University OF TAMPERE

Role of Human Endogenous Retrovirus type K (HERV-K) in immunosenescence: RNA-seq based transcriptomic analysis.

Binisha Hamal Mishra
Master's thesis
University of Tampere
Faculty of Social Sciences/Health Sciences
Public and Global Health
December 2018

University of Tampere
Faculty of Social Sciences/Health Sciences
BINISHA HAMAL MISHRA: ROLE OF HUMAN ENDOGENOUS RETROVIRUS TYPE K (HERV-K) IN IMMUNOSENESCENCE: RNA-SEQ BASED TRANSCRIPTOMIC ANALYSIS.
Master's thesis, 60 pages
Supervisors: Prof. Mikko Hurme and Prof. Marja Jyhlä
Public and Global Health
December 2018

Background: Immunosenescence, ageing related decline in immunity, leads to increased incident and severity of various infections, reactivation of latent viruses and decline in vaccine responses among the elderly. As the proportion of older population is rising sharply across the globe, fundamental understanding of immunosenescence is vital for prevention, control and treatment of ageing-related diseases among the vulnerable population. Cytomegalovirus (CMV) plays a substantial role in acceleration of immunosenescence. Interestingly, HERV-K (HML-2), the youngest group of human endogenous retroviruses, has been shown to be co-expressed with CMV in several diseases and cancers. Thus, it is rational to hypothesize that HERV-K (HML-2) potentially plays role in immunosenescence.

Subjects and methods: We investigated potential role of HERV-K (HML-2) in immunosenescence using whole genome RNA-sequencing technique with seven young subjects, age=26-32 years and seven older subjects from Vitality 90+ study cohort, age=90+ years. Gene set enrichment analysis (GSEA) was used to elucidate young and older group specific profile of gene ontology (GO) based biological processes. Direction (up or down) of regulation of the significant biological processes among older subjects was investigated with differential gene expression and subsequent over-representation analysis. Genes that were highly correlated (correlation > 0.80 or < -0.80) with HERV-K expression were used for over-representation analysis with GO based biological processes.

Results and Conclusion: Gene set enrichment analysis of nonagenarian and young groups identified significantly large number of biological processes enriched in nonagenarian as compared to young subjects. Differential gene expression analysis identified 347 genes highly correlated with the youngest provirus, 1q22 that were also

differentially expressed between the nonagenarian and young groups. Over-representation of those genes uncovered several immunity related genes and biological processes associated with provirus 1q22 among older subjects. Therefore, our results indicate that provirus 1q22 could be a potential player in immunosenescence. Further research with larger datasets with gender variation is needed for deeper understanding.

Keywords: Immunosenescence, HERV-K (HML-2), provirus 1q22, cytomegalovirus, differential gene expression, enrichment analysis, over-representation analysis

Acknowledgement

I would like to express my sincere gratitude to my supervisor Professor Mikko Hurme for offering me master's thesis position, his continuous support during the study and for his patience, motivation and immense knowledge that helped me in all the time of the thesis work. It would have not been possible without his help. My sincere thanks also goes to Professor Marja Jyhlä who was always there for any help I needed.

Contents

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

PBMC	Peripheral blood mononuclear cell
HERV-K (HML-2)	HERV-K
HERV	Human Endogenous Retrovirus
HCMV	Human Cytomegalovirus
EBV	Epstein-Barr-Virus
DC	Dendritic cells
NK cell	Natural killer cell
GO	Gene Ontology
PBMC	Peripheral blood mononuclear cell
GSA	Gene set analysis
GSEA	Gene set enrichment analysis
FPKM	Fragments per kilo base per million
ROS	Reactive oxygen species
IL	Interleukin
TNF-α	Tumor necrosis factor α
IFN- γ	Interferon γ

1 INTRODUCTION

The global population is ageing. In 2017, almost 962 million people (~13% of the global population) worldwide were 60 years or older. By 2050, people aged 60 or older is expected to increase to 2.1 billion and 3.1 billion in 2100 (United Nations Department of Economic and Social Affairs, 2017). One of the emerging problems due to ageing of population is increased vulnerability to infectious diseases as compared to the younger population. Studies have demonstrated that mortality rate increases 89-fold for pneumonia and influenza and 43-fold for cancer in people aged 65 in western countries (Bulati, Pellicanò, Vasto, & Colonna-Romano, 2008). Understanding and preventing these infections, thus, becomes important to public health care system. The most important cause for increased rate of infection and cancers in elderly population is due to diminished immunity. Deterioration of immunity due to ageing is known as immunosenescence. Immunosenescence leads to increased incident and severity of various infections, reactivation of latent viruses and decline in vaccine responses among the elderly. With the surge in proportion of older population across the globe, immunosenescence is emerging public health problem. Fundamental understanding of immunosenescence is thus vital for prevention and treatment of ageing-related diseases.

Immunosenescence affects both adaptive and innate immune system. However, innate immune system seems to be less affected by age. Markers of immunosenescence include decrease in naive T cell exiting thymus and declining function of haematopoietic cell. Naive T cell in the elderly shows numerous defects such as reduced interleukin-2, shorter telomeres, and restricted T cell receptor repertoire. Evidence suggests that epigenetic changes such as histone acetylation and DNA methylation accumulate with ageing contributing to immunosenescence (Ponnappan & Ponnappan, 2011).

Whether or not immunosenescence is a result of an adaptive response to infections is an open question. Evidences show that chronic T cells activation due to persistent viral infection in immune-competent hosts can potentially influence T cells (Aiello, Chiu, & Frasca, 2017). The most common pathogen known to induce immunosenescence is CMV (Pawelec, Derhovanessian, Larbi, Strindhall, & Wikby, 2009; Pawelec, Solana, & Wikby, 2005). Several studies have shown that HERV-K (HML-2) to be co-expressed along with CMV in many diseases (Assinger, Yaiw, Göttesdorfer, Leib-Mösch, & Söderberg-Nauclér, 2013; Bergallo et al., 2015). This raises interest in whether or not HERV-K (HML-2) is also involved in immunosenescence. The focus of this study is to investigate potential role HERV-K (HML-2) proviruses in immunosenescence.

HERV proviruses comprise about 8% of the human genome (Bannert & Kurth, 2004). The proviruses integrated themselves into human genome through infection long ago. Through the evolutionary period, the proviruses have accumulated mutations that have disabled them to produce infections. HERV-K(HML-2) are understood to be the youngest HERV family (less than 5 million years) that contains a complete set of genes (albeit defective due to mutations) (Bannert & Kurth, 2004; Dewannieux et al., 2006). Their expression has also been associated with breast and bladder cancer, sarcoma, malignant melanoma and lymphoma (Cegolon et al., 2013). Understanding their role in human diseases could establish them as a probable therapeutic targets as they represent virus-like tumour antigens (Assinger, Yaiw, Göttesdorfer, Leib-Mösch, & Söderberg-Nauclér, 2013). However, unlike CMV, association of HERV-K proviruses with immunosenescence has not been well studied.

We studied role of HERV-K proviruses in immunosenescence using whole genome RNA-sequencing technique with young healthy laboratory subjects (seven female between 26-32 years) and seven subjects from Vitality 90+ study cohort (seven female between 90-99 years). We performed gene set enrichment analysis (GSEA) to identify

young and older group specific Gene Ontology (GO) (Ashburner et al., 2000) based biological processes that are significantly associated with the proviruses. This approach, however, is not sufficient to elucidate the important question of whether biological processes are up or downregulated among older subjects. For that, differential gene expression analysis was done between young and nonagenarian groups (age of 90 and 99). We focused on only those genes that are correlated (correlation >0.80 or <-0.80) with the differentially expressed proviruses. We then performed over-representation analysis to identify significantly enriched GO biological processes. The study identified several immunity related genes and biological processes significantly associated with HERV-K provirus 1q22. Our results suggest that provirus 1q22 is potential player in mediating immunosenescence.

2 LITERATURE REVIEW

The focus of the study is to elucidate potential role of HERV-K (HML-2) provirus in immunosenescence. Therefore, in literature review, factors that effects immunosenescence in elderly will be focused. As immunosenescence is deterioration of immunity with age, information on ageing and its cellular and molecular manifestations is important to discuss. Following that, relation of inflammaging and immunosenescence will be briefly discussed. Types of immunity is important to discuss as immunosenescence affect both adaptive and innate immune system in different forms. Clinical implication of immunosenescence is important to be discussed, as it will explain the need to investigate more on immunosenescence. Immunosenescence has been known to be induced by pathogens like CMV. Furthermore, HERV-K has been shown to be co-expressed with CMV in several diseases and cancers. Therefore, HERV-K (HML-2) was chosen as pathogen of interest as there has not been any research done to investigate its role in immunosenescence.

2.1 Ageing

Ageing is defined as progressive deterioration of physiological ability leading to age-related mortality and decrease in the age related reproductive rate (Kirkwood, 1995; López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013; Rose, 1991). However, it is largely determined by genetics, environmental factor such as diet, exercise, exposure to microorganism, pollutants, and ionizing radiation (Sorrentino, Sanoff, & Sharpless, 2014). Hence, ageing occurs in different degrees in different individuals. Women live longer than men worldwide, indicating that longevity of life seems strongly influenced by gender (Sorrentino et al., 2014).

2.2 Cellular and Molecular manifestation of ageing

Several theories and mechanisms interpret the process and consequences of ageing. Due to the multiclausal dimension of ageing process, just one theory cannot not explain all its mechanism. López-Otín et al., 2013 have come up with nine cellular and molecular hallmark of ageing which together determine the ageing phenotype (Figure 2 -1).

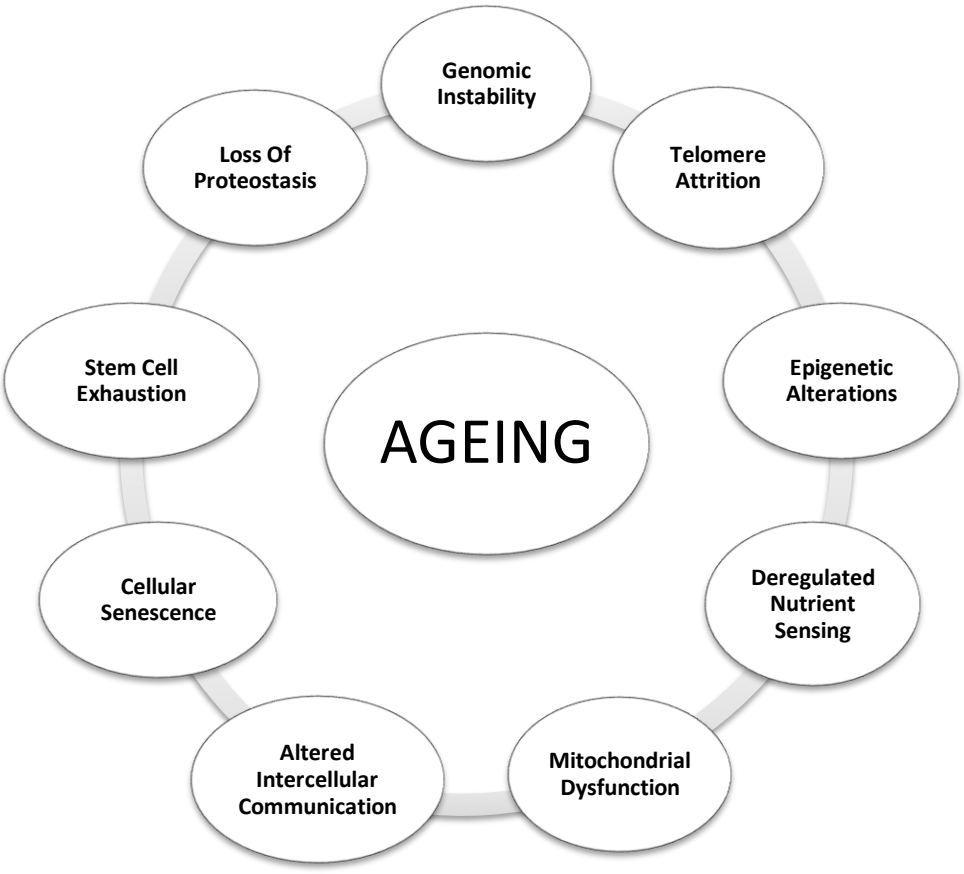

Figure 2-1 Hallmarks of Ageing (López-Otín et al., 2013).

2.3 Inflammaging

Inflammaging is illustrated as a chronic, low-grade, systemic inflammation that develop with age and are associated with frailty and mortality in elderly (Franceschi & Campisi, 2014). Studies revealed that during mild inflammation, levels of various pro inflammatory cytokines like interleukin 6 (IL-6), C-reactive protein (CPR), TNF-α and IFN- γ are elevated. This is associated and is predictive of many ageing phenotypes such as immune senescence, metabolic homeostatic, energy production and utilization and neuronal health (Franceschi & Campisi, 2014). The aetiology of inflammaging remain largely unknown. Acute inflammation has beneficial effect towards the neutralization of harmful pathogens during early life and adult. However, acute inflammatory response to pathogens become detrimental during ageing leading to increase susceptibility to infection.

Franceschi & Campisi, 2014 summarized the source of inflammaging into five following categories as,

i) destroyed cells and macromolecules that accumulate with age
ii) cellular senescence and secretion of proinflammatory cytokines
iii) immunosenescence that can be initiated by persistent infection such as CMV
iv) activation of the coagulation system with age and
v) harmful products and metabolites produced by oral or gut microbes that decreases with age.

2.4 Immunosenescence

Immunosenescence refers to the age-associated deterioration of the immune system (Boraschi & Italiani, 2014; Fülöp, Fóris, Wórum, & Leövey, 1985). This contributes to increasing levels of infection and poor response to vaccinations, as well as diseases like cardiovascular disease and cancer. Immunosenescence is also likely to contribute to inflammaging (Franceschi & Campisi, 2014).Immunosenescence leads to an

increased number of pro-inflammatory cytokine-producing cells like inflammatory cells. In addition, chronic inflammatory conditions can accelerate the immunosenescence process (Franceschi & Campisi, 2014). Inflammatory cells and immune cells have the same cellular foundation. Hence, Inflammaging and immunosenescence are connected and coexist, and is difficult to differentiate whether the inflammation-related diseases are caused by inflammaging or immunosenescence (Xia et al., 2016). Furthermore, internal relationship between inflamm-ageing and immunosenescence is not known, and the causes and pathological mechanisms of immune-inflammatory diseases are not understood. In addition, changes in adaptive and innate immunity is different with age. Adaptive immunity declines with age unlike innate immunity that will undergo minor changes that might cause mild hyperactivity. Furthermore, with age, adaptive immune system might give burden to the innate immunity. These changes are due to the lifelong exposure to pathogens and antigens such as CMV, human endogenous retrovirus, Epstein-Barr virus and intrinsic changes in immune cells and genetic predisposition (Franceschi, Bonafè, & Valensin, 2000; McElhaney & Effros, 2009). Therefore, we will discuss into details the changes in innate and adaptive immune system with age.

2.5 Changes in innate immune system

The innate immune system is the initial response of defense against pathogens. With age, there is compromised function in epithelial cells of the skin, lung, or gastro-intestinal tract, which facilitates pathogens to invade mucosal membrane resulting in health-related challenges among elderly. The innate immune system consists of phagocytic cells (neutrophils, and monocytes/macrophages, and dendritic cells), natural killer cells and soluble mediators like natural cytokines, chemokines, hormones, and oxygen-free radicals.

2.5.1 Neutrophils

Neutrophils are short-lived cells. They act as a host defense mechanism to bacterial, fungal infections and during acute inflammation. Neutrophils ingest pathogenic organism by phagocytosis. Generation of reactive oxygen species (ROS) and toxic granular enzymes work together to kill the ingested pathogens. (Lehrer et al., 1988). Neutrophils are recruited to the infection site by cytokine and chemokine mainly by IL-1 and IL-8(Lehrer, Ganz, Selsted, Babior, & Curnutte, 1988; Sica et al., 1990). Various studies showed no change in the number of circulating neutrophils with age (Chatta & Dale, 1996; Shekarabi & Asgari, 2014). However, neutrophils retrieved from elderly are less phagocytic than those from younger individuals (Butcher, Chahel, & Lord, 2000; Fülöp, Fóris, Wórum, & Leövey, 1985). Neutrophils also promote the maturation and migration of DC that will trigger the adaptive immune response (Solana, Campos, Pera, & Tarazona, 2014).

2.5.2 Natural killer cell

Natural killer (NK) cells are cytotoxic cell that play a major role in the elimination of virus-infected cells or tumor cells. It also helps in the regulation of the immune response by producing cytokines and chemokines that can stimulate cellular components of innate and adaptive immunity (Campos et al., 2014; Weiskopf, Weinberger, & Grubeck-Loebenstein, 2009). Evidence has shown the importance of NK cell activity in retaining health in ageing individual. Studies have reported that high NK cytotoxicity to be associate with healthy ageing and longevity (Solana & Mariani, 2000), whereas low NK cytotoxicity to be associated with increased morbidity and mortality due to infections, in elderly (K. Ogata et al., 2001). Other age related changes in NK include increased rate of rejuvenation of latent Mycobacterium tuberculosis (Pietilä et al., 2015) and slower resolution of inflammatory responses.

2.5.3 Macrophages

Macrophages trigger adaptive immune responses against pathogens by acting as antigen-presenting cells. Macrophages acts on the tissues in which they reside (Gomez, Nomellini, Faunce, & Kovacs, 2008). Macrophages are involved in phagocytosis of antigens and cellular debris and killing of invading pathogens and tumors through oxygen-dependent and independent mechanisms(Gomez et al., 2008). Furthermore, macrophages also secrete cytokines, which regulate multiple immune functions, specifically inflammatory responses. With age, antigen representation by macrophages is decreased likely due to decrease levels of MHC class II molecules (Plowden, Renshaw-Hoelscher, Engleman, Katz, & Sambhara, 2004; Zissel, Schlaak, & Muller-Quernheim, 1999). A study has shown decreased chemotaxis and defective phagocytosis in macrophages from elderly humans and mice. (Fietta et al., 1993). With age, macrophages significantly reduce secretion of cytokines such as TNF-α and IL-6 and decrease B7 receptor expression, which activates T cell (Davalos, Coppe, Campisi, & Desprez, 2010; van Duin et al., 2007).

2.5.4 Dendritic cells

Dendritic cells (DC) are the key component in linking the bridge between innate and adaptive immunity. DC are responsible for the first recognition of T and B lymphocytes and NK cell. Studies have reported that there will be decrease in the Langerhans cells in skin and plasmocytes with age (Agrawal et al., 2007). DC from healthy elderly is able to retain their capacity to efficiently activate antigen to T cells (Lung et al., 2000). However, expression of costimulatory molecules and IL-12 production is decreased in frail elderly and hence, have impaired ability to activate T cell proliferation (Lung, Saurwein-Teissl, Parson, Schönitzer, & Grubeck-Loebenstein, 2000). Furthermore, antigen-presenting capacity of aged plasmacytoid dendritic cell of elderly showed reduced capacity to activate T cell proliferation and IFN- γ secretion in CD4 and CD8 T cells as compared with plasmacytoid dendritic cell from young adults.

Table 2-1 Age related changes in innate immune system

Cell type	**effect of ageing**
Neutrophil	reduces chemotaxis and phagocytosis reduces superoxide anion production reduces activation signal transduction reduces production of IFN- γ
Macrophage	reduces phagocytosis reduces cytokine and chemokine secretion delayed wound healing Diminished TLR expression and function
Natural killer cell	decreases cytotoxic capacity decreases secretion of IFN- γ decreases response to IL-2
Dendritic cell	decreases ability to stimulate lymphocytes reduces chemotaxis decreases capacity to stimulate antigen specific T cell

2.6 Changes in adaptive immune system

Adaptive immune system is composed of the cellular immune response that is T lymphocytes, and the humoral immune response, which is the B lymphocytes.

2.6.1 B Lymphocytes

B lymphocytes develop in bone marrow and continues in the spleen for final maturation. B lymphocytes are responsible for the production of antibodies in response to infections. Number of peripheral B cell do not decline with age but the composition of the cell changes. The percentage of naïve B cells, accompanied with

the absence of CD27 cell, is significantly reduced with age. With age, there is defective in the production of antibody and impaired B cell responses. This will adequately reduce the ability to respond against pathogens (Buffa et al., 2011; Visentini et al., 2011). Furthermore, research showed that the antibody response against hepatitis vaccines was diminished dramatically in elderly adults (mean age 61 years) compared with a young adults (mean age of 33 years) (Rosenberg et al., 2013). Serum immunoglobulin levels are stable in elderly individual, however, the antibodies secretion in old age are of lower affinity because of antibody isotypes shift from IgG to IgM (Rosenberg et al., 2013).

2.6.2 T Lymphocytes

T lymphocyte develop in thymus. Thymus is most active in the early years, reaches peak size within the first year and undergoes decline with age almost complete by the age of 75.(George & Ritter, 1996) . This process is called involution. During involution, there is reduction of overall size of thymus and a replacement of the functional cortex and medulla tissue by fat and increase in volume of the perivascular space of human thymus with age (Flores, Li, Sempowski, Haynes, & Hale, 1999). This process results in decrease in the number of naïve T cells existing from thymus , which in turn reduces T-cell receptor diversity (Chou & Effros, 2013). This decrease in the naïve T- cell output is linked with the increased incidence of infections, cancers autoimmune diseases (Palmer, 2013). Furthermore, studies have suggested that naïve T cell of elderly individuals have numerous functional defect such as shorter telomere, restricted T cell receptor repertoire and reduced interleukin-2 production with compassion with younger adults (Kohler et al., 2005; Pfister et al., 2006). Shortening or sequential loss of telomeric DNA increases apoptosis with age (Aggarwal & Gupta, 1998). Shortening of telomere is linked to various diseases associated with ageing like diabetes, hypertension, Alzheimer's disease and cancer (Rizvi, Raza, & Mahdi, 2014). However, naïve T cell in elderly is associated with increased number of antigen-encounter memory cell, particularly effector T cells. This effector T-cell shows

phenotypic changes of co-stimulatory molecule such as CD28, which is said to be a predictor of immune incompetence in elderly (Vallejo, 2005).

Table 2-2 Age related changes in adaptive immune system

Cell type	**Effect of Ageing**
T- lymphocyte	decreases in CD4+/CD8+ lymphocyte decreases in membrane receptor increases release of pro-inflammatory cytokines like IL-1,IL-6 and TNF- α
B -lymphocyte	decreases in antibody production decreases in lymphocyte number decreases in production of IL-2 increases in Autoreactive serum antibodies

2.7 Clinical implications of immunosenescence

The clinical implications of immunosenescence include increased risk of infections, malignancy, autoimmune disorder, atherosclerosis and various neurodegenerative disorder. This is due to various mechanism discussed in previous chapters. With understanding of the mechanism of immune system, efficient vaccines and novel therapies could be developed to restore appropriate immunity treat or prevent these conditions by identifying the exact components to manipulate as needed. Therefore, understanding immunosenescence help to increase healthy life span of individual during old age.

2.7.1 Vaccination

Elderly people are susceptible to morbidity and mortality from infectious disease globally. Deaths among elderly due to pneumonia is twice as high and from tuberculosis is ten-folder higher compared with the young adults. Elderly people have less circulating monocytes and dendritic cells than that of young individuals. Furthermore, there is also fewer tissues associated with dendritic cells and macrophages and increased splenic dendritic cells. In addition, there is decrease in the population of naïve T and B cells with increase in mature T and B cells. Due to this, overall antibody production and specificity is significantly diminished in the elderly. There are vaccines to minimize the mortality from infectious condition, however, immunosenescence hinder the ability to develop immunity after vaccination in elderly. Vaccination will activate immune response in both humoral and cell mediated immunity. The effectiveness of vaccine depends on the immune response, therefore, immunocompromised individual such as elderly are at high risk of being inadequately protected by certain vaccine. Studies have demonstrated that the ability of influenza vaccine to induce protection depends on age (P. O. Lang et al., 2010; Pierre Olivier Lang, Govind, Mitchell, Siegrist, & Aspinall, 2011). It showed that efficacy of vaccine is between 70% and 90% in those under 65 years of age, but it was only 30–40% for those over 65 years of age. Thus, the current vaccines are less immunogenic and less efficient for the elderly. To develop effective vaccines for the 65+ population requires clear understanding of mechanism to stimulate a cell-mediated response that increases the protection in elderly population.

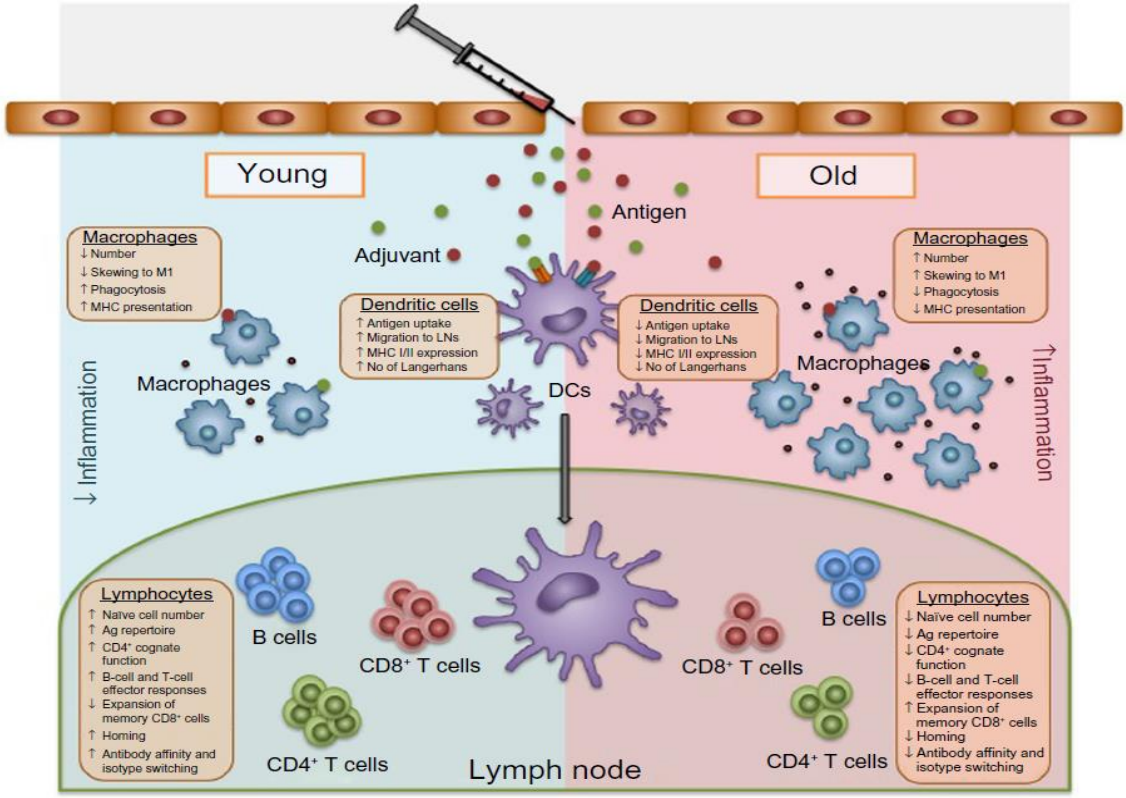

Figure 2-2 Elements of the immune response necessary to mount a response to vaccination in young and old (Bowdish, Loukov, & Naidoo, 2015).

2.7.2 Cancer

The incidence and prevalence of most cancers is higher with age (Thakkar, McCarthy, & Villano, 2014). Incidence of cancer in those over 65 is 10 times higher than in those younger (Thakkar et al., 2014). About 60% of all cancers and 70% of mortality due to cancer occur in people older than 65 years. Relation between ageing and cancer is not well- understood considering the existence of multiple different cancers. However, one of the reason to explain the incidence of cancer in older age is immunosenescence. It has been suggested that inflammatory markers increase with age, while immunological surveillances is less effective (Hoffe & Balducci, 2012). This could promote the growth of immunogenic tumors with age. Furthermore, diminished proliferation response and cytotoxicity of natural killer cells could also

increase the risk of tumor cells in elderly (Imai, Matsuyama, Miyake, Suga, & Nakachi, 2000).

There was fourteen million new cancer cases worldwide in 2012, and 2.2 million (15.4%) were due to infection (Plummer et al., 2016). Several infectious agents are considered to cause cancer, for example; Helicobacter pylori, hepatitis B virus, hepatitis C virus, human papillomavirus, Epstein-Barr virus, CMV, human herpesvirus and recently Human endogenous retrovirus. Cancer will remain a major cause of death around the world. Fraction of human cancer would be expected to be in control by preventing infection associated cancers (de Flora & Bonanni, 2011). However, further studies are needed to understand the underlying mechanism between infectious pathogens and cancer.

2.8 Human endogenous retroviruses (HERVs)

HERVs are remnants of ancient retrovirus infections that have been integrated into DNA via infection of germline cell. They are transmitted vertically through germline and through generations in a Mendelian manner. HERVs comprise about 8% of the human genome (Bannert & Kurth, 2004). There are more than 30 HERV families that have been identified during past two decade (Tristem, 2000). Through the evolutionary period, HERVs have accumulated mutations, deletions, termination signals within coding sequences that have disabled limited number of HERV to produce infections. However, a number of HERVs have intact open reading frames that code for functional proteins. HERVs are part of retrotransposon families or "jumping genes" within the human genome (Bannert & Kurth, 2004). Transposable elements may also be able to affect the function and expression of nearby genes depending on where they locate themselves in the genome. HERVs have frequently been proposed as etiological co-factors in chronic diseases such as cancer, autoimmunity and neurological disease (R. P. Subramanian, Wildschutte, Russo, &

Coffin, 2011). However, the causal relationship remains enigmatic (Nelson et al., 2004). HERV-K (HML-2) is understood to be the youngest HERV family (less than 5 million years) that contains a complete set of genes (albeit defective due to mutations) (Bannert & Kurth, 2004).

2.8.1 HERV-K (HML-2)

HERV-K (HML-2) (will be referred as HERV-K hereafter) is the newest of the HERV-K family (Barbulescu et al., 1999). Therefore, has had the least time to acquire mutations and comprises the most preserved set of retrovirus in the human genome (R. P. Subramanian et al., 2011). The "K" from HERV-K comes from the use of lysine tRNA to prime reverse transcription and (HML-2) comes from their relationship to their *pol* sequence to mouse mammary tumor virus, a causative agent for breast cancer in mice (Bittner, 1936). HERV-K is the only known HERV that is capable of expressing viral proteins and viral like particles. HERV-K contains more than 90 proviruses and around 950 long terminal repeats (LTRs) produced from recombination between the 5' and 3' LTRs of a provirus (Hughes & Coffin, 2004). More than 90 proviruses of the HERV-K family have intact and well preserved open reading frames in genes which sometimes encode for functional proteins that are expressed in various tissues (R. P. Subramanian et al., 2011). Some HERV-K has also maintained open reading frames for retroviral gene such as *gag, pro, pol* and *env.*

HERV-K is expressed at only very low levels in healthy tissues (Hohn, Hanke, & Bannert, 2013). Studies indicated that increased HERV- K expression to be linked with several diseases in humans. Diseases such as breast cancer (Bittner, 1936; Golan et al., 2008; Ono, Kawakami, & Ushikubo, 1987), germ cell tumor(Golan et al., 2008; Sauter et al., 1996), melanoma (Büscher et al., 2005; Hahn et al., 2008; Muster et al., 2003), ovarian cancer(Wang-Johanning et al., 2007), leukemia(Depil, Roche, Dussart, & Prin, 2002; Iwabuchi et al., 2004), schizophrenia (Huang et al., 2006; Karlsson et

al., 2001; Slokar & Hasler, 2016), rheumatoid arthritis (Freimanis et al., 2010; Reynier et al., 2009) as well as during HIV infection (Contreras-Galindo et al., 2012; Gonzalez-Hernandez et al., 2012; van der Kuyl, 2012). Various studies demonstrated the elevated expression level of HERV-K antibodies in blood of patients at early stage of cancer (Büscher et al., 2006; Karreth et al., 2011; Muster et al., 2003; Wang-Johanning et al., 2013). Therefore, presence of antibodies specific for HERV-K proteins could be promising additional diagnostic and prognostic approach towards these types of cancers (Kleiman et al., 2004).

Table 2-3List of malignant diseases that are associated with HERV-K activity.

Tissues	**cancers**	**HERV-K**
Breast	Breast cancer	Retroviral activity Virus particles Free viral RNA
Skin	Melanoma	Retroviral activity Enhanced transcription Retroviral particles
Testes	Germ cell tumors	Expression of Rec,Np9
Ovary	Ovarian cancer	Expression of Gag and Env
Prostrate	Prostate cancer	Expression of gag
Blood	Lymphoma	Free RNA Retroviral activity Virus particles

Until recently, HERV-K are related with induction of various diseases. However, Kleiman et al., 2004 suggested that targeting of neoplastic cell could be novel therapeutic approach to cure cancer. The authors suggested that reactivation and expression of HERV-K in healthy tissue might affect cellular physiology by activating HERV-K promotors that act on genes on cellular level. However, further investigation

on curing cancers by activating HERV-K to activate innate immune response is necessary.

Furthermore, it was shown that HERV-K expression can also be activated by EBV (Hsiao, Lin, Tai, Chen, & Huber, 2006; Sutkowski, Chen, Calderon, & Huber, 2004; Sutkowski, Conrad, Thorley-Lawson, & Huber, 2001), human herpesvirus (Tai, Luka, Ablashi, & Huber, 2009; Turcanova, Bundgaard, & Höllsberg, 2009) and human CMV (Assinger et al., 2013; Bergallo et al., 2015).

2.9 HERV-K and Cytomegalovirus

Human CMV infection infect 40-100% of human population globally (Jean Beltran & Cristea, 2014). After primary infection, CMV establish itself in some form in hematopoietic stem cells (Sindre et al., 1996). The mechanism in which CMV persists in human is still a poorly understood because it rarely causes problems in healthy people. CMV usually causes asymptomatic or mild infection in healthy individuals, but may cause severe and life-threatening conditions in immunocompromised individuals (Lancini, Faddy, Flower, & Hogan, 2014). CMV has been detected in various cancers such as, medulloblastoma, neuroblastoma, colon, breast and prostate cancer (Cobbs, Soroceanu, Denham, Zhang, & Kraus, 2008; Harkins et al., 2002, 2010; Samanta, Harkins, Klemm, Britt, & Cobbs, 2003). Studies have suggested that CMV and an inflammatory environmental condition could facilitate HERV-K activation in vivo in early tumor lesions (Assinger, Yaiw, Göttesdorfer, Leib-Mösch, & Söderberg-Nauclér, 2013; Bergallo et al., 2015). Furthermore, it has shown that HERV-K expression could be induced by several cytokines and growth factors induced by reactivation of CMV (Assinger et al., 2013).

Studies have suggested that persistent CMV infection initiates immunosenescence (Pawelec et al., 2009, 2005). One hallmark of latent CMV infection is substantial increment in number of CMV specific memory CD8+ T cells and decrease in CMV specific memory CD4+ T cells over time (Weltevrede, Eilers, de Melker, & van Baarle, 2016). This process of accumulation of CMV specific memory T cell during viral infection is known as memory inflation (Karrer et al., 2003). It has been suggested that CMV infection is linked with clonal extension of CD8+ T cells, an inverted CD4 to CD8 and increased numbers of CD8+CD28− T cells (Wikby et al., 2002). Furthermore, CMV infection may also inhibit the natural killer cell mediated lysis of infected cells (Wills et al., 2005) and modulates the natural killer cells receptor repertoire of natural killer and T cells (Gumá et al., 2004). In addition, several cross-sectional studies have reported that CMV has less effect on the naïve T cell pool than that of memory T cells (Mekker et al., 2012; Wertheimer et al., 2014).

2.10 HERV-K and Epstein Barr Virus

EBV infects 90% of the world's population. EBV is the first human virus related to oncogenesis (Epstein, Achong, & Barr, 1964). After initial primary infection, most individuals remain lifelong carriers of the disease without serious symptoms (Orem, Mbidde, Lambert, De Sanjose, & Weiderpass, 2007). Unlike CMV, EBV is not associated with immunosenescence (Geng & Wang, 2015). The important clinical impact of EBV infected people is an increased risk of lymphoproliferative disorders (Geng & Wang, 2015). However, various study suggested that EBV transactivates HERV-K (Hsiao et al., 2006; Sutkowski et al., 2004, 2001). Sutkowski et al., 2001 revealed that EBV activates *env* of HERV-K18 that possesses superantigen activity. Other viruses which are also been detected in various cancer along with CMV and HERV-K and EBV are Hepatitis B and C, human papillomavirus, human herpes virus 8 (Morales-Sánchez & Fuentes-Pananá, 2014; Parkin, 2006).

This study investigates whether or not HERV-K play a role in immunosenescence. We performed gene set enrichment analysis (GSEA) (also called as pathway analysis) to identify various biological processes/pathways that are significantly associated with HERV-K proviruses in nonagenarians. Group of seven young subjects will be used as control.

3 AIM AND OBJECTIVES

3.1 Aim

To investigate the role of HERV-K in immunosenescence and its underlying mechanisms.

3.2 Objectives

1. Identification of nonagenarian specific biological processes that are highly associated with HERV-K proviruses.
2. Identification of differentially expressed genes and HERV-K proviruses between nonagenarian and young control.
3. Identification of directionality (up or down-regulation) of significant biological processes through differential gene expression analysis between older and young subjects followed by over-representation analysis.

4 MATERIALS AND METHODS

4.1 Study populations

Study population consisted of seven young healthy laboratories personal aged between 26 and 32 (median 28) and seven nonagenarian (age 90-99) female from "The Vitality 90+ study". The decision to recruit seven individual is due to limited availability of funding resources. The recruitment and characterization of participants were accomplished as reported in the article by Goebeler, Jylhä, & Hervonen, 2003. A trained laboratory technician in the laboratory facilities collected blood samples. Blood samples were collected between 8:00 -12:00 am into EDTA containing tubes. Samples were directly placed into leucocyte separation on a Ficoll-Paque density gradient (Ficoll-Paque™ Premium (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). Peripheral blood mononuclear cell (PBMC) layer was collected for RNA extraction and suspended in 150 µl of RNAlater solution (Ambion Inc., Austin, TX, USA).

4.2 RNA extraction and sequencing

RNA used for RNA sequencing was cleansed using a miRNeasy mini kit (Qiagen, CA, USA) following manufacturer's protocol with on-column DNA digestion (Qiagen). The concentration and quality of the RNA was assessed with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE, USA). Agilent Bioanalyzer RNA nano chips (Agilent) were used to evaluate the integrity of total RNA and Qubit RNA –kit (Life Technologies) to quantitate RNA in samples. One µg RNA was used for ScriptSeq™ Complete Gold System (Epicentre) to ribodeplete rRNA and further for RNA-seq library preparation. SPRI beads (Agencourt AMPure XP, Beckman Coulter) were used for purification of RNAseq libraries. The library QC was evaluated on High Sensitivity chips by Agilent Bioanalyzer (Agilent). Paired-end

sequencing of RNAseq libraries was done using Illumina HiSeq technology with a minimum of 60 million-2x100bp paired-end reads per sample.

4.3 Data pre-processing

Raw reads were aligned to human genome reference build hg19 using TopHat v2.0.13 (Trapnell, Pachter, & Salzberg, 2009)with default parameters. Only uniquely mapped reads were considered in the transcript abundance estimation and to this end SAMtools (Li, 2011) was used to filter out reads mapping to multiple regions of the genome. The downstream analysis was all conducted using the tools in cufflinks2 v. 2.2.1 (Trapnell et al., 2012, 2013). The raw expression estimates were calculated using cuffquant and the expression were normalized using cuffnorm that gives the normalized read counts and the fragments per kilobase per million values (FPKM) for each gene as an output. The geometric normalization method was used which scales the read counts as well as the FPKM values according to procedure described (Hubbard et al., 2002). The annotation data for HERV-K(HML-2) was from (A. Subramanian et al., 2005).To ensure the robustness of the normalization, the expressions of HERV-K (HML-2) elements were quantified and normalized together with ENSEMBL v. 82 gene reference set (Cunningham et al., 2015; Hubbard et al., 2002).

4.4 Ethical Consideration

Study participants provided written informed consent. This study followed principles expressed in the declaration of Helsinki. The use of blood samples from nonagenarians and healthy controls was approved by the Ethics Committee of city of Tampere. (1592/403/1996).

4.5 Data cleaning

Only the genes and HERV-K proviruses that were expressed (normalized read count ≥ 16) in at least four subjects in each group (nonagenarian and young) were included in analysis. Number of HERV-K proviruses and human genes that survived the filtration criteria were 22 and 15912 respectively.

4.6 Gene Set Enrichment Analysis (GSEA)

Analysis of nonagenarian specific biological processes significantly associated with HERV-K proviruses was done using GSEA (Aravind Subramanian et al., 2005) implemented in R package clusterProfiler (version3.7). The analysis was done separately for nonagenarian and young subjects. Pearson correlation coefficients were calculated between all HERV-K elements and human genes. Genes were ranked by their correlation coefficients and used as input for GSEA. The Gene Ontology (GO) based biological processes, standardized annotation of genes and gene products (Ashburner et al., 2000) was used as gene sets (referred as GO terms here onwards). GO terms with size ranging from 25 to 500 member genes were included in the analysis. Statistical significance of GO term enrichment scores was accessed by calculating empirical P-values with 10000 permutations. P-values were corrected for multiple testing with Benjamini & Hochberg method (Benjamini & Hochberg, 1995).

4.7 Over-representation analysis

GSEA analysis was done using correlation scores calculated between human genes and provirus to profile nonagenarian specific biological process related GO terms. The approach, however, is not sufficient to elucidate the important question of whether biological pathways are up or downregulated among nonagenarian. For that, differential gene expression analysis was done between young and nonagenarian groups. Genes with adjusted P-value < 0.05 were considered significantly

differentiated between the compared groups. We performed over-representation analysis with only those significant genes that are hightly correlated with HERV-K provirus 1q22 (cor < -0.80 or > 0.80). GOstats R package (Beißbarth & Speed, 2004) was used for the over-representation analysis.

4.8 Methodological details

The advancement in biotechnology has led to generation of high throughput biological data. The main challenge for analysing such data is getting meaningful biological conclusion.

4.8.1 Gene expression

Genes encode for protein and proteins are essential for cellular functions. There are thousands of genes expressed in each particular type of cell that determines function of that particular type of cell. Gene expression involves two steps called transcription and translation. When genes are expressed, the genetic information on DNA is first transcribed to messenger RNA (mRNA). The mRNA molecules are then translated to a series of amino acids to build desired proteins. Regulation of gene expression plays a vital role in determining variety and amount of protein produced. There are various methods to quantify gene expression, for example microarrays (Agilent or Affymatrix) and RNA sequencing.

4.8.2 Gene expression data analysis pipeline

Aim of gene expression data analysis is to understand changes in gene activities across compared groups. Traditional methods focus on gene wise analysis that

involves analysis of one gene at a time. Gene expression data analysis pipeline involves following steps:

a. Biological hypothesis
b. Experimental design
c. Gene expression data generation
d. Data pre-processing
e. Differential gene expression analysis

Tradition approach had major limitations:

1. There can be large number of differentially expressed significant genes without unifying biological theme. Interpretation of these long gene lists can be quite a challenge.
2. Furthermore, no single gene might meet the threshold for statistical significance because of the noise (difference due to other factors than real biological difference) inherent to technology used.

Recently, the focus of Gene expression analysis has shifted from single gene to group of genes (gene set) that shares common biological theme. Certain phenotypes are often believed to be associated with multiple genes rather than a single gene (Aravind Subramanian et al., 2005). Furthermore, analyzing set of genes also averages out errors at single gene level and increases the statistical power by combining association signals from multiple genes in the same gene set. The approach is called as gene set analysis (GSA). GSA methods use gene sets that are defined based on the prior biological knowledge. Popular sources of gene sets include Gene Ontology (GO) (Beißbarth & Speed, 2004) terms, Molecular Signatures Database (MSigDB) (Liberzon et al., 2011)and Kyoto Encyclopedia of Genes and Genomes (KEGG) (H. Ogata et al., 1999).

4.8.3 Gene ontology

Gene ontology is structured and controlled vocabulary of functions (GO terms) and how these functions are related to each other. The GO Consortium currently has 3 organism databases - Fly Base (Drysdale & Consortium, 2008), Mouse Genome Informatics (Bult et al., 2008), and the *Saccharomyces* Genome Database (Ball et al., 2000). GO database is constantly revised and expanded as knowledge accumulates. Annotation in a GO term has a source (experimental evidence) and a database entry attributed to it. GO has three categories, i) molecular function (biochemical activity of a gene product, ii) biological process (biological function to which the gene or gene product contributes and iii) cellular component (the location where the gene product can be found or is active).

4.9 Gene set analysis methods

There are two major types of gene set analysis methods:

4.9.1 Over-representation analysis

Over-representation analysis, also called as enrichment analysis, is a technique for determining whether a list of gene sets (biological processes in our case) is present more than it would be expected in a subset of interesting genes (differentially expressed genes in our case). Arbitrary threshold is used to select interesting genes. For example, genes with adjusted p-value less than 0.05 can be considered as interesting genes. Although this method of GSA is quick and easy to implement, results are unstable as they depend on arbitrary threshold. Furthermore, there is massive reduction in sensitivity as the association measure is solely based on the count of the significantly associated genes (Khatri, Sirota, & Butte, 2012). DAVID, Ingenuity, GOstats, AmiGO and GeneGO are example of methods based on this approach.

4.9.2 Gene set analysis

Gene set analysis methods are evolved version of over-representation methods. These methods are threshold free. Meaning, all the genes included in the experiment is fed to the method and thus results are stable. The gene list is first ranked based on some gene level statistics like t-scores, correlation or p-values. The ranked gene list is then used as input for gene set analysis methods. In this study, we used gene set enrichment analysis (GSEA). GSEA is one of the most popular GSA method.

4.10 Analysis work flow

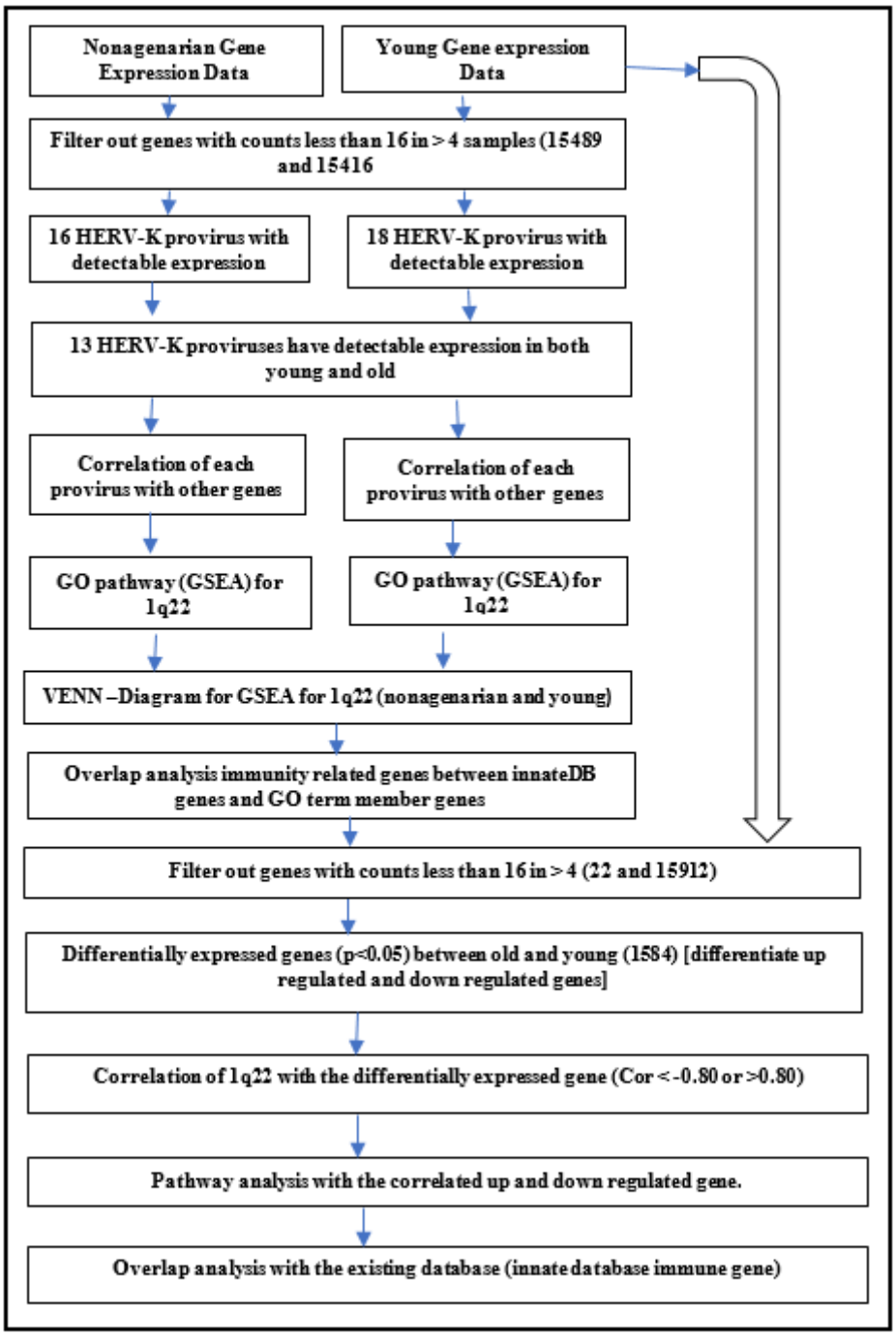

Figure 4-1 Analysis workflow.

5 RESULT

5.1 HERV-K proviruses expression analysis

We considered genes and proviruses detectable if the expression level is equal to or greater than normalized read count of 16. There were 13 proviruses with detectable expression level common in both nonagenarian and young. Three proviruses were specific to old and five proviruses was specific to young (Figure5-1).

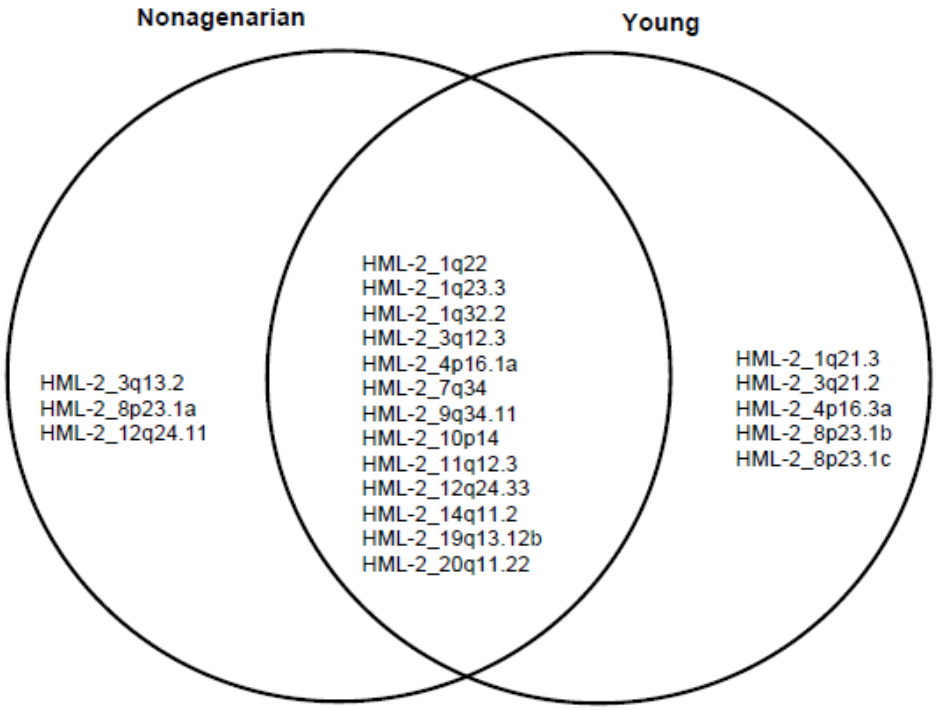

Figure 5-1 HERV-K proviruses with detectable expression level in nonagenarian and young subjects

We performed gene set enrichment analysis (GSEA) to identify young and older group specific GO based biological processes that are significantly associated with the proviruses. GSEA was performed in both nonagenarian and young group separately in order to identify age specific GO pathways. Table 5-1 summarizes expression levels and enriched GO terms in proviruses among nonagenarian and young. Median expression levels of proviruses were calculated in both young and nonagenarian groups (Table 5-1).

Table 5-1 Expression of HERV-K (HML-2) provirus in nonagenarian and young with GO pathway.

	Median expression level		Number of HERV-K(HML-2) expressed*		Number of GO term enriched	
HERV-K(HML-2) element	Nonagenarian	Young	Nonagenarian	Young	Nonagenarian	Young
1q22	373,71	279,40	7	7	805	46
1q23.3	94,50	93,55	7	7	495	35
1q32.2	40,58	42,30	7	7	744	96
3q12.3	786,72	948,35	7	7	81	211
4p16.1a	24,47	26,10	6	6	182	3
7q34	66,41	74,16	7	7	557	0
9q34.11	40,51	36,90	6	7	513	95
10p14	70,75	18,00	7	4	437	152
11q12.3	16,35	17,65	4	4	262	326
12q24.33	88,74	100,93	7	7	313	34
14q11.2	54,12	27,67	7	7	401	641
19q13.12b	123.07	145,24	7	7	530	112
20q11.22	27,73	21,91	5	6	232	252

***** Number of individuals where HERV-K (HML-2) is expressed (normalized read count ≥ 16)

Rest of the analysis will be focused only on provirus 1q22 because it is significantly differentially expressed between nonagenarian and young.

5.2 GSEA

GO pathway associated with provirus 1q22 was compared between nonagenarian and young individuals. There were 805 significant GO term enriched in nonagenarian and 46 significant GO term enriched in young. We focused on nonagenarian specific pathways i.e. 776 GO terms, which will be analyzed further.

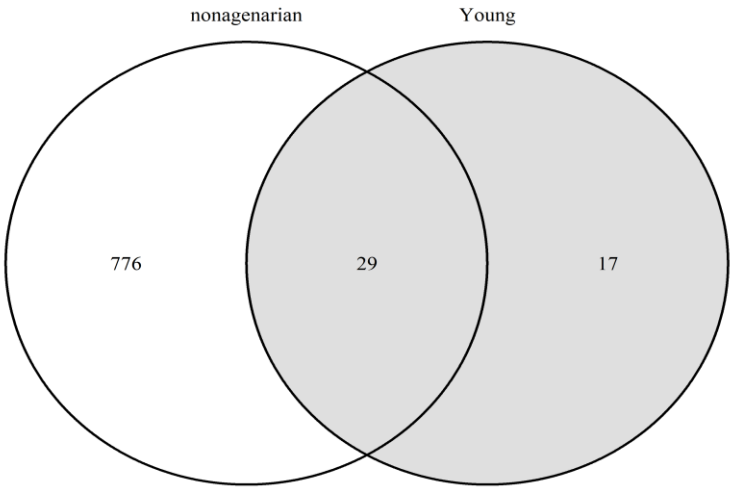

Figure 5-2 Distribution of provirus 1q22 related GO terms among nonagenarian and young individuals.

Table 5-2 Top 20 provirus 1q22 related GO terms specific to nonagenarians.

	ID	Description	enrichmentScore	pvalue	p.adjust
1	GO:0006521	regulation of cellular amino acid metabolic process	0,558059376	2,08751E-05	0,00054059
2	GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	0,559243566	2,10168E-05	0,00054059
3	GO:0002433	immune response-regulating cell surface receptor signaling pathway involved in phagocytosis	0,49232079	2,10371E-05	0,00054059
4	GO:0038096	Fc-gamma receptor signaling pathway involved in phagocytosis	0,49232079	2,10371E-05	0,00054059
5	GO:0033238	regulation of cellular amine metabolic process	0,525774017	2,10548E-05	0,00054059
6	GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	0,551476034	2,10886E-05	0,00054059
7	GO:0061418	regulation of transcription from RNA polymerase II promoter in response to hypoxia	0,516129041	2,11372E-05	0,00054059
8	GO:0002431	Fc receptor mediated stimulatory signaling pathway	0,491499357	2,11416E-05	0,00054059
9	GO:0038094	Fc-gamma receptor signaling pathway	0,493656878	2,11488E-05	0,00054059
10	GO:0044106	cellular amine metabolic process	0,446330969	2,12197E-05	0,00054059
11	GO:0002220	innate immune response activating cell surface receptor signaling pathway	0,426995167	2,12278E-05	0,00054059
12	GO:0007034	vacuolar transport	0,409165457	2,12395E-05	0,00054059
13	GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	0,441973548	2,12431E-05	0,00054059
14	GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	0,456991311	2,12504E-05	0,00054059
15	GO:0009308	amine metabolic process	0,440381559	2,12571E-05	0,00054059
16	GO:0014074	response to purine-containing compound	0,438915242	2,12571E-05	0,00054059
17	GO:0002223	stimulatory C-type lectin receptor signaling pathway	0,41561253	2,12608E-05	0,00054059
18	GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	0,454519951	2,12635E-05	0,00054059
19	GO:0043620	regulation of DNA-templated transcription in response to stress	0,418624822	2,12662E-05	0,00054059
20	GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	0,526230415	2,12666E-05	0,00054059

We investigated the significant GO terms with their relevance to immunity. This was achieved by investigating proportion of immunity related genes in each of the GO terms. We used innateDB database (Breuer et al., 2013) as reference for immunity related genes. We calculated simple overlap of immunity related genes between innateDB genes and GO term member genes. We found six GO terms with 50 percent immunity related genes.

Table 5-3 GO term with 50 percent of overlap with immunity gene GO term of 1q22 (nonagenarian specific)

ID	Description	Enrichment score	pvalue	p adjust	size	overlap
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	0,491	0,001	0,007	22	15
GO:0030593	neutrophil chemotaxis	0,405	0,002	0,014	23	12
GO:1990266	neutrophil migration	0,367	0,005	0,025	26	14
GO:0032755	positive regulation of interleukin-6 production	0,375	0,008	0,036	22	13
GO:0050766	positive regulation of phagocytosis	0,402	0,011	0,044	13	7
GO:0031663	lipopolysaccharide-mediated signaling pathway	0,376	0,012	0,047	21	12

5.3 Over-representation Analysis

GSEA provided nonagenarian specific profile of biological processes associated with provirus 1q22. However, directionality of immune related biological processes, which is the main question of this study, remains unanswered. In other words, whether immune related biological processes are up or down regulated is not inferred. We elucidated that information with over-representation analysis.

5.4 Differential expression analysis

Differential gene expression analysis was done to identify genes that are significantly differentially expressed between nonagenarian and young individuals. Fifteen hundred and eighty-four genes were differentially expressed between nonagenarian and young adults ($P < 0.05$). Among the differentially expressed genes, 673 were upregulated while 911 were down regulated in nonagenarian. The top 20 most significantly differentially expressed genes are shown in Table 5-4.

Table 5-4 Top 20 genes with the most significant differential expression

Rank	Ensembl Gene ID	Gene Symbol	Gene Description	P- value	Gene feature
1	ENSG00000173114	LRRN3	Leucine-rich repeat neuronal protein 3	1.327032e-07	down
2	ENSG00000104660	LEPROTL1	Leptin receptor overlapping transcript-like 1	2.168289e-07	up
3	ENSG00000174807	CD248	(Tumor endothelial marker 1) (CD antigen CD248)	2.296232e-07	down
4	ENSG00000166669	ATF7IP2	Activating transcription factor 7-interacting protein 2	5.223306e-07	up
5	ENSG00000115687	PASK	PAS domain-containing serine/threonine-protein kinase	8.643576e-07	down
6	ENSG00000164300	SERINC5	Serine incorporator 5	9.745879e-07	up
7	ENSG00000137054	POLR1E	DNA-directed RNA polymerase I subunit RPA49	1.863813e-06	up
8	ENSG00000047634	SCML1	Sex comb on midleg-like protein 1	1.895650e-06	down
9	ENSG00000103064	SLC7A6	Cationic amino acid transporter	1.953292e-06	up
10	ENSG00000129270	MMP28	Matrix metalloproteinase 28	2.474790e-06	down
11	ENSG00000136153	LMO7	LIM domain only protein 7	3.206203e-06	down
12	ENSG00000170365	SMAD1	Mothers against decapentaplegic homolog 1	3.390027e-06	up
13	ENSG00000112394	SLC16A10	Monocarboxylate transporter 10	3.877464e-06	down
14	ENSG00000157978	LDLRAP1	Low density lipoprotein receptor adapter protein 1	4.653414e-06	up
15	ENSG00000139998	RAB15	Ras-related protein Rab-15	4.792945e-06	down
16	ENSG00000154153	RETREG1	Reticulophagy regulator 1	5.507027e-06	down
17	ENSG00000136111	TBC1D4	C1 domain family member 4	5.675065e-06	down
18	ENSG00000116771	AGMAT	Agmatinase, mitochondrial, EC 3.5.3.11	6.487382e-06	down
19	ENSG00000143167	GPA33	Cell surface A33 antigen	6.885712e-06	down
20	ENSG00000182183	SHISAL2A	Protein shisa-like-2A	7.095193e-06	down

Among the proviruses, HML-2_1q22, HML-2_10p14, and HML-2_8p23.1a (hereafter referred as 1q22, 10p14, and 8p23.1a respectively) were significantly differentially expressed. We focus only on those genes that are highly positively or negatively correlated with the differentially expressed proviruses.

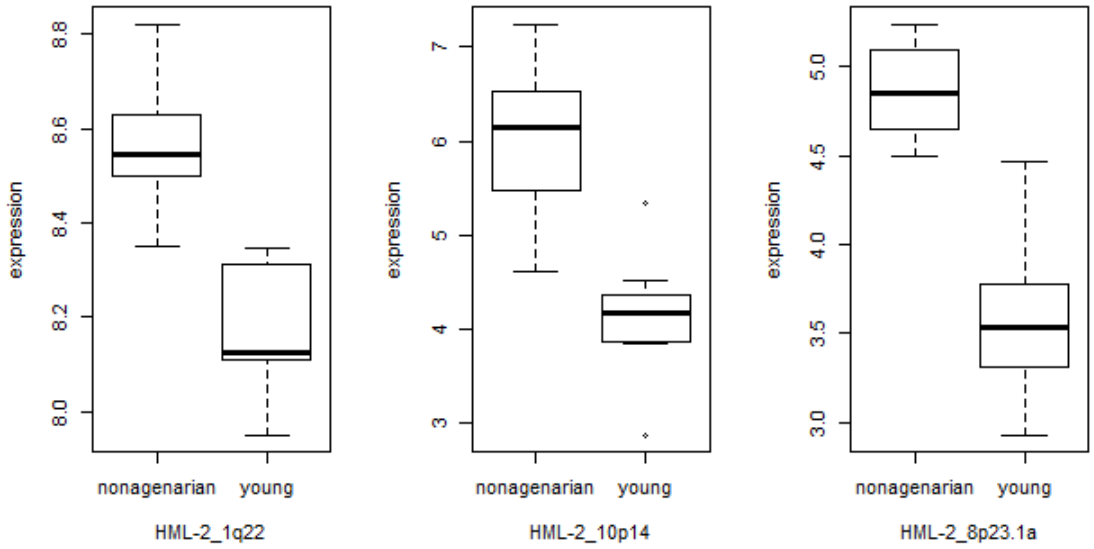

Figure 5-3 Expression levels of differentially expressed proviruses.

5.5 1q22 specific over-representation analysis

5.5.1 Correlation analysis

Based on our results, 347/1584 differentially expressed genes were highly correlated (>0.80) with provirus 1q22 (positively: 178, negatively: 169). Table 5-5 shows the top 20 of the highly correlated genes.

Table 5-5 Top 20 differentially expressed genes that are highly correlated with provirus 1q22

Rank	Ensembl Gene ID	Gene Symbol	Gene Description	P-value	adj.P-Value	Gene feature
1	ENSG00000104660	LEPROTL1	leptin receptor overlapping transcript like 1	2.168289e-07	0.001227489	up
2	ENSG00000103064	SLC7A6	solute carrier family 7 member 6	1.953292e-06	0.003480549	up
3	ENSG00000136153	LMO7	LIM domain	3.206203e-06	0.004530489	down
4	ENSG00000182183	SHISAL2A	Protein shisa-like-2A	7.095193e-06	0.005496089	down
5	ENSG00000184384	MAML2	mastermind like transcriptional coactivator 2	8.945131e-06	0.005948243	up
6	ENSG00000091409	ITGA6	integrin subunit alpha 6	9.798813e-06	0.005948243	down
7	ENSG00000113319	RASGRF2	Ras protein specific guanine nucleotide releasing factor 2	9.819542e-06	0.005948243	down
8	ENSG00000124374	PAIP2B	poly(A) binding protein interacting protein 2B	1.059053e-05	0.006065727	down
9	ENSG00000104689	TNFRSF10A	TNF receptor superfamily member 10a	1.480013e-05	0.006660920	up
10	ENSG00000182568	SATB1	SATB homeobox 1	1.495249e-05	0.006660920	up
11	ENSG00000166313	APBB1	amyloid beta precursor protein binding family B member 1	1.714967e-05	0.006718107	down
12	ENSG00000067533	RRP15	ribosomal RNA processing 15 homolog	1.919876e-05	0.006810315	up
13	ENSG00000140743	CDR2	cerebellar degeneration related protein 2	1.948537e-05	0.006810315	up
14	ENSG00000158987	RAPGEF6	Rap guanine nucleotide exchange factor 6	2.050003e-05	0.006881323	up
15	ENSG00000149308	NPAT	nuclear protein, coactivator of histone transcription	2.739645e-05	0.007606555	up
16	ENSG00000196960	RP11-192P3.5	N/A	2.893308e-05	0.007606555	down
17	ENSG00000081059	TCF7	transcription factor 7	3.046864e-05	0.007644637	down
18	ENSG00000162408	NOL9	nucleolar protein 9	3.520706e-05	0.007789457	up
19	ENSG00000152465	NMT2	N-myristoyltransferase 2	3.849521e-05	0.007789457	up
20	ENSG00000134709	HOOK1	hook microtubule tethering protein 1	3.986067e-05	0.007789457	down

5.5.2 Overlap with existing database

We investigated overlap between the interesting genes identified in our study with existing literature. For that purpose, we analyzed genes that are directly involved in immunological processes from innate database (Breuer et al., 2013). We identified that 30/347 1q22 correlated genes have also been reported as immune related genes in innate database. Twenty-one of the genes were positively correlated with 1q22, while nine of them were negatively correlated (Figure 5-4).

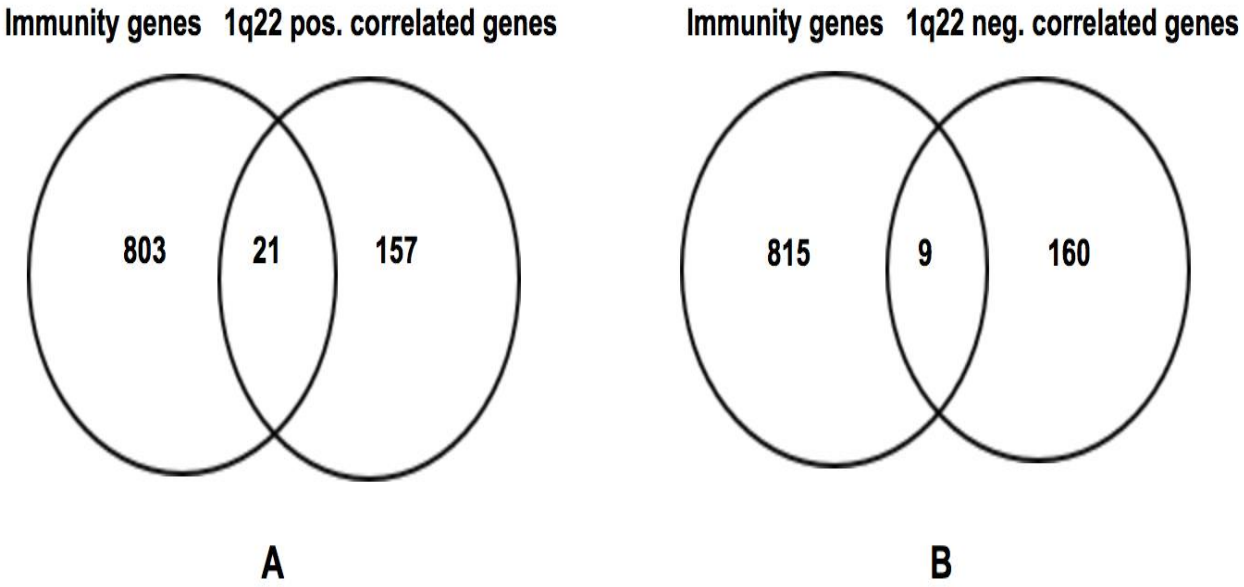

Figure 5-4 Overlap analysis between differentially expressed genes that are highly correlated with 1q22 and immune genes in innateDB

A. Overlap between immunity genes in innateDB and differentially expressed genes that are positively correlated with 1q22. B. Overlap between immunity genes in innateDB and differentially expressed genes that are negatively correlated with 1q22.

5.5.3 Pathways analysis with positively correlated genes

We identified 265 biological processes related GO terms significantly enriched (fdr < 0.05) in 1q22 positively correlated genes. Table 5-6 shows the top 20 GO terms.

Table 5-6 Top 20 biological processes related GO terms associated with differentially expressed genes that are positively correlated with provirus 1q22.

	GO ID	P value	FDR	Description
1	GO:0002446	2.182140e-19	6.718816e-16	neutrophil mediated immunity
2	GO:0002444	3.540878e-19	6.718816e-16	myeloid leukocyte mediated immunity
3	GO:0043312	7.777171e-19	8.899300e-16	neutrophil degranulation
4	GO:0002283	9.380027e-19	8.899300e-16	neutrophil activation involved in immune response
5	GO:0043299	1.509480e-18	1.097400e-15	leukocyte degranulation
6	GO:0042119	1.735020e-18	1.097400e-15	neutrophil activation
7	GO:0002275	2.262870e-18	1.182183e-15	myeloid cell activation involved in immune response
8	GO:0036230	2.492086e-18	1.182183e-15	granulocyte activation
9	GO:0002366	9.861345e-18	4.158200e-15	leukocyte activation involved in immune response
10	GO:0002263	1.193301e-17	4.528579e-15	cell activation involved in immune response
11	GO:0001775	2.189982e-17	7.555437e-15	cell activation
12	GO:0002274	2.836513e-17	8.970471e-15	myeloid leukocyte activation
13	GO:0016192	7.809046e-17	2.279641e-14	vesicle-mediated transport
14	GO:0006887	8.936820e-17	2.422517e-14	exocytosis
15	GO:0045321	1.442574e-16	3.649712e-14	leukocyte activation
16	GO:0045055	2.127137e-16	5.045304e-14	regulated exocytosis
17	GO:0002252	5.135083e-16	1.146332e-13	immune effector process
18	GO:0002443	5.429146e-15	1.144645e-12	leukocyte mediated immunity
19	GO:0032940	4.373867e-13	8.736223e-11	secretion by cell
20	GO:0046903	5.623539e-13	1.067066e-10	secretion

5.5.4 Pathways analysis with negatively correlated genes

We identified 39 biological processes related GO terms significantly enriched (fdr < 0.05) in 1q22 negatively correlated genes. Table 5-7 shows the top 20 GO terms.

Table 5-7 Top 20 biological processes related GO terms associated with differentially expressed genes that are negatively correlated with provirus 1q22.

	GO ID	P value	FDR	Description
1	GO:0016070	2.204722e-07	0.0005531647	RNA metabolic process
2	GO:0090304	4.399352e-06	0.0043406906	nucleic acid metabolic process
3	GO:0010467	5.190144e-06	0.0043406906	gene expression
4	GO:0006355	2.073308e-05	0.0074969260	regulation of transcription, DNA-templated
5	GO:0044271	2.237064e-05	0.0074969260	cellular nitrogen compound biosynthetic process
6	GO:1903506	2.428898e-05	0.0074969260	regulation of nucleic acid-templated transcription
7	GO:0051252	2.452295e-05	0.0074969260	regulation of RNA metabolic process
8	GO:2001141	2.578676e-05	0.0074969260	regulation of RNA biosynthetic process
9	GO:0006139	2.832718e-05	0.0074969260	nucleobase-containing compound metabolic process
10	GO:0006351	3.060941e-05	0.0074969260	transcription, DNA-templated
11	GO:0097659	3.462429e-05	0.0074969260	nucleic acid-templated transcription
12	GO:0046483	3.820915e-05	0.0074969260	heterocycle metabolic process
13	GO:0032774	3.884418e-05	0.0074969260	RNA biosynthetic process
14	GO:0034641	6.695794e-05	0.0106228514	cellular nitrogen compound metabolic process
15	GO:0010556	7.010868e-05	0.0106228514	regulation of macromolecule biosynthetic process
16	GO:0009059	7.029678e-05	0.0106228514	macromolecule biosynthetic process
17	GO:0010468	7.220962e-05	0.0106228514	regulation of gene expression
18	GO:0034654	7.632368e-05	0.0106228514	nucleobase-containing compound biosynthetic process
19	GO:2000112	8.044407e-05	0.0106228514	regulation of cellular macromolecule biosynthetic process
20	GO:0043170	8.525449e-05	0.0106951756	macromolecule metabolic process

6 DISCUSSION

Immunosenescence leads to increased incident and severity of various infections, reactivation of latent viruses and decline in vaccine responses among the elderly. As the proportion of older population is rising sharply across the globe, immunosenescence is an emerging public health problem. Fundamental understanding of immunosenescence is thus vital for prevention, control, and treatment of ageing-related diseases among the vulnerable population. Immunosenescence has been known to be induced by pathogens like CMV. Interestingly, HERV-K has been shown to be co-expressed with CMV in several diseases and cancers (See Section 1). Thus, it is rational to hypothesize that HERV-K might have role in induction of immunosenescence.

GSEA in our study revealed various nonagenarian specific GO terms that are correlated with provirus 1q22. This include several immunity related GO terms such as antigen processing and presentation of exogenous peptide antigen via MHC class I and class II, innate immune response activating cell surface receptor signalling pathway and Fc-gamma receptor-signalling pathway involved in phagocytosis. Metabolic process pathway such as regulation of cellular amino acid metabolic process, regulation of cellular amine metabolic process, cellular amine metabolic process, vascular transport, and amine metabolic process were also highly enriched in nonagenarian (Table 5-2). Overlap analysis between innateDB genes and 1q22 GO terms revealed six GO terms that consist at least 50% immunity related genes. The GO terms were related to Myd88 dependent toll like receptor signalling pathway, neutrophil chemotaxis, and migration, positive regulation of interleukin-6 production and phagocytosis. Study by Kovacs et al., 2010 suggested that interleukin-6 is a crucial factor in maintaining the age-related defects in macrophages through transformation of proinflammatory cytokines. A study showed that neutrophils from elderly migrated with similar chemokinesis as in young but with less accuracy (Sapey et al., 2014).

Our differential gene expression analysis revealed several human genes and three HERV-K proviruses (1q22, 10p14, and 8p23.1a) significantly differentially expressed between nonagenarian and young. However, as the focus of this study was to investigate role of HERV-K proviruses in immunosenescence, we analysed only the genes that are highly correlated with the proviruses. We hypothesized that the proviruses play role in immunosenescence by regulating the highly correlated genes. Interestingly, 30 of the differentially expressed genes that were highly correlated with 1q22 have also been reported as immune related genes in innate database (Breuer et al., 2013). This suggests that provirus 1q22 and correlated genes potentially play role in immunosenescence. Among the top 20 correlated genes presented in Table 5-5 (see Results Section), genes ITGA6, TCF7 and TNFRSF10A are present as immune related genes in innate database. ITGA6 and TCF7 are down regulated while TNFRSF10A is up regulated in nonagenarian. ITGA6 is expressed in immune cell such as appendix, spleen, and lymph node. Studies have shown elevated level of ITGA6 in different tumour cells (Jin et al., 2018; Kwon et al., 2013; Yamakawa, Kaneda, Saito, Ichihara, & Morishita, 2012). Furthermore, studies have shown that immunosenescence may contribute to increasing incidence of most cancers among elderly (Pawelec, 2017). TCF7 plays an important role in providing protective immunity after resolution of acute infections (Jeannet et al., 2010; Utzschneider et al., 2016; Zhou et al., 2010). TNFSF10 plays role in immune surveillance (Shin, Kang, Lee, & Kim, 2016). Immune surveillance is a mechanism of immune system to recognize and destroy invading pathogens but also host cell that become cancerous (Swann et al., 2007). Gene LEPROTL1, the most significant gene with provirus 1q22, negatively regulates growth hormone receptor (Touvier et al., 2009). Growth hormone is an important hormone, which effect immune system that decreases with age (Taub, Murphy, & Longo, 2010).

Various studies have shown that provirus 1q22 is upregulated in HIV patients (Brinzevich et al., 2014; Laderoute et al., 2007; Laderoute, Larocque, Giulivi, & Diaz-Mitoma, 2015). HIV causes immune suppression. If provirus 1q22 is upregulated in

HIV patients pertaining to immune suppression, it is rational to think that the provirus might also be upregulated in older people for similar reason. Furthermore, provirus 1q22 are able to encode complete *env* protein (Flockerzi et al., 2008) that has superantigenic property (Sutkowski et al., 2004). Superantigens cause non-specific activation of T cell causing polyclonal, antigen receptor independent activation of lymphocytes (Huber, Hsu, & Sutkowski, 1996). Expression of the superantigen is elevated in autoimmune diseases like rheumatoid arthritis (Sicat, Sutkowski, & Huber, 2005). Hence, it is possible that the superantigen contritbes to the development of autoimmune inflammatatory disease which goes alongside of the process of immunosenescence in elderly.

Two hundred and sixty five biological processes related GO terms were identified to be significantly enriched in genes positively correlated with 1q22. Similarly, we identified 39 biological processes related GO terms significantly enriched in genes negatively correlated with 1q22. As the three proviruses are upregulated in nonagenarian, genes that are positively correlated with proviruses are upregulated and those that are negatively correlated are downregulated in nonagenarian. The upregulated biological pathways are related to response to wound healing, phagocytosis, cell adhesion, immunity (eg., neutrophil mediated immunity, and myeloid leukocyte mediated immunity) regulation of vesicle mediated transportion, positive regualtion of endocytosis and regualtion of cell to cell adhesion. Studies have shown that there is no reduction in the neutrophil number with age (Butcher et al., 2000). Thus, there is no loss in the ability to generate neutrophil mediated response towards infections (Beerman et al., 2010). Furthermore, a study reported that larger number of neutrophils are needed for wound healing in older mice than in younger mice (Nishio, Okawa, Sakurai, & Isobe, 2008). Pathways that are seen to be upregulated in elderly are related with neutrophil mediated immunity. Pathways related to lymphocyte differentiation and proliferation are found to be downregulated in elderly, indicating lower response of aged lymphoid cell in elderly. It has been

suggested that reduced capability of lymphocyte to proliferate is one of the crucial age related change in immune system (Linton & Dorshkind, 2004).

This study is based on PBMC as it contains important cell type such as lymphocytes and monocytes needed for the analysis. Furthermore, the study uses freshly isolated PBMCs without any vitro activation, which probably represents the true in-vivo situation and therefore be useful to investigate the significance of the various effects of HERV-K.

7 CONCLUSION

We performed transcription-based analysis to uncover potential role of HERV-K proviruses in immunosenescence. GSEA analysis was done using correlation scores calculated between human genes and provirus to profile nonagenarian specific biological process related GO terms. We further, analysed differntial gene expression between young and nonagenarian groups. We also performed over-representation analysis (Gostats) with only those significant genes that are hightly correlated with HERV-K provirus 1q22 (cor < -0.80 or > 0.80). The study identified several immunity related genes and biological processes significantly associated with HERV-K provirus 1q22. Our results suggest that proviruse 1q22 is potential factor in mediating immunosenescence. Therefore, further research with larger datasets with gender variation is needed for deeper understanding.

8 REFERENCES

Aggarwal, S., & Gupta, S. (1998). Increased apoptosis of T cell subsets in aging humans: altered expression of Fas (CD95), Fas ligand, Bcl-2, and Bax. *Journal of Immunology (Baltimore, Md. : 1950)*.

Agrawal, A., Agrawal, S., Cao, J.-N., Su, H., Osann, K., & Gupta, S. (2007). Altered Innate Immune Functioning of Dendritic Cells in Elderly Humans: A Role of Phosphoinositide 3-Kinase-Signaling Pathway. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.178.11.6912

Aiello, A. E., Chiu, Y. L., & Frasca, D. (2017). How does cytomegalovirus factor into diseases of aging and vaccine responses, and by what mechanisms? *GeroScience*. https://doi.org/10.1007/s11357-017-9983-9

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*. https://doi.org/10.1038/75556

Assinger, A., Yaiw, K. C., Göttesdorfer, I., Leib-Mösch, C., & Söderberg-Nauclér, C. (2013). Human Cytomegalovirus (HCMV) induces Human Endogenous Retrovirus (HERV) transcription. *Retrovirology*. https://doi.org/10.1186/1742-4690-10-132

Ball, C. a, Dolinski, K., Dwight, S. S., Harris, M. a, Issel-Tarver, L., Kasarskis, a, … Cherry, J. M. (2000). Integrating functional genomic information into the Saccharomyces genome database. *Nucleic Acids Research*. https://doi.org/gkd065 [pii]

Bannert, N., & Kurth, R. (2004). Retroelements and the human genome: New perspectives on an old relation. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0404838101

Barbulescu, M., Turner, G., Seaman, M. I., Deinard, A. S., Kidd, K. K., & Lenz, J. (1999). Many human endogenous retrovirus K (HERV-K) proviruses are unique to

humans. *Current Biology*. https://doi.org/10.1016/S0960-9822(99)80390-X

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., & Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1000834107

Beißbarth, T., & Speed, T. P. (2004). GOstat: Find statistically overrepresented Gene Ontologies with a group of genes. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bth088

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. https://doi.org/10.2307/2346101

Bergallo, M., Galliano, I., Montanari, P., Gambarino, S., Mareschi, K., Ferro, F., … Ravanini, P. (2015). CMV induces HERV-K and HERV-W expression in kidney transplant recipients. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2015.04.018

Bittner, J. J. (1936). Some possible effects of nursing on the mammary gland tumor incidence in mice. *Science*. https://doi.org/10.1126/science.84.2172.162

Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., … Lynn, D. J. (2013). InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks1147

Brinzevich, D., Young, G. R., Sebra, R., Ayllon, J., Maio, S. M., Deikus, G., … Mulder, L. C. F. (2014). HIV-1 Interacts with Human Endogenous Retrovirus K (HML-2) Envelopes Derived from Human Primary Lymphocytes. *Journal of Virology*. https://doi.org/10.1128/JVI.00669-14

Bulati, M., Pellicanò, M., Vasto, S., & Colonna-Romano, G. (2008). Understanding

ageing: Biomedical and bioengineering approaches, the immunologic view. *Immunity and Ageing*. https://doi.org/10.1186/1742-4933-5-9

Bult, C. J., Eppig, J. T., Kadin, J. A., Richardson, J. E., Blake, J. A., Airey, M. T., … Zhu, Y. (2008). The Mouse Genome Database (MGD): Mouse biology and model systems. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm961

Büscher, K., Hahn, S., Hofmann, M., Trefzer, U., Özel, M., Sterry, W., … Denner, J. (2006). Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. *Melanoma Research*. https://doi.org/10.1097/01.cmr.0000215031.07941.ca

Büscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., & Denner, J. (2005). Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Research*. https://doi.org/10.1158/0008-5472.CAN-04-2983

Butcher, S., Chahel, H., & Lord, J. M. (2000). Ageing and the neutrophil: No appetite for killing? *Immunology*. https://doi.org/10.1046/j.1365-2567.2000.00079.x

Campos, C., Pera, A., Lopez-Fernandez, I., Alonso, C., Tarazona, R., & Solana, R. (2014). Proinflammatory status influences NK cells subsets in the elderly. *Immunology Letters*. https://doi.org/10.1016/j.imlet.2014.06.015

Cegolon, L., Salata, C., Weiderpass, E., Vineis, P., Palù, G., & Mastrangelo, G. (2013). Human endogenous retroviruses and cancer prevention: Evidence and prospects. *BMC Cancer*. https://doi.org/10.1186/1471-2407-13-4

Chatta, G. S., & Dale, D. C. (1996). Aging and haemopoiesis. Implications for treatment with haemopoietic growth factors. *Drugs Aging*. https://doi.org/10.2165/00002512-199609010-00004

Chou, J. P., & Effros, R. B. (2013). T cell replicative senescence in human aging. *Current Pharmaceutical Design*. https://doi.org/10.2174/1381612811319090016

Contreras-Galindo, R., Kaplan, M. H., Contreras-Galindo, A. C., Gonzalez-Hernandez,

M. J., Ferlenghi, I., Giusti, F., … Markovitz, D. M. (2012). Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *Journal of Virology*. https://doi.org/10.1128/JVI.00602-11

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gku1010

Davalos, A. R., Coppe, J. P., Campisi, J., & Desprez, P. Y. (2010). Senescent cells as a source of inflammatory factors for tumor progression. *Cancer and Metastasis Reviews*. https://doi.org/10.1007/s10555-010-9220-9

de Flora, S., & Bonanni, P. (2011). The prevention of infection-associated cancers. *Carcinogenesis*. https://doi.org/10.1093/carcin/bgr054

Depil, S., Roche, C., Dussart, P., & Prin, L. (2002). Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. *Leukemia*. https://doi.org/10.1038/sj.leu.2402355

Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., & Heidmann, T. (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Research*. https://doi.org/10.1101/gr.5565706

Drysdale, R., & Consortium, F. (2008). FlyBase : a database for the Drosophila research community. BT - Drosophila : methods and protocols. In *Drosophila : methods and protocols*. https://doi.org/10.1007/978-1-59745-583-1_3

Epstein, M. A., Achong, B. G., & Barr, Y. M. (1964). VIRUS PARTICLES IN CULTURED LYMPHOBLASTS FROM BURKITT'S LYMPHOMA. *The Lancet*. https://doi.org/10.1016/S0140-6736(64)91524-7

Fietta, A., Merlini, C., De Bernardi, P. M., Gandola, L., Piccioni, P. D., & Grassi, C. (1993). Non specific immunity in aged healthy subjects and in patients with chronic bronchitis. *Aging Milano*.

Flockerzi, A., Ruggieri, A., Frank, O., Sauter, M., Maldener, E., Kopper, B., … Mayer, J. (2008). Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissues and the need for a HERV Transcriptome Project. *BMC Genomics*. https://doi.org/10.1186/1471-2164-9-354

Flores, K. G., Li, J., Sempowski, G. D., Haynes, B. F., & Hale, L. P. (1999). Analysis of the human thymic perivascular space during aging. *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI7558

Franceschi, C., Bonafè, M., & Valensin, S. (2000). Human immunosenescence: The prevailing of innate immunity, the failing of clonotypic immunity, and the filling of immunological space. *Vaccine*. https://doi.org/10.1016/S0264-410X(99)00513-7

Franceschi, C., & Campisi, J. (2014). Chronic inflammation (Inflammaging) and its potential contribution to age-associated diseases. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. https://doi.org/10.1093/gerona/glu057

Freimanis, G., Hooley, P., Ejtehadi, H. D., Ali, H. A., Veitch, A., Rylance, P. B., … Nelson, P. N. (2010). A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: Investigating mechanisms of pathogenesis. *Clinical and Experimental Immunology*. https://doi.org/10.1111/j.1365-2249.2010.04110.x

Fülöp, T., Fóris, G., Wórum, I., & Leövey, A. (1985). Age-dependent alterations of Fc gamma receptor-mediated effector functions of human polymorphonuclear leucocytes. *Clinical and Experimental Immunology*.

Geng, L., & Wang, X. (2015). Epstein-Barr Virus-associated lymphoproliferative disorders: Experimental and clinical developments. *International Journal of Clinical and Experimental Medicine*.

George, A. J. T., & Ritter, M. A. (1996). Thymic involution with ageing: Obsolescence or good housekeeping? *Immunology Today*. https://doi.org/10.1016/0167-5699(96)80543-3

Goebeler, S., Jylhä, M., & Hervonen, A. (2003). Medical history, cognitive status and

mobility at the age of 90. A population-based study in Tampere, Finland. *Aging Clinical and Experimental Research*. https://doi.org/10.1007/BF03324494

Golan, M., Hizi, A., Resau, J. H., Yaal-Hahoshen, N., Reichman, H., Keydar, I., & Tsarfaty, I. (2008). Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia (New York, N.Y.)*. https://doi.org/10.1593/neo.07986

Gomez, C. R., Nomellini, V., Faunce, D. E., & Kovacs, E. J. (2008). Innate immunity and aging. *Experimental Gerontology*. https://doi.org/10.1016/j.exger.2008.05.016

Gonzalez-Hernandez, M. J., Swanson, M. D., Contreras-Galindo, R., Cookinham, S., King, S. R., Noel, R. J., … Markovitz, D. M. (2012). Expression of Human Endogenous Retrovirus Type K (HML-2) Is Activated by the Tat Protein of HIV-1. *Journal of Virology*. https://doi.org/10.1128/JVI.07215-11

Gumá, M., Angulo, A., Vilches, C., Gómez-Lozano, N., Malats, N., & López-Botet, M. (2004). Imprint of human cytomegalovirus infection on the NK cell receptor repertoire. *Blood*. https://doi.org/10.1182/blood-2004-05-2058

Gupta, S. (2014). Role of dendritic cells in innate and adaptive immune response in human aging. *Experimental Gerontology*. https://doi.org/10.1016/j.exger.2013.12.009

Hahn, S., Ugurel, S., Hanschmann, K.-M., Strobel, H., Tondera, C., Schadendorf, D., … Löwer, R. (2008). Serological Response to Human Endogenous Retrovirus K in Melanoma Patients Correlates with Survival Probability. *AIDS Research and Human Retroviruses*. https://doi.org/10.1089/aid.2007.0286

Hoffe, S., & Balducci, L. (2012). Cancer and Age: General Considerations. *Clinics in Geriatric Medicine*. https://doi.org/10.1016/j.cger.2011.09.001

Hohn, O., Hanke, K., & Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Frontiers in Oncology*. https://doi.org/10.3389/fonc.2013.00246

Hsiao, F. C., Lin, M., Tai, A., Chen, G., & Huber, B. T. (2006). Cutting edge: Epstein-Barr virus transactivates the HERV-K18 superantigen by docking to the human complement receptor 2 (CD21) on primary B cells. *Journal of Immunology (Baltimore, Md. : 1950)*. https://doi.org/10.4049/jimmunol.177.4.2056

Huang, W. J., Liu, Z. C., Wei, W., Wang, G. H., Wu, J. G., & Zhu, F. (2006). Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2006.01.007

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., … Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Res*. https://doi.org/10.1093/nar/30.1.38

Huber, B. T., Hsu, P. N., & Sutkowski, N. (1996). Virus-encoded superantigens. *Microbiological Reviews*.

Imai, K., Matsuyama, S., Miyake, S., Suga, K., & Nakachi, K. (2000). Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. *Lancet*. https://doi.org/10.1016/S0140-6736(00)03231-1

Iwabuchi, H., Kakihara, T., Kobayashi, T., Imai, C., Tanaka, A., Uchiyama, M., & Fukuda, T. (2004). A Gene Homologous to Human Endogenous Retrovirus Overexpressed in Childhood Acute Lymphoblastic Leukemia. *Leukemia & Lymphoma*. https://doi.org/10.1080/10428190412331272758

Jean Beltran, P. M., & Cristea, I. M. (2014). The life cycle and pathogenesis of human cytomegalovirus infection: Lessons from proteomics. *Expert Review of Proteomics*. https://doi.org/10.1586/14789450.2014.971116

Karlsson, H., Bachmann, S., Schroder, J., McArthur, J., Torrey, E. F., & Yolken, R. H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proceedings of the National Academy of Sciences*.

https://doi.org/10.1073/pnas.061021998

Karreth, F. A., Tay, Y., Perna, D., Ala, U., Tan, S. M., Rust, A. G., … Pandolfi, P. P. (2011). In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*. https://doi.org/10.1016/j.cell.2011.09.032

Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1002375

Kirkwood, T. B. L. (1995). The Evolution of Aging. *Reviews in Clinical Gerontology*. https://doi.org/10.1017/S0959259800003932

Kleiman, A., Senyuta, N., Tryakin, A., Sauter, M., Karseladze, A., Tjulandin, S., … Mueller-Lantzsch, N. (2004). HERV-K(HML-2) GAG/ENV antibodies as indicator for therapy effect in patients with germ cell tumors. *International Journal of Cancer*. https://doi.org/10.1002/ijc.11649

Kohler, S., Wagner, U., Pierer, M., Kimmig, S., Oppmann, B., Möwes, B., … Thiel, A. (2005). Post-thymic in vivo proliferation of naive CD4+ T cells constrains the TCR repertoire in healthy human adults. *European Journal of Immunology*. https://doi.org/10.1002/eji.200526181

Kovacs, E. J., Gomez, C. R., Karavitis, J., Palmer, J. L., Faunce, D. E., Ramirez, L., & Nomellini, V. (2010). Interleukin-6 contributes to age-related alteration of cytokine production by macrophages. *Mediators of Inflammation*. https://doi.org/10.1155/2010/475139

Laderoute, M. P., Giulivi, A., Larocque, L., Bellfoy, D., Hou, Y., Wu, H. X., … Diaz-Mitoma, F. (2007). The replicative activity of human endogenous retrovirus K102 (HERV-K102) with HIV viremia. *AIDS*. https://doi.org/10.1097/QAD.0b013e3282f14d64

Laderoute, M. P., Larocque, L. J., Giulivi, A., & Diaz-Mitoma, F. (2015). Further

Evidence that Human Endogenous Retrovirus K102 is a Replication Competent Foamy Virus that may Antagonize HIV-1 Replication. *The Open AIDS Journal*. https://doi.org/10.2174/1874613601509010112

Lancini, D., Faddy, H. M., Flower, R., & Hogan, C. (2014). Cytomegalovirus disease in immunocompetent adults. *Medical Journal of Australia*. https://doi.org/10.5694/mja14.00183

Lang, P. O., Govind, S., Mitchell, W. A., Kenny, N., Lapenna, A., Pitts, D., & Aspinall, R. (2010). Influenza vaccine effectiveness in aged individuals: The role played by cell-mediated immunity. *European Geriatric Medicine*. https://doi.org/10.1016/j.eurger.2010.07.002

Lang, P. O., Govind, S., Mitchell, W. A., Siegrist, C. A., & Aspinall, R. (2011). Vaccine effectiveness in older individuals: What has been learned from the influenza-vaccine experience. *Ageing Research Reviews*. https://doi.org/10.1016/j.arr.2010.09.005

Lehrer, R. I., Ganz, T., Selsted, M. E., Babior, B. M., & Curnutte, J. T. (1988). Neutrophils and host defense. *Annals of Internal Medicine*. https://doi.org/10.7326/0003-4819-109-2-127

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btr509

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btr260

Linton, P. J., & Dorshkind, K. (2004). Age-related changes in lymphocyte development and function. *Nature Immunology*. https://doi.org/10.1038/ni1033

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*. https://doi.org/10.1016/j.cell.2013.05.039

Lung, T. L., Saurwein-Teissl, M., Parson, W., Schönitzer, D., & Grubeck-Loebenstein, B. (2000). Unimpaired dendritic cells can be derived from monocytes in old age and can mobilize residual function in senescent T cells. *Vaccine*. https://doi.org/10.1016/S0264-410X(99)00494-6

McElhaney, J. E., & Effros, R. B. (2009). Immunosenescence: what does it mean to health outcomes in older adults? *Current Opinion in Immunology*. https://doi.org/10.1016/j.coi.2009.05.023

Mekker, A., Tchang, V. S., Haeberli, L., Oxenius, A., Trkola, A., & Karrer, U. (2012). Immune Senescence: Relative Contributions of Age and Cytomegalovirus Infection. *PLoS Pathogens*. https://doi.org/10.1371/journal.ppat.1002850

Morales-Sánchez, A., & Fuentes-Pananá, E. M. (2014). Human viruses and cancer. *Viruses*. https://doi.org/10.3390/v6104047

Muster, T., Waltenberger, A., Grassauer, A., Hirschl, S., Caucig, P., Romirer, I., … Wolff, K. (2003). An Endogenous Retrovirus Derived from Human Melanoma Cells. *Cancer Research*. https://doi.org/Published December 2003

Nelson, P. N., Hooley, P., Roden, D., Ejtehadi, H. D., Rylance, P., Warren, P., … Murray, P. G. (2004). Human endogenous retroviruses: Transposable elements with potential? *Clinical and Experimental Immunology*. https://doi.org/10.1111/j.1365-2249.2004.02592.x

Nishio, N., Okawa, Y., Sakurai, H., & Isobe, K. I. (2008). Neutrophil depletion delays wound repair in aged mice. *Age*. https://doi.org/10.1007/s11357-007-9043-y

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. https://doi.org/10.1093/nar/27.1.29

Ogata, K., An, E., Shioi, Y., Nakamura, K., Luo, S., Yokose, N., … Dan, K. (2001). Association between natural killer cell activity and infection in immunologically normal elderly people. *Clinical and Experimental Immunology*.

https://doi.org/10.1046/j.1365-2249.2001.01571.x

Ono, M., Kawakami, M., & Ushikubo, H. (1987). Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. *Journal of Virology*.

Orem, J., Mbidde, E. K., Lambert, B., De Sanjose, S., & Weiderpass, E. (2007). Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. *African Health Sciences*. https://doi.org/10.5555/afhs.2007.7.3.166

Palmer, D. B. (2013). The effect of age on thymic function. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2013.00316

Parkin, D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *International Journal of Cancer*. https://doi.org/10.1002/ijc.21731

Pawelec, G., Derhovanessian, E., Larbi, A., Strindhall, J., & Wikby, A. (2009). Cytomegalovirus and human immunosenescence. *Reviews in Medical Virology*. https://doi.org/10.1002/rmv.598

Pawelec, G., Solana, R., & Wikby, A. (2005). Human immunosenescence: is it infectious? *Immunological Reviews*. https://doi.org/10.1016/j.it.2004.05.006

Pfister, G., Weiskopf, D., Lazuardi, L., Kovaiou, R. D., Cioca, D. P., Keller, M., … Grubeck-Loebenstein, B. (2006). Naive T cells in the elderly: are they still there? *Annals of the New York Academy of Sciences*. https://doi.org/10.1196/annals.1354.018

Pietilä, M., Neuvonen, M., Borodulin, K., Korpela, K., Sievänen, T., & Tyrväinen, L. (2015). Relationships between exposure to urban green spaces, physical activity and self-rated health. *Journal of Outdoor Recreation and Tourism*. https://doi.org/10.1016/j.jort.2015.06.006

Plowden, J., Renshaw-Hoelscher, M., Engleman, C., Katz, J., & Sambhara, S. (2004). Innate immunity in aging: Impact on macrophage function. *Aging Cell*.

https://doi.org/10.1111/j.1474-9728.2004.00102.x

Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., & Franceschi, S. (2016). Global burden of cancers attributable to infections in 2012: a synthetic analysis. *The Lancet Global Health*. https://doi.org/10.1016/S2214-109X(16)30143-7

Ponnappan, S., & Ponnappan, U. (2011). Aging and Immune Function: Molecular Mechanisms to Interventions. *Antioxidants & Redox Signaling*. https://doi.org/10.1089/ars.2010.3228

Reynier, F., Verjat, T., Turrel, F., Imbert, P. E., Marotte, H., Mougin, B., & Miossec, P. (2009). Increase in human endogenous retrovirus HERV-K (HML-2) viral load in active rheumatoid arthritis. *Scandinavian Journal of Immunology*. https://doi.org/10.1111/j.1365-3083.2009.02271.x

Rizvi, S., Raza, S. T., & Mahdi, F. (2014). Telomere length variations in aging and age-related diseases. *Current Aging Science*. https://doi.org/10.2174/1874609808666150122153151

Rose, M. R. (1991). *Evolutionary biology of aging*. *American Journal of Human Genetics*. https://doi.org/10.1017/CBO9781107415324.004

Rosenberg, C., Bovin, N. V, Bram, L. V, Flyvbjerg, E., Erlandsen, M., Vorup-Jensen, T., & Petersen, E. (2013). Age is an important determinant in humoral and T cell responses to immunization with hepatitis B surface antigen. *Hum Vaccin Immunother*. https://doi.org/24480 [pii]\r10.4161/hv.24480

Sapey, E., Greenwood, H., Walton, G., Mann, E., Love, A., Aaronson, N., … Lord, J. M. (2014). Phosphoinositide 3-kinase inhibition restores neutrophil accuracy in the elderly: Toward targeted treatments for immunosenescence. *Blood*. https://doi.org/10.1182/blood-2013-08-519520

Sauter, M., Roemer, K., Best, B., Afting, M., Schommer, S., Seitz, G., … Mueller-Lantzsch, N. (1996). Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors. *Cancer*

Research. https://doi.org/10.1158/0008-5472.can-04-2983

Shekarabi, M., & Asgari, F. (2014). Aging immunity and infection. In *Immunology of Aging*. https://doi.org/10.1007/978-3-642-39495-9_17

Sica, A., Matsushima, K., Van Damme, J., Wang, J. M., Polentarutti, N., Dejana, E., … Mantovani, A. (1990). IL-1 transcriptionally activates the neutrophil chemotactic factor/IL-8 gene in endothelial cells. *Immunology*.

Sicat, J., Sutkowski, N., & Huber, B. T. (2005). Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. *The Journal of Rheumatology*.

Slokar, G., & Hasler, G. (2016). Human endogenous retroviruses as pathogenic factors in the development of schizophrenia. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2015.00183

Solana, R., Campos, C., Pera, A., & Tarazona, R. (2014). Shaping of NK cell subsets by aging. *Current Opinion in Immunology*. https://doi.org/10.1016/j.coi.2014.04.002

Solana, R., & Mariani, E. (2000). NK and NK/T cells in human senescence. *Vaccine*. https://doi.org/10.1016/S0264-410X(99)00495-8

Sorrentino, J. A., Sanoff, H. K., & Sharpless, N. E. (2014). Defining the toxicology of aging. *Trends in Molecular Medicine*. https://doi.org/10.1016/j.molmed.2014.04.004

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0506580102

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the*

National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.0506580102

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*. https://doi.org/10.1186/1742-4690-8-90

Sutkowski, N., Chen, G., Calderon, G., & Huber, B. T. (2004). Epstein-Barr virus latent membrane protein LMP-2A is sufficient for transactivation of the human endogenous retrovirus HERV-K18 superantigen. *Journal of Virology*. https://doi.org/10.1128/JVI.78.14.7852-7860.2004

Sutkowski, N., Conrad, B., Thorley-Lawson, D. A., & Huber, B. T. (2001). Epstein-Barr virus transactivates the human endogenous retrovirus HERV-K18 that encodes a superantigen. *Immunity*. https://doi.org/10.1016/S1074-7613(01)00210-2

Tai, A. K., Luka, J., Ablashi, D., & Huber, B. T. (2009). HHV-6A infection induces expression of HERV-K18-encoded superantigen. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2009.05.019

Thakkar, J. P., McCarthy, B. J., & Villano, J. L. (2014). Age-specific cancer incidence rates increase through the oldest age groups. *American Journal of the Medical Sciences*. https://doi.org/10.1097/MAJ.0000000000000281

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2450

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq

experiments with TopHat and Cufflinks. *Nature Protocols*. https://doi.org/10.1038/nprot.2012.016

Tristem, M. (2000). Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. *Journal of Virology*. https://doi.org/10.1128/JVI.74.8.3715-3730.2000

Turcanova, V. L., Bundgaard, B., & Höllsberg, P. (2009). Human herpesvirus-6B induces expression of the human endogenous retrovirus K18-encoded superantigen. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2009.05.015

United Nations Department of Economic and Social Affairs. (2017). *World Population Prospects: The 2017 Revision*. *Population Division*. https://doi.org/10.1017/CBO9781107415324.004

Vallejo, A. N. (2005). CD28 extinction in human T cells: Altered functions and the program of T-cell senescence. *Immunological Reviews*. https://doi.org/10.1111/j.0105-2896.2005.00256.x

van der Kuyl, A. C. (2012). HIV infection and HERV expression: A review. *Retrovirology*. https://doi.org/10.1186/1742-4690-9-6

van Duin, D., Mohanty, S., Thomas, V., Ginter, S., Montgomery, R. R., Fikrig, E., … Shaw, A. C. (2007). Age-Associated Defect in Human TLR-1/2 Function. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.178.2.970

Wang-Johanning, F., Li, M., Esteva, F. J., Hess, K. R., Yin, B., Rycaj, K., … Johanning, G. L. (2013). Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. *International Journal of Cancer*. https://doi.org/10.1002/ijc.28389

Wang-Johanning, F., Liu, J., Rycaj, K., Huang, M., Tsai, K., Rosen, D. G., … Johanning, G. L. (2007). Expression of multiple human endogenous retrovirus

surface envelope proteins in ovarian cancer. *International Journal of Cancer*. https://doi.org/10.1002/ijc.22256

Weiskopf, D., Weinberger, B., & Grubeck-Loebenstein, B. (2009). The aging of the immune system. *Transplant International*. https://doi.org/10.1111/j.1432-2277.2009.00927.x

Wertheimer, A. M., Bennett, M. S., Park, B., Uhrlaub, J. L., Martinez, C., Pulko, V., … Nikolich-Zugich, J. (2014). Aging and Cytomegalovirus Infection Differentially and Jointly Affect Distinct Circulating T Cell Subsets in Humans. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.1301721

Wikby, A., Johansson, B., Olsson, J., Löfgren, S., Nilsson, B. O., & Ferguson, F. (2002). Expansions of peripheral blood CD8 T-lymphocyte subpopulations and an association with cytomegalovirus seropositivity in the elderly: The Swedish NONA immune study. *Experimental Gerontology*. https://doi.org/10.1016/S0531-5565(01)00212-1

Williams, G. C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*. https://doi.org/10.2307/2406060

Wills, M. R., Ashiru, O., Reeves, M. B., Okecha, G., Trowsdale, J., Tomasec, P., … Sissons, J. G. P. (2005). Human Cytomegalovirus Encodes an MHC Class I-Like Molecule (UL142) That Functions to Inhibit NK Cell Lysis. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.175.11.7457

Xia, S., Zhang, X., Zheng, S., Khanabdali, R., Kalionis, B., Wu, J., … Tai, X. (2016). An Update on Inflamm-Aging: Mechanisms, Prevention, and Treatment. *Journal of Immunology Research*. https://doi.org/10.1155/2016/8426874

Zissel, G., Schlaak, M., & Muller-Quernheim, J. (1999). Age-related decrease in accessory cell function of human alveolar macrophages. *Journal of Investigative Medicine*.