THE HUMAN LUNG VIRAL MICROBIOME IN HEALTH AND DISEASE

Arwa Abbas

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Co-Supervisor of Dissertation

_____

Frederic D. Bushman, Ph.D., Professor of Microbiology

Co-Supervisor of Dissertation

_____

Ronald G. Collman, M.D., Professor of Medicine

Graduate Group Chairperson

_____

Daniel S. Kessler, Ph.D., Associate Professor of Cell and Developmental Biology

Dissertation Committee

Matthew D. Weitzman, Ph.D., Professor of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Chair

Beatrice H. Hahn, M.D., Professor of Medicine

Michael R. Betts, Ph.D., Associate Professor of Microbiology

Elizabeth A. Grice, Ph.D., Associate Professor of Dermatology

ACKNOWLEDGEMENTS

When asked if I struggle with having two mentors, I often say that the most difficult part is deciding who is more supportive or inspiring. Both Rick and Ron have guided me through the process of becoming a capable scientist in ways that synergize and complement the other's mentoring style. They've taught me how to balance being mindful of the experimental minutiae while also thinking about the broader aspects of the biological phenomena being studied. Their unabashed excitement about designing, executing and interpreting experiments has left a lasting impression on me. Similarly, I would like to thank my entire thesis committee and my rotation mentors, Sunny Shin and Mark Goulian, for their valuable feedback and insight, and also for their confidence in my (eventual) success.

The science produced and presented in this thesis would not have been possible without the contributions of many others. First, I truly appreciate Jacque Young's patience as she introduced me to the world of viruses. Before I became a senior graduate student myself, I looked up to Alexandra Bryson, Christel Chehoud, Erik Clarke, and Katie Sheehan-Wetzel to learn how to be a clear presenter, empathic lab citizen, and engaged member of the graduate student community. I would especially like to thank Louis Taylor, my ingenious partner in virus hunting and overall amazing human being, who has taught me numerous microbiological, computational and social skills. I'd also like to thank Meagan Rubel, my steadfast advocate and bottomless source of advice and knowledge, for encouraging me to think outside of the box. While the membership of the lab is always in a state of flux, the overall atmosphere of comradery and innovative thinking is constant. Therefore, I'd like to thank all the current and former members of the Bushman and Collman labs, the Penn Chop Microbiome Program, the Department of Microbiology, the Biomedical Graduate Studies and the Cellular and Molecular Biology (CAMB) Program directors and coordinators, and the clinical investigators at the Hospital of the University of Pennsylvania for contributing to the productive, collegiate environment where I've been lucky to train in.

Numerous friends and colleagues outside of the lab have positively shaped my graduate school experience. I had the privilege of working alongside and befriending inspiring student leaders like Seleeke Flingai, Ciara Gimblet, Brenda Salantes, Julianne Rieders, Kate Palozola and Neha Pancholi. My friends in the Mostly Virology Program and the larger CAMB community, especially Annie Chen, Tzvi Pollock, and Priya Chatterji, are amazing individuals whose company makes me feel safe and sane. I've enjoyed and learned much from working with eclectic groups including EMPOWER for College, the Science Education Academy, the Ernest E. Just Biomedical Society and the CAMB Student Newsletter.

Successfully surviving the Ph.D. process is undoubtedly difficult and I am deeply thankful and cognizant of how blessed I am to have come this far. Therefore, I'd like to thank the Saeeds and Zafars, my adoptive older siblings, for looking after me since my arrival in Philadelphia and for being my role models in many different spheres of life. I'd also like to thank all the friends and family who have given me encouragement and continuously held me to a high standard knowing that I could, and would, meet their expectations. Finally, I cannot thank my mother, father and sister enough for their unwavering, unconditional love. While I don't often believe in myself, I do believe in the others who believe in me.

ABSTRACT

THE HUMAN LUNG VIRAL MICROBIOME IN HEALTH AND DISEASE
Arwa Abbas
Frederic D. Bushman and Ronald G. Collman

Vast and diverse microbial communities (the microbiome) are distinct at different human body sites and strongly influence health and disease. Specifically, the respiratory tract microbiome is thought to influence outcomes after lung transplantation, the only therapeutic option for end-stage lung diseases. Studies dissecting the role of the microbiome on pulmonary health should also include the viral microbiome (virome), which is less-studied due to unique challenges in identifying these small, diverse, self-replicating genetic elements. Organ transplantation is accompanied by immunosuppression, which can result in reactivation of latent viruses, transfer of viruses from organ donor to recipient, and increased susceptibility to viral infections. We therefore used high-throughput metagenomic approaches to study the virome of lung transplant recipients (LTRs). We first characterized the virome in LTRs and its relationship to clinically defined adverse events. We discovered that a family of eukaryotic viruses (*Anelloviridae*) is abundant in the lung and blood of LTRs and that their levels in the lungs were associated with primary graft dysfunction, a form of acute lung injury. Next, we investigated the temporal and spatial dynamics of the virome during lung transplantation and identified herpesviruses, parvoviruses, polyomaviruses, bacteriophage and complex anellovirus populations. We focused on the abundant anelloviruses by assembling genomes from shotgun metagenomic sequences and tracking their representation in the lung and blood of LTRs post-transplantation using a metric that accounts for inter-and intra-subject viral diversity. This analysis revealed that anellovirus populations move between lung allografts and the peripheral blood of LTRs. However, many uncharacterized sequences still existed in the

metagenomic data generated in these studies. To address this, we developed a molecular and bioinformatics pipeline to mine public datasets and discovered a novel family of small, circular DNA viruses (*Redondoviridae*). Quantification of redondoviruses in human oro-respiratory samples showed an association with periodontal disease and acute illness. Overall, this work helps define the virome during lung transplantation and introduces a new family of human viruses, broadly demonstrating the importance of exploring the human virome.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ACR, acute cellular rejection
AIDS, acquired immunodeficiency syndrome
ATG, anti–thymocyte globulin
BAL, bronchoalveolar lavage
BLAST, Basic Local Alignment Search Tool
BOS, bronchiolitis obliterans syndrome
Bp, base pairs
cDNA, complementary DNA
CF, cystic fibrosis,
CMV, cytomegalovirus
COPD, chronic obstructive pulmonary disease
Cp, capsid
CRESS, circular, Rep-encoding ssDNA
CTOT-03, Clinical Trials in Organ Transplantation-03
dN/dS, non-synonymous to synonymous ratio of substitution
dsDNA, double-stranded DNA
EBV, Epstein-Barr virus
FDR, false discovery rate
GSVA, Gene Set Variation Analysis
HHV, human herpesvirus
HIV, human immunodeficiency virus
HOMD, Human Oral Microbiome Database
ICU, intensive care unit
IPF, idiopathic pulmonary fibrosis
IRB, institutional review board
ITS, internal transcribed spacer
Kb, kilobases
LTRs, lung transplant recipients
miRNA, micro RNA
MDA, multiple displacement amplification
MG-RAST, Metagenomic Rapid Annotations using Subsystems Technology
MRSA, methicillin-resistant *S. aureus*
ORF, open reading frame
PCV-2, porcine circovirus 2
PGD, primary graft dysfunction
PMWS, post-weaning multisystemic wasting syndrome
PoSCV-5, Porcine stool-associated circular virus 5
RBS, ribosomal binding site
Rep, replication-associated protein
RCR, rolling-circle replication
RPKM, reads per kilobase per million sequenced
rRNA, ribosomal RNA
SRA, Sequence Read Archive
ssDNA, single-stranded DNA

SWGA, selective whole-genome amplification
TLR, Toll-like receptor
TTV, torque teno virus
TTSu, torque teno sus virus
VLP, virus-like particle
Virome, viral microbiome

CHAPTER 1: INTRODUCTION

The human microbiome is a large and complex population of commensal and pathogenic microorganisms including bacteria, archaea, fungi, bacteriophage and eukaryotic viruses that live on and within us. The composition and temporal and spatial dynamics of this microbiome can be characterized through high-throughput DNA sequencing technologies. Unbiased nucleic-acid sequencing techniques have advantages over traditional culture-based methods which are limited in sensitivity and require prior knowledge of the microbial species of interest. Application of newer sequencing-based molecular techniques has revealed that many human body sites such as the gut, skin, mouth and urogenital tract harbor unique microbiomes. In some sites, these communities play an integral role in maintaining homeostasis locally and for the organism as a whole. Gut microbes benefit their host by promoting proper immune system development and function (Belkaid and Harrison, 2017), providing nutrition (Flint et al., 2012), metabolizing xenobiotics (Maurice et al., 2013) and competing against pathogens (Hibbing et al., 2010). On the other hand, disruptions in the microbiome (dysbiosis) have been linked to diverse infectious, metabolic, oncological, psychological and immunological diseases (Knight et al., 2017).

Bacteria, fungi and archaea within the microbiome are most efficiently catalogued by targeting shared sequence tags, such as the 16S ribosomal RNA (rRNA) or 18S rRNA genes. In contrast, viruses lack sequences that are shared amongst all types and thus shotgun metagenomic sequencing is needed to identify multiple viruses. One challenge that remains in the rapidly-growing microbiome field is the accurate classification of metagenomic sequences lacking similarity to reference database sequences. Fortunately, the generation of vast amounts of metagenomic data allows unclassifiable sequences to be

associated with certain environments, hosts and disease states using large-scale bioinformatics comparisons. Comprehensive characterization of novel sequences will broaden our understanding of the human microbiome and could reveal hitherto undiscovered links between the human microbiome and health and disease.

1.1 THE HUMAN LUNG MICROBIOME

The mammalian respiratory tract is a crucial organ system that spans from the nares to the lung alveoli and is responsible for the exchange of carbon and oxygen. The organ is divided at the larynx into the upper and lower respiratory tract, each containing niches with distinct physiological features that ultimately shape the microbial communities found there (Man et al., 2017). The Human Microbiome Project was the first large interdisciplinary effort that profiled the microbiomes of 15-18 body sites for 242 healthy individuals (Consortium, 2012). The anterior nares and several distinct sites within the oral cavity were the only respiratory tract sites targeted for microbial sequencing in this initial survey. It was found that *Staphylococcus* species including *S. aureus*, the cause of methicillin-resistant *S. aureus* (MRSA) infections, and commensals such as *S. epidermidis* are common and universal, respectively, in the nares. The oral cavity is generally dominated by *Streptococcus* spp., but also *Haemophilus* in the buccal mucosa, *Actinomyces* in the supragingival plaque, and *Prevotella* in the subgingival plaque (Consortium, 2012). Other studies employing both targeted and unbiased molecular techniques frequently detect viruses in the upper respiratory tract of asymptomatic individuals, including rhinovirus, bocavirus, polyomaviruses, adenovirus, coronavirus and anelloviruses (Man et al., 2017). The healthy upper respiratory tract also has a fungal microbiome that includes *Aspergillus* spp., *Penicillium* spp., *Candida* spp. and *Alternaria* spp. A key finding of these landmark studies was that ostensibly healthy individuals can have remarkable variation in their

microbiomes. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have been shown to play key roles.

Lungs are a unique organ system as they are constantly exposed to the external environment by respiration of exogenous particles and to the microbe-rich upper respiratory tract. The lower airways are lined with a similar respiratory epithelium to that found in the upper airways, while the alveoli, where gas-exchange takes place, have functionally distinct alveolar epithelial cells. Although the lower respiratory tract has historically been regarded as strictly sterile due to physiological barriers and clearance by immunological processes, recent culture-independent, nucleic-acid sequencing-based approaches have revealed bacterial and fungal populations both in health and in disease (Huang et al., 2013). In healthy individuals, the lower airway bacterial community composition is indistinguishable from that of the upper respiratory tract and total bacterial counts are low (Charlson et al., 2011). This suggests that the healthy lower respiratory tract does not harbor a thriving microbial community, but a sparse and transient population that likely derives from aspiration of the upper respiratory tract microbiome. However, even low levels of bacteria induce some immune activation (Segal et al., 2013, Segal et al., 2016, Huang and Lynch, 2011) which likely influences pulmonary development and homeostasis.

Individuals suffering from obstructive lung diseases such as cystic fibrosis (CF) (Willner et al., 2012b, Stokell et al., 2015), asthma (Hilty et al., 2010), and chronic obstructive pulmonary disorder (COPD) (Erb-Downward et al., 2011) have distinctly altered bacterial microbiomes whose specific features and members have been linked to clinical disease severity and progression (Huang et al., 2011, Zhao et al., 2012, Sze et al., 2012, Erb-Downward et al., 2011). The exact function of the lung microbiome in establishing and maintaining respiratory health and the consequences of disrupting these microbial communities are still mostly unknown and under active investigation.

Specific trans-kingdom interactions in the respiratory tract are known to cause severe disease, such as the deadly secondary bacterial pneumonia that occurred after initial influenza A infection during the pandemic in 1918 (Taubenberger et al., 2000), and have been extensively studied in animal models. In general, respiratory viruses can modulate innate and adaptive immune responses promoting bacterial colonization and infection. Similarly, certain bacteria have been shown to both promote and impede infections with specific viruses in the respiratory tract (Man et al., 2017).

Fewer studies have focused broadly on the lung virome in humans. An initial investigation of viral communities in expectorated sputum from individuals with and without CF (Willner et al., 2009) suggested that there is a core set of bacteriophages found in the healthy human respiratory tract. Consistent with the low bacterial biomass observed in other studies of healthy lungs, the phage communities in individuals without CF also appeared to represent a random, transient community related to bacteria from the external environment. A separate set of bacteriophages associated with pathogenic bacterial species such as *P. aeruginosa* were found in CF patients. This was followed by a case study of lung tissue sampled from two CF lungs which revealed distinct and diverse bacteriophages and eukaryotic viruses, including herpesviruses, adenoviruses, human papillomavirus, and anelloviruses (Willner et al., 2012a). Other studies which instead focused on upper respiratory tract samples from individuals with acute respiratory tract infections (Yang et al., 2011) (Lysholm et al., 2012, Wylie et al., 2012, Zoll et al., 2015), healthy adults (Wylie et al., 2014) and idiopathic pulmonary fibrosis (Wootton et al., 2011), also detect the viruses described above. These findings all support the existence of an authentic respiratory virome whose role can now be investigated in various pulmonary and systemic diseases.

1.2 LUNG TRANSPLANTATION AND PRIMARY GRAFT DYSFUNCTION

Lung transplantation is the only long-term option for patients suffering from end-stage lung diseases including COPD, CF, idiopathic pulmonary fibrosis (IPF) and emphysema. In 2015, more than 5,000 lung transplants were performed worldwide (Chambers et al., 2017). In addition to the obvious survival benefits for those individuals with fatal lung diseases, transplantation can substantially improve quality-of-life (Kotloff and Thabut, 2011). Unfortunately, the median survival for adult lung transplantation recipients (LTRs) is only 5.8 years (Lund et al., 2016), which is lower than most other solid-organ transplants (Watson and Dark, 2012).

Barriers to the short and long-term success of lung transplantation include graft failure and rejection. Primary graft dysfunction (PGD) is a form of acute lung injury which occurs within 72 hours of transplantation. PGD is characterized by hypoxemia, pulmonary edema and alveolar damage. Approximately 10-30% of all LTRs experience severe PGD, which is a leading cause of death immediately post transplantation (Lee and Christie, 2011, Suzuki et al., 2013).

The pathogenesis of PGD is multi-factorial and thought to be related to ischemia-reperfusion events. An inflammatory environment develops within the lung after donor brain death, driven by macrophages that release chemokines and cytokines. This inflammatory state is perpetuated after reperfusion by recruitment of T-cells and neutrophils. Activated immune cells, in conjunction with reactive oxygen species generated during ischemia-reperfusion, are thought to directly injure lung endo- and epithelium (Diamond and Wigfield, 2013). LTRs who survive severe PGD are more likely to develop donor-specific antibodies (Ius et al., 2014) and bronchiolitis obliterans syndrome (BOS), a clinical manifestation of chronic rejection (Daud et al., 2007). Treatment for PGD is limited to supportive therapy and no definitive preventive measures exist (Suzuki et al., 2013).

Reducing the incidence of PGD would dramatically improve outcomes following lung transplantation.

The pathology of PGD is incompletely understood, but has been associated with innate immune pathway activation. Pathways identified in a gene set enrichment analysis of mRNA from bronchoalveolar lavage (BAL) of LTRs who developed PGD included those involved in inflammasome activation and pattern recognition receptors signaling pathways, which are stimulated by danger- and pathogen-associated molecular patterns (Cantu et al., 2013). In addition, complement activation within the allograft has been implicated in PGD in isogenic rat models of transplantation (Naka et al., 1997). Plasma levels of C4a and C5a, small peptides that promote local inflammation, neutrophil recruitment and macrophage activation, are elevated during PGD (Shah et al., 2014). The complement cascade can be initiated by antigen-antibody complexes, direct binding of complement to a pathogen, or binding to microbial mannose residues. Together, these studies suggest a microbial contribution to the immune activity observed in PGD.

1.3 ALTERED MICROBIOMES FOLLOWING LUNG TRANSPLANTATION

In addition to the established negative impact of graft injury on the clinical success of lung transplantation, LTRs are particularly susceptible to infection because they are under lifelong administration of immunosuppressive drugs to minimize graft rejection (Bhorade and Stern, 2009). Normal airway clearance in LTRs is also compromised due to decreased sputum clearance, loss of the cough reflex due to vagal denervation of the transplanted organ and surgical disruption of lymphatic system. Increased propensity for micro-aspiration (Atkins et al., 2007), can also allow seeding of the lungs from the microbe-rich upper respiratory tract. Accordingly, clinically significant infections are the leading cause of death within the first year post-transplantation (Burguete et al., 2013). Bacterial pneumonia caused by *P. aeruginosa* or *S. aureus*, and opportunistic fungal and viral

infections are frequent within the first few months post-transplantation (Burguete et al., 2013). Furthermore, severe disease can arise from reactivation of common, latent viruses such as cytomegalovirus (CMV), which has a 60% seropositivity rate in the general United States population (Staras et al., 2006). As a result, patients now routinely receive CMV prophylaxis if they are seropositive or receive an organ from a seropositive donor. Pathologies due to poorly characterized viruses have also been discovered in immunosuppressed organ transplant recipients. Recently, two LTRs who suffered from epidermal hyperplasia with pruritic rashes exhibited affected tissue that was abundant in human polyomavirus 7 viral particles, DNA and protein (Ho et al., 2015). Studying the virome in these immunosuppressed populations can reveal novel disease manifestations during viral infection and shed light on host-viral interactions that are normally masked in an immunocompetent host.

In addition to immediate and direct injury to lung tissue, acute infections dispose LTRs to later development of BOS, a clinical manifestation of chronic rejection and leading cause of long-term graft failure. Indeed, a well-known risk factor for BOS is CMV infection (Kroshus et al., 1997). Additionally, recent studies suggest that BOS is more frequent in LTRs with prior clinical infection or documented colonization with *Aspergillus, P. aeruginosa, Chlamydia pneumoniae*, and community-acquired respiratory viruses including parainfluenza, respiratory syncytial virus, metapneumovirus, coronavirus, rhinovirus, and influenza (Martin-Gandul et al., 2015, Dickson et al., 2014a, Peghin et al., 2017, Yates et al., 2005, Botha et al., 2008).

Besides *de novo* infection and reactivation from reservoirs within the recipient, microorganisms can also be transmitted from the donor lung. Routine screening of organ donors is performed for a limited panel of pathogens (Grossi et al., 2009), although unexpected transmission can occur due to inadequate screening. While these events occur

in a minority of transplantations, the reported 1% incidence of transmission of infectious agents is likely an underestimate (Ison and Nalesnik, 2011), as only infections leading to significant morbidity and mortality are reported and because transmission can be influenced by inoculum, organ type, and types of immunosuppression and antibiotics given (Ison and Nalesnik, 2011). Pathogenic or benign microbes originally in the donor lung may persist in immunosuppressed LTRs, but the extent of persistence of viruses has not been fully probed. An unbiased analysis of viral transmission during lung transplantation is essential to anticipate disease occurrence and would improve general understanding of how the virome is affected by host immunology and immuno-suppressive drugs.

Previous investigations of microbial agents and lung transplant outcomes are often limited by their focus on single agents rather than communities. Recent pioneering high-throughput studies show that the respiratory tract bacterial and fungal microbiomes of LTRs are aberrant compared to healthy lungs; specifically there is decreased microbial diversity, increased microbial burden and outgrowths of recognized lung pathogens and atypical species (Charlson et al., 2012, Dickson et al., 2014b, Borewicz et al., 2013). Initial studies of the viral component of the lung microbiome from our group discovered high levels of anelloviruses (Young et al., 2015). Anellovirus levels in both the lung and upper respiratory tract were higher in LTRs compared to healthy adults and were associated with bacterial dysbiosis (Young et al., 2015). This pilot study analyzed heterogeneous time points post-transplantation; thus, the virome of the donor organ and temporal dynamics within individual LTRs after transplantation could not be ascertained.

1.4 ANELLOVIRIDAE: A FAMILY OF UBIQUITOUS HUMAN VIRUSES

One intriguing finding in various studies of the virome is that humans are nearly always colonized by members of *Anelloviridae,* a family of diverse, non-enveloped, small circular ssDNA eukaryotic viruses including alphatorqueviruses, betatorqueviruses and

gammatorqueviruses (Figure 1.1). The first anellovirus, torque teno virus, was originally discovered in a hepatitis patient in 1997 (Nishizawa et al., 1997). Its prevalence in the human population has since been estimated to be at least 50% and as high as 90% in certain populations (Spandole et al., 2015). In fact, anelloviruses are found in many mammalian species including non-human primates, cats, rodents, swine and bats.

Anelloviridae display remarkable genomic variability; genome sizes vary from 2.8 kilobases (kb) to 3.9 kb between different genera and the number of unique isolates from diverse animals deposited in reference databases increases each year (Spandole et al., 2015). While the untranslated region of the genome is somewhat conserved within the family, the amino acid sequences of the Open Reading Frame 1 (ORF1) protein can diverge by up to 70% (Spandole et al., 2015). *Anelloviridae* quasi-species have been described to exist both within and between individuals. There are several hypotheses for the extensive genome variation seen in *Anelloviridae*. One is that there is a high basal mutation rate; it is known that RNA viruses and single-stranded DNA (ssDNA) viruses accumulate mutations at high rates. Indeed, the *Circoviridae* family of eukaryotic ssDNA viruses drifts at rates of $1.2x10^{-3}$ substitutions/site/year (Firth et al., 2009). The three hypervariable regions of the putative capsid protein (Nishizawa et al., 1999) may represent sites of immune evasion, which is well-known to occur during chronic infection with other viruses such as human immunodeficiency virus (HIV) (Rambaut et al., 2004). Another possibility is co-infections with multiple isolates allowing for within-host recombination (Worobey, 2000, Fahsbender et al., 2017). Considering that anelloviruses are detected in infants within six months of birth (Lim et al., 2015, Komatsu et al., 2004), it's plausible that viral diversity is a combination of adaption within the host and frequent *de novo* infections that continue into adulthood.

It is known that changes in total anellovirus levels in blood correlate with altered immune states. Plasma levels of anellovirus DNA increase in HIV-positive individuals that progress to acquired immunodeficiency syndrome (AIDS) and are inversely correlated with CD4+ T-cell counts (Thom and Petrik, 2007, Christensen et al., 2000, Li et al., 2013). In contrast, anellovirus levels decrease in some HIV-positive patients whose immune system is reconstituted following highly active anti-retroviral therapy (Devalle et al., 2009, Madsen et al., 2002). An increase in anellovirus load is also seen in the elderly, who are thought to be immune deficient (Haloschan et al., 2014). Other studies describe expansion of *Anelloviridae* in plasma of solid-organ transplant recipients receiving immunosuppression therapy (Görzer et al., 2014, De Vlaminck et al., 2013, Görzer et al., 2015). In addition, De Vlaminck et al reported a positive correlation between *Anelloviridae* expansion and increased dosage of immunosuppressive drugs including tacrolimus, a T-cell activation inhibitor, and prednisone, a corticosteroid. These findings, which suggest a strong link between host immune competence and control of these viruses, have led some to suggest measuring anellovirus levels in the blood to monitor functional immune suppression in clinical settings (Focosi et al., 2016, De Vlaminck et al., 2013).

Although anelloviruses are ubiquitous, many tenets of their interaction with the host, including their target cells, are incompletely characterized. Several early *in vitro* studies suggested that anelloviruses replicate in activated peripheral blood mononuclear cells and other hematopoietic cells (Mariscal et al., 2002, Maggi et al., 2001). Recently, plasma anellovirus levels were tracked in kidney and pancreas transplant recipients who experienced T-cell depletion with anti-thymocyte globulin (ATG) given as induction immune suppression. In this study, viral levels decreased with decreasing lymphocyte counts in peripheral blood. Transplant recipients treated with basiliximab, an induction agent that only prevents T-cell activation and proliferation, did not experience a similar

drop of anelloviruses (Focosi et al., 2015). Another study tracking blood anellovirus DNA levels in allogeneic stem cell transplantation reported a similar decrease in viral DNA after myeloablative conditioning; this reduction was partially reversed after successful engraftment of donor stem cells (Albert et al., 2017). Similarly, torque teno sus viruses (TTSu) have been detected by *in situ* hybridization in CD3+ cells within the interfollicular region and mantle zone of swine lymphoid tissue (Lee et al., 2015) . Thus, it is postulated that T-cells are a major replication compartment of *Anelloviridae* species *in vivo*, although their growth in these cell types has not been successfully maintained *in vitro*. Additionally, it is unclear whether anellovirus replication in lymphocytes requires activation of these cells *in vivo*. Investigating viral dynamics and different facets of the immune response is necessary to address the apparent paradox of why anelloviruses, hypothesized to replicate in activated T-cells, increase in various compartments in immune suppressed states.

A limited number of studies have examined how anelloviruses trigger and counteract host immune responses. TTV DNA was shown to elicit an inflammatory response via Toll-like receptors (TLRs) (Rocchi et al., 2009), while other groups demonstrated that a virally encoded micro RNA (miRNA) and the putative non-structural ORF2 protein inhibit interferon and NF-kB signaling, respectively (Kincaid et al., 2013, Zheng et al., 2007). Antibodies against human anellovirus ORF1 and ORF2 have been detected in human sera (Ott et al., 2000, Chen et al., 2013, Kakkola et al., 2008, Mankotia and Irshad, 2014). High rates of seropositivity against TTSu (24-100%) have also been seen in global swine herds (Kekarainen and Segalés, 2012) indicating that wide-spread infection with *Anelloviridae* species are common in non-human hosts as well.

Virtually nothing is known about cellular immune responses controlling anelloviruses. Many other chronic infections, such as CMV, are known to be controlled by virus-specific T-cells. Considering that persistent CMV infection even in the absence of overt

disease promotes oligoclonal expansion of CMV-specific T-cells (Brunner et al., 2011, Pawelec and Derhovanessian, 2011), it is possible that chronic and repeated infections with diverse anelloviruses may also profoundly shape the human immune system.

Since their discovery, anelloviruses have been widely interrogated on whether they are the etiological agent of any disease. In swine, inoculation of gnotobiotic piglets with TTSu1-containing tissue homogenates prior to infection with porcine circovirus 2 (PCV-2) induced post-weaning multisystemic wasting syndrome (PMWS), although neither TTSu-1 nor PCV-2 alone recapitulated the disease (Ellis et al., 2008). Furthermore, TTSu-2 prevalence and viral load are higher in PMWS-affected animals (Kekarainen and Segalés, 2012, Meng, 2012) and are also correlated with PCV-2 viral levels in lymph tissue from diseased animals (Lee et al., 2015). As a result, it is postulated that co-infection with TTSu potentiates diseases by PCV-2. Thus far, human anelloviruses are not definitively associated with any disease, although there is some preliminary evidence of a role in respiratory and oral diseases. For example, one study reported increased viral levels in the lungs of children with asthma and ciliary dysmotility (Pifferi et al., 2008). Additionally, increased anellovirus levels in the blood have been correlated with COPD (Feyzioğlu et al., 2014) and worse outcomes of IPF (Bando et al., 2015, Bando et al., 2008, Bando et al., 2001). The presence of anelloviruses in gingival tissue has also been associated with periodontal disease (Rotundo et al., 2004, Zhang et al., 2017). It remains unclear whether these viruses contribute to disease pathogenesis or are merely bystanders reflecting immune cell recruitment. In either case, assaying viral levels in the appropriate compartment could serve as a personalized, functional marker of disease progression and/or prognosis. In order to explore this possibility, it is necessary to examine and compare viral dynamics in a compartment-specific manner that is able to distinguish between anellovirus strains.

1.5 METAGENOMIC METHODS TO STUDY THE HUMAN LUNG VIRAL MICROBIOME

While the molecular characterization of the bacterial and fungal members of the human microbiome at certain body sites has bloomed in the past two decades, studies of the viral component remain incomplete. More recently, advancements in enrichment of low inputs of viral nucleic acid and improvements in sequencing technologies and bioinformatics pipelines have opened up possibilities to interrogate the human virome. But characterizing the lung virome faces challenges not found in high microbial biomass or easily accessible sites such as the gut, skin, oral cavity or uro-genital tract. First, sampling the lower respiratory tract via bronchoscopy requires passage through the microbe-rich upper respiratory tract with potential for carryover. Second, it now well-known that contamination from the environment, sequencing instrument or reagents can confound analyses of relatively low biomass samples (Kim et al., 2017, Salter et al., 2014, Naccache et al., 2013, Clarke et al., 2017a). Third, viral nucleic acids can comprise a vanishingly small fraction of the total DNA of a sample. Well-designed cohorts, sampling approaches, negative controls, and specialized sample preparation techniques that optimize recovery of viral nucleic acids can allow for stringent comparisons in human virome studies.

To that end, our group and others have developed biochemical and analytical methods specifically adapted for studying the virome in low-biomass samples (Thurber et al., 2009, Young et al., 2015, Aggarwala et al., 2017). First, enrichment of virus-like particles (VLPs) is achieved by filtration or centrifugation of the liquid clinical sample to remove eukaryotic and prokaryotic cells. Second, non-encapsidated nucleic acids are digested by enzymes to reduce contaminating human and bacterial DNA and RNA. Proteinase treatment of viral capsids is followed by nucleic acid extraction. Third, whole-genome amplification is performed (described further below) to amplify minute amounts of DNA. To study RNA

viruses, cDNA is generated using random priming and reverse transcription (Wang et al., 2003).

Unlike bacterial and fungal taxa, which can be identified based on 16S rRNA and internal transcribed spacer (ITS) sequences of their genomes respectively, a single universal gene sequence cannot be used to identify all viruses. Therefore, a shotgun metagenomic approach, which attempts to capture and sequence all DNA in a sample, is employed. Advances in library preparation now allow use of small amounts of input and multiplexing of hundreds of samples using dual-indexed barcodes. Short-read (150-300 bp) sequencing is predominantly performed using the Illumina platform which can generate large outputs and has high accuracy (average error rate of 0.24% per base) (Pfeiffer et al., 2018). After generation of these large datasets, sophisticated computational tools are used to filter out sequencing errors and contamination and ultimately identify and characterize viral sequences. Different viruses can be present in distinct abundances in the original sample; as a result, more abundant genomes will be sequenced to greater depth, while rarer genomes will be sparsely covered, or not represented at all. By using mathematical models, one can use these differences to further infer other aspects of viral community structure and population dynamics.

1.6 CHALLENGES IN CHARACTERIZING HIGHLY DIVERGENT VIRAL SEQUENCES

There are specific challenges posed by the "viral dark matter", which refers to the vast majority of reads generated by shotgun metagenomic studies that cannot be annotated into functional or taxonomic categories (Aggarwala et al., 2017). The enormous size and diversity of global viral populations is not usually captured by the few thousand viral species available in databases. As a result, any new putative viral genomic sequence (which does not resemble any database entry) is usually missed in analysis, leading to an

incomplete understanding of the virome. Nevertheless, there are approaches that can be used to gain insight on these unclassifiable sequences. For example, sequence reads can be assembled into contigs representing viral genomes or partial genomes, which can then be used as queries for nucleotide and protein sequence searches. Additionally, when a novel sequence is discovered and characterized, it can effectively be used as "bait" to pull out other distantly homologous sequences. This iterative approach is one way to systematically illuminate the viral dark matter.

1.7 INCREASED DISCOVERY OF CIRCULAR SINGLE-STRANDED DNA VIRUSES IN THE METAGENOMICS ERA

Shotgun metagenomics is an approach that samples and sequences all nucleic acid present in an environment and interrogates the biological entities present there. While there are certain steps in the workflow that can enrich or exclude certain types of organisms or nucleic acids, in general it is considered an unbiased approach. The improvements in sequencing capacity and accuracy, bioinformatics tool development and the decrease in cost allow generation of large datasets. Together, these advancements have proven especially fruitful in the field of virology.

Viruses are numerically the most abundant biological entities on the planet, inhabiting every imaginable environmental niche. For example, quantification of VLPs using electron and epiflourescent microscopy reveals that sea water contains 10^4-10^8 VLPs/mL (Proctor, 1997) and human stool contains 10^8-10^9 VLPs/gram (Aggarwala et al., 2017, Kim et al., 2011). While some of these VLPs may not represent infectious particles or are bacteriophage, some may represent novel and diverse eukaryotic viruses. However, because some viruses are recalcitrant to *in vitro* culture and/or may not cause overt cell pathology, they have been less studied. Additionally, rare viruses with small genomes are often difficult to identify against the background of highly abundant and large host

genomes. Thus, the world of small eukaryotic viruses is only beginning to be explored. In this work, we focus on circular ssDNA viruses that can infect humans.

One such group is the *Anelloviridae* family, the first member of which (torque teno virus or TTV) was discovered in 1997 in a single subject with fulminant hepatitis in the absence of viral agents already known to cause hepatitis (Nishizawa et al., 1997). A fragment of the TTV genome was first identified using representational difference analysis which detects differences in DNA populations using subtractive hybridization between two related samples (Lisitsyn et al., 1993); in this case, serum from the same subject prior to and during the disease state. A subsequent study was able to amplify a full-length genome from human blood using primers based on the genome fragment discovered earlier (Mushahwar et al., 1999). Further description of the *Anelloviridae* family and its postulated roles in human health has been discussed previously. Overall, its discovery demonstrates the remarkable ability of an unbiased approach followed by targeted queries to unveil a ubiquitous biological phenomenon.

Recently, sensitive detection of circular ssDNA molecules that requires no prior knowledge of the target sequence has improved markedly with the use of multiple displacement amplification (MDA) during DNA preparation for low biomass metagenomic studies. In this process, random primers first anneal to a denatured ssDNA template. Then, the highly processive and accurate bacteriophage phi29 DNA polymerase synthesizes DNA in an isothermal reaction until it encounters a segment of double-stranded DNA (dsDNA). Its ability to displace the second strand of DNA allows it to continue synthesis until it reaches the end of the template or falls off stochastically. This ultimately results in linear accumulation of a DNA template over time. However, in the case of a circular DNA template, a single polymerase can proceed in a circular fashion resulting in many rounds of template amplification creating concatenated genomes that are also targets for primer annealing.

This process is illustrated in Figure 1.2. As a result of using this technique in metagenomic surveys of samples from varied environments and hosts, many different circular ssDNA sequences have been identified; the challenge that remains is in understanding what they are and what they do.

One way to evaluate whether the sequence is likely a replication competent virus is by detecting an ORF that encodes a Replication-associated protein (Rep). This is because comparing sequences based on nucleotide sequence is usually insufficient to characterize highly divergent sequences that have no homolog in databases. On the other hand, the Rep protein has distinct conserved motifs and domains (Rosario et al., 2012) that are important for rolling-circle replication (RCR), a common mechanism by which circular viruses and mobile DNA elements replicate their genomes. Although the core catalytic residues of the Rep protein are highly conserved, differences in the motif signature and presence or absence of additional domains allows distinction of Reps from different viral families including *Geminiviridae*, *Circoviridae*, *Nanoviridae* and circular DNA such as plasmids or satellite DNA. However, the presence of a Rep protein is not a guarantee to identify a putative ssDNA virus; for example, there are no consistent conserved RCR motifs in members of the *Anelloviridae* family (Rosario et al., 2012).

Nevertheless, analyses of circular Rep-encoding ssDNA (CRESS) sequences found in metagenomic studies have recently revealed three new distinct families of eukaryotic viruses; *Bacilladnaviridae, Smacoviridae* and *Genomoviridae*. Viruses in the *Bacilladnaviridae* family infect marine diatoms (Tomaru et al., 2011). In addition to having a covalently-closed ssDNA genome encoding a Rep protein, portions of bacilladnaviral genomes are double-stranded (Tomaru et al., 2013), demonstrating the range of genome architectures possible within CRESS viruses. Smacoviruses (43 species across 6 genera) have been found in fecal matter of vertebrates, primarily mammals and birds, but not in

animal tissue (Varsani and Krupovic, 2018). A single genomovirus, sclerotinia gemycircularvirus 1, has been shown to infect a fungus, but hundreds of similar genome sequences have also been found in environmental, plant and animal-associated samples (Krupovic et al., 2016). A limitation in understanding the host range of these viruses is a lack of experimental systems to grow viruses discovered by metagenomic methods.

While traditional culture-based techniques are needed to understand the biological properties of newly discovered viruses, the approaches to identify them in the metagenomics era are undoubtedly changing. In recognition of this, a recent workshop of experts and members of the International Committee of Viruses has updated the framework for classification (Simmonds et al., 2017) of viral sequences that have solely been identified by metagenomic means and/or without biological or experimental characterization (Simmonds et al., 2017). These amendments will facilitate placing diverse viral genomes into higher order taxonomic groups from which insight may be gained regarding their evolutionary histories and host relationships.

1.8 DISSERTATION AIMS

Since solid organ transplantation and accompanying immunosuppression disrupt host–virus interactions, it is a unique and valuable setting to interrogate the human virome in general and to determine its influence on outcomes post-transplantation. In addition, the features of viral communities associated with post-transplant complications, such as PGD, have yet to be delineated. Additionally, viral transmission, reactivation and novel infections can all cause complications in LTRs. The extent of donor virus persistence in particular has not been well studied, which warrants a longitudinal investigation of the virome in this vulnerable population. Finally, deep sequencing of understudied niches within the human body may reveal novel eukaryotic viruses with unknown links to human health and disease. Therefore, the aims of the work presented in this dissertation are as follows:

1) What is the structure of the viral microbiome prior to and following lung transplantation?

Previously, our group analyzed the DNA virome within the upper and lower respiratory tract of LTRs and healthy individuals. A striking finding was the large numbers of metagenomic sequences derived from diverse anelloviruses. Quantification of anellovirus DNA in BAL and oral wash revealed that levels were 50-fold higher in LTRs than in healthy adults. Higher anellovirus loads in the lung were also found to correlate with dysbiotic bacterial communities (Young et al., 2015). Based on the observations in this cross-sectional study of samples taken at heterogeneous time points following organ transplantation, we sought to define the virome, with a particular focus on *Anelloviridae* populations, in donated organs and in LTRs in the first year post-transplantation. In CHAPTER 3, we use metagenomic and targeted molecular approaches to describe markedly elevated levels of anellovirus DNA in the lungs of donors prior to organ recovery. In CHAPTER 4, we corroborate previous findings that LTRs have abundant anelloviruses in peripheral blood and expand upon the previous detections of these viruses in the pulmonary compartment. Specifically, we demonstrate that anellovirus populations present in the donor lung can persist in the allograft and disseminate into the peripheral blood, and that anelloviruses circulating in the recipient prior to transplantation can re-enter the allograft. Therefore, lung transplantation can result in transfer of whole anellovirus communities from graft to host and vice versa.

2) Is there a relationship between the viral microbiome and primary graft dysfunction?

Given that PGD is associated with innate immune activation, we hypothesized that lung viruses, including anelloviruses, might be associated with PGD. In CHAPTER 3, we analyze extracellular viruses present in acellular BAL sampled from lung donors

immediately prior to organ recovery and from the allograft one hour after transplant and reperfusion in a case-control study of 22 pairs of LTRs. We also assayed recipient serum samples obtained at the time of reperfusion and investigated whether anellovirus levels were correlated with host cell gene expression, which was previously reported to correlate with PGD in these patients. While absolute viral levels did not distinguish PGD cases from controls, changes in anellovirus levels during the perioperative period were significantly associated with PGD. This difference is consistent with the hypothesis that PGD is linked to tissue viability or immune activation that may restrict viral levels in the lung. However, we did not find significant associations between peri-transplant viral dynamics or lung levels and host gene expression patterns.

3) Can novel eukaryotic viruses be discovered by mining large metagenomic datasets?

Shotgun metagenomic sequencing is a powerful tool to study the virome. The extensive databases generated from virome studies allow newly discovered viruses to be associated with specific environments, body sites and disease states using bioinformatics comparisons over large numbers of publically available sequences. In CHAPTER 5, we introduce a new family of small, circular DNA viruses, *Redondoviridae*, originally discovered in BAL of LTRs. After querying over 6000 samples from publicly available shotgun viral metagenomic datasets for homology to redondoviruses, we report that they primarily occur in human oral and respiratory samples and are notably absent in contamination controls and non-human samples. Redondoviruses encode two proteins similar to capsid and Rep proteins from other CRESS viruses, and one highly conserved open-reading frame without homology to any known protein family. Comparison of the new genomes to samples from human disease states showed an association with periodontal disease and critical illness. Redondovirus sequences disappeared with successful therapy, motivating studies of the

association of these viruses with periodontal disease. Thus we propose that *Redondoviridae* comprise a novel, widespread family of human viruses.

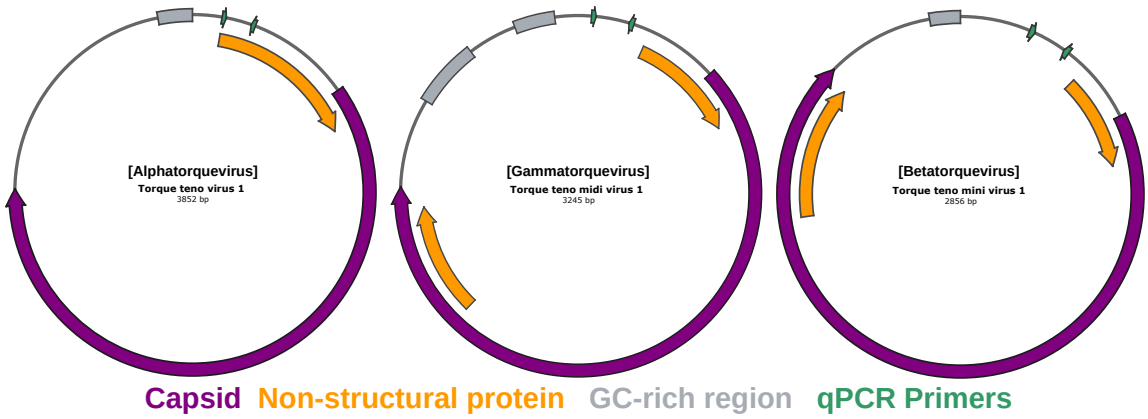


FIGURE 1.1: GENOME ORGANIZATION OF ANELLOVIRIDAE

Genomic organization of three human-infecting genera of *Anelloviridae;* Alphatorquevirus, Betatorquevirus and Gammatorquevirus, which each currently comprise of 31, 16, and 15 genomes respectively (NCBI Viral Genomes, 2018). Species demarcation is based on ORF1 (purple) nucleotide identity with a cutoff value of 35%. The circular genome is negative sense ssDNA with a single intergenic region containing one or more GC-rich regions. The putative capsid protein contains a conserved amino-terminus rich in basic amino acids, common in other viral coat proteins. The number of non-structural proteins predicted for each virus varies. While humans are hosts for all three genera, the definitive host range for each has not been determined and may include non-human animals as well. In this work, primers targeting the region upstream or within the ORF2 coding sequence were used to quantify diverse human anelloviruses. Modified from (Rosario et al., 2012).

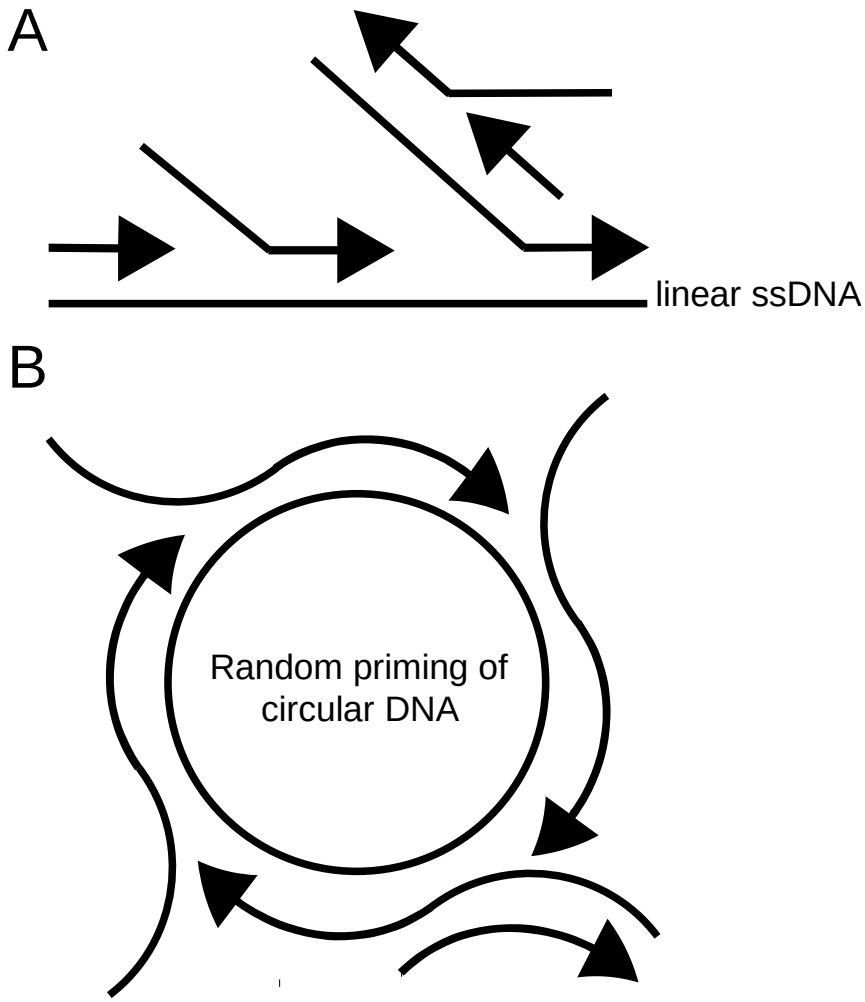


FIGURE 1.2: MULTIPLE DISPLACEMENT AMPLIFICATION STRATEGY
In A, the random-primed linear amplification of genomic DNA is shown. Secondary priming events are initiated from primary products. In B, the rolling circle amplification of circular DNA generates concatemers of the original template. DNA synthesis initiates from random oligonucleotide primers, indicated by arrowheads, and is performed by a DNA polymerase with strand-displacement ability. Secondary priming events occur on the displaced product DNA strands. Adapted from (Dean et al., 2002).

CHAPTER 2: MATERIALS AND METHODS

2.1 ETHICS STATEMENT

These studies were carried out in strict accordance of the regulations set forth in the Federal Policy for the Protection of Human Subjects or the "Common Rule". All participants provided written informed consent, and all specimens were obtained under protocols approved by institutional review boards (IRB) at their respective institutions.

2.2 STUDY POPULATION AND PRIMARY SAMPLE COLLECTION

Stored samples and clinical data were obtained from participants who had previously been prospectively enrolled in the multicenter Clinical Trials in Organ Transplantation-03 (CTOT-03) study (NCT00531921). Bronchoalveolar lavage (BAL) was collected in the operating room from donors immediately before organ procurement (donor BAL) and from allografts in recipients one hour after reperfusion (recipient BAL), as described previously (Cantu et al., 2013). Serum was also obtained from recipients 1 h after reperfusion. Additional BAL and serum were obtained during routine surveillance bronchoscopy and when indications for further testing arose (Figure 2.1). Specimens were prospectively collected following informed consent under IRB-approved protocols; samples analyzed here were retrieved retrospectively from stored specimens. A set of bronchoscope prewashes obtained for a separate study (Clarke et al., 2017a) and processed in an identical manner were included in the analysis.

In CHAPTER 3, we selected 23 grade 3 PGD cases and controls matched for donor age and pre-transplant recipient diagnosis as described in a previous report (Cantu et al., 2013). For one PGD case, no specimens remained from their matched control, and one additional PGD case was included for which there was no matching control. Consequently, samples were available from 22 case–control pairs and two unmatched PGD cases (Table 3.1). For 15 pairs, all perioperative BAL specimens were available for virome analysis,

whereas for seven pairs, insufficient sample remained from one or more lung samples (Tables 3.1 and Supplemental Table 3.1).

Subjects analyzed in CHAPTER 4 are described in (Table 4.1). All subjects except one (12-12) received induction immunosuppression (basiliximab). One subject (12-09) experienced grade 3 primary graft dysfunction within 72 hours of transplantation. Five subjects experienced one or more episodes of acute cellular rejection, which was treated with augmentation of immunosuppression. Common bacteria and fungi identified by routine clinical culture of donor BAL included *Staphylococcus sp., Streptococcus sp.,* and *Candida albicans*. Cytomegalovirus (CMV) serology was performed as indicated in Supplemental Table 4.3, but cultures were either negative or not performed for viral agents. Most subjects received microbial prophylaxis for CMV and *Pneumocystis jirovecii* as well as additional antibacterial or antifungal treatments.

Acellular BAL samples were obtained from healthy adult volunteers, as reported previously (Supplemental Table 3.2) (Charlson et al., 2011). Serum from a separate group of 11 healthy adult volunteers was also collected.

In CHAPTER 5, critically ill subjects were enrolled at the Hospital of the University of Pennsylvania medical Intensive Care Unit (ICU) within 24 hours of admission to the ICU. Informed consent was obtained under IRB protocol 823392. Oral swabs (n = 198), endotracheal aspirates (n = 87), and stool (n = 16) from 67 subjects were available to be queried by qPCR. Additional metagenomic sequence data (human, animal, and environmental) were derived from publicly available data repositories (Supplemental Table 5.3).

2.3 SAMPLE PROCESSING

To minimize cellular material, whole BAL was centrifuged at 960 x g for 10 minutes at room temperature and the supernatant was aliquoted and stored at -80°C. Serum was

collected in EDTA tubes, separated by centrifugation, and stored at -80°C. Virus-like particles (VLPs) were purified from 0.2 to 2 mL of acellular BAL or serum depending upon availability. BAL samples were thawed at 4°C and $MgSO_4$ and dithiothreitol were added to a final concentration of 10mM. To enrich for VLPs, the acellular BAL samples were concentrated using an Amicon Ultra-4 100kDA molecular weight cutoff filter (EMD Millipore; Billerica, MA) and the filters washed with 1-2 mL SM Buffer (Sambrook and Russell, 2001). The concentrated VLP preparations were treated with DNase I (225 Units) and RNase (7.5 µg) (Roche; Indianapolis, IN) at 37°C for 15 minutes to digest non-encapsulated (i.e. non-viral) nucleic acids, and the enzymes were inactivated at 70°C for 5 minutes. Whole serum samples were nuclease treated as above and the enzymes were inactivated by immediate addition of Buffer AC from the QIAmp Ultrasens Virus kit (Qiagen; Hilden, Germany).

Nucleic acids were extracted using QIAmp Ultrasens Virus kit (Qiagen; Hilden, Germany) per manufacturer instructions with the following modifications: 5 µg of linear polyacrylamide (Life Technologies; Carlsbad, CA) instead of carrier RNA was used and samples were eluted in 60µL of RNAse-free elution buffer containing 10mM Tris-HCL, 1mM EDTA at pH 7.5. DNA was stored at -20°C and RNA was stored at -80°C until processing. To detect RNA viruses, 20 µl of the DNA/RNA mixture were treated with 10 units of RNAse-free, recombinant DNAse (Roche; Indianapolis, MN) for 20 minutes at 37°C. 5 µL of each sample was then reverse transcribed. First strand cDNA synthesis was completed using SuperScript III First Strand Synthesis kit (ThermoFisher; Waltham, MA) and Primer A (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3'), to allow for random priming (Wang et al., 2003). Two rounds of second strand synthesis, again using Primer A for random priming, were performed using Sequenase Version 2.0 DNA polymerase (Affymetrix; Santa Clara, CA). The dsDNA product was then amplified by adding Primer B (5'-GTTTCCCAGTCACGATC-3')

(Wang et al., 2003) with AccuPrime Taq High Fidelity DNA polymerase (ThermoFisher; Waltham, MA) with the following reaction conditions: 75.5 µL of molecular grade H2O, 10 µl of 10x PCR Buffer I, 4 µL of 50 mM MgCl2, 2.5 µL 10mM dNTPs, 1 µL 100µM Primer B, 1 µL Taq and 6 µL dsDNA product. Products were amplified at 94°C for 2 min, 94°C 30sec, 40°C 30sec, 50°C 30 sec, 72°C 1 min for 40 cycles.

Extracted DNA was subjected to multiple displacement amplification (MDA) using random hexamers with Illustra GenomiPhi V2 DNA (GE Healthcare; Little Chalfont, UK). Whole genome amplified DNA and PCR amplified complementary DNA (cDNA) were quantified using Picogreen dsDNA Quantitation Reagent (ThermoFisher; Waltham, MA). Sequencing libraries were prepared using Nextera XT DNA Sample Preparation Kit (Illumina; San Diego, CA) using 1 ng of input DNA. Individual bar-coded samples were pooled in equimolar amounts into 4 separate libraries after quantification with KAPA Illumina Library Quantification (KAPA Biosystems; Wilmington, MA). Samples that did not have sufficient DNA for sequencing were excluded from the final pools. Libraries were sequenced either using a 2x125 paired-end run on the Illumina HiSeq 2500 platform or 2x250 paired-end run on the Illumina MiSeq platform. Each library also included 1-3 environmental controls subjected to the entire viral microbiome procedure.

DNA from oral swabs was extracted in single-tube DNeasy PowerSoil Kit (Qiagen; Hilden, German) and followed manufacturer's protocol except for two 50 µL elutions with buffer C6. Endotracheal aspirate was extracted with a 96-well format of the same kit.

2.4 TARGETED VIRAL ASSAYS

Anellovirus qPCR was performed on extracted DNA using primers NG779 (5'-ACWKMCGAATGGCTGAGTTT-3') and NG781 (5'-CCCKWGCCCGARTTGCCCT-3') that target human *Anelloviridae* members including torque teno virus (TTV), torque teno midi and torque teno mini virus (Ninomiya et al., 2008, Young et al., 2015). Values were

normalized to reflect equivalent volumes of starting BAL or serum and represent the average of three technical replicates. Verification of amplification of the target sequence was performed by gel visualization of PCR products on select samples using MDA DNA and by melt-curve analysis performed by 7500 Fast Real Time qPCR system (ThermoFisher; Waltham, MA) on all samples. The lower limit of detection was 38 copies/reaction; samples falling at or below this threshold were arbitrarily set to a minimum value.

In CHAPTER 5, extracted DNA from samples with contigs found to have homology to redondovirus genomes were amplified and cloned from 7 samples. Primers (Supplemental Table 5.1) were designed to amplify redondovirus genome sequences from DNA extracted from BAL that underwent MDA with Illustra GenomiPhi V2 DNA (GE Healthcare; Little Chalfont, UK). PCR was performed with AccuPrime™ Taq DNA Polymerase System (ThermoFisher; Waltham, MA, USA) using 1 μL of whole-genome-amplified product, 20 μM of forward and reverse primers and 0.2 μL Taq polymerase in a total volume of 50 μL. Products were visualized on 1-1.5% ethidium bromide agarose gels (Supplemental Figure 5.1). Amplicons were cloned and validated by using the Sanger sequence method on an ABI 3730XL (Applied BioSystems; Waltham, MA, USA) instrument. Full redondovirus genomes were either *de novo* synthesized (BioBasic; Markham, ON, CA) or cloned by Gibson assembly (NEB; Ipswitch, MA, USA) and also verified by Sanger sequencing.

To detect redondovirus sequences in BAL samples, a TaqMan-based qPCR assay (Supplemental Table 5.1) was designed targeting the genomic region encoding the capsid gene. For each sample, triplicate 20 μL reactions containing 4 μL of template DNA (depending on sample availability), 0.33 μL forward and reverse primers (18μM), 0.33 μL probe (5μM), 10 μL TaqMan Fast Universal PCR Master Mix (Applied Biosystems; Waltham, MA, USA) and 5 μL water were analyzed on a QuantStudio 5 Real Time PCR System (Applied Biosystems; Waltham, MA, USA) with the following cycling profile: 20 sec at 95°C for 1 cycle,

and 40 cycles of 95°C for 3 sec and 60°C for 30 sec (signal collection). A linearized plasmid containing the complete Human lung-associated brisavirus RC genome in a pUC57 vector was used as a 7-point standard curve ranging from 75 to 30,000,000 copies/reaction. Amplification signal was required in 2 out of 3 wells to be scored as positive.

To survey samples from critically ill and healthy individuals, extracted DNA was first subjected to selective whole-genome amplification (SWGA) using primers designed with the software described in (Clarke et al., 2017b). Each reaction contained 2 µL Phi29 10x Buffer (NEB; Ipswich, MA), 1 µL phi29 polymerase, 0.2 µL bovine serum albumin (10 mg/mL), 100 µM total of 20 primers (final concentration of each primer was 2µM), 2 µL of 10mM dNTPs, and 1 µL of template DNA in a total volume of 20 µL. Reactions underwent a step-down amplification process by incubating at 35°C for 5 min, 34°C for 10 min, 33°C for 15 min, 32°C for 20 min, 31°C for 30 min and then 30°C for 16 hours, followed by a heat inactivation step (65°C for 15 min) as previously described (Clarke et al., 2017b). After SWGA, 4 µL of product was queried in duplicate using a more sensitive TaqMan-based qPCR assay (Supplemental Table 5.1) that was designed targeting a conserved region of the capsid gene (*Cp)*. A linearized plasmid containing the complete genome of human lung-associated brisavirus RC was used for the 9-point standard curve ranging from 10 to 10^6 copies per reaction. Negative and positive controls were included in each run to evaluate inter-assay variability. The positive control was 10^4 copies of the standard curve plasmid containing the viral genome spiked into DNA extracted from a redondovirus-negative endotracheal aspirate sample and also subjected to SWGA.

2.5 BIOINFORMATICS PIPELINES

Paired-end sequence reads from the HiSeq and MiSeq machines were de-multiplexed and quality-trimmed. The Burrows-Wheeler Alignment tool (Li and Durbin, 2010) using hg18/19 (NCBI 36/37) or Genome Reference Consortium Human Build 38

(GRCH38) (downloaded 2 June 2016) as the reference genome identified human sequences which were then removed. FASTQ files were quality-trimmed (Q=33) and paired reads converted to FASTA format. Samples were further filtered to remove reads from PhiX whose genome is used during sequencing.

In the analyses presented in CHAPTER 3, high quality, non-human reads were aligned using the Basic Local Alignment Search Tool (BLAST) to the NCBI Viral Database (downloaded July 2015, containing 6,402 entries). The top hit for each read was recorded (Expected value < 10^{-5}). Consensus-based taxonomic assignments using output from the NCBI Viral Database BLAST search results were generated using BROCC (Dollive et al., 2012), using 80% minimum identity for species and genus consensus formation. Viral species with fewer than 20 BROCC-filtered BLAST hits per sample to a reference genome were excluded. Alignments to reference viral genomes were performed using Bowtie 2 (Langmead and Salzberg, 2012) and visualized using Integrated Genomics Viewer (Robinson et al., 2011).

For the analysis described in CHAPTER 4, high quality non-human reads were classified by Kraken (Wood and Salzberg, 2014) against the RefSeq standard database containing complete fungal, bacterial, archaeal, and viral genomes (downloaded 5 June 2016) and GRCH38. Hits to viruses known to be reagent contaminants or otherwise spurious (Clarke et al., 2017a) were removed from further downstream analysis. Multiple viruses were detected in extraction controls, although most hits to viral species were based on very few reads rendering validation by read alignment ineffective. Based on previous work, we believe these hits derive from multiple sources. For example, hits in extraction controls overwhelmingly consisted of bacteriophage of skin bacteria likely introduced during sample handling. Additionally, sparse reads of eukaryotic viruses likely originated from cross-contamination of highly abundant sequences from actual clinical samples

(Supplemental Figure 4.2). To filter out assignments likely arising from cross-contamination, while keeping those that could be validated by manual inspection of alignments to reference genomes, an empirical threshold of 10 read pairs assigned to a viral family was used to call a positive detection.

Contigs were assembled from reads within each sample using IDBA-UD (Peng et al., 2012), followed by CAP3 (Huang and Madan, 1999). Partial viral genomes were identified by performing nucleotide BLAST of contigs to the NCBI viral genomes database (downloaded 17 March 2016) retaining a single target genome hit with an E-value < 10^{-10}. This bioinformatics analyses described was performed using the Sunbeam pipeline (https://github.com/eclarke/sunbeam). A very sensitive global alignment of read pairs to *Anelloviridae* contigs was performed using Bowtie2 (Langmead and Salzberg, 2012) with the following additional settings (--mp 7 --no-mixed --np 2 --rdg 6,4 --rfg 6,4 --no-discordant) that increased mismatch, gap and ambiguity penalties. Reads that aligned with MAPQ scores > 23 were retained using SAMtools (version: 1.3.1)(Li et al., 2009) and the fraction of the genome covered was generated using BEDtools (Quinlan and Hall, 2010). Alignments were visually inspected using Integrative Genomics Viewer (Robinson et al., 2011). For samples positive for human viruses by the k-mer based approach, a very sensitive local alignment of read pairs to reference viral genomes was done using Bowtie2. Genome coverage and visualization was carried out as above.

To query viral metagenomic datasets for the presence of redondoviruses (CHAPTER 5), reads from viral metagenomic projects available in the Sequence Read Archive (SRA) or the Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) or the Human Oral Microbiome Database and from datasets from the University of Pennsylvania were processed in the following steps: 1) adaptor-trimmed single or paired-end reads were downloaded using fastq-dump (Leinonen et al., 2011); 2) a sensitive local alignment of

either single reads or read pairs to redondovirus genomes was performed using Bowtie2 (Langmead and Salzberg, 2012); 3) alignments were processed and genome coverage was calculated with SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010); and 4) alignments were visualized with a custom R (version 3.2.3) script (Ihaka and Gentleman, 1996) (R packages used: magrittr, ggplot2, reshape2).

Samples in which 25% of any redondovirus genome was covered were further analyzed using the Sunbeam pipeline (Clarke et al., 2018b) to process reads and to build and annotate contigs using MEGAHIT (Li et al., 2015a) and BLASTn (Altschul et al., 1990). Contigs were further refined by overlap consensus assembly using CAP3 (Huang and Madan, 1999) and manually checked for circularity and presence of key genomic features with CloneManager 9 (Scientific & Educational Software; Denver, CO).

2.6 GENE EXPRESSION ARRAY AND GENE SET VARIATION ANALYSIS

In CHAPTER 3, we tested correlations between anellovirus levels in the lung and differential host gene expression, employing gene expression data previously generated using Affymetrix Human Gene 1.0 ST Arrays (Affymetrix; Santa Clara, CA) to quantify global gene expression levels in the BAL cell pellet (Cantu et al., 2013) (Supplemental Tables 3.4–3.6). Robust Multi array Average normalized log2-transformed signal intensities from Affymetrix GeneChips were analyzed using Array Studio software (http://www.omicsoft.com/). Outliers implicated by Principal Components Analysis clustering of the first two principal components and median absolute deviation scores were excluded. Intensity cutoff for transcript clusters to be included in statistical analysis was determined by variance vs. mean intensity plots. In total, 16,009 transcript clusters were included in later statistical analyses. Gene Set Variation Analysis (GSVA) (Hänzelmann et al., 2013) was employed to statistically identify KEGG and Biocarta pathway gene sets (http://www.broadinstitute.org/gsea/msigdb/, extracted from the C2 curated gene set

collection) associated with specified phenotypes. GSVA is a non-parametric, unsupervised method used to calculate for each gene set an enrichment score for each individual sample in the study. The enrichment scores are a function of the relative expression of genes within vs. outside the gene set in a genome-wide analysis, with estimation of variation of gene set enrichment over the samples independent of any class label. Standard statistics were applied to the enrichment scores to assess differences (e.g., general linear modeling) and similarities (e.g., clustering methods) of enrichment of gene sets among groups of samples. Calculation of GSVA gene set enrichment scores was conducted in R version 3.1.1 (Ihaka and Gentleman, 1996).

2.7 DNA AND AMINO ACID SEQUENCE ANALYSIS

In CHAPTER 4, open reading frames (ORFs) were predicted and translated using a custom Python script from size-filtered *Anelloviridae* contigs discovered in Donor BAL and recipient serum samples. Proteins were identified by BLASTp (E-value < 10^{-5} against a database of 75 ORF1 sequences from representative human *Anelloviridae* members (Alphatorquevirus, Betatorquevirus, and Gammatorquevirus). A minimum alignment length of 250 amino acids was required. Proteins identified within an individual or within a cognate-donor recipient pair were aligned with MUSCLE (v3.8.31) (Edgar, 2004) using default settings. Alignments were then filtered to mask regions containing gaps longer than 2 amino acids present in 50% or more of the sequences and visualized. The percent identity between all protein sequences was calculated using a BLOSUM62 substitution matrix after pairwise alignment using the Biostrings package (Pages et al., 2009) in R version 3.2.3 (Ihaka and Gentleman, 1996).

In CHAPTER 5, the EMBOSS einverted utility (Rice et al., 2000), Mfold (Zuker, 2003) or CloneManager Professional 9 (Scientific & Educational Software; Denver, CO) was used to predict and visualize energetically favorable DNA structural features potentially important

for replication. Forna was used to plot stem loop structures (Kerpedjiev et al., 2015). Nucleotide and protein alignments were performed using MUSCLE (version 3.8.31) (Edgar, 2004). Phylogenetic trees were built using PhyML (Guindon et al., 2010) using sequences from 2-3 representative species of established viral families and all full-length protein sequences of novel redondoviruses. Branch support was quantified by the approximate likelihood-ratio test (Anisimova and Gascuel, 2006) and visualized using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Consensus motif logos were generated using WebLogo (Crooks et al., 2004). Conserved domains within Rep proteins were detected using NCBI's CD-search against the Pfam database (v30.0, E-value < 10-2). Protein folding predictions were done using PHYRE2 (Kelley et al., 2015) using default parameters.

To predict prokaryotic ribosomal binding sites (RBS), we implemented the algorithm described in (Krishnamurthy and Wang, 2018) in Python (version 3.6). Briefly, we extracted 18 nucleotides in the untranslated region immediately upstream of start codons and searched for prokaryotic RBS (full: AGGAGG; partial: AGGAG, GGAGG, AGGA, GGAG, GAGG).

To analyze synonymous and non-synonymous substitution rates, we aligned protein sequences using MUSCLE (Edgar, 2004), built phylogenetic trees with PhyML (Guindon et al., 2010) generated codon alignments with PAL2NAL (Suyama et al., 2006) and used HyPhy (Pond et al., 2005) to perform dN/dS analysis with FUBAR (Murrell et al., 2013) to predict sites under positive selection. For overlapping genes, only non-overlapping portions were considered—overlapping coding regions were identified and excluded using pyviko (Taylor and Strebel, 2017).

2.8 STATISTICAL ANALYSES

In CHAPTER 3, the Wilcoxon rank sum test, Wilcoxon signed rank test, Kruskal–Wallis test, paired two-tailed Student t-test and Spearman rank correlation tests were

performed in R version 3.2.3. Analysis of differences in TTV levels was performed on log10 normalized values. A false discovery rate (FDR) using the Benjamini–Hochberg method was applied for multiple testing correction.

In CHAPTER 4, we encountered several challenges in scoring representation of *Anelloviridae* lineages over time: 1) sequence depth differed among the samples after human filtering; 2) the degree of difference among genomes was variable, with some genomes in different individuals by chance showing similarity; and 3) different lineages were present in variable abundances. Therefore, we used several metrics to score the presence of *Anelloviridae* lineages and compare between samples.

To determine the relatedness of *Anelloviridae* sequences in longitudinal post-transplant lung and blood samples to *Anelloviridae* present in the initial donor lung and recipient serum samples, we generated *Anelloviridae* contigs from donor BAL and peri-transplant recipient serum, and aligned shotgun sequence reads from post-transplantation time points to those contigs. The fold coverage at each base and total mapped reads per kilobase per million sequenced reads were calculated using python and R scripts (https://github.com/sherrillmix/dnapy, https://github.com/sherrillmix/dnar). We then used a modification of the Gini index, a measure of inequality in a distribution of values (Gini, 1912), to holistically evaluate the evenness of read mapping to target genomes with the premise that authentic detections would result in deep and even coverage while uneven coverage represents alignment to conserved genomic regions or cross-contamination across samples. Differences in viral lineage detections, quantities, and homology between groups were tested using Wilcoxon Rank Sum Test and ANOVA. The relationship between *Anelloviridae* levels and clinical outcomes was tested using Student's t-test.

In CHAPTER 5, 20 datasets in which redondovirus genomes were found or were comprehensive studies of the human DNA virome were chosen for a targeted analysis of

human viruses. Specifically, reads from these datasets were aligned to 133 vertebrate viruses from the *Adenoviridae*, *Anelloviridae*, *Herpesviridae*, *Papillomaviridae*, *Parvoviridae* and *Polyomaviridae* families (downloaded from NCBI RefSeq on 20 August 2018). Alignments were done using the hisss pipeline (https://github.com/louiejtaylor/hisss) and analyzed in R (R packages used: tidyverse, reshape2, Biostrings, taxonomizr, UpSetR). Samples were considered positive for small (<10 kb) DNA viruses if greater than 25% of the target genome was covered. Samples were considered positive for large DNA viruses (>10 kb genomes), if greater than 10% of the target genome was covered (see Supplemental Figure 5.3). The distribution of the frequency of *Redondoviridae* and other viral family detection was tested using the Fisher's exact test with Bonferroni correction for multiple testing.

In studies of periodontitis, the difference in number of redondovirus reads in disease versus non-disease states were tested using the Wilcoxon signed-rank or rank-sum tests, depending on whether samples were paired or not.

2.9 DATA AVAILABILITY

Sequence data that have been filtered to remove contaminating human sequence are collected in BioProject records PRJNA390659 and PRJNA419524. The accession numbers for the viruses sequenced and reported in this paper are GenBank: MK059754-MK059772. Full details of each step of the Snakemake pipelines used in this work are available at https://github.com/sunbeam-labs/sunbeam and https://github.com/louiejtaylor/hisss. The script used for RBS analysis is available at https://github.com/louiejtaylor/find-prok-rbs.

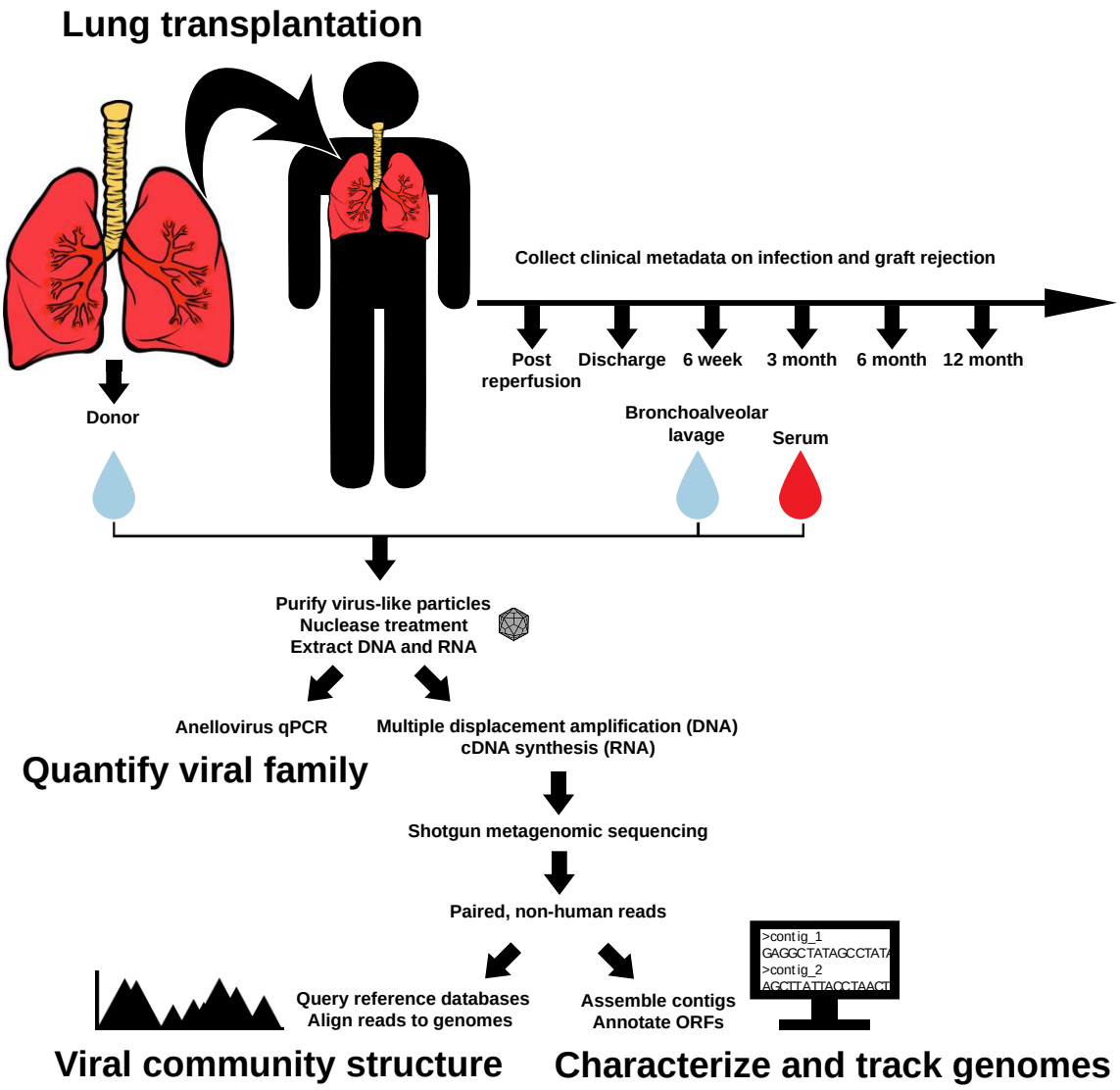


FIGURE 2-1: WORKFLOW FOR SEQUENCING THE VIRAL MICROBIOME OF LUNG TRANSPLANT RECIPIENTS

Bronchoalveolar lavage was collected from the organ donor prior to organ procurement, in the operating room after reperfusion in the lung transplant recipient and after transplantation in the Clinical Trials of Organ Transplantation 03 study. Contemporaneous serum was also collected from the lung transplant recipient. Samples were processed to enrich for viral DNA and RNA which were then extracted and prepared for shotgun metagenomic sequencing or assessed for anellovirus content using targeted qPCR. After sequences were filtered to remove low quality and identifiable human sequences, they are classified by comparison to reference viral genomes or assembled into longer sequences (contigs) whose gene content can be queried by predicting open-reading frames (ORFs).

CHAPTER 3: THE PERIOPERATIVE LUNG TRANSPLANT VIROME

The contents of this chapter have been previously published as:

Abbas, A.A., Diamond, J.M., Chehoud, C., Chang, B., Kotzin, J.J., Young, J.C., Imai, I., Haas, A.R., Cantu, E., Lederer, D.J., Meyer, K.C., Milewski, R.K., Olthoff, K.M., Shaked, A., Christie, J.D., Bushman, F.D., Collman, R.G., 2017. The Perioperative Lung Transplant Virome: Torque Teno Viruses Are Elevated in Donor Lungs and Show Divergent Dynamics in Primary Graft Dysfunction. Am. J. Transplant. 17, 1313–1324.

3.1 ABSTRACT

Primary graft dysfunction (PGD) is a principal cause of early morbidity and mortality after lung transplantation, but its pathogenic mechanisms are not fully clarified. To date, studies using standard clinical assays have not linked microbial factors to PGD. We previously used comprehensive metagenomic methods to characterize viruses in lung allografts >1 months after transplant and found that levels of anellovirus, mainly torque teno viruses (TTVs), were significantly higher than in nontransplanted healthy controls. We used quantitative polymerase chain reaction (qPCR) to analyze TTV and shotgun metagenomics to characterize full viral communities in acellular bronchoalveolar lavage from donor organs and postreperfusion allografts in PGD and non-PGD lung transplant recipient pairs. Unexpectedly, TTV DNA levels were elevated 100-fold in donor lungs compared with healthy adults ($p = 0.0026$). Although absolute TTV levels did not differ by PGD status, PGD cases showed a smaller increase in TTV levels from before to after transplant than did control recipients ($p = 0.041$). Metagenomic sequencing revealed mainly TTV and bacteriophages of respiratory tract bacteria, but no viral taxa distinguished PGD cases from controls. These findings suggest that conditions associated with brain death promote TTV replication and that greater immune activation or tissue injury associated with PGD may restrict TTV abundance in the lung

3.2 INTRODUCTION

Lung transplantation is the only long-term option for many patients with end-stage lung diseases. Median survival for lung transplant recipients (LTRs) is only 5.7 years (Yusen et al., 2015), the lowest among solid organ transplants (Watson and Dark, 2012). Factors limiting survival include primary graft dysfunction (PGD), infection, cellular and antibody-mediated rejection, and chronic lung allograft dysfunction (Yusen et al., 2015).

PGD is a form of acute lung injury characterized by hypoxemia, pulmonary edema and alveolar damage that occurs within 72 h of transplantation (Lee and Christie, 2011). Approximately 10–25% of LTRs experience severe PGD, which is a leading cause of death early after transplantation (Lee and Christie, 2011), and survivors of severe PGD are more likely to develop donor-specific antibodies and chronic rejection (Ius et al., 2014, Daud et al., 2007).

The pathogenesis of PGD is thought to involve both inflammatory and ischemia–reperfusion-related mechanisms. An inflammatory environment develops within the lung after donor brain death and is perpetuated after reperfusion by recruitment of lymphocytes, macrophages and neutrophils (Lee and Christie, 2011, Eppinger et al., 1997, Fiser et al., 2001, Eppinger et al., 1995, Naidu et al., 2003, Yang et al., 2009, Sharma et al., 2011, Johnston et al., 2012). Activated immune cells and reactive oxygen species generated during ischemia–reperfusion are thought to injure lung endothelium and epithelium (Lee and Christie, 2011). Host genetic variation in genes involved in oxidant stress responses and innate immunity correlate with risk of PGD (Diamond and Wigfield, 2013, Cantu et al., 2015). PGD has also been associated with inflammasome activation, pattern recognition receptor signaling and complement activation within the allograft (Cantu et al., 2013, Shah et al., 2014, Naka et al., 1997). These studies raise the possibility of a microbial contribution to pathogenesis.

Culture-based and molecular clinical assays have associated specific bacteria, fungi and community-acquired respiratory viruses with lung transplantation outcomes, such as the development of bronchiolitis obliterans syndrome (BOS) (Khalifah et al., 2004, Botha et al., 2008, Weigt et al., 2009, Gottlieb et al., 2009, Borewicz et al., 2013, Nakajima et al., 2011, Verleden et al., 2015, Charlson et al., 2012). Studies using high-throughput metagenomic sequencing have demonstrated that lung bacterial, fungal and viral communities are

aberrant in lungs of LTRs compared with healthy controls (Borewicz et al., 2013, Charlson et al., 2012, Young et al., 2015). LTRs have decreased bacterial diversity, increased bacterial load, outgrowths of pathogens and presence of atypical species (Borewicz et al., 2013, Charlson et al., 2012, Sharma et al., 2013, Dickson et al., 2014a). One study linked metagenomic bacterial community features and BOS (Willner et al., 2013).

Our group recently reported high levels of diverse torque teno viruses (TTVs) in acellular BAL fluid from LTRs sampled from 1 month to >10 years following transplant (Young et al., 2015). TTVs are small, nonenveloped, circular, negative-sense, single-stranded DNA (ssDNA), eukaryotic cell viruses belonging to the *Anelloviridae* family. These viruses are ubiquitous in the human population (Spandole et al., 2015), with initial infection occurring during childhood (Ninomiya et al., 2008, McElvania TeKippe et al., 2012, Chen et al., 2013). Changes in blood TTV levels correlate with altered immune states. Plasma TTV DNA levels increase in patients with human immunodeficiency virus who progress to AIDS and decrease following immune reconstitution with antiretroviral therapy (Thom and Petrik, 2007, Madsen et al., 2002, Devalle et al., 2009). Expansion of TTV is seen in blood of solid organ transplant recipients receiving immunosuppressive therapy (De Vlaminck et al., 2013, Görzer et al., 2014, Görzer et al., 2015). To date, TTV has not been identified as the etiological agent of any disease (Spandole et al., 2015, Hino and Miyata, 2007).

In this study, we investigated the hypothesis that lung viruses, including TTV, might be associated with PGD. To do this, we analyzed extracellular viruses present in acellular BAL sampled from lung donors immediately prior to organ recovery and from the allograft 1 h after transplantation and reperfusion. Two complementary approaches were used: 1) a TTV qPCR assay and 2) shotgun metagenomic sequence analysis of all viruses. These assays were also performed on recipient serum samples obtained at the time of reperfusion. We

also investigated whether TTV levels were correlated with host cell gene expression, which was previously reported to correlate with PGD in these patients (Cantu et al., 2013).

3.3 RESULTS

Study participants

Viral preparations from lung and blood specimens from 22 pairs of grade 3 PGD patients and matched controls plus two additional PGD patients were analyzed (Table 3.1 and Supplemental Table 3.1). Baseline demographics of study participants were reported previously (Cantu et al., 2013), and characteristics relevant to this study are summarized in Table 3.1. Most participants received cytomegalovirus (CMV) prophylaxis, and all but two (participants 12013 and 12016) received basiliximab for induction immunosuppression. Common bacteria and fungi identified by routine clinical culture included *Staphylococcus sp.*, *Streptococcus sp.*, and *Candida albicans*.

Analysis of TTV genome copy numbers by qPCR

High levels of TTVs were reported previously in the respiratory tract of LTRs in samples collected >1 mo after transplantation (Young et al., 2015) and in posttransplant blood (De Vlaminck et al., 2013, Görzer et al., 2014, Görzer et al., 2015). In this study, we investigated lung TTV levels at time of transplantation by quantifying TTV genomes in donor BAL and in recipient postreperfusion BAL and serum (Figure 3.1). Copies of TTV DNA per milliliter (mL) of acellular BAL varied up to 10,000-fold in both donor and recipient samples (Figure 3.1). Donor BAL obtained prior to lung procurement had a median of 670,600 TTV copies/mL, which was 100-fold higher than TTV levels in healthy adult BAL samples (median 6,631; $p = 0.0026$, Wilcoxon rank sum test).

Copies of TTV in postreperfusion BAL were approximately twofold greater than in samples taken before organ procurement, with a median of 1,102 000 copies/mL ($p = 0.89$ compared with donor BAL; $p < 0.001$ for comparison to healthy controls). There was a

positive correlation between TTV levels in lung samples obtained before and after transplantation ($p < 0.001$, Spearman's $q = 0.548$, Spearman rank correlation test) (Figure 3.1C). Consequently, donor lungs are markedly abnormal with regard to virome populations, even before transplantation.

Conversely, in serum, quantities of TTV immediately after reperfusion (median 35,530 copies/mL) were lower than in healthy adults (median 553,996 copies/mL; $p < 0.001$, Wilcoxon rank sum test). There was a positive trend correlating TTV levels in BAL and serum of LTRs immediately after transplantation ($p = 0.061$, Spearman's $q = 0.303$, Spearman rank correlation test) (Supplemental Figure 3.1).

The high levels of TTV DNA found in donor lungs prior to transplantation were surprising given that donors are selected because they lack overt lung disease. Therefore, we investigated clinical features of donors that might be associated with TTV levels (Supplemental Table 3.7). Lung TTV DNA levels were inversely correlated with donor age ($p = 0.036$; Spearman's $q = 0.328$, Spearman rank correlation test). In contrast, there was no correlation between lung TTV levels and donor cause of death, history of purulent secretions, aspiration or tobacco exposure. There was a positive trend correlating TTV levels in recipient BAL with organ ischemic time ($p = 0.065$; Spearman's $q = 0.298$, Spearman rank correlation test). However, no correlation was seen between postreperfusion TTV levels and use of cardiopulmonary bypass or nitric oxide during surgery, administration of blood products or fluids, transplant type, or recipient preoperative diagnosis (Supplemental Table 3.7).

We then investigated whether PGD was associated with TTV levels. When evaluated as absolute levels at a single time point, neither donor pretransplant nor postreperfusion BAL TTV levels correlated with PGD ($p = 0.89$ and $p = 0.82$, respectively, Wilcoxon signed rank test). Postreperfusion serum TTV levels also did not correlate with PGD ($p = 0.71$,

Wilcoxon signed rank test). TTV dynamics, however, differed significantly between PGD cases and controls (Figure 3.2). Non-PGD controls showed an average 3.5-fold increase in lung TTV levels over the perioperative period, whereas the increase was only 1.7-fold in PGD cases. When compared as pairs across the 15 complete sample sets, a 4-fold difference in TTV increase was observed between PGD cases and controls ($p = 0.041$; Wilcoxon signed rank test). Consequently, PGD is associated with a significantly lower rise in TTV levels during the peritransplant period.

Association of TTV levels and differences in mRNA expression in the lung

Gene set enrichment analysis in BAL cells from this cohort of patients previously identified several inflammasome and innate immune pathways for which perioperative change in expression correlated with PGD (Cantu et al., 2013). These pathways included NOD-like receptor signaling, Toll-like receptor signaling, the IL-1 receptor pathway and nuclear factor κB activation by nontypeable *Haemophilus influenza*.

We investigated whether TTV dynamics correlated with gene pathways enriched in PGD or with the top nine transcripts linked to these pathways. Neither the PGD associated pathways nor individual transcripts correlated with TTV dynamics (Supplemental Tables 3.4 and 3.5). We then tested whether absolute TTV levels in donor or recipient BAL correlated with transcript expression but again found no significant relationship (Supplemental Table 3.5). Finally, we queried the entire gene expression data set to determine whether any individual gene ($n = 16{,}009$) or pathway ($n = 403$) correlated with TTV levels in donor BAL, recipient BAL or perioperative dynamics. Several pathways and gene transcripts demonstrated a nominal correlation with lung TTV levels or dynamics, but none were significant after FDR correction (Supplemental Table 3.6). Consequently, lung TTV dynamics did not correlate detectably with host cell gene expression patterns.

Metagenomic shotgun sequencing of perioperative samples

To further investigate the lung virome around the time of transplant and to determine whether other viruses might be linked to PGD, we undertook metagenomic characterization of full viral populations in donor and recipient acellular BAL and in recipient serum. To validate recovery of RNA viruses, 10^6 plaque-forming units of bacteriophage Pseudomonas phage phi6 were spiked into a non-LTR BAL sample that was processed identically to LTR samples. Of the 369,669 nonhuman reads generated from this sample's cDNA library, 135,301 aligned by the Basic Local Alignment Search Tool (BLAST) with the tripartite phi6 genome. Coverage of the L, M, and S genomic segments was 71%, 82%, and 40%, respectively (Supplemental Table 3.3), confirming viral recovery and detection with our methods.

Applying our pipeline to LTR samples generated a median of 924, 291 filtered paired reads per sample (range 2,668–5,676,284 reads) in the BAL DNA library, 150, 150 reads (range 268–3,082,400 reads) in the BAL cDNA samples and 69,342 reads (range 5,241–815,079 reads) in serum DNA samples. On average, human sequences composed 54.6% (range 0.20–88.9%) of all high-quality reads in BAL DNA samples, 73% (range 1.18–98.6%) in BAL cDNA and 80.4% (range 13.2–95.8%) in serum DNA.

High-quality nonhuman reads were queried by BLAST against the NCBI viral database and viral taxonomic assignments generated using BROCC (Dollive et al., 2012). To control for contaminating sequences during sample and library preparation (Salter et al., 2014), buffer blanks were analyzed using the same workflow. On average, <1% of high quality nonhuman sequences could be assigned to known viral species. The remaining reads likely correspond to bacteria, poorly annotated bacteriophages, human reads that eluded the human filtering step, or potentially novel eukaryotic viruses.

We inspected read alignments to identify those that were spurious, that is, attributable to environmental contamination or unfiltered human sequences. Alignments

with Shamonda and Simbu viruses, for example, were confined to 90 and 40 base pair regions of the genomes, respectively, and had 92–100% identity to human 45S ribosomal RNA and thus were removed. The number of reads annotated as human endogenous retrovirus K tracked with the extent of human DNA in samples based on the number of reads aligning with the human genome and β-tubulin qPCR (data not shown) and were also removed.

Most participants had many reads aligning with human herpesviruses (HHVs) HHV-7, HHV-6A, and HHV-6B; however, sequence alignments involved regions of short direct repeats, and herpesvirus genome coverage never exceeded 1%. BLAST queries of these sequences against the human genome revealed high identity to human repetitive regions. Because these reads likely did not represent authentic herpesvirus detection, they were also removed. Finally, when very few reads align with a particular viral genome, low coverage makes it difficult to discriminate between coincidental alignment (e.g. to a short and/or non-unique region) and authentic virus detection; therefore, an empirical threshold was set at 20 reads in a sample aligning with a reference virus to confidently call virus detection. Assignments below this threshold are shown in Supplemental Tables 3.7 and 3.8.

After these filtering steps, a total of 105 viral species from 15 family-level groups were identified in BAL (two eukaryotic viruses, one plant virus and 12 bacteriophage families/unclassified species) (Figure 3.3 and Supplemental Figure 3.2). Anelloviruses, comprising multiple TTV species, were the most abundant eukaryotic virus in BAL. In serum, we identified reads that best aligned with 29 anellovirus species. In both lung and serum, populations of TTVs appeared highly diverse, both within and between individuals (Figure 3.4). The high levels of genome diversity within an individual sample could reflect multiple viral species or a few novel TTV viruses with low levels of identity to database reference strains.

In addition to TTV, the most abundant viruses seen in lung samples were DNA phages, including *Siphoviridae*, *Myoviridae* and *Podoviridae*, which infect a broad range of oral and respiratory tract bacteria such as *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Stenotrophomonas maltophilia* and *Pseudomonas aeruginosa* (Figure 3.3 and Supplemental Figure 3.2).

After filtering out HHV reads deemed spurious, as described above, we specifically queried the presence of authentic HHV sequences by aligning reads with ten HHV genomes; reactivation of HHVs, especially CMV, is a risk following organ transplantation (Kotloff and Thabut, 2011). No sample exceeded our threshold of 20 reads, but one donor BAL from a PGD case had 12 reads that uniquely aligned with CMV (Supplemental Table 3.8). Of note, our analysis of extracellular virions would not detect the presence of latent nonreplicating viral genomes within host cells.

One non-PGD donor BAL revealed sequences that aligned with circoviruses, which are recognized pathogens in birds and swine but in humans have been detected primarily in stool (Smits et al., 2014, Phan et al., 2014, Phan et al., 2015a). Other small single stranded circular DNA viruses in the proposed Cyclovirus genus within the *Circoviridae* family have been reported in cerebrospinal fluid and nasopharyngeal aspirates (Phan et al., 2014, Phan et al., 2015b). Our sequence reads covered only 9% of the closest circovirus reference genome, raising the possibility of a novel eukaryotic virus in this sample.

Although not reaching our threshold of 20 reads in any individual sample, we also detected a total of 26 reads across four samples that aligned with mimiviruses (Supplemental Table 3.8), which infect amoebae and have been reported recently in human lung samples but are of uncertain clinical significance (Saadi et al., 2013, Kutikhin et al., 2014).

Detection of RNA viruses was limited. The overwhelming majority of viral reads in the cDNA library were annotated as originating from DNA viruses. These reads might represent viral mRNA or incomplete removal of genomic DNA in cDNA preparation. In some samples, there were small numbers of short reads (reflecting genome coverage of 0.6%) aligning with a plant virus (*Tymoviridae*, Physalis mottle virus). The significance of these reads is uncertain. No RNA viruses known to infect humans were detected in the acellular BAL metagenomic analysis, including an absence of community-acquired respiratory viruses. No viral taxa distinguished PGD cases from controls.

3.4 DISCUSSION

In this study, we described markedly elevated levels of TTV in the lungs of donors prior to organ recovery. Although absolute levels of TTV did not distinguish PGD cases from controls, changes in TTV levels during the perioperative period were significantly associated with PGD.

Finding high levels of TTV in donor lungs was unexpected. Brain death is associated with profound hemodynamic, neurohumoral and inflammatory responses that result in proinflammatory cytokine elevations and leukocyte infiltration of the lung (Kutsogiannis et al., 2006, Faropoulos and Apostolakis, 2009). Although its host cell tropism is poorly understood, TTV is reported to reside in mononuclear immune cells (Maggi et al., 2001, Mariscal et al., 2002, Zhong et al., 2002, Takahashi et al., 2002, Focosi et al., 2015) and to increase in inflammatory states (Maggi et al., 2003, Walton et al., 2014). Consequently, local inflammation and leukocyte recruitment to the lung following brain death might result in enhanced TTV replication. In addition, corticosteroids are generally given to organ donors to modulate inflammation after brain death, and corticosteroid-induced immunosuppression could be another reason for the high TTV levels in donors because host immune function is thought to control virus replication (Thom and Petrik, 2007) (Görzer et

al., 2014, Focosi et al., 2010). Future studies will be needed to determine whether lung TTV levels increase in critically ill patients generally or if high levels are specifically associated with brain death and/or corticosteroid or other donor treatments.

TTV levels increased from pre- to posttransplant lung samples, but the magnitude of increase was lower in PGD cases than in controls. This difference is consistent with the hypothesis that PGD is linked to tissue viability or immune activation that may restrict TTV levels in the lung. PGD is associated with activation of multiple innate immune responses (Cantu et al., 2013, Naka et al., 1997, Shah et al., 2014) that might control TTV but that also mediate injury contributing to PGD. Alternatively, in addition to immune cells, TTV has been reported to replicate in respiratory epithelia and lung tissue (Pifferi et al., 2008, Okamoto et al., 2001). Because ischemia–reperfusion injury may result in decreased tissue viability (Eppinger et al., 1997, Eppinger et al., 1995, Fiser et al., 2001, Naidu et al., 2003, Yang et al., 2009, Sharma et al., 2011, Johnston et al., 2012), this may also limit TTV replication in PGD if permissive cell types are affected.

A prior analysis found that PGD was associated with enrichment in several gene pathways in this cohort; however, we did not find significant associations between peritransplant TTV dynamics or lung TTV levels and host gene expression patterns. This finding suggests that the association between TTV dynamics and PGD reflects a relationship distinct from previously identified mechanisms linked to host gene expression patterns.

The range of TTV levels in serum of our healthy controls was similar to previous reports (Haloschan et al., 2014, Christensen et al., 2000, Moen et al., 2003), but our postreperfusion serum levels were significantly lower (Figure 3.1). These levels immediately after transplant contrast with studies showing elevated blood TTV levels of chronic transplant recipients who were immunosuppressed (De Vlaminck et al., 2013, Görzer et al., 2014, Görzer et al., 2015). A recent report, however, described decreased TTV

levels at day seven in kidney–pancreas transplant recipients receiving anti-thymocyte globulin (ATG) that were ascribed to lymphocytolytic properties of ATG on mononuclear cells believed to support TTV replication (Focosi et al., 2015). Only slight decreases were reported at day seven in those receiving basiliximab, which is not lymphocytolytic but prevents T-cell proliferation. We studied samples immediately following induction with basiliximab, thus our results could reflect potential earlier effects of basiliximab on TTV replication due to acute suppression of T-cell proliferation. Two participants who did not receive induction immunosuppression had markedly lower serum TTV levels compared with other LTRs and healthy adults (Supplemental Table 3.7), in contrast to the predictions based on this idea; however, induction immunosuppression is withheld in recipients with intercurrent processes, which themselves might affect TTV levels and preclude definitive conclusions.

Apart from anelloviruses, eukaryotic DNA viruses detected by metagenomic sequencing of acellular BAL were sparse, and none correlated with PGD. We anticipated that HHVs might be prevalent, but only one sample was positive, with few reads aligning with CMV. One sample had sequences annotated as a circovirus, the significance of which is unclear but could reflect a novel virus. Most viral species detected were bacteriophages, and detection of phages known to infect respiratory tract bacteria implies the presence of their bacterial hosts during transplantation.

In contrast to DNA viruses, we did not find evidence of abundant human RNA viruses in these lung samples. Other studies have succeeded in identifying RNA viruses in human specimens using similar methods (Wang et al., 2003, Wylie et al., 2012, Holtz et al., 2014, Handley et al., 2012), and our internal-control spiked sample confirmed viral RNA recovery. We conclude that actively replicating RNA viruses, which would be released as

extracellular particles and thus be present in acellular BAL, are not common and abundant in lung allografts at the time of lung transplantation.

Our initial analysis revealed many sequences that aligned with human and other eukaryotic viruses, but on close inspection, some alignments appeared spurious. These findings underscore the need to carefully inspect viral assignments to distinguish authentic sequences from artifacts. In addition, alignments with very low read counts are difficult to authenticate, highlighting the challenge in interpreting such low-abundance sequences. Consequently, virome analysis based simply on automated read alignments may result in overcalling viruses that may not be genuinely present.

This study has several limitations. First, we used acellular BAL, which is suitable for analysis of the extracellular virome but cannot detect viral nucleic acids present within cells in nonreplicating or latent forms that are not releasing progeny into the alveolar or airway lumen. In addition, use of acellular BAL limits the ability to characterize bacterial and fungal populations in the sample. Second, there is likely geographic heterogeneity within the lung (Willner et al., 2012a), and any individual sample may represent only part of the viral community present. Third, because PGD can occur up to three days after transplant, it is possible that differences in viral communities could emerge later than our perioperative time points. Last, we did not have pretransplant recipient serum uniformly available to query whether TTV dynamics in the blood changed in the peritransplant period and were related to PGD incidence.

Finally, an important finding was that peritransplant samples did not contain evidence indicating active replication of RNA viruses or typical community-acquired respiratory viruses known to be associated with late transplant complications. Small numbers of sequences aligning with unexpected nonhuman viruses, as reported in this

study, might reflect novel human viruses. Further studies are warranted to determine whether TTV or novel viruses are linked to long-term outcomes of lung transplantation.

3.5 ACKNOWLEDGEMENTS

We are grateful to participants and volunteers for providing specimens, to the Clinical Trials in Organ Transplantation (CTOT) investigators (Jason D. Christie, Abraham Shaked, Kim M. Olthoff, Rita K. Milewski, Keith C. Meyer, David J. Lederer, Edward Cantu, Andrew R. Haas) and the staff participating in the CTOT-03 and Lung HIV Microbiome Project studies (Ize Imai), and to members of the Bushman and Collman laboratories for help and suggestions. Joshua M. Diamond and Bao-li Chang performed gene expression microarray analysis. Christel Chehoud provided bioinformatics support. Jacque C. Young and Jonathan J. Kotzin aided in sample processing. Frederic D. Bushman and Ronald G. Collman assisted in the design of experiments, interpretation of data, and writing of the manuscript. This work was supported by NIH grants R01-HL113252, U01-HL098957 and R01-HL087115, K24-HL115354, and received assistance from the Penn Center for AIDS Research (P30-AI045008) and the PennCHOP Microbiome Program. A.A.A. was supported by NSF grant DGE-1321851, and J.M.D. was supported by K23-HL121406.

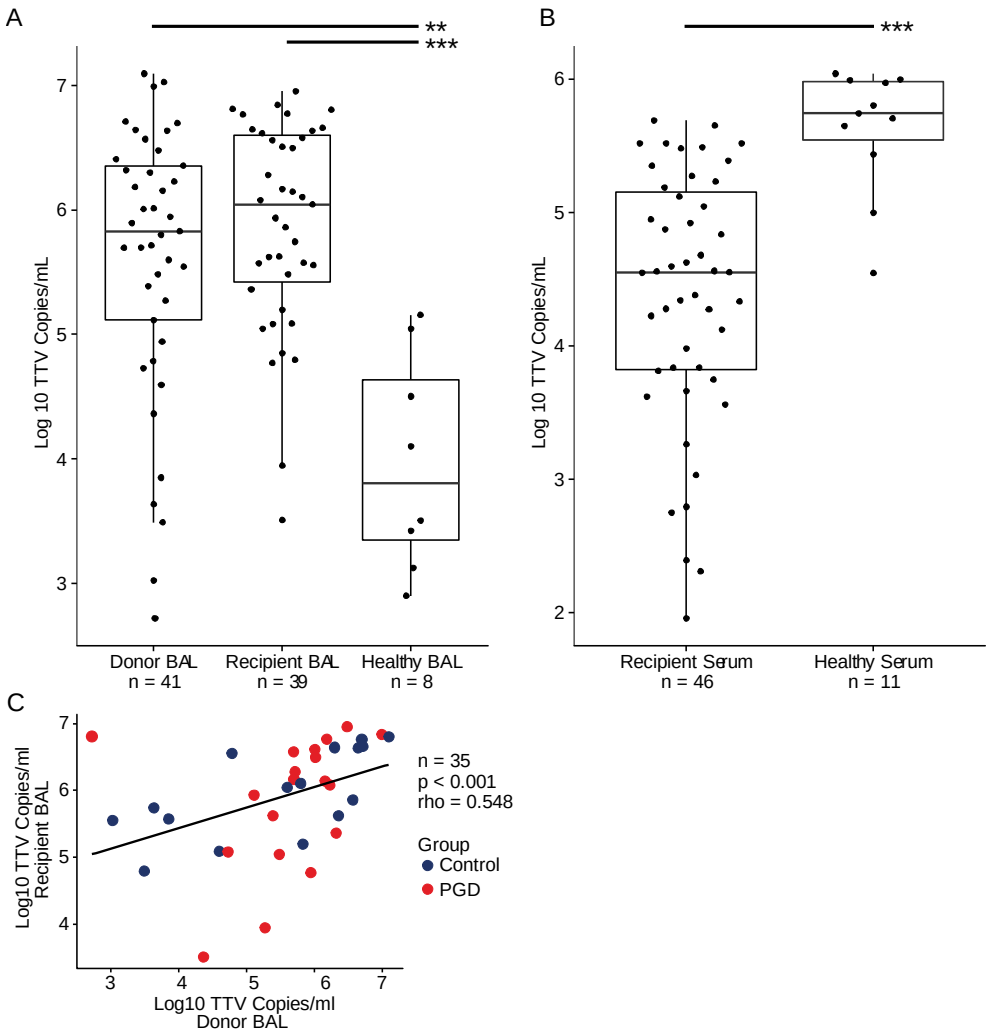


FIGURE 3.1: TORQUE TENO VIRUS (TTV) LEVELS IN LUNG TRANSPLANT DONORS, RECIPIENTS, AND HEALTHY ADULTS

A) TTV quantified by quantitative polymerase chain reaction (qPCR) in bronchoalveolar lavage (BAL). Boxes represent the middle two quartiles for each group, with the bold line representing the median value. Dots represent individual samples. Donor BAL was taken prior to organ procurement. Recipient BAL and recipient serum were taken 1 h after organ reperfusion. Quantities of TTV were higher in donor and recipient BAL compared with BAL of healthy adults, as determined by Wilcoxon rank sum test ($p = 0.0026$ and $p < 0.001$, respectively). (B) Quantities of TTV in lung transplant recipient serum were lower compared with healthy adults, as determined by the Wilcoxon rank sum test ($p < 0.001$). (C) Log10 TTV levels in paired donor and recipient BAL from individual samples were correlated ($p < 0.001$, Spearman's $q = 0.548$). A linear model was fitted to the data and is shown by the black line. The limit of quantification for the qPCR assay ranged from 11 to 65 copies per reaction. **$p < 0.01$, ***$p < 0.001$. PGD, primary graft dysfunction.

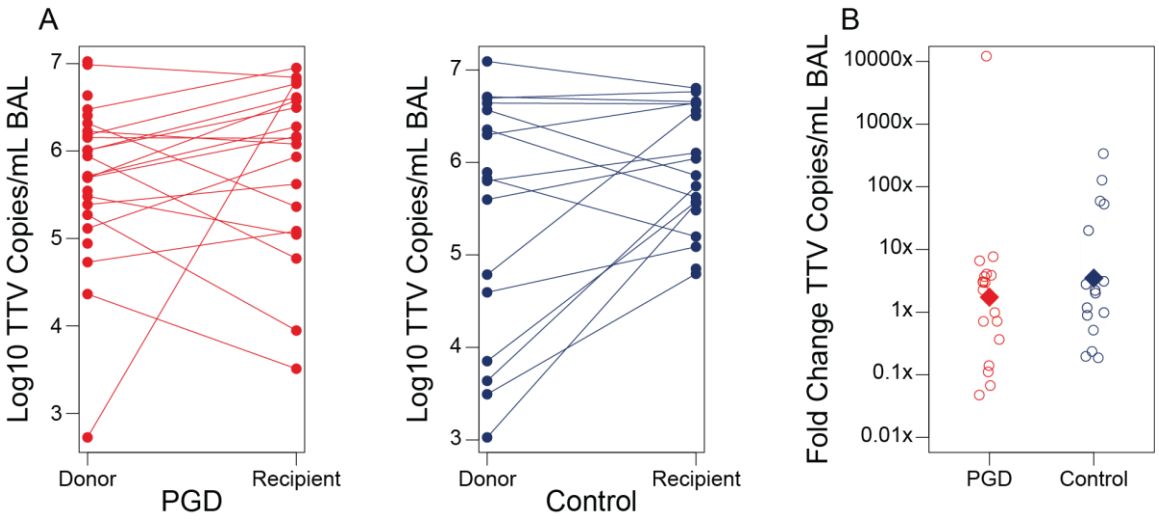


FIGURE 3.2: TORQUE TENO VIRUS (TTV) DYNAMICS IN PERIOPERATIVE PERIOD AND ASSOCIATION WITH PRIMARY GRAFT DYSFUNCTION (PGD)

(A) TTV levels from organ preprocurement (donor) and postreperfusion (recipient) bronchoalveolar lavage samples are shown for each PGD case and control. Samples from the same organ are connected by a line. (B) The fold change in viral levels in the lung for PGD cases and controls is shown. Empty circles represent individual samples, and the filled diamond represents the mean of the group. Both the average and median fold change was lower in PGD cases compared with paired controls ($p = 0.046$ and $p = 0.041$; paired Student t-test and Wilcoxon signed rank test, respectively). The number of paired samples analyzed was 15 because samples were missing for some patients.

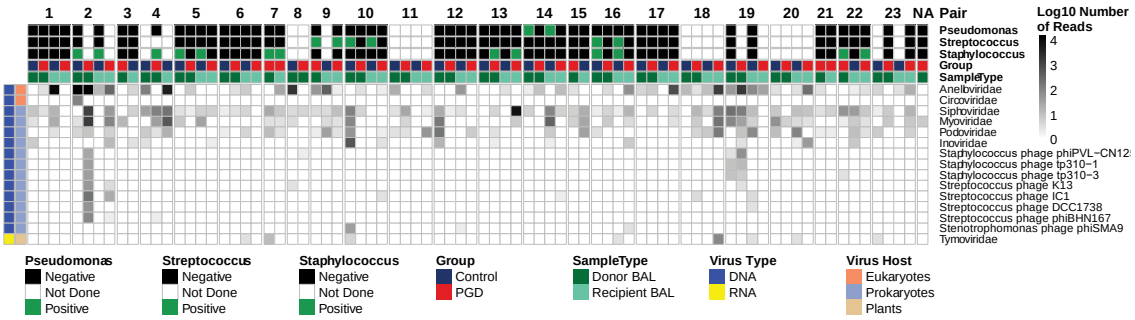


FIGURE 3.3: THE PERIOPERATIVE LUNG VIRAL MICROBIOME

Displayed is the number of shotgun metagenomic reads of DNA and cDNA libraries from each sample matching known viruses. Each column represents a different BAL sample. Each row represents a viral taxonomy at the family level or at the species level for hits that could not be classified into established families (species-level assignments for all viruses are shown in Supplemental Figure 3.2). Sequencing, processing of reads, alignments with viral genomes and removal of spurious hits was carried out as described in Methods. Columns are grouped by PGD–control pairs and labeled according to participant group, pair number and sample type, as shown by the color coding across the top of the figure. Results of standard bacterial culture in each sample for the most commonly identified bacteria are also annotated (top). Further information on viral type and host is given by the column at left. The intensity in each block represents the number of reads of each viral family in each sample on a log10 scale. BAL, bronchoalveolar lavage; PGD, primary graft dysfunction.

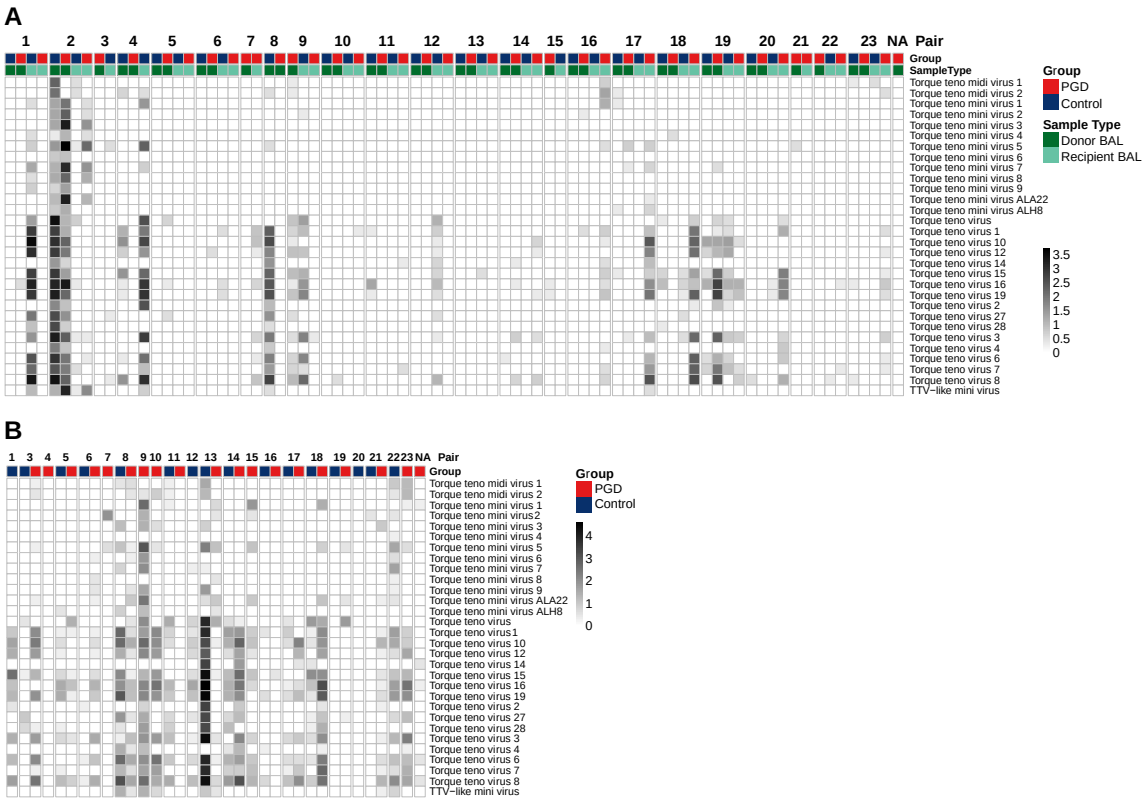


FIGURE 3.4: DIVERSITY OF ANELLOVIRUS ASSIGNMENTS IN LUNG AND SERUM DURING THE PERIOPERATIVE PERIOD

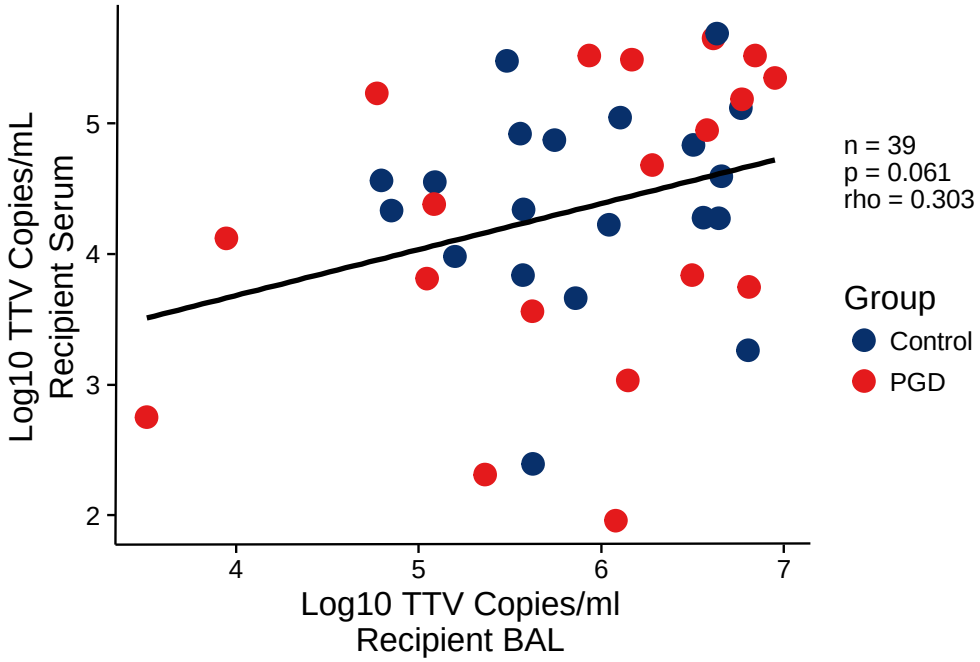
(A) Displayed are filtered metagenomic reads from perioperative BAL samples aligning with annotated human anelloviruses. Each column corresponds to an individual sample, and each row corresponds to the top-scoring reference anellovirus genome in the NCBI Viral Genomes database. The intensity of each block represents the log10 number of reads from that sample aligning with that reference species. Columns are grouped by PGD–control pairs and labeled according to participant group, pair number and sample type. (B) Displayed are filtered metagenomic reads from posttransplant recipient serum that aligned with human anellovirus entries in the NCBI Viral Genomes database. In serum, all viral hits remaining after stringent filtering steps were to anellovirus family members. Samples that had insufficient DNA after library preparation for Illumina sequencing were omitted. BAL, bronchoalveolar lavage; PGD, primary graft dysfunction; TTV, torque teno virus.

Participant identifier	Pair	Group	Samples	Age, years	Sex	Preoperative diagnosis	Donor BAL bacterial and fungal culture
11037	1	Control	DB, RB, RS	50	M	CF/Bronch	Negative
11012	1	PGD	DB, RB, RS	52	F	CF/Bronch	Negative
13017	2	Control	DB, RB, RS	36	F	CF/Bronch	*Staphylococcus*
12009	2	PGD	DB, RB, RS	53	F	CF/Bronch	MRSA
12004	3	Control	RB, RS	29	M	CF/Bronch	Negative
13026	3	PGD	DB, RS	29	M	CF/Bronch	Negative
12016	4	Control	DB, RB, RS	47	F	CF/Bronch	ND
13032	4	PGD	DB, RS	53	M	CF/Bronch	*Staphylococcus*
11005	5	Control	DB, RB, RS	63	M	COPD	*Staphylococcus, Candida albicans*
11004	5	PGD	DB, RB, RS	64	F	COPD	Negative
11016	6	Control	DB, RB, RS	61	F	COPD	Negative
11024	6	PGD	DB, RB, RS	61	M	COPD	Negative
13005	7	PGD	DB, RB, RS	64	F	COPD	*Staphylococcus*
12002	8	Control	DB, RS	62	F	COPD	ND
13024	8	PGD	DB, RS	65	M	COPD	ND
12001	9	Control	RB, RS	63	F	COPD	ND
13035	9	PGD	DB, RB, RS	63	M	COPD	*Streptococcus*
13021	10	Control	DB, RB, RS	39	F	ILD	*Streptococcus, Candida albicans*
11028	10	PGD	DB, RB, RS	49	M	ILD	Negative
11036	11	Control	DB, RB, RS	66	M	ILD	ND
11030	11	PGD	DB, RB, RS	56	M	ILD	ND
11010	12	Control	DB, RB, RS	49	F	ILD	Negative
11007	12	PGD	DB, RB, RS	51	M	ILD	Negative
11002	13	Control	DB, RB, RS	62	F	ILD	Negative
11008	13	PGD	DB, RB, RS	61	M	ILD	*Staphylococcus*
11013	14	Control	DB, RB, RS	64	M	ILD	*Pseudomonas*
11022	14	PGD	DB, RB, RS	64	M	ILD	Negative
13003	15	Control	RB, RS	57	F	ILD	Negative
11035	15	PGD	DB, RS	57	M	ILD	Negative
11032	16	Control	DB, RB, RS	53	M	ILD	*Streptococcus, Staphylococcus*
11042	16	PGD	DB, RB, RS	54	M	ILD	Negative
11006	17	Control	DB, RB, RS	59	M	ILD	Negative
11046	17	PGD	DB, RB, RS	60	M	ILD	Negative
13029	18	Control	DB, RB, RS	68	F	ILD	ND
12007	18	PGD	DB, RB, RS	66	M	ILD	ND
11040	19	Control	DB, RB, RS	63	M	ILD	Negative
12011	19	PGD	DB, RB, RS	61	M	ILD	*Klebsiella*
12015	20	Control	DB, RB, RS	62	M	ILD	ND
12013	20	PGD	DB, RB, RS	61	M	ILD	ND
11050	21	Control	RS	58	M	ILD	Negative
13030	21	PGD	DB, RB, RS	49	M	ILD	Negative
13011	22	Control	RB, RS	38	M	ILD	Negative
13037	22	PGD	DB, RB, RS	41	F	ILD	*Staphylococcus*
13027	23	Control	DB, RB, RS	64	F	ILD	ND
13039	23	PGD	DB, RB, RS	63	M	ILD	Negative
11011	NA	PGD	DB, RS	60	M	PPH	Negative

Table 3.1: Clinical Features of Lung Transplant Recipients

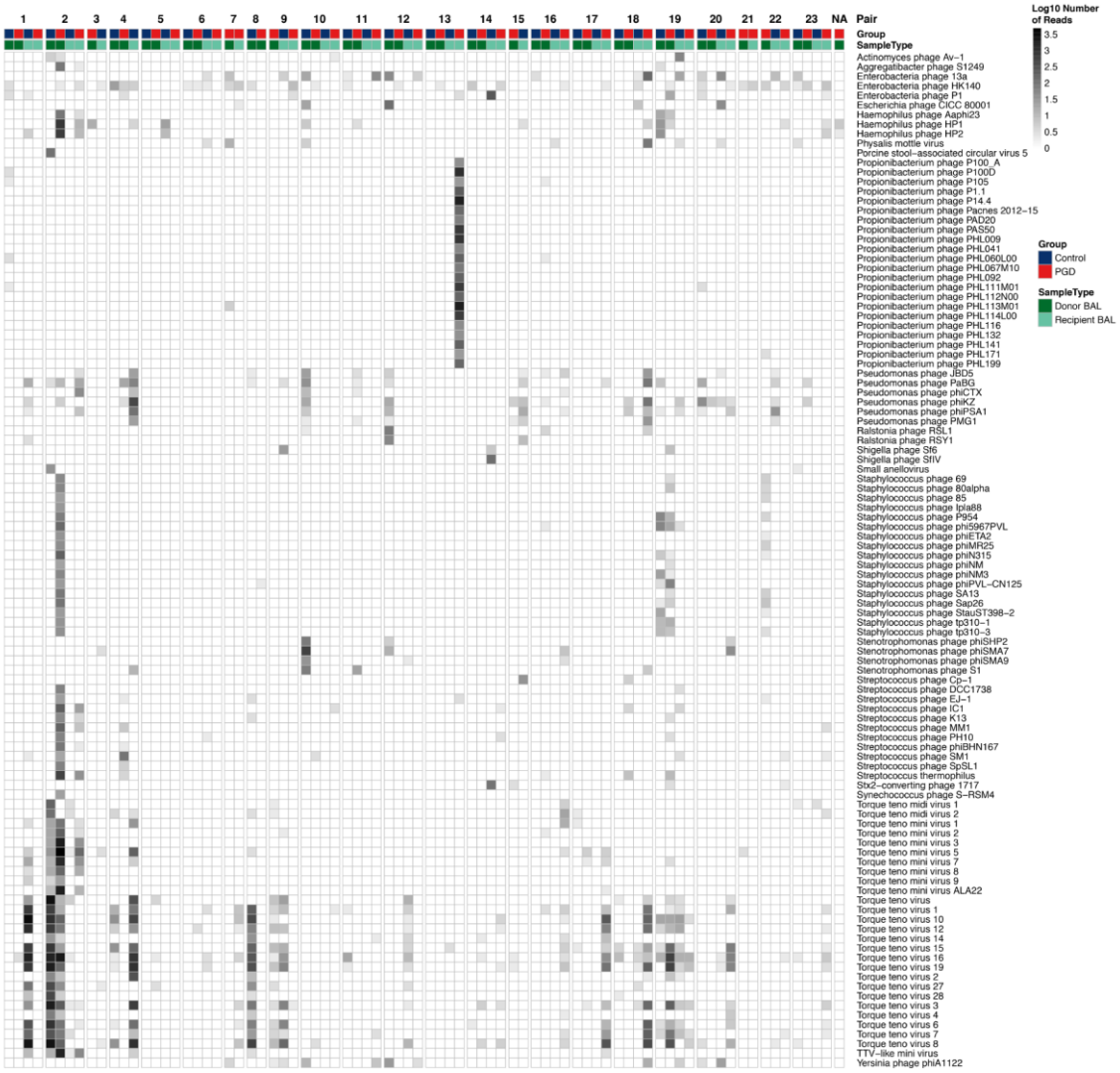
Participants were selected from the CTOT-03 study, and patients with PGD were matched with controls, as described previously (Cantu et al., 2013). PGD grade 3 occurring within the first 72 h following lung transplantation was used as the primary case definition (Yusen et al., 2015). BAL, bronchoalveolar lavage; CF/Bronch, cystic fibrosis or bronchiectasis; COPD, chronic obstructive pulmonary disease (including emphysema or a1-antitrypsin deficiency); DB, donor bronchoalveolar lavage; F, female; ILD, interstitial lung disease (including pulmonary fibrosis); M, male; MRSA, methicillin-resistant *Staphylococcus aureus*; NA, not applicable; ND, not done; PGD, primary graft dysfunction; PPH, primary pulmonary

hypertension; RB, recipient postreperfusion bronchoalveolar lavage; RS, recipient posttransplant serum.



SUPPLEMENTAL FIGURE 3.1: RELATIONSHIP BETWEEN TORQUE TENO VIRUS (TTV) LEVELS IN LUNG AND PERIPHERAL BLOOD WITHIN PARTICIPANTS

The relationship between transplant recipients lung and blood TTV burden (given as log10 copies/mL on each axis) is shown. TTV levels in the lung after organ reperfusion show a positive trend but were not significantly correlated with levels in serum at the same time point ($p = 0.061$, Spearman's $q = 0.303$, Spearman rank correlation). A linear model was fitted to the data and is shown by the black line.



SUPPLEMENTAL FIGURE 3.2: VIRAL SPECIES IN PERIOPERATIVE LUNGS

(A) Viral species with a minimum of 20 of metagenomic reads per sample aligning from perioperative bronchoalveolar lavage (BAL) DNA and cDNA libraries. Each column corresponds to an individual BAL sample, and each row corresponds to a reference viral genome. The intensity of each block represents the number of reads from that sample that match the reference species. Columns are grouped according to participant group, pair number and sample type. Data are the same as those shown in Figure 3.4 except individual species-level assignments are shown for all viral hits.

Pair Number	Control			PGD			Full Set
	Donor BAL	Recipient BAL	Control BAL pair	Donor BAL	Recipient BAL	PGD BAL pair	
1	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	Y	Y	Y
3		Y		Y			
4	Y	Y	Y	Y			
5	Y	Y	Y	Y	Y	Y	Y
6	Y	Y	Y	Y	Y	Y	Y
7	(n/a)	(n/a)		Y	Y		
8	Y			Y			
9		Y	Y	Y	Y	Y	
10	Y	Y	Y	Y	Y	Y	Y
11	Y	Y	Y	Y	Y	Y	Y
12	Y	Y	Y	Y	Y	Y	Y
13	Y	Y	Y	Y	Y	Y	Y
14	Y	Y	Y	Y	Y	Y	Y
15		Y		Y			
16	Y	Y	Y	Y	Y	Y	Y
17	Y	Y	Y	Y	Y	Y	Y
18	Y	Y	Y	Y	Y	Y	Y
19	Y	Y	Y	Y	Y	Y	Y
20	Y	Y	Y	Y	Y	Y	Y
21				Y	Y	Y	
22		Y		Y	Y	Y	
23	Y	Y	Y	Y	Y	Y	Y
11011	(n/a)	(n/a)		Y			
Total	**N = 17**	**N = 20**	**N = 17**	**N = 24**	**N = 19**	**N = 18**	**N = 15**

Supplemental Table 3.1: Bronchoalveolar Lavage Samples Utilized

Some subjects had no material remaining for analysis in this study, and two PGD subjects did not have matched controls. As a result, there were 17 complete Control Donor/Recipient BAL sets, 18 complete PGD Donor/Recipient BAL sets, and a total of 15 pairs had a complete set of all four BAL samples. N/A, not applicable (no matched control subject samples available).

Subject ID	Age	Sex	Smoker	Protocol
3B06	39	M	Y	2-scope
3B07	24	F	N	2-scope
3B08	31	M	N	2-scope
3B09	42	M	Y	2-scope
3B10	65	M	N	2-scope
3B11	50	F	Y	2-scope
3D01	48	F	Y	1-scope
3D02	37	F	N	1-scope

Supplemental Table 3.2: Features of Healthy Adult Lung Samples

Acellular BAL was obtained from healthy adult volunteers as previously described using a 2 scope bronchoscopy procedure (Charlson et al., 2011) or one-scope procedure (Beck et al., 2015). Serum from an independent group of healthy adults volunteers (n = 11) was also collected and analyzed by TTV qPCR.

	Spiked BAL
Paired HiSeq Reads	916,110
Reads Aligning to Human Genome (% of Total)	546,441 (59.6)
Quality and Human Filtered Paired Reads (% of Total)	369,669 (40.4)
NCBI Viral Database Hits (% of Total)	225,274 (24.6)
BROCC Filtered Viral Species Hits	213,946
Reads Aligning (% Coverage)	
Pseudomonas phage phi6 segment L	25,189 (70.5)
Pseudomonas phage phi6 segment M	108,471 (81.8)
Pseudomonas phage phi6 segment S	1,641 (39.8)
Total	135,301

Supplemental Table 3.3: Recovery of Viral RNA by Unbiased Metagenomic Sequencing of BAL

Whole BAL was collected from a non-LTR BAL sample and spiked with 10^6 plaque forming units of Pseudomonas phi6. The sample was centrifuged to remove eukaryotic and bacterial cells and then processed similarly to LTR acellular BAL. DNA/RNA were co-precipitated as detailed in Methods and cDNA synthesis carried out on DNAse-treated aliquot as previously described (Wang et al., 2003). The spiked sample was sequenced alongside 80 perioperative BAL samples from LTRs and a buffer blank.

Pathway ID	Source	Correlation Coefficient (vs log10 ΔTTV Copies/mL BAL)	NOM P value
NOD like receptor signaling pathway	KEGG	0.0217	0.5963
TOLL like receptor signaling pathway	KEGG	0.0119	0.7805
IL1R pathway	BIOCARTA	-0.0127	0.7937
NTHI pathway	BIOCARTA	-0.0223	0.6879
TOLL pathway	BIOCARTA	-0.0119	0.8033

Supplemental Table 3.4: Association of TTV Dynamics with Gene Sets and Pathways Enriched in PGD

Gene Set Variation Analysis (Hänzelmann et al., 2013) was employed to statistically identify enrichment scores of KEGG and Biocarta pathway gene sets in donor and recipient BAL. Specific innate immunity gene sets previously found to be enriched (Cantu et al., 2013) were examined for correlation with viral dynamics. The association between log10 transformed ΔTTV and Δ pathway enrichment scores measured in donor and recipient BAL (n=26) was tested using a general linear model. Definition of abbreviations: NOM P, nominal p-value (estimates the statistical significance of the association with a single gene set or transcript, uncorrected for multiple comparisons); NTHI, NFκB activation by nontypeable *Haemophilus influenza*.

Gene	Donor BAL (n=33)		Recipient BAL (n=38)		Δ Enrichment Score between Donor and Recipient (n=26)	
	Correlation Coefficient (vs Log10TTV Copies/mL BAL)	NOM P value	Correlation Coefficient (vs Log10TTV Copies/mL BAL)	NOM P value	Correlation Coefficient (vs log10 ΔTTV Copies/mL BAL)	NOM P value
IL1B	-0.0067	0.982	-0.0926	0.499	-0.098	0.801
NLRP3	-0.138	0.698	0.0355	0.838	-0.212	0.646
IL1A	-0.265	0.440	-0.137	0.463	-0.541	0.273
IL6	-0.272	0.422	-0.109	0.588	-0.404	0.429
CCL4	0.125	0.668	0.0171	0.909	0.321	0.377
TLR6	0.0998	0.617	0.0671	0.588	0.0464	0.889
TLR1	-0.0043	0.982	0.0015	0.992	-0.184	0.570
TNFAIP3	0.105	0.712	-0.0091	0.941	0.113	0.733
TLR4	0.0486	0.786	0.0648	0.471	-0.183	0.462

Supplemental Table 3.5: Association of TTV with Highest Ranked Transcripts Enriched in PGD in Perioperative BAL

The top 9 ranked individual transcripts, primarily from NRL and TLR signaling pathways, previously identified (Cantu et al., 2013) to be enriched in PGD were examined for any correlations with TTV during the perioperative period. Associations between log10 transformed TTV levels and dynamics measured in donor and recipient BAL and individual transcript expression levels and changes in enrichment score were tested using a general linear model.

Donor TTV levels and Donor Transcriptional Profile		Donor BAL (n=33)		
Pathway ID	**Source**	**Correlation Coefficient (vs log10 TTV/mL BAL)**	**NOM P value**	**FDR P value**
TOB1 pathway	BIOCARTA	-0.122	0.0038	0.995
NUCLEARRS pathway	BIOCARTA	-0.150	0.0257	0.995
Recipient TTV levels and Recipient Transcriptional Profile		**Recipient BAL (n=38)**		
Pathway ID	**Source**	**Correlation Coefficient (vs log10 TTV/mL BAL)**	**NOM P value**	**FDR P value**
Feeder pathway	BIOCARTA	0.148	0.0111	0.894
Alpha linolenic acid metabolism	KEGG	0.0892	0.0273	0.894
ACTINY pathway	BIOCARTA	0.0833	0.0278	0.894
PTDINS pathway	BIOCARTA	0.0851	0.0328	0.894
Vibrio cholerae infection	KEGG	0.0659	0.0331	0.894
Glycolysis pathway	BIOCARTA	0.122	0.0362	0.894
Type 1 diabetes mellitus	KEGG	-0.0769	0.0372	0.894
P53 pathway	BIOCARTA	0.0696	0.0443	0.894
Valine leucine and isoleucine biosynthesis	KEGG	0.120	0.0461	0.894
MPR pathway	BIOCARTA	0.0794	0.0487	0.894
ARF pathway	BIOCARTA	0.0763	0.0488	0.894
TTV Dynamics and Change in Transcriptional Profile		**Δ Enrichment Score between Donor and Recipient (n=26)**		
Pathway ID	**Source**	**Correlation Coefficient (vs ΔTTV Copies/mL BAL)**	**NOM P value**	**FDR P value**
D4GDI pathway	BIOCARTA	0.0983	0.0116	0.959
Feeder pathway	BIOCARTA	-0.124	0.0264	0.959
IL5 pathway	BIOCARTA	0.0894	0.0334	0.959
Alpha linolenic acid metabolism	KEGG	-0.0775	0.0341	0.959
CTLA4 pathway	BIOCARTA	0.106	0.0367	0.959
Asthma	KEGG	0.0922	0.0414	0.959

Supplemental Table 3.6: Association of TTV with Gene Set Variation Analysis (GSVA) of BAL mRNA

GSVA (Hänzelmann et al., 2013) was employed to statistically identify enrichment scores of KEGG and BIOCARTA pathway gene sets (n = 403) in donor and recipient BAL. Association analyses using general linear models were performed between GSVA pathway enrichment scores and TTV levels and dynamics during perioperative time points. The table shows the pathways that had a nominal P value <0.05 for each analysis, along with the P value after correction for multiple testing (Benjamini and Hochberg, 1995). FDR, False Discovery Rate (Benjamini–Hochberg procedure correction based on estimated probability that an association represents a false positive finding).

			Donor BAL (n=41)		Recipient BAL (n=39)		ΔTTV Copies/mL BAL (n=35)		Recipient Serum (n=46)	
Continuous Variables	**Median**	**Range**	**Correlation Coefficient**	**P**	**Correlation Coefficient**	**P**	**Correlation Coefficient**	**P**	**Correlation Coefficient**	**P**
Donor age	37	14-62	-0.328	0.0364	-0.00405	0.9811	0.196	0.259		
Total ischemic time (min)	488	18-821			0.298	0.0655	0.050	0.776	0.153	0.3111
Nitric oxide (min)	360	30-3700			-0.189	0.5588	0.018	0.973		
Platelets (mL)	268	90-450			1	0.3333	0.5	1	1	0.3333
Fresh frozen plasma (mL)	500	225-1206			1	0.3333	0.5	1	0.6	0.4166
Colloid fluids (mL)	1000	50-1750			-0.7	0.2333	-0.9	0.083	-0.1	0.95
Red blood cells (mL)	500	250-3150			-0.5	1	-0.5	1	0	1
Cardiopulmonary bypass time (min)	268	56-408			0.205	0.3855	0.181	0.485	0.106	0.6211
Categorical Variables	**Values**	**Total N**	**Median Log10 TTV Copies/mL**	**P**	**Median Log10 TTV Copies/mL**	**P**	**Median Δ Log10 TTV Copies/mL**	**P**	**Median Log10 TTV Copies/mL**	**P**
Donor cause of death	Anoxia	5	6.47	0.243	6.17	0.576	-0.099	0.52		
	Cerebro-vascular Stroke	19	5.48		5.93		0.329			
	Head trauma	16	6.01		5.94		0.474			
	Other	6	5.71		5.86		-0.132			
Donor history of aspiration	Yes	4	6.12	0.949	5.85	0.8411	0.949	1		
	No	42	5.80		6.04		0.345			
Donor history of purulent secretions	Yes	4	6.13	0.343	5.73	0.827	0.326	0.188		
	No	39	5.70		6.06		0.350			
	Not available	3	6.01		5.62		0.486			
Donor smoking status	Ever smoked	13	5.83	0.7	6.17	0.63	0.231	0.872		

				88		3				
	Never smoked	33	5.83		5.80		0.345			
Nitric oxide	Yes	16			6.08	0.417	-0.147	0.334		
	No	30			5.99		0.400			
Blood product administration	Yes	6			6.45	0.289	0.520	0.822	4.68	0.973
	No	40			5.93		0.303		4.54	
Platelets	Yes	4			6.28	0.497	0.474	0.846	4.68	0.966
	No	42			5.99		0.324		4.54	
Fresh frozen plasma	Yes	5			6.61	0.346	0.474	0.934	4.07	0.895
	No	41			5.99		0.324		4.55	
Colloid fluids	Yes	5			5.56	0.669	0.474	0.766	4.12	0.607
	No	41			6.06		0.324		4.55	
Red blood cells	Yes	5			6.61	0.346	0.474	0.934	4.07	0.895
	No	41			5.99		0.324		4.55	
Induction immunosuppression	Yes	44			6.04	0.875	0.355	0.165	4.56	0.004
	No	2			5.85		-0.439		2.18	
Cardiopulmonary bypass	Yes	25			6.08	0.606	0.113	0.303	4.55	0.153
	No	21			5.99		0.355		4.56	
Transplant type	Bilateral	30			6.22	0.182	0.47	0.396	4.56	0.936
	Single	16			5.86		0.19		4.36	
Recipient preoperative diagnosis	CF	8			5.49	0.469			4.42	
	COPD	9			5.93				4.62	
	ILD	28			6.09				4.55	
	PPH	1			NA				2.79	

Supplemental Table 3.7: Correlations between Clinical Variables and Torque Teno Virus

Correlations between TTV levels and continuous variables were tested using Spearman Rank Correlation. Associations with categorical variables containing two groups were tested with unpaired Wilcoxon Rank Sum test. Associations with categorical variables with greater than two groups were tested with Kruskal-Wallis test. Empty cells indicate analyses not tested because an association would not be logical for a given sample type or time point (eg: recipient preoperative diagnosis and Donor BAL). CF, cystic fibrosis/bronchiectasis; COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease; PPH, primary pulmonary hypertension.

Top Scoring Hit	Virus Type	Number of Samples Identified In	Total Reads	Examples of Known or Potential Hosts
Adenoviridae	dsDNA	1	2	Mammal
Alloherpesviridae	dsDNA	6	14	Mammal
Baculoviridae	dsDNA	2	4	Invertebrate
Bunyaviridae	ssRNA (–)	0	0	Mammal
Caliciviridae	ssRNA (+)	2	4	Mammal
Cystoviridae	dsRNA	4	8	Bacteria
Flaviviridae	ssRNA (+)	1	2	Mammal (Invertebrate vector)
Geminiviridae	ssDNA	1	14	Plant
Herpesviridae	dsDNA	18	55	Mammal
Microviridae	ssDNA (+)	1	1	Bacteria
Mimiviridae	dsDNA	4	26	Amoeba
Nudiviridae	dsDNA	1	2	Invertebrate
Papillomaviridae	dsDNA	2	6	Mammal, Amniote
Parvoviridae	ssDNA	7	41	Vertebrate, Invertebrate
Phycodnaviridae	dsDNA	3	10	Algae
Polyomaviridae	dsDNA	2	4	Mammal, Bird
Poxviridae	dsDNA	1	2	Vertebrate, Invertebrate
Retroviridae	ssRNA	12	58	Vertebrate
Sphaerolipoviridae	dsDNA	2	8	Bacteria, Archaea
Virgaviridae	ssRNA (+)	1	4	Plant
Unclassified *Salmonella* phage	dsDNA	2	2	Bacteria
Unclassified *Staphylococcus* phage	dsDNA	4	16	Bacteria
Unclassified *Streptococcus pyogenes* phage	dsDNA	4	29	Bacteria

Supplemental Table 3.8: Top Scoring Viral Families with Fewer than 20 Reads
After removal of shotgun metagenomic reads attributable to environmental contamination or unfiltered human sequences, viral species with greater than 20 reads by BROCC scoring (Dollive et al., 2012) in an individual sample were called authentic detections, whereas those with <20 reads in an individual sample are of uncertain significance. Presented here are viral families whose members did not meet this stringent threshold (individual species level assignment is shown in Supplemental Table 3.9).

Top Scoring Hit	Number of Samples Identified In	Total Reads
Abelson murine leukemia virus	4	21
Acanthamoeba polyphaga mimivirus	1	18
Achromobacter phage JWAlpha	2	16
Acinetobacter phage Bphi-B1251	1	1
Acinetobacter phage ZZ1	4	14
Adelie penguin polyomavirus	1	2
Bacillus phage BCP78	2	6
Bacillus phage G	2	6
Bacillus phage Moonbeam	1	18
Bacillus phage phIS3501	8	47
Bathycoccus sp. RCC1105 virus BpV	3	10
Betapapillomavirus 2	1	2
Betapapillomavirus 3	1	4
Blattodean ambidensovirus 2	2	14
Bovine herpesvirus 4	3	8
Burkholderia phage Bcep176	1	3
Burkholderia phage Bcep781	1	1
Burkholderia phage BcepMigl	1	4
Burkholderia phage BcepMu	1	1
Burkholderia phage KL3	3	11
Burkholderia phage KS14	1	1
Burkholderia phage KS5	3	9
Burkholderia phage phiE202	1	2
Campylobacter phage CP21	2	4
Cassava associated cicular DNA virus	1	2
Caulobacter phage Cr30	1	6
Caulobacter phage rogue	1	1
Caulobacter phage swift	5	11
Caviid herpesvirus 2	1	2
Chlamydia phage 1	1	1
Cronobacter phage vB_CsaM_GAP31	1	5
Cucurbit leaf crumple virus	1	14
Cyanophage S-TIM5	1	2
Cyprinid herpesvirus 1	3	6
Cyprinid herpesvirus 2	2	4
Enterobacteria phage cdtI	4	4
Enterobacteria phage fiAA91-ss	2	5
Enterobacteria phage HK022	1	7
Enterobacteria phage HK446	6	8
Enterobacteria phage HK542	1	2
Enterobacteria phage HK544	1	1
Enterobacteria phage HK629	2	8
Enterobacteria phage HK633	5	8
Enterobacteria phage HK97	3	4
Enterobacteria phage If1	2	8

Enterobacteria phage IME10	6	9
Enterobacteria phage lambda	6	9
Enterobacteria phage M13	7	18
Enterobacteria phage mEp043 c-1	5	6
Enterobacteria phage mEp235	1	2
Enterobacteria phage mEp237	3	5
Enterobacteria phage mEp460	2	3
Enterobacteria phage mEpX2	8	10
Enterobacteria phage Min27	8	9
Enterobacteria phage P2	1	1
Enterobacteria phage PsP3	1	16
Enterobacteria phage Sf101	1	1
Enterobacteria phage T3	1	2
Enterobacteria phage T4 sensu lato	1	5
Enterobacteria phage T7	1	4
Enterobacteria phage YYZ-2008	1	1
Equid herpesvirus 2	2	6
Erwinia phage phiEaH2	9	32
Erwinia phage vB_EamP-L1	1	2
Escherichia phage 121Q	1	2
Escherichia phage HK639	3	8
Escherichia phage PBECO 4	1	2
Escherichia phage phiV10	1	3
Escherichia phage TL-2011b	4	6
Escherichia phage wV7	1	2
Falconid herpesvirus 1	8	19
Fujinami sarcoma virus	1	2
Gallid herpesvirus 2	1	2
Gallid herpesvirus 3	3	6
Geobacillus virus E2	1	2
Human herpesvirus 5 (CMV)	1	12
Human immunodeficiency virus 1	1	1
Human mastadenovirus C	1	2
Ictalurid herpesvirus 1	2	4
Lactobacillus phage A2	1	4
Lactobacillus phage Lc-Nu	1	2
Lactobacillus phage Lrm1	1	2
Lactobacillus prophage Lj928	2	4
Lactococcus phage 1706	1	2
Mannheimia phage vB_MhM_1152AP	1	1
Mason-Pfizer monkey virus	2	6
Megavirus chiliensis	1	2
Merkel cell polyomavirus	1	2
Microbacterium phage Min1	1	4
Mongoose feces-associated gemycircularvirus c	1	2
Moumouvirus	2	6
Murine osteosarcoma virus	1	4

Mycobacterium phage ArcherS7	1	2
Mycobacterium phage PegLeg	1	9
Oryctes rhinoceros nudivirus	1	2
Ovine lentivirus	1	2
Pandoravirus inopinatum	2	4
Pandoravirus salinus	1	2
Parvovirus NIH-CQV	4	25
Pestivirus Giraffe-1	1	2
Phage Gifsy-1	1	1
Phage Gifsy-2	1	1
Primate erythroparvovirus 1	1	2
Propionibacterium phage P100_1	2	8
Propionibacterium phage PHL030	1	2
Propionibacterium phage PHL095	1	1
Propionibacterium phage PHL179	1	1
Pseudomonas phage DMS3	1	1
Pseudomonas phage F10	1	1
Pseudomonas phage JBD24	6	21
Pseudomonas phage JBD30	6	27
Pseudomonas phage KPP25	2	4
Pseudomonas phage LUZ7	1	2
Pseudomonas phage MP22	8	39
Pseudomonas phage MP29	1	2
Pseudomonas phage OBP	1	2
Pseudomonas phage PAJU2	1	2
Pseudomonas phage Pf1	1	9
Pseudomonas phage phi297	3	3
Pseudomonas phage phi6	4	8
Pseudomonas phage phiPsa374	1	2
Pseudomonas phage PPpW-3	3	10
Pseudomonas phage vB_PaeS_PAO1_Ab18	1	2
Ralstonia phage 1 NP-2014	1	1
Ralstonia phage RSA1	2	6
RD114 retrovirus	1	2
Rhodococcus phage ReqiPepy6	2	4
Salmonella phage epsilon34	2	3
Salmonella phage Fels-1	1	1
Salmonella phage Fels-2	2	4
Salmonella phage FSL SP-004	1	1
Salmonella phage g341c	1	1
Salmonella phage RE-2010	2	4
Salmonella phage SPN3UB	1	2
Salmonella phage SSU5	1	4
Salmonella phage ST64B	1	1
Salmonella phage ST64T	1	1
Salmonella phage STP4-a	1	2
Salmonella phage vB_SemP_Emek	2	2
Salmonella phage vB_SosS_Oslo	1	1

Salmonella phage Vi06	2	8
Sapporo virus	2	4
Sewage-associated circular DNA virus-28	1	2
Sewage-associated gemycircularvirus-1	1	2
Shewanella sp. phage 1/40	2	4
Spodoptera litura nucleopolyhedrovirus II	2	4
Staphylococcus phage 11	1	4
Staphylococcus phage 187	2	6
Staphylococcus phage 42E	3	19
Staphylococcus phage 52A	1	6
Staphylococcus phage 55	2	6
Staphylococcus phage 77	2	17
Staphylococcus phage 88	1	4
Staphylococcus phage CNPH82	1	2
Staphylococcus phage Ipla35	1	1
Staphylococcus phage Ipla7	1	1
Staphylococcus phage JS01	1	15
Staphylococcus phage PH15	1	1
Staphylococcus phage phi2958PVL	2	8
Staphylococcus phage phi7401PVL	2	8
Staphylococcus phage phiETA	2	16
Staphylococcus phage phiETA3	2	20
Staphylococcus phage phiMR11	3	18
Staphylococcus phage phiSa119	2	8
Staphylococcus phage PT1028	1	7
Staphylococcus phage PVL	2	5
Staphylococcus phage Pvl108	4	35
Staphylococcus phage SA11	1	1
Staphylococcus phage SA12	2	3
Staphylococcus phage SMSAP5	1	17
Staphylococcus phage StauST398-3	2	17
Staphylococcus phage StauST398-4	1	2
Staphylococcus phage StauST398-5	4	28
Staphylococcus phage StB20	1	2
Staphylococcus phage TEM123	3	20
Staphylococcus phage vB_SauM_Remus	1	1
Staphylococcus phage X2	3	9
Staphylococcus prophage phiPV83	3	5
Streptococcus phage 2972	1	1
Streptococcus phage 5093	1	3
Streptococcus phage 7201	1	1
Streptococcus phage 858	1	2
Streptococcus phage Alq132	2	4
Streptococcus phage Dp-1	3	3
Streptococcus phage PH15	1	16
Streptococcus phage phi3396	1	1
Streptococcus phage SMP	1	2

Streptococcus phage YMC-2011	5	8
Streptococcus pyogenes phage 315.2	1	14
Streptococcus pyogenes phage 315.4	2	7
Streptococcus pyogenes phage 315.5	1	8
Synechococcus phage ACG-2014i	1	1
Synechococcus phage KBS-M-1A	1	6
Synechococcus phage S-IOM18	1	2
Synechococcus phage S-SSM7	2	3
Thermus thermophilus bacteriophage P23-77	1	2
Thermus thermophilus phage IN93	1	6
Torque teno felis virus	1	1
Torque teno mini virus 4	5	18
Torque teno mini virus 6	2	12
Torque teno mini virus ALH8	4	23
Turnip vein-clearing virus	1	4
uncultured phage crAssphage	1	9
Vibrio phage JA-1	1	1
Vibrio phage pVp-1	3	6
Vibrio phage VH7D	1	2
Woolly monkey sarcoma virus	2	7
Y73 sarcoma virus	3	13
Yersinia phage L-413C	1	2
Yoka poxvirus	1	2

Supplemental Table 3.9: Top Scoring Viral Species with Fewer than 20 Reads
After removal of shotgun metagenomic reads attributable to environmental contamination or unfiltered human sequences, viral species with greater than 20 reads by BROCC scoring (Dollive et al., 2012) in an individual sample were called as authentic detections, whereas those with <20 reads in an individual sample are of uncertain significance. Presented here are viral species that did not meet this stringent threshold (the same data collapsed at the family level is shown in Supplemental Table 3.8).

CHAPTER 4: BIDIRECTIONAL TRANSFER OF ANELLOVIRIDAE LINEAGES BETWEEN GRAFT AND HOST DURING LUNG TRANSPLANTATION

The contents of this chapter have been previously published as:

Abbas, A.A., Young, J.C., Clarke, E.L., Diamond, J.M., Imai, I., Haas, A.R., Cantu, E., Lederer, D.J., Meyer, K., Milewski, R.K., Olthoff, K.M., Shaked, A., Christie, J.D., Bushman, F.D., Collman, R.G., 2018. Bidirectional transfer of Anelloviridae lineages between graft and host during lung transplantation. Am. J. Transplant.

4.1 ABSTRACT

Solid organ transplantation disrupts virus-host relationships, potentially resulting in viral transfer from donor to recipient, reactivation of latent viruses, and new viral infections. Viral transfer, colonization, and reactivation are typically monitored using assays for specific viruses, leaving the behavior of full viral populations (the "virome") understudied. Here we sought to investigate the temporal behavior of viruses from donor lungs and transplant recipients comprehensively. We interrogated the bronchoalveolar lavage and blood viromes during the peritransplant period and 6-16 months posttransplant in 13 donor-recipient pairs using shotgun metagenomic sequencing. *Anelloviridae*, ubiquitous human commensal viruses, were the most abundant human viruses identified. Herpesviruses, parvoviruses, polyomaviruses, and bacteriophages were also detected. *Anelloviridae* populations were complex, with some donor organs and hosts harboring multiple contemporaneous lineages. We identified transfer of *Anelloviridae* lineages from donor organ to recipient serum in 4 of 7 cases that could be queried, and immigration of lineages from recipient serum into the allograft in 6 of 10 such cases. Thus, metagenomic analyses revealed that viral populations move between graft and host in both directions, showing that organ transplantation involves implantation of both the allograft and commensal viral communities.

4.2 INTRODUCTION

Solid organ transplantation exposes recipients to viruses present in donor tissues, and also requires immunosuppression, which facilitates virus reactivation and *de novo* infection. Organ donors are routinely screened for viruses of known clinical concern (Grossi et al., 2009), and viral infections are common in lung transplant recipients (LTRs) (Burguete et al., 2013). Viruses may also contribute indirectly to acute cellular rejection (ACR) and chronic lung allograft dysfunction (Peghin et al., 2017, Vu et al., 2011).

Advances in metagenomic DNA sequencing methods now enable queries of whole viral populations (the "virome"), including viruses not monitored by current clinical assays. The virome of various human body sites, including the respiratory tract (Lewandowska et al., 2017, Wylie et al., 2012, Taboada et al., 2014, Willner et al., 2009), is only beginning to be characterized. Recent studies have shown that Anelloviridae are ubiquitous in the human eukaryotic virome (Spandole et al., 2015). *Anelloviridae* is a family of highly diverse, non-enveloped, small circular single-stranded DNA (ssDNA) viruses that infect humans and other mammals (Spandole et al., 2015). Human *Anelloviridae* are not associated with any diseases, although other small circular ssDNA viruses are important veterinary pathogens (Meng, 2012, Ellis, 2014, Todd, 2004). *Anelloviridae* are likely under chronic immune control, given that levels in blood increase in immunosuppressed states (Thom and Petrik, 2007, De Vlaminck et al., 2013, Görzer et al., 2014). Lower *Anelloviridae* burden in blood has been associated with solid organ rejection, suggesting that *Anelloviridae* DNA copy numbers may be a useful "functional" indicator of immune status (De Vlaminck et al., 2013, Focosi et al., 2016, Masouridi-Levrat et al., 2016). Whether lung allograft rejection might be associated with low *Anelloviridae* levels within the lung itself, perhaps as an indicator of local immune function, has not been investigated.

We previously investigated the virome in bronchoalveolar lavage (BAL) from healthy subjects and a cross-sectional sample of LTRs (Abbas et al., 2017) and found that lung allografts had markedly elevated *Anelloviridae* DNA levels compared to healthy adults. We also observed unexpectedly high levels of *Anelloviridae* DNA in the lungs of donors prior to organ recovery, and found an association between perioperative *Anelloviridae* dynamics and primary graft dysfunction. In these studies, it was unclear whether *Anelloviridae* genomes in recipients' allografts were derived from the donated organ or by entry of circulating viruses present in the recipient before transplantation. It was also unknown

whether *Anelloviridae* from the donor organ could disseminate and establish infection outside the lung. We hypothesized that the abundance of *Anelloviridae* in the donor lungs would result in virus transfer to immunosuppressed LTRs and that viruses already present in the recipient might also transfer into the allograft. To test this, we investigated the dynamics of whole viral populations from donor lungs before procurement, and in lung allograft and blood during the first 6-16 months posttransplant.

4.3 RESULTS

Study subjects

A total of 114 samples (13 donor BAL, 49 recipient BAL, 52 recipient serum) from 13 organ donors and their respective LTRs were available for metagenomic sequence analysis (Table 4.1). These included perioperative samples (donor BAL, and recipient BAL and serum taken immediately following organ implantation and reperfusion) and samples taken during routine posttransplant surveillance (approximately 1, 6, 12, 24, and 48 weeks posttransplant) and for clinical indications. The sampling schedule and relevant clinical events are summarized in Figure 4.1 and Table 4.1.

Metagenomic shotgun sequencing of donor and recipient samples

To analyze the virome, virus-like particles were isolated from acellular BAL and serum. Total nucleic acid was extracted, whole-genome amplified, and then analyzed by metagenomic sequencing. Initial analysis for RNA viruses on a subset of 51 samples yielded no authentic RNA virus detection as determined by the limited extent of genome coverage when sequence reads were mapped onto reference genomes. Therefore, subsequent analysis focused on DNA viruses.

Human sequences and technical artifacts were removed as described previously (Clarke et al., 2017a). Further filtering removed artefactual calls due to misannotation of human reads (CHAPTER 2), database errors, barcode hopping during Illumina (San Diego, CA)

sequencing (Lauder et al., 2016), and contamination from environmental sources (Salter et al., 2014, Naccache et al., 2013).

A median of 272,230 filtered read pairs in the donor BAL sample (range 14,390-4,667,852), 1,913,041 median read pairs in recipient BAL (range 189-36,920,976) and 351,211 in serum DNA samples (range 4,023 - 2,246,428) (Supplemental Table 4.1 and Supplemental Figure 4.1). Read counts in negative controls were lowest in sterile water extractions (maximum of 1310) and highest in SM buffer (range 4311-2,660,621) (Supplemental Table 4.1 and Supplemental Figure 4.1). Although efforts were taken to enrich for viral nucleic acids, bacterial, fungal and human sequences persisted after sample processing and library preparation. On average, human sequences comprised 43% of all high quality reads in BAL samples (range 0.2-99%) and 54% (range 0.07-99%) in serum samples. Human sequences averaged 12% in extraction controls (range 0.02-45%). To account for contaminating sequences introduced during acquisition, processing and library preparation of low biomass samples, bronchoscope prewashes, buffer and sterile water blanks were analyzed using the same workflow.

We acquired $>5.25\times10^{10}$ base pairs (bp) of DNA sequence from 114 clinical samples, six environmental controls (Supplemental Table 4.1), and 24 bronchoscope prewashes (Clarke et al., 2017a). Samples differed considerably in the number of reads remaining after filtration of human and artefactual sequences (Supplemental Table 4.1), likely reflecting both the abundance of the filtered sequences, and differences in the authentic content of viruses.

We assessed the representation of known viruses by analyzing sequence reads using a k-mer-based classification scheme (Wood and Salzberg, 2014), revealing a range of abundances (0%-52% in BAL and 0.06%-33% in serum). We identified viruses from four families known to infect humans: *Anelloviridae*, *Herpesviridae*, *Polyomaviridae*, and

Parvoviridae (Figure 4.2 and Table 4.2). The most frequently identified human-cell virus was *Anelloviridae*, consistent with previous reports (De Vlaminck et al., 2013, Young et al., 2015, Abbas et al., 2017). On average, 8.7% of all classifiable reads in BAL and 12.7% in serum were derived from *Anelloviridae* species.

To validate assignments from the k-mer analysis, reads from virus-positive samples were aligned to reference viral genomes and inspected for depth and evenness of coverage (Table 4.2, Supplemental Figure 4.1). In donor BAL 13-28, Epstein-Barr virus (EBV) was detected (1,800 reads covering 45% of the genome). The organ donor had positive serology for EBV (Supplemental Table 4.2); however, EBV was not detected in subsequent recipient BAL or serum samples. A complete genome of KI polyomavirus was also detected in BAL of this subject at 83 days posttransplant (Supplemental Figure 4.1A). Merkel cell polyomavirus was detected once each in BAL of subject 12-09 and serum of subject 13-17 (Supplemental Figure 4.1A).

Cytomegalovirus (CMV) was also detected in the lung of subject 13-17 at two consecutive time points and in serum, even though they were receiving CMV prophylaxis. A near-complete human bocavirus genome was detected in BAL 91 days posttransplant in subject 13-19. Two subsequent samples from this subject (one BAL and one serum) were positive for parvovirus B19 (Supplemental Figure 4.1B).

We also identified reads that matched to regions of several nonhuman eukaryotic viruses (Supplemental Table 4.3). These included two species within the *Circoviridae* family and micromonas pusilla virus, which infects green algae. The significance of sequences aligning to these viruses is uncertain.

In addition to eukaryotic viruses, we identified sequences from four bacteriophage families in BAL (*Inoviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*; Figure 4.2A). These bacteriophages have diverse bacterial hosts, including members of common oropharyngeal

flora and respiratory pathogens (Edlund et al., 2015). In contrast, in serum there were only a few hits to bacteriophages, which mainly overlapped with species or families, found in environmental controls (Figure 4.2B and Supplemental Figure 4.2). Thus, we lack convincing evidence for authentic detection of bacteriophage in blood.

Most viral hits in background samples matched bacteriophages (Supplemental Figure 4.2), predominantly bacteriophages of *Propionibacterium*, a skin inhabitant and common reagent contaminant (Salter et al., 2014, Clarke et al., 2017a). Several environmental controls yielded *Anelloviridae* reads, but at far lower levels than in clinical samples, likely reflecting "barcode hopping" during sequence acquisition (Kim et al., 2017). Additional hits in controls were to nonhuman eukaryotic-cell viruses reported previously in metagenomic analyses of bronchoscope prewash controls such as *Genomoviridae* (Clarke et al., 2017a), which we did not study further.

Sequences from each sample were then assembled into contigs that ranged in size from 300-49,800 bp, with a median of 344-1639 bp among the samples. In total, 75,935 contigs larger than 1,000 bp were built across all samples. Most of these (54,878) mapped to bacterial genomes, consistent with bacteriophages annotated as bacteria (since bacteriophage can integrate into bacterial genomes), or bacterial DNA that was incompletely removed during VLP preparations. In total, 1,919 contigs were annotated as viral based on alignment to the NCBI viral database (E-value<1x10-10).

Quantitative analysis of *Anelloviridae* dynamics in BAL and serum

Anelloviridae were the most prevalent viruses, and were present in donor lungs prior to transplantation. Therefore, we focused on investigating their temporal dynamics and relationship to clinical outcomes.

We first quantified absolute genome copy numbers using qPCR (Figure 4.3). Consistent with our previous work (Abbas et al., 2017), *Anelloviridae* genome copies in BAL

from donor lungs prior to transplantation were higher than those in healthy adults ($p = 0.006$; Wilcoxon Rank Sum Test, Supplemental Table 4.4). Levels in BAL generally remained elevated compared to healthy adults both early and late posttransplant ($p < .05$ for visits 1, 3, 4, and 6; Wilcoxon Rank Sum Test, Supplemental Table 4.4), extending prior observations from our cross-sectional study of LTRs (Young et al., 2015). In contrast to BAL levels, *Anelloviridae* genome copy numbers in serum of LTRs immediately after organ reperfusion were lower than levels in healthy adults ($p = 0.01$; Wilcoxon Rank Sum Test, Supplemental Table 4.4), consistent with a previous report (Focosi et al., 2015). Serum *Anelloviridae* levels then increased above the levels of healthy subjects ($p < 0.05$ for all time points >60 days posttransplant; Wilcoxon Rank Sum Test), as previously shown with iatrogenic immunosuppression (Görzer et al., 2014, Görzer et al., 2015, Moen et al., 2003).

Blood *Anelloviridae* levels have been shown to correlate with ACR in pediatric lung transplantation (Blatter et al., 2018), but no reports have addressed levels in lung. Therefore, we investigated whether there was a relationship between ACR and *Anelloviridae* levels in BAL, as well as serum. Consistent with prior reports (Blatter et al., 2018) , we found lower *Anelloviridae* levels in serum during the month preceding episodes of ACR compared to all non-ACR sera (average $10^{5.5}$ copies/mL vs average of $10^{6.9}$ copies/mL; $p = 0.039$; Student 2-tailed t-test). In addition, we also found lower *Anelloviridae* levels in BAL during the 30 days preceding a diagnosis of ACR (average $10^{4.8}$ copies/mL vs average of $10^{5.8}$ copies/mL; $p = 0.014$; Student 2-tailed t-test).

In contrast, neither levels of *Anelloviridae* in the donor organ prior to procurement nor diversity of viral lineages were associated with ACR (data not shown). We previously reported that *Anelloviridae* dynamics during the peritransplant period were also associated with primary graft dysfunction (Abbas et al., 2017).

Assessing transfer of Anelloviridae populations between allograft and recipient

Given the presence of *Anelloviridae* in both donor lungs and recipients at the time of transplantation, we investigated transfer of *Anelloviridae* lineages between the graft and transplant recipient. To do this, we assembled sequence reads from the donor lung and initial recipient serum samples into contigs and asked whether these contigs were represented by reads appearing in later samples. Contigs >2,000 bp were chosen to allow sufficient length to query the presence of each *Anelloviridae* lineage against the background of other variants. Such contigs could be assembled in seven donor BAL samples, and in ten postreperfusion LTR serum samples (used to represent viral lineages present in recipients at the time of transplantation).

We first analyzed the *Anelloviridae* Open Reading Frame 1 (ORF1) amino acid sequences to understand the baseline similarity of *Anelloviridae* swarms within and between subjects (Supplemental Figure 4.3), and determine whether there would be adequate diversity to enable lineage-specific tracking. Across all 140 perioperative contigs, ORF1 sequences exhibited 36% amino acid identity between subjects, and 37% identity within subjects. The low within- and between-subject identity emphasizes the extreme diversity of *Anelloviridae*, and allows tracking of initial lung and blood lineages at later time points.

We aligned reads from subsequent BAL and blood to these initially-present donor lung and blood contigs to track their appearance posttransplant. Representation at later time points was calculated using the Gini index (Lows, 1984). The Gini index scores evenness of contig coverage and is sensitive to regions of genomic divergence while accommodating different sequence depths in each sample and for each *Anelloviridae* genome. The Gini index has been used previously for analyses of metagenomic data (Bhattacharya et al., 2015, Kobayashi and Andoh, 2018).

To validate this approach, we first investigated whether sharing of *Anelloviridae* lineages based on Gini index values was more likely in cognate donor-recipient pairs than in unrelated pairs. Indeed, *Anelloviridae* lineages present in initial donor lung or recipient serum were more likely to be found in later BAL or blood specimens from cognate recipients than in samples from unrelated subjects ($p < 0.001$; Wilcoxon Rank Sum Test on Gini values, Figures 4.4 and 4.5).

We also compared samples using several additional approaches. First, we compared lineage representation in each sample by calculating reads per kilobase of alignment target per million reads sequenced, using reads remaining after removing human reads and low-quality reads. This normalizes for variable recovery of reads after filtration (Supplemental Table 4.5). Second, we calculated the total fraction of the contig covered by reads in subsequent samples (Supplemental Table 4.5). We also observed that full and even coverage of a contig by sequences from an unrelated individual was rare (Supplemental Tables 4.7 and 4.8). Finally, we assessed temporal dynamics of specific *Anelloviridae* contigs within a donor-recipient pair by visually inspecting alignments (Supplemental Figure 4.4). These approaches yielded conclusions similar to the analysis using the Gini index.

Detection of donor lung *Anelloviridae* lineages in recipient allograft and serum

To assess transfer of *Anelloviridae* lineages from donor lungs to recipients, we focused on the seven LTRs for whom donor BAL samples yielded contigs >2,000 bp (range 1-28 contigs per donor). Figure 4.4A illustrates the persistence of *Anelloviridae* lineages from donor lung in longitudinal BAL of recipients, and Figure 4.4B illustrates the appearance of donor lung lineages in serum of recipients. As a group, donor lung contigs were more similar to sequences in subsequent cognate BAL and serum samples than samples from other transplant recipients ($p < 0.001$; Wilcoxon Rank Sum Test on Gini index). Tracking individual donor viruses revealed that they were more evenly covered in

BAL specimens from their cognate recipients than in random comparisons in 4 of 7 pairs (Figure 4.4A). The same analysis comparing donor lung contigs and longitudinal recipient serum reads revealed the appearance of donor lung *Anelloviridae* lineages in recipient blood in 4 of 7 pairs (Figure 4.4B). Thus, posttransplant *Anelloviridae* persistence within the lung and dissemination to serum was detectable in the majority of cases where this could be evaluated.

Detection of recipient *Anelloviridae* lineages in lung allograft and recipient serum over time

We next asked whether *Anelloviridae* lineages present in recipients' blood at the time of transplant appeared in the allograft after transplantation, and whether these lineages persisted in serum. Postreperfusion serum contigs were used as a representation of recipient-derived lineages. Contigs >2000 bp were available for 10 LTRs (range 1-20 contigs per subject). Recipient serum *Anelloviridae* lineages were more closely related to both cognate serum and BAL after transplantation, compared with unrelated subjects ($p < 0.001$; Wilcoxon Rank Sum test on Gini index). Recipient *Anelloviridae* lineages present at time of transplantation were then tracked individually, and could be detected in serum at later time points in 8 of 10 pairs (Figure 4.5A), indicating persistence in blood. Entry of recipient blood *Anelloviridae* lineages into the allograft was investigated by comparing contigs from recipient postreperfusion serum samples to longitudinal recipient BAL samples. As expected, in most cases there was no significant coverage of recipient serum lineages in donor BAL pretransplant (before exposure to recipient lineages) (Supplemental Table 4.7). After transplantation, recipient *Anelloviridae* lineages were detectable in subsequent BAL samples in 6 of 10 subjects (Figure 4.5B). Thus, *Anelloviridae* lineages present systemically in recipients at the time of transplantation commonly entered and populated the graft.

4.4 DISCUSSION

In this study, we investigated dynamics of the lung and blood virome in lung transplant donors and recipients. Both human viruses and bacteriophages were identified. One viral family, *Anelloviridae*, was ubiquitous and abundant. *Anelloviridae* have not been reported to cause disease in humans, but they appear to be under immune control and monitoring their abundance and diversity provides a window on viral interactions with the host immune system.

Anelloviridae blood levels increase in states of immune deficiency (such as AIDS and iatrogenic immunosuppression) and decrease with immune reconstitution (Thom and Petrik, 2007, Görzer et al., 2014, Görzer et al., 2015, Madsen et al., 2002). Transplant recipients with episodes of organ rejection have been shown to have lower *Anelloviridae* levels than those without rejection, consistent with inadequate immune suppression (De Vlaminck et al., 2013, Blatter et al., 2018). Thus, while the specific immune mechanisms responsible for regulating *Anelloviridae in vivo* are unknown, it has been proposed that monitoring levels in blood might serve as a "functional" measure of immune activity to manage organ transplant immunosuppression (De Vlaminck et al., 2013, Focosi et al., 2016). Our data are consistent with previous findings in blood. In addition, a novel observation here is that *Anelloviridae* within the lung allograft are also relatively decreased prior to ACR. Since controlling immune activity within the allograft is the key goal of organ transplantation management, further studies appear warranted to determine whether lung *Anelloviridae* levels might reflect local compartmentalized immune function, and/or could offer actionable information useful in immunosuppression management.

The extensive intrasubject *Anelloviridae* sequence diversity observed here is consistent with previous reports of diverse lineages coexisting within individuals (Young et al., 2015, Okamoto et al., 2001, Devalle et al., 2009, Bzhalava et al., 2012, Jelcic et al., 2004,

Ninomiya et al., 2008). The origin of multiple independent lineages within individuals is unclear and could arise from intrahost evolution, initial infection by multiple lineages, or episodes of superinfection (Spandole et al., 2015). Additionally, certain individuals had, by chance, some *Anelloviridae* sequences that were highly similar to those in unrelated individuals. Tracking populations of related yet distinct viral genomes within subjects is challenging, so we combined breadth and depth of genome coverage using the Gini index and compared representation of specific *Anelloviridae* lineages within and between individuals. We show that some *Anelloviridae* lineages contained within the donor organ emigrate from the allograft and circulate systemically in recipients. Conversely, some *Anelloviridae* species from the recipient entered and populated the allograft. Donor-recipient transmission of clinically important viruses during organ transplantation is well described (Green et al., 2015), but engraftment of an entire viral population along with the organ, with bidirectional flow from the allograft to the recipient and vice-versa, is novel.

We found that some *Anelloviridae* lineages present either in the donor organ or the recipients' systemic circulation persisted for months, while others disappeared, possibly being replaced by new variants. The consequences of donor- or recipient-derived *Anelloviridae* within the graft are unknown. It is plausible that introduction of novel *Anelloviridae* lineages may have subtle deleterious consequences such as triggering localized inflammation within the allograft. Conversely, *Anelloviridae* have been implicated in suppressing NFκB-mediated cellular activation (Zheng et al., 2007). Unfortunately, we did not have sufficient power to ask what outcomes might be linked to donor- versus recipient-derived lineages within the graft, which should be a topic of future studies.

Only one donor BAL revealed a known human virus other than *Anelloviridae*, which was EBV. In recipients, herpesviruses (CMV, EBV), parvoviruses (B19, human bocavirus), and polyomaviruses (KI, Merkel cell) were identified in lung and blood. Detection of Merkel

cell polyomavirus and KI polyomavirus in whole BAL or tissue of lung transplantation patients has been described (Bergallo et al., 2010), although DNA in acellular BAL is a novel finding that suggests production of extracellular viral particles, consistent with active viral replication. Subject 13-17 yielded CMV in one serum and two BAL specimens despite being on CMV suppressive therapy, and another subject's BAL revealed human bocavirus. Although infection with traditional community acquired respiratory viruses is common after lung transplantation, none of our subjects had clinically recognized viral infection at the time of sample collection. Thus, our limited detection of known pathogenic viruses in acellular BAL and serum suggest that asymptomatic or subclinical infection in LTRs is uncommon, or that these metagenomic methods have limited sensitivity for their detection.

Our study has several limitations. The use of acellular BAL is appropriate for analysis of the extracellular (replicating) virome, but cannot detect intracellular viral nucleic acids of nonreplicating or latent forms. Due to geographic heterogeneity within the lung (Willner et al., 2012a), any single sample may report only part of the virome. Because replicate samples were not available, the effect of sampling stochasticity could not be quantified. Long-term storage of BAL and our methods of sample processing may have reduced our sensitivity to detect low abundance enveloped DNA and RNA viruses. Our small sample size limited our ability to compare the magnitude of association of *Anelloviridae* and ACR in lung versus blood, and the number of subjects in whom we could track individual lineages precluded an analysis of donor versus recipient strain-specific relationship to outcomes. Finally, our findings report on viruses present in reference databases. Indeed, between 12% and 94% of metagenomic reads in each sample could not be classified into any known kingdom, similar to other metagenomic studies (Quince et al., 2017, Krishnamurthy and Wang, 2017). Illuminating this viral dark matter may reveal uncharacterized viruses that may play a role in lung health and disease.

In summary, our study shows that lung transplantation is associated with engraftment not just of a donor organ, but of its endogenous population of *Anelloviridae* as well. Future studies will be needed to determine factors that regulate viral transfer between graft and host, and further define the relationship between compartmentalized lung *Anelloviridae* and transplantation outcomes.

4.5 ACKNOWLEDGEMENTS

We are grateful to participants and volunteers for providing specimens, to the CTOT-03 investigators (Jason D. Christie, Abraham Shaked, Kim M. Olthoff, Rita K. Milewski, Keith C. Meyer, David J. Lederer, Edward Cantu, Andrew R. Haas) and the staff participating in the CTOT-03 and Lung HIV Microbiome Project studies and to members of the Bushman and Collman (Ize Imai) laboratories for help and suggestions. Erik L. Clarke developed bioinformatics software used in this study and Jacque C. Young assisted in sample processing. Frederic D. Bushman and Ronald G. Collman assisted in the design of experiments, interpretation of data, and writing of the manuscript. This work was supported by NIH grants R01-HL113252, R61-HL137063, U01-HL098957, R01-HL087115, K24-HL115354, and received assistance from the Penn Center for AIDS Research (P30-AI045008) and the PennCHOP Microbiome Program. A.A.A. was supported by NSF grant DGE-1321851 and J.M.D. was supported by K23-HL121406.

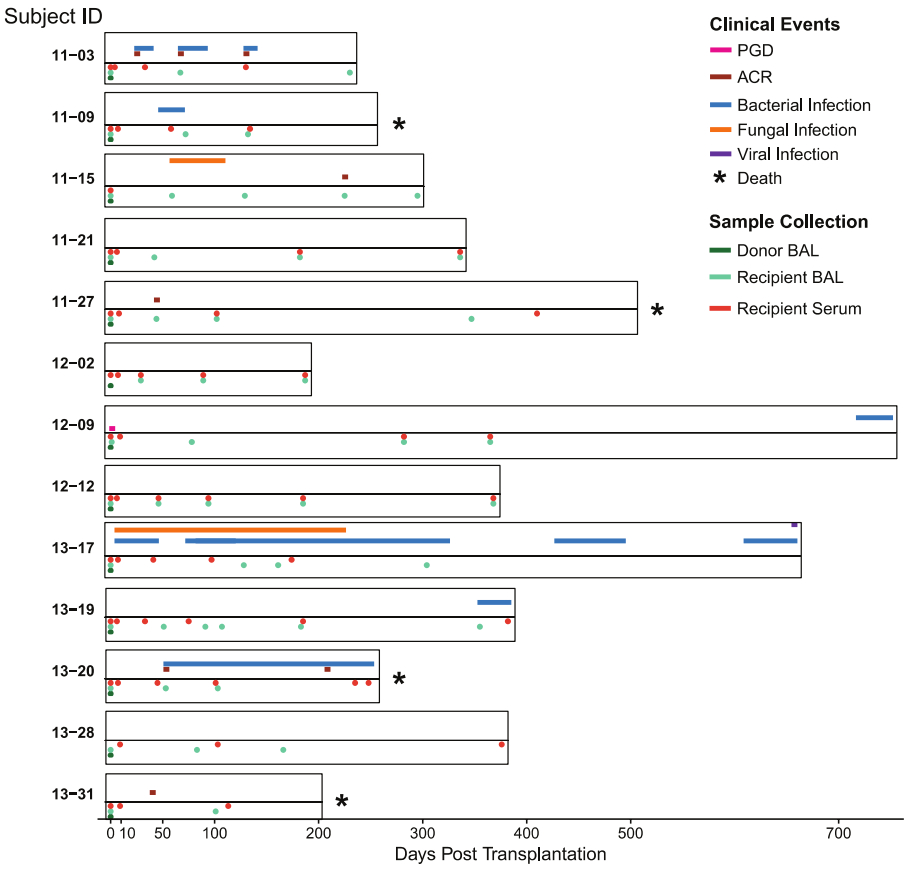


FIGURE 4.1: SAMPLE COLLECTION AND CLINICAL EVENTS

The duration of each subject's enrollment in the study is shown. Sample collection (bronchoalveolar lavage [BAL] and blood specimens) time points are displayed below the timeline and indicated by the color code. The timing and duration of adverse clinical events for each subject are displayed above the timeline and similarly annotated. For subject 12-09, primary graft dysfunction (PGD) grade 3 occurred within the first 72 hours following lung transplantation (Yusen et al., 2015). Five subjects who experienced acute cellular rejection (ACR) had a maximum grade of A2 (Stewart et al., 2007).

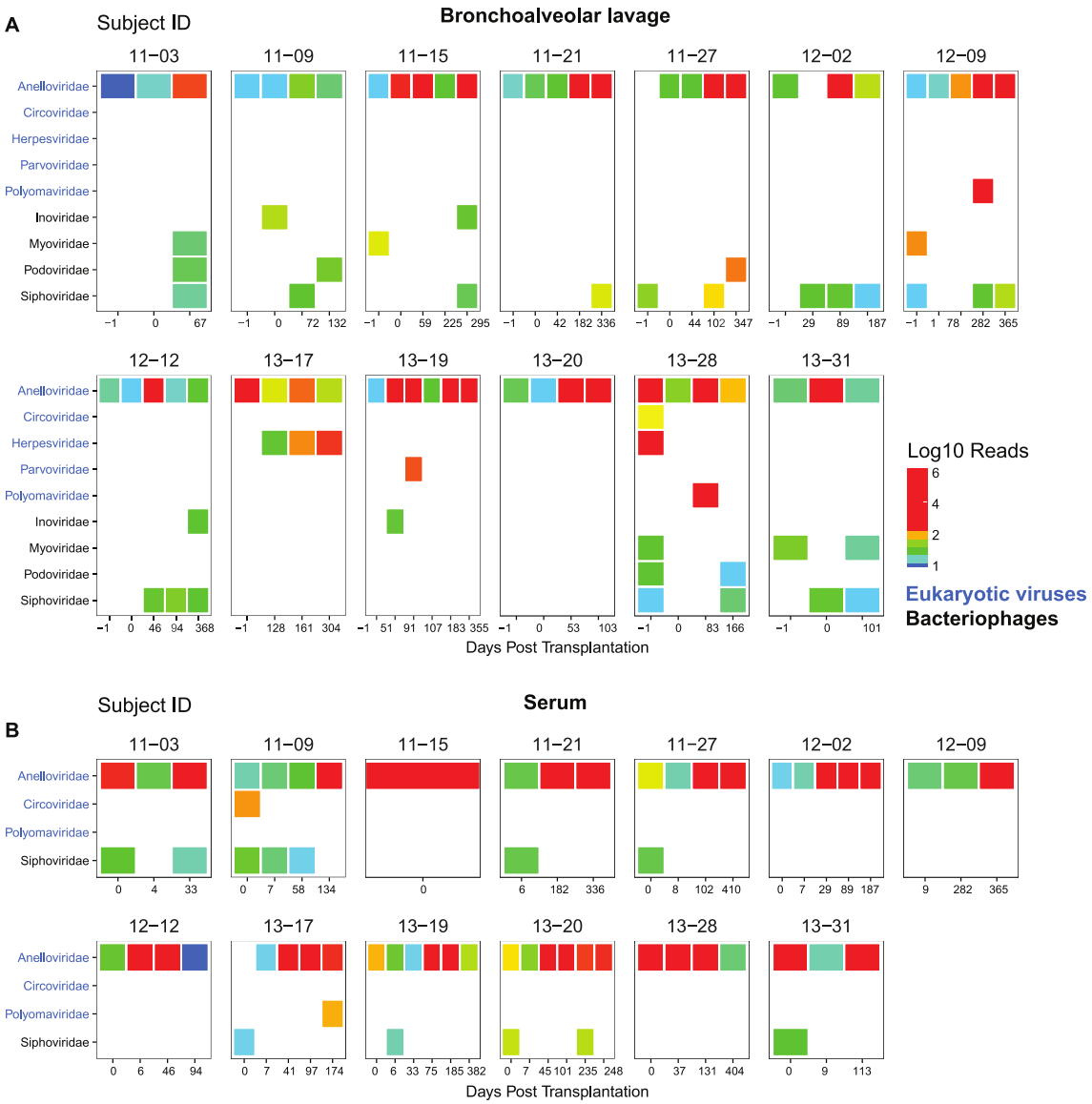


FIGURE 4.2: THE VIRAL MICROBIOME BEFORE AND AFTER TRANSPLANTATION
Displayed are the distribution and number of read pair assignments (on a log10 scale) from shotgun metagenomic sequencing of bronchoalveolar lavage (BAL) (A) and serum (B). Shown are results from DNA sequences that match known viruses, filtered to remove spurious hits. Each box represents a different donor-recipient pair and each column a different BAL (A) or serum (B) sample. BAL of the donor lung prior to procurement and transplantation is the first column and indicated as "-1" days posttransplant. Posttransplant BAL samples were taken during routine surveillance bronchoscopy or for other indications. Reads that could not be classified at the species level are not included for display. Only hits with a minimum of 10 reads assigned to a viral family per sample are included. Hits believed to be spurious or derived from environmental contamination are not displayed. Viral families are grouped according to target host. Sequencing, preprocessing of reads, classification, and quality control were carried out as described in CHAPTER 2.

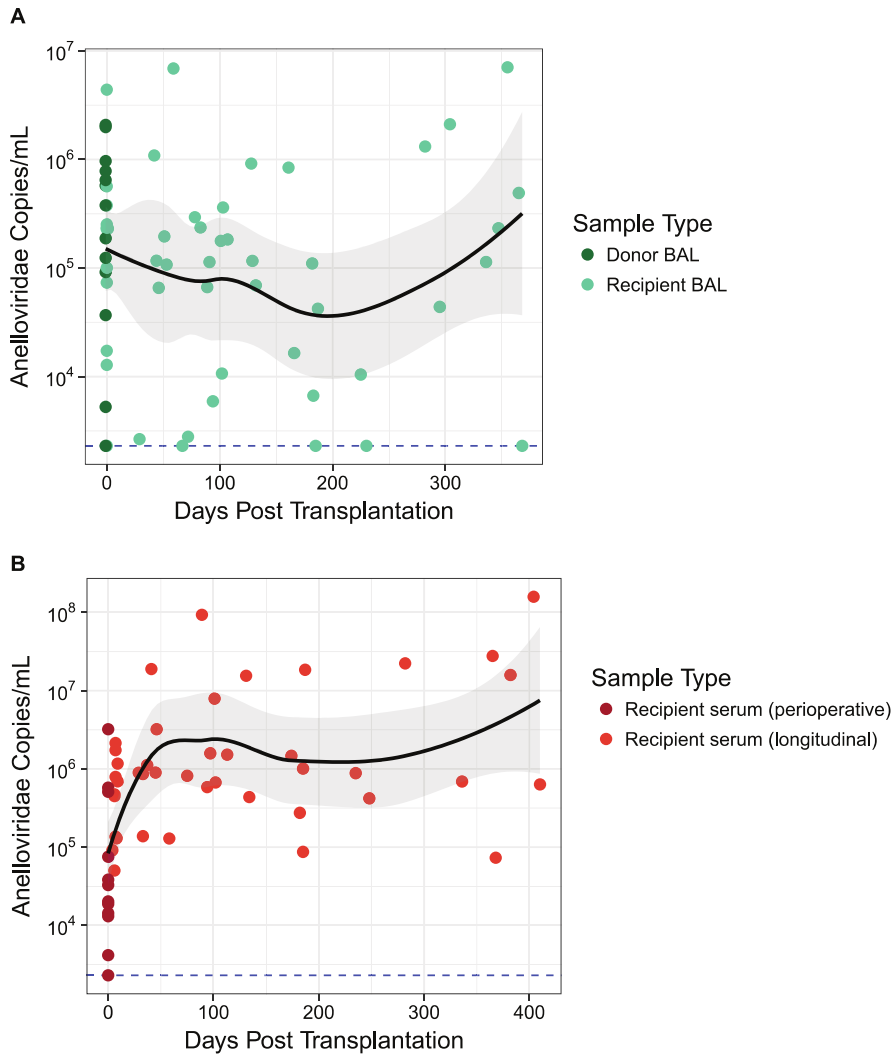


FIGURE 4.3 ANELLOVIRIDAE DYNAMICS BY BODY SITE

Anelloviridae species were quantified in bronchoalveolar lavage (BAL) and serum by qPCR that targets torque teno viruses, torque teno midi viruses, and torque teno mini viruses (*Anelloviridae* family members). The black line represents the local regression curve and standard error for all longitudinal samples. The dotted line represents the limit of detection for the qPCR assay (38 target copies/reaction). Samples at or below this limit were assigned this minimal value. Dots represent individual samples. (A) Longitudinal analysis of *Anelloviridae* genome copies in BAL. Donor BAL was taken prior to organ procurement. Recipient BAL was first obtained an hour after organ reperfusion and at various time points posttransplant. (B) Longitudinal analysis of *Anelloviridae* genome copies in serum from lung transplant recipients. Perioperative serum from the transplant recipients was obtained an hour after organ reperfusion.

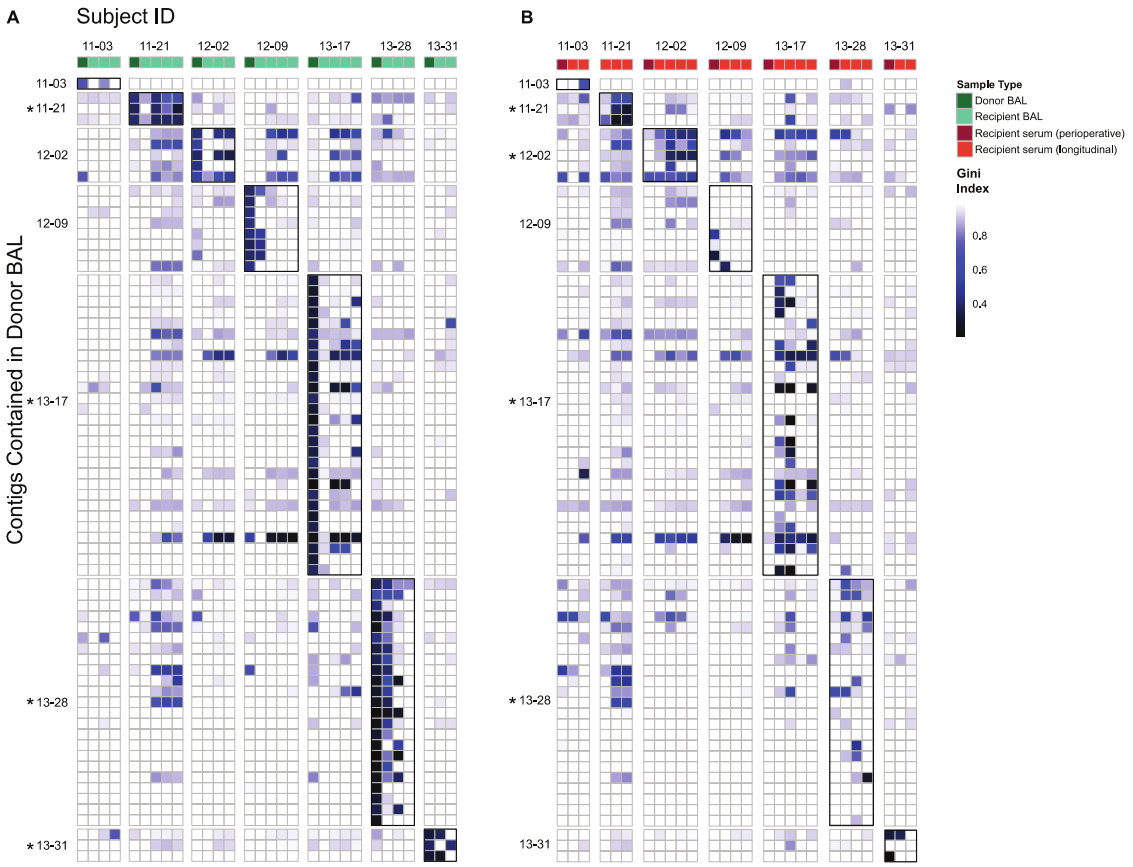


FIGURE 4.4 LONGITUDINAL MONITORING OF DONOR LUNG ANELLOVIRIDAE IN TRANSPLANT RECIPIENTS' SUBSEQUENT LUNG AND SERUM SAMPLES

Contigs >2,000 bp that could be assembled from 7 organ donor bronchoalveolar lavage (BAL) were aligned with reads found in posttransplant BAL (A) and serum (B) samples. Each row is a contig representing an *Anelloviridae* partial genome in donor BAL. Rows are grouped according to the donor organ in which the contig was found, shown on the left. Columns represent individual samples of BAL (A) or serum (B) arranged chronologically, grouped by recipient ID and annotated based on the color key. Recipient perioperative serum in B was sampled 1 hour after organ reperfusion. The color in each block represents the Gini index of each comparison of initial sample contig to subsequent samples' reads. A value of 1 is highly uneven coverage suggesting lack of detection, while 0 is even coverage across the genome suggesting highly confident detection. Alignments of samples to contigs in cognate donor-recipient pairs (donor lung and recipient of that specific donor organ) are outlined in black for ease of visualization. As a group, *Anelloviridae* lineages present in initial donor lungs were significantly more likely to be found in BAL and blood specimens from cognate recipients than in samples from unrelated subjects ($p < 0.001$; Wilcoxon Rank Sum Test). Individual donor-recipient pairs where there was significant detection of donor lineages in either BAL (A) or serum (B) are indicated by the asterisk ($p < 0.05$; Wilcoxon Rank Sum Test).

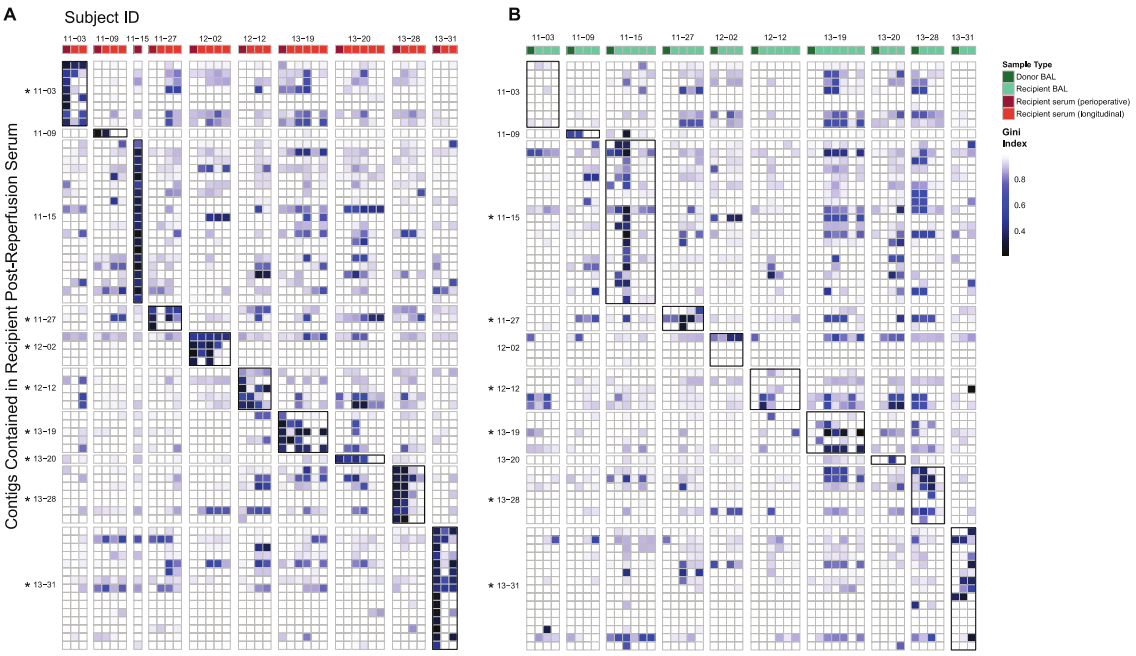


FIGURE 4.5: LONGITUDINAL MONITORING OF INITIAL RECIPIENT SERUM ANELLOVIRIDAE IN TRANSPLANT RECIPIENTS' SUBSEQUENT SERUM AND LUNG SAMPLES

Contigs >2,000 bp that could be assembled from 10 perioperative serum samples (obtained within an hour after transplantation and taken to represent recipient *Anelloviridae* at the time of transplantation) were aligned with reads found in all posttransplant serum (A) and bronchoalveolar lavage (BAL) (B) samples. Each row is a contig representing an Anelloviridae partial genome present in perioperative serum, grouped and annotated by subject. Columns represent individual serum (A) or BAL (B) samples arranged chronologically and grouped by subject (subject 11-15 only had serum from the perioperative time point available). In (B), BAL of the donor organ is the first column. The color in each block represents the Gini index of each genome in each sample. A value of 1 is highly uneven coverage suggesting lack of detection while 0 is even coverage across the genome suggesting highly confident detection. Alignments of samples to contigs in cognate donor-recipient pairs are outlined in black for ease of visualization. As a group, *Anelloviridae* lineages present in initial recipient serum were significantly more likely to be found in BAL and blood specimens at later time points from that subject than in samples from unrelated subjects ($P < .001$; Wilcoxon Rank Sum Test). Individual recipients where there was significant detection of these perioperative lineages in either BAL (A) or serum (B) are indicated by the asterisk ($P < .05$; Wilcoxon Rank Sum Test).

Subject ID	Age (Years)	Sex	Preoperative Diagnosis	Donor Age (Years)	Transplant Type	Infectious Clinical Events (Organisms)	Non-Infectious Events
11-03	28	M	Cystic Fibrosis	36	BLT	*Pseudomonas, Haemophilus influenzae,* MRSA	ACR
11-09	64	M	Pulmonary Fibrosis	40	SLT	*Staphylococcus epidermidis*	Death
11-15	66	F	Pulmonary Fibrosis	28	SLT	*Aspergillus terreus*	ACR
11-21	44	F	Cystic Fibrosis	50	BLT	None	None
11-27	22	M	Cystic Fibrosis	32	BLT	None	ACR, death
12-02	62	F	Emphysema	43	SLT	None	None
12-09	53	F	Bronchiectasis	20	BLT	*Streptococcus*	PGD
12-12	36	M	Cystic Fibrosis	52	BLT	None	None
13-17	36	F	Cystic Fibrosis	25	BLT	*Pseudomonas, Stenotrophomonas maltophila, Haemophilus influenzae, Candida glabrata,* Parainfluenza	None
13-19	26	F	Cystic Fibrosis	42	BLT	*Pseudomonas*, MRSA	None
13-20	21	M	Cystic Fibrosis	13	BLT	*Escherichia coli*, *Pseudomonas*	ACR, death
13-28	54	M	Pulmonary Fibrosis	19	SLT	None	None
13-31	45	M	Pulmonary Fibrosis	34	SLT	None	ACR, death

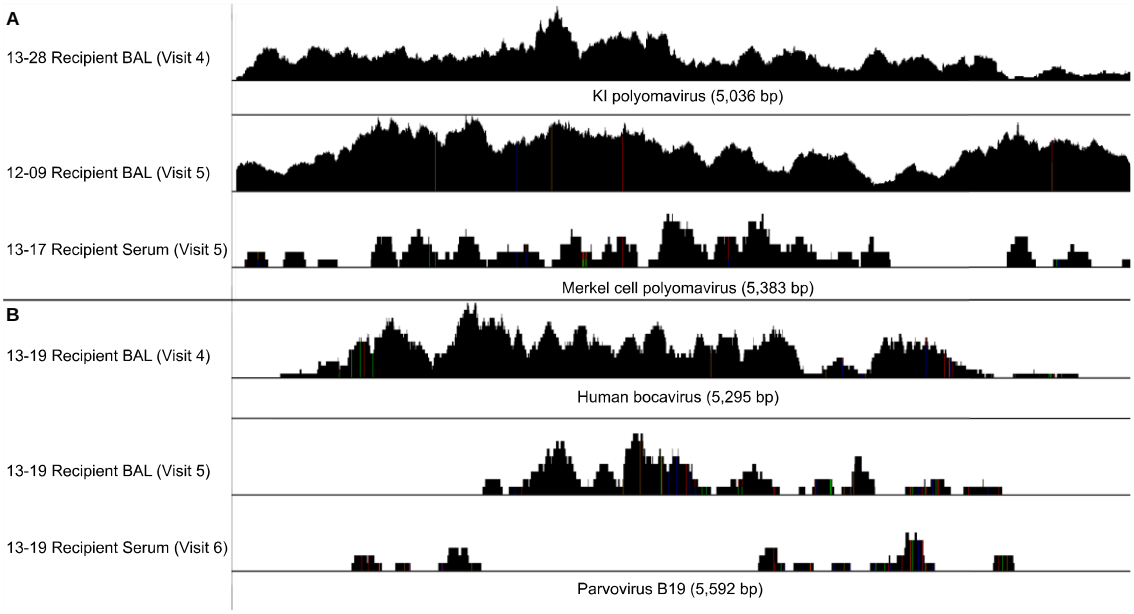
Table 4.1: Clinical Features of Lung Transplant Recipients

Subjects were derived from the multicenter Clinical Trials in Organ Transplantation-03 study. All subjects received maintenance immunosuppression post–lung transplantation and all but subject 12-12 received induction immunosuppression (Simulect/basiliximab). Infectious events reported were for the lung only and indicate any and all events in the duration of subject enrollment. The virome of peritransplant BAL samples from three subjects (12-02, 12-09, 13-17) was previously examined in a cross-sectional study of primary graft dysfunction (Abbas et al., 2017). ACR, acute cellular rejection; BLT, bilateral lung transplant; SLT, single lung transplant; F, female; M, male; MRSA, methicillin-resistant *Staphylococcus aureus*; PGD, primary graft dysfunction.

Subject	Time Point	Species	Mapped Reads	% Coverage
12-09	Visit 5 BAL	Merkel cell polyomavirus	5296	100
13-17	Visit 4 BAL	CMV	24	1
13-17	Visit 5 BAL	CMV	184	7
13-17	Visit 5 Serum	Merkel cell polyomavirus	82	72
13-17	Visit 6 Serum	CMV	348	11
13-19	Visit 4 BAL	Human bocavirus	262	87
13-19	Visit 5 BAL	Parvovirus B19	58	48
13-19	Visit 6 Serum	Parvovirus B19	22	27
13-28	Donor BAL	EBV	1848	45
13-28	Visit 4 BAL	KI polyomavirus	702	100

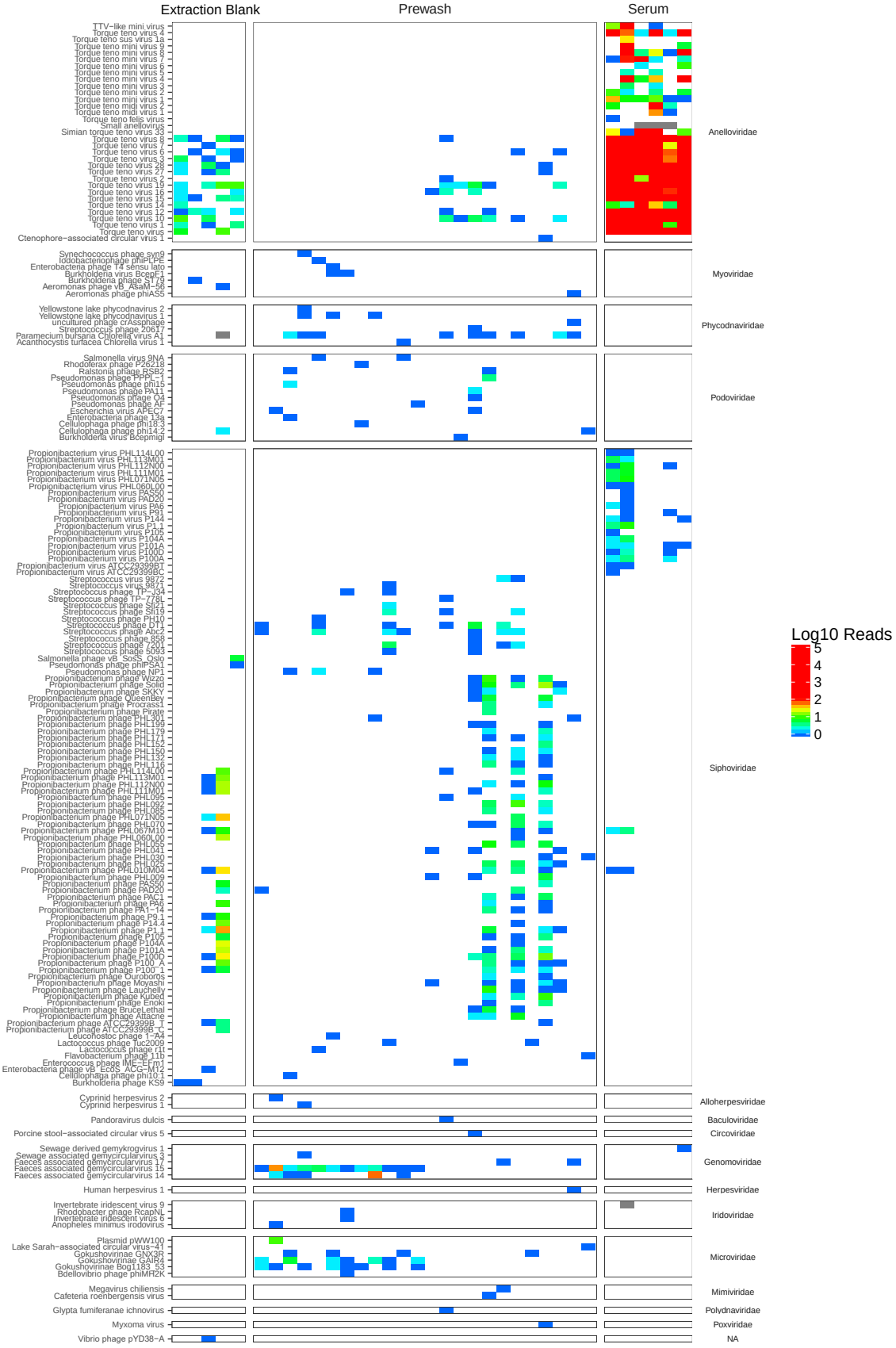
Table 4.2: Detection of Human Viruses in BAL and Serum

Viral hits discovered through the k-mer-based approach (Wood and Salzberg, 2014) were validated by aligning read pairs to reference viral genomes. The number of mapped reads and coverage of the reference genome were calculated as described in CHAPTER 2. BAL, bronchoalveolar lavage; CMV, cytomegalovirus; EBV, Epstein-Barr virus.



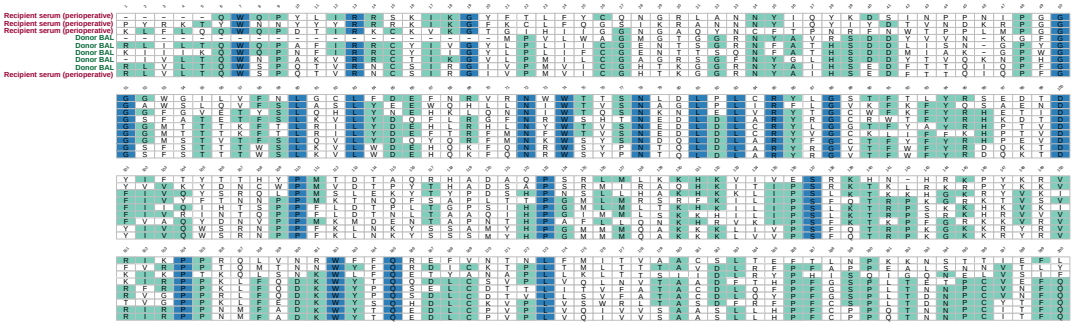
SUPPLEMENTAL FIGURE 4.1: DETECTIONS OF POLYOMAVIRIDAE AND PARVOVIRIDAE

Reads from all samples that yielded *Polyomaviridae* or *Parvoviridae* hits using the k-mer based method for viral identification were locally aligned to reference genomes of (A) *Polyomaviridae* species KI polyomavirus Stockholm 60 (NC_009238.1) and Merkel cell polyomavirus (NC_010277.2) and (B) *Parvoviridae* species human parvovirus B19 (NC_000883.2) and human bocavirus (NC_007455.1). Each row represents a BAL or serum sample, the X-axis represents the genome position (with name and length annotated on bottom) and the Y-axis represents the relative depth of coverage with vertical colored bars indicating nucleotide mismatches to the reference genome.


Extraction Blank
Prewash
Serum
Anelloviridae
Myoviridae
Phycodnaviridae
Podoviridae
Siphoviridae
Alloherpesviridae
Baculoviridae
Circoviridae
Genomoviridae
Herpesviridae
Iridoviridae
Microviridae
Mimiviridae
Polydnaviridae
Poxviridae
NA
Log10 Reads
5
4
3
2
1
0

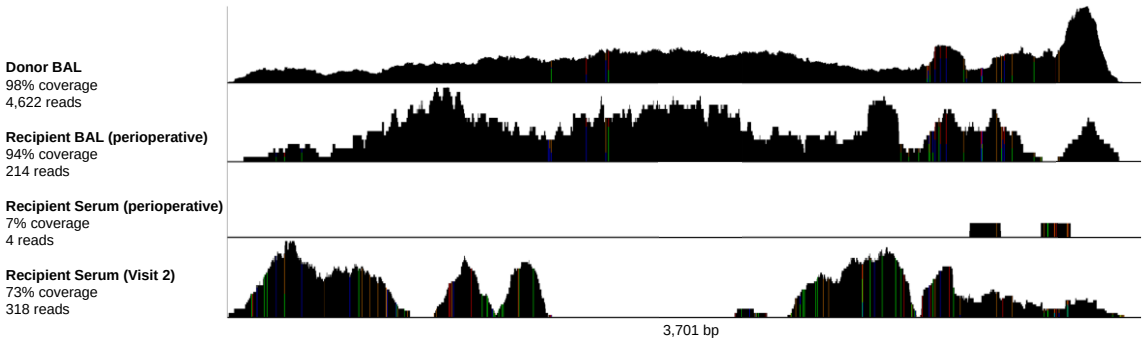
SUPPLEMENTAL FIGURE 4.2: SEQUENCING BACKGROUND OF EXTRACTION CONTROLS

The identity of reads classified as viral in two sets of negative control samples are displayed. One set (Extraction Blanks) included buffer and sterile water controls that were processed and sequenced alongside clinical samples used in this study. A second set of negative controls consisted of bronchoscope prewashes (Prewash) from bronchoscopies of subjects enrolled in a separate study of lung disease that took place at the University of Pennsylvania (Clarke et al., 2017a). This set was processed in the same laboratory and using a similar workflow. Hits to sequencing artifacts and known reagent contaminants as described in CHAPTER 2 were removed, but no minimal read count threshold was imposed for display. Each row represents a viral species and the color in each block represents the number of read hits. To contextualize the findings in negative controls, clinical samples of recipient serum from the third posttransplantation follow-up visit are also displayed.



SUPPLEMENTAL FIGURE 4.3: DIVERSITY OF ANELLOVIRIDAE LINEAGES WITHIN AND BETWEEN AN ORGAN DONOR AND RECIPIENT

Multiple sequence alignment of 200 residues of the N terminus of nine *Anelloviridae* ORF1 protein sequences (excluding the arginine-rich basic region and masking positions with 50% or more gaps) from donor BAL and perioperative serum of subject 12-02. Residues that are 50% conserved are highlighted in green, and 100% conserved are in blue. Sequences are arranged on the Y-axis based on amino acid similarity.



SUPPLEMENTAL FIGURE 4.4: DETECTION OF DONOR ANELLOVIRIDAE IN LONGITUDINAL POST-TRANSPLANTATION SAMPLES

Shotgun reads from all samples available from subject 13-28 were aligned to a representative *Anelloviridae* contig (1 of 23) found in BAL of the donor organ. Each row represents a BAL or serum sample, the X-axis represents the genome position (length annotated on bottom) and the Y-axis represents the relative depth of coverage. Vertical colored bars indicate nucleotide mismatches in the read alignments.

Sample Name	Sample Type	Raw	Human	Non-Human
11-03-V1-DB	Donor BAL	1956424	816239	165582
11-09-V1-DB	Donor BAL	3160112	1943233	130830
11-15-V1-DB	Donor BAL	3264543	11310	2346882
11-21-V1-DB	Donor BAL	860103	536864	113861
11-27-V1-DB	Donor BAL	1872248	371717	1035522
12-02-V1-DB	Donor BAL	2289969	1454800	777544
12-09-V1-DB	Donor BAL	12686623	6787759	5651326
12-12-V1-DB	Donor BAL	1874283	1603131	41609
13-17-V1-DB	Donor BAL	1047349	310244	719731
13-19-V1-DB	Donor BAL	1457706	1265293	32626
13-20-V1-DB	Donor BAL	1861666	1249573	97345
13-28-V1-DB	Donor BAL	1587655	484308	906492
13-31-V1-DB	Donor BAL	1679478	415616	1042980
11-03-R-B	Recipient BAL	439275	770	328811
11-03-V1-B	Recipient BAL	1257761	516313	301310
11-03-V4-B	Recipient BAL	1597279	109320	1112054
11-09-R-B	Recipient BAL	2346319	592297	1159872
11-09-V1-B	Recipient BAL	3181946	2545912	325010
11-09-V4-B	Recipient BAL	2419697	1705701	181870
11-15-R-B	Recipient BAL	1575492	6201	1237528
11-15-V1-B	Recipient BAL	2078888	311514	1106728
11-15-V3-B	Recipient BAL	2542535	1262714	422425
11-15-V4-B	Recipient BAL	619	153	158
11-15-V5-B	Recipient BAL	2244404	44855	1382057
11-21-V1-B	Recipient BAL	600009	229370	128567
11-21-V3-B	Recipient BAL	3786466	1139998	1811183
11-21-V5-B	Recipient BAL	3553060	83879	2619361
11-21-V6-B	Recipient BAL	3572215	158824	2388694
11-27-V1-B	Recipient BAL	1661821	1164846	34145
11-27-V3-B	Recipient BAL	2370271	1059538	624717
11-27-V4-B	Recipient BAL	2334735	526790	1194534
11-27-V6-B	Recipient BAL	1501218	122433	1020961
12-02-R-B	Recipient BAL	114865	52303	61372
12-02-V4-B	Recipient BAL	13191916	798995	12249474
12-02-V5-B	Recipient BAL	5984585	766493	5145501
12-09-V1-B	Recipient BAL	4867349	1920911	2859500
12-09-V4-B	Recipient BAL	98496	26590	70613
12-09-V5-B	Recipient BAL	37472310	277772	36921061
12-09-V6-B	Recipient BAL	2952888	361745	2563874
12-12-V1-B	Recipient BAL	1979248	1464794	39472
12-12-V3-B	Recipient BAL	1895686	611686	1011135
12-12-V4-B	Recipient BAL	1633177	175098	1228640
12-12-V5-B	Recipient BAL	2069927	13106	1781569
12-12-V6-B	Recipient BAL	1210768	70207	1001866
13-17-V1-B	Recipient BAL	1231200	1006652	211496
13-17-V4-B	Recipient BAL	22395421	600844	21552339

13-17-V5-B	Recipient BAL	12450857	1122519	11171179
13-17-V6-B	Recipient BAL	13635219	13443883	108286
13-19-R-B	Recipient BAL	1305542	188576	964193
13-19-V1-B	Recipient BAL	984399	855431	42835
13-19-V3-B	Recipient BAL	1078451	499158	500639
13-19-V4-B	Recipient BAL	1289450	901772	278211
13-19-V5-B	Recipient BAL	1168656	651485	421303
13-19-V6-B	Recipient BAL	6011680	66034	4945747
13-20-V1-B	Recipient BAL	727	260	169
13-20-V3-B	Recipient BAL	1029236	445410	528666
13-20-V4-B	Recipient BAL	789504	143840	591342
13-28-V1-B	Recipient BAL	2014647	1608986	106997
13-28-V4-B	Recipient BAL	2355760	190512	1882743
13-28-V5-B	Recipient BAL	2641574	320513	2148774
13-31-V1-B	Recipient BAL	5167797	3427281	1174587
13-31-V4-B	Recipient BAL	2981873	2075337	200046
11-03-V1-S	Recipient Serum	1813270	1640412	158761
11-03-V2-S	Recipient Serum	425917	405555	16783
11-03-V3-S	Recipient Serum	1781094	1296168	464950
11-09-V1-S	Recipient Serum	2056805	1933155	109357
11-09-V2-S	Recipient Serum	2489671	2236378	236039
11-09-V3-S	Recipient Serum	2034829	1312370	707137
11-09-V4-S	Recipient Serum	1443676	122172	1307199
11-15-V1-S	Recipient Serum	1672140	1357089	302346
11-21-V2-S	Recipient Serum	2015563	1758079	240512
11-21-V5-S	Recipient Serum	1868480	27234	1818959
11-21-V6-S	Recipient Serum	2279056	11157	2246428
11-27-V1-S	Recipient Serum	2654848	2460302	176975
11-27-V2-S	Recipient Serum	1647913	1613074	22980
11-27-V4-S	Recipient Serum	1606629	3970	1583848
11-27-V6-S	Recipient Serum	1280010	217956	1049626
12-02-R-S	Recipient Serum	897347	51614	834704
12-02-V1-S	Recipient Serum	474984	426271	44295
12-02-V2-S	Recipient Serum	705979	681949	19249
12-02-V4-S	Recipient Serum	904866	643	892271
12-02-V5-S	Recipient Serum	1083500	3606	1064928
12-09-V1-S	Recipient Serum	201763	195688	4121
12-09-V2-S	Recipient Serum	543074	526041	8657
12-09-V5-S	Recipient Serum	1474761	33149	1423914
12-09-V6-S	Recipient Serum	1009859	7403	990695
12-12-V1-S	Recipient Serum	1259010	1196317	54460
12-12-V2-S	Recipient Serum	2286524	2223321	44055
12-12-V3-S	Recipient Serum	1999146	24701	1955242
12-12-V4-S	Recipient Serum	1364719	22028	1322745
13-17-V1-S	Recipient Serum	876558	849118	19333
13-17-V2-S	Recipient Serum	1362875	1345168	5560
13-17-V3-S	Recipient Serum	1139257	6678	1115890
13-17-V4-S	Recipient Serum	1021584	1780	1004316

13-17-V5-S	Recipient Serum	1036023	708744	317118
13-19-V1-S	Recipient Serum	1978949	1911840	49523
13-19-V2-S	Recipient Serum	2621610	2061085	535953
13-19-V3-S	Recipient Serum	1343451	1300973	31399
13-19-V4-S	Recipient Serum	1440690	16797	1402755
13-19-V5-S	Recipient Serum	1088715	73275	1003314
13-19-V6-S	Recipient Serum	1207139	4985	1189693
13-20-V1-S	Recipient Serum	551122	306020	239443
13-20-V2-S	Recipient Serum	1249709	1196692	42962
13-20-V3-S	Recipient Serum	1751744	503455	1230607
13-20-V4-S	Recipient Serum	1875473	99959	1752866
13-20-V5-S	Recipient Serum	2370143	2201543	151746
13-20-V6-S	Recipient Serum	2175066	1933715	227776
13-28-V1-S	Recipient Serum	7208036	6953925	187731
13-28-V2-S	Recipient Serum	24507228	24106426	211626
13-28-V4-S	Recipient Serum	1694135	153232	1522217
13-28-V6-S	Recipient Serum	1807297	4610	1784467
13-31-V1-S	Recipient Serum	2777189	1584123	1167929
13-31-V2-S	Recipient Serum	1856278	1838968	4109
13-31-V4-S	Recipient Serum	1741316	1340777	385397
BufAve-12-2015	Buffer	4721	146	4311
SM-12-2015	Buffer	53850	8455	44927
SM-1	Buffer	319451	16680	299080
SM-D-1	Buffer	1633660	295	1360248
GenomiphiNTC	PCR Water	1443	106	1310
H20NegCtrl1	PCR Water	72	27	33

Supplemental Table 4.1: Sequencing Pre-processing Summary

Shotgun sequencing libraries were prepared from DNA extracted from acellular BAL, serum and environmental controls, pooled in equimolar amounts and sequenced on single lanes in three independent 2x125 base pair Illumina HiSeq 2500 runs. Samples were de-multiplexed, and quality-filtered as described in Supplemental Methods. Reads that aligned to the human genome were removed prior to microbiome analysis. Displayed are the number of read pairs for each sample at various stages of bioinformatics processing. The bottom six rows represent various types of environmental control specimens. BAL, Bronchoalveolar lavage; B, Recipient BAL; DB, Donor BAL; NTC, No template control; R, For Cause; S, Recipient Serum; V1, Post-reperfusion; V2, Visit 2; V3, Visit 4; V5, Visit 5; V6, Visit 6.

Subject	Virus	Clinical Detection (recipient)	Metagenomic Detection	Donor IgG	Recipient IgG	Maintenance CMV prophylaxis
13-17	CMV	None	Visit 4 BAL Visit 5 BAL Visit 6 BAL	NA	NA	Yes
13-28	EBV	None	Donor BAL	Positive	Negative	Yes
13-31	CMV	None	Visit 4 Serum	Positive	Positive	Yes

Supplemental Table 4.2: Metagenomic Detection of Human Herpesviruses
Metagenomic detection was called by aligning reads of all samples from subjects with any putatively positive hits to reference Herpesvirus genomes and inspecting coverage depth and breadth. CMV prophylaxis consisted of either Ganciclovir or Valganciclovir. In some cases, data regarding seropositivity was not available (NA). BAL, Bronchoalveolar lavage; CMV, cytomegalovirus; EBV, Epstein-Bar virus.

Subject ID	Time Point	Family	Species	Read Pairs
11-09	For Cause BAL	Unclassified dsDNA virus	Micromonas pusilla virus 12T	13
11-09	Post-reperfusion serum	*Circoviridae*	Circoviridae SFBeef	62
13-28	Donor BAL	*Circoviridae*	Porcine stool-associated circular virus 5	67

Supplemental Table 4.3: Detection of Non-Human Eukaryotic Viruses in BAL and Serum

Human-filtered read pairs from each sample were assigned to a viral taxonomic class using a k-mer based (Wood and Salzberg, 2014) search against a database containing complete RefSeq viral genomes. A threshold of tem read pairs per genome per sample was used to call a positive detection. BAL, Bronchoalveolar lavage; dsDNA, double-stranded DNA.

Sample	Total N	Median Log10 Anelloviridae Copies/mL	P-value (Compared to Healthy)
Healthy BAL	8	3.90	NA
Donor BAL	13	5.58	0.006
Post-reperfusion (Visit 1) BAL	12	5.36	0.0124
Visit 3 BAL	6	5.07	0.0127
Visit 4 BAL	12	5.07	0.0157
Visit 5 BAL	8	4.47	0.279
Visit 6 BAL	6	5.56	0.0293
Healthy Serum	21	5.44	NA
Post-reperfusion (Visit 1) Serum	12	4.42	0.01
Visit 2 Serum	12	5.77	0.104
Visit 3 Serum	6	5.94	0.0973
Visit 4 Serum	9	6.18	0.0004
Visit 5 Serum	7	6.00	0.0199
Visit 6 Serum	7	5.84	0.0364

Supplemental Table 4.4: Comparison of Anelloviridae Levels in Lung Transplant Recipients and Healthy Adults

Comparison between *Anelloviridae* DNA copies at each time point and healthy samples were tested with a two-sided Wilcoxon Rank Sum test. BAL, Bronchoalveolar lavage.

Peri-Transplant Anelloviridae contig source	Recipient Longitudinal Sample Type	Cognate versus Unrelated Pair (p-value)		
		Gini Index	RPKM	Fraction Coverage
Donor Lung	Recipient BAL	<0.001	<0.001	<0.001
Donor Lung	Recipient Serum	<0.001	<0.001	<0.001
Donor Lung	All Recipient Samples	<0.001	<0.001	<0.001
Recipient Initial Blood	Recipient BAL	<0.001	<0.001	<0.001
Recipient Initial Blood	Recipient Serum	<0.001	<0.001	<0.001
Recipient Initial Blood	All Recipient Samples	<0.001	<0.001	<0.001

Supplemental Table 4.5: Transfer of Donor Lung and Recipient Serum Anelloviridae between Compartments and Persistence within Lung and Serum

The relatedness of *Anelloviridae* reads in longitudinal BAL and serum samples to the *Anelloviridae* contigs present in the initial lung allograft ("Donor Lung") or recipient serum at the time of transplant ("Recipient Initial Blood") was determined by calculation of Gini Index, RPKM and overall coverage. These coverage metrics in longitudinal samples were compared for cognate versus unrelated pairs using the Wilcoxon Rank Sum Test. Reads from the donor BAL or post-reperfusion serum samples in which contigs were originally discovered were excluded from analysis. Donor Lung versus Recipient Serum reflects transfer from lung allograft to recipient serum, whereas Donor Lung versus Recipient BAL reflects persistence of initial lung *Anelloviridae* within that recipient's lung compartment. Recipient Initial Blood versus Recipient BAL reflects migration of recipient serum *Anelloviridae* into the lung. Recipient Initial Blood versus Recipient Serum reflects persistence of *Anelloviridae* in longitudinal serum samples. BAL, bronchoalveolar lavage; RPKM, reads per kilobase per million sequenced.

Subject	Donor Contig ID	Gini index	% Coverage
11-03	398	NA	0
11-21	130	NA	NA
11-21	143	NA	NA
11-21	144	NA	NA
12-02	1	0.922	11
12-02	2	NA	0
12-02	3	NA	0
12-02	4	NA	0
12-02	5	0.703	91
12-09	1083	NA	0
12-09	1084	NA	0
12-09	1136	NA	0
12-09	1341	0.489	73
12-09	1415	0.957	5
12-09	1779	0.327	86
12-09	2352	NA	0
12-09	998	NA	0
13-17	103	NA	0
13-17	132	NA	0
13-17	146	NA	0
13-17	30	NA	0
13-17	31	NA	0
13-17	32	NA	0
13-17	33	NA	0
13-17	34	NA	0
13-17	35	NA	0
13-17	41	NA	0
13-17	42	NA	0
13-17	52	NA	0
13-17	54	NA	0
13-17	57	NA	0
13-17	60	NA	0
13-17	61	NA	0
13-17	62	NA	0
13-17	63	NA	0
13-17	64	NA	0
13-17	65	NA	0
13-17	68	NA	0
13-17	70	NA	0
13-17	71	NA	0
13-17	73	NA	0
13-17	74	NA	0
13-17	86	NA	0
13-17	93	NA	0
13-17	97	0.913	9
13-28	1007	NA	0

13-28	1048	0.912	9
13-28	1118	NA	0
13-28	1145	NA	0
13-28	1149	NA	0
13-28	1200	NA	0
13-28	1224	NA	0
13-28	1253	NA	0
13-28	1263	NA	0
13-28	1276	NA	0
13-28	1329	NA	0
13-28	144	0.953	32
13-28	1468	NA	0
13-28	357	0.974	4
13-28	657	NA	0
13-28	735	0.932	7
13-28	758	0.973	4
13-28	765	NA	0
13-28	775	0.906	13
13-28	938	NA	0
13-28	949	NA	0
13-28	965	NA	0
13-28	971	0.611	78
13-31	252	0.321	100
13-31	303	NA	0
13-31	391	0.215	100

Supplemental Table 4.6: Coverage of Donor Lung Anelloviridae Contigs by Initial Recipient Serum Sequences

Contigs were identified as *Anelloviridae* by nucleotide alignment to the NCBI viral database (E-value $<10^{-5}$). Reads from recipient post-reperfusion serum samples were uniquely mapped to all 71 donor lung *Anelloviridae* contigs. The percentage of each genome covered by aligned reads and the evenness of coverage (Gini index) was determined. A Gini index of 1 is highly uneven coverage, while 0 is more even coverage, suggesting greater similarity to the target genome. This allows identification of those contigs whose definitive detection post-transplantation might be confounded by baseline similarity between donor and recipient viruses. In some cases (NA), metrics were not able to be calculated due to lack of samples or contig sequences.

Subject	Recipient Contig ID	Gini index	% Coverage
11-03	11	NA	0
11-03	12	NA	0
11-03	15	NA	0
11-03	16	NA	0
11-03	20	NA	0
11-03	21	0.939	7
11-03	8	NA	0
11-03	9	NA	0
11-09	19	0.527	75
11-15	19	NA	0
11-15	22	0.581	67
11-15	23	NA	0
11-15	24	NA	0
11-15	29	NA	0
11-15	35	0.761	34
11-15	41	NA	0
11-15	42	NA	0
11-15	43	0.830	24
11-15	45	NA	0
11-15	47	NA	0
11-15	48	NA	0
11-15	52	NA	0
11-15	56	NA	0
11-15	64	0.951	5
11-15	65	NA	0
11-15	71	NA	0
11-15	91	NA	0
11-15	92	NA	0
11-15	93	NA	0
11-27	5	NA	0
11-27	7	0.775	26
11-27	8	NA	0
12-02	1	0.786	58
12-02	3	NA	0
12-02	4	NA	0
12-02	5	NA	0
12-12	3	NA	0
12-12	4	0.981	2
12-12	6	NA	0
12-12	7	NA	0
12-12	8	NA	0
13-19	1	NA	0
13-19	3	NA	0
13-19	4	0.962	4
13-19	5	NA	0

13-19	7	NA	0
13-20	17	NA	0
13-28	12	NA	0
13-28	13	NA	0
13-28	14	0.737	33
13-28	16	NA	0
13-28	7	0.567	89
13-28	8	0.919	20
13-28	9	0.906	19
13-31	10	NA	0
13-31	11	NA	0
13-31	12	NA	0
13-31	13	0.314	97
13-31	14	NA	0
13-31	15	NA	0
13-31	16	NA	0
13-31	17	NA	0
13-31	19	NA	0
13-31	20	NA	0
13-31	5	NA	0
13-31	6	0.338	95
13-31	7	NA	0
13-31	8	NA	0
13-31	9	NA	0

Supplemental Table 4.7: Coverage of Recipient Serum Anelloviridae Contigs by Donor Lung Sequences

Contigs were identified as *Anelloviridae* by alignment to the NCBI viral database. Reads from donor BAL were uniquely mapped to all 69 recipient *Anelloviridae* contigs. The fraction of each genome covered by aligned reads and the evenness of coverage (Gini index) was determined. This allows identification of those contigs whose definitive detection post-transplantation might be confounded by baseline similarity between donor and recipient viruses. BAL, bronchoalveolar lavage.

CHAPTER 5: REDONDOVIRIDAE: A NOVEL FAMILY OF SMALL, CIRCULAR DNA VIRUSES OF THE HUMAN ORO-RESPIRATORY TRACT

The contents of this chapter are under review for publication as:

Abbas, A.A.*, Taylor, L.J.*, Dothard, M.I., Leiby, J.S., Fitzgerald, A.S., Khatib L.A., Collman, R.G., Bushman, F.D., 2019. Redondoviridae: A Novel Family of Small, Circular, DNA Viruses of the Human Oro-Respiratory Tract.

* These authors contributed equally

5.1 ABSTRACT

This chapter introduces a new family of small, circular DNA viruses—named *Redondoviridae*—identified using metagenomic sequence-based methods. We first identified two redondovirus genomes by shotgun sequencing of viral particle preparations from bronchoalveolar (lung) lavage from human organ donors. We then queried 6,377 metagenomic samples for *Redondoviridae*, recovering 17 additional complete genomes and detecting redondovirus sequences in 67 human samples, mostly from respiratory tract and oro-pharyngeal sites. *Redondoviridae* was the second most prevalent eukaryotic DNA virus family in oro-respiratory sites, after *Anelloviridae*. We quantified redondovirus genomes in samples from critically ill patients and found that they were abundant in some patients. Analysis of redondovirus sequences in metagenomic data sets revealed an association with periodontal disease. Thus we propose that redondoviruses are widespread human viruses colonizing oro-respiratory sites and blooming in several clinical conditions.

5.2 INTRODUCTION

Viruses are the most abundant biological entities on Earth, but global viral populations (the "virome") are still mostly uncharacterized. Identifying novel viruses can be difficult if they have limited sequence homology to viral genomes in reference databases. Recent advances in sample preparation and sequencing techniques have uncovered a world of new viruses (Paez-Espino et al., 2016, Simmonds et al., 2017, Rosario and Breitbart, 2011, Minot et al., 2013, Minot et al., 2011). However the majority of reads in most studies remain unclassified (Aggarwala et al., 2017, Krishnamurthy and Wang, 2017), leaving our understanding of the virome incomplete. Here we describe the identification of a new human-specific viral family, its localization in oro-respiratory sites, and its association with disease states.

Metagenomic sample preparation often involves multiple displacement amplification (MDA) of virome DNA fractions with a highly processive, strand-displacing DNA polymerase. This process enriches for small, circular, single-stranded DNA (ssDNA) molecules and has facilitated the discovery of viruses with such genomes (Rosario et al., 2012, Labonté and Suttle, 2013, Krupovic et al., 2016). Many of these ssDNA viruses encode a replication initiation protein (Rep)—thus this group is collectively known as circular Rep-encoding single-stranded DNA (CRESS) viruses (Rosario et al., 2012). The genome architecture and functional domains of major viral proteins of CRESS viruses are conserved, though pairwise nucleotide identities between viral families are often low. A well-studied group of animal CRESS viruses is the Circovirus genus within the *Circoviridae* family, which includes pathogenic viruses of swine and birds (Ellis, 2014, Todd, 2000). The *Circoviridae* family also contains the genus Cyclovirus, which consists of viruses identified by metagenomic sequencing in samples from several mammalian species (Breitbart et al., 2017, Li et al., 2010), including some sporadically identified in human disease states (Phan et al., 2014, Smits et al., 2014). The recently identified *Smacoviridae* family has been detected in mammalian feces, though the definitive hosts are unknown (Varsani and Krupovic, 2018). Other CRESS families include plant pathogens *Geminiviridae* and *Nanoviridae* (Harrison et al., 1977, Fauquet et al., 2005), *Genomoviridae* which infect fungi (Krupovic et al., 2016, Varsani and Krupovic, 2017), and many additional apparent viral genomes with unknown hosts and limited resemblance to established taxa (Simmonds et al., 2017).

We and others have previously investigated the human respiratory tract virome in health and disease. Typically anelloviruses, herpesviruses and bacteriophages dominate human respiratory tract samples (Willner et al., 2009, Abeles et al., 2015, Wylie et al., 2012, Pérez-Brocal and Moya, 2018, Young et al., 2015, Abbas et al., 2017, Abbas et al., 2018,

Clarke et al., 2017a). Recently, we identified short sequence reads with limited homology to a swine-associated CRESS virus (Cheung et al., 2014) in bronchoalveolar lavage (BAL) of organ donors (Abbas et al., 2018, Abbas et al., 2017), raising the possibility that we had detected a new human CRESS virus. However, the nature, origin, distribution, and clinical significance of these sequences were unclear from these initial studies.

Following up on this lead, we ultimately identified a widespread group of highly divergent CRESS viruses in human respiratory and oral samples. After our initial detection of weak resemblance to the porcine CRESS virus in two BAL samples, we assembled reads into two complete genomes, which were verified by PCR amplification and Sanger sequencing. Using these genomes as alignment targets, we found homologous sequences in 67 human samples. From these we recovered 17 additional complete genomes, mainly from lung and oropharynx. These CRESS genomes are sufficiently different from previously described taxa that we propose that they are members of a new family, which we name *Redondoviridae* (*redondo*—Spanish for "round") containing the new genera Vientovirus and Brisavirus (from the Spanish words for "wind" and "breeze", alluding to their discovery in the respiratory tract). A recently described genome identified in upper respiratory secretions of a febrile individual (Cui et al., 2017) is homologous to these new sequences and is likely also a redondovirus. We validated that redondovirus sequences were authentically present in the original clinical samples and not found in contamination controls. Analysis of the distribution of redondoviruses showed that they were the second most prevalent virus in respiratory samples, after anelloviruses. Finally, using metagenomic alignments and targeted quantitative PCR (qPCR) assays to query human samples, we identified an association of *Redondoviridae* with periodontal disease and acute critical illness.

5.3 RESULTS

Initial Discovery of Redondoviruses in Human Bronchoalveolar Lavage Fluid

In two previous studies of lung transplant recipients (Abbas et al., 2017, Abbas et al., 2018), samples of BAL were enriched for viral particles, then RNA and DNA was purified and subjected to metagenomic sequencing and analysis. DNA fractions were amplified using MDA. Alignment of reads from two organ donor BAL samples to the NCBI viral genome database showed modest (14%) coverage of Porcine stool-associated circular virus 5 (PoSCV-5) isolate CP3 (GenBank: NC_023878) (Figure 5.1). PoSCV-5 is currently an unclassified and unstudied member of the *Circoviridae* family.

After assembling reads from these samples into contigs, we found that sequences matching PoSCV-5 were incorporated into circular contigs of approximately 3000 base pairs (bp). Thus, whole viral genomes were present in the initial BAL samples, but only a small region of these genomes resembled PoSCV-5. Several sets of nested primers (Supplemental Table 5.1 and Supplemental Figure 5.1) were used to amplify overlapping fragments from the original BAL samples. These fragments were sequenced using the Sanger method and assembled to construct two circular genomes of 3,026 bp (Human lung-associated brisavirus RC; accession MK059757) and 3,056 bp (Human lung-associated vientovirus FB; accession MK059763) (Figure S1).

Contigs assembled from shotgun metagenomic reads of other BAL samples processed by our group were then queried for DNA sequence similarity to the two novel genomes. In total, seven complete *Redondoviridae* genomes were discovered and cloned from independent BAL samples from organ donors and patients with sarcoidosis (Supplemental Figure 5.1). The full set of new genomes was then used as alignment targets to interrogate publicly available datasets. A total of 12 more samples had sufficient coverage of redondovirus sequences to allow assembly, yielding 19 complete genomes (Figure 5.1A, Supplemental Table 5.2).

A danger is that small circular viruses may be derived from environmental contamination in clinical or laboratory reagents (Naccache et al., 2013, Salter et al., 2014). We queried 144 contamination controls from seven studies analyzed by shotgun metagenomics, and found no evidence for redondovirus genomes. This included 24 bronchoscope prewash controls, which consist of sterile saline solution passed through bronchoscopes before insertion into a patient; prewashes were also subjected to MDA prior to shotgun metagenomic sequencing (Clarke et al., 2017a).

We next used a qPCR assay targeting redondovirus genomes to further test the 24 bronchoscope prewashes and two additional DNA extraction controls subjected to MDA. All were negative by qPCR analysis. As positive controls, we detected robust qPCR signals in MDA-amplified DNA extracted from the original acellular BAL samples from which these genomes were cloned (Supplemental Figure 5.1C).

As a further check on our interpretation of the putative human origin of *Redondoviridae*, we investigated whether they could be bacteriophages. The presence of prokaryotic ribosomal binding sites (RBS) upstream of viral open reading frames (ORFs) can provide evidence for a prokaryotic host. We implemented the algorithm described in Krishnamurthy and Wang, 2018 and identified no prokaryotic RBS proximal to any redondovirus protein coding sequence. These data support the ideas that redondovirus sequences were not derived from environmental contamination and are not bacteriophages.

Redondovirus Genomes Contain Conserved Features of CRESS Viruses and a Novel ORF

The redondovirus genome is larger than Circovirus genomes (3.0 kb versus 1.7-2.0 kb; Table 5.1) and contains two ambisense ORFs encoding a 334-363 amino acid Rep and a 449-531 amino acid capsid (Figure 5.1B). These proteins are only 10-15% identical to those of porcine circoviruses 1 and 2, and 40-55% for PoSCV-5. All redondovirus genomes also contain a third ORF overlapping the capsid gene (*Cp*). ORF3 is not found in porcine

circoviruses or in PoSCV-5 and has no homology to any described protein sequence or family. PoSCV-5 is the most closely related known virus, but it is sufficiently divergent in genome architecture and protein identity that we do not consider it a member of the *Redondoviridae* family.

Redondoviruses display considerable sequence divergence when comparing their capsid and Rep proteins. The range of pairwise amino acid identities of capsid is 67.5-99.6% (median 82.3%) while the range of Rep amino acid identity is 36.6-99.7% (median 54%) (Figure 5.1B). Surprisingly, *Cp* is more highly conserved than *Rep*. One might have expected that the capsid protein, which is presumably recognized by host antibodies, would be under stronger diversifying (positive) selection. Part of *Cp* overlaps ORF3, and so could be constrained in sequence drift for that reason, but even in the non-overlapping carboxy-terminal coding region (Figure 5.1B), the variability is still lower than in *Rep*.

To clarify the phylogenetic relationships between viral proteins within the *Redondoviridae* family and compare to those from other CRESS virus families, we built maximum-likelihood phylogenetic trees of *Rep* and *Cp* protein sequences. Redondoviruses are more similar to each other than to other CRESS families by protein identity and genome organization (Figure 5.2, Table 5.1). The capsid and Rep protein phylogenies show different relationships between the isolates, suggesting that recombination is common in redondoviruses, as in other circular ssDNA viruses (Ma et al., 2007, Lefeuvre et al., 2009, Fahsbender et al., 2017, Leppik et al., 2007). Based on the definitions in Varsani and Krupovic, 2018 and analysis of the diversity of viral Rep proteins, redondovirus genomes can be grouped in two genera, demarcated by 50% Rep protein identity, which we propose be called Vientovirus and Brisavirus (Supplemental Table 5.2).

The redondovirus Rep protein (Figure 5.3B) contains two domains found in many small DNA and RNA viruses: one involved in rolling-circle replication (Pfam: PF00799) and

a second helicase domain within the P-loop NTPase superfamily (Pfam: PF00910). Multiple sequence alignments revealed conserved motifs (Ilyina and Koonin, 1992, Gorbalenya et al., 1990) characteristic of these domains.

Redondovirus capsids, like those of other ssDNA viruses, contain a basic amino-terminus. Protein modeling by PHYRE2 (Kelley et al., 2015) weakly predicted (58% confidence over 7% of sequence) that it contains folds similar to coat proteins of ssRNA viruses that infect plants (Figure 5.3C).

By analyzing the synonymous to non-synonymous mutation ratio (dN/dS), we identified four clusters of sites potentially under positive selection in the carboxy-terminal portion of capsid that does not overlap with ORF3 (Figure 5.3C). We also identified two sites of possible diversifying selection in Rep (Figure 5.3B), suggesting a possible response to innate immune pressure. It should be noted that dN/dS as a marker of selective pressure is untested in CRESS viruses and may be confounded by overlap of unidentified coding sequence and/or functionally important DNA secondary structure elements (Zanini and Neher, 2013, Muhire et al., 2014).

Circovirus genomes typically contain a conserved motif ("NANTATTAC") within a stem-loop structure followed by short direct repeats, located in the intergenic region at the 5' end of *Cp* and *Rep*. Such sequences are candidates for the origin of replication (Mankertz et al., 1997) where the viral-encoded Rep binds and cleaves, mediating replication by host polymerases. Such stem-loops are found in other CRESS virus families. In the one previous report of a single redondovirus genome (Cui et al., 2017), the authors suggested a hairpin in the large intergenic region as the origin of replication. However, analysis of all 20 genomes showed that a conserved, stable stem loop structure is predicted to form in the smaller intergenic region, partially overlapping *Rep*. Although the length of the stem, size of the loops, and presence of downstream direct repeats vary, most redondovirus genomes

contain a nonanucleotide motif ("TATTATTAT") (Figure 5.1B) similar but not identical to that of other CRESS viruses. This structure is highly conserved among redondoviruses (Figure 5.3A), while the sequence of the alternative intergenic hairpin is not, suggesting that this is a more likely candidate for the replication origin.

Redondovirus Genomes Identified in Shotgun Metagenomic Data

To investigate these divergent sequences further and determine their distribution in the biosphere, we surveyed our own and publicly available metagenomic datasets for homology to redondoviruses (Supplemental Table 5.3 presents the datasets queried). Studies were favored for analysis if they 1) biochemically enriched for viral nucleic acids, 2) used MDA, which enriches for small circular viral genomes (Kim and Bae, 2011, Kim et al., 2008), 3) reported detection of Circovirus-like sequences, and/or 4) included a diverse range of sample types. In total, we queried 6,377 samples from 97 datasets covering 30 organisms or environments. Within human metagenomes, 18 body sites or fluids were examined. A positive hit was defined as 25% coverage of any redondovirus genome. Redondoviruses were detected in metagenomic sequences from human oral cavity (3.8% of datasets), lung (3.3%), nasopharyngeal (0.95%), and gut (0.59%). The most frequent sites of detection were the mouth and respiratory tract (Figure 5.4A). Redondovirus sequences were rare in other human body sites, and not found in other animals, freshwater, marine, or soil samples, nor in laboratory reagents. We thus conclude that redondoviruses are CRESS viruses that are authentically present in the human oro-respiratory tract; whether infrequent detection in gut samples reflects an authentic site of replication or transient passage after swallowing is uncertain.

Redondovirus Co-occurrence with Human DNA Viruses

Adeno-associated virus, a ssDNA virus of the *Parvoviridae* family that encodes capsid and Rep proteins, is known to require coinfection with a helper virus such as

adenovirus to replicate, leading us to ask whether redondoviruses co-occurred with any other eukaryotic viral family. We analyzed a subset (20) of the 97 datasets we previously screened for redondoviruses for the presence of common human DNA virus families (*Adenoviridae, Anelloviridae, Herpesviridae, Papillomaviridae, Parvoviridae* and *Polyomaviridae*) from eight human body sites. *Redondoviridae* was the second-most frequent human DNA virus family detected, exceeded only by *Anelloviridae*, which are known to be ubiquitous in humans (Spandole et al., 2015). Figure 5.4C shows the representation of additional human DNA viruses that co-occurred with redondoviruses in metagenomic datasets. Only anelloviruses were found to co-occur significantly with redondoviruses (Figure 5.4C, p=5.7x10-7, Fisher's Exact Test with Bonferroni correction). Anelloviruses are small ssDNA viruses that seem unlikely to contribute helper functions to redondovirus replication. We speculate that the inflammatory milieu known to favor anellovirus replication (Maggi et al., 2001, Mariscal et al., 2002) may be similarly favorable for redondoviruses. Alternatively, given the ubiquitous nature of anelloviruses in humans, this association may reflect the fact that MDA enriches for both anelloviruses and redondoviruses, resulting in their co-detection. Rarely, other human viruses were found in redondovirus positive samples; these included Human mastadenovirus C and Epstein-Barr virus.

Redondoviruses in the Respiratory Tract are Elevated in Abundance in Critical Illness

Several sample sets were further queried using metagenomic analysis and qPCR to assess redondovirus abundance in the respiratory tract. We investigated 916 selected oro-respiratory samples using metagenomic analysis of datasets described above, reflecting a mixture of health and disease states, and found that redondoviruses were still the second-most frequent DNA virus detected, after anelloviruses (Supplemental Figure 5.3).

To investigate the presence of redondovirus in healthy subjects further, we tested DNA isolated from oropharyngeal swabs from 60 adults using qPCR (Charlson et al., 2010). DNA was subjected to selective whole genome amplification (SWGA) (Clarke et al., 2017b) to enrich for redondovirus sequences over the human genome background, followed by redondovirus qPCR. Nine of 60 healthy subjects were positive (15%), although quantities even following SWGA amplification showed generally modest levels (Figure 5.5A).

We then tested samples from 67 critically ill individuals using SWGA and qPCR (Figure 5.5A and B). Six (9%) had oropharyngeal samples positive for redondovirus. Post-SWGA quantities were, on average, 104-fold greater than in healthy subjects, although the use of SWGA prohibits comparison of absolute quantities between groups. Four critically ill subjects also had lung secretions (endotracheal aspirates) available for testing; three were positive for redondovirus. In subjects with serial samples, redondovirus was generally detectable over a period of 2-3 weeks, suggesting persistent colonization or infection. We conclude that redondoviruses are found in both healthy and critically ill individuals, but levels are elevated in illness. Furthermore, the upper and lower respiratory tracts appear to represent common niches with stable redondovirus detection over time.

Redondovirus Sequence Reads are Associated with Periodontitis

The set of 97 metagenomic studies assessed for redondovirus sequences (Supplemental Table 5.3) contained samples from several disease states, allowing us to assess possible associations of redondoviruses with human disorders. In addition to our initial detections in BAL from organ donors and lung transplant recipients (Abbas et al., 2017, Abbas et al., 2018), redondoviruses were found in 1) BAL from subjects with sarcoidosis, 2) healthy controls (Clarke et al., 2017a), 3) gingival samples from subjects with periodontitis (Wang et al., 2013, Shi et al., 2015, Califf et al., 2017), 4) oropharyngeal and nasopharyngeal samples from febrile subjects (Mokili et al., 2013, Wang et al., 2016), 5) oral

samples from subjects with rheumatoid arthritis (Zhang et al., 2015), 6) stool samples from healthy individuals, 7) stool samples from subjects with inflammatory bowel disease (Norman et al., 2015) and 8) stool samples from subjects with HIV-associated immunodeficiency (Monaco et al., 2016) (Figure 5.4B).

A considerable proportion of redondovirus-positive samples were from studies of periodontal disease (Figure 5.4B), so we analyzed redondoviruses in periodontitis further. Three studies queried gingival or oral samples from subjects suffering or recovered from periodontitis. One study queried samples before and after corrective treatment by scaling and root planing together with improved oral hygiene (Shi et al., 2015). Redondovirus representation was high prior to treatment, and then fell substantially after treatment, as measured by the number of reads aligning to the most broadly covered redondovirus genome in each sample (Figure 5.5C). We averaged redondovirus reads across all individual tooth sites sampled for each subject and found lower redondovirus prevalence after recovery (Figure 5.5C, $p=0.014$, Wilcoxon signed-rank test). The second study compared disease severity in two groups of subjects with chronic periodontitis; one group received treatment with 0.25% sodium hypochlorite rinse, while the other received a water rinse (Califf et al., 2017). We compared redondovirus representation in sub and supra-gingival sites from subjects whose periodontitis did or did not improve, and found that subjects that did not show improvement had greater number of reads mapping to redondovirus genomes (Figure 5.5D, $p=0.028$, Wilcoxon rank-sum test). A third study analyzed two patients with severe periodontal disease, before and after treatment; both subjects were positive for redondovirus prior to treatment but no detections were found in samples taken after successful treatment (Kumar et al., 2018). Thus we conclude that redondoviruses are associated with periodontitis in multiple studies, and that levels are reduced with effective treatment.

5.4 DISCUSSION

In this study, we introduce *Redondoviridae*, a family of small, circular DNA viruses that appears to be restricted to humans and selectively found in lung and oro-pharyngeal sites. Our first detection of redondovirus genomes resulted from aligning metagenomic sequences from lungs of two organ donors to a viral genome database, resulting in weak hits to PoSCV-5. Assembly of shotgun metagenomic reads yielded complete circular genomes, which were then used to interrogate our collection of lung virome samples, allowing us to identify seven genomes. We used these genomes to interrogate 6,377 metagenomic samples encompassing multiple environmental sites, hosts, body sites, and disease states, detecting redondoviruses in 67 human samples and building 12 additional genomes. Independently, another group reported a single genome (Cui et al., 2017) most closely related to Human oral-associated brisavirus YH (accession MK059758), in the throat of a febrile patient. Of the DNA viruses we surveyed in 20 human virome datasets, redondoviruses were the second most abundant, exceeded only by anelloviruses, recognizing that the methods used in virome sequencing are designed to sample viral particles, indicating replicating rather than latent viruses. The prevalence of redondoviruses was similar in cohorts of healthy subjects and critically ill subjects, although higher post-SWGA genome quantities imply higher absolute levels in the ill subjects. Analysis of metagenomic samples revealed an association of redondoviruses with periodontal disease.

Previous studies have tentatively implicated viruses in periodontitis based on alterations of subgingival bacteriophage communities (Ly et al., 2014) and increased representation of some eukaryotic viruses including HIV, HCMV and HSV-1(Cappuyns et al., 2005, Li et al., 2017). It is possible that redondovirus infection and replication may help maintain the inflammatory state associated with periodontitis and contribute to disease

progression. A role in disease initiation seems less likely given the established roles of bacteria and oral hygiene (Edlund et al., 2015, Costalonga and Herzberg, 2014). The role of redondoviruses in periodontitis warrants further study. Similarly, what role redondoviruses play in diseases of the respiratory tract are unknown but can now be investigated.

Do redondoviruses require helper viruses to replicate? The Dependoparvoviruses, such as adeno-associated virus, are small, linear ssDNA viruses that require co-infection with larger DNA viruses to condition cells for efficient replication. Samples containing redondoviruses were scanned for other DNA viruses, but no large double stranded DNA viruses were consistently identified. Anelloviruses, small ssDNA viruses, did co-occur. While we do not rule out that anelloviruses support redondovirus replication, it seems more likely that the inflammatory states known to promote anellovirus replication may do the same for redondoviruses, or alternatively, that the methods for virome sampling preferentially recover both redondoviruses and anelloviruses.

The high level of sequence variation in redondovirus Rep proteins is intriguing. Viruses encoding Reps are ubiquitous in both prokaryotes and eukaryotes. There are even transposon families that mobilize via ssDNA intermediates using Rep-like enzymes (Grabundzija et al., 2016). Cells have likely been opposing parasitism by Rep-encoding elements since the origins of cellular life. We conjecture that Rep amino acid variation reflects an ongoing Red Queen's Race between host intrinsic immunity and Rep enzymes. If so, there should be active host cell mechanisms targeting and inhibiting Rep proteins. The redondovirus Rep enzymes reported here provide an entry point to investigating this possibility.

5.5 ACKNOWLEDGEMENTS

This work was supported by NIH grants R61-HL137063 (Lung Virome in Health and Disease) and R01-HL113252 (Lung Transplant Microbiome and Chronic Allograft Dysfunction). A.A.A was supported by NSF grant DGE-1321851 and L.J.T by T32-AI-007324. We thank Jason D. Christie and members of the lung transplant program for access to transplant specimens and ICU staff for assistance with specimen collection. We are grateful to members of the Bushman and Collman laboratories for suggestions, Arvind Varsani for help with virus taxonomy, and for the laboratories of X.J. Meng and Matthew Weitzman for assistance and support. Louis J. Taylor was an equal contributor to the design and execution of the experimental methodology, interpretation of data and writing the manuscript. Ayannah S. Fitzgerald, Lalya A. Khatib and Jacob S. Leiby collected and processed human clinical samples obtained from the ICU. Marisol I. Dothard performed and analyzed qPCR assays on ICU samples. Frederic D. Bushman and Ronald G. Collman supervised all experiments, data analysis and assisted in writing the manuscript.

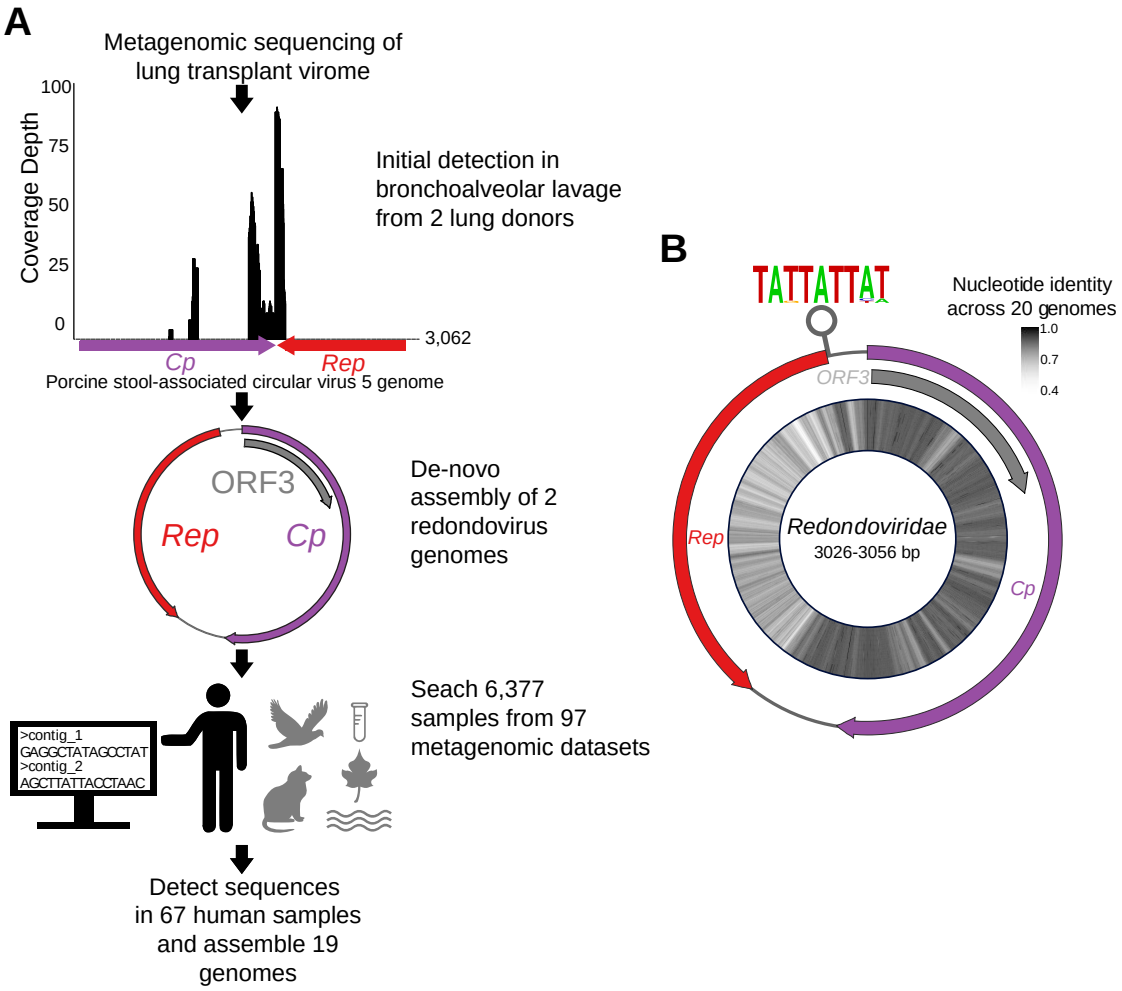


FIGURE 5.1: DISCOVERY OF REDONDOVIRUS GENOMES IN METAGENOMIC SAMPLES

A-Several hundred shotgun metagenomic reads from two organ donor BAL virome samples were identified as having limited homology to Porcine stool-associated circular virus 5 (PoSCV-5). Reads from these samples were assembled into two contigs, which were then cloned from multiple displacement amplified sample DNA using target specific primers and Sanger sequenced. The complete circular genomes were used to query additional internal and public microbial metagenomic datasets. Target-specific amplification, sequencing and genome assembly was repeated for additional samples with sequences homologous to these novel genomes if the original DNA was available. In cases where original samples were not available, metagenomic contigs were checked for circularity and completeness. A total of 19 complete genomes were recovered from 67 human samples (bottom). See also Supplemental Figure 5.1.

B-The genomic architecture of redondoviruses shows ambisense open-reading frames (ORFs) encoding a conserved capsid, Replication associated protein (Rep) and unknown protein (ORF3). The average nucleotide identity of 20 *Redondoviridae* members (19 genomes discovered here and one genome previously reported (Cui et al., 2017) is shown on the inside of the genome map as a heatmap. A putative origin of replication stem-loop structure with a conserved nonanucleotide motif is predicted to form in the 5' end of the Rep coding region. The height of the letter in the motif represents its frequency.

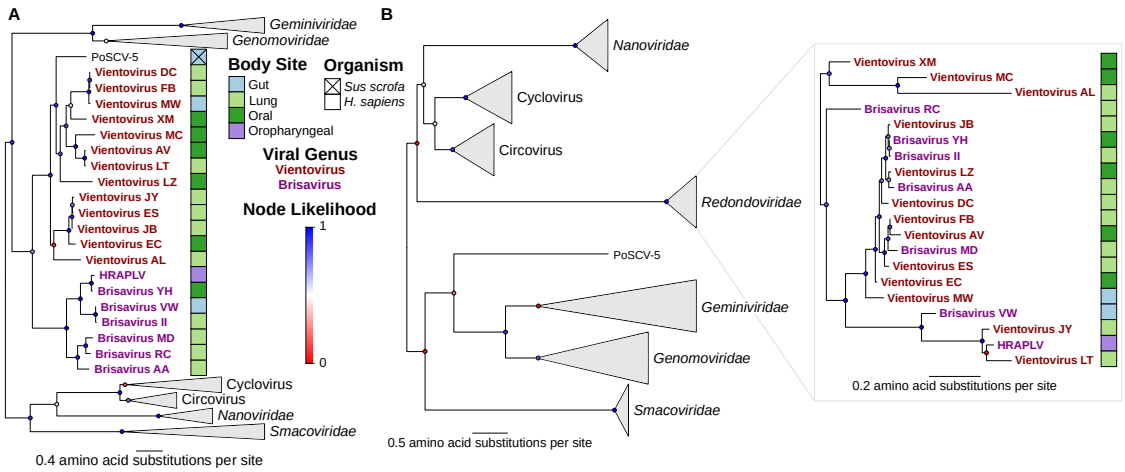


FIGURE5.2 REDONDOVIRIDAE IS A DISTINCT VIRUS FAMILY BASED ON CAPSID AND REP IDENTITIES

Phylogenetic trees of redondovirus Rep (A) and capsid (B) proteins from CRESS DNA viruses. Collapsed viral genera or families are indicated by grey triangles. Branch likelihood, determined by approximate likelihood ratio test, is shown by colored circles at each node and the scale shows amino acid substitutions per site. The sample type of origin for each redondovirus is shown as colored boxes next to each virus' name, which is colored to reflect genus designation. See also Supplemental Table 5.2.

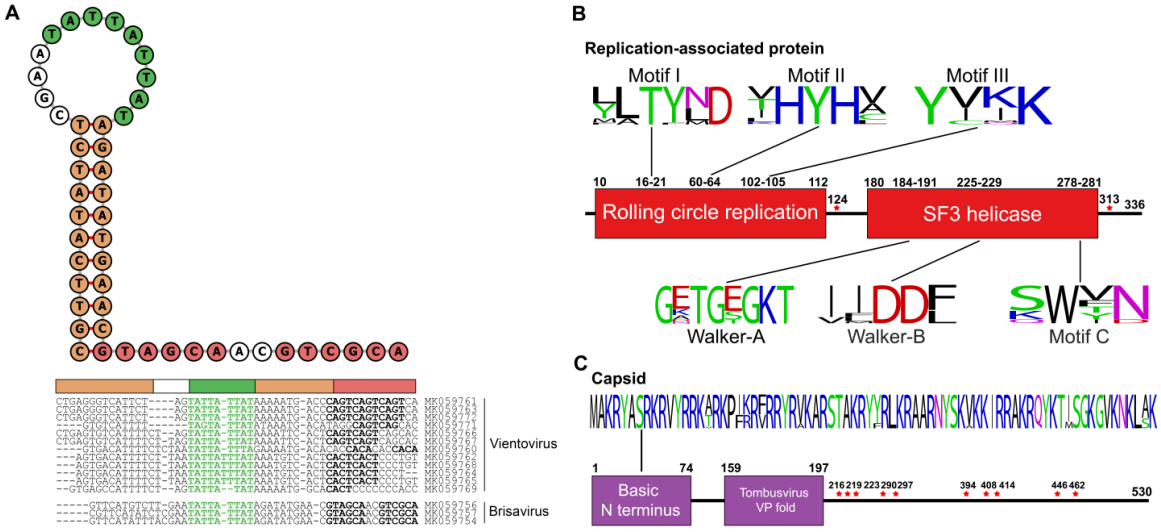


FIGURE 5.3: REDONDOVIRUS GENOMES CONTAIN CONSERVED MOTIFS IMPLICATED IN ROLLING-CIRCLE REPLICATION

A-The sequence and predicted structure of the putative replication origin of Human lung-associated brisavirus AA is shown in the top left. The inverted repeat forming the stem is shown in orange, the nonanucleotide motif within the loop is shown in green, and an imperfect six bp direct repeat sequence is shown in purple. Individual predicted stem loop sequences (threshold for stability: ΔG°<-5 kcal/mol) are shown to the right of the folded sequence. The calculated ΔG° of melting for the predicted stem-loops ranges from -5.0 to -9.45 kcal/mol.

B-Conserved rolling circle replication and superfamily 3 (SF3) helicase motifs were found in redondoviruses. The positions for the motifs are given using the Human lung-associated brisavirus AA genome sequence (Accession MK059754). The height of each letter represents its frequency. Single amino acid positions predicted to be under positive selection pressure are marked by a red star.

C-The putative redondovirus capsid protein contains a basic amino-terminus and a predicted virus coat protein-like fold. The positions for the motifs correspond to Human lung-associated brisavirus AA, as above. Amino acid positions predicted to be under positive selection pressure are marked by a red star.

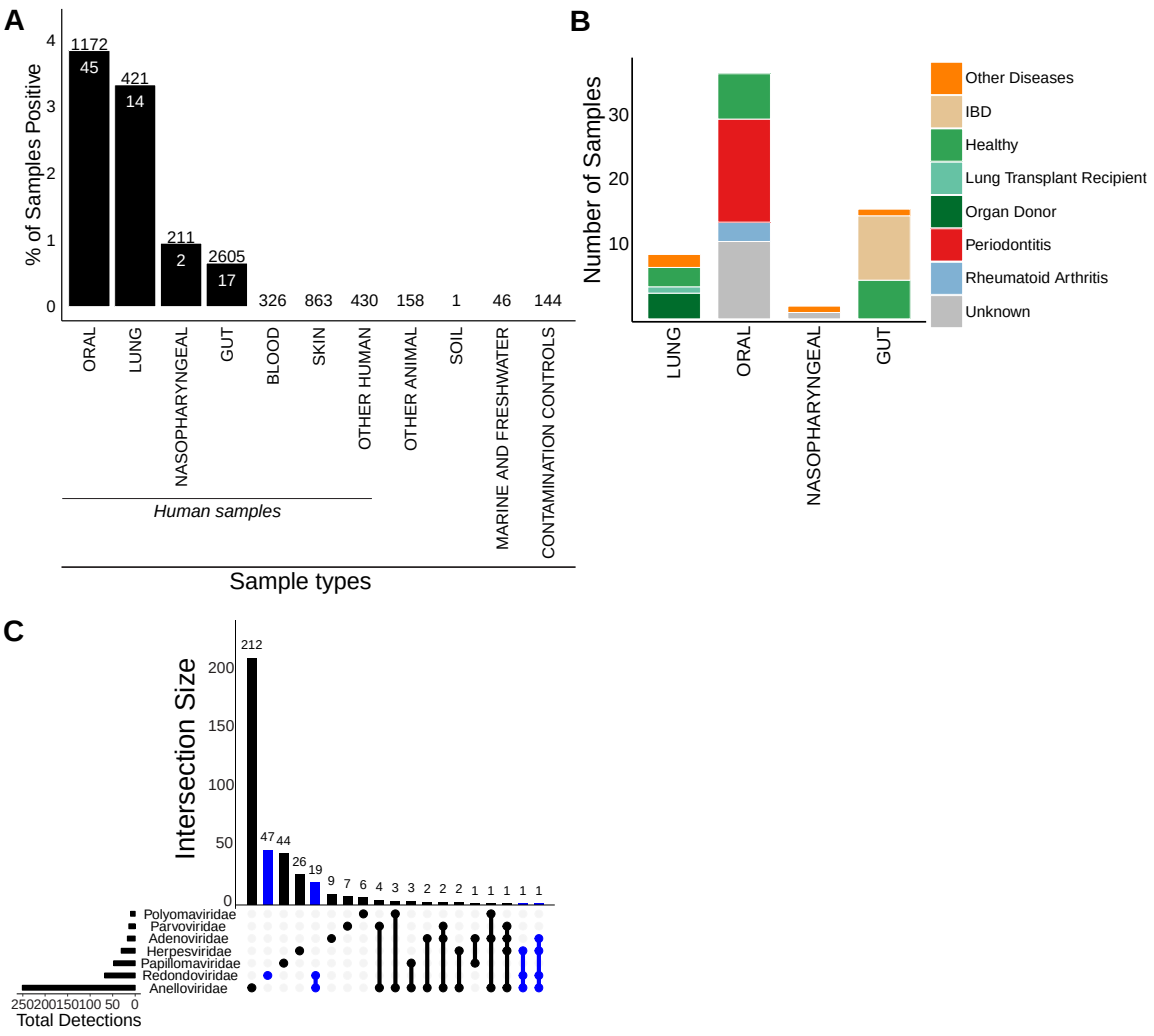


FIGURE 5.4: FREQUENCY OF REDONDOVIRUS DETECTION AND CO-OCCURRENCE WITH HUMAN DNA VIRUSES

A-Reads from 97 metagenomic datasets encompassing different human and non-human sample types were aligned to redondovirus genomes. A positive hit was determined based on 25% coverage of any redondovirus genome by short-read alignment. The percentage of samples that were positive is plotted on the y-axis and human body sites and other sample types are shown on the x-axis. The total number of samples analyzed in each category is annotated above and the total number of positive samples is indicated within the bar. B-The clinical status breakdown, if available, of redondovirus-positive samples is shown. C-Reads from a subset of 20 datasets across nine body sites were analyzed for homology to 20 redondovirus genomes and 133 animal-cell DNA viruses from six viral families. The height of each column represents the total number of samples that had detections of single or multiple viral families (rows). The viral families included in the co-detections are depicted as filled dots connected with lines below. The size of the bars on the left represents the total number of samples in which that viral family was detected. Cases where redondoviruses were detected are indicated in blue. See also Supplemental Figure 5.2 and 5.3. IBD, inflammatory bowel disease

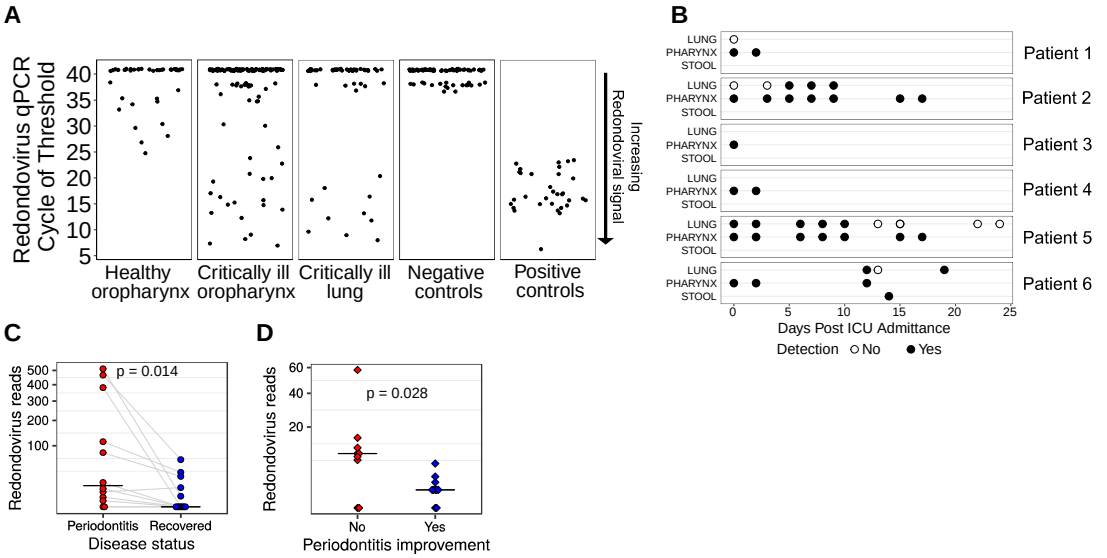


FIGURE 5.5: REDONDOVIRUSES IN THE ORO-RESPIRATORY TRACT IN HUMANS WITH CRITICAL ILLNESS AND PERIODONTITIS

A-Quantification of redondovirus genome sequences in post-SWGA DNA from oropharyngeal swabs (oropharynx) from 60 healthy volunteers, and oropharyngeal swabs and endotracheal aspirates (lung secretions) from 67 critically ill subjects. The average cycle of threshold (Ct) value of replicates is plotted on the y-axis. Samples with undetermined (i.e: no amplification) value in all 3 replicates are assigned an arbitrary value above the Ct value of the limit of resolution of the assay (37) which corresponds to 11 target copies per reaction. Samples below this value are counted as authentic detections. Negative controls included extraction blanks, reagent blanks and no template controls. Positive controls represent replicates of 10^4 copies of Human lung-associated brisavirus RC spiked into DNA extracted from a redondovirus-negative lung sample, subjected to SWGA, and assayed by qPCR.

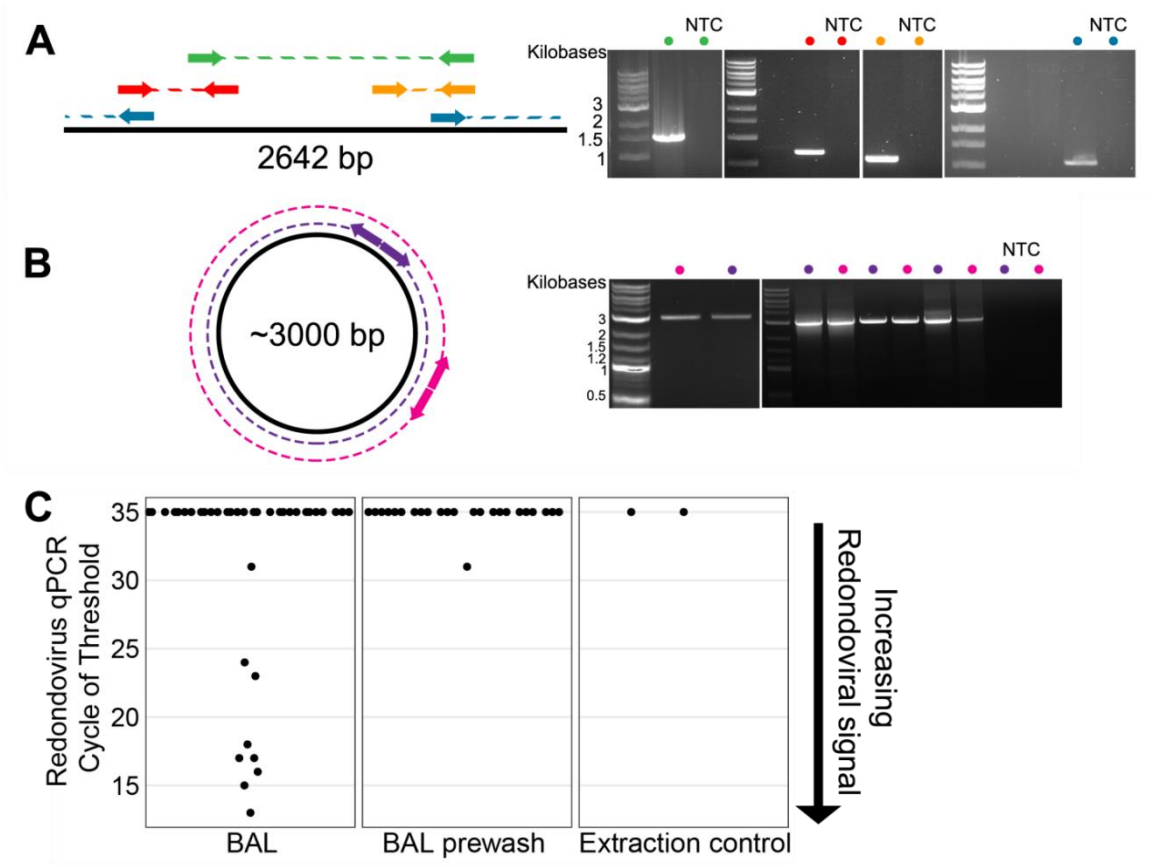
B-qPCR was used for redondovirus detection in respiratory and/or stool samples from 67 subjects in the medical intensive care unit (ICU). Six total subjects were positive for redondoviruses. The time point and type of sample surveyed for these six subjects is shown on the x- and y-axis, respectively. Positive samples are indicated by a filled-circle and negative samples by an open circle.

C-Number of reads mapping to a redondovirus in periodontitis samples from (Shi et al., 2015). Each point represents the average of all samples from a particular individual either before treatment (red) or after disease resolution (blue). Points from the same subject are connected by grey lines. The horizontal black line indicates the median. The Wilcoxon signed-rank test was used to test for paired differences between groups.

D-Each point represents the number of reads mapping to a redondovirus in samples from (Califf et al., 2017) from subjects with periodontitis whose disease either did (blue) or did not (red) improve during the study. The horizontal black line indicates the median. The Wilcoxon rank-sum test was used to test for differences between groups.

Feature	*Redondoviridae*	*Circoviridae*	*Nanoviridae*	*Geminiviridae*	*Genomoviridae*	*Smacoviridae*
Size (kb)	3.0-3.1	1.7-2.0	1.0 * 6 segments	2.5-3.0	2.1-2.2	2.6-2.9
ORFs	Cp, Rep, ORF3	Cp, Rep, ORF3/4	Cp, Rep, others	Cp, Rep, others	Cp, Rep	Cp, Rep
ORF orientation	Ambisense	Ambisense	Segmented	Ambisense (or segmented)	Ambisense	Ambisense
Origin sequence	TATTATTAT	TAGTATTAC	TATTATTAC	TAATATTAC	TAATATTAT	NAGTATTAC
Origin location	Noncoding (upstream) / in Rep	Noncoding (upstream) / in Rep	Noncoding (upstream)	Noncoding (upstream)	Noncoding (upstream)	Noncoding (downstream)

Table 5.1: Comparison of Genomic Features between Redondoviridae and Other CRESS-DNA Viruses



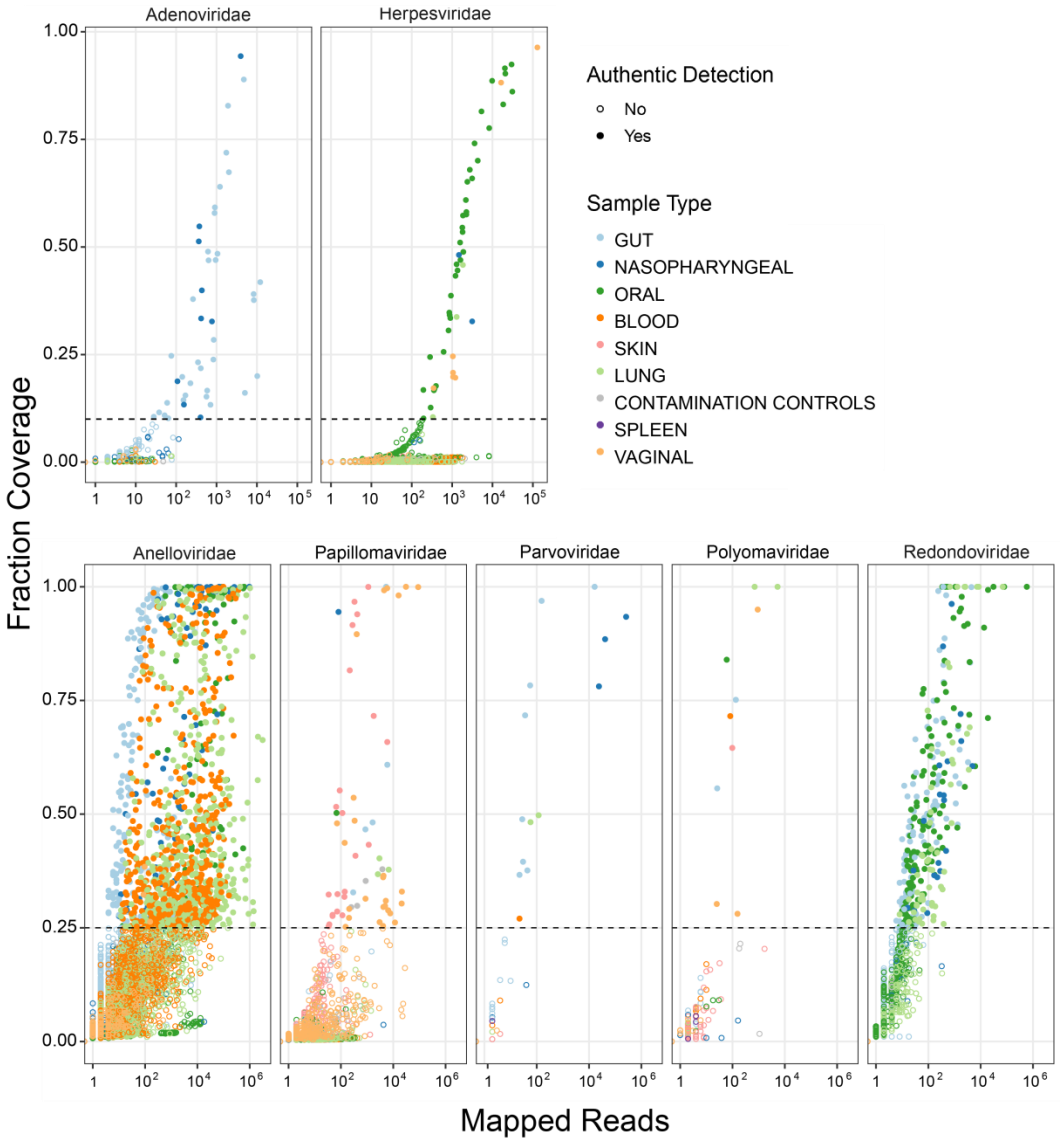
SUPPLEMENTAL FIGURE 5.1: PCR AMPLIFICATION AND DETECTION OF REDONDOVIRUS GENOMES, RELATED TO FIGURE 5.1

A-PCR amplicons of the expected size were observed from whole-genome amplified DNA from the BAL sample where Human lung-associated brisavirus RC was discovered. The outward facing primer set (blue) yielded a 600 bp product which was sequenced by the Sanger method and used to complete genome assembly.

B-Examples of approximately 3,000 bp products of two different outward facing sets of primers are shown. These represent, from left to right, the complete genomes of Human lung-associated vientovirus FB and Human lung-associated brisavirus MD, AA and II. DNA species visualized with ethidium bromide on a 1% agarose gel are shown.

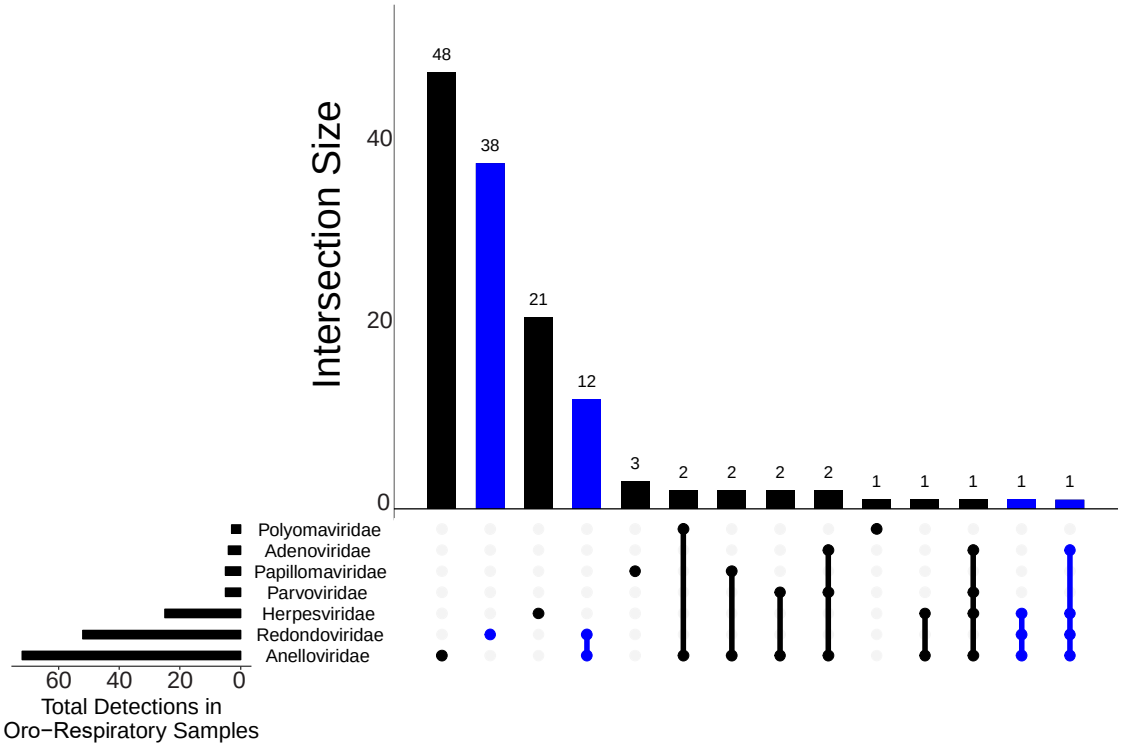
C-qPCR, performed in triplicate, was used to detect redondovirus sequences in acellular human bronchoalveolar lavage (BAL) samples after multiple displacement amplification. The average Ct value of replicates with any detection of redondoviruses is plotted on the y-axis. Samples with undetermined (i.e: no amplification signal) values in all three replicates are plotted at an arbitrarily high value of 35. The cycle of threshold value of the limit of quantification of the assay was 31, corresponding to 75 target copies per reaction. Samples falling below this value were counted as authentic detections. Sample types surveyed included BAL from organ donors, lung transplant recipients, and individuals with various lung diseases. BAL prewash samples represent the saline solution passed through the bronchoscope before insertion into patient. The BAL prewash point near the limit of detection represents a single replicate, with the other two replicates yielding values below

the limit of detection. Extraction controls represent sterile water processed through DNA extraction kits. NTC; no template control



SUPPLEMENTAL FIGURE 5.2: SUMMARY OF READ ALIGNMENTS TO NOVEL AND KNOWN HUMAN DNA VIRUSES, RELATED TO FIGURE 5.4

Reads from 2,675 human samples were aligned to 20 Redondovirus genomes and 133 human DNA viral genomes from six families from the NCBI RefSeq database. The log10 number of mapped reads (x-axis) and the fraction of target genome covered (y-axis) for each sample are plotted. Sample types are colored based on human body site or control. Each panel represents detection of a separate viral family. The dotted line represents the empirically determined threshold, based on examining depth and breadth of aligning reads across viral genomes, for calling a positive detection of a given viral family.



SUPPLEMENTAL FIGURE 5.3: PREVALENCE OF HUMAN DNA VIRUS FAMILIES IN LUNG SAMPLES, RELATED TO FIGURE 5.4

Reads from 13 metagenomic datasets encompassing 916 oro-respiratory sample types (lung, oral, or nasopharyngeal) were aligned to 20 redondovirus genomes and 133 animal-cell DNA viruses from six viral families. A positive hit was determined as described previously. The height of each column represents the total number of samples that had detections of single or multiple viral families (rows). The viral families included in the co-detections are depicted as filled dots connected with lines below. The size of the bars on the left represents the total number of samples in which that viral family was detected. Cases where redondoviruses were detected are indicated in blue.

Identifier	Sequence (5'-3')	Description
LJT-011	CCTTTGGTCTCGAAATCTTCCTATACTGG	Redondovirus whole genome amplification F, set A (3bp overlap with LJT-035)
LJT-035	AGGCCTCTCTCCCTTCCATTTGG	Redondovirus whole genome amplification R, set A (3bp overlap with LJT-011)
LJT-036	GGTTATCGTTCATTTGATCATGCATTAGTACC	Redondovirus whole genome amplification F, set B (3bp overlap with LJT-037)
LJT-037	ACCAAGATGTTTAAGCCCTTTAGTTAATGTTTC	Redondovirus whole genome amplification R, set B (3bp overlap with LJT-036)
LJT-001	TTCACTTAAATGGTTTGATGGCTACTGCG	Redondovirus Sanger sequencing primer
LJT-002	CCAACTGGAGTAACTTTACTACAAACCATAGAAG	Redondovirus Sanger sequencing primer
LJT-003	GACTTGCTTCTATGGTTTGTAGTAAAGTTACTCC	Redondovirus Sanger sequencing primer
LJT-004	GTGACGCATAAATTTGCATTTTTTGACCGC	Redondovirus Sanger sequencing primer
LJT-005	GGCTATAATCCCATACTTACGCCGG	Redondovirus Sanger sequencing primer
LJT-006	CGTAGATCATCAAGTACTGCTATTTCTTGGC	Redondovirus Sanger sequencing primer
LJT-007	GATGAAACTGCTCTATTCCTTGGATGCC	Redondovirus Sanger sequencing primer
LJT-008	GACTTGCTTCTATGGTTTGTAGTAAAGTTACTCC	Redondovirus Sanger sequencing primer
LJT-009	GGAAATTTTTAACATCATTCTCAGGAAGATGGTAACC	Redondovirus Sanger sequencing primer
LJT-010	ATTTATGCGTCACTTTATTCCAAATTTAAATATGTTGG	Redondovirus Sanger sequencing primer
LJT-012	CTACTCCAGAGCAAGCTTTCAAGTGG	Redondovirus Sanger sequencing primer
LJT-013	TTAAGAATTCCTCTTCTGGTAGCCCTAGG	Redondovirus Sanger sequencing primer
LJT-014	AGGCGAAGGGAAGACAAAAACAGC	Redondovirus Sanger sequencing primer
LJT-015	CGAGCTCATCCAGTAGCTGTTATCG	Redondovirus Sanger sequencing primer
LJT-016	GTTGATAAACCAGTAACACCTGTTTCTGAAGG	Redondovirus Sanger sequencing primer
LJT-017	AAGCGTCTCGTTAATTGTACTTGCTTATCC	Redondovirus Sanger sequencing primer
LJT-018	ATAATCCGTGCAGAGAAATTGAGAAACATATTCC	Redondovirus Sanger sequencing primer
LJT-019	GATCAATTTCTTTTGGTATTTGTCAGCGGG	Redondovirus Sanger sequencing primer
LJT-020	CTTTCCAATGGGTGAACAAGGAAGG	Redondovirus Sanger sequencing primer
LJT-021	CGAAGAGCTAGAGGATCCCCGGGTACC	SapI/NruI addition for virus cloning into pUC18 F (*comp Tm: 57.7, overall Tm 64.7)
LJT-022	CGAAGAGCGAGTCGACCTGCAGGCATG	SapI/NruI addition for virus cloning into pUC18 R (*comp Tm: 58.5, overall Tm 66.2)
LJT-046	GTCGACTCAAAAGCTCTTCGCCTTTGGTCTCGAAATCTTCCTATACTGG	Gibson assembly primers for amplifying vector, Set A-R (complements LJT-011)
LJT-047	GGGGATCCTCTAGCTCTTCGAGGCCTCTCTCCCTTCCATTTGG	Gibson assembly primers for amplifying vector, Set A-F (complements LJT-035)
LJT-048	GTCGACTCAAAAGCTCTTCGGGTTATCGTTCATTTGATCATGCATTAGTACC	Gibson assembly primers for amplifying vector, Set B-R (complements LJT-036)
LJT-049	GGGGATCCTCTAGCTCTTCGACCAAGATGTTTAAGCCCTTTAGTTAATGTTTC	Gibson assembly primers for amplifying vector, Set B-F (complements on LJT-037)
CircoqPCR-Fwd	GGATGCCATGAAACTTTGATAC	Redondovirus qPCR forward primer

CircoqPCR-Rev	TCTTCCTCCTTATTTGTATGGC	Redondovirus qPCR reverse primer
CircoqPCR-Probe	CCCATACTTACGCCGGTTACCTGC	Redondovirus qPCR TaqMan probe
Pan-HCRV-AA-Fwd	GCAGAGTTGTCAGCACATTT	Redondovirus qPCR forward primer
Pan-HCRV-AA-Rev	ATACCAGTATAGGAAGATTTCGAG	Redondovirus qPCR reverse primer
Pan-HCRV-AA-Probe	AGGGCTGCTAGGAATTATTCAAAAGTCAAGAAGATTAGAAGG	Redondovirus qPCR TaqMan probe
D13017_4-IntF-229	GATGCTCCACTAGTAATCTG	Redondovirus Sanger sequencing primer
D13017_4-IntR-1036	GCCTTTGGTCTCGAAATC	Redondovirus Sanger sequencing primer
D13017_4-IntF-1708	CGGTAAATCCCATGTCTTTG	Redondovirus Sanger sequencing primer
D13017_4-IntR-2419	GGCTGCTAGGAATTATTCG	Redondovirus Sanger sequencing primer
D13017_4-IntF-849	CGTCCGTATAGAGTAAGACTG	Redondovirus Sanger sequencing primer
D13017_4-Ext-359	TAGCCGTCAAACCATTTCAG	Redondovirus Sanger sequencing primer
D13017_4-Ext-2341	AACACCTTTGCCAGACATTG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntF-200	CGTACATCTTGCTCAGGAATAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntR-1036	ACTTCTGCGAAACTTCCATC	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntF-797	AGAACTGGCAGAACGATTAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntR-1345	TAGCCGTCGAACCATTTAAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntF-1891	CGTCGTCCGAATTAAAGTAAGG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntR-3032	TCATGGCATCCAAGGAATAGAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntF-1155	ATCCGTGCAGAGAAATTGAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-IntR-2036	ACAGGTGTCACTGGTTTATC	Redondovirus Sanger sequencing primer
13-28-contig-100_967-Ext-319	GGCTAAGCTCACAATATCAG	Redondovirus Sanger sequencing primer
13-28-contig-100_967-Ext-2919	ATCTGAAGTTGGGTCTGTTG	Redondovirus Sanger sequencing primer
HCRVswga1	TACGAATATTA	Redondovirus SWGA primer
HCRVswga2	TATCGTAATAT	Redondovirus SWGA primer
HCRVswga3	GTAATAATCTAT	Redondovirus SWGA primer

HCRVswga4	ATTATAATACG	Redondovirus SWGA primer
HCRVswga5	TATTACGATAA	Redondovirus SWGA primer
HCRVswga6	TAATAATACTAG	Redondovirus SWGA primer
HCRVswga7	TAGTATAACTC	Redondovirus SWGA primer
HCRVswga8	TTATCGTAATA	Redondovirus SWGA primer
HCRVswga9	ATATTACGATA	Redondovirus SWGA primer
HCRVswga10	GAGTTATACTA	Redondovirus SWGA primer
HCRVswga11	CAATATTACG	Redondovirus SWGA primer
HCRVswga12	CGTAATATTG	Redondovirus SWGA primer
HCRVswga13	ATTAGTATTATG	Redondovirus SWGA primer
HCRVswga14	ATATTATTGTAG	Redondovirus SWGA primer
HCRVswga15	CTACAATAATAT	Redondovirus SWGA primer
HCRVswga16	CTAGTATTATTA	Redondovirus SWGA primer
HCRVswga17	GTATTATTAGAA	Redondovirus SWGA primer
HCRVswga18	CATAATACTAAT	Redondovirus SWGA primer
HCRVswga19	TAATATTCGTA	Redondovirus SWGA primer
HCRVswga20	GTTATTATATTG	Redondovirus SWGA primer

Supplemental Table 5.1: List of Primers Used, Related to Figure 5.1, Supplemental Figure 5.1 and Figure 5.5

Genus	Isolate	Accession #	Host Body Site	Host species	Country
Brisavirus	Human lung-associated brisavirus AA	MK059754	Lung	Homo sapiens	USA
Brisavirus	Human lung-associated brisavirus II	MK059755	Lung	Homo sapiens	USA
Brisavirus	Human lung-associated brisavirus MD	MK059756	Lung	Homo sapiens	USA
Brisavirus	Human lung-associated brisavirus RC	MK059757	Lung	Homo sapiens	USA
Brisavirus	Human oral-associated brisavirus YH	MK059758	Oral	Homo sapiens	USA
Brisavirus	Human respiratory-associated PoSCV5-like virus	NC_023878	Oral	Homo sapiens	China
Brisavirus	Human gut-associated brisavirus VW	MK059759	Gut	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus AL	MK059760	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus DC	MK059761	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus ES	MK059762	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus FB	MK059763	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus JB	MK059764	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus JY	MK059765	Lung	Homo sapiens	USA
Vientovirus	Human lung-associated vientovirus LT	MK059766	Lung	Homo sapiens	USA
Vientovirus	Human oral-associated vientovirus AV	MK059767	Oral	Homo sapiens	Spain
Vientovirus	Human oral-associated vientovirus EC	MK059768	Oral	Homo sapiens	USA
Vientovirus	Human oral-associated vientovirus LZ	MK059769	Oral	Homo sapiens	USA
Vientovirus	Human oral-associated vientovirus MC	MK059770	Oral	Homo sapiens	USA
Vientovirus	Human oral-associated vientovirus XM	MK059771	Oral	Homo sapiens	China
Vientovirus	Human gut-associated vientovirus MW	MK059772	Gut	Homo sapiens	UK

Supplemental Table 5.2: Summary of Taxa within Redondoviridae, Related to Figure 5.2

Based on definitions in (Varsani and Krupovic, 2018) and the analysis of the diversity of viral Rep proteins, *Redondoviridae* genomes can be grouped into two genera, demarcated by 50% Rep protein identity, which we propose be called Vientovirus and Brisavirus, from the Spanish words for “wind” and “breeze”, alluding to their discovery in the respiratory tract.

Title	Samples	Database Source	Accession Number
Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach.	8	SRA	SRP000287
Viruses in the faecal microbiota of monozygotic twins and their mothers.	8	SRA	SRP002523
Diversity and abundance of single-stranded DNA viruses in human feces.	6	SRA	SRP005097
Hypervariable loci in the human gut virome.	12	SRA	SRP094782
Metagenomic exploration of viruses throughout the Indian Ocean.	26	SRA	SRP003580
Symptomatic atherosclerosis is associated with an altered gut metagenome.	27	SRA	SRP016067
Identification of a novel human papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness.	8	SRA	SRP012061
Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool.	14	SRA	SRA054231
Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease.	28	SRA	SRP033553
Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach.	19	SRA	ERP001706
Rapid evolution of the human gut virome.	27	SRA	SRP021107
Metagenomic analysis of tuberculosis in a mummy.	1	SRA	SRP018736
Temporal response of the human virome to immunosuppression and antiviral therapy.	153	SRA	SRP032345
Human oral viruses are personal, persistent and gender-consistent.	121	SRA	SRP033575
Changes in abundance of oral microbiota associated with oral cancer.	12	SRA	ERP004294
The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease.	320	SRA	SRP067755
Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients.	22	SRA	SRP098739
Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses.	1	SRA	SRP047227
Disease-specific alterations in the enteric virome in inflammatory bowel disease.	171	SRA	ERP008725
Gut virome sequencing in children with early islet autoimmunity.	96	SRA	SRP059576
Dynamic changes in the subgingival microbiome and their potential for diagnosis and prognosis of periodontitis.	48	SRA	SRP052958
The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment.	530	SRA	ERP006678
Early life dynamics of the human gut virome and bacterial microbiome in infants.	143	SRA	SRP058399
Direct sequencing of human gut virome fractions obtained by flow cytometry.	2	SRA	ERP007252
Virome analysis of antiretroviral-treated HIV patients shows no correlation between T-cell activation and anelloviruses levels.	2	SRA	SRP059897
The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome.	675	SRA	SRP049645
Metagenomic analysis of viral genetic diversity in respiratory samples from children with severe acute respiratory infection in China.	4	SRA	SRP058055
RNA-sequencing study of peripheral blood monocytes in chronic periodontitis.	10	SRA	SRP047232
Metagenomic Sequencing of the Chronic Obstructive Pulmonary Disease Upper Bronchial Tract Microbiome Reveals Functional Changes Associated with Disease Severity.	18	SRA	ERP010088
Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses.	60	SRA	SRP066046
Viral Outbreak in Corals Associated with an In Situ Bleaching Event: Atypical Herpes-Like Viruses and a New Megavirus Infecting Symbiodinium.	1	SRA	ERP013548
Preliminary analysis of salivary microbiome and their potential roles in oral lichen planus.	53	SRA	SRP067603
Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome.	66	SRA	ERP010635

Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation.	18	SRA	SRP073516
Proteomic Characterization of Middle Ear Fluid Confirms Neutrophil Extracellular Traps as a Predominant Innate Immune Response in Chronic Otitis Media.	176	SRA	SRP069302
Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology.	14	SRA	SRP045601
Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans.	785	SRA	SRP090628
Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system.	11	SRA	SRP071354
Microbial diversity in individuals and their household contacts following typical antibiotic courses.	192	SRA	SRP077685
Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens.	8	SRA	SRP078679
Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype.	2	SRA	SRP065327
Healthy human gut phageome.	4	SRA	SRP090453
Identification of Viruses in Cases of Pediatric Acute Encephalitis and Encephalopathy Using Next-Generation Sequencing.	3	SRA	DRP003259
The Perioperative Lung Transplant Virome: Torque Teno Viruses Are Elevated in Donor Lungs and Show Divergent Dynamics in Primary Graft Dysfunction.	185	SRA	SRP109620
Viral Metagenomic Analysis Displays the Co-Infection Situation in Healthy and PMWS Affected Pigs.	8	SRA	SRP076383
Transmission of viruses via our microbiomes.	192	SRA	SRP077685
The balance of metagenomic elements shapes the skin microbiome in acne and health.	78	SRA	SRP101642
MetaSort untangles metagenome assembly by reducing microbial community complexity.	3	SRA	SRP095074
Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery.	189	SRA	SRP078001
Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling.	28	SRA	SRP082656
Virome Assembly and Annotation: A Surprise in the Namib Desert.	1	SRA	ERP015045
Circadian oscillations of microbial and functional composition in the human salivary microbiome.	36	SRA	DRP003804
A metagenomics study for the identification of respiratory viruses in mixed clinical specimens: an application of the iterative mapping approach.	2	SRA	SRP100814
A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer.	10	SRA	ERP010087
Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children.	169	SRA	SRP107965
A Different Microbiome Gene Repertoire in the Airways of Cystic Fibrosis Patients with Severe Lung Disease.	12	SRA	SRP072212
Virome comparisons in wild-diseased and healthy captive giant pandas.	1	SRA	SRP108570
Microbial Lineages in Sarcoidosis. A Metagenomic Analysis Tailored for Low-Microbial Content Samples.	109	SRA	SRP110811
Virome analysis for identification of novel mammalian viruses in bats from Southeast China.	12	SRA	SRP102052
Host Genetic Control of the Oral Microbiome in Health and Disease.	88	SRA	SRP104798
Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters.	8	SRA	SRP092902
Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota in patients with periodontitis and dental caries.	60	SRA	SRP114751
Identification of sapovirus GV.2, astrovirus VA3 and novel anelloviruses in serum from patients with acute hepatitis of unknown aetiology.	9	SRA	SRP102763
Temporal dynamics of uncultured viruses: a new dimension in viral diversity.	11	SRA	SRP091978
Sera of Peruvians with fever of unknown origins include viral nucleic acids from non-vertebrate hosts.	2	SRA	PRJNA382858
Discovering viral genomes in human metagenomic data by predicting unknown protein families.	17	SRA	ERP019738
Plasma virome of cattle from forest region revealed diverse small circular ssDNA viral genomes.	1	SRA	SRX3235902

The analysis of the oral DNA virome reveals which viruses are widespread and rare among healthy young adults in Valencia (Spain).	72	SRA	ERP104599
Comprehensive virome analysis reveals the complexity and diversity of the viral spectrum in pediatric patients diagnosed with severe and mild hand-foot-and-mouth disease.	1	SRA	SRP090926
Selective maternal seeding and environment shape the human gut microbiome.	216	SRA	SRP100409
Short term dynamics of the sputum microbiome among COPD patients.	26	SRA	SRP124904
Metagenomic data of DNA viruses of poultry affected with respiratory tract infection.	16	SRA	SRP075600
Antibiotic Treatment Leads to Fecal *Escherichia coli* and Coliphage Expansion in Severely Malnourished Diarrhea Patients.	27	SRA	SRP100895
The vaginal eukaryotic DNA virome and preterm birth.	128	SRA	SRP068239
Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community.	36	SRA	SRP128680
Viromes of one year old infants reveal the impact of birth mode on microbiome diversity.	40	SRA	SRP106048
Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses.	13	SRA	SRP119893
Temporal dynamics of the lung and plasma viromes in lung transplant recipients.	10	SRA	ERP107081
Bidirectional transfer of Anelloviridae lineages between graft and host during lung transplantation.	120	SRA	SRP125483
metagenomic sequencing of human nasal swab from acute respiratory infection	1	SRA	SRP062772
WGS of viruses: adult saliva	16	SRA	SRP074878
Perioperative nasal microbiome	54	SRA	SRP089889
Culture of human cervical microbiota	19	SRA	SRP097289
metagenome: family 190LO infant 1 month	96	SRA	SRP109960
DNA sequencing of circular virome of Seabass tissue	73	SRA	SRP112556
Environmental sampling of the METATRANSCRIPTOMIC for Subject	40	SRA	SRP116887
Vaginal virome of patient with Bacterial Vaginosis_day 0 after Clindamycin uptake	4	SRA	SRP126271
MiSeq of human metagenome:Blood from W07	125	SRA	SRP128102
Cafe and FSM: Virome of human gut	9	SRA	SRX020505
Bronchoalveolar lavage samples from idiopathic pneumonia syndrome patients and controls	36	SRA	SRP162048
Altered oral viral ecology in association with periodontal disease.	91	MG-RAST	mgp7236
Multi-omics Analysis of Periodontal Pocket Microbial Communities Pre- and Posttreatment.	23	MG-RAST	mgp21311
Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis.	43	HOMD	20141024
Metagenomic analysis uncovers strong relationship between periodontal pathogens and vascular dysfunction in American Indian population	22	MG-RAST	mgp15104
Healthy adult serum	28	Unpublished (Internal)	NA
Infant fecal virome	156	Unpublished (Internal)	NA
Medical intensive care unit	621	Unpublished (Internal)	NA

Supplemental Table 5.3: List of Studies Queried, Related to Figure 5.1 and Figure 5.4
The number of samples in each study represents the total samples uploaded to the central database. However, only samples that were determined not to be 16S, ITS or other targeted amplicons were analyzed by the pipeline described here. SRA; Sequence Read Archive, MG-RAST; Metagenomic Rapid Annotations using Subsystems Technology, HOMD, Human Oral Microbiome Database

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

Overall, this work has increased our understanding of the human lung viral microbiome, particularly in the states where the immune system is perturbed. While metagenomics has revolutionized studies of viral communities and their relationship to disease states in gastrointestinal, dermatological and oral body sites, its application in hard-to-access body sites has been more limited. Balancing the ability of metagenomic approaches to retrieve rare viral sequences with their increased sensitivity to contamination in low biomass samples is necessary to draw meaningful conclusions about viral entities actually present versus false positives. Careful quantitative approaches can reveal the complex behavior of viral communities within a human host over time. Viruses within body sites are embedded in a different community and environment than those circulating in the blood. Thus, site-specific studies of viral communities can reveal insights on viral behavior and impact on the host (Virgin, 2014, Cadwell, 2015).

6.1 ANELLOVIRUS DYNAMICS IN PERIOPERATIVE PERIOD ARE ASSOCIATED WITH PRIMARY GRAFT DYSFUNCTION

In CHAPTER 3, we first used an unbiased metagenomic sequencing approach to survey the viral microbiomes in a cross sectional study of lung donors and recipients. Previous studies had only analyzed post-transplantation time points and most had only surveyed for viruses in blood. Here, we took advantage of having paired samples of the donor organ and allograft recipient to interrogate viral dynamics before and immediately after organ transplantation. We found that in LTRs, the lung virome consists of several families of bacteriophage and is dominated by a single family of “benign” eukaryotic DNA

viruses (*Anelloviridae*). Upon this discovery, we followed up with a targeted molecular assay to quantify anellovirus levels in the lungs and blood, as extrapolating absolute abundance from viral shotgun metagenomic data has major caveats such as enriching for small circular DNA with MDA. We demonstrate that anellovirus levels are increased in the lung both before and after transplantation but were not associated with donor cause of death or various clinical variables regarding organ management. Anellovirus expansion in the donor organ followed donor brainstem death and organ reperfusion, which both trigger inflammatory cascades (Faropoulos and Apostolakis, 2009). Therefore, potential organ donors are administered corticosteroids to promote organ preservation (Kutsogiannis et al., 2006) and lung transplant recipients are given induction immune suppression. Thus, we hypothesize that the dramatic increase in anelloviruses in the lung during this relatively short period is a combination of complex immune activation promoting viral replication in target cells and immune suppression failing to control replication and spread.

We also showed that viral dynamics in the lung at the time of transplantation are associated with PGD. Specifically PGD was found to be associated with a significantly lower rise in anellovirus levels in the lung after transplantation. Our findings suggest that anelloviruses may reflect or impact local innate immune function in the lung that is responsible for PGD pathology (Diamond and Wigfield, 2013). Quantifying anellovirus burden, especially in a non-invasive manner, to predict organ transplantation outcomes is appealing since patients vary in their sensitivity to immune suppression (Wieland et al., 2010) and this unpredictability impedes efforts to balance the risk of organ rejection and infection.

Several groups have investigated the utility of using anelloviruses as biomarkers. For example, in a study of solid organ transplant recipients (heart or lung), Quake et al., 2015, compare time-normalized blood anellovirus levels in organ-rejecting and non-

rejecting individuals and show that lower anellovirus burden is associated with organ rejection. They then tested the performance of anellovirus levels in classifying rejecting and non-rejecting subjects with positive, albeit modest, results. Another study of pediatric lung transplantation specifically asked whether levels of two genera within the *Anelloviridae* family (Alphtorquevirus and Betatorquevirus) were associated with either acute rejection (Blatter et al., 2018). Again, low anellovirus levels (specifically of alphatorqueviruses) early in transplantation, possibly indicative of ineffective immune suppression, were associated with acute rejection. Finally, both Blatter et al., 2017 and Gorzer et al., 2017 show that long-term graft dysfunction or failure occur significantly more frequently in LTRs with lower anellovirus loads. Taken together with our findings in CHAPTER 3, it's possible that surveillance of site-specific of all or specific genera of anelloviruses could inform management of immunosuppressive therapy for organ transplant recipients and other scenarios where exogenous immune modulation is needed.

6.2 FUTURE DIRECTION: QUERYING MULTIDIMENSIONAL LUNG MICROBIAL COMMUNITIES IN LUNG TRANSPLANTATION OUTCOMES

Here, we investigated whether a viral feature of the lung microbiota was associated with PGD. Others have investigated bacterial features. For example, a recent retrospective study using culture-based techniques showed that LTRs who received a donor organ positive for potentially pathogenic bacteria required longer duration of mechanical ventilation after transplantation (Ahmad et al., 2018). While no association with donor bacterial culture and PGD was observed in this small cohort (n=32), given the decreased sensitivity of culture compared to molecular techniques and a lack of interrogation of viruses and fungi, a role for the microbiome in PGD cannot yet be ruled out. Indeed, in our metagenomic approach, we discovered that donor lungs commonly have sequences of bacteriophages that infect *Staphylococcus, Streptococcus,* and *Haemophilus influenza* species.

Bacteriophages indirectly mark the presence of their bacterial hosts. We thus speculate that the lung microbiome of LTRs who develop PGD may be defined by a general increase in microbial burden, which could stimulate innate pathways detecting conserved microbial elements such as cell wall components and bacterial DNA. Alternatively, there may be increased abundance of certain species with virulence factors that impact tissue viability. This microbial burden would likely be bacterial and/or fungal in origin as our work did not discover enrichment of pathogenic eukaryotic viruses in PGD compared to non-PGD controls. Targeted qPCR for candidate taxa discovered by sequencing, as was done for anelloviruses, can more accurately enumerate microbial burden in the lung. The inflammatory response in the lower respiratory tract characteristic of PGD could be triggered by either live or dead microorganisms that were transferred by aspiration from the upper respiratory tract. In such a scenario, we might expect increased abundance of upper respiratory tract flora in the lungs of PGD cases compared to non-PGD controls. The rate and extent of immune activity of both epithelial and immune cells within the lung may then impact anellovirus replication, explaining the phenomenon we describe in CHAPTER 3.

To build a complete picture of the role of the lung microbiome in other clinical outcomes after transplantation, larger, prospective longitudinal studies are needed. To date, there have been no studies of LTRs integrating analyses of viruses, bacteria and fungi. Such an approach is important because disease pathogenesis may be multifactorial and not easily prescribed to a single etiological agent. Since it is now well understood that the microbiome is altered in LTRs (Cribbs and Beck, 2017, Becker et al., 2014), specific features of community composition (microbial burden, structure, presence of specific taxa) should be assessed through systematic and stringent culture-independent methods in settings of important clinical outcomes.

6.3 BIDIRECTIONAL TRANSFER OF ANELLOVIRUSES BETWEEN GRAFT AND HOST DURING LUNG TRANSPLANTATION

To study extent and persistence of viral transmission in solid organ transplantation, including viruses not yet linked to any disease, we used metagenomic approaches to define the lung and blood viromes in a longitudinal study of 13 donor-organ recipient pairs (described in CHAPTER 4). Based on findings from previous studies from our group (Young et al., 2015, Abbas et al., 2017) and others (De Vlaminck et al., 2013, Görzer et al., 2014, Görzer et al., 2015), we focused on the behavior of anelloviruses, which are known to be abundant in both lung and blood of individuals receiving organ transplants. After assembling full and partial viral genomes from metagenomic reads, we tracked individual lineages of anelloviruses by scoring their abundance within an individual over time. This approach allowed us to account for the high intra-individual diversity of viruses and population mixing that took place after organ transplantation.

We specifically highlight the potential of the virome derived from the organ donor lung to persist within the immunosuppressed host. This insight into the transmissibility of the lung virome establishes a foundation for studying the donor virome's association with clinical outcomes in LTRs and in other lung disease settings. In addition, complex populations of anelloviruses already present in an individual prior to lung transplantation were observed to repopulate the allograft. Our findings corroborate and expand upon those recently reported in a smaller, independent cohort of LTRs (Segura-Wang et al., 2018). In this study, anellovirus populations in the lung and blood were followed starting at 30-90 days post-transplantation in seven individuals. To overcome the same challenge we encountered in discriminating between closely related anellovirus strains, this group calculated normalized abundance after mapping reads to reference anellovirus genomes. Despite the slight differences in statistical approach, similar conclusions were reached; namely that anelloviruses are the most common eukaryotic virus detected in the lung and

blood, that individuals harbor diverse groups of these viruses, a wide variety of viral dynamics are observed post-transplantation and many strains detected in the lung are also detectable in the blood. A primary future direction is to use these findings as the foundation for understanding the genomic variability of anellovirus populations at different body sites.

6.4 FUTURE DIRECTION: DEFINING THE HOST CELLULAR IMMUNE RESPONSE TO ANELLOVIRUSES

Given the ubiquity of anelloviruses in the healthy adult population and their consistent expansion in immunosuppressed organ transplant populations, it is surprising that so little is known about the molecular immunology of these small DNA viruses. This is likely due to the difficulty in studying chronic infection with diverse viral quasi-species (Nishizawa et al., 1999, Madsen et al., 2002, Segura-Wang et al., 2018, Li et al., 2015b, Lim et al., 2015, Clarke et al., 2018a). Notably, studies examining longitudinal anellovirus dynamics (including diversity and absolute levels) have done so in adults with co-morbidities or in developing infants, leaving a paucity of information about infection in asymptomatic, immune-competent adults. Fewer studies have characterized the concurrent host immune state. Thus far, there is no evidence of a correlation between anellovirus levels and T-cell activation in HIV-infected individuals (Li et al., 2015b) or in hemodialized patients (Fodor et al., 2002), although serum levels of certain cytokines (IFN-γ, TNF-α, FGF-basic and MCP-3) were found to be associated with anellovirus levels in children receiving bone marrow transplants (Zanotta et al., 2015). Additionally, the adaptive cellular immune response to anelloviruses is understudied, despite it likely playing a role in keeping chronic and recurrent infections in check.

Therefore, we propose that future work should determine whether targets against these small DNA viruses occupy a substantial portion of the human T-cell repertoire. We hypothesize that ORF1, which encodes the putative capsid, is the most likely target for

cellular immune responses, although the degree of homology across the 600-800 amino acid ORF is exceedingly low. This precludes construction of peptide libraries of a consensus sequence for immune stimulation as these epitopes likely would not represent a substantial proportion of viral variants within an individual. Developing a novel, personalized approach to assess T-cell immune responses to hypervariable viruses might thus entail first defining an individual's viral repertoire by sequencing clones obtained by limiting dilution amplification. This can be followed with transfection of autologous ORF1 mRNA or peptides into lymphocytes which can then carry out antigen presentation. This allows enumeration and characterization of ORF1-antigen-specific responder cells using established immunological assays such as ELISPOT and intracellular cytokine staining. Addressing the major gap in knowledge on the type and extent of the host immune response to anelloviruses in healthy humans would establish a baseline to compare how these responses are altered or absent in LTRs.

6.5 DISCOVERY OF A NEW FAMILY OF CIRCULAR HUMAN DNA VIRUSES THROUGH METAGENOMIC DATA MINING

Studying viruses, the most abundant biological entities on the planet, by DNA sequencing is difficult due to the great sequence diversity present in global viral communities and because no common gene or sequence can be used to identify all of them. In CHAPTER 5, we characterize a family of highly divergent sequences originally found in samples of the lower respiratory tract of organ donors and lung transplant recipients. These samples were processed using multiple displacement amplification (MDA), a method to amplify small amounts of nucleic acids in low-biomass samples. MDA, in conjunction with metagenomic sequencing, has facilitated the discovery of many new, circular ssDNA viral sequences. After observing an initial alignment of many reads from two BAL samples to a limited portion of an experimentally uncharacterized circovirus, we asked whether these

sequences might actually originate from a different genome, as circoviruses are not known to infect humans. De novo assembly of reads yielded two circular genomes which were used to further identify homologous sequences in other human lung samples processed by our group that may have been missed by classification as they lacked similarity to reference genomes. Ultimately, we identified and cloned genomes from seven independent BAL samples. Comparison of the genomic architecture and protein sequences of these genomes revealed that they are most closely related to each other than to other known viral families (Figure 5.2). Thus, we propose that these divergent viral genomes form their own family termed *Redondoviridae*.

To determine the prevalence of redondoviruses in human and non-human samples, a bioinformatics pipeline was developed to query over 6000 samples from publicly available metagenomic datasets. We detected redondoviruses exclusively in human samples, primarily from the respiratory tract, and recovered 12 additional complete genomes. Additionally, during our analysis, a complete genome homologous to redondoviurses was reported in the oro-respiratory tract of a febrile individual by an independent group (Cui et al., 2017). This strengthens the idea that the redondoviral niche is the human respiratory tract. Intriguingly, redondoviruses were the most common human DNA virus detected in human oral samples, although many of these were processed by MDA prior to shotgun sequencing, which preferentially amplifies circular DNA. Importantly, these sequences were absent in contamination controls and non-human samples.

Taking advantage of the sizable amount of clinical metadata annotated in publicly available datasets, we investigated whether redondoviruses were correlated with any human disease. Analysis of redondovirus sequences in two independent studies revealed a positive association with periodontal disease. Next, we quantified redondoviruses using a targeted molecular assay in oral samples of healthy and critically ill individuals and found

that although prevalence was similar between the two groups (9-15%), relative levels during illness were higher and redondoviruses were detected longitudinally in several individuals. Thus we propose that redondoviruses are a novel group of viruses colonizing human oro-respiratory sites.

6.6 FUTURE DIRECTION: DETERMINING ROLE OF REDONDOVIRUSES IN RESPIRATORY TRACT HEALTH AND DISEASE

This work demonstrated the utility for shotgun metagenomics followed by experimental validation in discovering novel human viruses. An immediate future direction will be to explore and validate the association of redondovirus in periodontal disease. Given the established role of oral hygiene and bacterial infection in periodontitis, it is unlikely redondoviruses directly contribute to disease pathogenesis and may be a reflection of an inflammatory environment in the mouth. This hypothesis is consistent with the observation that redondoviruses are present in the upper and lower respiratory tract and stool of a heterogeneous population of critically ill subjects. Nonetheless, since these viruses are described here for the first time, it is imperative to ask whether redondoviral infection and host immune responses to them contribute to tissue injury in various oral and pulmonary diseases. Demonstrating adaptive (humoral and cellular) immune responses to these viruses would lend further evidence that they authentically infect humans and shed light on infection history of the general public and in specific disease cohorts.

6.7 FUTURE DIRECTION: DEVELOPING REDONDOVIRUS CULTURE SYSTEM; PREDICTIONS FOR REPLICATION STRATEGY AND HOST CELL TROPISM

Redondoviruses encode two known proteins; a capsid and Rep and also contain a highly conserved open-reading frame of unknown function. Based on this simple genome arrangement, which is similar to that of circoviruses, we predict that redondoviruses

replicate by rolling circle replication (RCR), initiated by the Rep and completed by host cell DNA polymerases. The distinct catalytic activities of Rep that are known to facilitate replication, stem-loop nicking and rejoining of genome ends, can be tested *in vitro* using purified protein and model DNA substrates. Understanding the requirements and mechanisms of redondovirus Rep activity can shed light on their replication strategy *in vivo* and be useful in developing drug targets if these viruses prove pathogenic in humans.

Thus far, we have not been able to recover redondovirus particles by transfecting cloned, circularized dsDNA genomes (a strategy that works for circoviruses) into a modest panel of adherent mammalian cell lines. Indeed, there has been historic difficulty in establishing *in vitro* systems to propagate small circular DNA viruses such as papillomaviruses and polyomaviruses. Redondoviruses may be restricted by cell-intrinsic immunity in which case permissive cell types should be tried. Other possibilities include using primary cells from the upper and lower respiratory tract or inoculating cells with viral particles isolated from human clinical samples. However, it does appear possible to isolate and propagate identified viruses originally identified only by metagenomic sequencing (Shkoporov et al., 2018). Concerted effort is thus needed to establish a robust replication system to answer many basic biological questions about these newly discovered viruses.

6.8 CONCLUDING REMARKS

The work described in this dissertation lays a foundation for using metagenomic approaches to ask diverse questions about the human viral microbiome. In the future, metagenomic sequence analysis may become a core diagnostic and clinical tool, as it can be highly sensitive and informative about the unique viral species present. In order for this to be a reality, concurrent investigations about the host and other microbial kingdoms are

needed to understand how the behavior of endogenous viral population reflects the holistic health of its host. Furthermore, large, well-curated viral metagenomic datasets can and should be exploited in ways beyond what they were originally conceived for. Doing so will slowly, but steadily, illuminate the vast world of omnipresent viruses and enhance our ability to interrogate their role in human health and disease.

REFERENCES

ABBAS, A. A., DIAMOND, J. M., CHEHOUD, C., CHANG, B., KOTZIN, J. J., YOUNG, J. C., IMAI, I., HAAS, A. R., CANTU, E., LEDERER, D. J., MEYER, K. C., MILEWSKI, R. K., OLTHOFF, K. M., SHAKED, A., CHRISTIE, J. D., BUSHMAN, F. D. & COLLMAN, R. G. 2017. The Perioperative Lung Transplant Virome: Torque Teno Viruses Are Elevated in Donor Lungs and Show Divergent Dynamics in Primary Graft Dysfunction. *Am J Transplant,* 17**,** 1313-1324.

ABBAS, A. A., YOUNG, J. C., CLARKE, E. L., DIAMOND, J. M., IMAI, I., HAAS, A. R., CANTU, E., LEDERER, D. J., MEYER, K., MILEWSKI, R. K., OLTHOFF, K. M., SHAKED, A., CHRISTIE, J. D., BUSHMAN, F. D. & COLLMAN, R. G. 2018. Bidirectional transfer of anelloviridae lineages between graft and host during lung transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*.

ABELES, S. R., LY, M., SANTIAGO-RODRIGUEZ, T. M. & PRIDE, D. T. 2015. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PloS one,* 10.

AGGARWALA, V., LIANG, G. & BUSHMAN, F. D. 2017. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mobile DNA,* 8**,** 12.

AHMAD, O., SHAFII, A. E., MANNINO, D. M., CHOATE, R. & BAZ, M. A. 2018. Impact of donor lung pathogenic bacteria on patient outcomes in the immediate post-transplant period. *Transplant infectious disease : an official journal of the Transplantation Society,* 20.

ALBERT, E., SOLANO, C., PASCUAL, T., TORRES, I., MACERA, L., FOCOSI, D., MAGGI, F., GIMÉNEZ, E., AMAT, P. & NAVARRO, D. 2017. Dynamics of Torque Teno virus plasma DNAemia in allogeneic stem cell transplant recipients. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 94**,** 22-28.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* 215**,** 403-410.

ANISIMOVA, M. & GASCUEL, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology,* 55**,** 539-552.

ATKINS, B. Z., TRACHTENBERG, M. S., PRINCE-PETERSEN, R., VESS, G., BUSH, E. L., BALSARA, K. R., LIN, S. S. & DAVIS, R. D. 2007. Assessing oropharyngeal dysphagia after lung transplantation: altered swallowing mechanisms and increased morbidity. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation,* 26**,** 1144-1148.

BANDO, M., NAKAYAMA, M., TAKAHASHI, M., HOSONO, T., MATO, N., YAMASAWA, H., OKAMOTO, H. & SUGIYAMA, Y. 2015. Serum torque teno virus DNA titer in idiopathic pulmonary fibrosis patients with acute respiratory worsening. *Internal medicine (Tokyo, Japan),* 54**,** 1015-1019.

BANDO, M., OHNO, S., OSHIKAWA, K., TAKAHASHI, M., OKAMOTO, H. & SUGIYAMA, Y. 2001. Infection of TT virus in patients with idiopathic pulmonary fibrosis. *Respiratory medicine,* 95**,** 935-942.

BANDO, M., TAKAHASHI, M., OHNO, S., HOSONO, T., HIRONAKA, M., OKAMOTO, H. & SUGIYAMA, Y. 2008. Torque teno virus DNA titre elevated in idiopathic pulmonary fibrosis with primary lung cancer. *Respirology (Carlton, Vic.),* 13**,** 263-269.

BECK, J. M., SCHLOSS, P. D., VENKATARAMAN, A., TWIGG, H., JABLONSKI, K. A., BUSHMAN, F. D., CAMPBELL, T. B., CHARLSON, E. S., COLLMAN, R. G., CROTHERS, K., CURTIS, J. L.,

DREWS, K. L., FLORES, S. C., FONTENOT, A. P., FOULKES, M. A., FRANK, I., GHEDIN, E., HUANG, L., LYNCH, S. V., MORRIS, A., PALMER, B. E., SCHMIDT, T. M., SODERGREN, E., WEINSTOCK, G. M., YOUNG, V. B. & PROJECT, L. 2015. Multicenter Comparison of Lung and Oral Microbiomes of HIV-infected and HIV-uninfected Individuals. *American journal of respiratory and critical care medicine,* 192**,** 1335-1344.

BECKER, J., POROYKO, V. & BHORADE, S. 2014. The lung microbiome after lung transplantation. *Expert review of respiratory medicine,* 8**,** 221-231.

BELKAID, Y. & HARRISON, O. J. 2017. Homeostatic Immunity and the Microbiota. *Immunity,* 46**,** 562-576.

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)***,** 289-300.

BERGALLO, M., COSTA, C., TERLIZZI, M. E., ASTEGIANO, S., CURTONI, A., SOLIDORO, P., DELSEDIME, L. & CAVALLO, R. 2010. Quantitative detection of the new polyomaviruses KI, WU and Merkel cell virus in transbronchial biopsies from lung transplant recipients. *Journal of clinical pathology,* 63**,** 722-725.

BHATTACHARYA, T., GHOSH, T. S. & MANDE, S. S. 2015. Global Profiling of Carbohydrate Active Enzymes in Human Gut Microbiome. *PloS one,* 10.

BHORADE, S. M. & STERN, E. 2009. Immunosuppression for Lung Transplantation. *Proceedings of the American Thoracic Society,* 6**,** 47-53.

BLATTER, J. A., SWEET, S. C., CONRAD, C., DANZIGER-ISAKOV, L. A., FARO, A., GOLDFARB, S. B., HAYES, D., MELICOFF, E., SCHECTER, M., STORCH, G., VISNER, G. A., WILLIAMS, N. M. & WANG, D. 2018. Anellovirus loads are associated with outcomes in pediatric lung transplantation. *Pediatric transplantation,* 22.

BOREWICZ, K., PRAGMAN, A. A., KIM, H. B., HERTZ, M., WENDT, C. & ISAACSON, R. E. 2013. Longitudinal analysis of the lung microbiome in lung transplantation. *FEMS microbiology letters,* 339**,** 57-65.

BOTHA, P., ARCHER, L., ANDERSON, R. L., LORDAN, J., DARK, J. H., CORRIS, P. A., GOULD, K. & FISHER, A. J. 2008. Pseudomonas aeruginosa colonization of the allograft after lung transplantation and the risk of bronchiolitis obliterans syndrome. *Transplantation,* 85**,** 771-774.

BREITBART, M., DELWART, E., ROSARIO, K., SEGALÉS, J., VARSANI, A. & ICTV REPORT, C. 2017. ICTV Virus Taxonomy Profile: Circoviridae. *The Journal of general virology,* 98**,** 1997-1998.

BRUNNER, S., HERNDLER-BRANDSTETTER, D., WEINBERGER, B. & GRUBECK-LOEBENSTEIN, B. 2011. Persistent viral infections and immune aging. *Ageing research reviews,* 10**,** 362-369.

BURGUETE, S. R., ELLI, D. J., FEANDEZ, J. F. & LEVINE, S. M. 2013. Lung transplant infection. *Respirology,* 18**,** 22-38.

BZHALAVA, D., EKSTRÖM, J., LYSHOLM, F., HULTIN, E., FAUST, H., PERSSON, B., LEHTINEN, M., DE VILLIERS, E.-M. M. & DILLNER, J. 2012. Phylogenetically diverse TT virus viremia among pregnant women. *Virology,* 432**,** 427-434.

CADWELL, K. 2015. The virome in host health and disease. *Immunity,* 42**,** 805-813.

CALIFF, K. J., SCHWARZBERG-LIPSON, K., GARG, N., GIBBONS, S. M., CAPORASO, J. G., SLOTS, J., COHEN, C., DORRESTEIN, P. C. & KELLEY, S. T. 2017. Multi-omics Analysis of Periodontal Pocket Microbial Communities Pre- and Posttreatment. *mSystems,* 2.

CANTU, E., LEDERER, D. J., MEYER, K., MILEWSKI, K., SUZUKI, Y., SHAH, R. J., DIAMOND, J. M., MEYER, N. J., TOBIAS, J. W., BALDWIN, D. A., VAN DEERLIN, V. M., OLTHOFF, K. M., SHAKED, A., CHRISTIE, J. D. & INVESTIGATORS, C. 2013. Gene set enrichment analysis identifies key innate immune pathways in primary graft dysfunction after

lung transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 13**,** 1898-1904.

CANTU, E., SHAH, R. J., LIN, W., DAYE, Z. J., DIAMOND, J. M., SUZUKI, Y., ELLIS, J. H., BORDERS, C. F., ANDAH, G. A., BEDUHN, B., MEYER, N. J., RUSCHEFSKI, M., APLENC, R., FENG, R., CHRISTIE, J. D. & INVESTIGATORS, L. 2015. Oxidant stress regulatory genetic variation in recipients and donors contributes to risk of primary graft dysfunction after lung transplantation. *The Journal of thoracic and cardiovascular surgery,* 149**,** 596-602.

CAPPUYNS, I., GUGERLI, P. & MOMBELLI, A. 2005. Viruses in periodontal disease - a review. *Oral diseases,* 11**,** 219-229.

CHAMBERS, D. C., YUSEN, R. D., CHERIKH, W. S., GOLDFARB, S. B., KUCHERYAVAYA, A. Y., KHUSCH, K., LEVVEY, B. J., LUND, L. H., MEISER, B., ROSSANO, J. W., STEHLIK, J., FOR & TRANSPLANTATION, I. 2017. The Registry of the International Society for Heart and Lung Transplantation: Thirty-fourth Adult Lung And Heart-Lung Transplantation Report-2017; Focus Theme: Allograft ischemic time. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation,* 36**,** 1047-1059.

CHARLSON, E. S., BITTINGER, K., HAAS, A. R., FITZGERALD, A. S., FRANK, I., YADAV, A., BUSHMAN, F. D. & COLLMAN, R. G. 2011. Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract. *American Journal of Respiratory and Critical Care Medicine,* 184**,** 957-963.

CHARLSON, E. S., CHEN, J., CUSTERS-ALLEN, R., BITTINGER, K., LI, H., SINHA, R., HWANG, J., BUSHMAN, F. D. & COLLMAN, R. G. 2010. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PloS one,* 5.

CHARLSON, E. S., DIAMOND, J. M., BITTINGER, K., FITZGERALD, A. S., YADAV, A., HAAS, A. R., BUSHMAN, F. D. & COLLMAN, R. G. 2012. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. *American journal of respiratory and critical care medicine,* 186**,** 536-545.

CHEN, T., VÄISÄNEN, E., MATTILA, P. S., HEDMAN, K. & SÖDERLUND-VENERMO, M. 2013. Antigenic diversity and seroprevalences of Torque teno viruses in children and adults by ORF2-based immunoassays. *The Journal of general virology,* 94**,** 409-417.

CHEUNG, A. K., NG, T. F. F., LAGER, K. M., ALT, D. P., DELWART, E. L. & POGRANICHNIY, R. M. 2014. Identification of a novel single-stranded circular DNA virus in pig feces. *Genome announcements,* 2.

CHRISTENSEN, J. K., EUGEN-OLSEN, J., SŁRENSEN, M., ULLUM, H., GJEDDE, S. B., PEDERSEN, B. K., NIELSEN, J. O. & KROGSGAARD, K. 2000. Prevalence and prognostic significance of infection with TT virus in patients infected with human immunodeficiency virus. *The Journal of infectious diseases,* 181**,** 1796-1799.

CLARKE, E. L., CONNELL, A. J., SIX, E., KADRY, N. A., ABBAS, A. A., HWANG, Y., EVERETT, J. K., HOFSTAEDTER, C. E., MARSH, R., ARMANT, M., KELSEN, J., NOTARANGELO, L. D., COLLMAN, R. G., HACEIN-BEY-ABINA, S., KOHN, D. B., CAVAZZANA, M., FISCHER, A., WILLIAMS, D. A., PAI, S.-Y. Y. & BUSHMAN, F. D. 2018a. T cell dynamics and response of the microbiota after gene therapy to treat X-linked severe combined immunodeficiency. *Genome medicine,* 10**,** 70.

CLARKE, E. L., LAUDER, A. P., HOFSTAEDTER, C. E., HWANG, Y., FITZGERALD, A. S., IMAI, I., BIERNAT, W., RĘKAWIECKI, B., MAJEWSKA, H., DUBANIEWICZ, A., LITZKY, L. A., FELDMAN, M. D., BITTINGER, K., ROSSMAN, M. D., PATTERSON, K. C., BUSHMAN, F. D. & COLLMAN, R. G. 2017a. Microbial Lineages in Sarcoidosis: A Metagenomic

Analysis Tailored for Low Microbial Content Samples. *American journal of respiratory and critical care medicine*.
CLARKE, E. L., SUNDARARAMAN, S. A., SEIFERT, S. N., BUSHMAN, F. D., HAHN, B. H. & BRISSON, D. 2017b. swga: a primer design toolkit for selective whole genome amplification. *Bioinformatics (Oxford, England),* 33**,** 2071-2077.
CLARKE, E. L., TAYLOR, L. J., ZHAO, C., CONNELL, J., BUSHMAN, F. D. & BITTINGER, K. 2018b. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *bioRxiv*.
CONSORTIUM, H. 2012. Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-214.
COSTALONGA, M. & HERZBERG, M. C. 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology letters,* 162**,** 22-38.
CRIBBS, S. K. & BECK, J. M. 2017. Microbiome in the pathogenesis of cystic fibrosis and lung transplant-related disease. *Translational research : the journal of laboratory and clinical medicine,* 179**,** 84-96.
CROOKS, G. E., HON, G., CHANDONIA, J.-M. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator. *Genome research,* 14**,** 1188-1190.
CUI, L., WU, B., ZHU, X., GUO, X., GE, Y., ZHAO, K., QI, X., SHI, Z., ZHU, F., SUN, L. & ZHOU, M. 2017. Identification and genetic characterization of a novel circular single-stranded DNA virus in a human upper respiratory tract sample. *Archives of virology,* 162**,** 3305-3312.
DAUD, S. A., YUSEN, R. D., MEYERS, B. F., CHAKINALA, M. M., WALTER, M. J., ALOUSH, A. A., PATTERSON, G. A., TRULOCK, E. P. & HACHEM, R. R. 2007. Impact of immediate primary lung allograft dysfunction on bronchiolitis obliterans syndrome. *American journal of respiratory and critical care medicine,* 175**,** 507-513.
DE VLAMINCK, I., KHUSH, K. K., STREHL, C., KOHLI, B., LUIKART, H., NEFF, N. F., OKAMOTO, J., SNYDER, T. M., CORNFIELD, D. N., NICOLLS, M. R., WEILL, D., BERNSTEIN, D., VALANTINE, H. A. & QUAKE, S. R. 2013. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell,* 155**,** 1178-1187.
DEAN, F. B., HOSONO, S., FANG, L., WU, X., FARUQI, A. F., BRAY-WARD, P., SUN, Z., ZONG, Q., DU, Y., DU, J., DRISCOLL, M., SONG, W., KINGSMORE, S. F., EGHOLM, M. & LASKEN, R. S. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 5261-5266.
DEVALLE, S., RUA, F., MORGADO, M. G. & NIEL, C. 2009. Variations in the frequencies of torque teno virus subpopulations during HAART treatment in HIV-1-coinfected patients. *Archives of Virology,* 154**,** 1285-1291.
DIAMOND, J. M. & WIGFIELD, C. H. 2013. Role of innate immunity in primary graft dysfunction after lung transplantation. *Current opinion in organ transplantation,* 18**,** 518-523.
DICKSON, R. P., ERB-DOWNWARD, J. R., FREEMAN, C. M., WALKER, N., SCALES, B. S., BECK, J. M., MARTINEZ, F. J., CURTIS, J. L., LAMA, V. N. & HUFFNAGLE, G. B. 2014a. Changes in the lung microbiome following lung transplantation include the emergence of two distinct Pseudomonas species with distinct clinical associations. *PloS one,* 9.
DICKSON, R. P., ERB-DOWNWARD, J. R., PRESCOTT, H. C., MARTINEZ, F. J., CURTIS, J. L., LAMA, V. N. & HUFFNAGLE, G. B. 2014b. Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. *Journal of clinical microbiology,* 52**,** 3605-3613.
DOLLIVE, S., PETERFREUND, G. L., SHERRILL-MIX, S., BITTINGER, K., SINHA, R., HOFFMANN, C., NABEL, C. S., HILL, D. A., ARTIS, D., BACHMAN, M. A., CUSTERS-ALLEN, R.,

GRUNBERG, S., WU, G. D., LEWIS, J. D. & BUSHMAN, F. D. 2012. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biology,* 13.
EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32**,** 1792-1797.
EDLUND, A., SANTIAGO-RODRIGUEZ, T. M., BOEHM, T. K. & PRIDE, D. T. 2015. Bacteriophage and their potential roles in the human oral cavity. *Journal of oral microbiology,* 7**,** 27423.
ELLIS, J. 2014. Porcine circovirus: a historical perspective. *Veterinary pathology,* 51**,** 315-327.
ELLIS, J. A., ALLAN, G. & KRAKOWKA, S. 2008. Effect of coinfection with genogroup 1 porcine torque teno virus on porcine circovirus type 2-associated postweaning multisystemic wasting syndrome in gnotobiotic pigs. *American journal of veterinary research,* 69**,** 1608-1614.
EPPINGER, M. J., DEEB, G. M., BOLLING, S. F. & WARD, P. A. 1997. Mediators of ischemia-reperfusion injury of rat lung. *The American journal of pathology,* 150**,** 1773-1784.
EPPINGER, M. J., JONES, M. L., DEEB, G. M., BOLLING, S. F. & WARD, P. A. 1995. Pattern of injury and the role of neutrophils in reperfusion injury of rat lung. *The Journal of surgical research,* 58**,** 713-718.
ERB-DOWNWARD, J. R., THOMPSON, D. L., HAN, M. K., FREEMAN, C. M., MCCLOSKEY, L., SCHMIDT, L. A., YOUNG, V. B., TOEWS, G. B., CURTIS, J. L., SUNDARAM, B., MARTINEZ, F. J. & HUFFNAGLE, G. B. 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PloS one,* 6.
FAHSBENDER, E., BURNS, J. M., KIM, S., KRABERGER, S., FRANKFURTER, G., EILERS, A. A., SHERO, M. R., BELTRAN, R., KIRKHAM, A., MCCORKELL, R., BERNGARTT, R. K., MALE, M. F., BALLARD, G., AINLEY, D. G., BREITBART, M. & VARSANI, A. 2017. Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica. *Virus evolution,* 3.
FAROPOULOS, K. & APOSTOLAKIS, E. 2009. Brain death and its influence on the lungs of the donor: how is it prevented? *Transplantation proceedings,* 41**,** 4114-4119.
FAUQUET, C. M., MAYO, M. A., MANILOFF, J., DESSELBERGER, U. & BALL, L. 2005. *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses*, Academic Press.
FEYZIOĞLU, B., TEKE, T., OZDEMIR, M., KARAIBRAHIMOĞLU, A., DOĞAN, M. & YAVŞAN, M. 2014. The presence of Torque teno virus in chronic obstructive pulmonary disease. *International journal of clinical and experimental medicine,* 7**,** 3461-3466.
FIRTH, C., CHARLESTON, M. A., DUFFY, S., SHAPIRO, B. & HOLMES, E. C. 2009. Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *Journal of virology,* 83**,** 12813-12821.
FISER, TRIBBLE, C. G., LONG & KAZA, A. K. 2001. Pulmonary macrophages are involved in reperfusion injury after lung transplantation. *Pulmonary macrophages are involved in reperfusion injury after lung transplantation*.
FLINT, H. J., SCOTT, K. P., LOUIS, P. & DUNCAN, S. H. 2012. The role of the gut microbiota in nutrition and health. *Nature reviews. Gastroenterology & hepatology,* 9**,** 577-589.
FOCOSI, D., ANTONELLI, G., PISTELLO, M. & MAGGI, F. 2016. Torquetenovirus: the human virome from bench to bedside. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases,* 22**,** 589-593.
FOCOSI, D., MACERA, L., BOGGI, U., NELLI, L. C. & MAGGI, F. 2015. Short-term kinetics of torque teno virus viraemia after induction immunosuppression confirm T

lymphocytes as the main replication-competent cells. *The Journal of general virology,* 96**,** 115-117.
FOCOSI, D., MAGGI, F., ALBANI, M., MACERA, L., RICCI, V., GRAGNANI, S., DI BEO, S., GHIMENTI, M., ANTONELLI, G., BENDINELLI, M., PISTELLO, M., CECCHERINI-NELLI, L. & PETRINI, M. 2010. Torquetenovirus viremia kinetics after autologous stem cell transplantation are predictable and may serve as a surrogate marker of functional immune reconstitution. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 47**,** 189-192.
FODOR, B., LADÁNYI, E., ALEKSZA, M., TAKÁCS, M., LAKOS, G., ARKOSSY, O., KOÓS, A., NAGY, A., SZÉLL, J., KLENK, N., SÁRVÁRY, E. & SIPKA, S. 2002. No effect of transfusion transmitted virus viremia on the distribution and activation of peripheral lymphocytes in hemodialyzed patients. *Nephron,* 92**,** 933-937.
GINI, C. 1912. Variabilità e mutabilità. Contributi allo studio dele relazioni e delle distribuzioni statistiche. *Studi Economico-Giuridici della Università di Cagliari*.
GORBALENYA, A. E., KOONIN, E. V. & WOLF, Y. I. 1990. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS letters,* 262**,** 145-148.
GÖRZER, I., HALOSCHAN, M., JAKSCH, P., KLEPETKO, W. & PUCHHAMMER-STÖCKL, E. 2014. Plasma DNA levels of Torque teno virus and immunosuppression after lung transplantation. *The Journal of Heart and Lung Transplantation,* 33**,** 320-323.
GÖRZER, I., JAKSCH, P., KUNDI, M., SEITZ, T., KLEPETKO, W. & PUCHHAMMER-STÖCKL, E. 2015. Pre-transplant plasma Torque Teno virus load and increase dynamics after lung transplantation. *PloS one,* 10.
GOTTLIEB, J., SCHULZ, T. F., WELTE, T., FUEHNER, T., DIERICH, M., SIMON, A. R. & ENGELMANN, I. 2009. Community-acquired respiratory viral infections in lung transplant recipients: a single season cohort study. *Transplantation,* 87**,** 1530-1537.
GRABUNDZIJA, I., MESSING, S. A., THOMAS, J., COSBY, R. L., BILIC, I., MISKEY, C., GOGOL-DÖRING, A., KAPITONOV, V., DIEM, T., DALDA, A., JURKA, J., PRITHAM, E. J., DYDA, F., IZSVÁK, Z. & IVICS, Z. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature communications,* 7**,** 10716.
GREEN, M., COVINGTON, S., TARANTO, S., WOLFE, C., BELL, W., BIGGINS, S. W., CONTI, D., DESTEFANO, G. D., DOMINGUEZ, E., ENNIS, D., GROSS, T., KLASSEN-FISCHER, M., KOTTON, C., LAPOINTE-RUDOW, D., LAW, Y., LUDROSKY, K., MENEGUS, M., MORRIS, M. I., NALESNIK, M. A., PAVLAKIS, M., PRUETT, T., SIFRI, C. & KAUL, D. 2015. Donor-derived transmission events in 2013: a report of the Organ Procurement Transplant Network Ad Hoc Disease Transmission Advisory Committee. *Transplantation,* 99**,** 282-287.
GROSSI, P. A., FISHMAN, J. A. & OF PRACTICE, A. S. T. 2009. Donor-Derived Infections in Solid Organ Transplant Recipients. *American Journal of Transplantation,* 9.
GUINDON, S., DUFAYARD, J.-F. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology,* 59**,** 307-321.
HALOSCHAN, M., BETTESCH, R., GÖRZER, I., WESESLINDTNER, L., KUNDI, M. & PUCHHAMMER-STÖCKL, E. 2014. TTV DNA plasma load and its association with age, gender, and HCMV IgG serostatus in healthy adults. *Age (Dordrecht, Netherlands),* 36**,** 9716.
HANDLEY, S. A., THACKRAY, L. B., ZHAO, G., PRESTI, R., MILLER, A. D., DROIT, L., ABBINK, P., MAXFIELD, L. F., KAMBAL, A., DUAN, E., STANLEY, K., KRAMER, J., MACRI, S. C., PERMAR, S. R., SCHMITZ, J. E., MANSFIELD, K., BRENCHLEY, J. M., VEAZEY, R. S.,

STAPPENBECK, T. S., WANG, D., BAROUCH, D. H. & VIRGIN, H. W. 2012. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell,* 151**,** 253-266.
HÄNZELMANN, S., CASTELO, R. & GUINNEY, J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics,* 14**,** 7.
HARRISON, B. D., RKER, H., BOCK, K. R., GUTHRIE, E. J., MEREDITH, G. & ATKINSON, M. 1977. Plant viruses with circular single-stranded DNA. *Nature,* 270.
HIBBING, M. E., FUQUA, C., PARSEK, M. R. & PETERSON, S. B. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology,* 8**,** 15-25.
HILTY, M., BURKE, C., PEDRO, H., CARDENAS, P., BUSH, A., BOSSLEY, C., DAVIES, J., ERVINE, A., POULTER, L., PACHTER, L., MOFFATT, M. F. & COOKSON, W. O. 2010. Disordered microbial communities in asthmatic airways. *PloS one,* 5.
HINO, S. & MIYATA, H. 2007. Torque teno virus (TTV): current status. *Reviews in medical virology,* 17**,** 45-57.
HO, J., JEDRYCH, J. J., FENG, H., NATALIE, A. A., GRANDINETTI, L., MIRVISH, E., CRESPO, M. M., YADAV, D., FASANELLA, K. E., PROKSELL, S., KUAN, S.-F. F., PASTRANA, D. V., BUCK, C. B., SHUDA, Y., MOORE, P. S. & CHANG, Y. 2015. Human polyomavirus 7-associated pruritic rash and viremia in transplant recipients. *The Journal of infectious diseases,* 211**,** 1560-1565.
HOLTZ, L. R., CAO, S., ZHAO, G., BAUER, I. K., DENNO, D. M., KLEIN, E. J., ANTONIO, M., STINE, O. C., SNELLING, T. L., KIRKWOOD, C. D. & WANG, D. 2014. Geographic variation in the eukaryotic virome of human diarrhea. *Virology,* 468-470**,** 556-564.
HUANG, X. & MADAN, A. 1999. CAP3: A DNA sequence assembly program. *Genome research,* 9**,** 868-877.
HUANG, Y. J., CHARLSON, E. S., COLLMAN, R. G., COLOMBINI-HATCH, S., MARTINEZ, F. D. & SENIOR, R. M. 2013. The Role of the Lung Microbiome in Health and Disease. A National Heart, Lung, and Blood Institute Workshop Report. *American journal of respiratory and critical care medicine,* 187**,** 1382-1387.
HUANG, Y. J. & LYNCH, S. V. 2011. The emerging relationship between the airway microbiota and chronic respiratory disease: clinical implications. *Expert review of respiratory medicine,* 5**,** 809-821.
HUANG, Y. J., NELSON, C. E., BRODIE, E. L., DESANTIS, T. Z., BAEK, M. S., LIU, J., WOYKE, T., ALLGAIER, M., BRISTOW, J., WIENER-KRONISH, J. P., SUTHERLAND, E. R., KING, T. S., ICITOVIC, N., MARTIN, R. J., CALHOUN, W. J., CASTRO, M., DENLINGER, L. C., DIMANGO, E., KRAFT, M., PETERS, S. P., WASSERMAN, S. I., WECHSLER, M. E., BOUSHEY, H. A., LYNCH, S. V. & AND HEART, B. 2011. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *The Journal of allergy and clinical immunology,* 127**,** 372.
IHAKA, R. & GENTLEMAN, R. 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics,* 5**,** 299-314.
ILYINA, T. V. & KOONIN, E. V. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research,* 20**,** 3279-3285.
ISON, M. G. & NALESNIK, M. A. 2011. An Update on Donor-Derived Disease Transmission in Organ Transplantation. *American Journal of Transplantation,* 11**,** 1123-1130.
IUS, F., SOMMER, W., TUDORACHE, I., KÜHN, C., AVSAR, M., SIEMENI, T., SALMAN, J., HALLENSLEBEN, M., KIENEKE, D., GREER, M., GOTTLIEB, J., HAVERICH, A. & WARNECKE, G. 2014. Early donor-specific antibodies in lung transplantation: Risk

factors and impact on survival. *The Journal of Heart and Lung Transplantation,* 33**,** 1255-1263.
JELCIC, I., HOTZ-WAGENBLATT, A., HUNZIKER, A., ZUR HAUSEN, H. & DE VILLIERS, E.-M. M. 2004. Isolation of multiple TT virus genotypes from spleen biopsy tissue from a Hodgkin's disease patient: genome reorganization and diversity in the hypervariable region. *Journal of virology,* 78**,** 7498-7507.
JOHNSTON, L. K., RIMS, C. R., GILL, S. E., MCGUIRE, J. K. & MANICONE, A. M. 2012. Pulmonary Macrophage Subpopulations in the Induction and Resolution of Acute Lung Injury. *American journal of respiratory cell and molecular biology,* 47**,** 417-426.
KAKKOLA, L., BONDÉN, H., HEDMAN, L., KIVI, N., MOISALA, S., JULIN, J., YLÄ-LIEDENPOHJA, J., MIETTINEN, S., KANTOLA, K., HEDMAN, K. & SÖDERLUND-VENERMO, M. 2008. Expression of all six human Torque teno virus (TTV) proteins in bacteria and in insect cells, and analysis of their IgG responses. *Virology,* 382**,** 182-189.
KEKARAINEN, T. & SEGALÉS, J. 2012. Torque teno sus virus in pigs: an emerging pathogen? *Transboundary and emerging diseases,* 59 Suppl 1**,** 103-108.
KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols,* 10**,** 845-858.
KERPEDJIEV, P., HAMMER, S. & HOFACKER, I. L. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics (Oxford, England),* 31**,** 3377-3379.
KHALIFAH, A. P., HACHEM, R. R., CHAKINALA, M. M., SCHECHTMAN, K. B., PATTERSON, G. A., SCHUSTER, D. P., MOHANAKUMAR, T., TRULOCK, E. P. & WALTER, M. J. 2004. Respiratory viral infections are a distinct risk for bronchiolitis obliterans syndrome and death. *American journal of respiratory and critical care medicine,* 170**,** 181-187.
KIM, D., HOFSTAEDTER, C. E., ZHAO, C., MATTEI, L., TANES, C., CLARKE, E., LAUDER, A., SHERRILL-MIX, S., CHEHOUD, C., KELSEN, J., CONRAD, M., COLLMAN, R. G., BALDASSANO, R., BUSHMAN, F. D. & BITTINGER, K. 2017. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome,* 5**,** 52.
KIM, K.-H. & BAE, J.-W. 2011. Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. *Applied and Environmental Microbiology,* 77**,** 7663-7668.
KIM, K.-H. H., CHANG, H.-W. W., NAM, Y.-D. D., ROH, S. W., KIM, M.-S. S., SUNG, Y., JEON, C. O., OH, H.-M. M. & BAE, J.-W. W. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and environmental microbiology,* 74**,** 5975-5985.
KIM, M. S., PARK, E. J., ROH, S. W. & AND ENVIRONMENTAL, B.-J. W. 2011. Diversity and abundance of single-stranded DNA viruses in human faeces. *Applied and environmental ….*
KINCAID, R. P., BURKE, J. M., COX, J. C., DE VILLIERS, E.-M. M. & SULLIVAN, C. S. 2013. A human torque teno virus encodes a microRNA that inhibits interferon signaling. *PLoS pathogens,* 9.
KNIGHT, R., CALLEWAERT, C., MAROTZ, C., HYDE, E. R., DEBELIUS, J. W., MCDONALD, D. & SOGIN, M. L. 2017. The Microbiome and Human Biology. *Annual review of genomics and human genetics,* 18**,** 65-86.
KOBAYASHI, T. & ANDOH, A. 2018. Numerical analyses of intestinal microbiota by data mining. *Journal of clinical biochemistry and nutrition,* 62**,** 124-131.
KOMATSU, H., INUI, A., SOGO, T., KURODA, K., TANAKA, T. & FUJISAWA, T. 2004. TTV infection in children born to mothers infected with TTV but not with HBV, HCV, or HIV. *Journal of medical virology,* 74**,** 499-506.

KOTLOFF, R. M. & THABUT, G. 2011. Lung transplantation. *American journal of respiratory and critical care medicine,* 184**,** 159-171.
KRISHNAMURTHY, S. R. & WANG, D. 2017. Origins and challenges of viral dark matter. *Virus research*.
KRISHNAMURTHY, S. R. & WANG, D. 2018. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology,* 516**,** 108-114.
KROSHUS, T. J., KSHETTRY, V. R., SAVIK, K., JOHN, R., HERTZ, M. I. & BOLMAN, R. M. 1997. Risk factors for the development of bronchiolitis obliterans syndrome after lung transplantation. *The Journal of thoracic and cardiovascular surgery,* 114**,** 195-202.
KRUPOVIC, M., GHABRIAL, S. A., JIANG, D. & VARSANI, A. 2016. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Archives of virology,* 161**,** 2633-2643.
KUMAR, P. K., GOTTLIEB, R. A., LINDSAY, S., DELANGE, N., PENN, T. E., CALAC, D. & KELLEY, S. T. 2018. Metagenomic analysis uncovers strong relationship between periodontal pathogens and vascular dysfunction in American Indian population. *bioRxiv***,** 250324.
KUTIKHIN, A. G., YUZHALIN, A. E. & BRUSINA, E. B. 2014. Mimiviridae, Marseilleviridae, and virophages as emerging human pathogens causing healthcare-associated infections. *GMS hygiene and infection control,* 9.
KUTSOGIANNIS, D. J., PAGLIARELLO, G., DOIG, C., ROSS, H. & SHEMIE, S. D. 2006. Medical management to optimize donor organ potential: review of the literature. *Canadian journal of anaesthesia = Journal canadien d'anesthésie,* 53**,** 820-830.
LABONTÉ, J. M. & SUTTLE, C. A. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. *The ISME journal,* 7**,** 2169-2177.
LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods,* 9**,** 357-359.
LAUDER, A. P., ROCHE, A. M., SHERRILL-MIX, S., BAILEY, A., LAUGHLIN, A. L., BITTINGER, K., LEITE, R., ELOVITZ, M. A., PARRY, S. & BUSHMAN, F. D. 2016. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome,* 4**,** 29.
LEE, J. C. & CHRISTIE, J. D. 2011. Primary Graft Dysfunction. *Clinics in Chest Medicine,* 32**,** 279-293.
LEE, Y., LIN, C.-M. M., JENG, C.-R. R., CHANG, H.-W. W., CHANG, C.-C. C. & PANG, V. F. 2015. The pathogenic role of torque teno sus virus 1 and 2 and their correlations with various viral pathogens and host immunocytes in wasting pigs. *Veterinary microbiology,* 180**,** 186-195.
LEFEUVRE, P., LETT, J. M. M., VARSANI, A. & MARTIN, D. P. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. *Journal of virology,* 83**,** 2697-2707.
LEINONEN, R., SUGAWARA, H., SHUMWAY, M. & COLLABORATION, I. 2011. The sequence read archive. *Nucleic acids research,* 39**,** 21.
LEPPIK, L., GUNST, K., LEHTINEN, M., DILLNER, J., STREKER, K. & DE VILLIERS, E.-M. M. 2007. In vivo and in vitro intragenomic rearrangement of TT viruses. *Journal of virology,* 81**,** 9346-9356.
LEWANDOWSKA, D. W., SCHREIBER, P. W., SCHUURMANS, M. M. M., RUEHE, B., ZAGORDI, O., BAYARD, C., GREINER, M., GEISSBERGER, F. D., CAPAUL, R., ZBINDEN, A., BÖNI, J., BENDEN, C., MUELLER, N. J., TRKOLA, A. & HUBER, M. 2017. Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PloS one,* 12.

LI, D., LIU, C.-M. M., LUO, R., SADAKANE, K. & LAM, T.-W. W. 2015a. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England),* 31**,** 1674-1676.
LI, F., ZHU, C., DENG, F.-Y. Y., WONG, M. C. M. C. M., LU, H.-X. X. & FENG, X.-P. P. 2017. Herpesviruses in etiopathogenesis of aggressive periodontitis: A meta-analysis based on case-control studies. *PloS one,* 12.
LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England),* 26**,** 589-595.
LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & SUBGROUP 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England),* 25**,** 2078-2079.
LI, L., DENG, X., DA COSTA, A. C., BRUHN, R., DEEKS, S. G. & DELWART, E. 2015b. Virome analysis of antiretroviral-treated HIV patients shows no correlation between T-cell activation and anelloviruses levels. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 72**,** 106-113.
LI, L., DENG, X., LINSUWANON, P. & OF …, B.-D. 2013. AIDS alters the commensal plasma virome. *Journal of ….*
LI, L., KAPOOR, A., SLIKAS, B., BAMIDELE, O. S., WANG, C., SHAUKAT, S., MASROOR, M. A., WILSON, M. L., NDJANGO, J.-B. N. B., PEETERS, M., GROSS-CAMP, N. D., MULLER, M. N., HAHN, B. H., WOLFE, N. D., TRIKI, H., BARTKUS, J., ZAIDI, S. Z. & DELWART, E. 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *Journal of virology,* 84**,** 1674-1682.
LIM, E. S., ZHOU, Y., ZHAO, G., BAUER, I. K., DROIT, L., NDAO, I. M., WARNER, B. B., TARR, P. I., WANG, D. & HOLTZ, L. R. 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature medicine,* 21**,** 1228-1234.
LISITSYN, N., LISITSYN, N. & WIGLER, M. 1993. Cloning the differences between two complex genomes. *Science*.
LOWS, R. L. 1984. Measurement of Inequality: The Gini Coefficient and School Finance Studies. *Journal of Education Finance,* 10**,** 83-94.
LUND, L. H., EDWARDS, L. B., DIPCHAND, A. I., GOLDFARB, S., KUCHERYAVAYA, A. Y., LEVVEY, B. J., MEISER, B., ROSSANO, J. W., YUSEN, R. D., STEHLIK, J., FOR & TRANSPLANTATION, I. 2016. The Registry of the International Society for Heart and Lung Transplantation: Thirty-third Adult Heart Transplantation Report-2016; Focus Theme: Primary Diagnostic Indications for Transplant. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation,* 35**,** 1158-1169.
LY, M., ABELES, S. R., BOEHM, T. K., ROBLES-SIKISAKA, R., NAIDU, M., SANTIAGO-RODRIGUEZ, T. & PRIDE, D. T. 2014. Altered oral viral ecology in association with periodontal disease. *mBio,* 5**,** 14.
LYSHOLM, F., WETTERBOM, A., LINDAU, C., DARBAN, H., BJERKNER, A., FAHLANDER, K., LINDBERG, A. M., PERSSON, B., ALLANDER, T. & ANDERSSON, B. 2012. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PloS one,* 7.
MA, C.-M. M., HON, C.-C. C., LAM, T.-Y. Y., LI, V. Y., WONG, C. K., DE OLIVEIRA, T. & LEUNG, F. C. 2007. Evidence for recombination in natural populations of porcine circovirus type 2 in Hong Kong and mainland China. *The Journal of general virology,* 88**,** 1733-1737.
MADSEN, C. D., EUGEN-OLSEN, J., KIRK, O., PARNER, J., KAAE CHRISTENSEN, J., BRASHOLT, M. S., OLE NIELSEN, J. & KROGSGAARD, K. 2002. TTV viral load as a marker for

immune reconstitution after initiation of HAART in HIV-infected patients. *HIV clinical trials,* 3**,** 287-295.
MAGGI, F., FORNAI, C., ZACCARO, L., MORRICA, A., VATTERONI, M. L., ISOLA, P., MARCHI, S., RICCHIUTI, A., PISTELLO, M. & BENDINELLI, M. 2001. TT virus (TTV) loads associated with different peripheral blood cell types and evidence for TTV replication in activated mononuclear cells. *Journal of medical virology,* 64**,** 190-194.
MAGGI, F., PIFFERI, M., FORNAI, C., ANDREOLI, E., TEMPESTINI, E., VATTERONI, M., PRESCIUTTINI, S., MARCHI, S., PIETROBELLI, A., BONER, A., PISTELLO, M. & BENDINELLI, M. 2003. TT virus in the nasal secretions of children with acute respiratory diseases: relations to viremia and disease severity. *Journal of virology,* 77**,** 2418-2425.
MAN, W. H., DE STEENHUIJSEN PITERS, W. A. A. & BOGAERT, D. 2017. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nature reviews. Microbiology,* 15**,** 259-270.
MANKERTZ, A., PERSSON, F., MANKERTZ, J., BLAESS, G. & BUHK, H. J. 1997. Mapping and characterization of the origin of DNA replication of porcine circovirus. *Journal of virology,* 71**,** 2562-2566.
MANKOTIA, D. S. & IRSHAD, M. 2014. Cloning and expression of N22 region of Torque Teno virus (TTV) genome and use of peptide in developing immunoassay for TTV antibodies. *Virology journal,* 11**,** 96.
MARISCAL, L. F., LÓPEZ-ALCOROCHO, J. M., RODRÍGUEZ-IÑIGO, E., ORTIZ-MOVILLA, N., DE LUCAS, S., BARTOLOMÉ, J. & CARREÑO, V. 2002. TT virus replicates in stimulated but not in nonstimulated peripheral blood mononuclear cells. *Virology,* 301**,** 121-129.
MARTIN-GANDUL, C., MUELLER, N. J., PASCUAL, M. & MANUEL, O. 2015. The Impact of Infection on Chronic Allograft Dysfunction and Allograft Survival After Solid Organ Transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 15**,** 3024-3040.
MASOURIDI-LEVRAT, S., PRADIER, A., SIMONETTA, F., KAISER, L., CHALANDON, Y. & ROOSNEK, E. 2016. Torque teno virus in patients undergoing allogeneic hematopoietic stem cell transplantation for hematological malignancies. *Bone marrow transplantation,* 51**,** 440-442.
MAURICE, C. F., HAISER, H. J. & TURNBAUGH, P. J. 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell,* 152**,** 39-50.
MCELVANIA TEKIPPE, E., WYLIE, K. M., DEYCH, E., SODERGREN, E., WEINSTOCK, G. & STORCH, G. A. 2012. Increased prevalence of anellovirus in pediatric patients with fever. *PloS one,* 7.
MENG, X. J. 2012. Emerging and re-emerging swine viruses. *Transboundary and emerging diseases,* 59 Suppl 1**,** 85-102.
MINOT, S., BRYSON, A., CHEHOUD, C., WU, G. D., LEWIS, J. D. & BUSHMAN, F. D. 2013. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 12450-12455.
MINOT, S., SINHA, R., CHEN, J., LI, H., KEILBAUGH, S. A., WU, G. D., LEWIS, J. D. & BUSHMAN, F. D. 2011. The human gut virome: inter-individual variation and dynamic response to diet. *Genome research,* 21**,** 1616-1625.
MOEN, E. M., SAGEDAL, S., BJØRO, K., DEGRÉ, M., OPSTAD, P. K. & GRINDE, B. 2003. Effect of immune modulation on TT virus (TTV) and TTV-like-mini-virus (TLMV) viremia. *Journal of medical virology,* 70**,** 177-182.

MOKILI, J. L., DUTILH, B. E., LIM, Y. W., SCHNEIDER, B. S., TAYLOR, T., HAYNES, M. R., METZGAR, D., MYERS, C. A., BLAIR, P. J., NOSRAT, B., WOLFE, N. D. & ROHWER, F. 2013. Identification of a novel human papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness. *PloS one,* 8.

MONACO, C. L., GOOTENBERG, D. B., ZHAO, G., HANDLEY, S. A., GHEBREMICHAEL, M. S., LIM, E. S., LANKOWSKI, A., BALDRIDGE, M. T., WILEN, C. B., FLAGG, M., NORMAN, J. M., KELLER, B. C., LUÉVANO, J. M. M., WANG, D., BOUM, Y., MARTIN, J. N., HUNT, P. W., BANGSBERG, D. R., SIEDNER, M. J., KWON, D. S. & VIRGIN, H. W. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell host & microbe,* 19**,** 311-322.

MUHIRE, B. M., GOLDEN, M., MURRELL, B., LEFEUVRE, P., LETT, J.-M. M., GRAY, A., POON, A. Y., NGANDU, N. K., SEMEGNI, Y., TANOV, E. P., MONJANE, A. L. L., HARKINS, G. W., VARSANI, A., SHEPHERD, D. N. & MARTIN, D. P. 2014. Evidence of pervasive biologically functional secondary structures within the genomes of eukaryotic single-stranded DNA viruses. *Journal of virology,* 88**,** 1972-1989.

MURRELL, B., MOOLA, S., MABONA, A., WEIGHILL, T., SHEWARD, D., KOSAKOVSKY POND, S. L. & SCHEFFLER, K. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution,* 30**,** 1196-1205.

MUSHAHWAR, I. K., ERKER, J. C., MUERHOFF, S. A., LEARY, T. P., SIMONS, J. N., BIRKENMEYER, L. G., CHALMERS, M. L., PILOT-MATIAS, T. J. & DEXAI, S. M. 1999. Molecular and biophysical characterization of TT virus: Evidence for a new virus family infecting humans. *Proceedings of the National Academy of Sciences,* 96**,** 3177-3182.

NACCACHE, S. N., GRENINGER, A. L., LEE, D., COFFEY, L. L., PHAN, T., REIN-WESTON, A., ARONSOHN, A., HACKETT, J., DELWART, E. L. & CHIU, C. Y. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *Journal of virology,* 87**,** 11966-11977.

NAIDU, B. V., KRISHNADASAN, B., FARIVAR, A. S., WOOLLEY, S. M., THOMAS, R., ROOIJEN, N., VERRIER, E. D. & MULLIGAN, M. S. 2003. Early activation of the alveolar macrophage is critical to the development of lung ischemia-reperfusion injury. *The Journal of thoracic and cardiovascular surgery,* 126**,** 200-207.

NAKA, Y., MARSH, H. C., SCESNEY, S. M., OZ, M. C. & PINSKY, D. J. 1997. Complement activation as a cause for primary graft failure in an isogeneic rat model of hypothermic lung preservation and transplantation. *Transplantation,* 64**,** 1248-1255.

NAKAJIMA, T., PALCHEVSKY, V., PERKINS, D. L., BELPERIO, J. A. & FINN, P. W. 2011. Lung transplantation: infection, inflammation, and the microbiome. *Seminars in immunopathology,* 33**,** 135-156.

NINOMIYA, M., TAKAHASHI, M., NISHIZAWA, T., SHIMOSEGAWA, T. & OKAMOTO, H. 2008. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *Journal of clinical microbiology,* 46**,** 507-514.

NISHIZAWA, T., OKAMOTO, H., KONISHI, K., YOSHIZAWA, H., MIYAKAWA, Y. & MAYUMI, M. 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and biophysical research communications,* 241**,** 92-97.

NISHIZAWA, T., OKAMOTO, H., TSUDA, F., AIKAWA, T., SUGAI, Y., KONISHI, K., AKAHANE, Y., UKITA, M., TANAKA, T. & MIYAKAWA, Y. 1999. Quasispecies of TT virus (TTV) with sequence divergence in hypervariable regions of the capsid protein in chronic TTV infection. *Journal of virology,* 73**,** 9604-9608.

NORMAN, J. M., HANDLEY, S. A., BALDRIDGE, M. T., DROIT, L., LIU, C. Y., KELLER, B. C., KAMBAL, A., MONACO, C. L., ZHAO, G., FLESHNER, P., STAPPENBECK, T. S., MCGOVERN, D. P., KESHAVARZIAN, A., MUTLU, E. A., SAUK, J., GEVERS, D., XAVIER, R. J., WANG, D., PARKES, M. & VIRGIN, H. W. 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell,* 160**,** 447-460.

OKAMOTO, H., NISHIZAWA, T., TAKAHASHI, M., ASABE, S., TSUDA, F. & YOSHIKAWA, A. 2001. Heterogeneous distribution of TT virus of distinct genotypes in multiple tissues from infected humans. *Virology,* 288**,** 358-368.

OTT, C., DURET, L., CHEMIN, I., TRÉPO, C., MANDRAND, B. & KOMURIAN-PRADEL, F. 2000. Use of a TT virus ORF1 recombinant protein to detect anti-TT virus antibodies in human sera. *The Journal of general virology,* 81**,** 2949-2958.

PAEZ-ESPINO, D., ELOE-FADROSH, E. A., PAVLOPOULOS, G. A., THOMAS, A. D., HUNTEMANN, M., MIKHAILOVA, N., RUBIN, E., IVANOVA, N. N. & KYRPIDES, N. C. 2016. Uncovering Earth's virome. *Nature,* 536**,** 425-430.

PAGES, H., ABOYOUN, P., GENTLEMAN, R. & DEBROY, S. 2009. String objects representing biological sequences, and matching algorithms. *R package version,* 2.

PAWELEC, G. & DERHOVANESSIAN, E. 2011. Role of CMV in immune senescence. *Virus research,* 157**,** 175-179.

PEGHIN, M., HIRSCH, H. H., LEN, Ó., CODINA, G., BERASTEGUI, C., SÁEZ, B., SOLÉ, J., CABRAL, E., SOLÉ, A., ZURBANO, F., LÓPEZ-MEDRANO, F., ROMÁN, A. & GAVALDÁ, J. 2017. Epidemiology and Immediate Indirect Effects of Respiratory Viruses in Lung Transplant Recipients: A 5-Year Prospective Study. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 17**,** 1304-1312.

PENG, Y., LEUNG, H. C. M., YIU, S. M. & CHIN, F. Y. L. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics,* 28**,** 1420-1428.

PÉREZ-BROCAL, V. & MOYA, A. 2018. The analysis of the oral DNA virome reveals which viruses are widespread and rare among healthy young adults in Valencia (Spain). *PloS one,* 13.

PFEIFFER, F., GRÖBER, C., BLANK, M., HÄNDLER, K., BEYER, M., SCHULTZE, J. L. & MAYER, G. 2018. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports,* 8**,** 10950.

PHAN, T. G., DA COSTA, A. C., ZHANG, W., POTHIER, P., AMBERT-BALAY, K., DENG, X. & DELWART, E. 2015a. A new gyrovirus in human feces. *Virus genes,* 51**,** 132-135.

PHAN, T. G., LUCHSINGER, V., AVENDAÑO, L. F., DENG, X. & DELWART, E. 2014. Cyclovirus in nasopharyngeal aspirates of Chilean children with respiratory infections. *The Journal of general virology,* 95**,** 922-927.

PHAN, T. G., MORI, D., DENG, X., RAJINDRAJITH, S., RANAWAKA, U., FAN NG, T. F., BUCARDO-RIVERA, F., ORLANDI, P., AHMED, K. & DELWART, E. 2015b. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology,* 482**,** 98-104.

PIFFERI, M., MAGGI, F., DI CRISTOFANO, C., CANGIOTTI, A. M., NELLI, L. C., BEVILACQUA, G., MACCHIA, P., BENDINELLI, M. & BONER, A. L. 2008. Torquetenovirus infection and ciliary dysmotility in children with recurrent pneumonia. *The Pediatric infectious disease journal,* 27**,** 413-418.

POND, S. L., FROST, S. D. & MUSE, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England),* 21**,** 676-679.

PROCTOR, L. M. 1997. Advances in the study of marine viruses. *Microscopy research and technique,* 37**,** 136-161.

QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN, N. J. & SEGATA, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature biotechnology,* 35**,** 833-844.
QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England),* 26**,** 841-842.
RAMBAUT, A., POSADA, D. & REVIEWS …, C. K. A. 2004. The causes and consequences of HIV evolution. *The causes and consequences of HIV evolution*.
RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG,* 16**,** 276-277.
ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nature biotechnology,* 29**,** 24-26.
ROCCHI, J., RICCI, V., ALBANI, M., LANINI, L., ANDREOLI, E., MACERA, L., PISTELLO, M., CECCHERINI-NELLI, L., BENDINELLI, M. & MAGGI, F. 2009. Torquetenovirus DNA drives proinflammatory cytokines production and secretion by immune cells via toll-like receptor 9. *Virology,* 394**,** 235-242.
ROSARIO, K. & BREITBART, M. 2011. Exploring the viral world through metagenomics. *Current opinion in virology,* 1**,** 289-297.
ROSARIO, K., DUFFY, S. & BREITBART, M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Archives of virology,* 157**,** 1851-1871.
ROTUNDO, R., MAGGI, F., NIERI, M., MUZZI, L., BENDINELLI, M. & PRATO, G. P. P. 2004. TT virus infection of periodontal tissues: a controlled clinical and laboratory pilot study. *Journal of periodontology,* 75**,** 1216-1220.
SAADI, H., PAGNIER, I., COLSON, P., CHERIF, J. K., BEJI, M., BOUGHALMI, M., AZZA, S., ARMSTRONG, N., ROBERT, C., FOURNOUS, G., LA SCOLA, B. & RAOULT, D. 2013. First isolation of Mimivirus in a patient with pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America,* 57**,** 34.
SALTER, S. J., COX, M. J., TUREK, E. M., CALUS, S. T., COOKSON, W. O., MOFFATT, M. F., TURNER, P., PARKHILL, J., LOMAN, N. J. & WALKER, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology,* 12**,** 1-12.
SAMBROOK, J. & RUSSELL, D. W. 2001. *Molecular cloning : a laboratory manual,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.
SEGAL, L. N., ALEKSEYENKO, A. V., CLEMENTE, J. C., KULKARNI, R., WU, B., GAO, Z., CHEN, H., BERGER, K. I., GOLDRING, R. M., ROM, W. N., BLASER, M. J. & WEIDEN, M. D. 2013. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome,* 1**,** 19.
SEGAL, L. N., CLEMENTE, J. C., TSAY, J.-C. J. C., KORALOV, S. B., KELLER, B. C., WU, B. G., LI, Y., SHEN, N., GHEDIN, E., MORRIS, A., DIAZ, P., HUANG, L., WIKOFF, W. R., UBEDA, C., ARTACHO, A., ROM, W. N., STERMAN, D. H., COLLMAN, R. G., BLASER, M. J. & WEIDEN, M. D. 2016. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nature microbiology,* 1**,** 16031.
SEGURA-WANG, M., GÖRZER, I., JAKSCH, P. & PUCHHAMMER-STÖCKL, E. 2018. Temporal dynamics of the lung and plasma viromes in lung transplant recipients. *PloS one,* 13.
SHAH, R. J., EMTIAZJOO, A. M., DIAMOND, J. M., SMITH, P. A., ROE, D. W., WILLE, K. M., ORENS, J. B., WARE, L. B., WEINACKER, A., LAMA, V. N., BHORADE, S. M., PALMER, S. M., CRESPO, M., LEDERER, D. J., CANTU, E., ECKERT, G. J., CHRISTIE, J. D. & WILKES, D. S. 2014. Plasma complement levels are associated with primary graft dysfunction

and mortality after lung transplantation. *American journal of respiratory and critical care medicine,* 189**,** 1564-1567.
SHARMA, A. K., LAPAR, D. J., ZHAO, Y., LI, L., LAU, C. L., KRON, I. L., IWAKURA, Y., OKUSA, M. D. & LAUBACH, V. E. 2011. Natural Killer T Cell–derived IL-17 Mediates Lung Ischemia–Reperfusion Injury. *American Journal of Respiratory and Critical Care Medicine,* 183**,** 1539-1549.
SHARMA, P., DIENE, S. M., THIBEAUT, S., BITTAR, F., ROUX, V., GOMEZ, C., REYNAUD-GAUBERT, M. & ROLAIN, J.-M. M. 2013. Phenotypic and genotypic properties of Microbacterium yannicii, a recently described multidrug resistant bacterium isolated from a lung transplanted patient with cystic fibrosis in France. *BMC microbiology,* 13**,** 97.
SHI, B., CHANG, M., MARTIN, J., MITREVA, M., LUX, R., KLOKKEVOLD, P., SODERGREN, E., WEINSTOCK, G. M., HAAKE, S. K. & LI, H. 2015. Dynamic changes in the subgingival microbiome and their potential for diagnosis and prognosis of periodontitis. *mBio,* 6**,** 14.
SHKOPOROV, A., KHOKHLOVA, E. V., FITZGERALD, C. B., STOCKDALE, S. R., DRAPER, L. A., ROSS, P. R. & HILL, C. 2018. ΦCrAss001, a member of the most abundant bacteriophage family in the human gut, infects Bacteroides. *bioRxiv*.
SIMMONDS, P., ADAMS, M. J., BENKŐ, M., BREITBART, M., BRISTER, J. R., CARSTENS, E. B., DAVISON, A. J., DELWART, E., GORBALENYA, A. E., HARRACH, B., HULL, R., KING, A. M., KOONIN, E. V., KRUPOVIC, M., KUHN, J. H., LEFKOWITZ, E. J., NIBERT, M. L., ORTON, R., ROOSSINCK, M. J., SABANADZOVIC, S., SULLIVAN, M. B., SUTTLE, C. A., TESH, R. B., VAN DER VLUGT, R. A. A., VARSANI, A. & ZERBINI, F. M. 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nature reviews. Microbiology,* 15**,** 161-168.
SMITS, S. L., SCHAPENDONK, C. M., VAN BEEK, J., VENNEMA, H., SCHÜRCH, A. C., SCHIPPER, D., BODEWES, R., HAAGMANS, B. L., OSTERHAUS, A. D. & KOOPMANS, M. P. 2014. New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerging infectious diseases,* 20**,** 1218-1222.
SPANDOLE, S., CIMPONERIU, D., BERCA, L. M. & MIHĂESCU, G. 2015. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Archives of virology,* 160**,** 893-908.
STARAS, S. A. S., DOLLARD, S. C., RADFORD, K. W., FLANDERS, D. W., PASS, R. F. & CANNON, M. J. 2006. Seroprevalence of Cytomegalovirus Infection in the United States, 1988–1994. *Clinical Infectious Diseases,* 43**,** 1143-1151.
STEWART, S., FISHBEIN, M. C., SNELL, G. I., BERRY, G. J., BOEHLER, A., BURKE, M. M., GLANVILLE, A., GOULD, K. F., MAGRO, C. & MARBOE, C. C. 2007. Revision of the 1996 working formulation for the standardization of nomenclature in the diagnosis of lung rejection. *The Journal of heart and lung transplantation,* 26**,** 1229-1242.
STOKELL, J. R., GHARAIBEH, R. Z., HAMP, T. J., ZAPATA, M. J., FODOR, A. A. & STECK, T. R. 2015. Analysis of Changes in Diversity and Abundance of the Microbial Community in a Cystic Fibrosis Patient over a Multiyear Period. *Journal of Clinical Microbiology,* 53**,** 237-247.
SUYAMA, M., TORRENTS, D. & BORK, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research,* 34**,** 12.
SUZUKI, Y., CANTU, E. & CHRISTIE, J. D. 2013. Primary Graft Dysfunction. *Seminars in Respiratory and Critical Care Medicine,* 34**,** 305-319.
SZE, M. A., DIMITRIU, P. A., HAYASHI, S., ELLIOTT, W. M., MCDONOUGH, J. E., GOSSELINK, J. V., COOPER, J., SIN, D. D., MOHN, W. W. & HOGG, J. C. 2012. The lung tissue

microbiome in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine,* 185**,** 1073-1080.
TABOADA, B., ESPINOZA, M. A., ISA, P., APONTE, F. E., ARIAS-ORTIZ, M. A., MONGE-MARTÍNEZ, J., RODRÍGUEZ-VÁZQUEZ, R., DÍAZ-HERNÁNDEZ, F., ZÁRATE-VIDAL, F., WONG-CHEW, R. M., FIRO-REYES, V., DEL RÍO-ALMENDÁREZ, C. N., GAITÁN-MEZA, J., VILLASEÑOR-SIERRA, A., MARTÍNEZ-AGUILAR, G., DEL SALAS-MIER, M. C., NOYOLA, D. E., PÉREZ-GÓNZALEZ, L. F., LÓPEZ, S., SANTOS-PRECIADO, J. I. & ARIAS, C. F. 2014. Is there still room for novel viral pathogens in pediatric respiratory tract infections? *PloS one,* 9.
TAKAHASHI, M., ASABE, S., GOTANDA, Y., KISHIMOTO, J., TSUDA, F. & OKAMOTO, H. 2002. TT virus is distributed in various leukocyte subpopulations at distinct levels, with the highest viral load in granulocytes. *Biochemical and biophysical research communications,* 290**,** 242-248.
TAUBENBERGER, J. K., REID, A. H. & FANNING, T. G. 2000. The 1918 influenza virus: A killer comes into view. *Virology,* 274**,** 241-245.
TAYLOR, L. J. & STREBEL, K. 2017. Pyviko: an automated Python tool to design gene knockouts in complex viruses with overlapping genes. *BMC microbiology,* 17**,** 12.
THOM, K. & PETRIK, J. 2007. Progression towards AIDS leads to increased torque teno virus and torque teno minivirus titers in tissues of HIV infected individuals. *Journal of Medical Virology,* 79**,** 1-7.
THURBER, R. V., HAYNES, M., BREITBART, M., WEGLEY, L. & ROHWER, F. 2009. Laboratory procedures to generate viral metagenomes. *Nature protocols,* 4**,** 470-483.
TODD, D. 2000. Circoviruses: immunosuppressive threats to avian species: a review. *Avian pathology : journal of the W.V.P.A,* 29**,** 373-394.
TODD, D. 2004. Avian circovirus diseases: lessons for the study of PMWS. *Veterinary microbiology,* 98**,** 169-174.
TOMARU, Y., TAKAO, Y., SUZUKI, H., NAGUMO, T., KOIKE, K. & NAGASAKI, K. 2011. Isolation and characterization of a single-stranded DNA virus infecting Chaetoceros lorenzianus Grunow. *Applied and environmental microbiology,* 77**,** 5285-5293.
TOMARU, Y., TOYODA, K., SUZUKI, H., NAGUMO, T., KIMURA, K. & TAKAO, Y. 2013. New single-stranded DNA virus with a unique genomic structure that infects marine diatom Chaetoceros setoensis. *Scientific reports,* 3**,** 3337.
VARSANI, A. & KRUPOVIC, M. 2017. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus evolution,* 3.
VARSANI, A. & KRUPOVIC, M. 2018. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Archives of virology,* 163**,** 2005-2015.
VERLEDEN, S. E., RUTTENS, D., VANDERMEULEN, E., BELLON, H., VAN RAEMDONCK, D. E., DUPONT, L. J., VANAUDENAERDE, B. M., VERLEDEN, G. & VOS, R. 2015. Restrictive chronic lung allograft dysfunction: Where are we now? *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation,* 34**,** 625-630.
VIRGIN, H. W. 2014. The virome in mammalian physiology and disease. *Cell,* 157**,** 142-150.
VU, D. L. L., BRIDEVAUX, P. O. O., AUBERT, J. D. D., SOCCAL, P. M. & KAISER, L. 2011. Respiratory viruses in lung transplant recipients: a critical review and pooled analysis of clinical studies. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 11**,** 1071-1078.
WALTON, A. H., MUENZER, J. T., RASCHE, D., BOOMER, J. S., SATO, B., BROWNSTEIN, B. H., PACHOT, A., BROOKS, T. L., DEYCH, E., SHANNON, W. D., GREEN, J. M., STORCH, G. A.

& HOTCHKISS, R. S. 2014. Reactivation of multiple viruses in patients with sepsis. *PloS one,* 9.
WANG, D., URISMAN, A., LIU, Y.-T., SPRINGER, M., KSIAZEK, T. G., ERDMAN, D. D., MARDIS, E. R., HICKENBOTHAM, M., MAGRINI, V., ELDRED, J., LATREILLE, P. J., WILSON, R. K., GANEM, D. & DERISI, J. L. 2003. Viral Discovery and Sequence Recovery Using DNA Microarrays. *PLoS Biology,* 1.
WANG, J., QI, J., ZHAO, H., HE, S., ZHANG, Y., WEI, S. & ZHAO, F. 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Scientific reports,* 3**,** 1843.
WANG, Y., ZHU, N., LI, Y., LU, R., WANG, H., LIU, G., ZOU, X., XIE, Z. & TAN, W. 2016. Metagenomic analysis of viral genetic diversity in respiratory samples from children with severe acute respiratory infection in China. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases,* 22**,** 4580-4589.
WATSON, C. J. E. & DARK, J. H. 2012. Organ transplantation: historical perspective and current practice. *BJA: British Journal of Anaesthesia,* 108.
WEIGT, S. S., ELASHOFF, R. M., HUANG, C., ARDEHALI, A., GREGSON, A. L., KUBAK, B., FISHBEIN, M. C., SAGGAR, R., KEANE, M. P., SAGGAR, R., LYNCH, J. P., ZISMAN, D. A., ROSS, D. J. & BELPERIO, J. A. 2009. Aspergillus colonization of the lung allograft is a risk factor for bronchiolitis obliterans syndrome. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 9**,** 1903-1911.
WIELAND, E., OLBRICHT, C. J., SÜSAL, C., GURRAGCHAA, P., BÖHLER, T., ISRAELI, M., SOMMERER, C., BUDDE, K., HARTMANN, B., SHIPKOVA, M. & OELLERICH, M. 2010. Biomarkers as a tool for management of immunosuppression in transplant patients. *Therapeutic drug monitoring,* 32**,** 560-572.
WILLNER, D., FURLAN, M., HAYNES, M., SCHMIEDER, R., ANGLY, F. E., SILVA, J., TAMMADONI, S., NOSRAT, B., CONRAD, D. & ROHWER, F. 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PloS one,* 4.
WILLNER, D., HAYNES, M. R., FURLAN, M., HANSON, N., KIRBY, B., LIM, Y. W., RAINEY, P. B., SCHMIEDER, R., YOULE, M., CONRAD, D. & ROHWER, F. 2012a. Case studies of the spatial heterogeneity of DNA viruses in the cystic fibrosis lung. *American journal of respiratory cell and molecular biology,* 46**,** 127-131.
WILLNER, D., HAYNES, M. R., FURLAN, M., SCHMIEDER, R., LIM, Y. W., RAINEY, P. B., ROHWER, F. & CONRAD, D. 2012b. Spatial distribution of microbial communities in the cystic fibrosis lung. *The ISME journal,* 6**,** 471-474.
WILLNER, D. L., HUGENHOLTZ, P., YERKOVICH, S. T., TAN, M. E., DALY, J. N., LACHNER, N., HOPKINS, P. M. & CHAMBERS, D. C. 2013. Reestablishment of recipient-associated microbiota in the lung allograft is linked to reduced risk of bronchiolitis obliterans syndrome. *American journal of respiratory and critical care medicine,* 187**,** 640-647.
WOOD, D. E. & SALZBERG, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology,* 15.
WOOTTON, S. C., KIM, D. S., KONDOH, Y., CHEN, E., LEE, J. S., SONG, J. W., HUH, J. W., TANIGUCHI, H., CHIU, C., BOUSHEY, H., LANCASTER, L. H., WOLTERS, P. J., DERISI, J., GANEM, D. & COLLARD, H. R. 2011. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine,* 183**,** 1698-1702.
WOROBEY, M. 2000. Extensive homologous recombination among widely divergent TT viruses. *Journal of virology*.

WYLIE, K. M., MIHINDUKULASURIYA, K. A., SODERGREN, E., WEINSTOCK, G. M. & STORCH, G. A. 2012. Sequence analysis of the human virome in febrile and afebrile children. *PloS one,* 7.

WYLIE, K. M., MIHINDUKULASURIYA, K. A., ZHOU, Y., SODERGREN, E., STORCH, G. A. & WEINSTOCK, G. M. 2014. Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC biology,* 12**,** 71.

YANG, J., YANG, F., REN, L., XIONG, Z., WU, Z., DONG, J., SUN, L., ZHANG, T., HU, Y., DU, J., WANG, J. & JIN, Q. 2011. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *Journal of clinical microbiology,* 49**,** 3463-3469.

YANG, Z., SHARMA, A. K., LINDEN, J., KRON, I. L. & LAUBACH, V. E. 2009. CD4+ T lymphocytes mediate acute pulmonary ischemia–reperfusion injury. *The Journal of Thoracic and Cardiovascular Surgery,* 137**,** 695-702.

YATES, B., MURPHY, D. M., FORREST, I. A., WARD, C., RUTHERFORD, R. M., FISHER, A. J., LORDAN, J. L., DARK, J. H. & CORRIS, P. A. 2005. Azithromycin reverses airflow obstruction in established bronchiolitis obliterans syndrome. *American journal of respiratory and critical care medicine,* 172**,** 772-775.

YOUNG, J. C., CHEHOUD, C., BITTINGER, K., BAILEY, A., DIAMOND, J. M., CANTU, E., HAAS, A. R., ABBAS, A., FRYE, L., CHRISTIE, J. D., BUSHMAN, F. D. & COLLMAN, R. G. 2015. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 15**,** 200-209.

YUSEN, R. D., EDWARDS, L. B., KUCHERYAVAYA, A. Y., BENDEN, C., DIPCHAND, A. I., GOLDFARB, S. B., LEVVEY, B. J., LUND, L. H., MEISER, B., ROSSANO, J. W. & STEHLIK, J. 2015. The Registry of the International Society for Heart and Lung Transplantation: Thirty-second Official Adult Lung and Heart-Lung Transplantation Report—2015; Focus Theme: Early Graft Failure. *The Journal of Heart and Lung Transplantation,* 34**,** 1264-1277.

ZANINI, F. & NEHER, R. A. 2013. Quantifying selection against synonymous mutations in HIV-1 env evolution. *Journal of virology,* 87**,** 11843-11850.

ZANOTTA, N., MAXIMOVA, N., CAMPISCIANO, G., DEL SAVIO, R., PIZZOL, A., CASALICCHIO, G., BERTON, E. & COMAR, M. 2015. Up-regulation of the monocyte chemotactic protein-3 in sera from bone marrow transplanted children with torquetenovirus infection. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 63**,** 6-11.

ZHANG, X., ZHANG, D., JIA, H., FENG, Q., WANG, D., LIANG, D., WU, X., LI, J., TANG, L., LI, Y., LAN, Z., CHEN, B., LI, Y., ZHONG, H., XIE, H., JIE, Z., CHEN, W., TANG, S., XU, X., WANG, X., CAI, X., LIU, S., XIA, Y., LI, J., QIAO, X., AL-AAMA, J. Y., CHEN, H., WANG, L., WU, Q.-J. J., ZHANG, F., ZHENG, W., LI, Y., ZHANG, M., LUO, G., XUE, W., XIAO, L., LI, J., CHEN, W., XU, X., YIN, Y., YANG, H., WANG, J., KRISTIANSEN, K., LIU, L., LI, T., HUANG, Q., LI, Y. & WANG, J. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature medicine,* 21**,** 895-905.

ZHANG, Y., LI, F., CHEN, X., SHAN, T.-L. L., DENG, X.-T. T., DELWART, E. & FENG, X.-P. P. 2017. Detection of a new species of torque teno mini virus from the gingival epithelium of patients with periodontitis. *Virus genes,* 53**,** 823-830.

ZHAO, J., SCHLOSS, P. D., KALIKIN, L. M., CARMODY, L. A., FOSTER, B. K., PETROSINO, J. F., CAVALCOLI, J. D., VANDEVANTER, D. R., MURRAY, S., LI, J. Z., YOUNG, V. B. & LIPUMA, J. J. 2012. Decade-long bacterial community dynamics in cystic fibrosis airways.

Proceedings of the National Academy of Sciences of the United States of America, 109**,** 5809-5814.
ZHENG, H., YE, L., FANG, X., LI, B., WANG, Y., XIANG, X., KONG, L., WANG, W., ZENG, Y., YE, L., WU, Z., SHE, Y. & ZHOU, X. 2007. Torque teno virus (SANBAN isolate) ORF2 protein suppresses NF-kappaB pathways via interaction with IkappaB kinases. *Journal of virology,* 81**,** 11917-11924.
ZHONG, S., YEO, W., TANG, M., LIU, C., LIN, X.-R. R., HO, W. M., HUI, P. & JOHNSON, P. J. 2002. Frequent detection of the replicative form of TT virus DNA in peripheral blood mononuclear cells and bone marrow cells in cancer patients. *Journal of medical virology,* 66**,** 428-434.
ZOLL, J., RAHAMAT-LANGENDOEN, J., AHOUT, I., DE JONGE, M. I., JANS, J., HUIJNEN, M. A., FERWERDA, G., WARRIS, A. & MELCHERS, W. J. 2015. Direct multiplexed whole genome sequencing of respiratory tract samples reveals full viral genomic information. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology,* 66**,** 6-11.
ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research,* 31**,** 3406-3415.