Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Cory Arrouzet Date

Association Between Population-Level Secretor Status Prevalence and Genogroup 2 Genotype 4 Norovirus Prevalence

By

Cory Joseph Arrouzet

Master of Public Health

Department of Epidemiology

_____

Ben Lopman, MSc, PhD
Committee Chair

Association Between Population-Level Secretor Status Prevalence and Genogroup 2 Genotype 4 Norovirus Prevalence

By

Cory Joseph Arrouzet

Bachelor of Science
Santa Clara University, 2015

Faculty Thesis Advisor: Ben Lopman, MSc, PhD

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2018

Abstract

Association Between Population-Level Secretor Status Prevalence and Genogroup 2 Genotype 4 Norovirus Prevalence

By Cory Joseph Arrouzet

Noroviruses are the leading cause of acute gastroenteritis worldwide. The GII.4 genotype has been the dominant genotype worldwide since its global emergence in 1994. Individuals with a functional FUT2 gene, termed secretors, have consistently shown increased susceptibility to certain noroviruses, including GII.4. To investigate the association between population level secretor status and GII.4 prevalence, we conducted a systematic literature review of studies reporting norovirus genotypes in outbreaks or sporadic cases for which we had data on secretor prevalence from a prior systematic review. 2528 references were identified from searching the literature. 219 genotype and 112 secretor studies with data from 38 countries were included in the final analysis. Using inverse variance-weighted linear regression modeling, we observed a significant, positive association between country-level secretor and GII.4 prevalence. An increase in secretor prevalence is associated with a 0.60% (95% CI: 0.11, 1.08) increase in GII.4 prevalence, controlling for study type, age of patients, pandemic variant period, and Human Development Index. These results have implications for vaccine interventions and future research to understand the effects of population level host genetic heterogeneity on country level genotype distributions.

Association Between Population-Level Secretor Status Prevalence and Genogroup 2 Genotype 4 Norovirus Prevalence

By

Cory Joseph Arrouzet

Bachelor of Science
Santa Clara University, 2015

Faculty Thesis Advisor: Ben Lopman, MSc, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health
in Epidemiology

2018

Acknowledgements

I would like to thank my thesis advisor, Dr. Ben Lopman for his constant guidance, support, and mentorship over the past year, and for never losing faith in me when the systematic review dragged on forever. I would also like to thank Karen Ellis, with whom I endured a marathon of data collection for this thesis, for her endless support and camaraderie. Also, thank you to Molly Steele for her expertise and advice throughout this process. This project would not have been possible without any of them.

Furthermore, thank you to Anita Kambhampati and Yingxi Chen for generously providing the secretor status data, to Drs. Janet Mans, Leesa Bruggink, Miao Jin, Shobha Chitambar, Yasumasa Iwatani, Aksara Thongprachum, and Ushijima Hiroshi for generously providing genotype data at the request of an email from across the globe, and to all of the norovirus researchers around the world whose published works were included in our analysis.

Thank you to my parents, Mark and Nancy Arrouzet, and my brother, Kyle Arrouzet, for their ever-undying love and support through all of my endeavors. Finally, thank you to my fellow MPH candidates, all of my professors, and all of the other Rollins faculty and staff for creating the ideal educational setting to learn and grow into public health professional.

Table of Contents

Chapter I: Background & Literature Review

Introduction

Norovirus was first identified from a school outbreak of acute, nonbacterial gastroenteritis in Norwalk, Ohio in 1968 (1). By the 1990s, noroviruses were recognized as the predominant cause of epidemic acute gastroenteritis (AGE) worldwide (2). They are genetically diverse viruses that have evolved over time with new pandemic variants emerging every two to four years. Individuals with a particular genetic mutation, termed secretors, have shown greater genetic susceptibility to some norovirus genotypes compared to nonsecretors. Prevalence of secretor status varies greatly by ethnicity and geography. At the individual level secretor status mediates susceptibility to certain noroviruses, including the most common GII.4 genotype, and therefore influences the distribution of genotypes from which an individual can be infected. It is unclear, however, how the prevalence of secretor status at the population level influences the genotype distribution of circulating and epidemic noroviruses.

Burden and Epidemiology

Burden

Although diarrheal disease mortality continues to decline globally, diarrheal diseases are the seventh leading cause of global mortality and the number two cause of infectious disease deaths. Diarrheal diseases are also the fifth leading cause of disability adjusted life years (DALYs) globally. Among children under the age of five, they are the fifth leading cause of death and DALYs (3). Within diarrheal diseases, noroviruses are the leading cause of both sporadic cases and outbreaks of AGE globally, accounting for 18%

(95% CI: 17-20%) of all cases of gastroenteritis according to a recent systematic literature review on norovirus prevalence by Ahmed et al (4). In this study, the prevalence of norovirus among diarrheal diseases was the same for both children aged 0-4 and for cases older than 4 years. The incidence of diarrheal disease in general is highest among young children, and therefore norovirus incidence is also higher among young children (5). Each year in the United States, noroviruses cause approximately 570–800 deaths, 56,000–71,000 hospitalizations, 400,000 ED visits, 1.7–1.9 million outpatient visits, and 19–21 million illnesses (6). Seroprevalence studies have demonstrated that up to 50% of children under 5 years of age and up to 100% of adults have been exposed to at least one norovirus during their lifetime (7).

Epidemiology

Due to their extremely low infectious dose, noroviruses can be easily transmitted in droplets from person to person via the fecal-oral route, fecal-vomitus route, through the environment via fomites, and through contaminated food and water (5, 8). While there are noroviruses that circulate in animal populations, humans are the only known reservoir for human noroviruses, and zoonotic transmission of norovirus has not been observed (9). Viral shedding precedes symptoms in 30% of exposed individuals enabling transmission in asymptomatic individuals which is of particular concern among food handlers and close contacts (8). Additionally, noroviruses are highly environmentally stable, contributing to transmission in a range of settings (8, 10).

Norovirus outbreaks occur in people of all ages suggesting immunity is either not long-lasting or not cross protective against different virus strains (11). Outbreaks occur

primarily in nursing homes, other healthcare settings, cruise ships, and restaurants (8). In the United States, outbreaks are most common in long term care facilities and less common in acute care settings while the frequency of outbreaks in long term and acute care settings is roughly equal in other developed nations (5). Groups at increased risk of norovirus infection include young children, elderly, travelers, soldiers, close contacts of people ill with norovirus, and immunocompromised patients (12).

In the United States, endemic norovirus disease occurs year round, with seasonal peaks in winter months (6). Formerly referred to as "winter vomiting disease," norovirus outbreaks and sporadic cases have shown a distinct seasonality in many parts of the globe. The highest rates of transmission occur in the winter months (December-February in the Northern Hemisphere and June-August in the Southern Hemisphere) with the lowest rates occurring during the summer months (13).

Natural History

Norovirus symptoms include vomiting and diarrhea and typically resolve within a couple days, although chronic disease has been observed in immunocompromised patients (14). Most norovirus infections are asymptomatic or cause mild symptoms, but norovirus also causes considerable amounts of severe disease (5). In a systematic review published in 2015, Devasia et al. estimated an average incubation period for norovirus infection of 32.8 (95% CI: 30.9 – 34.6) hours and an average symptomatic period of 44.2 (95% CI: 38.9-50.7) hours (15). The authors also found no meaningful associations between both incubation period and symptom duration and environment, host, and pathogen factors. Other studies have estimated similar incubation periods for norovirus (16).

Vaccines

Currently, there is no licensed vaccine for noroviruses, but multiple trials are ongoing to evaluate the safety and effectiveness of candidate virus-like particle (VLP) vaccines (17). Positive immunogenicity and safety results of VLP vaccine randomized clinical trials provide promise of norovirus vaccines in the future (18). Further research is required to know whether vaccines will provide protection to strains not included in the vaccine, how long vaccine-induced immunity will last, and whether vaccines will be able to keep pace with the rapid evolution of noroviruses (17).

Genetic and Molecular Epidemiology

Genome

Noroviruses are nonenveloped, positive sense, single-stranded RNA viruses in the *Caliciviridae* family (19, 20). The norovirus genome contains three open reading frames (ORF1-ORF3). ORF1 contains the polymerase gene that encodes the RNA-dependent RNA polymerase (RdRp) enzyme responsible for viral replication. ORF2 contains the capsid gene that encodes the major structural protein VP1, and ORF3 encodes a minor structural protein VP2 (19).

Detection and Typing Methods

Noroviruses are typically detected using reverse-transcriptase polymerase chain reactions (RT-PCR and are categorized into genogroups and genotypes by sequencing the RT-PCR products from either ORF1, ORF2, or the ORF1/ORF2 junction. Recombination events

are being increasingly recognized, and thus noroviruses are more frequently categorized according to both the polymerase and capsid genes in recent years (21).

Genogroups

Noroviruses are grouped into at least seven genogroups (GI-GVII) based on ORF1 and ORF2 sequences with strains from genogroups GI, GII, and GIV causing disease in humans (19). GII is the most prevalent genogroup globally followed by GI, with GIV rarely causing disease in humans. GII viruses are responsible for over 95% of sporadic infections worldwide while GI cause approximately 4% (22). Among outbreaks, GII viruses account for about 90%, GI cause less than 10% and GIV cause less than 0.1% (23)

Genotypes

Within genogroups, human noroviruses are categorized into at least nine GI genotypes, 22 GII genotypes, and one GIV genotype based upon the polymerase and/or capsid genes (21, 24). GII.4 is the most prevalent genotype causing a majority of sporadic infections globally followed by GII.3 as the second most prevalent genotype (22). Other previously less common GII genotypes have emerged in various recent seasons and have become more prevalent globally including GII.16, GII.21, GII.13, and GII.17 (22, 25, 26). In the United States, GII.4 cause the majority of outbreaks while GI.3, GI.5, GII.2, GII.3, GII.6, GII.13, and GII.17 caused over 5% of outbreaks in at least one season each between 2013 and 2016 (27).

Differences in Epidemiology by Genotype

Studies have shown that GII.4 viruses are transmitted from person to person more often than GI and GII.non-4 viruses and are also detected more often in healthcare settings than GI and GII.non-4 genotypes (23). Patients under five are more often infected by non-GII.4 noroviruses than patients over five years of age. Likewise, GII.4 is more common in healthcare settings compared to community settings where more genotype diversity is common (28). Between 2013 and 2016 in the U.S. norovirus winter seasonality was primarily driven by GII.4 winter seasonality, although other genotypes peaked in winter months in some seasons as well (27). GII.4 outbreaks are typically associated with higher mortality and hospitalization rates than other genotypes (29). Devasia et al observed longer, yet insignificant, durations of symptoms in GII.4 infections compared to non-GII.4 infections. They observed no significant difference in incubation periods by genotype (15).

Evolution

Noroviruses are constantly evolving with new genotypes emerging every few years. GII.4 noroviruses have been the most prevalent genotype globally since their emergence in 1994. Novel GII.4 variants have emerged every two to four years and replaced previously prevalent strains as the dominant norovirus globally (2). There have been seven pandemic GII.4 variants named for their location and year of origin: US 95/96, Farmington Hills 2002, Hunter 2004, Yerseke 2006a, Den Haag 2006b, New Orleans 2009, and Sydney 2012 (19). Each variant descended directly from the previously dominant variant until 2006 when two new variants emerged, one descending from the 2004 variant (2006a) and

one from the 2002 variant (2006b) (2). There are various explanations for the rapid evolutionary pattern of GII.4 variants. One hypothesis is that population herd immunity against GII.4 selectively drives antigenic shift of norovirus binding sites, allowing them to bind to different receptors in humans (2). A second reason for the evolution of GII.4 relates to herd immunity duration. Given the ability of new GII.4 variants to spread globally in under three months and evidence suggesting immunity lasts 6-12 months, herd immunity may force rapid evolution of GII.4 variants before they reach an equilibrium state allowing new variants to effectively increase their susceptible population sizes (2). Intense selective pressure placed on GII.4 viruses could lead to rapid evolution in binding capabilities to avoid human immune responses. In the face of short term host immunity, this selective pressure may explain the explosive pandemic nature of emergent GII.4 variants (30, 31).

In recent norovirus seasons, other genotypes have caused large epidemics in some countries and have replaced GII.4 as the dominant genotype during particular norovirus seasons. GII.17 noroviruses emerged as the predominant genotype, rapidly replacing GII.4 in some parts of Asia including Hong Kong, South Korea, and China during the 2014-2015 season (25, 26, 32-34). During the 2016-2017 season, two novel GII polymerase genotype 16 (GII.P16) recombinants, GII.P16-GII.2 and GII.P16-GII.4_Sydney, emerged as substantial causes of norovirus hospitalizations and outbreaks various regions of the world including Asia, Europe, and the United States.

Host Genetic Susceptibility and Resistance

Norovirus Binding Biology

Noroviruses bind to histo-blood group antigens (HBGAs) on cellular surfaces of gastroduodenal epithelial cells (35). HBGAs are oligosaccharides found in gut, urinary tract, and respiratory cells, as well as other bodily fluids including milk, blood, and saliva (36). Two fucosyltransferase enzymes encoded by the fucosyltransferase-1 (FUT1) and fucosyltransferase-2 (FUT2) genes are responsible for the synthesis of certain HBGAs in humans. Individuals without a function FUT2 gene are termed non-secretors and lack specific HBGAs that certain noroviruses bind to on certain cell surfaces including gastroduodenal epithelial cells compared to secretors who have functional FUT2 genes (37).

Association Between Secretor Status and Genotype Susceptibility

Several studies have demonstrated an association between secretor status and genotype susceptibility, confirming a genetic component to norovirus susceptibility for some genotypes (38). For the prototype Norwalk Virus (Genotype GI.1), binding is mediated by the secretor phenotype (35). In both human challenge studies and observational outbreak studies, secretors have shown increased susceptibility to GII.4 noroviruses while nonsecretors display strong genetic resistance to GII.4 norovirus disease (39, 40). In a systematic review and meta-analysis on the association between secretor status and susceptibility published in 2016, Kambhampati et al. found that secretors had 4.2 times the odds of norovirus infection compared with nonsecretors (95% CI, 2.3-7.9). and were

9.9 times more frequently infected with GII.4 (95% CI, 3.9-24.8) compared to nonsecretors (41).

Nonsecretors, however, have not been observed to be absolutely resistant to GII.4 noroviruses. In a GII.4 human challenge study published in 2012, Frenck et al. found one of 17 nonsecretors became ill with diarrhea and vomiting after being exposed to GII.4 while another shed the virus for a day without displaying symptoms (42). Additionally, in a prospective study in Burkina Faso conducted in 2009 and 2010, Nordgren et al. observed a symptomatic GII.4 infection in a nonsecretor child while Carlsson et al. found one symptomatic GII.4 infection among 17 nonsecretor residents in an outbreak study conducted in a Spanish nursing home in 2004 (43, 44). Likewise, Liu et al. observed one GII.4 infection among 18 nonsecretor children in prospective study in China conducted from 2009 to 2011 (45).

Other genotypes have shown associations between secretor status and susceptibility. Along with GII.4, GII.3 noroviruses were detected significantly more often among secretors by Liu et al (45). Nordgren et al. also observed significantly fewer GII infections among nonsecretors compared to secretor-positive children (44). A 2002 hospital-based study by Thorven et al. in Sweden also showed resistance to GII nosocomial infections and outbreaks among nonsecretors (46). However, these associations are not universal across all norovirus genotypes. For other genotypes including the GII.2 Snow Mountain virus, secretors have shown equal levels of susceptibility as nonsecretors (38, 47). This finding suggests that for some noroviruses, susceptibility is not mediated by secretor status. Additionally, observational studies have shown that nonsecretors are infected with non-GII.4 noroviruses significantly more often

than secretors suggesting nonsecretors have potentially increased susceptibility to certain genotypes (38, 39).

Distribution of Secretor Status by Ethnicity

Prevalence of the secretor phenotype varies across different ethnic groups, and therefore also varies geographically. In some European and African populations, secretor status prevalence is approximately 80% (38). In other European nations including the United Kingdom and Ireland, secretor prevalence may be closer to 70% (48-52). In some eastern and southeastern Asian nations including China, India, and Pakistan, secretor prevalence is less than 70% (53-55). Studies have shown secretor prevalences around 90% in some Latin American countries including Nicaragua and Ecuador, while others are about 80% including Brazil and Chile (39, 56-58).

Summary

In summary, noroviruses are highly genetically diverse viruses that cause acute gastrointestinal symptoms. The most common genotype, GII.4, emerged globally in 1994. Since then, new pandemic GII.4 variants have emerged every two to four years. Susceptibility to certain norovirus genotypes, including GII.4, is mediated by the presence or absence of a functional FUT2 gene which determines the secretor phenotype. Secretors exhibit increased susceptibility to GII.4 noroviruses compared to nonsecretors. Noroviruses evolve rapidly with recombination events being increasingly recognized with improved genotyping and diagnostic techniques. Due to individual level immunity

to certain genotypes among nonsecretors, it is hypothesized that country level secretor prevalence can influence a country's genotype distribution.

Chapter II: Manuscript

Title

Association Between Population-Level Secretor Status Prevalence and Genogroup 2 Genotype 4 Norovirus Prevalence

Authors

Cory J. Arrouzet; Karen Ellis; Anita Kambhampati; Yingxi Chen; Molly Steele; Ben Lopman

Abstract

Noroviruses are the leading cause of acute gastroenteritis worldwide. The GII.4 genotype has been the dominant genotype worldwide since its global emergence in 1994. Individuals with a functional FUT2 gene, termed secretors, have consistently shown increased susceptibility to certain noroviruses, including GII.4. To investigate the association between population level secretor status and GII.4 prevalence, we conducted a systematic literature review of studies reporting norovirus genotypes in outbreaks or sporadic cases for which we had data on secretor prevalence from a prior systematic review. 2528 references were identified from searching the literature. 219 genotype and 112 secretor studies with data from 38 countries were included in the final analysis. Using inverse variance-weighted linear regression modeling, we observed a significant, positive association between country-level secretor and GII.4 prevalence. An increase in secretor prevalence is associated with a 0.60% (95% CI: 0.11, 1.08) increase in GII.4 prevalence, controlling for study type, age of patients, pandemic variant period, and Human Development Index. These results have implications for vaccine interventions

and future research to understand the effects of population level host genetic heterogeneity on country level genotype distributions.

Introduction

Noroviruses are the leading cause of acute viral gastroenteritis worldwide. Noroviruses are nonenveloped, positive sense, single-stranded RNA viruses in the *Caliciviridae* family (19, 20). The norovirus genome contains three open reading frames (ORF1-ORF3). ORF1 contains the polymerase gene that encodes the RNA-dependent RNA polymerase (RdRp) enzyme responsible for viral replication. ORF2 contains the capsid gene that encodes the major structural protein VP1, and ORF3 encodes a minor structural protein VP2 (19). Noroviruses are grouped into at least seven genogroups (GI-GVII) based on ORF1 and ORF2 sequences with strains from genogroups GI, GII, and GIV causing disease in humans (19). GII is the most prevalent genogroup globally followed by GI, with GIV rarely causing disease in humans. Within genogroups, human noroviruses are categorized into at least nine GI genotypes, 22 GII genotypes, and one GIV genotype based upon the polymerase and/or capsid genes (21, 24). GII.4 is the most prevalent genotype causing a majority of sporadic infections globally followed by GII.3 as the second most prevalent genotype (22).

Noroviruses bind to histo-blood group antigens (HBGAs) on cellular surfaces of gastroduodenal epithelial cells (35). HBGAs are oligosaccharides found in gut, urinary tract, and respiratory cells, as well as other bodily fluids including milk, blood, and saliva (36). Two fucosyltransferase enzymes encoded by the fucosyltransferase-1 (FUT1) and fucosyltransferase-2 (FUT2) genes are responsible for the synthesis of certain HBGAs in

humans. Individuals without a functional FUT2 gene are termed non-secretors and lack specific HBGAs that certain noroviruses bind to on certain cell surfaces including gastroduodenal epithelial cells compared to secretors who have functional FUT2 genes (37).

GII.4 viruses evolve rapidly and are able to bind to a variety of HBGAs. One explanation for GII.4 evolution is that population herd immunity against GII.4 selectively drives antigenic shift of norovirus binding sites, allowing them to bind to different receptors in humans (2). A second reason for the evolution of GII.4 relates to herd immunity duration. Given the ability of new GII.4 variants to spread globally in under three months and evidence suggesting immunity lasts 6-12 months, herd immunity may force rapid evolution of GII.4 variants before they reach an equilibrium state allowing new variants to effectively increase their susceptible population sizes (2). Intense selective pressure placed on GII.4 viruses could lead to rapid evolution in binding capabilities to avoid human immune responses. Coupled with short term host immunity, this selective pressure may explain the explosive pandemic nature of emergent GII.4 variants (30, 31).

Several studies have demonstrated an association between secretor status and genotype susceptibility, confirming a genetic component to norovirus susceptibility for some genotypes (38). In both human challenge studies and observational outbreak studies, secretors have shown increased susceptibility to GII.4 noroviruses while nonsecretors display strong genetic resistance to GII.4 norovirus disease (39, 40). In a systematic review and meta-analysis on the association between secretor status and susceptibility published in 2016, Kambhampati et al. found that secretors had 4.2 times

the odds of norovirus infection compared with nonsecretors (95% CI, 2.3-7.9). and were 9.9 times more frequently infected with GII.4 (95% CI, 3.9-24.8) compared to nonsecretors (41). Nonsecretors, however, have not been observed to be absolutely resistant to GII.4 noroviruses.

To investigate the association between human secretor prevalence and norovirus GII.4 prevalence, we developed a dataset by conducting a systematic literature review on studies reporting genotype distributions and synthesized it with estimates of national secretor prevalence by updating Kambhampati et al. Our hypothesis was that greater secretor status prevalence is associated with increased GII.4 prevlence.

Methods

Genotype Systematic Literature Review

We searched the literature in PubMed for studies published in English starting in 1994 using the search terms "norovirus" coupled with any of "surveillance," "genotype," or "strain." All identified references from the search were initially screened by title. Included titles were then screened by abstract. We subsequently reviewed all included abstracts by full text and extracted data from all included full text articles. Individual outbreak reports, studies of environmental samples that did not include any human samples, and studies in animals were all excluded. Studies that selected only specific genotypes for genetic or epidemiologic analyses without reporting a genotype distribution were also excluded. Finally, studies that did not report norovirus genotypes and studies from which a distribution of genotypes that at least included a proportion of GII.4 among all genotypes could not be determined were excluded. Included studies were ones those in

which a GII.4 prevalence could be calculated that came from countries from which secretor prevalence was also available. Genotyping studies from countries without secretor data were also excluded.

We used Evidence Partners DistillerSR (Ottawa, Canada) systematic review software to automate metadata collection and enter data from the literature into a database. We extracted genogroup, genotype, and GII.4 variant counts or percent from every included full text review. For studies where data of interest were only contained in figures, we requested data from corresponding authors by email. Other variables collected from the studies included sampling unit (i.e. whether the study included outbreaks, sporadic cases, or both), setting, age distribution, study period, typing method, and typing region. In cases where data were presented stratified by time period, typing region, or setting, multiple data entries were made from the same study.

Duplicated data was confirmed by reviewing full text articles of any observations with common within-country locations and overlapping time periods. Papers that presented different analyses using data that came from the same study were considered duplicated. Generally, when duplicated data was identified, larger studies were kept unless there were reasons to keep smaller ones. Smaller studies were included over larger ones if they had better data quality, were more representative of the population data, or were published in multiple installments over a larger period of time than the larger study.

Secretor Status Prevalence Data

Secretor status prevalence data was collected and updated prior to this study from the systematic literature review described in Kambhampati et al (41). Search terms used in

this review included “histo blood group antigens”, “secretor”, “FUT2”, and other related terms. Included studies reported data secretor status prevalence. Studies that reported results of secondary data analysis or nonhuman data, were written in a language other than English, or that lacked a control group were excluded. To combat potential selection bias, studies were carefully selected to reduce the possibility of biased data. For example, studies assessing secretor types in norovirus-positive or rotavirus-positive subjects were systematically excluded as these studies would likely have inflated secretor prevalence compared to the general population. Furthermore, any studies where selection was based on a factor that is likely related to secretor status were also excluded. Country-level secretor status prevalence estimates were calculated by summing the number of secretors across every included study for each country and dividing by the summed study populations across all included studies within that country. This aggregated secretor dataset was then merged with the genotype dataset assembled from this current systematic literature review.

Data Management

All studies that reported genotypes in percent were converted to counts by dividing each genotype percent by 100 and multiplying by the sample size. GII.4 prevalence was calculated as the count of outbreaks or sporadic cases genotyped as GII.4 divided by the total number of genotyped outbreaks or sporadic cases. Both GII.non-4 and GI norovirus prevalences were calculated by summing the counts of all GII.non-4 and GI genotypes respectively and dividing by the total sample size of the observation. Because GI and GII.non-4 could not be differentiated in some studies, GII.non-4 and GI prevalences were

only calculated for 349 observations where the GI and GII.non-4 were presented separately.

Covariates assessed in the analysis were study type, age category, pandemic variant period, WHO region, Human Development Index (HDI), genotyping method, and genotyping region. The study type variable combined sampling unit (outbreaks vs. sporadic cases) and setting into one classification scheme that included outbreak studies separately from different settings of sporadic case studies. Study population ages were categorized into ≤5 years, >5 years, and mixed. Observations were classified according to time period of dominant pandemic GII.4 variant such that each time period category is mutually exclusive. Time period categories used were 1994-2001, 2002-2003, 2004-2005, 2006-2008, 2009-2011, 2012-2017, and observations whose study period spanned multiple pandemic periods (19). HDI data were obtained from the United Nations Human Development Programme (UNDP) website and both the HDI categories and numeric indices were merged with the secretor and genotype data by country after changing alternative country names(59). Likewise, WHO region classifications were obtained from the WHO website for each country in the dataset and merged with the full dataset by country (60). Typing method was categorized as RT-PCR, other methods, and unknown. Typing region categories used are ORF1, ORF2, ORF1 & ORF2, ORF1/ORF2 overlap only, and unknown. For this categorization scheme, studies that used both ORF1 and the ORF1/ORF2 overlap or ORF2 and the ORF1/ORF2 overlap were categorized as ORF1 and ORF 2 respectively.

A dataset aggregated by country created from the genotype and secretor prevalence datasets was used for mapping and scatter plot generation. For the

aggregation, country-level GII.4 prevalence was calculated as the sum of all GII.4 counts across each country divided by the sum of sample sizes across each country. Maps were created in ESRI ArcGIS 10.5.1 and projected using the WGS 1984 World Mercator projection.

Data Analysis

All statistical analysis was performed using SAS 9.4 (Cary, NC). Multiple observations for some studies were included due to time period, setting, or typing region stratifications presented in the study. Medians and inter-quartile ranges (IQR) were calculated for GII.4, GII.non-4, and GI prevalences across all observations and for each predictor variable category or quartile.

We used weighted ordinary least squares regression to investigate the association between country-level secretor prevalence and observation GII.4 prevalence. All regression models were weighted by inverse variance of the GII.4 prevalence outcome so that weights were proportional to observation sample size with larger studies receiving more weight than smaller studies. Inverse variance weights were calculated using the inverse of the binomial proportion variance formula:

$$1/variance = n/[p(1-p)]$$

where p is the observation GII.4 prevalence and n is the observation sample size. For observations where GII.4 prevalence was 1 or 0, prevalences were adjusted by adding or subtracting 0.05 in order to calculate a defined, positive inverse variance. Univariate regression models were fit for secretor prevalence and each covariate individually to examine each variable's independent association with GII.4 prevalence. Bivariate

regression models were fit to examine the association between secretor prevalence and GII.4 prevalence while controlling for each covariate individually. We used a multiple linear regression model to estimate the association between population secretor prevalence and GII.4 prevalence while controlling for important confounders and predictors of GII.4 prevalence. Associations between categorical predictors were assessed using Chi-square or Fisher's exact tests when Chi-square was not appropriate. Highly associated predictors were not included in the final model together. ANOVAs were conducted to test for hypothesized associations between numerical and categorical covariates to determine if specific variables were highly correlated and should not be included in the final model together. Covariates considered for the final model were hypothesized *a priori* to be confounders or meaningfully changed the secretor status regression coefficient in bivariate compared to univariate analysis were considered for the final model. Among highly correlated covariates, the more important predictors in bivariate analysis were assessed for inclusion in the final model according to R^2 values and changes in the secretor status regression coefficient. We fit the same MLR model and excluded observations where GII.4 prevalence was equal to 1 or 0 to assess the influence of inflated weights at the bounds of the binomial variance equation.

Results

Description of Included Studies

The search identified a total of 2528 references, one of which was duplicated. After screening titles and abstracts, 456 full text articles with potentially relevant information were reviewed. Data from 248 articles were initially included. After screening for

duplicated data, 219 articles representing 38 countries were included (Figure 1). Scotland was included as a separate country from the United Kingdom because we have a separate secretor prevalence estimate and genotype data specifically from Scotland. Secretor status data from 112 studies from the same 38 countries were also included.

Data Description

Due to multiple stratifications, 411 observations were included in the final dataset. Most observations used RT-PCR detection and about half used ORF2 to characterize norovirus strains. Observation sample sizes (number of sporadic cases or outbreaks) ranged from one to 3,960 with a median of 45. Almost half of all observations came from Western Pacific nations with the majority of those coming from China, Japan, South Korea, and Australia. About a quarter of the observations' study periods coincided with global dominance of the Sydney 2012 pandemic GII.4 variant. 118 observations were from outbreak studies (28.7%), with 243 (59.1%) being sporadic case studies and the rest being mixed or unspecified (12.2%). Most included data came from high human development nations according to the UNDP (Table 1).

Unweighted median GII.4 prevalences were higher in outpatient sporadic case studies, studies with subjects over 5 years, and studies that coincide with 2006a and 2006b GII.4 variants. By WHO region, the highest unweighted median GII.4 prevalence was in Europe, and in very high development nations (Table 1).

Aggregated GII.4 prevalence is highest in North American and some European countries (Figure 2). Secretor prevalence is highest the Americas and parts of Europe

while lower secretor prevalences are more common in Southeast Asia (Figure 3). Figure 4 illustrates the relationship between secretor and GII.4 prevalence.

Univariate Analysis

In univariate analysis country-level secretor prevalence was significantly, positively associated with GII.4 prevalence (Table 2). HDI was also positively associated with GII.4 prevalence. Categorical variables that varied significantly with GII.4 prevalence were WHO region, age category, typing region and pandemic variant period, with the latter variable able to explain the most variance ($R^2 = 0.24$).

Bivariate Linear Regressions

Controlling for any covariate included in the analysis individually, secretor prevalence was significantly, positively associated with weighted GII.4 prevalence. Covariates that most affected the association between secretor prevalence and GII.4 prevalence were pandemic variant period, and HDI (Table 3).

Relationships Between Predictors

We hypothesized that certain variables would be highly associated, and thus should not be included in multivariable models together. WHO region and HDI measure similar constructs, and indeed were highly associated (ANOVA $F = 120.1$, $p < 0.01$). Likewise, the sampling unit (outbreaks v. sporadic cases) and setting were highly associated due to study design (Fisher's Exact $p < 0.01$). Therefore, these two variables were combined into one comprehensive study type variable in order to control for differences in

norovirus outbreaks and sporadic cases as well as different settings. Additionally, typing region and pandemic variant period were highly associated as genotyping practices have shifted over time from ORF1 to ORF2 being the most common, to increasingly typing both regions and the ORF1/ORF2 overlap (Fisher's Exact $p < 0.01$).

Multiple Linear Regression

Co-associated predictors were chosen in the final model based on univariate and bivariate regression results. Pandemic variant period was determined to be a more important predictor of GII.4 prevalence based on R^2 and a more important confounder than typing region. According to the bivariate regressions, HDI is a more important confounder than WHO region.

In multiple regression, for every percent increase in population secretor prevalence, there is a corresponding 0.60% increase in GII.4 prevalence (95% CI 0.11%, 1.08%; $R^2 = 0.34$), controlling for study type, pandemic variant period, age category, and HDI (Table 4). Removing age category has no meaningful effect on the secretor-GII.4 relationship ($\beta = 0.57$; 95% CI 0.10, 1.05; $R^2 = 0.34$). We observed similar model results when excluding 1 or 0 GII.4 prevalence values ($\beta = 0.53$; 95% CI 0.05, 1.01; $R^2 = 0.33$). Inpatient and outpatient sporadic studies also were observed to have higher GII.4 prevalence compared to outbreaks. Relative to the most recent pandemic period (2012-17), GII.4 prevalence was lower prior to 2002 ($\beta = -0.09$; 95% CI -0.20, 0.01), but higher from 2002-2011 with 2004-2005 having the greatest prevalence ($\beta = 0.31$; 95% CI 0.19, 0.43).

Discussion

Main Findings

We identified an independent positive association between country-level secretor status prevalence and GII.4 norovirus prevalence. This finding is consistent with our previously-stated hypothesis that populations with higher proportion of secretors will primarily be infected with GII.4; a higher frequency of non-secretor alleles is associated with more norovirus diversity (38). Second, we found a significant positive correlation between HDI and GII.4 prevalence with higher HDI countries having higher GII.4 prevalence in both the univariate and bivariate model with secretor prevalence (Tables 2-3). Third, GII.4 prevalence is significantly different across pandemic periods lowest prevalence prior to 2002 (43%) and higher prevalence after 2002. The Hunter variant (2004-2005) and Den Haag and Yerseke variant (2006-2008) periods had the highest GII.4 prevalence at 84% and 87%, respectively (Table 2).

Limitations

This analysis is subject to at least four limitations. First, as a secondary analysis our findings are limited by design of the original studies. This analysis is subject to any biases of studies included in both the secretor status or genotype data, which is likely more relevant for secretor data where selection bias could be present. Careful selection of secretor studies likely limited selection bias among included references. Second, only 38 countries were included in the analysis due to availability of data for both secretor and GII.4 prevalence. Additionally, about half of the observations came from the Western Pacific region. Having more data from different regions of the globe would make this

analysis more representative. Third, country level secretor prevalence was estimated from limited data and does not account for any within-country heterogeneity. Included studies on secretor prevalence may not be representative of the entire country, especially for countries with little data. Finally, this analysis does not consider spatial effects. Human and viral diversity do not conform to national boundaries, but for practicality, nations were chosen as the spatial unit. Taken together, these limitations likely would bias the relationship between secretor status and GII.4 to the null.

Strengths

There are three key strengths for our analysis. First, the study makes use of a whole body of literature and includes all available published genotype data from countries with secretor data published at the time of the review. This dataset spans a long period of GII.4 dominance since its global emergence during the mid-1990s (61). Second, although only 38 countries were included, the countries are geographically diverse, with every continent represented. Third, multivariable methods allowed us to control for study design and demographic factors that may confound the relationship between secretor and GII.4 prevalence including HDI, age category, study setting, sampling of sporadic cases versus outbreaks, and time period. For example, higher income countries likely have better infrastructure for outbreak surveillance, and certain genotypes may be more common in outbreaks than sporadic cases. Additionally, GII.4 infections are likely more prevalent in inpatient studies, and secretor prevalence is likely higher among those infected with GII.4.

Relationship to Literature

Several studies have identified individual level risk associated with secretor status and resistance to some noroviruses, but none have explored the relationship between population-level secretor prevalence and the genotype within a population. Resistance to GII.4 among nonsecretors is well established, and secretor prevalence clearly varies globally. Our findings show that this relationship also operates at the population level and suggest that human diversity drives genotype prevalences. Similar population-level relationships have been observed in several other infectious diseases including cholera and malaria for which there is an individual-level relationship between host genetics and susceptibility. Individuals with blood group O are at increased risk of cholera compared to A and B blood groups. Cholera-endemic regions are associated with low blood group O prevalence in cholera-endemic regions due to interaction between population genetics and cholera (62). Similarly, sickle cell heterozygosity confers resistance to malaria and higher prevalence of sickle cell trait has been observed in areas of higher malaria prevalence (63).

Implications

The specific implications of these findings for viral evolution remain to be elucidated. New genotypes may be more likely to emerge in areas of lower secretor prevalence where there are more non-GII.4 viruses and perhaps more opportunity for recombination. Conversely, GII.4 evolution could be more driven in settings of higher secretor prevalence where population immunity develops as a result of high GII.4 exposure and may exert selective pressure on GII.4 viruses.

The interaction between secretor prevalence and GII.4 prevalence may have implications for vaccine interventions. GII.4 vaccines may lower the prevalence of GII.4 in populations with higher secretor prevalence. Populations with lower secretor prevalence may benefit less from GII.4 vaccines, suggesting than different vaccine formulations would be appropriate for different populations. Perhaps vaccine introduction could also increase selective pressure on GII.4 noroviruses, subsequently causing more rapid evolution.

Prior observational evidence suggests that higher nonsecretor prevalence coincides with more diverse norovirus genotype distributions, but further research is needed to quantify the effects that secretor prevalence may have on viral diversity (38). An outstanding question is whether higher secretor prevalence populations have higher incidence of norovirus gastroenteritis. Among higher nonsecretor prevalence populations, GII.4 prevalence may be lower due to population resistance, but the incidence of non-GII.4 genotypes may be the same as populations with more secretors even though their prevalence is relatively higher. Future studies should examine whether population-level resistance predicts norovirus genotype-specific incidence.

References

1. Adler JL, Zickl R. Winter vomiting disease. *J Infect Dis* 1969;119(6):668-73.
2. Bull RA, White PA. Mechanisms of GII.4 norovirus evolution. *Trends Microbiol* 2011;19(5):233-40.
3. GBD Compare. 2016. (https://vizhub.healthdata.org/gbd-compare/). (Accessed).

4. Ahmed SM, Hall AJ, Robinson AE, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect Dis* 2014;14(8):725-30.

5. Lopman B, Atmar R, Baric R, et al. Global Burden of Norovirus and Prospects for Vaccine Development. 2015.

6. Hall AJ, Lopman BA, Payne DC, et al. Norovirus disease in the United States. *Emerg Infect Dis* 2013;19(8):1198-205.

7. Donaldson EF, Lindesmith LC, Lobue AD, et al. Viral shape-shifting: norovirus evasion of the human immune system. *Nat Rev Microbiol* 2010;8(3):231-41.

8. Teunis PF, Moe CL, Liu P, et al. Norwalk virus: how infectious is it? *J Med Virol* 2008;80(8):1468-76.

9. Wilhelm B, Waddell L, Greig J, et al. A scoping review of the evidence for public health risks of three emerging potentially zoonotic viruses: hepatitis E virus, norovirus, and rotavirus. *Prev Vet Med* 2015;119(1-2):61-79.

10. Charles KJ, Shore J, Sellwood J, et al. Assessment of the stability of human viruses and coliphage in groundwater by PCR and infectivity methods. *J Appl Microbiol* 2009;106(6):1827-37.

11. Glass RI, Noel J, Ando T, et al. The epidemiology of enteric caliciviruses from humans: a reassessment using new diagnostics. *J Infect Dis* 2000;181 Suppl 2:S254-61.

12. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. *N Engl J Med* 2009;361(18):1776-85.

13. Ahmed SM, Lopman BA, Levy K. A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS One* 2013;8(10):e75922.

14. Bok K, Green KY. Norovirus gastroenteritis in immunocompromised patients. *N Engl J Med* 2012;367(22):2126-32.

15. Devasia T, Lopman B, Leon J, et al. Association of host, agent and environment characteristics and the duration of incubation and symptomatic periods of norovirus gastroenteritis. *Epidemiol Infect* 2015;143(11):2308-14.

16. Lee RM, Lessler J, Lee RA, et al. Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis* 2013;13:446.

17. Cortes-Penfield NW, Ramani S, Estes MK, et al. Prospects and Challenges in the Development of a Norovirus Vaccine. *Clin Ther* 2017;39(8):1537-49.

18. Leroux-Roels G, Cramer JP, Mendelman PM, et al. Safety and Immunogenicity of Different Formulations of Norovirus Vaccine Candidate in Healthy Adults: A Randomized, Controlled, Double-Blind Clinical Trial. *J Infect Dis* 2018;217(4):597-607.

19. Vinje J. Advances in laboratory methods for detection and typing of norovirus. *J Clin Microbiol* 2015;53(2):373-81.

20. Xi JN, Graham DY, Wang KN, et al. Norwalk virus genome cloning and characterization. *Science* 1990;250(4987):1580-3.

21. Kroneman A, Vega E, Vennema H, et al. Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol* 2013;158(10):2059-68.

22. Hoa Tran TN, Trainor E, Nakagomi T, et al. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global

distribution of genogroups, genotypes and GII.4 variants. *J Clin Virol* 2013;56(3):185-93.

23. van Beek J, de Graaf M, Al-Hello H, et al. Molecular surveillance of norovirus, 2005-16: an epidemiological analysis of data collected from the NoroNet network. *Lancet Infect Dis* 2018.

24. Zheng DP, Ando T, Fankhauser RL, et al. Norovirus classification and proposed strain nomenclature. *Virology* 2006;346(2):312-23.

25. Chan MC, Lee N, Hung TN, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat Commun* 2015;6:10061.

26. Hoa-Tran TN, Nakagomi O, Dao AT, et al. Molecular epidemiology of noroviruses detected in Vietnamese children with acute gastroenteritis from 2012 to 2015. *J Med Microbiol* 2017;66(1):34-45.

27. Cannon JL, Barclay L, Collins NR, et al. Genetic and Epidemiologic Trends of Norovirus Outbreaks in the United States from 2013 to 2016 Demonstrated Emergence of Novel GII.4 Recombinant Viruses. *J Clin Microbiol* 2017;55(7):2208-21.

28. Franck KT, Fonager J, Ersboll AK, et al. Norovirus epidemiology in community and health care settings and association with patient age, Denmark. *Emerg Infect Dis* 2014;20(7):1123-31.

29. Desai R, Hembree CD, Handel A, et al. Severe outcomes are associated with genogroup 2 genotype 4 norovirus outbreaks: a systematic literature review. *Clin Infect Dis* 2012;55(2):189-93.

30. Lopman B, Zambon M, Brown DW. The evolution of norovirus, the "gastric flu". *PLoS Med* 2008;5(2):e42.

31. Lindesmith LC, Donaldson EF, Lobue AD, et al. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med* 2008;5(2):e31.

32. Jung S, Hwang BM, Jung H, et al. Emergence of Norovirus GII.17-associated Outbreak and Sporadic Cases in Korea from 2014 to 2015. *Osong Public Health Res Perspect* 2017;8(1):86-90.

33. Fu J, Ai J, Jin M, et al. Emergence of a new GII.17 norovirus variant in patients with acute gastroenteritis in Jiangsu, China, September 2014 to March 2015. *Euro Surveill* 2015;20(24).

34. Zhang P, Chen L, Fu Y, et al. Clinical and molecular analyses of norovirus-associated sporadic acute gastroenteritis: the emergence of GII.17 over GII.4, Huzhou, China, 2015. *BMC Infect Dis* 2016;16(1):717.

35. Marionneau S, Ruvoen N, Le Moullac-Vaidye B, et al. Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. *Gastroenterology* 2002;122(7):1967-77.

36. Tan M, Jiang X. Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle. *Trends Microbiol* 2005;13(6):285-93.

37. Marionneau S, Cailleau-Thomas A, Rocher J, et al. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* 2001;83(7):565-73.

38. Nordgren J, Sharma S, Kambhampati A, et al. Innate Resistance and Susceptibility to Norovirus Infection. *PLoS Pathog* 2016;12(4):e1005385.

39. Lopman BA, Trivedi T, Vicuna Y, et al. Norovirus Infection and Disease in an Ecuadorian Birth Cohort: Association of Certain Norovirus Genotypes With Host FUT2 Secretor Status. *J Infect Dis* 2015;211(11):1813-21.

40. Rydell GE, Kindberg E, Larson G, et al. Susceptibility to winter vomiting disease: a sweet matter. *Rev Med Virol* 2011;21(6):370-82.

41. Kambhampati A, Payne DC, Costantini V, et al. Host Genetic Susceptibility to Enteric Viruses: A Systematic Review and Metaanalysis. *Clin Infect Dis* 2016;62(1):11-8.

42. Frenck R, Bernstein DI, Xia M, et al. Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. *J Infect Dis* 2012;206(9):1386-93.

43. Carlsson B, Kindberg E, Buesa J, et al. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS One* 2009;4(5):e5593.

44. Nordgren J, Nitiema LW, Ouermi D, et al. Host genetic factors affect susceptibility to norovirus infections in Burkina Faso. *PLoS One* 2013;8(7):e69557.

45. Liu P, Wang X, Lee JC, et al. Genetic susceptibility to norovirus GII.3 and GII.4 infections in Chinese pediatric diarrheal disease. *Pediatr Infect Dis J* 2014;33(11):e305-9.

46. Thorven M, Grahn A, Hedlund KO, et al. A homozygous nonsense mutation (428G-->A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *J Virol* 2005;79(24):15351-5.

47. Lindesmith LC, Beltramello M, Donaldson EF, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog* 2012;8(5):e1002705.

48. Dickey W, Wylie JD, Collins JS, et al. Lewis phenotype, secretor status, and coeliac disease. *Gut* 1994;35(6):769-70.

49. Heneghan MA, Kearns M, Goulding J, et al. Secretor status and human leucocyte antigens in coeliac disease. *Scand J Gastroenterol* 1996;31(10):973-6.

50. Mitchell RJ, Sunderland E. ABH secretion in Northumberland, England. *Human heredity* 1982;32(3):181-4.

51. Procter J, Crawford J, Bunce M, et al. A rapid molecular method (polymerase chain reaction with sequence-specific primers) to genotype for ABO blood group and secretor status and its potential for organ transplants. *Tissue antigens* 1997;50(5):475-83.

52. Toft AD, Blackwell CC, Saadi AT, et al. Secretor status and infection in patients with Graves' disease. *Autoimmunity* 1990;7(4):279-89.

53. Naidu JM, Veerraju P, Mukherjee DP. ABO blood groups and ABH saliva secretion in Reddis and Kammas of Southern Andhra Pradesh. *American journal of physical anthropology* 1978;48(4):533-4.

54. Saboor M, Ullah A, Qamar K, et al. Frequency of ABH secretors and non secretors: A cross sectional study in Karachi. *Pakistan journal of medical sciences* 2014;30(1):189-93.

55. Sun XM, Yan XF, He ZG, et al. Norovirus Infection and Histo-blood Group Antigens in Children Hospitalized with Diarrhea in Lulong and Chenzhou in China. *Biomed Environ Sci* 2016;29(4):286-9.

56. Bernardo CR, Camargo AVS, Ronchi LS, et al. ABO, Secretor and Lewis histo-blood group systems influence the digestive form of Chagas disease. *Infect Genet Evol* 2016;45:170-5.

57. Bucardo F, Kindberg E, Paniagua M, et al. Genetic susceptibility to symptomatic norovirus infection in Nicaragua. *J Med Virol* 2009;81(4):728-35.

58. O'Ryan ML, Rabello M, Cortes H, et al. Dynamics of Helicobacter pylori detection in stools during the first 5 years of life in Chile, a rapidly developing country. *Pediatr Infect Dis J* 2013;32(2):99-103.

59. UNDP. Human Development Data (1990-2015). United Nations Development Programme; 2016. (http://hdr.undp.org/en/data). (Accessed April 1, 2018).

60. WHO. WHO Regional Offices. World Health Organization; 2018. (http://www.who.int/about/regions/en/). (Accessed April 1, 2018).

61. Noel JS, Fankhauser RL, Ando T, et al. Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *J Infect Dis* 1999;179(6):1334-44.

62. Glass RI, Holmgren J, Haley CE, et al. Predisposition for cholera of individuals with O blood group. Possible evolutionary significance. *Am J Epidemiol* 1985;121(6):791-6.

63. Elguero E, Delicat-Loembet LM, Rougeron V, et al. Malaria continues to select for sickle cell trait in Central Africa. *Proc Natl Acad Sci U S A* 2015;112(22):7051-4.

Tables

Table 1. Characteristics of studies included in final review

Characteristics	N	Median GII.4 Prevalence (IQR)	Median GII.non-4 Prevalence (IQR)[a]	Median GI Prevalence (IQR)[a]
Total	411	0.64 (0.38-0.80)	0.25 (0.11-0.48)	0.03 (0.00-0.11)
Secretor Prevalence Quartiles[b]				
Quartile 1	119	0.58 (0.36-0.79)	0.30 (0.14-0.51)	0.01 (0.00-0.11)
Quartile 2	96	0.67 (0.48-0.78)	0.23 (0.11-0.35)	0.04 (0.02-0.15)
Quartile 3	105	0.67 (0.40-0.81)	0.26 (0.12-0.48)	0.02 (0.00-0.10)
Quartile 4	91	0.67 (0.31-0.88)	0.25 (0.10-0.56)	0.02 (0.00-0.12)
Study Type				
Outbreaks	118	0.66 (0.33-0.84)	0.21 (0.08-0.44)	0.06 (0.00-0.16)
Inpatient Cases	76	0.66 (0.45-0.81)	0.26 (0.14-0.54)	0.00 (0.00-0.05)
Outpatient Cases	48	0.70 (0.51-0.85)	0.27 (0.12-0.47)	0.00 (0.00-0.09)
Inpatient & Outpatient Cases	80	0.64 (0.43-0.75)	0.27 (0.14-0.46)	0.04 (0.00-0.11)
Community Cases	39	0.57 (0.12-0.80)	0.32 (0.15-0.51)	0.01 (0.00-0.13)
Mixed Setting	39	0.56 (0.33-0.73)	0.28 (0.14-0.57)	0.06 (0.00-0.15)
Not Specified	11	0.77 (0.55-0.85)	0.21 (0.19-0.32)	0.03 (0.00-0.07)
Age Category				
≤5 Years	105	0.57 (0.37-0.75)	0.28 (0.18-0.54)	0.00 (0.00-0.11)
>5 Years	15	0.79 (0.50-0.87)	0.10 (0.05-0.26)	0.08 (0.03-0.19)
Mixed	291	0.67 (0.38-0.83)	0.25 (0.10-0.48)	0.03 (0.00-0.11)
Pandemic Variant Period				
1994-2001	52	0.49 (0.14, 0.67)	0.35 (0.20, 0.69)	0.08 (0.00, 0.25)
2002-2003	20	0.85 (0.67, 0.90)	0.10 (0.00, 0.19)	0.03 (0.02, 0.25)
2004-2005	28	0.72 (0.52, 0.88)	0.16 (0.06, 0.26)	0.01 (0.00, 0.13)
2006-2008	69	0.78 (0.59, 0.89)	0.14 (0.05, 0.28)	0.02 (0.00, 0.06)
2009-2011	56	0.67 (0.49, 0.82)	0.29 (0.13, 0.44)	0.01 (0.00, 0.06)
2012-2017	103	0.59 (0.25, 0.73)	0.32 (0.18, 0.58)	0.03 (0.00, 0.11)
Multiple Periods[c]	83	0.59 (0.35-0.73)	0.30 (0.18-0.50)	0.04 (0.00-0.13)
WHO Region				
Africa	20	0.53 (0.46-0.65)	0.28 (0.00-0.45)	0.12 (0.00-0.33)
Americas	79	0.67 (0.31-0.84)	0.24 (0.08-0.50)	0.02 (0.00-0.12)
South-East Asia	19	0.44 (0.25-0.75)	0.43 (0.25-0.62)	0.00 (0.00-0.13)
Europe	83	0.72 (0.47-0.85)	0.16 (0.10-0.33)	0.04 (0.00-0.11)
Eastern Mediterranean	14	0.33 (0.15-0.70)	0.53 (0.17-0.72)	0.11 (0.00-0.17)
Western Pacific	196	0.64 (0.44-0.78)	0.28 (0.17-0.48)	0.02 (0.00-0.10)
UNDP Human Development Index Category				
Very High Development	292	0.68 (0.48-0.82)	0.23 (0.11-0.40)	0.03 (0.00-0.11)
High Development	73	0.42 (0.12-0.78)	0.41 (0.11-0.87)	0.00 (0.00-0.11)
Medium Development	32	0.49 (0.24-0.67)	0.43 (0.21-0.60)	0.07 (0.00-0.15)
Low Development	14	0.50 (0.29-0.67)	0.25 (0.00-0.34)	0.19 (0.00-0.38)
Typing Method				
RT-PCR	384	0.64 (0.38-0.80)	0.26 (0.11-0.48)	0.03 (0.00-0.11)
Other	25	0.67 (0.36-0.80)	0.21 (0.06-0.49)	0.00 (0.00-0.12)
Unknown	2	0.76 (0.72-0.80)	0.13 (0.10-0.11)	0.10 (0.10-0.11)
Typing Region				
ORF 1	90	0.70 (0.38-0.87)	0.19 (0.08-0.38)	0.03 (0.00-0.13)
ORF 2	203	0.63 (0.41-0.79)	0.28 (0.13-0.49)	0.03 (0.00-0.11)
ORF 1 and ORF 2	89	0.57 (0.33-0.75)	0.33 (0.16-0.56)	0.02 (0.00-0.12)
ORF 1/ORF 2 Overlap Only	16	0.68 (0.49-0.83)	0.23 (0.12-0.48)	0.00 (0.00-0.12)
Unknown	13	0.60 (0.51-0.89)	0.15 (0.07-0.28)	0.04 (0.00-0.11)

Abbreviations: IQR, inter-quartile range; WHO, World Health Organization; UNDP, United Nations Development Programme; RT-PCR, reverse-transcription polymerase chain reaction; ORF, open reading frame.

[a]Prevalence of GI and GII.non-4 were only calculated for 349 observations

[b]Secretor prevalence quartiles were determined from country-level secretor prevalence.

[c]Observations that span multiple pandemic variant periods.

Table 2. Weighted univariate linear regressions of predictors on GII.4 prevalence

Predictor	R^2	β	95% CI
Secretor Prevalence	0.05	**1.17**[a]	**0.69, 1.64**
Study Type	0.02		
Intercept		**0.66**	**0.61, 0.70**
Outbreaks		0.00	Reference
Inpatient Cases		0.03	-0.06, 0.11
Outpatient Cases		0.00	-0.10, 0.11
Inpatient & Outpatient Cases		-0.03	-0.12, 0.06
Community Cases		-0.10	-0.23, 0.04
Mixed Setting		0.06	-0.03, 0.14
Not Specified		0.12	-0.10, 0.34
Age Category	0.02		
Intercept		**0.57**	**0.49, 0.64**
>5 Years		0.00	Reference
≤5 Years		0.05	-0.11, 0.20
Mixed		**0.11**	**0.03, 0.19**
Pandemic Variant Period	0.24		
Intercept		**0.55**	**0.51, 0.59**
2012-2017		0.00	Reference
1994-2001		**-0.12**	**-0.22, -0.02**
2002-2003		**0.15**	**0.03, 0.27**
2004-2005		**0.29**	**0.17, 0.42**
2006-2008		**0.32**	**0.26, 0.39**
2009-2011		**0.12**	**0.01, 0.23**
Multiple Periods		**0.13**	**0.06, 0.20**
WHO Region	0.05		
Intercept		**0.64**	**0.60, 0.67**
Western Pacific		0.00	Reference
Africa		-0.18	-0.46, 0.09
Americas		0.04	-0.03, 0.11
South-East Asia		-0.11	-0.36, 0.14
Europe		**0.11**	**0.04, 0.18**
Eastern Mediterranean		**-0.36**	**-0.61, -0.10**
HDI	0.10	**1.01**	**0.71, 1.31**
Detection Method	0.00		
Intercept		**0.66**	**0.63, 0.69**
RT-PCR		0.00	Reference
Other		0.00	-0.14, 0.15
Unknown		0.06	-0.05, 0.17
Typing Region	0.04		
Intercept		**0.72**	**0.66, 0.78**
ORF1		0.00	Reference
ORF2		**-0.11**	**-0.18, -0.04**
ORF1 and ORF2		-0.03	-0.13, 0.06
Overlap Only		-0.01	-0.12, 0.10
Unknown		0.03	-0.07, 0.14

Abbreviations: CI, confidence interval; WHO, World Health Organization; HDI, human development index; RT-PCR, reverse transcription polymerase chain reaction; ORF, open reading frame.

[a]Bold values indicate statistically significant parameter estimates (α=0.05)

Table 3. Weighted bivariate linear regressions of secretor status and each covariate on weighted GII.4 prevalence

Model Variables	Categories	R^2	β	95% CI
Secretor Prevalence		0.07	**1.23**[a]	**0.73, 1.73**
Study Type	Outbreaks		0.00	Reference
	Inpatient Cases		0.04	-0.05, 0.12
	Outpatient Cases		0.03	-0.07, 0.13
	Inpatient & Outpatient Cases		-0.03	-0.12, 0.06
	Community Cases		-0.12	-0.25, 0.01
	Mixed Setting		0.01	-0.07, 0.09
	Not Specified		0.13	-0.08, 0.35
Secretor Prevalence		0.08	**1.28**	**0.80, 1.75**
Age Category	≤5 Years		0.00	Reference
	>5 Years		0.09	-0.06, 0.24
	Mixed		**0.14**	**0.06, 0.22**
Secretor Prevalence		0.27	**0.93**	**0.50, 1.37**
Pandemic Variant Period	2012-2017		0.00	Reference
	1994-2001		**-0.13**	**-0.23, -0.03**
	2002-2003		**0.17**	**0.05, 0.29**
	2004-2005		**0.29**	**0.17, 0.42**
	2006-2008		**0.30**	**0.23, 0.36**
	2009-2011		**0.12**	**0.02, 0.23**
	Multiple Periods		**0.11**	**0.04, 0.18**
Secretor Prevalence		0.08	**0.87**	**0.33, 1.41**
WHO Region	Western Pacific		0.00	Reference
	Africa		-0.08	-0.36, 0.20
	Americas		0.01	-0.06, 0.08
	South-East Asia		-0.07	-0.31, 0.18
	Europe		**0.08**	**0.01, 0.15**
	Eastern Mediterranean		**-0.28**	**-0.53, -0.02**
Secretor Prevalence		0.11	**0.69**	**0.19, 1.18**
HDI			**0.85**	**0.52, 1.17**
Secretor Prevalence		0.06	**1.18**	**0.70, 1.66**
Detection Method	RT-PCR		0.00	Reference
	Other		0.05	-0.10, 0.19
	Unknown		0.06	-0.05 0.17
Secretor Prevalence		0.08	**1.08**	**0.58, 1.58**
Typing Region	ORF1		0.00	Reference
	ORF2		**-0.07**	**-0.15, -0.00**
	ORF1 and ORF2		0.02	-0.07, 0.12
	Overlap Only		0.02	-0.09, 0.13
	Unknown		0.04	-0.07, 0.14

Abbreviations: CI, confidence interval; WHO, World Health Organization; HDI, human development index; RT-PCR, reverse transcription polymerase chain reaction; ORF, open reading frame.

[a]Bold values indicate statistically significant regression coefficients (α=0.05)

Table 4. Weighted multiple linear regression model estimating the association between secretor and GII.4 prevalence, controlling for study type, pandemic variant period, age category, and Human Development Index. $R^2 = 0.34$

Variable	Category	β	95% CI
Secretor Prevalence		0.60	0.11, 1.08
Study Type	Outbreaks	0.00	Reference
	Inpatient Cases	0.10	0.02, 0.17
	Outpatient Cases	0.14	0.04, 0.23
	Inpatient & Outpatient Cases	0.06	-0.02, 0.14
	Community Cases	-0.04	-0.15, 0.08
	Mixed Setting	0.02	-0.06, 0.09
	Not Specified	0.08	-0.10, 0.27
Pandemic Variant Period	2012-2017	0.00	Reference
	1994-2001	-0.09	-0.20, 0.01
	2002-2003	0.15	0.04, 0.27
	2004-2005	0.31	0.19, 0.43
	2006-2008	0.29	0.23, 0.36
	2009-2011	0.14	0.04, 0.25
	Multiple Periods	0.11	0.05, 0.18
Age Category	≤5	0.00	Reference
	>5	0.06	-0.09, 0.20
	Mixed	0.02	-0.07, 0.10
HDI		0.88	0.54, 1.22

Abbreviations: CI, confidence interval; HDI, human development index.

Figures

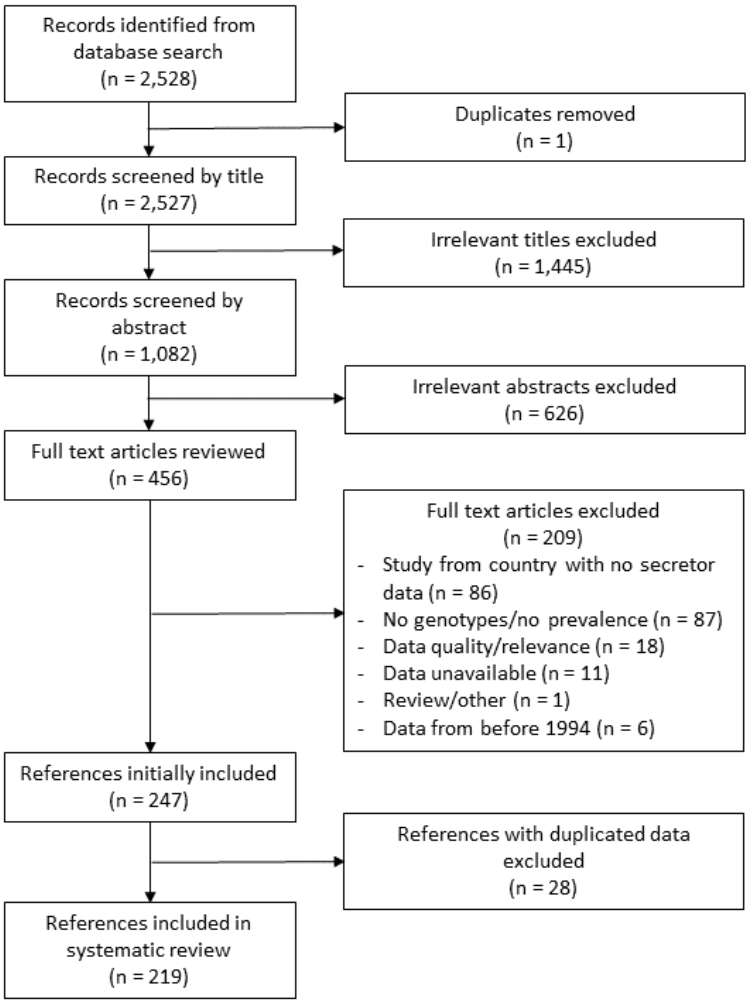

Figure 1. Flow diagram of references included in systematic literature review

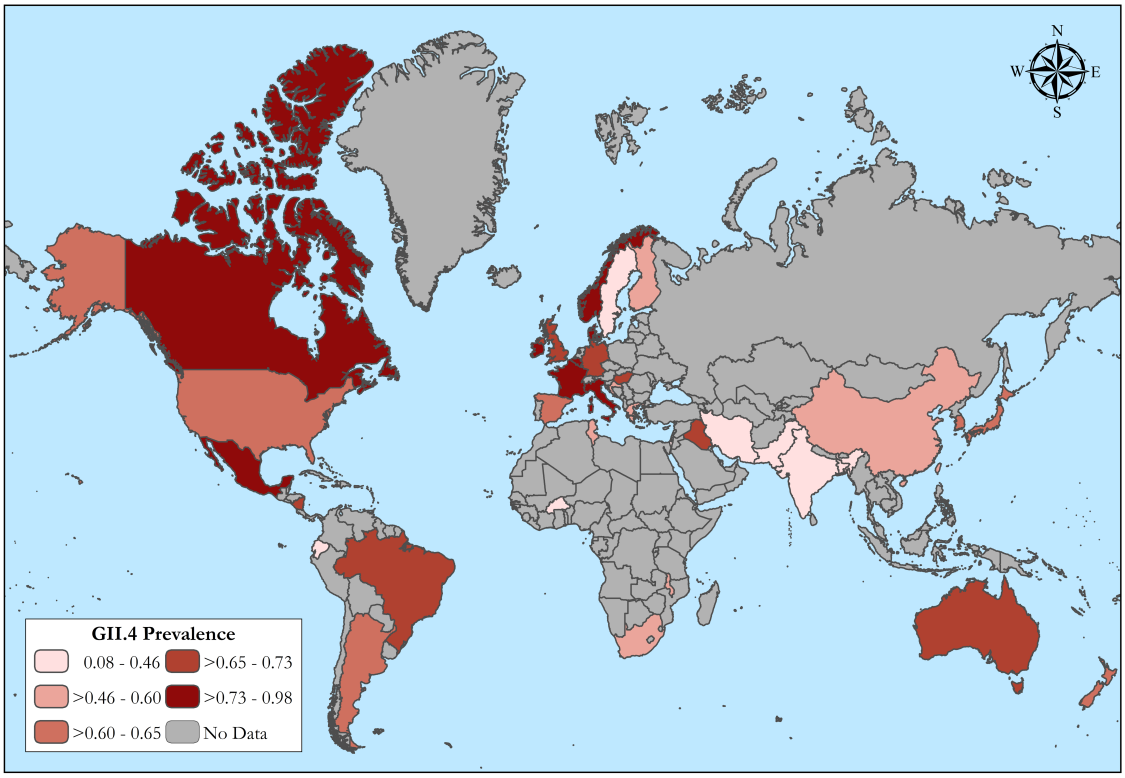

Figure 2. Distribution of aggregated GII.4 norovirus prevalence for all countries included in the analysis

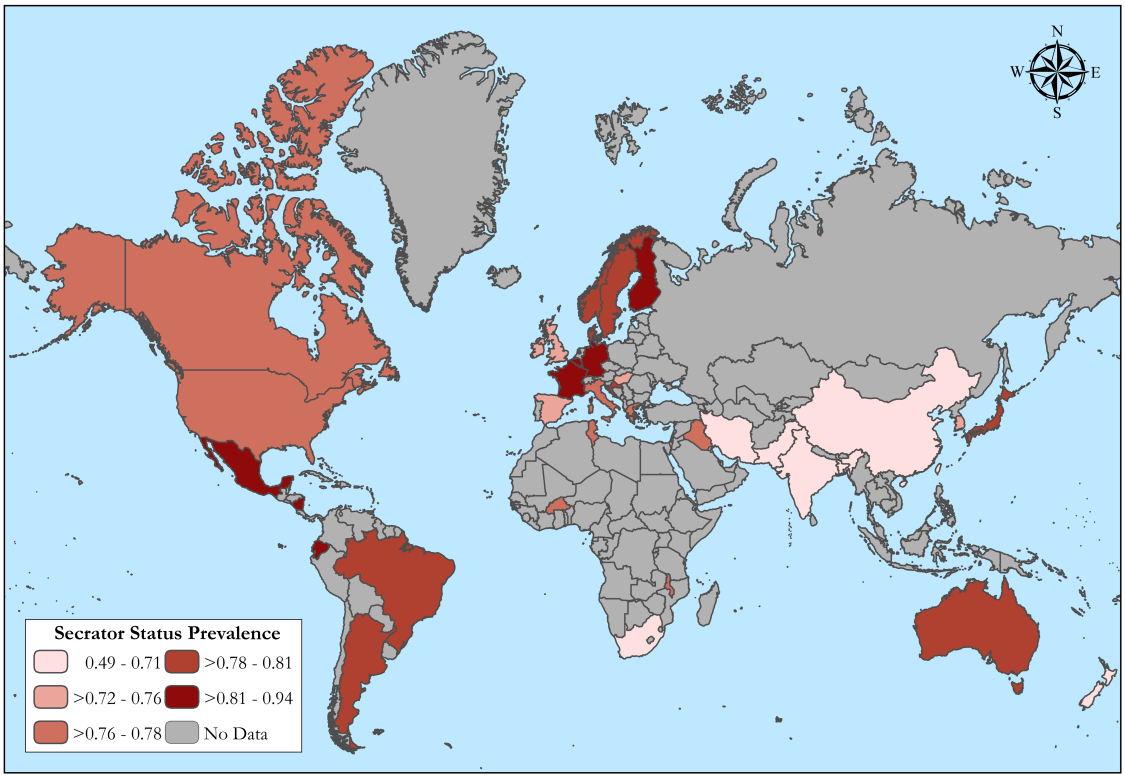

Figure 3. Distribution of aggregated secretor status prevalence for all countries included in the analysis.

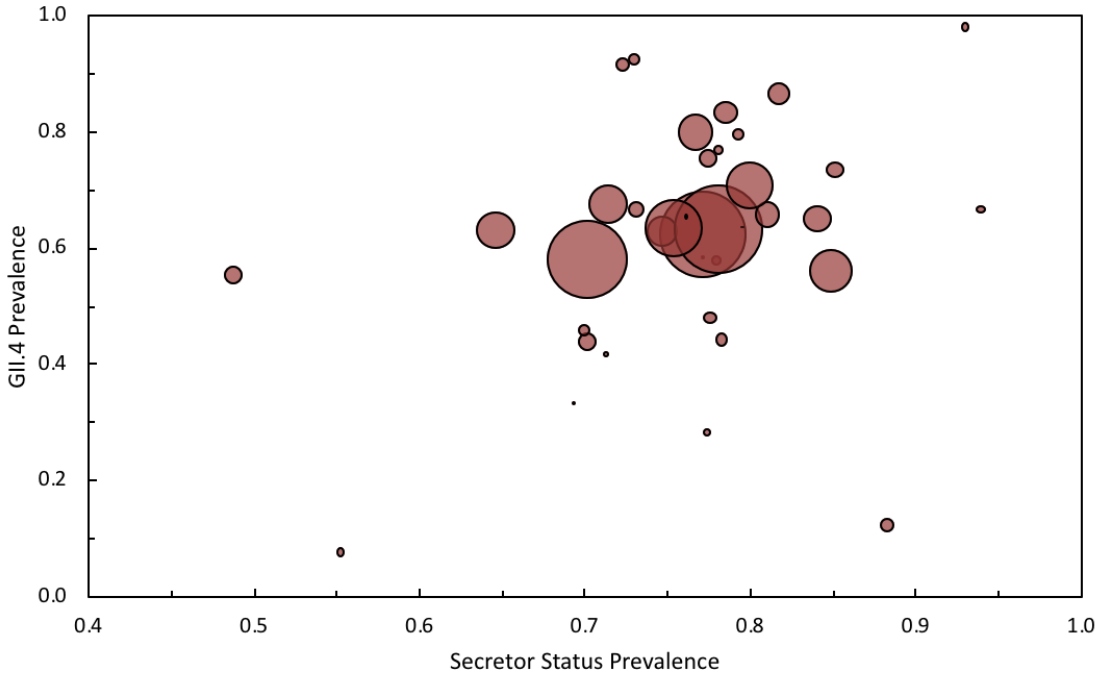

Figure 4. Scatterplot of aggregated country level GII.4 prevalence and secretor prevalence. Marker size is proportional to total country sample size.

Chapter III: Public Health Implications and Future Directions

Several studies have observed individual level associations between secretor prevalence and GII.4 norovirus susceptibility, yet none have systematically assessed the effects of population level secretor prevalence on a country's norovirus genotype diversity. In this study, we observed a significant positive association between population level host genetic susceptibility and the prevalence of the most common norovirus genotype, GII.4. This finding has important public health implications and suggests further research on the effects of population level resistance on the prevalence and incidence of norovirus genotypes.

Population-level resistance to GII.4 may decrease the prevalence of GII.4 in a within a given population. This notion could be important for vaccine interventions especially among high risk populations including patients in healthcare settings and nursing homes. Since GII.4 noroviruses are more prevalent in nursing homes and healthcare settings and are also associated with higher mortality and severe disease rates than other genotypes, they present a serious public health threat in these settings. GII.4 noroviruses are also more likely to cause outbreaks in these settings. Vaccination against GII.4 in nursing homes and healthcare settings among both patients and staff may reduce the prevalence of GII.4 noroviruses in these settings and thus also reduce the norovirus-specific mortality and severe disease rates, as well as the incidence of norovirus outbreaks. Consideration of how the genotype distribution and genotype evolution may change in the face of decreasing susceptible populations is crucial for planning vaccine and antiviral intervention programs.

The most important next step in this research is to examine this association with data that is more representative of other parts of the globe. Due to lack of either secretor data or genotype data, few studies were included in this dataset from Africa and, with the exception of Brazil, South America, relative to Eastern Asia, Australia, and the United States. Critical to further studying this relationship between secretor and GII.4 prevalence are representative, country-level secretor status prevalence estimates and additional molecular surveillance studies.

Another next step in this research would be to explore how the prevalence of non-GII.4 genotypes differ with secretor prevalence. Since GII.4 is the most common genotype, all non-GII.4 viruses would be expected to have higher prevalences with lower GII.4 prevalence. However, it is possible that specific genotypes could proliferate in the absence of GII.4 dominance. Understanding whether specific genotypes may proliferate with a reduction in the GII.4 susceptible population is important for understanding the implications of vaccination against GII.4 noroviruses.

This study did not address the relationship between secretor prevalence and norovirus incidence. It is possible that high secretor prevalence populations have higher incidence of norovirus overall due to a larger susceptible population. Therefore, among higher nonsecretor prevalence populations, GII.4 prevalence may be lower due to population resistance, but the incidence of non-GII.4 genotypes may be the same as populations with more secretors even though their prevalence is relatively higher. A further step in this research would be to study the relationship between secretor prevalence and incidence of norovirus genotypes.