proteinases, 2A and 3C / 3CD. P1 is then further divided to provide VP0, VP1 and VP3 (Heweat, 2010). VP0 is cleaved to VP4 and VP2 during viral assembly (Figure 5-C). Among the other proteins synthesized are the RNA-dependent viral RNA polymerase and the accessory proteins required for genome replication and mRNA synthesis. Positive-stranded RNA is copied to a negative chain intermediate (Figure 5-D), an RNA molecule with a complementary nucleotide sequence. The synthesis of the complementary (-) strain is initiated at the 3 'end of the RNA and triggered by the VPg

protein. This complementary strain (-) is subsequently used as a model to synthesize a large number of copies of the viral genome (Figure 5-E), which are used as mRNA for the translation of multiple viral proteins or, later in the infection cycle, packaged into new viral particles. The new infectious viruses are brought together in the cytoplasm and finally released by lysis of the infected cell (Figure 5) (Rancaniello, 2001).

The replication kinetics of HRVs is rapid (the cycle is completed in 5-10 hours) and very efficient (the number of virions produced by a single infected cell ranges from 10,000 to 100,000) although only a small part of the virions produced are infectious (from 1/10 to 1/1000) (Caprioli, 2005).

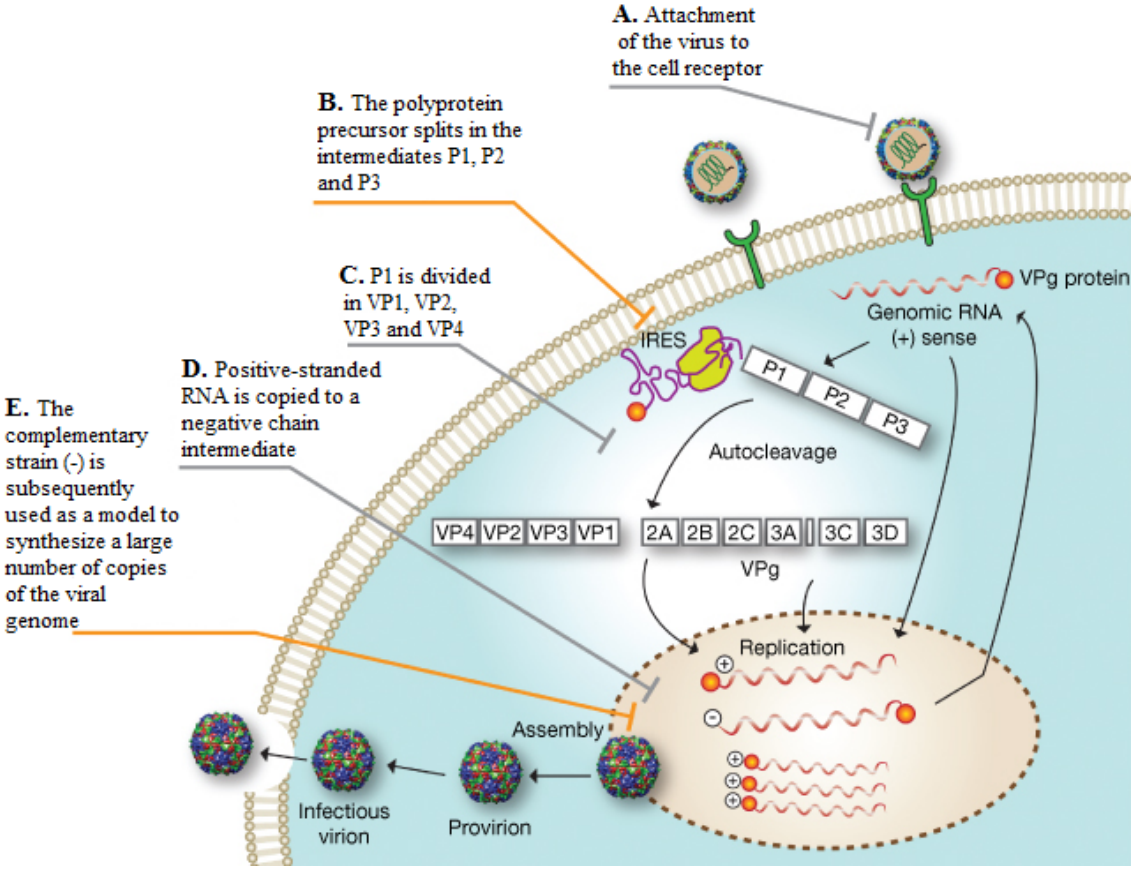

Figure 5. Rhinovirus life cycle (Benschop, 2015)

1.1.3. Transmission

The HRV mainly attacks the nasal mucosa. This may be related to the fact that the cells of the nasal epithelium are closer to the optimal growth temperature (33°C) of most HRVs (Coleman, 2015).

Transmission of infection occurs by the efficient entry of the virus into the nose or tear ducts of a susceptible subject (Johannessen, 2012).

Transmission can occur through three ways:

1) By direct contact with the hands of an infected person even without respiratory symptoms, even for as short a time as 10 seconds; in this way the particles are transferred to the hands of the next subject. The recipient is infected by taking contaminated fingers on the nasal or conjunctival mucosa (Johannessen, 2012).
2) By indirect contact, transmission occurs by contact with contaminated objects. like glasses, coffee cups handle and door knobs. In fact, HRVs can remain viable on these objects for hours or even days (Johannessen, 2012).
3) The virus can be transferred from infectious aerosols, such as particles produced by coughing, sneezing, chattering, etc. (Johannessen, 2012).

The first symptom is often a sore throat and may occur immediately after virus entry in the nose. The disease incubation period may be as short as 2 or 3 days but may be as long as 5 days post-infection (Lessler et al. 2009). The median duration of the disease is seven days in young adults but may be up to two weeks in a quarter of cases or even longer in children and the elderly (Monto, 2002), leaving rhinovirus presence in one infection episode usually extending to almost one month.

Infections with HRVs occur throughout the year, but mostly in spring and autumn. In particular, the highest peak is found in September, and coincides with the start of school, which certainly improves the means for efficient transmission (Blomqvists, 2004).

The reasons for seasonal behaviour may be due to the fact that HRVs persists better in an environment where the relative humidity is above 50% (Blomqvist, 2004).

1.2. Single nucleotide polymorphisms (SNPs)

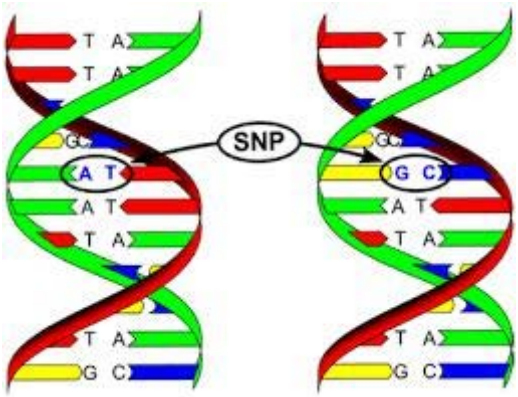

Figure 6. Single nucleotide polymorphism

(http://www.martaliveonlus.org/sites/default/files/allegati/news-ricerca/SNP_square.jpg)

A genetic variation can be defined as the presence of multiple alleles for the same locus. These include polymorphisms and mutations, the latter are pathology associated genetic variations, whereas polymorphisms do not usually cause pathologies and show the least frequent allele in the population in a proportion greater than 1% (Brookes, 1999).

Single nucleotide polymorphisms (SNPs) are variations in a single base in the DNA sequence (Figure 6). These variations occur when one nucleotide is replaced by another. In general, 2/3 of the SNPs involve the change from C to T (both pyrimidines) (Ameur et al., 2009).

SNPs are the most common genetic variants within the genome and have therefore become the most frequently used markers in association study maps. According to some estimates, in the world population there are about ten million SNPs for which both alleles are present with a frequency of more than 1%. These SNPs (common SNPs) make up

90% of the variability in the world population, while the remaining 10% is made up of a diversified set of other rare variants (Kruglyak and Nickerson, 2001). It is estimated that the polymorphisms are distributed along the genome at irregular intervals, about every 300bp. It is hypothesized that during the evolution, whole groups of polymorphisms were selected in functional regions of the genome. Recently the position of polymorphisms has become the object of great attention. Non-synonymous polymorphisms would cause an amino acid to change and thus the synthesis of an altered protein. Polymorphisms in promoters, splice regions, 3'- and 5'-UTR regions and in intragenic regions could alter protein expression, through changes in its regulation, alternative splicing or mRNA stability. Although intergenic polymorphisms are common, only a few showed they could alter gene expression (Ameur et al., 2009). It has been estimated that around 50,000 to 200,000 polymorphisms could have a biological effect (Bernig and Chanock, 2006).

The formation of a genetic variant occurs following a single mutational event that occurs with a very low frequency (in the order of 2.5×10^{-8} mutations per site or 175 mutations per diploid genomes per generation). Each new variant is initially associated with the entire genomic region in the vicinity of the chromosomal locus in which the mutation occurred. The specific set of alleles that are on the same chromosome or in the same sub chromosomal region is called a haplotype, therefore the onset of a mutation leads to the formation of a new haplotype, just like the formation of a mosaic haplotype in the case of recombination between two different haplotypes present on the chromosome of maternal and paternal origin (Nachman, 2000).

During the evolution of humanity many mutations have occurred. Association studies are based on the assumption that a mutation occurring in a single ancestral chromosome is inherited from past generations and transmitted with the entire adjacent genomic region, as a single block. This mutation therefore represents a distinctive sign, a marker, which

allows us to identify that particular segment of the ancestral chromosome (Pääbo, 2003). It turns out that neighbouring genes are more easily co-inherited in blocks of haplotypes, resulting in association between populations that originate from a common ancestor. This phenomenon is called linkage disequilibrium (LD) (Reich and Lander, 2001).

The haploblocks therefore originate from short segments of an ancestral chromosome passed through numerous generations without recombination. SNPs in strong linkage disequilibrium along the genome are called SNP-haploblocks. By virtue of this it is not necessary to genotype all 13 SNPs within an haploblock, but only a few are called tagSNPs, through which all the other SNPs can be deduced (Cheng, 2007). It is assumed that the genotype of non-directly genotyped polymorphisms can be correlated with one or more markers. This non-random correlation indicates that these markers are in linkage disequilibrium between them (Carlson et al., 2003). Since the probability of recombination between two SNPs increases with increasing physical distance between the two, it is clear that the degree of association between SNPs progressively decreases with distance (Reich and Lander, 2001).

However, after transmission through various generations the markers undergo various genetic recombination. As a result, only those markers in close proximity to a particular locus (for example the locus of a predisposition gene for a particular pathology) will remain in linkage disequilibrium in the generations of individuals affected by the same pathology (Pääbo, 2003). However, several studies have highlighted how the Linkage Disequilibrium framework along the genome is very complex, variable and difficult to predict only through theoretical models (Gabriel et al., 2002).

The polymorphisms are therefore genetic markers that can be used in maps of complex traits of a disease but require related analyses to establish the causes or functional meaning of the variant. The situation is complicated by the fact that many polymorphisms

seem to be population specific: about 15% of the polymorphisms could differ in different populations (Berning and Chanock, 2006).

1.2.1. **Interleukin 1B Gene**

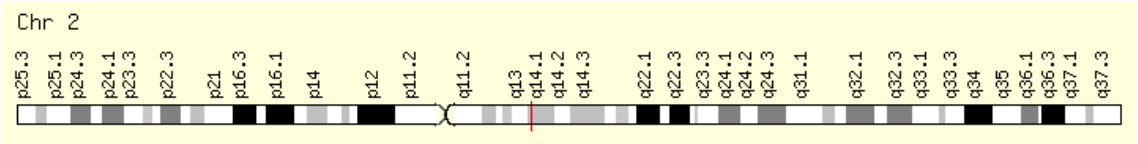

Figure 7. Genomic Locations for IL1B Gene on chromosome 2 (red line) (www.genecards.org)

The Interleukin 1 genes are located on chromosome 2 (Figure 7) where there is an aggregate of genes encoding both for IL-1β, IL-1α and for the receptor of these two molecules (Figure 8).

Interleukins-1 are pluripotent cytokines, that are able to perform and regulate many immune functions and are specially involved in the activation of inflammatory responses. They are present in two isoforms, IL-1α and IL-1β.

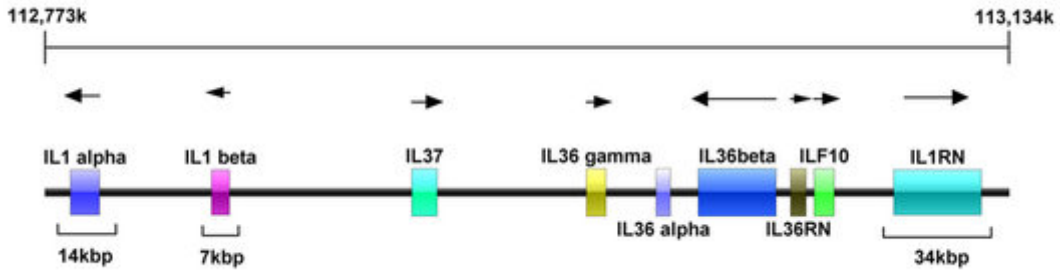

Figure 8. Representation of the human chromosome-2 locus containing the IL-1 gene cluster showing the relative size and positions of the known IL-1 genes within the IL-1 gene cluster (Khazim, 2018)

It seems that both forms act as mediators of biological activities of the same type, which includes synthesis of acute phase proteins by the hepatocytes, polymorphonuclear chemotaxis and release of polymorphonuclear cells from the blood and bone marrow (Mattila, 2002).

IL-1β, a cytokine of 17.5 kDa is an important mediator of the inflammatory response, is released into the blood stream exercising widespread actions in the body; it is involved in a variety of cellular activities, including cell proliferation, differentiation, apoptosis and the increase of collagenase synthesis. It is one of the factors capable of inducing fever. It induces the expression of cyclooxygenase 2 (COX2) and nitric oxide synthase (iNOS), which induces nitric oxide synthesis to contribute to the decrease of vascular resistance. Once in the circulation, it binds to the receptors present in the hypothalamic vessels determining the synthesis of COX-2 which in turn induces an increase in the levels of prostaglandin E2 (PGE2). The latter activates the hypothalamic centre of thermoregulation leading to the appearance of fever (Van der Poll e Van Deventer, 1999). It is an essential cytokine for host response and resistance to pathogens, but it can also exacerbate damage during chronic diseases and acute tissue injury.

It is mainly produced by monocytes/macrophages, but also by dendritic cells, neutrophils, T and B lymphocytes, endothelial cells, keratinocytes, fibroblasts, astrocytes and microglia cells. In the peripheral circulation, IL-1β activates endothelial receptors so as to allow the expression of adhesion molecules and chemokines that facilitate the migration of neutrophils to tissues. This interleukin also amplifies the expression of other cytokines, such as TNF-α and IL-6. At bone marrow level, IL-1β results in an increase in myeloid progenitor cells by promoting neutrophil release resulting in neutrophilia, through the production of Granulocyte-macrophage colony-stimulating factor (GM-CSF) and causes a reduction in erythropoietin response to a state of anaemia.

In many studies, interleukin-1β has been reported to be involved in the genesis of asthma (Karjalainen, 2003), and of chronic rhinosinusitis with nasal polyposis in a Turkish population (Erbek, 2007). Most studies attempted to establish the association of polymorphisms in the IL1B promoter gene such as rs16944. In 2007, Erbek et al.

described a susceptibility for the development of nasal polyps associated with the rs16944 polymorphism. Tests for the role of IL-1β in pulmonary immune responses were collected in mouse models; the changes observed in these mice include a significant reduction in pulmonary eosinophilic inflammation, decreased goblet cell hyperplasia and reduced cell recruitment in the lungs (Schmitz, 2003).

In another study, it was shown that the frequency of IL1β -511 polymorphism (AA rs16944 genotype) is triple in patients with Aspirin-exacerbated respiratory disease (AERD) compared to Aspirin-tolerant asthma (ATA) patients, suggesting that patients carrying this polymorphism may show genetic susceptibility to develop AERD (Falfan-Valencia, 2011).

Being this protein involved in countless biological processes, but particularly in the immune mechanisms, it is possible that polymorphisms in this gene can modulate the susceptibility to HRV infection.

1.2.2. TOLLIP Gene

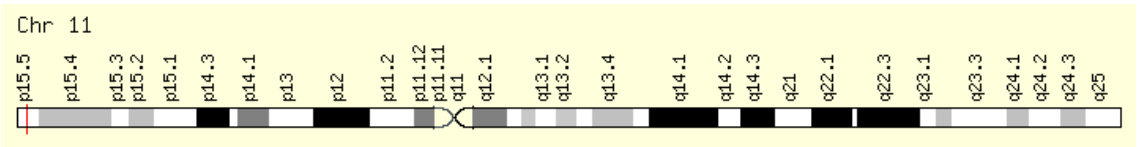

Figure 9. Genomic Locations for TOLLIP Gene on chromosome 11 (red line) (www.genecards.org)

The TOLLIP gene encodes the Toll interacting protein, also known as TOLLIP, a 30 kDa protein, which is an inhibitory adapter protein. It is recognized as a negative regulator of TLR signalling, thus regulates inflammatory signalling and is involved in trafficking of interleukin-1 receptors and IL1R-associated kinase turnover.

Toll-Like Receptors (TLRs) are a class of proteins that play a key role in the body's defence, particularly in innate immunity. They are single-pass, non-catalytic,

transmembrane receptors, expressed primarily on the membrane of sentinel cells such as macrophages and dendritic cells (Liu, 2014).

They recognize certain typical structures of pathogens and microbes and are part of the superfamily of "receptors that recognize molecular profiles" (Pattern Recognition Receptors or PRR). Once the pathogen has breached the anatomical barriers of the host (e.g. man's skin or intestinal mucosa), it is recognized by TLRs that activate the immune responses of sentinel cells (Abul, 2012).

Recent studies suggest that TOLLIP is also involved in autophagy. For example, human TOLLIP is necessary for the correct clearance of polyQ proteins related to Huntington's disease through ubiquitin-dependent autophagy (Lu, 2014).

Various pathways of innate immunity including Toll-like receptors (TLR) and autophagy are involved in HRV infection, and since TOLLIP is a negative regulator of these mechanisms, genetic variations may partly explain the variation in airway response to HRV infection (Lu, 2014).

Several TOLLIP SNPs have been implicated in sepsis (Song, 2011), tuberculosis (Shah, 2012) and idiopathic pulmonary fibrosis (Kropski, 2015), but the function of TOLLIP SNPs in HRVs infections is not yet clear. We hypothesized that TOLLIP SNPs (e.g., rs5743899) and changes associated with TOLLIP expression levels contribute to varying degrees of human airway epithelial responses to HRVs.

1.3. Polymerase chain reaction

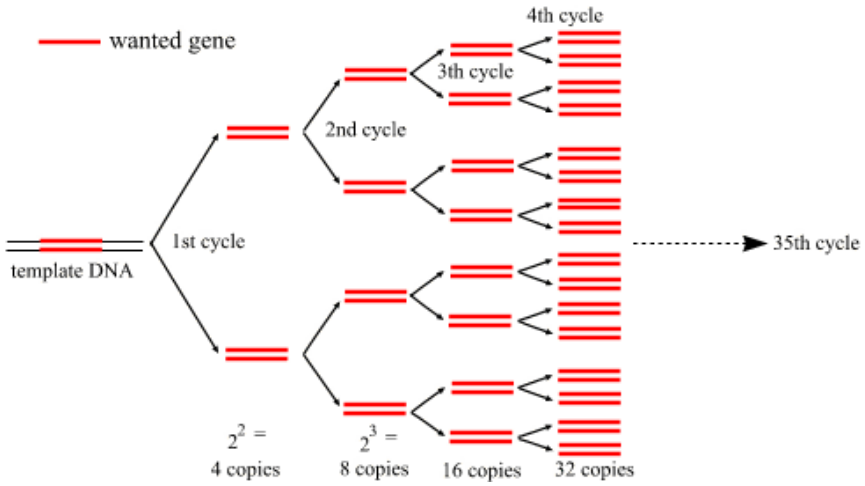

Figure 10. Theorical cycle in a PCR reaction (Hakhverdyan, 2018)

The polymerase chain reaction (PCR) is an in vitro nucleic acid amplification method. PCR allows the exponential synthesis of a DNA segment.

In every PCR cycle, theoretically, the double of the initial DNA is obtained; in general, there are 35 cycles, which translates into a theoretical amplification of over 34 billion ($34x10^9$) times (Figure 10). Virtually the efficiency is never 100%, so the amplification obtained will usually be lower than this theoretical limit (Cabeda et al., 2012).

The amplification mixture must contain (Scialpi, 2008):

- Target sequence, consisting of the genetic material containing the target regions to be amplified;
- Specific primers. DNA polymerase is not able to initiate DNA synthesis (polymerization) without having an oligonucleotide primer. This primer is then designed and synthesized in such a way that its sequence is complementary to that of the initial area to be amplified, providing a free 3 'end for the nucleotides to be added by the polymerase. Since the molecule we are amplifying is double-stranded, each strand will have its own primer. Primers are designed to limit the ends of the target sequence to be amplified with the 3' ends facing each other.

- Thermostable DNA polymerase
- dNTPs (deoxynucleoside triphosphates)
- Buffer, that ensures optimal conditions of activity of the enzyme used are added to the DNA.
- Magnesium chloride plays an important role in the solution as it affects specificity and sensitivity. An important parameter to be evaluated is its concentration because it influences both the specificity of the primer coupling and the activity of the enzyme (Scialpi, 2008).

For each cycle of the PCR reaction we can distinguish three phases: denaturation, the pairing or annealing and finally extension.

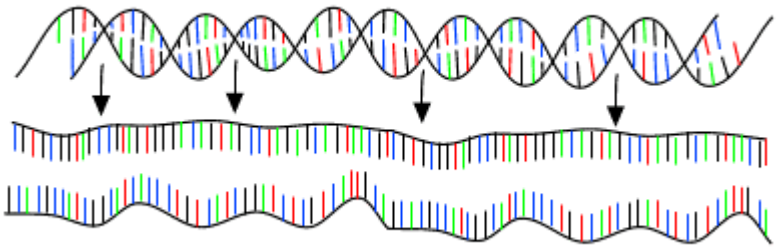

Figure 11. Step 1 of PCR: Denaturation (Hakhverdyan, 2018)

During denaturation (Figure 11), high temperatures of about 95°C are reached. In this phase the double helix of the DNA is separated into two single strands (thermal denaturation). This happens because the high temperature makes the hydrogen bonds, that link the chains, unstable, allowing them to break. Denaturation is favoured by the presence of relatively high saline concentrations. Since the half-life of "Taq pol" is about 30 minutes in a 95°C environment, the enzyme activity remains stable for the 30-40 cycles required to obtain millions of copies of the same target sequence (Clewley, 1995).

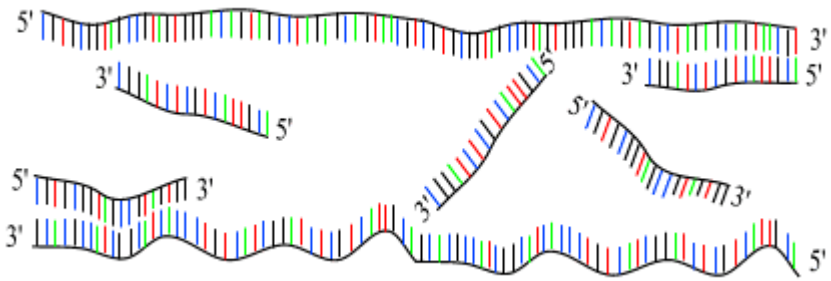

Figure 12. Step 2 of PCR: Annealing (Hakhverdyan, 2018)

In the second phase (Figure 12), the temperature is lowered to the annealing temperature (Ta) (reaching temperatures between 55 and 72°C): at these temperatures primers bind to their complementary sequences in the DNA mould, at the ends of the fragment of interest. The Ta is conditioned by primer sequence and their length. The annealing temperature (temperature at which 50% of the primer is single stranded) is often estimated, based on the primers melting temperature. However, Ta should always be experimentally defined (Weissensteiner, 2004).

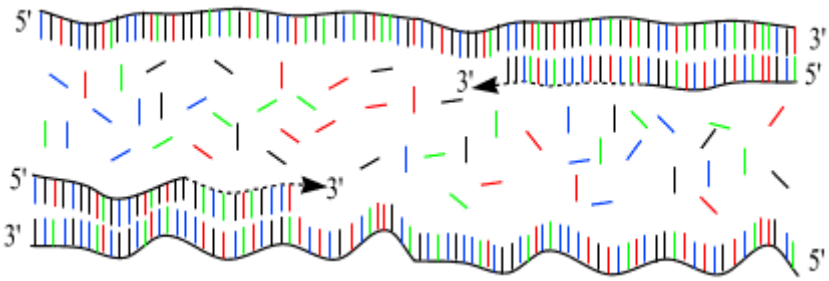

Figure 13. Step 3 of PCR: Extension (Hakhverdyan, 2018)

At the last phase, the extension (Figure 13), the temperature is slightly increased to maximize the action of the enzyme (the working temperature of the enzyme is usually between 68 and 72°C). In this phase the oligonucleotide primers are extended by the DNA

polymerase which incorporates the individual nucleotide complementary to the DNA template. The amplification cycle, consisting of the three phases just described, is repeated 30-40 times to obtain the exponential amplification of the target sequence.

In the last amplification cycles the reaction slows down, due to the consumption of reagents and the reduction of polymerase activity, until a plateau is reached, in which there is no more amplification due to the exhaustion of the reagents and the reaction product accumulation (Cabeda et al., 2012).

1.3.1. Real Time PCR

The conventional PCR technique is a qualitative technique. In fact, it is not possible to correlate the quantity of the final product with the quantity of DNA initially present, because in the final phase the efficiency of the reaction can be variable. A development of the PCR technique consists of real-time PCR (Bustin, 2004).

Real-time PCR, or qPCR (quantitative PCR), allows to quantify the synthesis of the PCR product at each amplification cycle in real time. This allows a quantitative estimate of the initial DNA to be made (Bustin, 2004).

The signal that is quantified is the fluorescence emitted by fluorophores (fluorescent dyes capable of binding to the DNA molecules produced at each amplification cycle). Fluorophores can bind to DNA in a nonspecific manner, or act as markers of oligonucleotide probes complementary to specific sequences (Filion, 2012).

Non-specific fluorophores used in real-time PCR are characterized by the development of fluorescence only when they are bound to the neo-synthesized DNA strands (Figure 14). The emitted fluorescence increases proportionally to the number of double stranded DNA molecules produced. The quantity of amplified product can therefore be determined

at each amplification cycle, detecting the emitted radiation of the fluorophore (Filion, 2012).

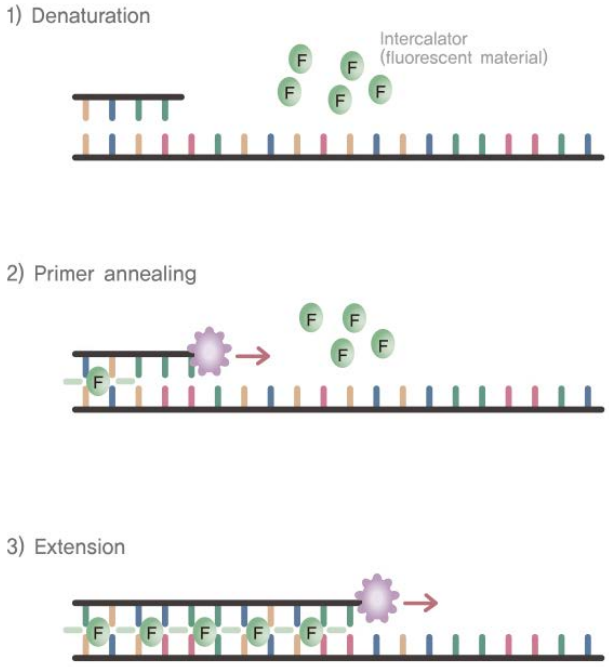

Figure 14. SYBR® Green detection (http://eng.bioneer.com)

One of the most used fluorophores is SYBR® Green, a fluorescent molecule that during the PCR reaction is intercalated in the DNA double strand. The DNA-SYBR® Green complex absorbs UV light (λmax = 488nm) and emits green light (λmax = 522nm). SYBR® Green intercalates in double-stranded DNA, which is why it is not suitable for quantifying single-stranded RNA. During the denaturing phase the SYBR green is free in the reaction mixture (Figure 14-1), then in the annealing phase and particularly at the end of polymerization stage it intercalates in the DNA molecule (Figure 14-2, 3) and when excited emits a fluorescence proportional to the number of copies of DNA produced (Bustin, 2004).

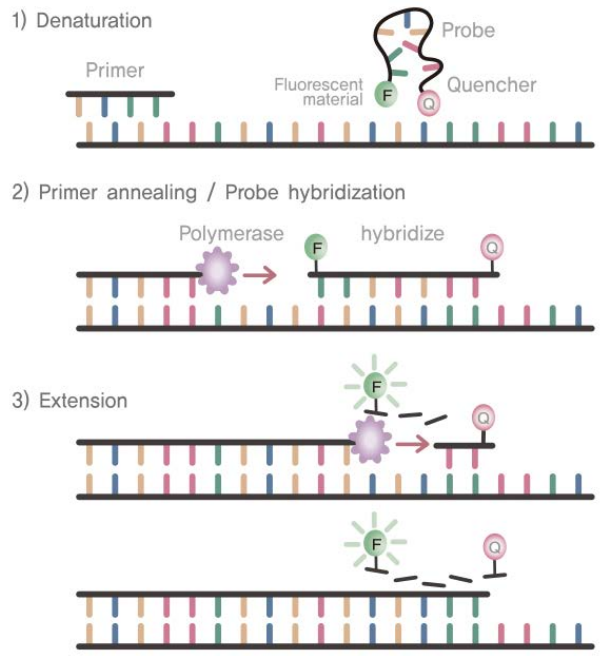

Figure 15. Taq-man Probe detection (http://eng.bioneer.com)

In analyses based on the use of specific probes, the fluorescent signal is detected only as a result of the probe's pairing to the target sequence. The probes typically used can be of several types, but the most frequent are: hydrolytic or hybridization probes (Bustin, 2004).

The hydrolytic probes have a fluorophore at each end: high-energy (reporter) and a fluorophore (quencher) inhibitor (Figure 15-1). When the probe is intact (Figure 15-1 and 15-2), the inhibitor is sufficiently close to the fluorophore to block the emission of the fluorescent signal. During the elongation (Figure 15-3), in each amplification cycle, the polymerase hydrolyses the probe. In this way the fluorophore is released in the reaction mixture and moves away from the inhibitor action, with a consequent emission of the fluorescent signal (Bustin, 2000).

So, in the analyses using hydrolytic probes the fluorescent signal emission depends on the activity of Taq DNA polymerase. Hydrolytic probes are also called TaqMan probes. The hybridisation probes, on the other hand, allow detection of the signal at the moment in which they bind to the target sequence. There are different models of hybridization probes. Among these we find a type of probe that exploits the transfer of fluorescent resonance energy (FRET) (Filion, 2012).

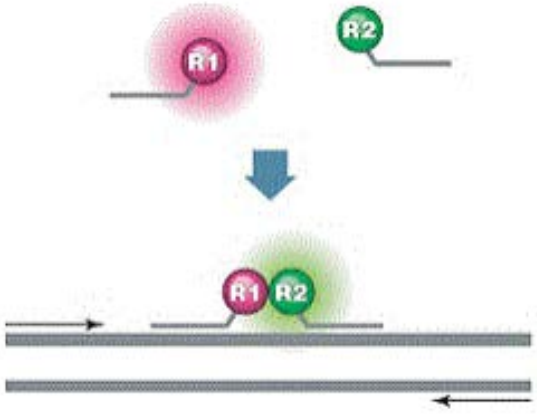

Figure 16. FRET probes. R1 is donator fluorophore, R2 the acceptor (Bio-Rad.com)

During the FRET process, the excitation of a donor fluorophore causes the transfer of energy to an acceptor fluorophore (Figure 16). The FRET probes are formed by a pair of oligonucleotides, designed in such a way as to hybridize close to one another on the target sequence. The signal that is detected is proportional to the quantity of hybridized probe and, consequently, to the amount of amplified product. Another variant of the hybridization probes consists of the "molecular beacon" (probes beacons) (Logan, 2009).

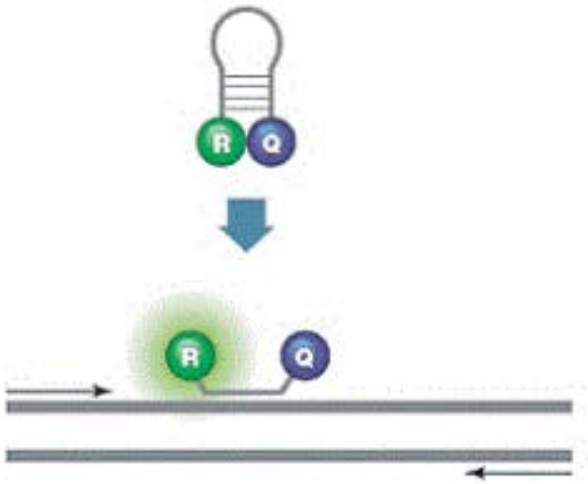

Figure 17. Molecular beacon probes (Bio-Rad.com).

These probes are designed to have the ends self-complementary, so that the molecule assumes a hairpin conformation (Figure 17).

At each end, a fluorophore and its inhibitor are bound. The signal is emitted when the probe binds to the target sequence. In fact, with the distension of the hairpin, the fluorophore moves away from the inhibitor and is therefore able to emit light (Logan, 2009).

The information obtained with this technique is therefore greater than that of a classic PCR. Real-time PCR is often used in combination with the retro-transcription reaction (RT), to quantify the levels of expression of specific genes of interest, and to genotype SNP's (Bustin, 2004).

1.3.2 Melting Temperature Analysis

At the end of the PCR reaction melting curves can be constructed. These curves describe the fluorescence variation as a function of temperature.

Each PCR amplicon (a specific PCR product) has its characteristic melting curve, it is obtained at the end of the amplification cycle reaction by increasing the temperature from 50 to 95°C at a continuous rate (usually between 0.2-0.5ºC7second) and carrying out the fluorescence reading at the fastest rate allowed by the equipment (Bustin, 2004).

As the temperature increases, little decrease in fluorescence is observed until the melting temperature is reached. At this point the double stranded amplicons will denature with a consequent sudden decrease in the detected fluorescence.

Melting curve analysis is a practical and useful control of the real-time PCR amplification specificity. The dynamic control of the entire melting curve as the temperature changes completely defines the melting dynamic (S. Kennedy et al., 2001).

There are two ways of displaying melting curves: the first (Figure 18) consists in visualizing the fluorescence curve as a temperature function; the second one (Figure 19) represents the fluorescence negative derivative in respect to temperature $-dF/dT$ (the fluorescence unit is the RFU "relative fluorescence units") (Kennedy et al., 2001).

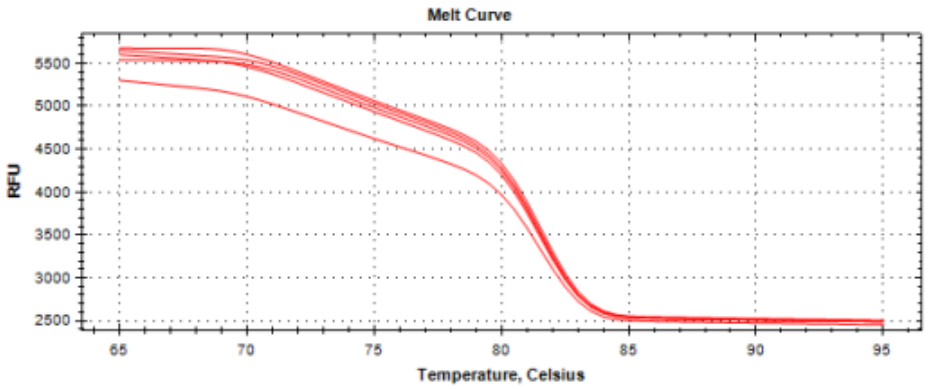

Figure 18. A set of samples melting curves evaluated for the same target in a real-time PCR experiment (MolecularLab.it)

Figure 19 gives us a complete view on the fusion dynamics. The second inflection point of Figure 18 corresponds to peak of the Figure 19.

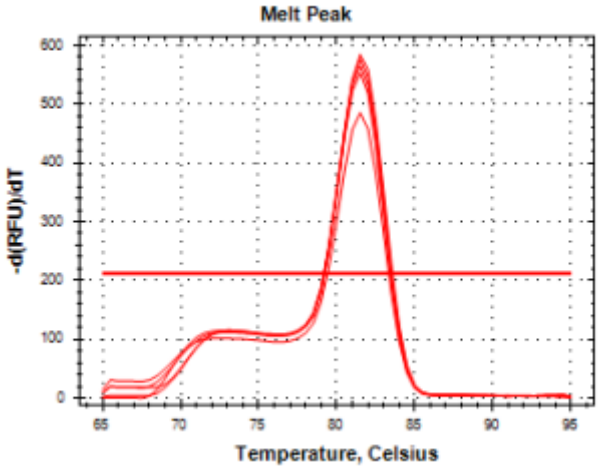

Figure 19. A set of samples melting curves in negative derivate form evaluated for the same target in a real-time PCR experiment (MolecularLab.it)

Melting curve profiles provide us with a tool to evaluate the specificity of the amplification. A specific PCR amplified (usually in length around one to five hundred base pairs) has a melting temperature Tm>80°C; the curve peak will therefore be in correspondence with this values range. Primer dimers, a-non-specific product due to the link between partially complementary primers, giving shorter amplicons, denature at a temperature Tm <80°C; their presence is therefore identified in the curve by a peak at temperatures below 80°C. Other non-specific amplification products have different melting profiles than the amplified sample, usually there are more peaks at different temperatures (Bustin, 2004).

1.4. Restriction Fragment Length Polymorphism

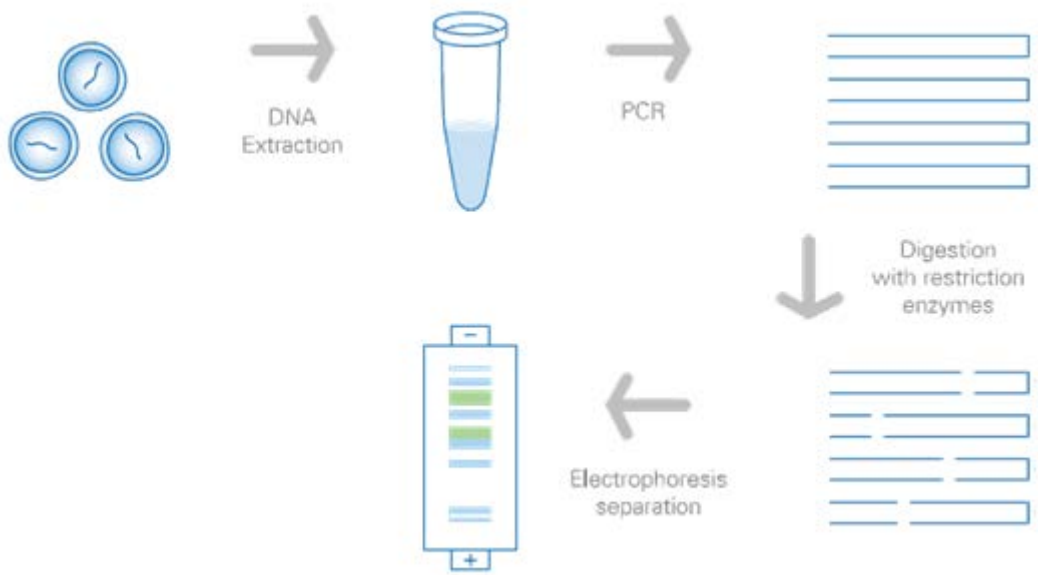

Figure 20. PCR-RLFP sequence of steps (adapted from www.thermofisher.com)

The RFLP technique is based on DNA hydrolysis with restriction endonucleases followed by the separation of the fragments generated by electrophoresis.

Restriction enzymes break down DNA sequences by recognizing specific nucleotide sequences, thus differentially identifying DNA fragments that include areas with polymorphisms/mutations (Scialpi, 2008).

The RFLP technique can be applied after the PCR (which amplifies the specific nucleic acid fragment of interest) and in this way direct the analysis to one particular polymorphism of interest (Figure 20) (S. Kennedy et al., 2001).

1.5. Agarose gel electrophoresis

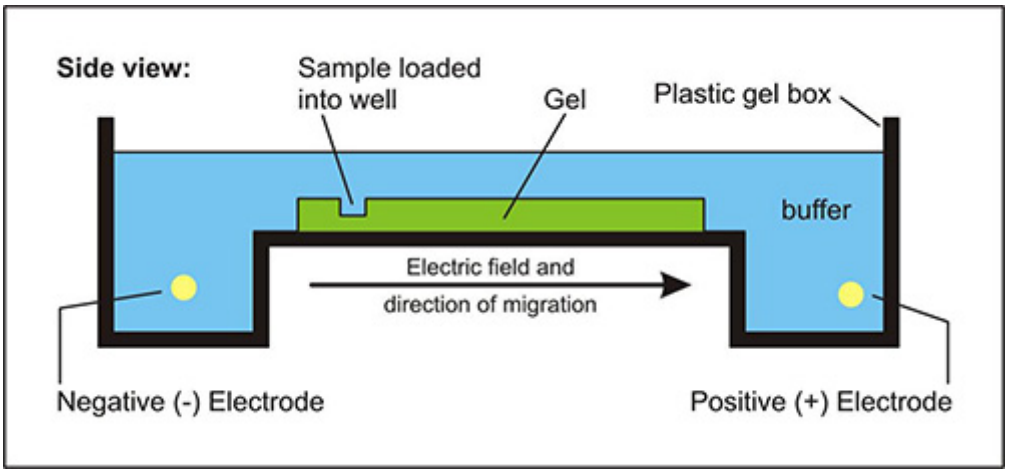

Figure 21. A schematic example of an agarose gel electrophoresis apparatus (orbitbiotech.com)

Agarose gel electrophoresis allows the separation of DNA fragments. The DNA is negatively charged by the presence of the terminal phosphate groups of the DNA sequence. Applying an electric field, DNA migrates towards the positive pole (anode) (Scialpi, 2008).

The migration speed depends on (Bustin, 2004):

• DNA size. The DNA molecules are separated by size in such a pattern that the distance travelled is inversely proportional to the logarithm of its molecular weight. So, the larger molecules migrate more slowly, the smaller ones faster;

• agarose concentration in the gel. Agarose is a linear and neutral polysaccharide formed by units of D-galactose and 3,6-anhydro-L-galactose alternately bonded with glycosidic bonds. It is a water-soluble sugar at boiling point, while it becomes solid as it cools, forming a matrix through hydrogen bonds between the linear chains. During gelling, the agarose polymers associate in a non-covalent way to form a network of bundles, in which the pores formed determine the molecular properties of the sieving gel. The size of the pores depends on the agarose concentration.

• DNA conformation. The superwired, linear and circular DNA have different migration speeds even if of equal size: The superwired form runs faster because it is more compact; The circular shape runs slower because it is the most "cumbersome" and it is more difficult to move inside the pores of the gel; The linear form is placed in half (the linear form is, for example, the one that is found as a product in the PCR).

• applied voltage, about 5Volt/cm (anode-cathode distance) is usually applied to prevent overheating or excessively slow migration.

• presence of ethidium bromide or SYBRGreen, which are fluorescent intercalating dye that allow visualization of DNA. When exposed to UV light, the electrons of the aromatic ring of these fluorophores are activated, which leads to the release of energy (light) when the electrons return to the ground state. They work by intercalating in the DNA molecule in a concentration dependent manner. This allows an estimate of the amount of DNA based on its intensity. Because of its positive charge, the use of EtBr reduces the rate of DNA migration by 15%. EtBr is a suspected mutagen and carcinogen, so care should be taken when handling agarose gel containing it.

• composition (ionic strength) of the buffer. The lowering strength the higher the resistance and eat produced and the slower the migration (Bustin, 2004).

To separate the DNA using agarose gel electrophoresis, the DNA is loaded into prefabricated wells in the gel and an electric field is applied (Figure 21). The DNA phosphate skeleton (and RNA) is negatively charged, so when placed in an electric field, the DNA fragments will migrate to the anode (Scialpi, 2008).

The DNA samples are mixed with a dense dye mix (Gel loading dye), before being loaded into the wells, this increase density allowing the deposition in the well. The dye also allows indication of the electrophoretic front (Bustin, 2004).

A size marker is also used which is made up of DNA fragments of known dimensions and allows to determine the size of the sample DNA. This is done by migrating along with the DNA samples, it is usually loaded into the first well, and is used as a dimensional reference.

Once the electrophoretic run is complete, the results are displayed using a UV lamp (Scialpi, 2008).

1.6. Objectives of the study

In the present study we considered two SNPs, rs5743899 of the TOLLIP gene and rs16944 of the gene encoding Interleukin 1β (IL-1β). Being the two genes involved in the body's immune response against external pathogens, it is hypothesized that these two SNPs are involved in susceptibility to RHV viral infections.

For over a year, our group has characterized the presence of Rhinovirus in the nostrils of healthy volunteers using RT-qPCR.

In the present study we genotyped these immune-related SNPs and correlated them to the frequency of detection and relative quantification of Rhinovirus in the volunteers' nostrils.

2. MATERIAL AND METHODS

2.1 Population studied

In the study a total of 89 volunteers were enrolled, recruited from among the student population of the Faculty of Health Sciences of the Fernando Pessoa University in Porto, Portugal.

The volunteers were aged between 20 and 41 years (Figure 22), the gender distribution (females account for 2/3 of the volunteers) is shown in Table 1, and other characteristics are shown in Table 2.

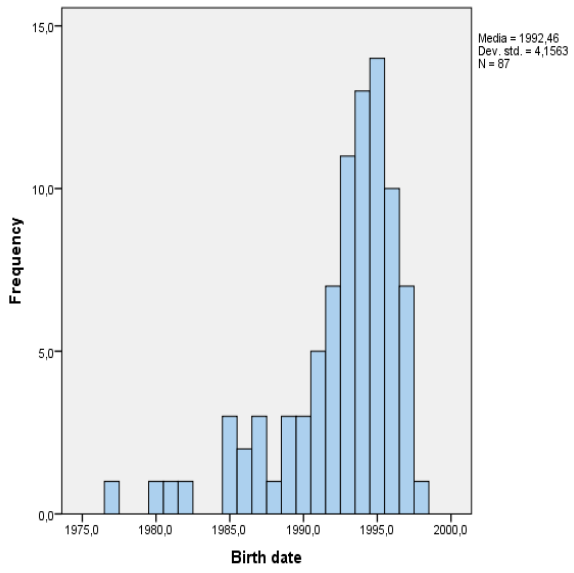

Figure 22. Sample distribution by age

Table 1. Sample distribution by gender

Gender	N	%
Male	24	26,7%
Female	64	73,3%

Table 2. Population characteristics

Self-Reported Allergies	%
Chemicals	4,4%
Contact	13,3%
Insects	7,8%
Foods	13,3%
Other	15,8%
Smoker subjects	34,4%
Asthmatic subjects	8,9%

Each student provided a monthly swab sample collected for the characterization of the year-long presence of rhinovirus in the nostrils, and a blood sample or mouth endothelial cell cotton swab (if unwilling to donate blood) for DNA extraction and purification. A total of 33 samples of blood and 15 samples of mouth epithelial cell cotton swab were used in the present study.

2.2 DNA extraction

Genomic DNA was extracted from blood or epithelial mouth cells using Qiamp mini Blood DNA kit (Qiagen) according to the manufacturers instruction.

The principle on which this kit extraction is based depends on DNA binding to silica particles. These particles were contained in columns and allow DNA binding and elimination of RNA and proteins by means of a series of washes with appropriate buffers. The DNA bound to the filter can then be eluted by water or a Tris-HCl and EDTA based buffer.

To perform the extraction, 200µl of whole blood at room temperature, and 200µl of lysis buffer were added to 20µl of proteinase K (600 IU/ml). Subsequently, the sample was incubated at 56°C for 10 minutes, after which, 200µl of absolute ethanol (96-100%) were added to the sample and, the mixture transferred to the silica spin column. The column was then centrifuged and washed with two different buffers of increasing astringency.

Finally, the genomic DNA was eluted with 200µl of nuclease free water or with a Tris EDTA based buffer and stored at -20°C until the subsequent analytical procedures were performed.

2.3 Detection of the rs5743899 (TOLLIP) by PCR-RFLP

The polymorphism rs5743899 of the TOLLIP gene present in the intron and exon 4 was typed by PCR-RFLP using the restriction enzyme HhaI, essentially as previously described (Araujo, Silva, Mesquita, et al., 2015).

A previously described primer pair (table 3) was used to amplify the regions encompassing the polymorphism. PCR was performed in a final volume of 30µL containing 1x IQ SybrGreen Supermix (BioRad, USA), 1µM forward and reverse primers and 1µL of blood sample DNA.

Table 3. Primer pair used for TOLLIP polymorphism study

Gene	Primer name	Primer sequence[a)]
TOLLIP	rs5743899F	5'-GGC AAT GGC AGT GGC CAC CAG TGA-3'
	rs5743899R	5'-CCG ATG CCC GCA CAC CTG TGT GAT-3'

a) Araujo, Silva, Mesquita, et al., 2015

The PCR cycling conditions were the subject of extensive optimization varying step temperatures and duration. The optimum cycling conditions were determined as including an initial denaturation step of 3 min at 98ºC, followed by 35 cycles of denaturation (30s at 98ºC) and annealing/amplification (45s at 72ºC). The generated PCR amplicons showed the expected length of 279bp (figure 23).

A total volume of 15uL of PCR products was digested with 20 IU of the restriction enzyme HhaI (New England Biolabs, USA), in CutSmart buffer (New England Biolabs, USA) at 37°C for 2 hours.

PCR restriction fragments were size separated by electrophoresis in 4% NuSieve 3:1 TAE agarose gel (FMC, USA). Presence of the G allele was detected by fragments of 125bp, 93bp and 61bp whereas the A allele generated bands of 218bp and 61bp (figure 23).

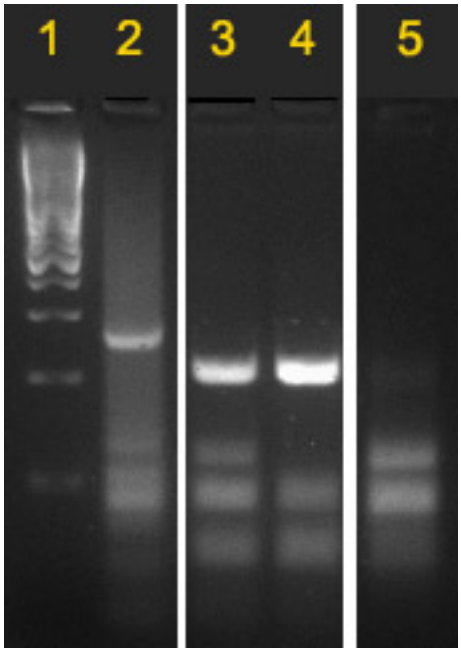

Figure 23. Electrophoresis with example of TOLLIP PCR amplicon and Restriction Fragments of each genotype. Well 1 shows an 100bp ladder; well 2 shows the PCR amplicon; well 3 shows a Heterozygous individual; well 4 represents results for a homozygous A allele; well 5 indicates an homozygous G genotype result.

2.4 Detection of the rs16944 (IL1B) by PCR-RFLP

The polymorphism rs16944 of the Interleukin-1 beta (-511) was determined using polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP), essentially as previously described (Bhat, Naykoo, Quasim, et al., 2014). Amplification of the target region was carried out by polymerase chain reaction using the specific forward and reverse primers shown in Table 4.

Table 4. Primer pair used for IL1B polymorphism study

Gene	Primer name	Primer sequence[a)]
IL-1B	rs16944F	5’-TGG CAT TGA TCT GGT TCA TC-3’
	rs16944R	5’-GTT TAG GAA TCT TCC CAC TT-3’

a) Bhat, Naykoo, Quasim, et al., 2014

PCR was performed in a final volume of 30µL containing 1x Mix quantitec iProof High-Fidelity Master Mix (Bio-Rad 172-5310), 1µM forward and reverse primers and 1µL of blood sample DNA or 13 µL of mouth swab samples DNA.

The PCR cycling conditions were extensively optimized. Optimum conditions were determined to include an initial denaturing step of 3 min at 98°C, then 45 cycles of denaturation (30s at 98°C), annealing (30s at 60°C) and extension (30s at 72°C).

The generated PCR amplicons showed the expected size of 305bp (figure 24).

Digestion of 10µL the amplyfied products of IL1B (-511) was done by using 10IU of restriction endonuclease Ava1 (New England Biolabs, USA), in CutSmart buffer (New England Biolabs, USA) and incubated at 37°C for 2 hours.

PCR restriction fragments were size separated by electrophoresis in 4% NuSieve 3:1 agaroses gel (FMC, USA). Presence of the T allele was detected by fragments of 305bp, whereas the C allele generated bands of 190bp and 115bp (figure 24).

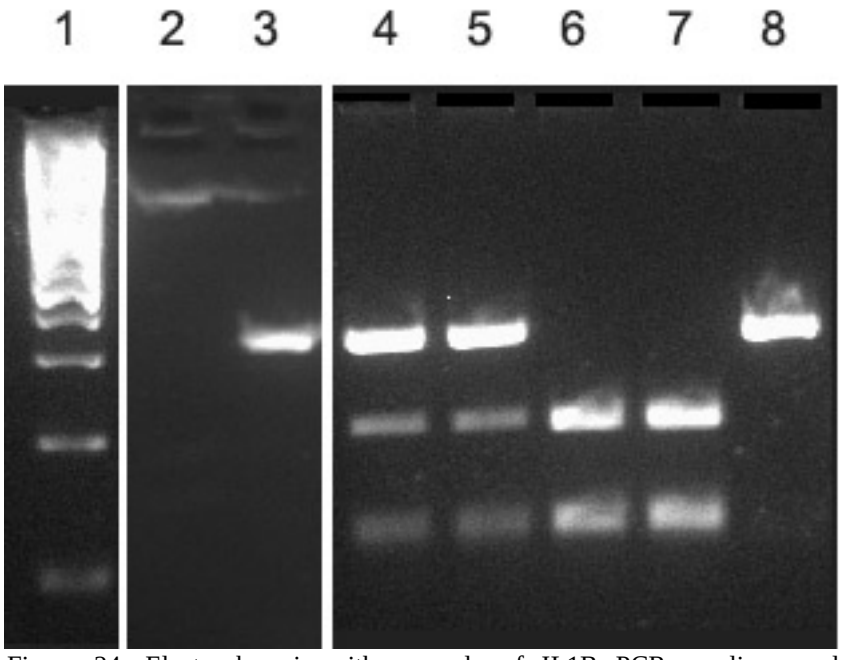

Figure 24. Electrophoresis with example of IL1B PCR amplicon and Restriction Fragments of each genotype. Well 1 shows an 100bp ladder; wells 2 and 3 show the PCR amplicon of a negative control and a positive DNA sample; wells 4 and 5 show results of heterozygous individuals; wells 6 and 7 represent results of homozygous C allele individuals; well 8 indicates an homozygous T genotype result.

2.5 Nasal Sample Collection

Voluntaries came monthly for nasal swab collection. Samples were collected from each nostril using sterile cotton swabs. Swabs from left and right nostrils were stored together in 200mL of RNALatter (Invitrogen) solution for 24h at 4ºC and then transferred to a deep freezer (-70ºC) until RNA extraction procedures. Nasal sample collection was never performed by the author of the present thesis.

Of the 89 voluntaries enrolled in the study, a total of 56 remained actively engaged in the study at the end of the 12 month observation period (major reason for dropping the study was discontinuity of attendance to University campus due to end of studies program).

Most of the voluntaries (62%) donated 8 to 12 nasal swab samples (figure 25).

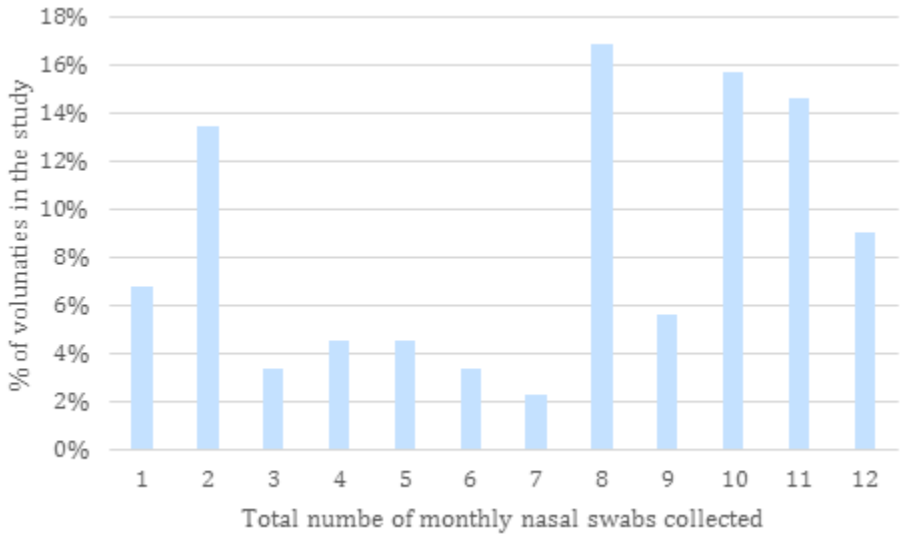

Figure 25. Distribution of the number of collected nasal swab samples per voluntary

2.6 Viral RNA extraction

Viral RNA extraction was performed using QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer instructions. Final elution volume was 60µL. RNA samples were stored at -80ºC. Viral RNA extraction was never performed by the author of the present thesis.

2.7 Rhinovirus RNA detection and quantification

Rhinovirus RNA amplification was performed by qPCR on a Lightcycler 1.1 (Roche, Germany) as previously described (Myatt, et al.2004). Semi-quantification was obtained relative to a dilution series of a positive sample. Rhinovirus qPCR was never performed by the author of the present thesis.

2.8 Statistical analysis

Data was statistically analysed using Statistical Package for Social Sciences (SPSS) version 22. Variable normality test was done using the Kolmogoronov-Smirnov test against a gaussian distribution. Comparisons among genotype groups were performed using the Kruskal-Wallis H test, whereas comparisons among allele groups were done using the Mann-Whitney U test. When differences between groups were expected to fall in the extremities of the observed values, group comparisons were also done using the Moses extreme reactions test. Tests were always done to compare ranks and results were always interpreted using a significance level of 0,05.

3. RESULTS

3.1 Allele frequencies, Genotype frequencies and Hardy-Weinberg equilibrium

A total of 33 DNA samples were used for the genotyping of TOLLIP (all samples extracted from blood as the TOLLIP PCR was found too inefficient for the very diluted DNA obtained from mouth epithelial cotton swabs) and 48 DNA samples were used for the genotyping of interleukin 1 beta (33 blood samples and 15 mouth epithelial cotton swab samples).

Allele and genotypic IL1B frequencies showed that, as expected, the most present genotype was TC (50%; table 5), whereas the T (wild type) allele was the most frequent in the studied population (65%; table 5).

The observed genotype frequencies of the TOLLIP polymorphism were notoriously different from the IL1B, as the heterozygous genotype was not the most abundant one. In fact, The AA genotype was dominant (42,4%; table 5) and the GG genotype was the least frequent with only 24,2% (table 5). However, allele frequencies were not very different from the observed in IL1B with the A allele being the most frequent (59%; table 5). The observed genotypes for the population were however not significantly different from the expected ones if the populations were in Hardy-Weinberg equilibrium. Thus, the observed deviations were as expected for such a small population.

Table 5. Genotypes and allele frequencies for TOLLIP and IL1B

Genotypes frequencies				Allele frequencies			
IL1B		TOLLIP		IL1B		TOLLIP	
TT	39,6%	AA	42,4%	T	65,0%	A	59,0%
TC	50,0%	AG	33,3%	C	35,0%	G	41,0%
CC	10,4%	GG	24,2%				

3.2 Variables Normality Check

First, we tested whether a gaussian distribution could be used to describe the RHV determinations (frequency of positive samples, average RHV positive sample concentration, number of RHV+ episodes, duration of RHV+ episodes).

As expected, distributions of rhinovirus determinations were found to be significantly different ($p<0.001$) from the gaussian distribution (see example in Figure 26). For this reason, non-parametric statistical tests were used in the next stages of data analysis.

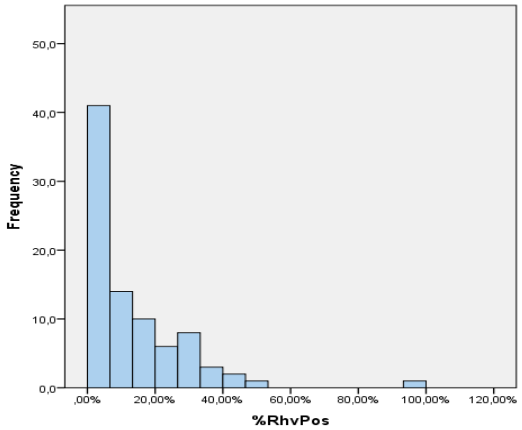

Figure 26. Rhinoviruses frequencies distribution

3.2 IL1B and TOLLIP genotype and frequency of rhinovirus detected in nostrils of volunteers

Next, we evaluated if IL1B or TOLLIP genotype groups showed differences in the frequency of nostril Rhinovirus detection. As can be seen in figures 27 and 28, there was a wide variation in the frequency of RHV positive nostril samples in all genotype groups, and no statistically significant differences between genotype groups were found (table 6).

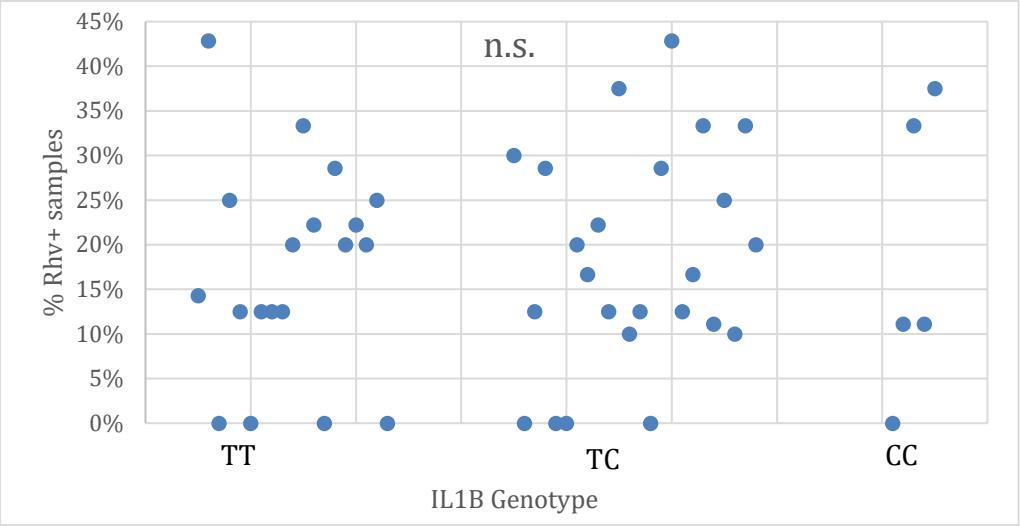

Figure 27. Percentage of RHV+ samples as a function of IL1B genotype. No significant differences (Kruskall-Wallis test) between genotype groups were observed.

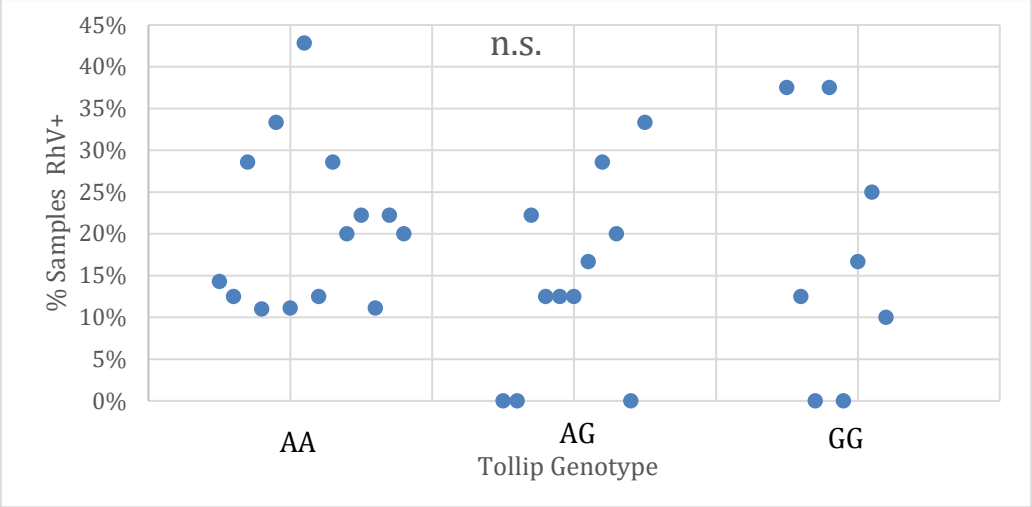

Figure 28. Percentage of RHV+ samples as a function of TOLLIP genotype. No significant differences (Kruskall-Wallis test) between genotype groups were observed.

However, as seen in figure 28 and table 6, TOLLIP genotype AA was the only one with no individuals showing year-long absence of rhinovirus (0% positive samples). Nevertheless, the low number of such individuals (5) is not sufficient for statistical significance. No such effect was observed for the IL1B genotypes (figure 27) as similar frequencies of year-long RHV free individuals were present in each genotype group (table 6).

Table 6. Percentage of year-long RHV free individuals per genotype groups

	TOLLIP			IL1B		
	AA	AG	GG	TT	TC	CC
Year long RHV-free individuals	0/14 (0%)	3/11 (27%)	2/8 (25%)	4/18 (22%)	4/24 (17%)	1/5 (20%)
p	n.s.			n.s.		

n.s.=not significant (Kruskal-Wallis independent samples test)

3.3 IL1B and TOLLIP genotype and concentration of rhinovirus in the nostrils of volunteers

Taking advantage of the quantitative nature of the qPCR, we also quantified the rhinovirus titre in the nostril swab samples. To that effect, a tenfold serial dilution of a positive control was used as a quantification standard, with results expressed as relative to the original control concentration.

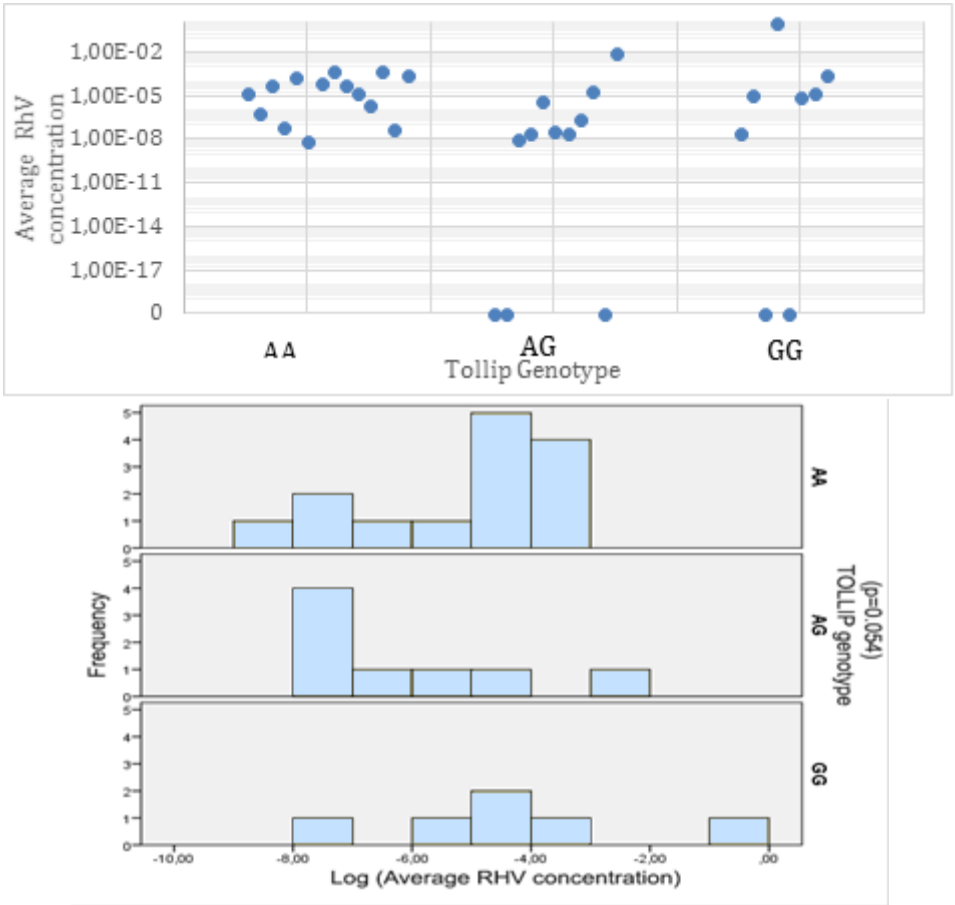

Figure 29. Correlation between TOLLIP genotype and rhinovirus concentration

The rhinovirus concentration detected in the volunteers' nostrils ranged from a minimum of zero (negative samples) to a maximum of $9{,}95x10^{-1}$ (equivalent to the positive control). Observed rhinovirus nostril concentrations in each genotype group are shown in figures 29 (TOLLIP genotypes) and 30 (IL1B genotypes).

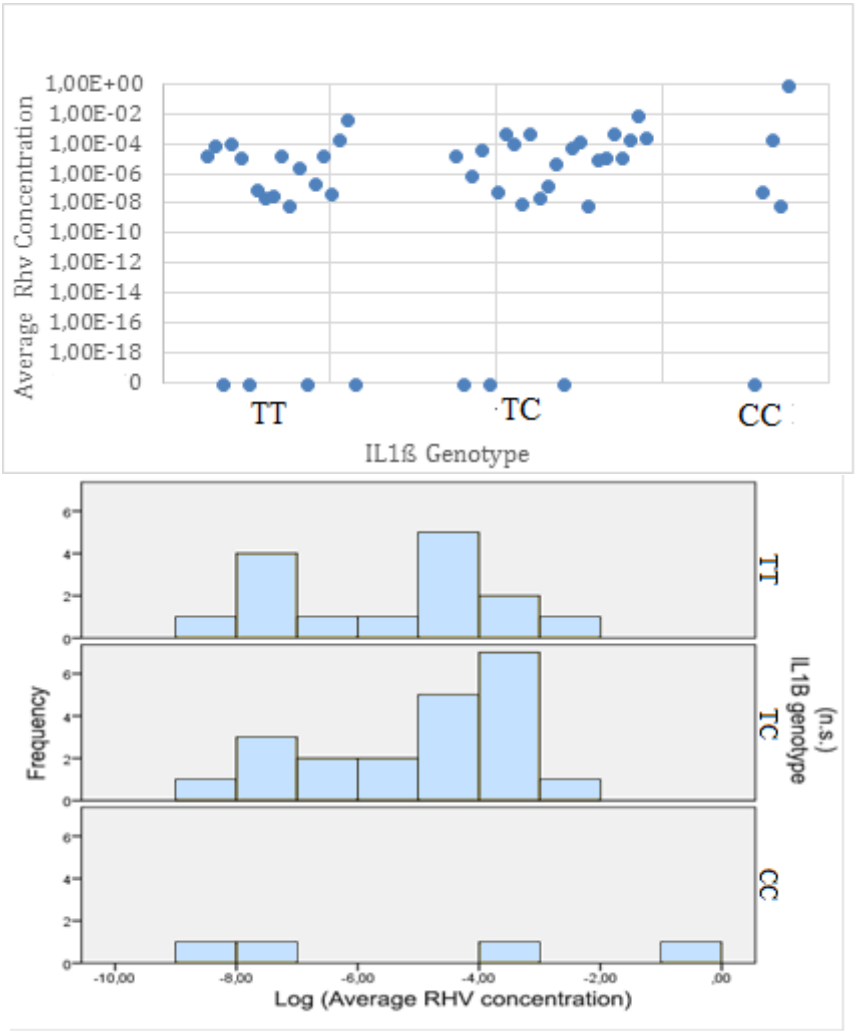

Figure 30. Correlation between IL1B genotype and rhinovirus concentration (n.s. = not significant)

As can be seen in figure 30, no IL1B genotype effects on the relative RHV concentration were observed. However, albeit not reaching statistical significance, a tendency (p=0.054) for TOLLIP genotypes to influence viral titres was found (Figure 29). Apparently, individuals with the homozygous G genotype have a tendency to have higher viral titres, in positive samples (and/or to show less frequently very low viral titres in positive samples).

3.4 IL1B and TOLLIP genotypes and the number and duration of nasal RHV positive episodes

As mentioned in the introduction (section 1.1.3) usually Rhinovirus incubation and cold disease duration allow the healthy human body to clear viruses in less than a month. Thus, two consecutive positive samples may either indicate a reinfection, or be the result of an immune system less capable of clearing the virus. In an attempt to evaluate whether immune system related genetic polymorphisms could support the latter hypothesis, we tested whether considering consecutive positive samples as abnormally long but single episodes would correlate to any particular genotype.

As can be seen in table 7, no significant differences were observed among genotype groups in relation to the frequency of long (>1 month) RHV positive episodes.

Table 7. Frequency of voluntaries with long RHV+ episodes per genotype groups

	TOLLIP			IL1B		
	AA	AG	GG	TT	TC	CC
Voluntaries with episodes lasting >1 month	4/14 (29%)	3/11 (27%)	2/8 (25%)	5/19 (26%)	7/24 (29%)	2/5 (40%)
P	n.s.			n.s.		

n.s.=not significant

However, close analysis of data (figure 31) reveals that TOLLIP AA genotype simultaneously shows the longest episode (3 months) and absence of RHV free individuals.

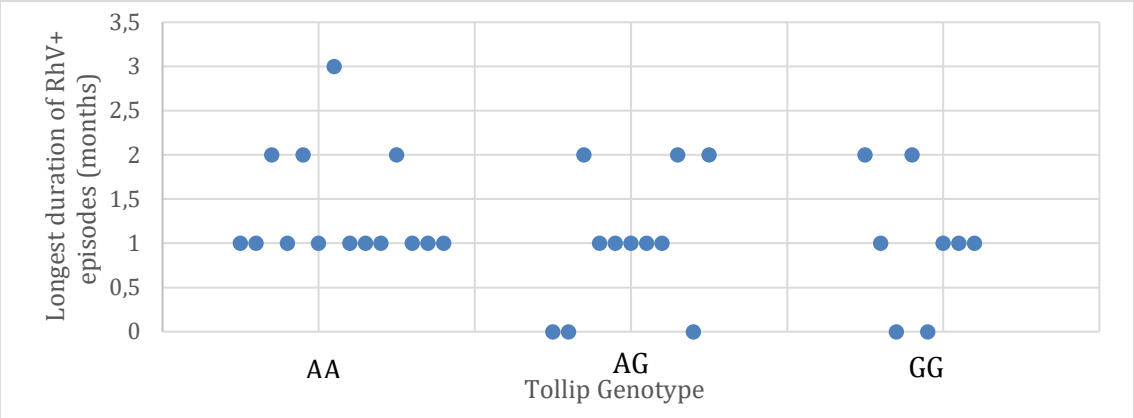

Figure 31. Duration of RHV+ episodes according to TOLLIP rs5743899 genotype

Similarly, the IL1B TT genotype was the only one with episodes lasting 3 months, but no similar association to the year-long RHV-free voluntaries (0-month longest episode) was found (figure 32).

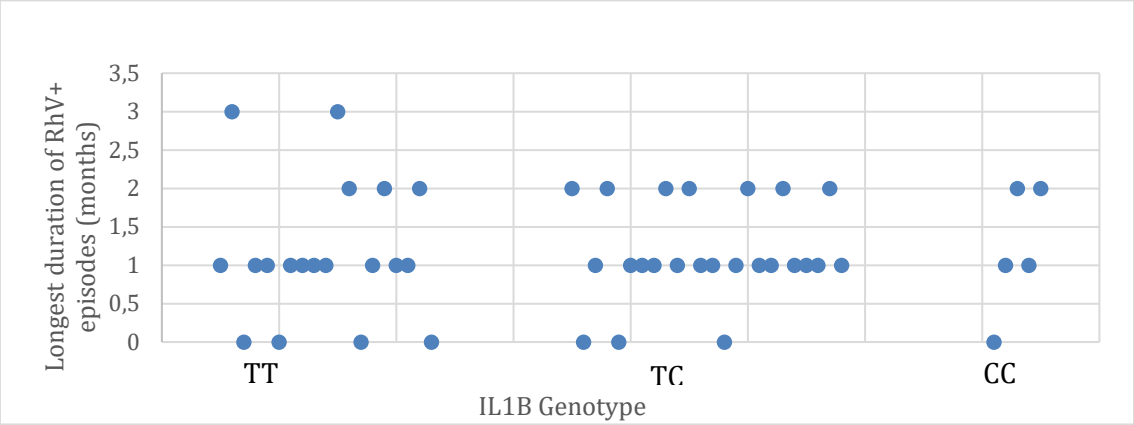

Figure 32. Duration of RHV+ episodes according to IL1B rs16944 genotype

Next, we tested genotype influence on total number of RHV+ episodes (if consecutive positive samples were assumed to be one infection/colonization episode). As expected in an adult population, voluntaries were found to have only up to 2 rhinovirus positive episodes. As can be seen on table 8 no statistically significant differences were observed between the genotype groups relative to the frequency of RHV reinfections.

Table 8. Frequency of voluntaries with more than one RHV+ episode per genotype group

	TOLLIP			IL1B		
	AA	AG	GG	TT	TC	CC
Voluntaries with >1 RHV+ episode	4/14 (29%)	2/11 (18%)	3/8 (38%)	5/18 (28%)	8/24 (33%)	2/5 (40%)
P	n.s.			n.s.		

n.s.=not significant

3.5 individual IL1B and TOLLIP allele presence and rhinovirus detection in the nostrils of volunteers

Since genotype analyses indicated that some alleles could have protective or risk effects on the rhinovirus presence (see 3.2 and 3.4) we decided to compare the detection of RHV in the nostrils of volunteers according to the individual presence of each studied allele.

Table 9. Percentage of year-long RHV free individuals as function of TOLLIP Allele groups

	TOLLIP rs5743899			
	A allele [a]		G allele	
	A+	A-	G+	G-
Year-long RHV-free individuals	3/25 (12%)	2/8 (25%)	5/19 (25%)	0/14 (0%)
Kruskal-Wallis[b]	n.s.		p<0.040	

a) + indicates presence of allele; - indicates absence of allele
b) n.s. = not significant

The results shown in tables 9 and 10 indicated that TOLLIP G allele could indeed have a protective effect as only individuals with this allele were found to be free of RHV for the entire observation period (one year). No other allele showed a similar difference (Tables 9 and 10).

Table 10. Percentage of year-long RHV free individuals as function of IL1B Allele groups

	IL1B rs16944			
	T allele [a]		C allele	
	T^+	T^-	C^+	C^-
Year-long RHV-free Individuals	8/43 (24%)	1/5 (20%)	5/29 (17%)	4/19 (21%)
p [b]	n.s.		n.s.	

a) + indicates presence of allele; - indicates absence of allele
b) n.s. = not significant

3.6 individual IL1B and TOLLIP allele presence and rhinovirus concentration in the nostrils of volunteers

Next, we tested if the studied alleles showed any influence on the RHV titres in the nostrils of the voluntaries. As can be seen in figure 33 and 34, no statistically significant effect was observed of any allele on the RHV positive samples nasal titre for the A Allele. However, the grouping of the negative samples on the TOLLIP G+ allele group, together with the presence of the highest viral concentrations on positive samples renders the presence of the G allele as giving rise to extreme values. In fact, voluntaries with this allele are the only ones that seem to resist RHV for the entire period, but when infected show the highest nostril concentrations of rhinovirus (Man Whitney U test p=0.031; Moses test p=0.020; figure 33).

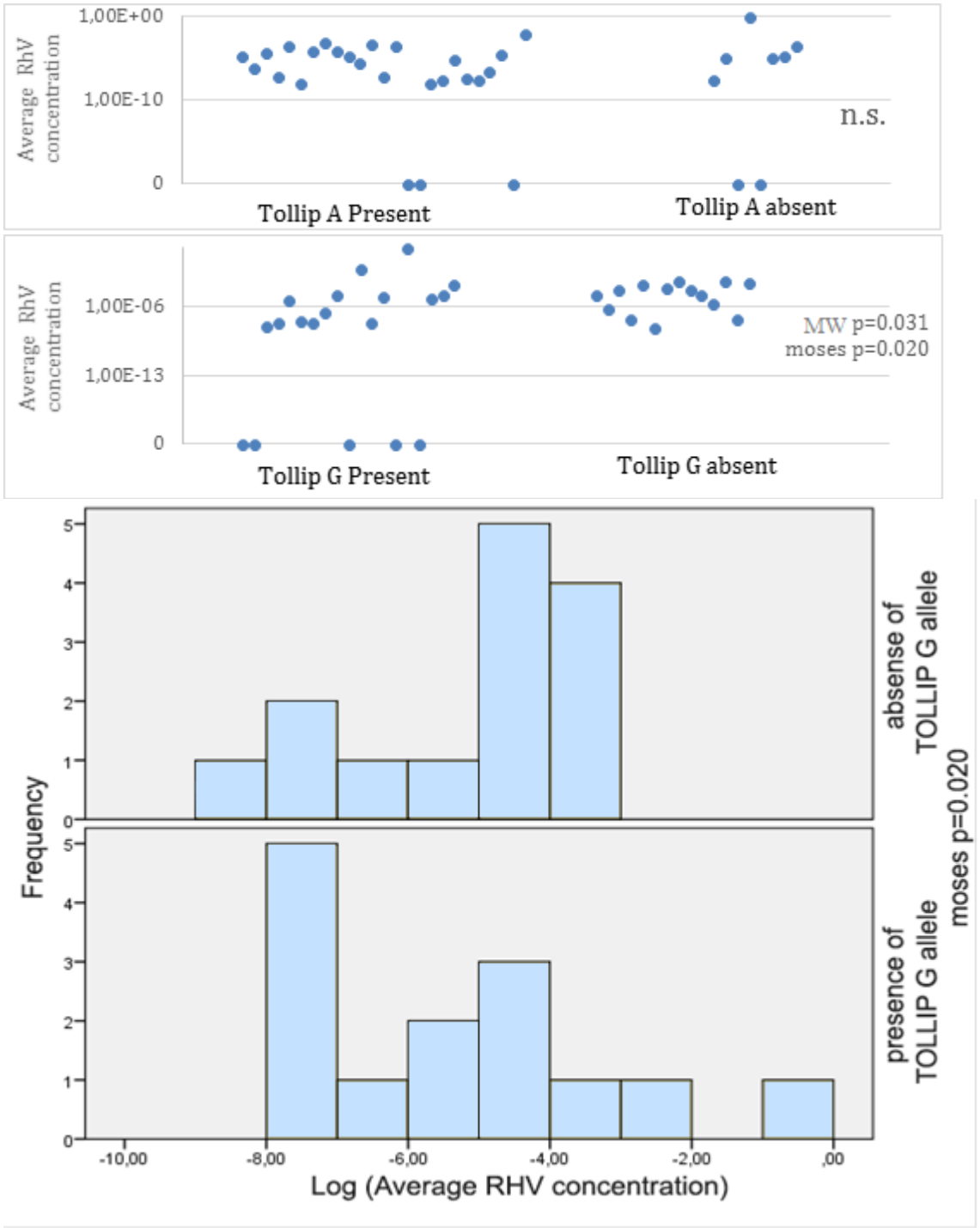

Figure 33. Rhinovirus concentration as a function of the presence of specific TOLLIP alleles (KW=Kruskal-Wallis test; moses = moses test for extreme reactions; n.s.= not significant)

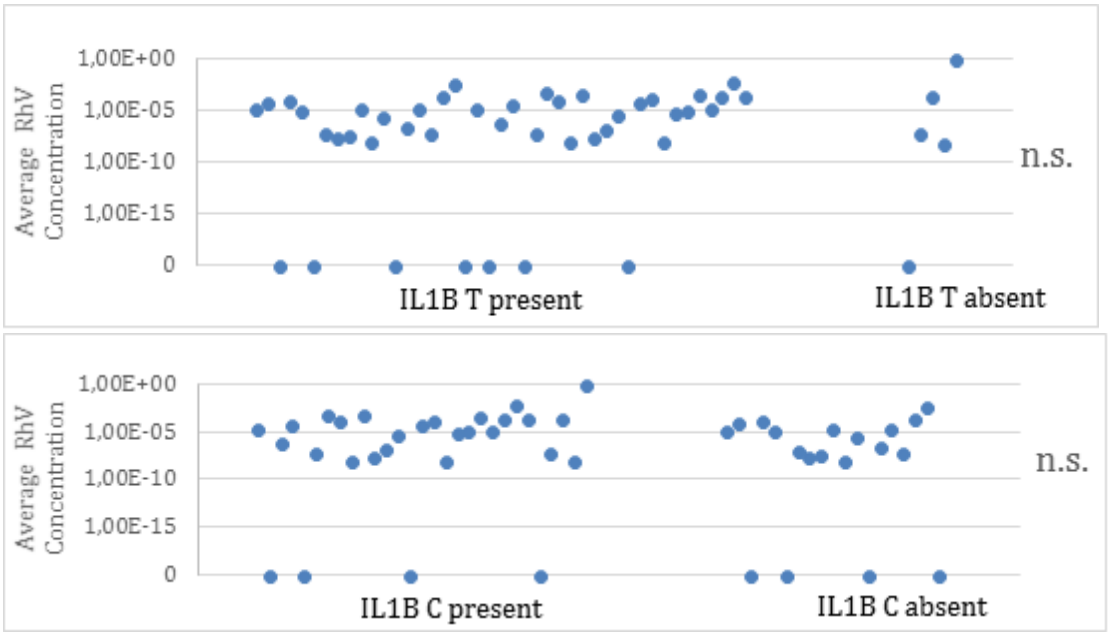

Figure 34. Rhinovirus concentration as a function of the presence of specific IL1B alleles (n.s.= not significant)

3.7 IL1B and TOLLIP alleles and number and duration of RHV$^+$ episodes

As can be seen on figure 35, the TOLLIP G allele did correlate with extreme values for the number of episodes ($p<0.001$). Presence of the G allele was correlated with shorter (or even total absence of) RHV$^+$ episodes, whereas the absence of this allele correlated to extreme lengthy RHV+ episodes.

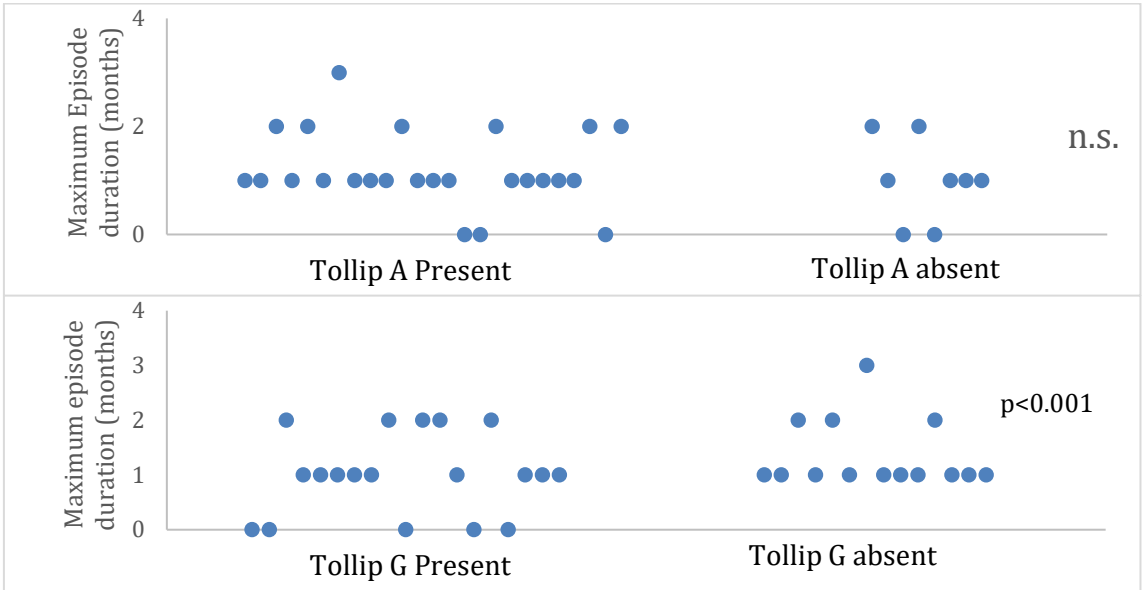

Figure 35. RHV+ episode duration as a function of the presence of specific TOLLIP Alleles. Statistical significant differences shown are for the Moses test for extreme reactions (n.s.= not significant)

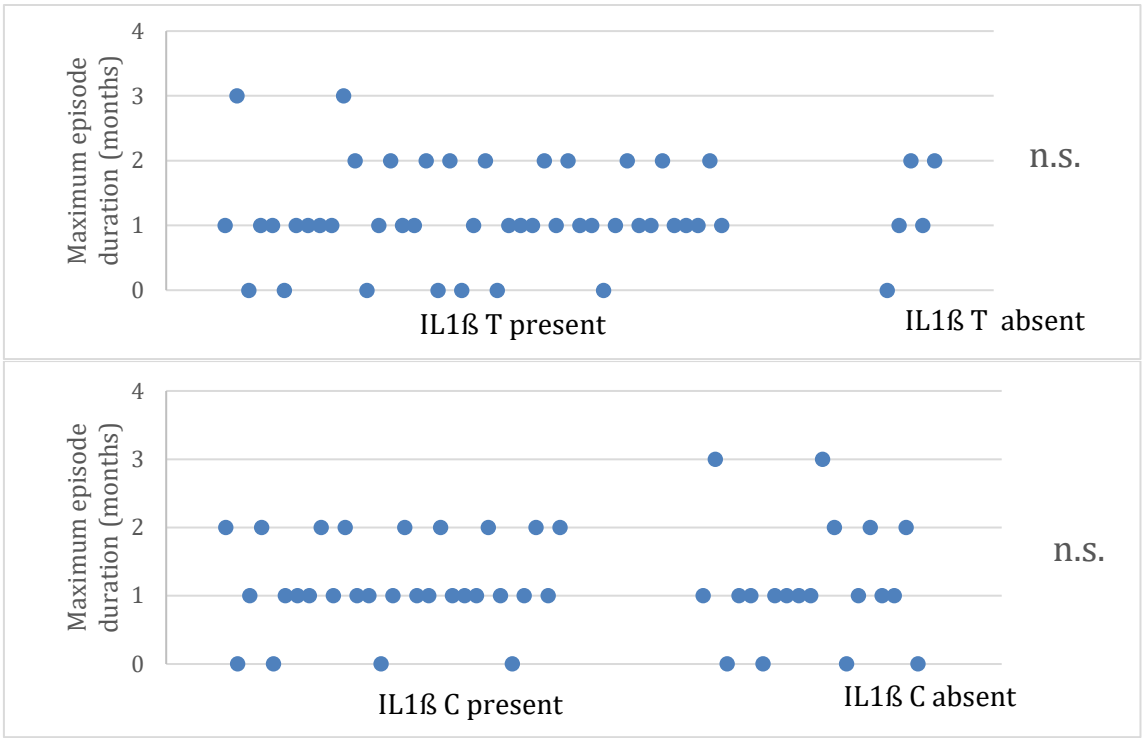

Figure 36. RHV+ episode duration as a function of the presence of specific IL1B Alleles. Statistical significant differences shown are for the Moses test for extreme reactions (n.s.= not significant)

No such difference was observed for the TOLLIP A allele (Figure 35) or any of the IL1B alleles (Figure 36)

Similarly, the TOLLIP G allele was the only one where absence of RHV+ episodes across the entire observation period was ever found ($p<0.001$; figure 37). No other difference was found between allele groups regarding the number of RHV+ episodes (Figures 37 and 38).

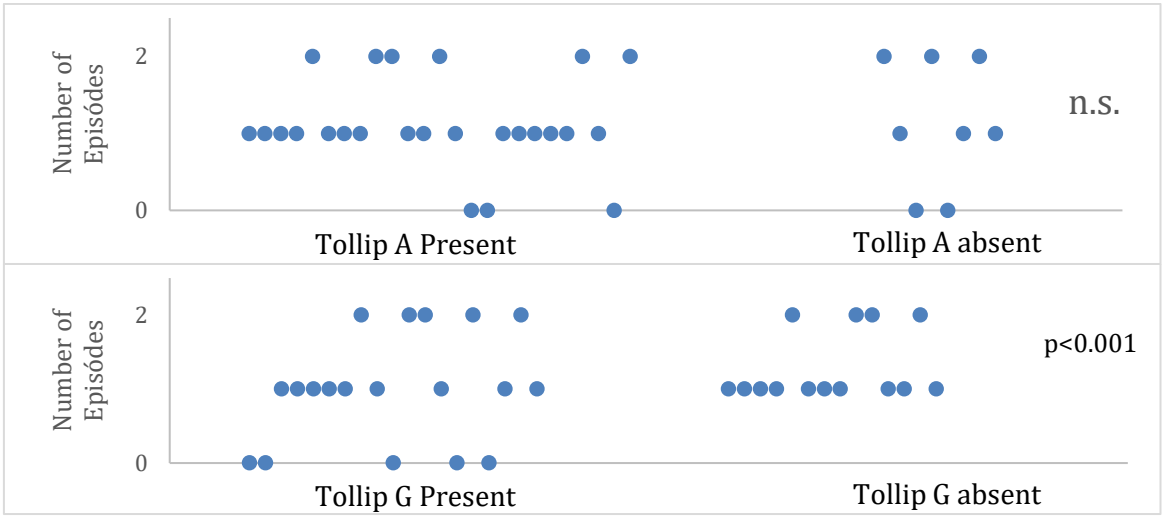

Figure 37. Number of RHV+ episode as a function of the presence of specific TOLLIP Alleles. Statistical significant differences shown are for the Moses test for extreme reactions (n.s.= not significant)

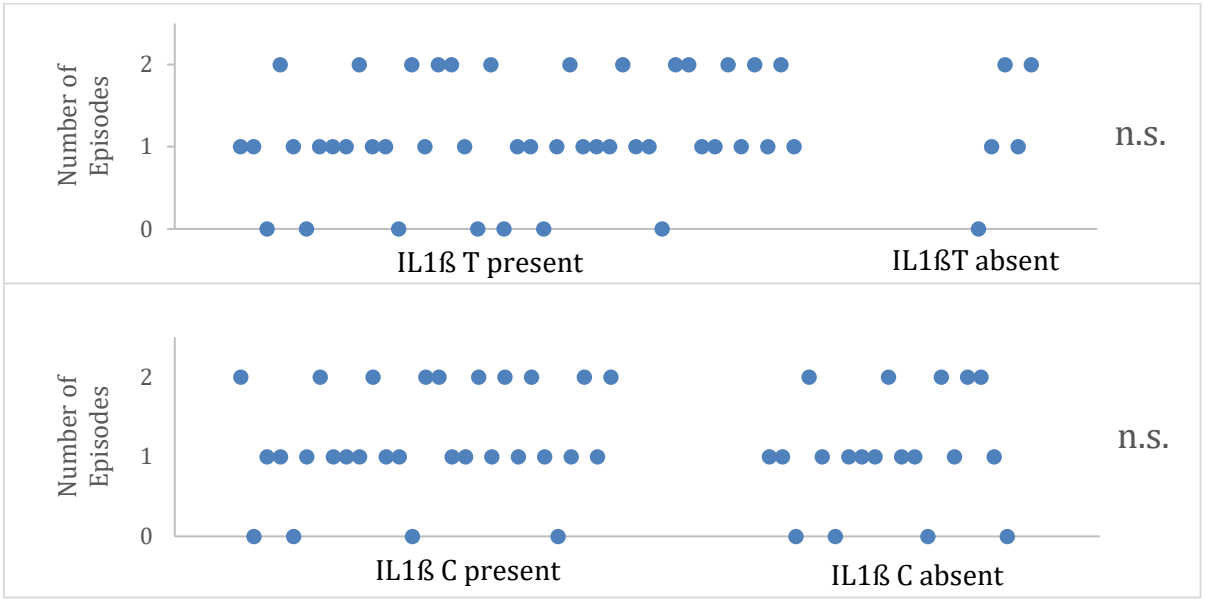

Figure 38. Number of RHV+ episodes as a function of the presence of specific TOLLIP Alleles. Statistical significant differences shown are for the Moses test for extreme reactions (n.s.= not significant)

4. DISCUSSION

TOLLIP and IL1B genes are involved in the innate immunity, participating in inter-cellular signalling in numerous inflammatory processes. It is therefore possible that genetic polymorphisms in these genes can impact viral infection and clearance.

In the present study, we aimed at investigating whether polymorphisms in these two genes (rs5743899 and rs16944) have any impact on the nostril detection of rhinovirus, a common virus that usually produces a mild but recurring infection (common cold).

Over a period of one year, a total of 86 healthy young volunteers donated monthly nasal swab samples for the detection of Rhinovirus RNA (56 voluntaries remained active at the end of the study). Most of the voluntaries (62%) donated 8 or more samples (figure 23), with a lower success rate of sample collection during the summer period when students are more often absent from University, but also when rhinovirus presence is less frequent. A subset of the voluntaries also agreed to donate a blood sample (others only agreed to epithelial mouth cotton samples) for the genotyping of Immune System related genetic polymorphisms. In the present study, we report the results obtained with a subset of 33 blood samples and 15 mouth epithelial cells derived DNA for the genotyping of rs5743899 and rs16944, two polymorphic variants of TOLLIP and IL1B genes respectively.

The results did not support involvement of IL1B polymorphism (rs16944) on the frequency of rhinovirus positive samples, their respective relative rhinovirus concentration or the number and duration of rhinovirus positivity episodes.

However, TOLLIP rs5743899 polymorphism was found to influence rhinovirus infectivity as individuals who never tested positive for rhinovirus were found to harbour the TOLLIP G allele (figures 28, and 33). This effect, at least in the small sample yet available does not seem to be dependent on allele copy number, since

absence of RHV^+ episodes was found both among heterozygous and Homozygous G individuals (Figure 28). Interestingly this allele does not seem to be fully protector as when rhinovirus is found among these individuals it tends to show a higher titre then among TOLLIP G- individuals (figure 33).

A similar effect was observed when the number and duration of rhinovirus positive episodes was analysed. Individuals with the G Allele showed shorter episodes (figure 35) as well as the absence of episodes (figure 37).

These results must be interpreted in the context of TOLLIP gene biological action. It was described that the TOLLIP protein acts as a negative regulator of Toll-like receptor (TLR) mediated signalling (Zhang & Ghosh 2002). This is interesting in light of the finding that HRV capsid is recognized before viral infection by TLR2 (Triantafilou et al,2011) which was among the TLR's proven to be effectively inhibited by TOLLIP (Zhang & Ghosh 2002; Shah et al., 2012). Also, interestingly is the observation that upon HRV internalization, the viral recognition routes change towards TLR7, TLR8 and MDA-5 (Triantafilou et al,2011), with different signalling mechanisms, explaining our observation of a different effect of TOLLIP polymorphism on the prevention of infection and on infection progression. Of all polymorphisms described in the TOLLIP gene, rs5743899 has proven to be interesting in TLR dependent immune reactions as the G allele decreases gene expression (Shah et al., 2012) which in turn leads to increased and sustained production of IL6 upon TLR stimulation (Shah et al., 2012). This is an interesting finding in light of the fact that IL6 is an important immediate but usually transiently expressed inflammation mediator whose continued expression has pathological effects on chronic inflammation (Tanada et al., 2014). This could in turn explain why rs5743899 G allele was found to correlate with increased inflammatory responses, but

lower levels of anti-viral gene (IFN-λ1) expression in response to in vitro viral infection among asthmatic subjects (Huang et al., 2016). In fact, this could represent a similar effect to the observed in our healthy population, as we found that these individuals resist better to infection (possibly due to higher TLR mediated capacity to fight the virus before cellular infection) but have higher viral titres (due to more ineffective immunological responses lacking the action of antiviral genes as IFN-λ1).

5. CONCLUSION

The results here described support the hypothesis that some immunological system related gene polymorphisms may be among the factors influencing the individual differential susceptibility to rhinovirus infection, in particular the capacity to remain virus free for long periods of time. The present results support such a role for the TOLLIP rs5743899 polymorphism, with the G allele presenting as a factor decreasing likelihood of rhinovirus nostril detection. It must however be stressed that even with the incomplete data here described (33/89 TOLLIP typed individuals) the G allele is not a magic bullet. In fact, most of the individuals harbouring a G allele did show positive rhinovirus samples, and in these individuals, the rhinovirus titres are high. Thus, it seems that although TOLLIP G+ individuals seem to resist more efficiently rhinovirus infection, once the infection is established in these individuals, rhinovirus titre seems to be higher than among TOLLIP G- individuals. This allele is thus not a definite protection for rhinovirus, but rather seems to decrease rhinovirus associated risk for infection, although it could be associated with a poorer response to the infection once it is established. This is in agreement with previous work showing that the lower expression of TOLLIP observed among rs5743899 G allele carriers is associated with higher TLR2 mediated signalling (a crucial receptor for RHV capsid identification prior to viral entry into cells), but lower expression of viral response immunological genes such as IFN-λ1 once infection is established.

No influence of the IL1B rs16944 polymorphism on the rhinovirus nostril presence was detected in the present study.

The present results should nevertheless be looked upon with some degree of caution due to the low number of patients this far typed (33/89 for TOLLIP and 48/89 for IL1B). This

fact, together with the low number of individuals (5) typed for TOLLIP that never tested positive for rhinovirus suggests that the present results should be confirmed by extending the study to the remainder of the available voluntaries, and to the voluntaries of a previous proof-of-concept pilot study that the group performed in 2009.

6. BIBLIOGRAPHY

Abul K., Lichtman A., Shiv P., Immunologia cellulare e molecolare, Milano, ELSEVIER, 2012

Ameur A., Rada-Iglesias A., Komorowski J., Wadelius C. Identification of candidate regulatory SNPs by combination of transcription-factor-binding site prediction, SNP genotyping and haploChIP. Nucleic Acids Res. 2009 Jul;37(12):e85. Epub 2009 May 18.

Benschop K.S.M., H. van der Avoort, E. Duizer, M. Koopmans, Antivirals against enteroviruses: a critical review from a public-health perspective, 2015

Bernig T., Chanock SJ. Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. Expert Rev Mol Diagn. 2006May;6(3):319-31.

Bhat, I.A., Naykoo, N.A., Qasim, I., Ganie, F.A., Yousuf, Q., Bhat, B.A., Rasool, R., Aziz, S.A., Shah, Z.A., 2014. Association of interleukin 1 beta (IL-1β) polymorphism with mRNA expression and risk of non small cell lung cancer. Meta Gene 2, 123–133. doi:10.1016/j.mgene.2013.12.002

Blomqvist, S., 2004. Epidemiology of Human Rhinoviruses, Environmental Sciences

Brookes AJ. The essence of SNPs. Gene. 1999 Jul 8;234(2):177-86.

Bustin S. A., A-Z of Quantitative PCR, International University Line, 2004

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, Journal of Molecular Endocrionology, 25, pp. 169-193.

Cabeda J.M. et al. (2012). A Genética Molecular na Investigação e Prática Clínicas: Fundamentos Teòrico e Tecnologicos

Caprioli, A., Morabito, S., Brugère, H., Oswald, E., 2005. Enterohaemorrhagic Escherichia coli: Emerging issues on virulence and modes of transmission. Veterinary Research 36, 289–311. doi:10.1051/vetres:2005002

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet. 2003 Apr;33(4):518-21. Epub 2003 Mar 24

Cheng JC, Tang NL, Yeung HY, Miller N. Genetic association of complex traits: using idiopathic scoliosis as an example. Clin Orthop Relat Res. 2007 Sep;462:38-44.

Clewley J. P., The Polymerase Chain Reaction (PCR) for Human Viral Diagnosis, 1995 CRC Press Inc

Coleman A., Gern J. E., Allergy, Immunity and Tolerance in Early Childhood, Madison, WI, USA, 2015

Cordey, S., Gerlach, D., Junier, T., Zdobnov, E.M., Kaiser, L., Tapparel, C., 2008. The cis-acting replication elements define human enterovirus and rhinovirus species. RNA 14, 1568–1578. doi:10.1261/rna.1031408

de Araujo, F.J., da Silva, L.D.O., Mesquita, T.G., Pinheiro, S.K., Vital, W. de S., Chrusciak-Talhari, A., de Oliveira Guerra, J.A., Talhari, S., Ramasawmy, R., 2015. Polymorphisms in the TOLLIP gene influence susceptibility to cutaneous leishmaniasis caused by leishmania guyanensis in the amazonas state of Brazil. PLoS Neglected Tropical Diseases 9. doi:10.1371/journal.pntd.0003875

Erbek S. S., Yurtcu E., ErbekS., Atac F. B., Sahin F. I., Cakmak O., Proinflammatory cytokine single nucleotide polymorphisms in nasal polyposis. Archives of Otolaryngology—Head and Neck Surgery. 2007; 133(7):705–709.

Filion M., Quantitative Real-Time PCR in Applied Microbiology, Caister Academic Press, Norfolk 2012

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, LochnerA, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ,Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296: 2225-9.

Hakhverdyan, Mikhayil. (2018). Basics of the Polymerase Chain Reaction.

Hewat A., Handbook of Cell Signaling (Second Edition), 2010

Huang C., Jiang D., Francisco D., et al., Tollip SNP rs5743899 Modulates Human Airway Epithelial, U.S.A., 2016

Responses to Rhinovirus Infection

Jawetz, Melnick, Adelberg, Medical Microbiology

Johannessen I., Burns S. M., in Medical Microbiology (Eighteenth Edition), 2012

Karjalainen J., Joki-Erkkilä V. P., Hulkkonen J., et al., The IL1A genotype is associated with nasal polyposis in asthmatic adults. Allergy. 2003; 58(5):393–396.

Kennedy S., N. Oswald, PCR Troubleshooting and Optimization: The Essential Guide, Caister Academic Press Norfolk 2011

Khazim, Khaled & Azulay, Etti & Kristal, Batya & Idan, Cohen. (2018). Interleukin 1 gene polymorphism and susceptibility to disease. Immunological Reviews. 281. 40-56. 10.1111/imr.12620.

Komar, A. A. (2009). Single Nucleotide Polymorphisms - Methods and Protocols. Edited by Komar, A. A. Cleveland: Humana Press

Kropski JA, Lawson WE, Blackwell TS. Personalizing Therapy in Idiopathic Pulmonary Fibrosis: A Glimpse of the Future? *American Journal of Respiratory and Critical Care Medicine*. 2015;192(12):1409-1411. doi:10.1164/rccm.201509-1789ED.

Kruglyak L., Nickerson DA. Variation is the spice of life. Nat Genet. 2001 Mar;27(3):234-6

Lessler, J., Reich, N.G., Brookmeyer, R., Perl, T.M., Nelson, K.E., Cummings, D.A., 2009. Incubation periods of acute respiratory viral infections: a systematic review. The Lancet Infectious Diseases. doi:10.1016/S1473-3099(09)70069-6

Lewis-Rogers, N., Crandall, K.A., 2010. Evolution of Picornaviridae: An examination of phylogenetic relationships and cophylogeny. Molecular Phylogenetics and Evolution 54, 995–1005. doi:10.1016/j.ympev.2009.10.015

Linden, Lonneke & Wolthers, Katja & Kuppeveld, F. (2015). Replication and Inhibitors of Enteroviruses and Parechoviruses. Viruses. 7. 4529-62. 10.3390/v7082832.

Liu Y, Jiang X-L, Liu Y, et al. Toll-interacting protein (Tollip) negatively regulates pressure overload-induced ventricular hypertrophy in mice. *Cardiovascular Research*. 2014;101(1):87-96. doi:10.1093/cvr/cvt232.

Logan J. M. J., Edwards K. J., Saunders N. A., Real-Time PCR: Current Tecnology and Applications, Norfolk 2009 Caister Academic Press

Lu K., Kefeng I. et al.**, A**utophagic Clearance of PolyQ Proteins Mediated by Ubiquitin-Atg8 Adaptors of the Conserved CUET Protein Family, Volume 158 , Issue 3 , 549 – 563, 2014

Mattila K.M., 2002. Association of an interleukin 1B gene polymorphism (-511) with Parkinson's disease in Finnish patients. Journal of Medical Genetics 39, 400–402. doi:10.1136/jmg.39.6.400

Myatt, T. A. et al, 2004. Detection of airborne rhinovirus and its relation to outdoor air supply in office environments Am. J. Respir. Crit. Care Med. 169, 1187–1190. Doi: 10.1164/rccm.200306-760OC

Monto A. S., Epidemiology of viral respiratory infections, Michigan, USA, 2002

Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000 Sep;156(1):297-304.

Pääbo S., The mosaic that is our genome. Nature. 2003 Jan 23;421(6921):409-12.

Rancaniello I., Freymuth, F., Vabret, A., Rozenberg, F., Dina, J., Petitjean, J., Gouarin, S., Legrand, L., Corbet, S., Brouard, J., Lebon, P., 2001. Replication of respiratory viruses, particularly influenza virus, rhinovirus, and coronavirus in HuH7 hepatocarcinoma cell line. Journal of Medical Virology 77, 295–301. doi:10.1002/jmv.20449

Reich DE, Lander ES. On the allelic spectrum of human disease. Trends genet. 2001; 17: 502-10.

Schmitz N., Kurrer M., Kopf M., The IL-1 receptor 1 is critical for Th2 cell type airway immune responses in a mild but not in a more severe asthma model. European Journal of Immunology. 2003; 33(4):991–1000.

Scialpi A., Mengoni A., La PCR e le sue varianti: quaderno di laboratorio. Firenze 2008, Firenze University Press

Semler, B.L., Ertel, K.J., Editors-in-Chief: B.W.J. Mahy, M.H.V. van Regenmortel, 2004. Picornaviruses: Molecular Biology, in: Applied and Environmental Microbiology. Academic Press, Oxford, pp. 129–140. doi:10.1128/AEM.69.1.350-357.2003

Shah JA, Vary JC, Chau TT, et al. Human TOLLIP regulates TLR2 and TLR4 signaling and its polymorphisms are associated with susceptibility to tuberculosis. *Journal of immunology (Baltimore, Md : 1950)*. 2012;189(4):1737-1746. doi:10.4049/jimmunol.1103541.

Song Z, Yin J, Yao C, et al. Variants in the Toll-interacting protein gene are associated with susceptibility to sepsis in the Chinese Han population. *Critical Care*. 2011;15(1):R12. doi:10.1186/cc9413.

Stobart Christopher C., Nosek Jenna M., Moore Martin L., Rhinovirus Biology, Antigenic Diversity, and Advancements in the Design of a Human Rhinovirus Vaccine, Frontiers in Microbiology, 2017

Tanaka T., Narazaki M., Kishimoto T., IL-6 in Inflammation, Immunity, and Disease, Cold Spring Harb Perspect Biol 2014;6:a016295

Triantafilou K., Vakakis E., Richer E.A.J., Human rhinovirus recognition in non-immune

cells is mediated by Toll-like receptors and MDA-5, which trigger a synergetic

pro-inflammatory immune response, UK, 2011

Van Der Poll T., Van Devender S. J. H., Cytokines and anticytokines in the pathogenesis of sepsis. Infectious disease clinics of North America; 13: 413-426, 1999

Waris, M., Österback, R., Lahti, E., Vuorinen, T., Ruuskanen, O., Peltola, V., 2008. Comparison of sampling methods for the detection of human rhinovirus RNA. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 58, 200–4. doi:10.1016/j.jcv.2013.06.002

Weissensteiner T., Nolan T., Bustin,H S. A., Griffin G., Griffin A., PCR Technology: Current Innovations, Second Edition CRC Press 2004

Zhang G., Ghosh S., Negative Regulation of Toll-like Receptor-mediated Signaling by Tollip, U.S.A, 2002