

https://dspace.org/

# Is there an association between genetic polymorphisms and immune responses to rotavirus vaccination in South African infants?

**Miya, Thabiso**

**2021**
https://hdl.handle.net/10539/33177

Downloaded from a DSpace repository. https://dspace.org/




# Is there an association between genetic polymorphisms and immune responses to rotavirus vaccination in South African infants?

by

**Thabiso Miya**

**(488744)**

**Thesis**

Submitted in fulfilment of the requirements for the degree

**Philosophiae Doctor**

in

**Molecular and Cell Biology**

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr Debbie de Assis Rosa

Co–supervisor: Dr Michelle Groome

November 2021

**DECLARATION**

I, Thabiso Victor Miya (488744), am a student registered for the degree of Doctor of Philosophy of Science in Genetics (Thesis) in the academic year 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis **HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

NRF declaration

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.



_____ _____

(Signature of candidate)

____02____day of ___Novmber___2021___at__Wits___

## DEDICATION

This thesis is dedicated to the memory of my grandmother Mamotshoane Miya and my eldest sister Jabulile Ngwenya. Thank you for your unconditional love and wonderful memories.

**ABSTRACT**

Rotavirus (RV) vaccines have demonstrated low efficacy in sub–Saharan Africa. Recent studies have reported that RVs bind to histoblood group antigens (HBGAs), and thus they may act as receptors for this virus. Variation in HBGA phenotypes have been suggested as the cause of RV vaccines efficacy reduction. Thus, the aim of this study was to investigate the associations between polymorphisms in HBGA genes (*ABO*/*FUT2/FUT3)* and selected innate immune response genes, and ABO/Secretor (Se)/Lewis (Le) phenotypes with immune response (seroconversion) to RV vaccine called Rotarix (RV1) in Black South African infants.

A total of 215 saliva samples collected before and after exposure to RV vaccination in infants were available from a previous study. Previously published RV–specific IgA titres pre– and post–vaccination were used to classify samples as seroconverters or non–seroconverters. Pre–vaccination IgA titre in babies was significantly associated with seroconversion and was used as a covariate in multivariable analysis of factors influencing baby immune response to vaccination.

We reported minor allele frequencies (MAF) of the 50 SNPs. A haplotype in *FUT2* linked to the se428 mutation (rs601338) was the most common haplotype in the cohort (42.8%); overall the frequency of non–Secretor status predicted from genetic data was 16%. The most common *FUT3* haplotype causing predicted Le absence was linked to the le508 (rs3745635) mutation, found in 29% of haplotypes. Overall, the frequency of Le negative phenotype predicted from genetic data was 20%. ABO genotypic data were used to predict ABO blood type frequencies in the cohort, that were similar to previous frequencies reported in the SA Black population.

Se and Le phenotypes, *FUT2* genotypes and ABO genotypes were not significantly associated with seroconversion in the current study. Only one *FUT3* SNP rs761121538 (le862) was associated significantly with seroconversion in an allelic model (P= 0.002, OR= 0.070; P= 0.042 after Bonferroni correction, P= 0.051 in multivariable analysis). Other missense mutations in *FUT3* were not significantly associated with seroconversion in the current study. However, when *FUT2* and *FUT3* genetic data were combined and used to predict Le antigen phenotypes, there was an association between predicted Le phenotype and seroconversion (P= 0.026). Epistasis analysis also suggested that the interaction between some SNPs in the *FUT2* se428 haplotype and the *FUT3* le667 SNP (creating the Le (a–b–) phenotype) was significantly associated with seroconversion. These findings support the hypothesis that combinations of *FUT2–FUT3* variants may contribute to seroconversion after exposure to Rotarix vaccine.

With regards to our study of genetic variation in genes influencing innate immunity, the *TLR8* SNP rs2129377 was significantly associated with seroconversion (P= 0.0146, OR= 5.351; P > 0.05 after Bonferroni correction) and one six SNP haplotype found on the X chromosome in the *TLR7–TLR8* region was significantly associated with seroconversion (P= 0.042, P > 0.05 after Bonferroni correction).

Overall, it can be concluded that interaction between SNPs in *FUT2* and *FUT3* genes, and variation in innate sensors of viral RNA, might play a significant role towards variation in immune response to Rotarix vaccine in this South African cohort.

**ACKNOWLEDGEMENTS**

I would like to express my heartfelt gratitude and appreciation to the following people and institutions:

- To my supervisor Dr Debbie de Assis Rosa. Thank you for your patience, inspiration, guidance, and help. I will forever be grateful for you.
- To my co–supervisor Dr Michelle Groome. Honestly, none of this would have been possible without you. Thank you for pioneering this project and also providing the study sample.
- To my advisor Dr Nikki Gentle. Thank you for your assistance and guidance in this project.
- To Prof. Shabir Madhi. Thank you for providing me with a partial bursary for year of my studies.
- To Prof. Jeffrey Dorfman. Thank you for helping out with ELISA optimization and overall guidance on this project.
- To the study participants and their parents/guardians. This project would have been impossible without you. Thank you.
- To RMPRU and all the staff members. Thank you for your warm welcome and for allowing me to use your facilities and equipment.
- I acknowledge the National Research Fund (NRF) for financial support. I acknowledge the Fogarty International Center of the National Institutes of Health (NIH) for awarding Dr Michelle Groome with Emerging Global Leader Award which was used to fund this project.
- I acknowledge the University of the Witwatersrand. Thank you for affording me the opportunity to carry out this project.
- To the staff members and postgraduate students at the School of Molecular and Cell Biology (MCB). Thank you for all the assistance and encouragement.
- To my incredible parents Mr. Mmatluoa Paulus Miya and Mrs. Emma Paulina Miya. Thank you for your incredible love, support, and encouragement. I will forever be grateful for you both.
- To my wonderful brothers Molemo Elvis Miya and Molefi Lawrence Miya. Thank you for endless motivation and support.
- To my friends and extended family. Thank you for your love and encouragement.

Above all I would like to thank The Almighty God for his wonderful grace and wisdom to successfully complete this project.

## CONFERENCE PRESENTATIONS AND PUBLICATIONS ASSOCIATED WITH THIS STUDY

Miya, T.V., Groome, M., Madhi, S.A., and de Assis Rosa, D (2018). “Is there an association between Lewis and Secretor phenotypes (*FUT2* and *FUT3* genotypes) and immune response one month after administration of two doses of the Rotarix vaccine in South African infants?” Poster presentation at the Molecular Biosciences Research Thrust (MBRT) research day at the University of the Witwatersrand, Johannesburg [29 November 2017].

Miya, T.V., Groome, M., Madhi, S.A., and de Assis Rosa, D (2019). “No association between Secretor phenotypes (*FUT2* genotypes) and immune response to Rotarix vaccination in South African infants”. Poster presentation at the 12$^{th}$ African Rotavirus Symposium at the Emperors Palace, Johannesburg [30 July – 01 August 2019].

Miya, T.V., Groome, M., Madhi, S.A., and de Assis Rosa, D. (2019). “No association between Secretor phenotypes (*FUT2* genotypes) and immune response to Rotarix vaccination in South African infants”. Poster presentation at the 10$^{th}$ Cross Faculty Postgraduate Symposium at the University of the Witwatersrand, Johannesburg [03 September 2019].

Miya, T.V., Groome, M., and de Assis Rosa, D (2021). “*TLR* genetic variation is associated with Rotavirus–specific IgA seroconversion in South African Black infants after two doses of Rotarix vaccine”. *Vaccine* (accepted October 2021).

## TABLE OF CONTENTS

Is there an association between genetic polymorphisms and immune responses to rotavirus vaccination in South African infants? ..... i

**DECLARATION** ..... i

**DEDICATION** ..... ii

**ABSTRACT** ..... iii

**ACKNOWLEDGEMENTS** ..... v

**CONFERENCE PRESENTATIONS AND PUBLICATIONS ASSOCIATED WITH THIS STUDY** ..... vii

**TABLE OF CONTENTS** ..... viii

**LIST OF FIGURES** ..... xi

**LIST OF TABLES** ..... xiii

**LIST OF ABBREVIATIONS** ..... xvii

**1. INTRODUCTION** ..... 21

**1.1 Rotavirus: the main cause of severe acute gastroenteritis** ..... 1

**1.2 Rotaviruses structure, classification, life cycle, and pathogenesis** ..... 3

**1.2.1 Rotavirus structure** ..... 3

**1.2.2 Rotavirus classification** ..... 4

**1.2.3 Rotavirus life cycle and pathogenesis** ..... 5

**1.4 HBGA phenotypes** ..... 7

**1.4.1 Formation of the ABO blood types** ..... 9

**1.4.2 Variations in the *ABO* gene** ..... 9

**1.4.3 ABO alleles common in South Africa** ..... 11

**1.4.4 Formation of the Secretor (Se) phenotype** ..... 13

**1.4.5 Formation of the Lewis (Le) phenotype** ..... 13

**1.4.6 Changes in HBGA expression with age** ..... 15

**1.4.7 Using haplotypes to define HBGA variation** ..... 16

**1.5 HBGAs and susceptibility/ resistance to rotavirus infection** ..... 16

**1.6 Host immune response to rotavirus infection** ..... 20

**1.7 Vaccines against rotavirus infection** ..... 25

**1.8 Correlates of protection against rotavirus infection/ correlate of vaccine efficacy** ..... 27

**1. 9 Aims and Objectives** ..... 31

**1.9.1 Aim** ..... 31
**1.9.2 Hypothesis** ..... 31
**1.9.3 Objectives** ..... 32
**2. METHODS AND MATERIALS** ..... 33
**2.1 Study population and sample usage** ..... 33
**2.2 DNA extraction from serum or saliva** ..... 36
**2.3 Polymerase Chain Reaction (PCR)** ..... 36
**2.3.1 PCR Optimization** ..... 36
**2.4 Agarose gel electrophoresis** ..... 41
**2.5 Sanger sequencing of *FUT3* and *FUT2* amplicons** ..... 41
**2.6 SNP selection** ..... 42
**2.6.1 Literature review of candidate genes and SNPs of interest** ..... 42
**2.6.2 Minor allele frequency (MAF)** ..... 43
**2.6.3 Tag–SNPs selection** ..... 43
**2.6.4 SNPs selected for iPLEX® MassARRAY®** ..... 43
**2.7 SNP genotyping** ..... 44
**2.8 Saliva phenotyping assay** ..... 47
**(1) First time–point ELISA:** ..... 48
**(2) Second time–point ELISA:** ..... 50
**2.9 Rotavirus–specific antibodies – Serology testing** ..... 50
**2.10 Statistical analysis** ..... 51
**2.10.1 Comparison of seroconverters and non–seroconverters to detect other variables influencing antibody response.** ..... 51
**2.10.2 Analysis of association between demographic or clinical data and ABO/Se/Le phenotypes** ..... 51
**2.10.3 SNP data quality control (QC)** ..... 52
**2.10.4 Allele, genotype, and haplotype frequency analysis of the cohort** ..... 53
**2.10.5 Genotype–phenotype association analysis** ..... 54
**2.10.6 Univariate analysis** ..... 55
**2.10.7 Multivariable analysis** ..... 56
**2.10.8 Epistasis** ..... 56
**3. RESULTS** ..... 57
**3.1 Secretor and Lewis phenotypes and genotypes, and their association with antibody responses to rotavirus vaccination in a South African cohort** ..... 58
**3.1.1 Antibody responses to Rotarix vaccination** ..... 57
**3.1.2 Secretor and Lewis phenotyping by ELISA in saliva samples** ..... 60

**3.1.3 Associations between Secretor and Lewis phenotypes and IgA response to Rotarix vaccination for the two time–points at six weeks and 18 months of age, respectively** ..... 62

**3.1.4 Results of DNA extraction** ..... 66

**3.1.5 Characterisation of *FUT2* and *FUT3* genetic variation in a SA cohort by Sanger sequencing and by MassArray analysis** ..... 67

**3.1.6 Associations between *FUT2* and *FUT3* genetic variation and antibody responses to Rotarix vaccination** ..... 98

**3.1.7 Epistasis** ..... 106

**3.2 ABO phenotypes and genotypes, and their association with antibody responses to rotavirus vaccination in a South African cohort** ..... 107

**3.2.1 ABO phenotypes by ELISA in Secretors, compared at two different time–points** ..... 107

**3.2.2 Characterisation of *ABO* genetic variation by MassArray analysis** ..... 109

**3.3 Genetic variation in innate receptor genes, and their association with antibody responses to rotavirus vaccination in a South African cohort** ..... 118

**3.3.1 Characterisation of innate receptor genetic variation by MassArray analysis** ..... 118

**3.3.2 Associations between genetic variation of innate immune response genes and antibody responses to Rotarix vaccination** ..... 129

**4. DISCUSSION AND CONCLUSION** ..... 133

**4.1 Non–genetic factors influencing seroconversion in our cohort** ..... 134

**4.2 Genetic variation in the current study cohort** ..... 136

**4.3 Do *FUT2* variation and Secretor phenotypes associate with seroconversion to RV1 in our current study cohort?** ..... 137

**4.4 Do *FUT3* variation and Lewis phenotypes associate with seroconversion to RV1 in a SA cohort?** ..... 133

**4.5 Is there an association between ABO phenotypes or genotypes in our current study cohort with seroconversion?** ..... 136

**4.6 Is there an association between genetic variation in innate immune receptors in our current study cohort and seroconversion?** ..... 137

**4.7 Strengths and limitations of the current study** ..... 138

**4.7.1 Strengths** ..... 138

**4.7.2 Limitations** ..... 138

**4.8 Conclusion** ..... 139

**4.9 Implications for future vaccine design and research** ..... 140

**5. REFERENCES** ..... 142

**6. APPENDICES** ..... 167

**Appendix A** ..... 167

**Appendix B** ..... 168

**Appendix C** ..... 169

**Appendix D** ..... 186
**Appendix E** ..... 188
**Appendix F** ..... 192

## LIST OF FIGURES



| | | |
|---|---|---|
| **Figure 1.1** | The structure of RV showing various viral proteins and their functions. | 4 |

| | | |
|---|---|---|
| **Figure 1.2** | Phylogeny of major RV P genogroups (P[I] – P[V]) showing hosts (animal pictures) and host receptors (in red text) recognized by various RV strains. | 7 |
| **Figure 1.3** | Formation of secreted and membrane bound ABOH antigens, Se and Le phenotypes. | 8 |
| **Figure 1.4** | The *ABO* gene structure. | 9 |
| **Figure 1.5** | RNA sensors involved in the detection of pathogens including RNA viruses and some bacteria. | 21 |
| **Figure 1.6** | Innate immune signalling pathways involved in the detection of RV and subsequent recruitment of innate immune responses. | 24 |
| **Figure 2.1** | Different materials sample types and sample sizes in the current study. | 35 |
| **Figure 2.2** | A schematic representation of SNP genotyping using iPLEX® MassARRAY® technology. | 47 |
| **Figure 2.3** | A model for genotype–phenotype association analysis in gPLINK. From Henry et al. (2015). | 54 |
| **Figure 3.1** | A scatter plot showing IgA median and range at six weeks (pre–vaccine), 14 weeks (post dose 1), and 18 weeks (post dose 2). | 57 |
| **Figure 3.2** | Se phenotype frequencies between seroconverters and non–seroconverters for the first time–point. | 62 |
| **Figure 3.3** | Se phenotype frequencies between seroconverters and non–seroconverters for the second time–point. | 63 |
| **Figure 3.4** | Le phenotype frequencies compared between seroconverters and non–seroconverters. | 64 |
| **Figure 3.5** | Le phenotype frequencies compared between seroconverters and non–seroconverters. | 65 |
| **Figure 3.6** | A bar graph showing significant difference in DNA yield from plasma samples (n= 40) vs. saliva samples (n= 78). | 66 |
| **Figure 3.7** | *FUT2* sequence alignment of the current study. | 74 |
| **Figure 3.8** | Examples of electropherograms showing the three novel *FUT3* missense SNPs at positions c.601, c602 and c.607 which occurred in the region of the FUT3_Ex3_21F primer binding site in heterozygous state in our cohort. | 78 |
| **Figure 3.9** | *FUT3* sequence alignment of the current study showing different haplotypes formed by common SNPs found in our cohort. | 84 |
| **Figure 3.10** | LD plot of *FUT2* and *FUT3* markers (Pairwise LD). | 92 |

| | | |
|---|---|---|
| **Figure 3.11** | Predicted Se phenotype counts compared between seroconverters and non–seroconverters. | 102 |
| **Figure 3.12** | Predicted Le antigen counts between seroconverters and non–seroconverters. | 103 |
| **Figure 3.13** | Predicted Le phenotype counts between seroconverters and non–seroconverters predicted from Se phenotypes. | 104 |
| **Figure 3.14** | Combined Le phenotype counts between seroconverters and non–seroconverters predicted from Se phenotypes. | 105 |
| **Figure 3.15** | ABO blood type frequency comparison between current study cohort and the SA Bantu speaking cohort. | 114 |
| **Figure 3.16** | ABO blood type counts between seroconverters and non–seroconverters. | 117 |


## LIST OF TABLES


| | | |
|---|---|---|
| **Table 1.1** | ABO blood groups and their respective genotype frequencies in the Black SA cohort. | 10 |

| | | |
|---|---|---|
| **Table 1.2** | *ABO* gene alleles (occurring in SA) with their corresponding rs ID, phenotypes (ABO subtypes), as well as their frequencies in LWK proxy cohort. | 11 |
| **Table 1.3** | The possible serological ABO phenotypes, genotypes, and respective phenotype frequencies in the Black SA cohort. | 12 |
| **Table 1.4** | The major Le phenotypes caused by the epistatic interaction of the Se and Le antigens. | 14 |
| **Table 1.5** | Se and Le phenotype frequencies in African and non–African populations. | 15 |
| **Table 1.6** | A summary of studies showing the association between HBGAs and susceptibility to RV infection as stratified by the P genotypes. | 17–18 |
| **Table 1.7** | Pre–licensure clinical trials efficacy data of RV1 and RV5 collected from high and middle income, middle–low– and low–income countries. | 26 |
| **Table 2.1** | Seven *FUT2* and *FUT3* primer sets sequences and their respective base pair lengths used for PCR optimization. | 37 |
| **Table 2.2** | PCR master mix components for a 25 µl reaction according to NEB OneTaq® recommendations. | 37 |
| **Table 2.3** | *FUT2* and *FUT3* targeted amplicons and calculated annealing temperatures that were used as references for each of the primer sets during PCR optimization. | 38 |
| **Table 2.4** | NEB OneTaq® PCR protocol used during PCR optimization. | 38 |
| **Table 2.5** | Touch Down PCR protocol used during optimization. | 39 |
| **Table 2.6** | Optimal annealing temperatures and PCR protocols used for *FUT2* and *FUT3*. | 40 |
| **Table 2.7** | Reaction components for PCR of target regions containing the selected SNPs of interest. | 44 |
| **Table 2.8** | PCR protocol for amplification of target regions using thermo cycler. | 45 |
| **Table 2.9** | iPLEX® gold reaction components for single base extension reaction. | 45 |
| **Table 2.10** | Thermal cycler conditions for single base extension reaction using iPLEX® gold reaction components. | 46 |
| **Table 2.11** | Three different models of allelic dominance. | 55 |
| **Table 3.1** | Comparison of seroconverters to non–seroconverters before and after two doses of RV1 vaccine in a South African cohort of 215 infants. | |
| **Table 3.1** | Comparison of seroconverters to non–seroconverters before and after two doses of RV1 vaccine in the current study consisting of 215 infants. | 58 |

| | | |
|---|---|---|
| **Table 3.2** | Comparison of seroconverters to non–seroconverters before and after two doses of RV1 vaccine in a South African cohort of 138 infants with MassArray SNP data. | 59 |
| **Table 3.3** | Se phenotype frequency counts for time–point 1 and time–point 2. | 60 |
| **Table 3.4** | Le phenotype frequency counts at time–point 1 and time–point 2. | 61 |
| **Table 3.5** | Optimal temperatures, PCR protocols used, expected band sizes for the seven *FUT2* and *FUT3* primers used during amplification of cohort DNA samples. | 68 |
| **Table 3.6** | *FUT2* SNPs found through sequencing of 19 samples in our cohort. | 71–72 |
| **Table 3.7** | List of 13 *FUT2* SNPs selected for genotyping using iPLEX® MassARRAY® for the current study. | 76 |
| **Table 3.8** | *FUT3* SNPs in exon 3 found through sequencing in a Black SA cohort. | 80–81 |
| **Table 3.9** | List of 11 *FUT3* SNPs selected for genotyping using iPLEX® MassARRAY® for the current study. | 86 |
| **Table 3.10** | Allele frequencies of *FUT2* and *FUT3* in the current study cohort (SA) (n= 138) compared to LWK from Kenya and YRI from Nigeria | 89–91 |
| **Table 3.11** | Haplotype table of 12 *FUT2* SNPs. | 94 |
| **Table 3.12** | Haplotype table of nine *FUT3* SNPs. | 96 |
| **Table 3.13** | *FUT2* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database. | 97 |
| **Table 3.14** | *FUT3* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database. | 97 |
| **Table 3.15** | Statistical results of univariate analysis of association between *FUT2* and *FUT3* allele frequencies and seroconversion using Chi–square test ($\chi^2$). | 99 |
| **Table 3.16** | Statistical results of univariate analysis of association between *FUT2* and *FUT3* haplotype (21 SNPs) frequencies and seroconversion using Chi–square test ($\chi^2$). | 101 |
| **Table 3.17** | The predilected Le antigens and their respective counts in the current study. | 103 |
| **Table 3.18** | Epistasis analysis results between *FUT2* and *FUT3* SNPs. | 106 |
| **Table 3.19** | Current study cohort first time–point 1 and time–point 2 ABO blood type distribution compared to other SA Bantu speaking cohorts. | 108 |
| **Table 3.20** | *ABO* allele frequencies in the current study cohort (SA) vs. other two African populations i.e., LWK and YRI. | 110–111 |
| **Table 3.21** | Haplotype table of nine *ABO* SNPs. | 113 |

| | | |
|---|---|---|
| **Table 3.22** | *ABO* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database. | 115 |
| **Table 3.23** | Allele frequencies of SNPs in innate immune response genes in the current study cohort (SA) vs. other two African populations. | 119–121 |
| **Table 3.24** | Haplotype table of three *DDX58* SNPs. | 122 |
| **Table 3.25** | Haplotype table of five *IF1H1* SNPs. | 124 |
| **Table 3.26** | Haplotype table of six *TLR3* SNPs. | 126 |
| **Table 3.27** | Haplotype table of six SNPs on chromosome X from *TLR7* and *TLR8*. | 128 |
| **Table 3.28** | Statistical results of univariate analysis of association between innate immune response genes allele frequencies and seroconversion using Chi–square test ($\chi^2$). | 130 |
| **Table 4.1** | Various studies investigating the association between seroconversion (vaccine–take) and HBGAs. | 133 |

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AGE | Acute Gastroenteritis |
| AME | Aseptic Meningoencephalitis |
| BSA | Bovine Serum Albumin |
| CARD | Caspase Activation and Recruitment Domains |
| CI | Confidence Interval |
| DC | Dendritic Cells |
| ddNTPs | Dideoxyribonucleotide Triphosphates |
| DLPs | Double–Layered Particles |
| DNA | Deoxyribonucleic Acid |
| dsDNA | Double–Stranded DNA |
| ELISA | Enzyme–Linked Immunosorbent Assay |
| ENGINES | Entire Genome Interface for Exploring SNPs |
| FBS | Fetal Bovine Serum |
| FUC | Fucose |
| FUT | Fucosyltransferase |
| GAL | Galactose |
| GENO | Maximum SNP Missingness Rate |
| GMT | Geometric Mean Titre |
| GWAS | Genome–Wide Association Studies |
| HBGAs | Histo–Blood Group Antigens |
| HCl | Hydrochloric Acid |
| HIV | Human Immunodeficiency Virus |
| HREC | Human Research Ethics Committee |

| | |
|---|---|
| Hsc70 | Heat Shock Cognate Protein 70 |
| HSE | HSV–1 Encephalitis |
| HSV–2 | Herpes Simplex Virus Type 2 |
| HWE | Hardy–Weinberg Equilibrium |
| IFN | Interferons |
| Ig | Immunoglobulin |
| ISG | IFN Stimulated Gene |
| LD | Linkage Disequilibrium |
| Le | Lewis Phenotype |
| LLR | Lamb Rotavirus Vaccine |
| LMW | Low Molecular Weight |
| LWK | Luhya Webuye Kenya |
| MAF | Minor Allele Frequency |
| MALDI–TOF | Matrix–Assisted Laser Desorption/Ionization–Time–Of–Flight |
| MAVS | Mitochondrial Antiviral Signalling Protein |
| MAX–MAF) | Maximum Minor Allele Frequency |
| MDA5 | Melanoma Differentiation–Associated Gene 5 |
| MIND | Maximum Individual Missingness Rate |
| mRNA | Messenger RNA |
| nAB | Neutralizing Antibodies |
| NF–κB | Nuclear Factor–κB |
| NS | Non–Synonymous |
| NSP | Non–Structural Protein |
| OPD | O–Phenylenediamine |

| | |
|---|---|
| OPV | Oral Polio Vaccine |
| OR | Odds Ratio |
| GRS | Genetic Risk Score |
| PAMP | Pathogen Associated Molecular Patterns |
| PBS | Phosphate–Buffered Saline |
| PCR | Polymerase Chain Reaction |
| $P_{EMPs}$ | Empirical P values |
| PRR | Pattern Recognition Receptor |
| QC | Quality Control |
| RBC | Red Blood Cell |
| RIG–I | RNA Sensors Retinoic Acid–Inducible Gene–I |
| RLR | Retinoic Acid–Inducible Gene I (RIG–I)–Like Receptors |
| RNA | Ribonucleic Acid |
| rs | Reference SNP Identity Numbers |
| RT | Room Temperature |
| RV | Rotavirus |
| RV1 | Rotarix |
| RV5 | Rotateq |
| SA | South Africa |
| SAP | Shrimp Alkaline Phosphatase |
| Se | Secretor Phenotype |
| SIFT | Sorting Intolerant from Tolerant |
| SNP | Single Nucleotide Polymorphism |
| ssDNA | Single–Stranded DNA |

| | |
|---|---|
| ssRNA | Single Stranded Ribonucleic Acids |
| STAT | Signal Transducer and Activator of Transcription |
| T1D | Type 1 Diabetes |
| TBE | Tris/Borate/EDTA |
| TBEV | Tick–Borne Encephalitis Virus |
| TLPs | Triple–Layered Particles |
| TLR | Toll Like Receptor |
| UEA–l | Ulex Europaeus Agglutinin Lectin |
| VLPs | Virus–Like Particles |
| VP | Viral Protein |
| WHO | World Health Organization |
| YRI | Yoruba In Ibadan |
| β–TrCP | β–Transducin Repeat–Containing Protein |

# 1. INTRODUCTION

### 1.1 Rotavirus: the main cause of severe acute gastroenteritis

Rotavirus (RV) is a major cause of severe acute gastroenteritis (AGE) in children and infants below five years of age. Global child mortality rates due to RV infection have declined from four million in the 1970s to 500 000 by 2015 since the introduction of RV vaccines (GBD Diarrhoeal Diseases Collaborators, 2017). In 2010, global mortality rates of children below five years of age were estimated at 1.24 million for 2008 (Black et al., 2010). This estimation was later updated to 752 000 mortalities by the World Health Organization (WHO) for the same year in 2013 (WHO, 2013). A study published in 2016 reported that the number of global RV–induced deaths in children below the age of five decreased from 528 000 in the year 2000 to 215 000 in 2013 (Tate et al., 2016). These estimates are different from those published previously (Fischer–Walker et al., 2012; Lozano et al., 2012; Lanata et al., 2013; GBD, 2013). These differences are attributed to dissimilar data sources and analytical assumptions used in those studies. According to the 2007 WHO global report, 527 000 deaths occurred yearly among children due to RV infection prior to the introduction of RV vaccines. Furthermore, 230 000 of these deaths were reported in Sub–Saharan Africa (WHO, 2007). Moreover, six of the seven countries with highest rates of RV–induced AGE–related deaths are in Africa (Madhi et al., 2010). The latest RV–related mortality rates among infants below five years (from 2013 to 2018) are estimated between 122 322 to 215 757 (GBD, 2016; Tate et al., 2016; Troeger et al., 2018).

In 2009, diarrhoea was the leading cause of mortality in South African children under the age of five and it accounted for 18% of deaths in this age group (Steele et al., 2003). In the same year, RV was found in ~25% of children who were admitted in hospitals for diarrhoea prior to the introduction of RV vaccines in August 2009 (Steele et al., 2003). In 2004, WHO estimated that 271 (95% CI (confidence interval), 224–318) of South African infants died due to RV induced diarrhoea (Seheri et al., 2010). The 2006 South African Health Review indicated that diarrheal diseases were responsible for 10.2% deaths amongst South African children below five years of age in the year 2000. These mortality rates highlight the importance of development of new safer and efficacious RV vaccines.

Whether they live in high–income or low–income countries, virtually all children below five years of age are infected with RV at least once (Gurwith et al., 1981; Velázquez et al., 1996). However, RV epidemiology is different among various settings around the world (Levy et al., 2009). For example, RV infection is common during the winter season in countries with

temperate climates, whilst showing less precise seasonality in tropical climate settings. In these settings, RV infection tends to occur throughout the year and RV strain diversity is often greater (Levy et al., 2009). Studies have shown that 75% of infants in developing countries usually acquire their first RV infection at birth as compared to those in high–income countries (Bahl et al., 2005; Cunliffe et al., 1998).

RVs are commonly spread in a faecal–oral manner and also by close contact between individuals (Estes et al., 2007). Furthermore, RVs can also be spread via contaminated fomites, usually in hospitals and out–of–home settings (Yen et al., 2011). Fomites are contaminated inanimate objects that facilitate transfer of pathogens such as viruses and bacteria into a new host. In susceptible hosts, very few infectious RV virions are needed to induce an infection (Bishop, 1996). RV–induced AGE manifests itself in the form of mild (watery diarrhoea) to severe diarrhoea which involves fever and vomiting (Yen et al., 2011). These symptoms subsequently lead to severe dehydration which is in fact what causes death of the infected individuals (Matson et al., 2003). However, less common cases of meningoencephalitis caused by RV infection have previously been observed (Dickey et al., 2009). Furthermore, complications such as hepatitis, hemophagocytic lymphohistiocytosis, polio–like paralysis, and acute myositis have also been reported (Gilger et al., 1992). However, their relationship with RVs has not yet been clearly described (Yen et al., 2011).

RV viremia and antigenemia in blood samples have been discovered in children with RV infection. Thus, RV infection is not specifically confined to the gut, however clinical importance of these findings is still unclear (Blutt et al., 2003; Fischer et al., 2005; Blutt and Conner, 2007). It has been reported that viral shedding is prolonged in RV infected patients who are severely immunocompromised, for instance; those who had undergone solid organ and bone marrow transplantation (Yen et al., 2011). In contrast, children who have human immunodeficiency virus (HIV) infection do not appear to suffer more severe symptoms because of RV infection (Cunliffe et al., 2001; Jere et al., 2001; Steele et al., 2009).

## 1.2 Rotaviruses structure, classification, life cycle, and pathogenesis

### 1.2.1 Rotavirus structure

RVs are non–enveloped viruses which are double–stranded (ds) ribonucleic acid (RNA) and are 100 nm in diameter. Their structure involves a triple–layered protein capsid that contains 11 segments of the dsRNA (which is the viral genome) and is 18 555 nucleotides in total length (Estes and Kapikian, 2007; Greenberg and Estes, 2009;) (**Figure 1.1**). Each helix, or segment, is a gene which is numbered from one up to 11 by their decreasing sizes. Each gene codes for one viral protein (VP), except gene 9, which codes for two proteins. Six of the 12 viral proteins are structural VPs: VP1–VP4, VP6, and VP7, and six are non–structural proteins (NSPs) (NSP1–NSP6). VP1 encodes RNA polymerase (which produces messenger RNA (mRNA) transcripts) while VP3 encodes guanylyl transferase which adds the 5' cap to the mRNA. VP2 encodes the inner capsid protein and VP6 encodes the intermediate capsid (middle layer) or group antigen. On the other hand, VP4 and VP7 encode the outer capsid or the outer shell. VP4 is usually cleaved into VP5* and VP8* through proteolysis (Settembre et al, 2011). In turn, VP5* and VP8* bind to host receptors and are important for viral attachment leading to RV infection. VP8* of RVs infecting humans particularly interacts with human histo–blood group antigens (HBGAs) (Settembre et al., 2011; Liu et al, 2012). Subsequent research has shown that polymorphisms and developmental regulation of HBGA expression may lead to susceptibility to various RV strains (Hu et al., 2012; Imbert–Marcille et al., 2013; Ramani et al., 2013). VP5* and VP8* induce neutralizing antibodies (nAb) and also forms the “P” serotype. VP7 also induces nAb and is responsible for the “G” serotype.

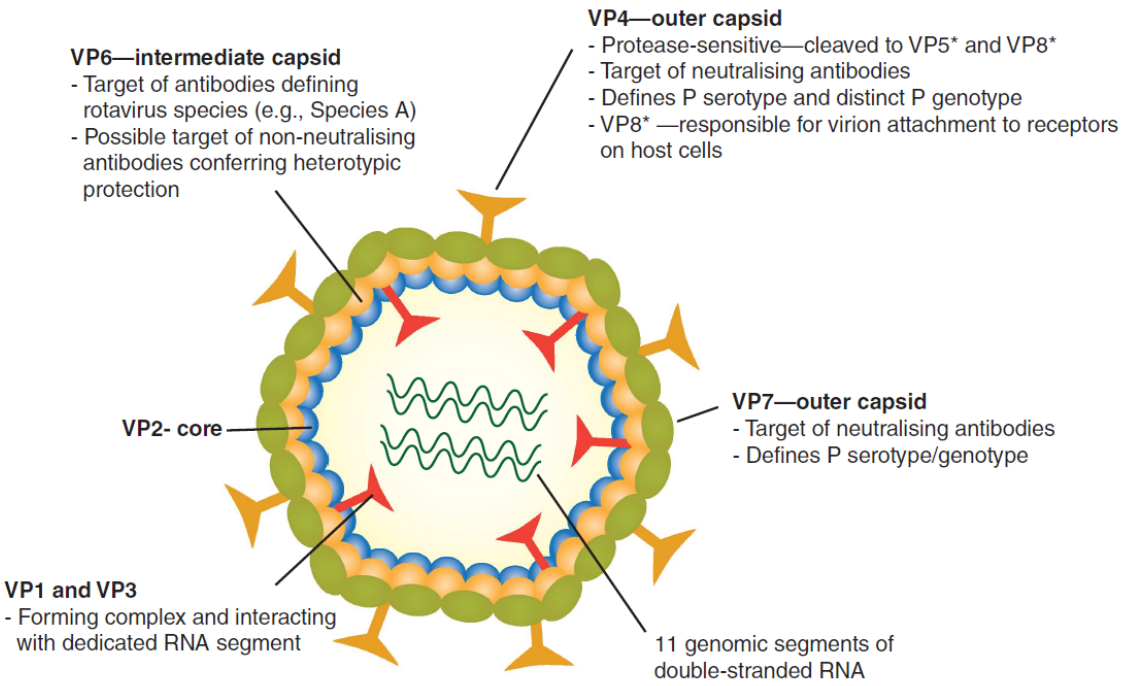



**Figure 1.1** The structure of RV showing various viral proteins and their functions. From Clarke and Desselberger, (2015).

In summary, the infectious RV particle comprises of three concentric layers of protein containing a genome which is made up of eleven segments of dsRNA.

### 1.2.2 Rotavirus classification

RVs are commonly classified into groups by different methods and according to different antigenic variation:

**1.** Group (A–G) and subgroup [I] or [II]: are determined by VP6 inner capsid variation. Group A, and less commonly B and C RVs account for most of the RV disease in humans, globally (Hoshino and Kapikian, 2000).

**2.** Serotype: antibodies against the VP7 protein determine the G serotype, while antibodies against VP4 determine the P serotype. There are at least 14 RV G serotypes and approximately 11 RV P serotypes in humans and animals (Hoshino and Kapikian, 2000).

**3.** P and G genotypes and P genogroups: P and G genotypes have been determined from VP4 and VP7 amino acid sequences. There is a significant concordance between G genotypes and G serotypes, but P genotypes have different numbering from P serotypes (Hoshino and Kapikian, 2000). The complete genotype of an RV is written as the G genotype number, then

the P genotype number e.g., G1P[8]. More than 70 combinations of VP4 and VP7 were reported to infect humans (Santos and Hoshino, 2005). In particular, G1P[8], G2P[4], G3P[8], G4P[8], and G9P[8] account > 90% of circulating viruses worldwide, with Africa having the highest RV strain diversity (Santos and Hoshino, 2005). The most common RV genotypes recorded in South Africa (SA) in 2014 and 2015 included G1P[8], G2P[4], and G9P[8] (Page et al., 2018). RV strain diversity is mainly due to zoonotic transmission and reassortment events between co–infecting strains (Santos and Hoshino, 2005; Todd et al., 2010).

P genotypes are further grouped into five genogroups according to their VP8* sequences (VP4 subdomain) (Jiang et al., 2017). The five genogroups are:

- P[I] – comprises of many P genotypes infecting humans and humans
- P[II] – comprises of common human P [4], [6], [8], and [9] strains
- P[III] – comprises of P[14], [9] and [25] strains
- P[IV] – compromises of P[11] strain only
- P[V] – compromises of P genotypes that infect birds P[17], [30], [31], and [35]

Lastly, each genogroup uses a unique set of host receptors which enables them access to the host cell (see **Figure 1.2**) as discussed further in **Section 1.3**.

### 1.2.3 Rotavirus life cycle and pathogenesis

RVs infect enterocytes located on the villi of small intestine and this allows them to replicate in the gut. Subsequently, the epithelium becomes subjected to functional and structural modifications (Desselberger, 2014). RVs are resistant to the acidic pH of the stomach and the gut digestive enzymes, due to their triple layered capsid structure. VP4 is then modified into VP5* and VP8* subdomains by trypsin. RV life cycle involves viral attachment, which is followed by penetration and uncoating, and then mRNA synthesis, followed by viroplasm formation, and then RNA synthesis and double–layer particle (DLP) formation, followed by maturation and release (**Figure 1.2**) (Desselberger, 2014).

Attachment RV to the host cell starts by the interaction of triple–layered particle (TLP) (infectious RV particle) VP4 and VP8* proteins with the host cell receptors (**Figure 1.2**) (López and Arias, 2004). Due to their precise cell tropism, RVs only infect intestinal villi enterocytes, which suggests that specific host receptors exist. A certain number of animal RVs recognize sialic acid on the host cell surface during infection. However, human RVs are virtually sialidase–insensitive (Haselhorst *et al*., 2009). Human RV strain (Wa) recognizes an

internal sialic acid. Studies have shown that some sialic acid–sensitive RV strains use other receptors containing sialic acid such as GM1 gangliosides (Haselhorst *et al*., 2009). Furthermore, certain cell surface molecules can also act as co–receptors during after the attachment process. These molecules include certain integrins and human heat shock cognate protein 70 (Hsc70) (Desselberger, 2014).

RVs cause villous blunting due to obliteration of enterocytes, resulting in absorption reduction during their lifecycle (Ramani and Atmar, 2015). In addition, NSP4 (an enterotoxin) plays a critical role in the secretion component of diarrhoea. It does this by activating calcium channels in the infected cells (Estes and Morris, 1999), which is followed by exacerbation of chloride secretion, subsequent reduction in border enzymes e.g., sucrase isomaltase. Chloride secretion leads to an osmotic gradient resulting in diarrhoea (Jourdan *et al*., 1998).

### 1.3 HBGAs as rotavirus receptors

Saliva and oligosaccharide binding assays studies showed that HBGAs are receptors for RVs (Huang et al., 2012). RV strains use HBGAs as receptors during an infection (Huang et al., 2012). For example, RV P[8] and P[4] strains particularly recognize Lewis (Le) B antigen, whilst P[6] recognizes H type 1 antigen (Secretor phenotype) as shown in **Figure 1.2**. On the other hand, P[11] from genogroup [IV] attach to HBGAs, specifically H precursor type 2 (Hu et al., 2012; Huang et al., 2012; Ramani et al., 2013).

Nearly all RV P genotypes of the genogroups P[II]–P[IV] that usually infect humans, bind HBGAs (**Figure 1.2**). Furthermore, RVs have also been reported to bind mucin cores of mucin O–glycans (Liu et al., 2016), which can be extended with saccharide residues such as N–acetylglucosamine (GlcNac), galactose (Gal), sialic acid, and fucose (Fuc) (Taliford et al., 2015).

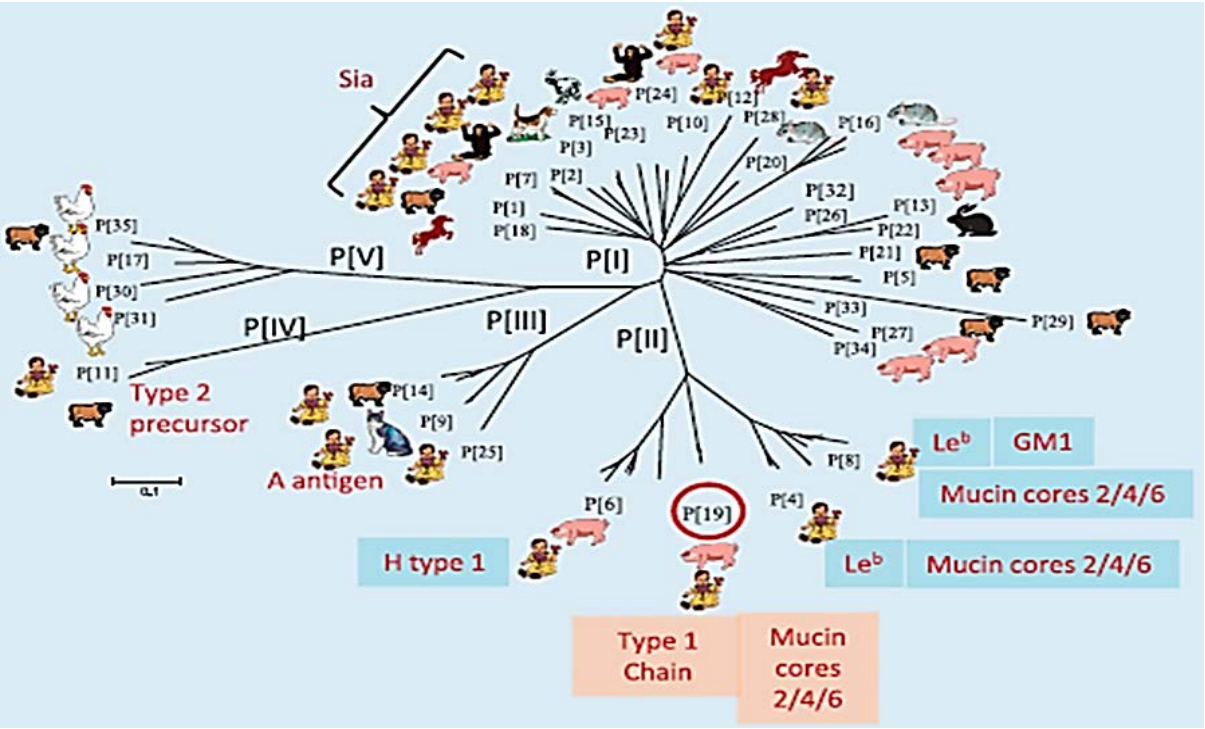



**Figure 1.2** Phylogeny of major RV P genogroups (P[I] – P[V]) showing hosts (animal pictures) and host receptors (in red text) recognized by various RV strains. From Jiang et al. (2017). **Le B = Lewis B, Sia = sialic acid dependent, A and H = histo–blood group antigens A and H, GM1 = gangliosidoses.*

### 1.4 HBGA phenotypes

HBGAs are defined as oligosaccharides commonly found on the surfaces of erythrocytes and mucosal epithelia. HBGAs are particularly known for their role in blood transfusion compatibility and include the ABO blood group antigens which form the A, B, AB, and O blood types. In some individuals, ABO antigens are secreted into the blood serum, intestinal contents, saliva, and milk (Ravn and Dabelsteen, 2000). This phenomenon is known as having a "Secretor" ABO phenotype. Lewis blood phenotypes are also related to the ABO and Secretor phenotypes as explained below.

The ABO, Secretor (Se), and Lewis (Le) phenotypes are produced via a series of enzymatic reactions which add various monosaccharides to a substance known as "precursor" or paragloboside. Many of the enzymes in this pathway are fucosyl transferases. The resulting oligosaccharides consist of various combinations of Gal, GlcNAc, and Fuc. Precursor (lacto–N–neotetraosyl ceramide or I antigen or paraglobiside), the biosynthetic precursor of the ABOH antigens, consists of GlcNAc and galactose sugars with no Fuc residues. The precursor

exists in two forms: precursor type 1 and precursor type 2 (**Figure 1.3**). GlcNAc connected to Gal by β–1–3 linkage creates precursor type 1 (Galβ3GlcNAcβ), while GlcNAc connected to Gal by β–1–4 linkage create precursor type 2 (Galβ4GlcNAcβ) (**Figure 1.3**). Precursor type 2 is the receptor for the P[IV] genogroup of RV (Ravn and Dabelsteen, 2000, **Figure 1.3**).

Precursor type 2 is modified into membrane–bound H substance type 2 by the fucosyltransferase–1 (FUT1) enzyme in red blood cells (RBCs) (**Figure 1.3**, Jiang et al., 2017). The unmodified H substance type 2 becomes the O blood group antigen (**Figure 1.3**). Further modification of H substance type 2 by α–1,3–N–acetyl galactosamine transferase (A Enzyme) (addition of GalNac) or α–1,3–galactosyl transferase (B enzyme) (addition of Gal) forms the A and B blood types or antigens, respectively (Ravn and Dabelsteen, 2000). The A antigen is the receptor for the RV P[III] genogroup (Jiang et al., 2017, **Figure 1.2** above). Furthermore, human group C RV VP8*s have been shown to recognize type A HBGAs as ligands (Sun et al., 2018). A lack of FUT1 forms the Bombay phenotype whereby O, A, nor B blood antigens are absent.

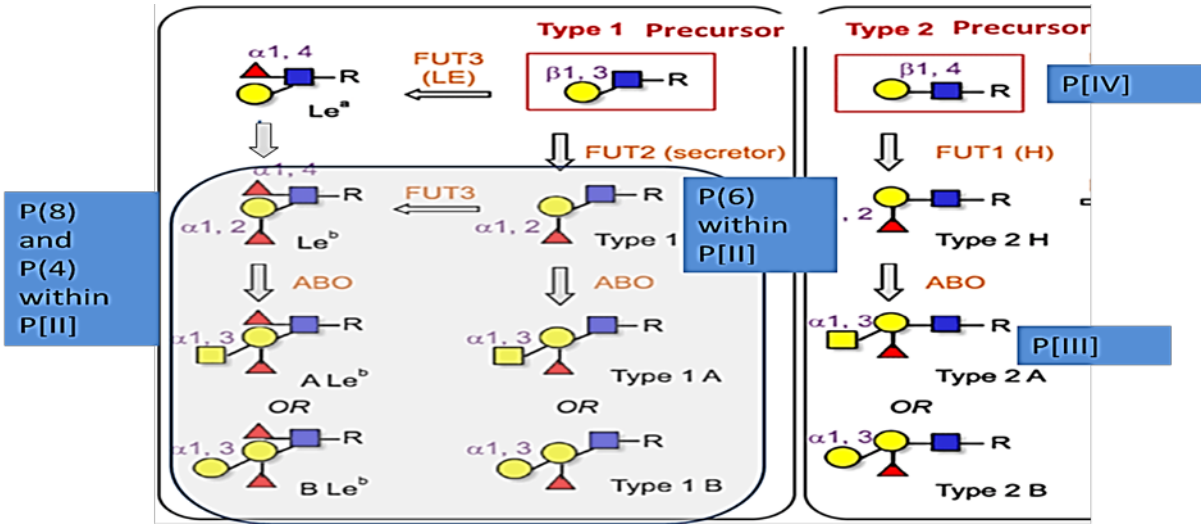



**Figure 1.3** Formation of secreted and membrane bound ABOH antigens, Se and Le phenotypes. The grey box indicates Se phenotypes. Membrane–bound H type 2 (O antigen) is made from type 2 precursor by FUT1, and secreted H antigen type 1 (secreted O antigen) is made from the type 1 precursor by FUT2. Addition of Gal to H1 or H2 by galactosyl transferase creates blood antigen B, while addition of GalNac to H by galactosamine transferase causes blood antigen A. Le phenotypes are created by the action of FUT3 on either precursor type 1 or H type 1. *Yellow circle = Gal; Blue square = GlcNAc; Red triangle = fucose; Yellow square = GalNAc. The blue boxes indicate RV genogroups that can bind to this HBGA.* Adapted from Kwon et al. (2015).

### 1.4.1 Formation of the ABO blood types

The A and B enzymes (mentioned previously) are responsible for the formation of the A, B, AB, or O blood types from the type 2 H antigen as encoded by the *ABO* gene (Daniels, 2005). *ABO* is a 19kb gene comprising of seven exons found on chromosome 9q34.2 (**Figure 1.4**). According to Daniels (2005), exons 6 and 7 makes up 77% of the coding region of the *ABO* gene.

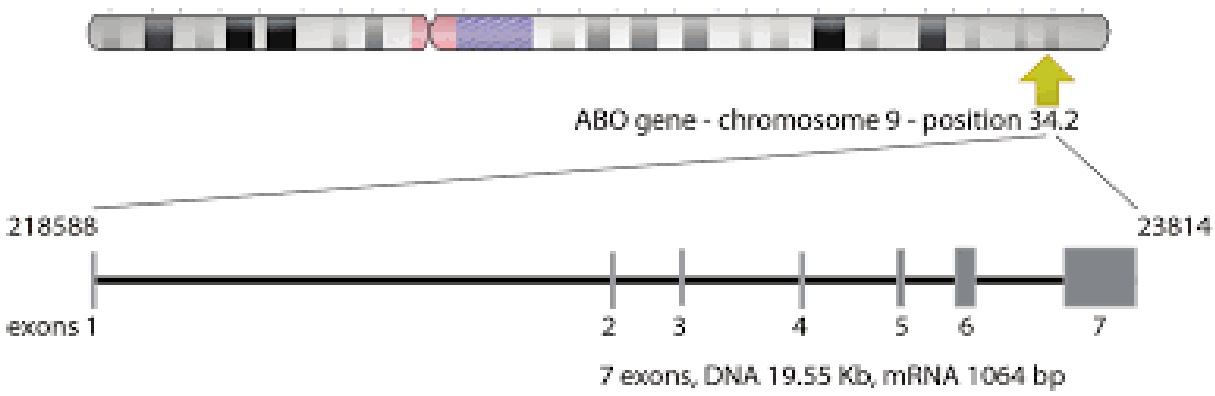



**Figure 1.4** The *ABO* gene structure. Located on chromosome 9 at position 34.2, the ABO gene is 19kb long and comprises of seven exons which encode 1064 bp mRNA. From Misevic (2018).

### 1.4.2 Variations in the *ABO* gene

Variations in the *ABO* gene have previously been documented and three major alleles have been reported: $A_1$, $B_1$, and $O_1$. In particular, the $A_1$ allele has been reported to be the ancestral allele and thus it is used as a reference allele (Hosseini–Maaf et al., 2005; Kitano et al., 2012). The $A_1$ and $B_1$ alleles carry four major SNPs: 526C>G (rs7853989), 703G>A (rs8176743), 796C>A (r8176746), and 803G>C (rs8176747) (Cozzi et al., 2017). These SNPs are found in exon 7 of the ABO gene whereby they induce missense mutations (Cozzi et al., 2017). The amino acids at position 266 of 796C>A (r8176746) and 268 of 803G>C (rs8176747) determine whether A enzyme or B enzyme is formed (Dean, 2005a). On the other hand, the $O_1$ allele contains a base deletion of guanine at position 261 of (261delG, rs8176719). This deletion induce a frameshift mutation leading to an inactive enzyme (Yamamoto et al., 2012; Yip, 2002). Therefore, the A and B enzymes synthesise four major blood antigens or blood types: A, B, AB, and O encoded. These blood types are encoded by the following genotypes, respectively: AA/AO, BB/BO, OO, and AB (Dean, 2005a). **Table 1.1** below shows genotype frequencies for a Black SA cohort from the Brain et al. (1966) study.

**Table 1.1** ABO blood groups and their respective genotype frequencies for a Black SA cohort (n= 867) from the Brain et al. (1966) study.

| Blood group | Genotype | Frequency in Black SA population (%) |
|---|---|---|
| A | AA or AO | 30.7 |
| B | BB or BO | 19.4 |
| O | OO | 43.8 |
| AB | AB | 6.1 |

*SA – South Africa

Various ABO subtypes induced by different SNP combinations and variants of the major alleles have been discovered and described. Furthermore, over 200 *ABO* gene alleles have been found (Yamamoto et al., 2012). ABH antigens are known to be expressed ubiquitously across different populations around the world. However, specific alleles, subtypes, and frequencies vary across different populations (Jiang et al., 2017). For example, the following alleles have been specifically documented in the SA population: $A_1$, $A_2$, $A_{bantu}$, $B_1$, $O_1$, $O_{1v}$, and $O_{1bantu}$ (Brain, 1966; Moores, 1991) (**Table 1.2**).

**Table 1.2** *ABO* gene alleles for the current study cohort with their corresponding rs ID, phenotypes (ABO subtypes), as well as their frequencies in LWK proxy cohort. These frequencies were obtained from Ensembl. (Moller et al., 2016; "SNP linked to Gene (geneID:28) Via Contig Annotation," n.d.; Yip, 2002; Zhang et al., 2012).

| Exon | SNP | rs ID | Phenotype (ABO subtypes) | Frequency in LWK cohort |
|---|---|---|---|---|
| 4 | 188G>A | rs549446 | $O_{1v}$ | 0.35 |
| 4 | 189C>T | rs549443 | $O_{1v}$ | 0.23 |
| Exon 4/Intron 4 | 203+1delG | rs782023144 | $A_{bantu}$, $O_{1bantu}$ | 0.001 |
| 6 | 261delG | rs8176719 | O | 0.71 |
| 6 | 297A>G | rs8176720 | $O_{1v}$, $O_{1bantu}$, $B_1$ | 0.5 |
| **7** | 467C>T | rs1053878 | $A_2$, $A_{bantu}$, $A_{1.02}$ | 0.2 |
| **7** | 526C>G | rs7853989 | $B_1$ | 0.17 |
| **7** | 646T>A | rs8176740 | $O_{1v}$ | 0.25 |
| **7** | 657C>T | rs8176741 | $B_1$ | 0.17 |
| **7** | 681G>A | rs8176742 | $O_{1v}$ | 0.13 |
| **Exon** | **SNP** | **rs ID** | **Phenotype (ABO subtypes)** | **Frequency in LWK cohort** |
| **7** | 703G>A | rs8176743 | B1 | 0.17 |
| **7** | 771C>T | rs8176745 | O1v | 0.26 |
| **7** | 796C>A | rs8176746 | B1 | 0.16 |
| **7** | 803G>C | rs8176747 | B1 | 0.16 |
| **7** | 829G>A | rs8176748 | O1v | 0.25 |
| **7** | 1061delC | rs56392308 | $A_2$, $A_{bantu}$ | 0.05 |

*LWK – Luhya Kenya

### 1.4.3 ABO alleles common in South Africa

As previously mentioned, the $A_1$ allele is commonly used as a reference allele and has a frequency of 18.62% in the Black SA population (**Table 1.3**) (Brain , 1966). The $A_2$ allele has a single base deletion (1061delC, rs56392308) which leads to a synthesis of an A enzyme with 30–50 fold less enzymatic activity compared to the B enzyme (Daniels, 2005; Yip; 2002). Encoded by $A_1A_2$ genotype, the $A_{int}$ allele has a frequency of 4.25% and it is serologically stronger than $A_2$ but weaker than $A_1$ (**Table 1.3**) (Brain, 1966). According to Yoshida et al.

(1985), approximately 80 percent of the H antigens remain unglycosylated in the presence of the $A_{int}$ allele. On the other hand, the $O_{1v}$ has the same 261delG SNP as the major $O_1$ variant, however, it has nine additional SNPs the makes different from the $O_1$ allele (**Table 1.3**) (Daniels, 2005). Nevertheless, the $O_1$ and $O_{1v}$ alleles are functionally similar and induce the O phenotype, despite their differences (Daniels, 2005). The $A_{bantu}$ allele is globally rare, however, it was reported to be common in the Black SA cohort at a frequency of 1.26% (Brain, 1966). $A_{bantu}$ occurred as result of recombination process between $A_2$ and $O_{1bantu}$ alleles since it shares two similar SNPs with $A_2$ allele (1061delC, rs8176750, 467C>T, rs1053878) and a G deletion mutation with $O_{1bantu}$ allele. The deletion occurs at a junction between exon 4 and intron 4 (Hosseini–Maaf et al., 2005; Yip, 2002). The $A_{bantu}$ allele produces a weak A enzyme such that when it is heterozygous with an O allele, the blood type can stereotypically be misclassified as O (Hosseini–Maaf et al., 2005). The order of the A antigen strength from least to the strongest on RBCs is as follows: $A_1$, $A_{int}$, $A_2$, and $A_{bantu}$ (Grundbacher Summerlin, 1971).

**Table 1.3** The possible serological ABO phenotypes, genotypes, and respective phenotype frequencies in a Black SA cohort (n= 867) (Brain, 1966).

| **Phenotype** | **Genotypes** | **Frequency in a Black SA cohort** |
|---|---|---|
| $A_1$ | $A_1A_1/O_1A_1/O_{1v}A_1/O_{1bantu}A_1$ | 0.187 |
| $A_{int}$ | $A_2A_1/A_{bantu}A_1/$ | 0.043 |
| $A_2$ | $A_2A_2/$ $A_{bantu}$A2/O1$A_2$/ $O_1vA_2/O_{1bantu}A_2$ | 0.068 |
| $A_{bantu}$ | $A_{bantu}Ab_{antu}$ | 0.013 |
| $B_1$ | $B_1B_1/O_1B_1/O_{1v}B_1/O_{1bantu}B_1$ | 0.2 |
| AB | $A_1B_1$ | |
| $A_{weak}$B | $A_2B_1/A_{bantu}B_1$ | 0.06 |
| O | $O_1O_1/O_1O_{1v}/O_{1v}O_{1v}/O_1A_{bantu}/O_{1v}A_{bantu}/$ $O_{1bantu}A_{bantu}$ /$O_{1bantu}O_1$/ $O_{1bantu}O_{1v}/O_{1bantu}O_{1bantu}$ | 0.44 |

### 1.4.4 Formation of the Secretor (Se) phenotype

The Se phenotype is formed when precursor type 1 is modified into the H antigen type 1 by addition of a Fuc residue to the last Gal of type I precursor at the α–1,2 position, mediated by enzyme fucosyltransferase–2 (FUT2) (**Figure 1.4**). Presence of H type 1 confers the Se phenotype. Non–Secretors (also referred to as "se" phenotype) usually express inert FUT2 enzyme because of mutations in the *FUT2* gene and therefore, those individuals cannot express the ABH antigens in their saliva or other fluids (Koda et al., 2001). Mutations in the *FUT2* gene may be called by their identifying rs number, or by their coding position, as well as by the prefix "Se" or 'se" indicating that the mutation causes an active or inactive FUT2 enzyme, respectively. For example, rs601338 or se428 is an inactivating missense mutation causing a non–Secretor phenotype that is common in African populations (Koda et al., 2001).

H type 1 antigen is the receptor for P[6] RV strain within the P[II] genogroup (Koda et al., 2001, **Figure 1.3**). Again, further modification of H substance type 1 by α–1,3–N–acetyl galactosamine transferase (addition of GalNac) or α–1,3–galactosyl transferase (addition of Gal) creates the A and B blood antigens, respectively. However, these are secreted antigens rather than bound to the RBC membrane (**Figure 1.4**). Koda et al. (2001) reported that approximately 20% of the world population are non–Secretors and 80% are Secretors (with active *FUT2*). Chanzu et al. (2015) demonstrated that 212 (75.7%) of Africans are Secretors compared to the 68 (24.3%) who are non–Secretors in a study comprising of 280 females between 18–65 years of age from Kenya.

### 1.4.5 Formation of the Lewis (Le) phenotype

Enzyme fucosyltransferase–3 (FUT3) can cause presence or absence of the Le antigens, depending on genetic polymorphisms and whether the enzyme is active or not. Mutations in the *FUT3* gene may be called by their identifying rs number, or by their coding position, as well as by the prefix "Le" or 'le" indicating that the mutation causes an active or inactive FUT3 enzyme, respectively. For example, rs3745635 or le508 is an inactivating missense mutation that is common in African populations (Koda et al., 2001).

The action of FUT3 and the type of antigens created, is usually considered in combination with the action of FUT2 (which creates secreted H substance type 1 if the FUT2 enzyme is functional). Enzyme fucosyltransferase–3 (FUT3) can add Fuc to precursor type 1 (from non–functional FUT2) or H antigen type 1 (from functional FUT2) to create the Le A and Le B

antigens respectively (**Figure 1.4**). Le antigens are passively adsorbed onto the surface of RBCs (Koda et al., 2001) and are not produced by the cell itself. Secreted Le A is formed when FUT3 enzyme adds Fuc to precursor type 1, whilst secreted Le B is formed when FUT3 adds Fuc to H antigen type 1. Secreted Le B antigen can also be formed by the action of FUT2 on the secreted Le A antigen (Chandrasekaran et al., 1995; Chandrasekaran et al., 2002). Secreted Le B antigen is the receptor for the P[II] genogroup of RVs (**Figure 1.3**) (Jiang et al., 2017). Le A or Le B phenotypes remain present if H antigen is then converted to A or B antigen.

When the combined action of the FUT2 and FUT3 enzymes are considered, there are four possible Le phenotypes: Le (A+B–), Le (A–B+), Le (A+B+) and Le (A–B–). If only the FUT2 locus is functional (FUT3 inactivated), the phenotype will be Se+ and Le (A–B+). If only the FUT3 locus is functional (FUT2 inactivated) the phenotype will be a non–Secretor and Le (A+B–). If both FUT2 and FUT3 are functional, the phenotype will be Se+ and theoretically could be Le (A+B+), however nearly all the Le A is converted to Le B and so the phenotype appears as Se+ Le (A–B+) (Jiang et al., 2017). In summary, the presence of the Le A antigen correlates with a non–Secretor phenotype, whilst the presence of Le B antigen correlates with the Se phenotype. The presence of both Le A and Le B suggests a weak Secretor (weak FUT2) enzyme and Le– individuals could be either Secretors or non–Secretors (**Table 1.4**).

**Table 1.4**: The major Le phenotypes caused by the epistatic interaction of the Se and Le antigens.

| | **Se+** | **Se– (also called se)** |
|---|---|---|
| **Le+** | Le (a–b+)<br>Weak Se+ might have Le (a+b+) | Le (a+b–) |
| **Le– (also called le)** | Le (a–b–) | Le (a–b–) |

Le and Se phenotype frequencies in different African populations vs. non–African populations are shown in **Table 1.5**. The Se+ Le B+ phenotype (RV P[II] receptor caused by combined FUT2 presence and FUT3 presence) is the most common phenotype in all groups below, but less frequent in populations with African ancestry than other ethnicities. The Se+ Le– phenotype (caused by FUT2 presence and FUT3 absence) is significantly higher in African populations than in non–African populations, while the Se– Le (A–B–) phenotype (combined FUT2 absence and FUT3 absence) is only found in populations with African ancestry.

**Table 1.5** Se and Le phenotype frequencies in African and non–African populations.

| Se/Le phenotype | Le phenotype | Caucasians | Nicaragua (Central America) | Af–Am | Burkina Faso (West Africa) | SA |
|---|---|---|---|---|---|---|
| **Reference** | | Miller et al. (1954) | Nordgren et al. (2014) | Miller et al. (1954) | Nordgren et al. (2014) | Moores and Brain (1968) |
| **N** | | 460 | 34 (all RV infected) | 211 | 208 | 171 |
| **Se+ Le+** | Includes Le a+ Le b+ Le a– Le b+ | 71.5% | 94% | 54.5% | 54% | 53% |
| **Se– Le+** | Le a+ Le b– | 22.8% | 0 | 23.2% | 14% | 22% |
| **Se+ Le–** | Le a– Le b– | 5.7% | 6% | 22.3% | 25% | 20% |
| **Se– Le–** | Le a– Le b– | 0 | 0 | 0 | 6.7% | 5% |

**Se – Secretor, Le – Lewis, Af–Am – African American, SA – South Africa*

Se and Le variation has been associated with several infectious and non–infectious diseases. The incidence of HIV, *Neisseria gonorrhoea*, *Trichomonas vaginalis*, and bacterial vaginosis, noroviruses, *Helicobacter pylori*, and *Campylobacter jejuni*, was much higher in Secretors compared to non–Secretors (Chanzu et al., 2015) due to changes in pathogen adherence to the mucous layer. FUT2 (Se) status has previously been implicated in the susceptibility to RV and norovirus infections (Nordgren et al., 2014; Vang Trang et al., 2014; Payne et al., 2015), as well as diabetes type 1 (Perrett et al., 2019) and cancer (Akhtar et al., 2018) susceptibility.

### 1.4.6 Changes in HBGA expression with age

The otherwise unusual phenotype Le (A+B+) may also appear in Se infants because of partial FUT2 activity during infancy (Jiang et al., 2017 ). Le antigen is weakly expressed during pregnancy and (Le (A–B–) phenotype is commonly seen during gestation (Jiang et al., 2017). Thus, most newborns will type as Le (A–B–) (Jiang et al., 2017). ABO, FUT2, and FUT3 are major enzymes driving the expression of HBGAs leading the formation of A/B/H, Se/se/ and Le A/B antigens.

### 1.4.7 Using haplotypes to define HBGA variation

A haplotype can be described as a set of alleles which are located on the same segment of a chromosome and are usually transmitted together as a block (Strachan and Read, 2011). Use of haplotypes allows us to define functional Se, Le, and ABO alleles rather than just defining by individual SNPs (each allele can be defined by multiple SNPs that are linked in a haplotype).

A tag–SNP can be described as a representative SNP with high LD located a specific chromosomal region. Therefore, tag–SNPs with high LD are usually in close promixmity with each other on a loci and form a haplotype (Strachan and Read, 2011).

## 1.5 HBGAs and susceptibility/ resistance to rotavirus infection

As previously discussed, several RV P genogroups have been shown to use HBGAs as host cell receptors (**Figure 1.2** and **Figure 1.3**). The relevant genogroups are P[II] including genotypes P[8] and P[4] with HBGA Le B antigen and P[6] (part of P[II]) with H type 1 antigen, P[III] including genotype P[9] with the A antigen, and P[IV] including genotype P[11] with type 2 precursor. Different studies reported a significant association between susceptibility to P[8] RV infection (genogroup [PII]) and Se+ Le+ phenotypes, thereby demonstrating the role of Le B antigen as a host cell receptor for P[8] strains, as summarized in **Table 1.6** and discussed in more details below.

**Table 1.6** A summary of studies showing the association between HBGAs and susceptibility to RV infection as stratified by the P genotypes. Adapted from Sharma et al. (2020).

| P–genotype | Country | Lewis | ABO | Secretor | Reference |
|---|---|---|---|---|---|
| **P[4]** | Nicaragua | Only Le B infected | Not investigated | Only Secretors infected | Nordgren et al., 2014 |
| | China | Le A less infected | No association | Secretors more infected | Zhang et al., 2016 |
| | China | No association | No association | Only Secretors/partial Secretors infected | Sun et al., 2016 |
| | Malawi | Le– less susceptible | No association | Secretors more infected | Pollock et al., 2018 |
| | Bangladesh | Only Le B infected | Not investigated | Only Secretors infected | Lee et al., 2018 |
| | Vietnam | Le B present in all infected | No association | Only Secretors/partial Secretors infected | Vang Trang et al., 2014 |
| | USA | Not investigated | Not investigated | Only Secretors infected | Payne et al., 2015 |
| | | | | | |
| **P[6]** | Burkina Faso | Le– more susceptible | No association | No association | Nordgren et al., 2014 |
| | Malawi | Le– more susceptible | No association | No association | Pollock et al., 2018 |
| | Bangladesh | Le– more susceptible | Not investigated | No association | Lee et al., 2018 |
| | Vietnam | Le– more susceptible | No association | Not investigated | Vang Trang et al., 2014 |
| | | | | | |
| **P[8]** | Burkina Faso | Only Le B infected | No association | Only Secretors infected | Nordgren et al., 2014 |

| P–genotype | Country | Lewis | ABO | Secretor | Reference |
|---|---|---|---|---|---|
| **P[8]** | Nicaragua | Only Le B infected | Not investigated | Only Secretors infected | Nordgren et al., 2014 |
| | China | Le A less infected | No association | Secretors more infected | Zhang et al., 2016 |
| | China | Le A less infected | No association | Secretors more infected | Sun et al., 2016 |
| | Spain | Le B more infected | Blood group A and AB more infected compared to O | Secretors more infected | Perez–Ortin et al., 2019 |
| | Malawi | Le– less infected | No association | Secretors more infected | Pollock et al., 2018 |
| | Bangladesh | Le– less infected | Not investigated | No association | Lee et al., 2018 |
| | Vietnam | No Le A infected | No association | Only Secretors/partial Secretors infected | Vang Trang et al., 2014 |
| | France | Not investigated | Not investigated | Secretors more infected | Imbert–Marcille et al., 2014 |
| | USA | Not investigated | Not investigated | Apart from 1, only Secretors infected | Payne et al., 2015 |
| | Tunisia | Only Le positives infected | No association | No association | Ayouni et al., 2015 |

Regarding P[8] RV infections and Se (FUT2) status; a study based on a French cohort showed that 100% of 52 patients with RV were Se+ (Imbert–Marcille et al., 2013). In contrast, only 19% of the healthy adult control samples (n= 95) were non–Secretors. There was significant difference reported between genotypes of the controls and cases (P< 0.0007). These results suggest that lack of *FUT2* expression may render resistance to the P[8] strain (Imbert–Marcille et al., 2013). Payne et al. (2015) demonstrated that 188 (99%) of 189 RV positive cases were Secretors while, 630 (77%) of the 818 control participants were Secretors (P< 0.001). Multivariable analysis showed 98% protection (95% CI, 84%–100%) against RV infection amongst non–Secretor *FUT2* null individuals. Epidemiologic association between RV in cases and control may be attributed to RV strain and HBGA diversity (Payne et al, 2015).

Nordgren et al. (2014) found an association between Le/Se phenotype/genotype and susceptibility to RV infection in a study based in Burkina Faso and Nicaragua. In that study, 100% of 27 children from Burkina Faso infected with RV P[8] were Se+ Le B+, whereas only 22% of 27 children infected with P[6] were Se+ Le B+ (P< 0.001). RV P[8] strain (n= 22) also infected only Le+/Se+ children from Nicaragua, thus confirming the Burkina Faso results. The observation that Le– phenotype is very common among African populations may provide an explanation for the prevalence of P[6] strains in the African continent (Nordgren et al., 2014).

Another recent study based on a Southern China cohort involving cases (n= 266) and controls (n= 151) reported an association between Se phenotypes and the risk of RV infection (Zhang et al., 2016). In this study, P[8] (n= 70) and P[4] (n= 5) strains were predominant amongst individuals with a confirmed RV infection and all 75 RV positive individuals were Le+ (P< 0.004). These findings confirm the fact that the Le and Se statuses are crucial host susceptibility factors to P[8] and P[4] RV strains. Van Trang et al. (2014) also reported an association between RV infection and Le–Se phenotypes in a study based on a Vietnamese cohort. In that study, 74 of the 85 children who tested positive for RV infection were all Se+ (P< 0.001). Of the 116 individuals, only 31 tested negative for RV infection and were all Le– and Se–. RV strains detected in this cohort were as follows: P[8] (n= 74), P[6] (n= 6), and P[4] (n= 5) (Van Trang et al., 2014).

However, Ayouni et al. (2015) showed that RV P[8] strain infected both Se (n= 28) and non–Secretor (n= 4) children (P= 0.017) in Tunisia. The RV strains detected were as follows; P[8] (n= 30) and P[4] (n= 2). Overall, these findings suggest that RV infection may be associated with Se or Le status of individuals and that other health factors might be playing a role in RV

susceptibility. However, this need to be further investigated because the population size of that study restrict tangible conclusions (Ayouni et al., 2015).

Contrastingly, RV P[6] was reported to infect children who were both Le+ and Le– and P[6] was more likely to infect those that were Le– (Nordgren et al., 2014). These results were expected since P[6] is thought to recognizes H type 1 (Secretor) antigen which is Le A– and Le B– (Jiang et al., 2017).

Se phenotype has also shown to be associated with development of RV–specific immunoglobulins against P[8] and P[4] RV strains after infection. Interestingly, Zhang et al. (2016) measured the concentration of RV–specific IgG and found that IgG levels were higher for Le+ and Se+ phenotypes vs. non–Secretor phenotypes ($P < 0.0001$). These findings confirm the association of RV susceptibility and Le– Se statuses in Chinese populations. Günaydin et al. (2016) also showed in Swedish individuals that Secretors had increased ($P < 0.01$) RV–specific serum immunoglobulins, and increased susceptibility to RV infections vs. non–Secretors. In that study, 783 IgA–deficient patients and 1009 healthy blood donors were recruited. Overall, these results suggest that *FUT2* and *FUT3* genes are putative susceptibility determinants for RV infection.

### 1.6 Host immune response to rotavirus infection

The innate immune system can be described as a set of non–specific pathogen sensors that utilizes cellular pattern recognition receptors (PRRs) to sense and detect and respond to pathogen–associated molecular patterns (PAMPs) (Morelli et al., 2015). PRRs are cellular pathways that are particularly involved in sensing of foreign RNA PAMPs. RNA–sensing PRRs include endosomal toll–like receptors (TLRs) such as TLR3, TLR7, TLR8, and cytosolic retinoic acid–inducible gene I (RIG–I)–like receptors (RLRs) such as RIG-I, MDA5, NLRP3, and NOD2 (**Figure 1.5**). These receptors are important for recognition of RNA virus infections and initiation of protective innate immune responses. TLR3 detects cells infect by the virus, whereas TLR7 (and TLR8 in humans) detect RNA of the virus absorbed by the endosomes of sentinel cells (i.e., cell–extrinsic recognition). Lastly, RIG–I and NLRP3 detect virus in the cytosol of infected cells (i.e., cell–intrinsic recognition) (Chen et al., 2017).

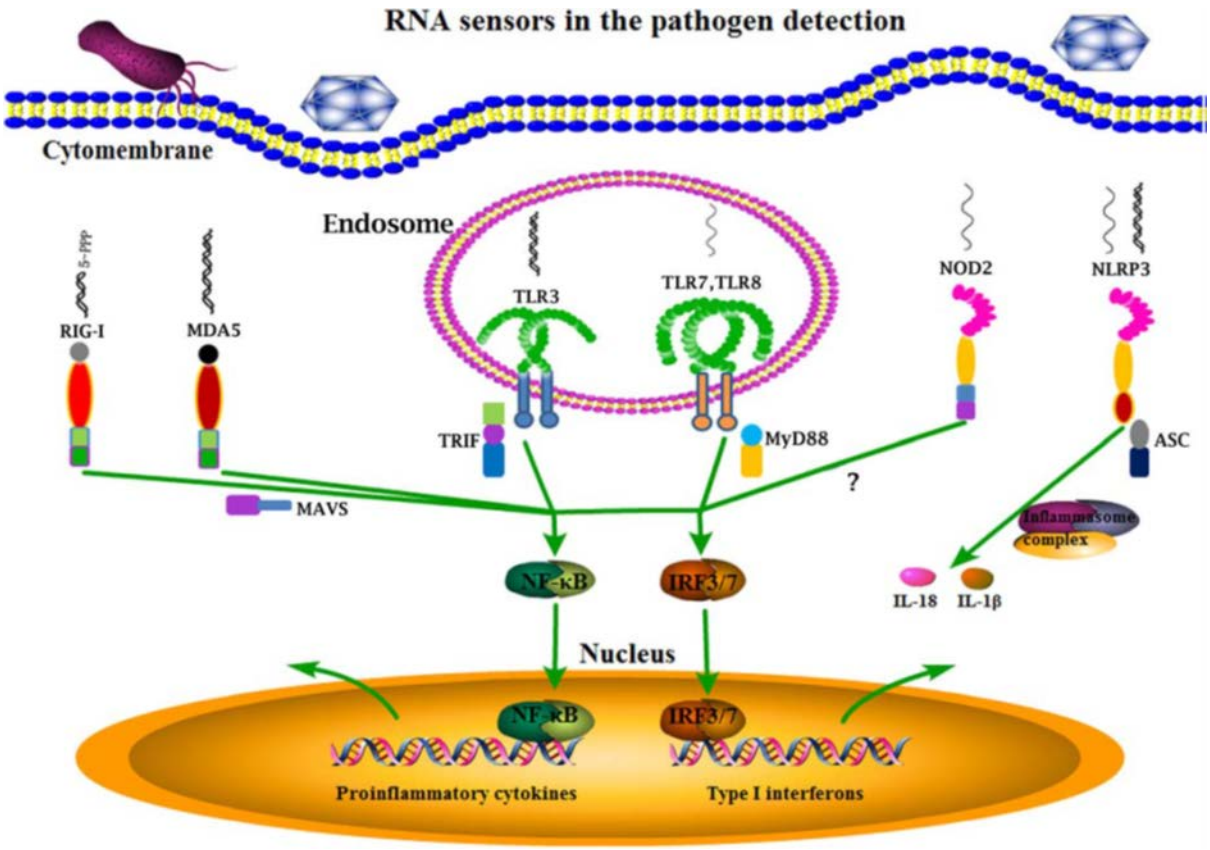



**Figure 1.5** RNA sensors involved in the detection of pathogens including RNA viruses and some bacteria. From Chen et al. (2017).

RV infection induces different signalling pathways which lead to an innate immune response, post infection. The detection of RV strains and subsequent activation of the innate immune responses are shown in **Figure 1.5** above. TLR2, TLR3, TLR4, TLR7, and TLR8 expression are induced by RV infection (Xu et al., 2006; Xu et al., 2009) and these influence gut inflammation, interferon, and cytokine production during RV infection (Xu et al., 2006). TLR signalling consists of at least two distinct pathways: (**1**) MyD88–independent pathway that produces inflammatory cytokines, and (**2**) MyD88–dependent pathway which is associated with stimulation of IFNs as well as maturation of dendritic cells (DCs).

As shown in **Figure 1.5**, the MyD88–independent pathway involves detection of RV dsRNA inside the endosome by TLR3 which, via TRIF, activates the nuclear factor–κB (NF–κB) signalling pathway. This subsequently induce the expression of interferons (IFNs) which are key antiviral cytokines in innate immune response. However, age dependent TLR3 expression in the intestinal epithelium can contribute to RV susceptibility (Pott et al., 2012). Null expression of TLR3 in the neonate gut suggests that it is not involved in protection from RV infection during infancy. Moreover, even adult mice expressing TLR3 remain vulnerable to

RV infection, suggesting that TLR3–mediated signalling may not be enough to infer protect against this RV. This suggests that additional means of innate immunity to RV exist (Uchiyama, et al., 2014).

Apart from TLR3, viral single–stranded (ss) RNA and short viral dsRNA can also be detected by TLR7 and TLR8 inside the endosome (Payne et al., 2014). Unlike TLR3, TLR7 and TLR8 signal via the MyD88 signalling pathway (**Figure 1.5**). MyD88 is an adaptor protein that mediates a major portion of the signalling by all TLRs, except for TLR3, as well as all signalling by inflammasome cytokine receptors (IL–1b and IL–18). While no studies have directly examined the recognition of RV by TLR7 or TLR8, MyD88 signalling mediated by TLR was essential for control of RV infection in a mouse model, driving both innate and adaptive immune responses (Uchiyama, et al., 2014). Genetic variation in both the TLR3 TRIF–mediated signalling pathway, and the TLR7/8 MyD88–mediated pathway, might influence innate immune responses to RV infection.

Other proteins that detect viral RNA include RLRs, such as RIG–I (encoded by *DDX58* gene) and melanoma differentiation–associated gene 5 (MDA5) (encoded by *IFIH1* gene). RIG–I detects ss and ds viral RNA, whilst MDA5 only detects ds viral RNA (**Figure 1.5**) (Jensen and Thomsen, 2012). Both proteins contain common DECH–box helicase domains which are crucial for ATP hydrolysis and RNA binding. They also contain a C–terminal domain (CTD) which is essential for RNA recognition and autoregulation (only in RIG–I). These two RLRs express different affinities to RNA and thus, respond distinctly to viruses. RNA binding triggers oligomerization of RIG–I or MDA5, subsequently inducing oligomerization of adaptor mitochondrial antiviral signalling protein (MAVS). This interaction is facilitated by the protein's mutual caspase activation and recruitment domains (CARD) (Bruns et al, 2014). After polymerization, MAVS activates a signalling cascade that activates interferon–regulating factor (IRF)3 and IRF7 and nuclear factor–kappa B (NF–κB). In turn, these transcription factors activate antiviral type I and III IFNs and other chemokines and cytokines which subsequently regulate cell survival (Besch et al, 2009). Genetic variation in the RLR pathway may influence innate immune signalling and responses to RV infection.

As demonstrated in **Figure 1.6** below, the outcomes of RV infection stimulating TLR and RLR signalling include production of IFN α, β, and γ (Jensen and Thomsen, 2012). IFNs subsequently activate the signal transducer and activator of transcription (STAT)1/STAT2 signalling complex which induce the expression of IFN–stimulated genes (ISGs) in neaby

neighbouring cells, leading to an antiviral condition. Subsequently, ISG products blocks RV replication and maturation thereby containing the virus (**Figure 1.6**) (Holloway and Couson 2013).

RV and the induced IFN response activate adaptive immune response by activating DCs (Holloway and Coulson, 2013). DCs can swiftly produce high levels of IFNs and pro–inflammatory cytokines when encountering viruses. With their ability to present antigens and activate B and T cells, DCs can connect the innate and adaptive immune responses. Upon RV infection, DCs promote the expression of gut homing receptors on T and B cells (Soloff and Barratt–Boyes, 2010). These homing cells are pivotal in resolution of murine RV infection and defence against future infections. During adaptive immune response, RV–specific antibodies including immunoglobulin (Ig) A, IgG, IgM, as well as neutralising antibodies (nAb) are produced. IgA is the main Ig class which providing viral protection in mucosal surfaces such as the gut which is targeted by RVs. Several studies have implicated RV–specific serum IgA levels as correlate of protection against RV infection in humans (Holloway and Coulson, 2013) (see **Section 1.8**).

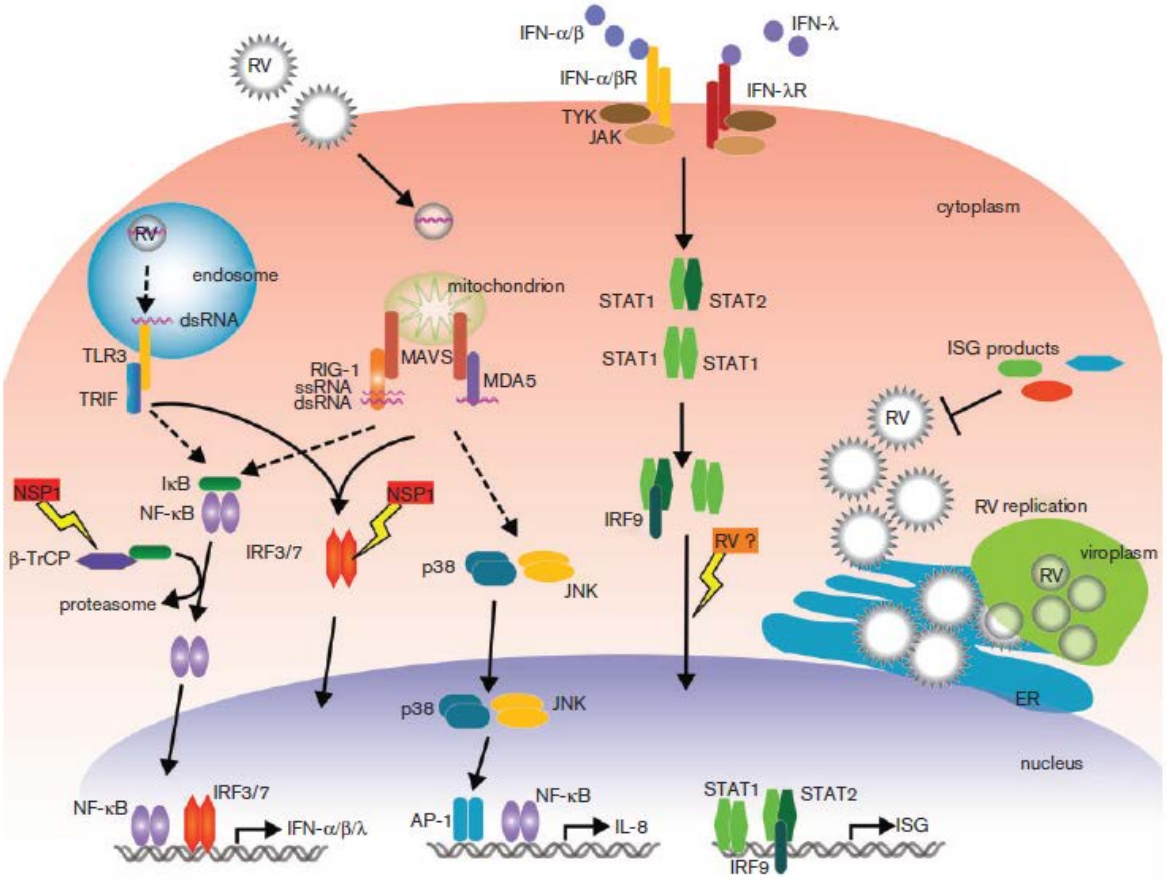



**Figure 1.6** Innate immune signalling pathways involved in the detection of RV and subsequent recruitment of innate immune responses. From Holloway and Coulson (2013). **STAT = Signal Transducer and Activator of Transcription proteins, IRF = IFN response factors, JNK = c–Jun N–terminal kinase, ISG = IFN stimulated gene, JAK = Janus kinas, TRIF = TIR–domain–containing adaptor–inducing IFN–β, MAVS = mitochondrial antiviral signalling protein, MDA5 = melanoma differentiation–associated protein 5, RIG–1 = retinoic acid–inducible gene–I, NF–κB = Nuclear factor– κB , AP1 = activator protein–1, IF = Interleukin, β–TrCP = β transducin repeat–containing protein (β–TrCP), NSP1 = non–structural protein 1. TYK =*

RVs can evade innate immunity using various mechanisms (Randall and Goodbourn, 2008). For example, NSP1 (putative E3 ubiquitin ligase) carried out the degradation of cellular factors involved in both IFN induction and downstream signalling (Morelli et al., 2015). For example, RVs digest β–transducin repeat–containing protein (β–TrCP) through RV NSP1 causing IFN α/β/γ repression (**Figure 1.6**) (Graff et al., 2009). In another mechanism, RV NSP1 can directly digest IFN response factor (IRF) 3/7 complex also leading to the repression of IFN α/β/γ (**Figure 1.6**). Holloway et al. (2009) hypothesized that that RV strains can inhibit the nuclear accumulation of STAT1/2/IRF9 complex (**Figure 1.6**).

### 1.7 Vaccines against rotavirus infection

High mortality rates due to RV infection among children around the world, especially in Africa, led to formulation of different vaccines as a way of providing protection against this virus. Since these are live–attenuated oral vaccines, they mimic the immune response to natural infection without inducing significant side effects. Thus, oral RV vaccines provide protection against subsequent future infections.

The first licenced RV vaccine was called RotaShield (Wyeth Lederle Vaccines). It was made in the United States and was licensed in 1998. Licensure was granted after successful clinical trials in the US, Venezuela, and Finland. RotaShield showed vaccine efficacy of 82–91% against severe RV–induced AGE. However, this vaccine was later withdrawn from the US market later in the same year due to a significant intussusception rate as its side effect. This occurred at a rate of one case/10 000 vaccinated children after only the first dose (Kramarz et al., 2001; Murphy et al., 2001; Peter et al., 2002). Intussusception is defined as bowel obstruction characterized by invagination of intestines into each other, subsequently leading to ischemia and edema (Yen et al., 2011).

Removal of RotaShield lead to development of two alternative live oral RV vaccines namely, Rotarix (RV1) (GlaxoSmithKline Biologicals, Rixensart ) and Rotateq (RV5) (Merck Vaccines Whitehouse, Pennsylvania) (Yen et al., 2011). Both vaccines were introduced into the immunization programmes of various countries around the world as recommended by WHO (WHO, 2009). RV1 is a monovalent vaccine which is made up of human G1P[8] RV strain. Early in infancy, two doses of RV1 are orally administered (first dose at 14 weeks, second dose at 14 weeks of age). Pre–licensure RV1 clinical trials were performed in high– and middle–income settings including Latin America, Finland, Europe, Asia (Hong Kong, Singapore, Taiwan, Japan), and Africa (SA and Malawi) (Yen et al., 2011). Data collected from these trials showed that high– and middle–income countries had high RV1 vaccine efficacy (at an average of 90%) compared to Africa with lower RV1 vaccine efficacy (at 59%). In contrast to RV1, RV5 is a pentavalent human–bovine reassortant vaccine which is made up of five reassortant viruses. These are four human G–type protein (G1–G4) with bovine P[5], and bovine G[6] with human P–type protein (P[8]) (Vesikari et al., 2006). RV5 is also administered orally early in infancy (first dose administered from 14 weeks of age). Similar to RV1, RV5 pre–licensure clinical trials were also carried out and data showed that US and Finland (high and middle

income) had higher RV5 efficacy compared to Africa (Kenya, Ghana, Mali) and Asia (Vietnam and Bangladesh) which are middle–low and low–income continents (**Table 1.7**).

**Table 1.7** Pre–licensure clinical trials efficacy data of RV1 and RV5 collected from high and middle income, middle–low– and low–income countries. Adapted from Yen et al. (2011).

| **Location** | **Vaccine** | **Vaccine Efficacy (95% CI)** | **Rotavirus Gastroenteritis Severity** |
|---|---|---|---|
| **High and middle income** | | | |
| Latin America and Finland | RV1 | 85% (72–92) | Severe (Vesikari score ≥11) |
| Europe | RV1 | 87% (80–92) | Any severity |
| Asia (Hong Kong, Singapore, Taiwan) | RV1 | 96% (85–100) | Severe (Vesikari score ≥11) |
| Japan | RV1 | 92% (62–99) | Severe (Vesikari score ≥11) |
| United States and Finland | RV5 | 98% (88–100)<br>74% (67–80) | Severe G1–G4 (Clark score >16)<br>Any severity (G1–G4) |
| **Middle–low and low income** | | | |
| Africa (South Africa and Malawi) | RV1 | 59% (36–74) | Severe (Vesikari score ≥11) |
| Africa (Kenya, Ghana, Mali) | RV5 | 64% (40–79) | Severe (Vesikari score ≥11) |
| Asia (Vietnam and Bangladesh) | RV5 | 51% (13–73) | Severe (Vesikari score ≥11) |

Despite low efficacies demonstrated by these vaccines, the number of RV hospitalisations and deaths have significantly been reduced (Burnett et al., 2017; Jonesteller et al., 2017). In African trials (for example), two or three doses of RV1 reduced AGE by at least 77% and 45% during first year of inception in SA and Malawi, respectively (Madhi et al., 2010). RV5 was also shown to reduce severe AGE cases by at least 64% in Kenya, Ghana, and Mali (Armah et al., 2010). In a separate study, severe AGE cases were reduced by 51% in Bangladesh and Vietnam (Zaman et al., 2010). Both RV1 and RV5 did not show any significant increased risk of intussusception during pre–licensure trials (Ruiz–Palacios et al., 2006; Vesikari et al., 2006). Despite these significant reductions in disease burden, there is still a need for further investigation into the exact cause of the reduced protection in Africa and Asia compared to Europe and America. Few reasons for this observation have been proposed and these include

malnutrition (e.g., zinc, vitamin A and B deficiency), differences in gut microbiota, environmental enteropathy, co–infections, passive transfer of maternal antibodies, immaturity/functional reduction of the infant's immune system, and various genetic factors (Desselberger et al., 2017). Observed associations between HBGAs and RV infection raise the question of whether HBGA and Se/Le phenotypes may influence RV vaccine efficacy.

Development of new RV vaccine candidates utilizing various RV strains, formulations, and routes of administration can potentially improve immunogenicity, efficacy, and safety of RV vaccines. A number of parenterally administered, non–replicating RV vaccines are in various stages of development, including inactivated virus, protein subunits, and virus–like particles which bypass the intestine and can potentially increase efficacy and safety profile (Glass et al., 2018). The most advanced candidate consists of a truncated VP8 subunit protein fused with the P2 epitope from tetanus toxin which is expressed in *Escherichia coli*. A monovalent formulation of the P2–VP8 vaccine was shown to be well–tolerated and immunogenic among toddlers and infants in SA, and results from a trial assessing the trivalent P2–VP8 vaccine are expected soon (Groome et al., 2017).

### 1.8 Correlates of protection against rotavirus infection/ correlate of vaccine efficacy

IgA is the main protective antibody class at mucosal surfaces, where it inhibits viral entry or blocks viral transcription and thus, acts as a first line of defence (Günaydin et al., 2014). In cases where the virus successfully infects the host cell, IgA antibodies neutralizes the virus. Correlation between RV–specific IgA antibody titres in serum and protection against RV infection is not commonly accepted (Angel et al., 2012). However, serum IgA titres are commonly used as correlate of protection against RV infections in children after they have been vaccinated and at the start of natural infections by RV (Patel et al., 2013; Desselberger et al., 2013). This is specifically based on neutralizing capacity of IgA immunoglobulins triggered by various RV antigens. Velazquez et al. (2000) reported that IgA geometric mean titre >1: 800 and an IgG titre >1: 6400 appeared to be significantly associated with protection against RV infection after a second infection in a study based on new–born infants from Mexico City recruited from birth and monitored for two years for signs and symptoms of RV infection. IgA and IgG immunoglobulins protect against RV infections (Velazquez et al., 2000). Therefore, they can be used as indicators of RV immune response after natural infection or vaccination using immunoassays (Dennehy, 2008).

Observations of associations between HBGA and RV infection raise the question of whether Se and/or Le phenotypes may be used as a correlates of RV vaccine efficacy. Several studies have found associations between the presence of Se phenotype and RV infection by P[4] or P[8] strains, but not by P[6] strains in various studies (Sharma et al., 2020). In addition, several studies have shown that non–Secretors (or Le A phenotypes associated with non–Secretion status) experience less seroconversion in response to RV vaccination containing P[8] (Sun et al., 2016; Zhang et al., 2016). Overall, this suggests that RV P[8], either live or in RV vaccines, is better recognised (bound) by hosts with Se+ phenotype (including Lewis–B phenotype), leading to increased infection and increased vaccine–take in these individuals, and decreased infection and vaccine–take in Se– individuals. Africans have a relatively low frequency of Se+ Le B+ (FUT2+/ FUT3+) HBGA phenotypes compared to Caucasian and Central American groups, which lead Nordgren et al. (2014) to deduce that a larger proportion of African infants cannot mount adequate immune response to RV P[8] vaccination. This may provide an explanation as to why most people are resistant to P[8] infection but susceptible to P[6] strain and why oral vaccines with P[8] have lower effectiveness in Africa (Nordgren et al., 2014).

Pollock et al. (2018) reported that non–Secretor HBGA phenotype was associated with a reduction in the risk of RV1 vaccine failure in a case–control study based on Malawian cohort (n= 202) of infants. A study by Armah et al. (2019) reported that seroconversion after two to three doses of RV1 was associated with the Se (41%) status (95% CI, 1.2–8.1; P= 0.016) as compared to non–seroconverters (13%). Neither salivary ABO blood type nor Le status were associated with seroconversion in that study. These findings therefore suggest that RV1 vaccine–take (seroconversion) in infants may be associated with the Se status and not the Le status, however this might vary across populations. That study was based on 166 Ghanaian infants exposed to RV1 vaccine during a trial conducted between 2012 and 2013 (Armah et al., 2019). Lee et al. (2018) reported that RV1 protected both Secretors and non–Secretors, however unvaccinated non–Secretors had a reduced RV–induced diarrhoea and were completely protected from P[4], but not P[8] strain. Furthermore, vaccine efficacy was reduced in non–Secretors vs. Secretors (Lee et al., 2018). That study involved 700 Bangladesh infants who were part of RV1 trial conducted between 2010 and 2014. Lastly, Bucardo and colleagues reported some discrepancies in terms of seroconversion in a study involving 236 Nicaraguan infants exposed to both RV1 (n= 168) and RV5 (n= 68) (Bucardo et al., 2018). In this study, significantly fewer individuals with Le A phenotype (i.e., non–Secretors) (0/14) were seroconverters as compared to Le B phenotype (45/175) and Le– (115/47) individuals (P<

0.01), regardless of the vaccine administered. Furthermore, Se+ blood group B individuals exposed to RV1 showed less seroconversion (5%) as compared to Se+ blood group A (26%) and O (27%) ($P< 0.05$) individuals. These findings suggests that indeed differences in HBGA phenotypes across populations might play a role in terms of RV vaccine–take and efficacy (Bucardo et al., 2018).

ABO phenotypes may also possibly be considered as markers of RV vaccine efficacy. Blood type A (presence of the A antigen) has been shown to be used as a receptor by RV type P[III] (Wolpin et al., 2010). There are different subtypes of the A antigen, including the common A2 variant which is classified as a "weakly expressed" phenotype (Wolpin et al., 2010), but the impact of A subtypes on RV infection or vaccination response has not been examined to date. In a recent study, Kazi et al. (2017) postulated that varying HBGAs may play a crucial role in terms of the discrepancies observed in the level of protection for current RV vaccines based on IgA seroconversion rates amongst individuals in low–income versus high–income settings. This is because in this study, authors found that IgA seroconversion rates varied significantly based on HBGA phenotypes after three doses of RV1 vaccine in a Pakistani based cohort (N=181). Highest seroconversion rate (51%) was observed in individuals who were Secretors (with blood type O) and intermediate rate (30%) was observed in individuals who were Secretors (but with non–blood type O) (Kazi et al., 2017). A study by Bucardo et al. (2018) involving Nicaraguan children exposed to RV1 reported a significantly lower seroconversion rate amongst Secretors with blood type B vs. those with blood type O and A. In the same study, blood type A and O individuals exposed to RV15 vaccine had the highest seroconversion rate compared to blood type B (Bucardo et al., 2018). Armah and colleagues reported low seroconversion rate in individuals with blood type O compared to B blood type in a study involving a Ghanaian infants cohort exposed to RV1, however, the difference was not statistically significant (Armah et al., 2018). A study by Bucardo et al. (2019) involving Nicaraguan infants reported high vaccine shedding (indirect measure of seroconversion) in blood type A individuals exposed to RV1 vaccine, but low shedding in blood type A individuals exposed to RV5 vaccine compared to blood type O and B participants.

Genetic variation in the HBGA pathways may be useful predictors of susceptibility to RV infection and as predictors of RV vaccine efficacy. For example, some variants in the *FUT2/FUT3* genes inactivate these loci, causing the non–Secretor status or Le– phenotypes, respectively. Associations between *FUT2/FUT3* gene polymorphisms and RV1 vaccine efficacy (Patel et al., 2009; Armah et al., 2010; Madhi et al., 2010) have been documented in

the US, Vietnam, Burkina Faso, Nicaragua, Tunisia, and China (Payne et al., 2015; Van Trang et al., 2014; Nordgren et al., 2014; Bucardo et al., 2018; Ayouni et al., 2015; Zhang et al., 2016 and Sun et al., 2016, respectively). However, no studies carried out to investigate genetic variation in *FUT2/FUT3* genes and the risk response to RV infection in SA children.

Genetic variation in the immune responses to RV infection may also be useful predictors of RV vaccine efficacy. We hypothesize that variation in the innate immune system *TLR3, TLR7*, *TLR8, RIG–1* and *MDA5* genes may influence immune response to RV vaccination. TLR genetic variants have been shown to influence efficacy of other vaccines (Broquet et al., 2011; Pott et al., 2012; Uchiyama et al., 2015; Dou et al., 2017) e.g., *TLR7* and *TLR8* and measles vaccine. There are currently no published studies in any ethnicity examining *TLR or RLR* gene polymorphisms with risk of RV infection or with RV vaccine response.

In summary, the current study seeks to determine whether there is an association between polymorphisms in *ABO*, *FUT2/3*, *DDX58*, *IFIH1*, and *TLR7/8/3* genes, as well as ABO, Se and Le phenotypes in saliva, with immune response after exposure to live–attenuated RV1 vaccine in SA children below the age of five.

## 1. 9 Aims and Objectives

### 1.9.1 Aim

To determine whether there is an association between polymorphisms in HBGAs (*ABO, FUT2/FUT3)* and innate immune RNA–sensor genes (*DDX58, IFIH1, TLR3, TLR7, TLR8),* and ABO, Se and Le phenotypes in saliva, with immune response to RV1 vaccination in South African children.

### 1.9.2 Hypothesis

The hypothesis for the current study is that RV P[8] binds to Le B during natural infection also during RV1 vaccination. Furthermore, Le B requires both FUT2 and FUT3 to be synthesised and functional. If either or both FUT2 and FUT3 are inactivated, then this might result in lack of Le B antigen. Subsequently, we predict that:

1. Non–Secretors (Le A) have poor seroconversion.

2. Le– individuals have poor seroconversion. If the Le B antigen is absent, there is nothing for vaccine P[8] to bind to – leading to poor seroconversion i.e. non–Secretors or Le A, cannot bind RV P[8] in vaccine. Therefore, they will have poorer response to the vaccine. Alternatively, all Le– individuals might have poorer response to the RV1 vaccine. This is because P[8] is the main component of RV1 which requires Le+ antigens to replicate, whereas Le– phenotype is very prominent in Africa.

3. Genetic variation in the *ABO* gene might be associated with low seroconversion after exposure to RV1.

4. Genetic variation in the innate immune response genes might be associated with low seroconversion after exposure to RV1.

### 1.9.3 Objectives

- To sequence *FUT2* and *FUT3* genes in a subset of cohort DNA samples to determine common and/or unique SNPs and alleles in the cohort.
- To select and genotype candidate SNPs present in *FUT2* and *FUT3* genes in the whole cohort and relate these to post–vaccine serum IgA levels.
- To determine the ABO, Le, and Se phenotype in a cohort of infants who received two doses of oral RV1 through saliva ELISA phenotyping at an early timepoint when children were being breastfed and assess the association between HBGA phenotypes and post–vaccine serum IgA levels.
- To determine the ABO, Le, and Secretor phenotype in a cohort of infants who received two doses of RV1 through saliva ELISA phenotyping at a later timepoint when children were no longer being breastfed and assess the association between HBGA phenotypes and post–vaccine serum IgA levels.
- To assess the association between *FUT2*/*FUT3* genotypes and age related FUT2/FUT3 phenotypes at two time–points after vaccination.
- To select and genotype SNPs in *TLR3/7/8*, *IFIH1*, and *MDA5* genes in the whole cohort and correlate these with post–vaccine serum IgA levels.
- To select and genotype SNPs in the *ABO* gene in the whole cohort and correlate these with post–vaccine serum IgA levels.
- To validate the designed *ABO* SNP genotyping method in samples with known *ABO* serotypes.

# 2. METHODS AND MATERIALS

## 2.1 Study population and sample usage

The study cohort included saliva and serum samples from 250 Black HIV–negative SA infants who were co–enrolled into two studies previously conducted by the Respiratory and Meningeal Pathogens Research Unit under the supervision of Prof. S.A. Madhi and colleagues (**Figure 2.1**).

The studies are:

> **1. Protocol Title**: Immunogenicity of Pentaxim, Prevnar, Rotarix and other childhood vaccines included in the new immunization schedule for South African infants. Ethics Reference Number: M090824. Serum samples were collected and stored as part of this protocol.
>
> **2. Protocol Title**: Dynamics of pneumococcal colonization in HIV unexposed infants and their mothers and exploring putative salivary and serum correlates of protection against colonization. Protocol Number: M2C08. Ethics Reference Number: M081035. Saliva samples were collected and stored as part of this protocol.

The recruited infants were enrolled from 02 December 2009 to 09 April 2010 during their routine immunization visits at Diepkloof Primary Health Clinic, Soweto, SA. Enrolled infants were five to eight weeks of age with follow up to five years of age. Of the 250 infant samples, 221 infants had serum samples and 168 infants had saliva samples. Infants were vaccinated with RV1 vaccine at six and 14 weeks of age. IgA data were collected at six weeks of age (before vaccine), at six week (after dose 1) and at 14 weeks of age (after dose 2).

A total of 220 infants had pre–vaccine IgA data and 215 infants had IgA data post dose 1 and dose 2 of RV1, respectively (**Figure 2.1**). Informed consent forms were signed by the parent/guardian of the participants in both studies. Both informed consent forms included consent for storage of samples for future testing. This current study was approved by the Human Research Ethics Committee (HREC) (Medical) of the University of the Witwatersrand, SA, ethics reference number: M090824 (**Appendix A**).

Stored saliva samples (n= 215) were used to test for ABO, Le, and Se phenotypes by enzyme–linked immunosorbent assay (ELISA) (**Figure 2.1**). This was carried out at two different time–points:

(**1**) Time–point when the infants were six weeks of age and exclusively breastfeeding.

(**2**) Time–point when the infants were 15–18 weeks of age and no longer breastfeeding or exclusively breastfeeding.

Stored saliva was used for deoxyribonucleic acid (DNA) extraction, PCR, sequencing, and MassArray genotyping experiments.

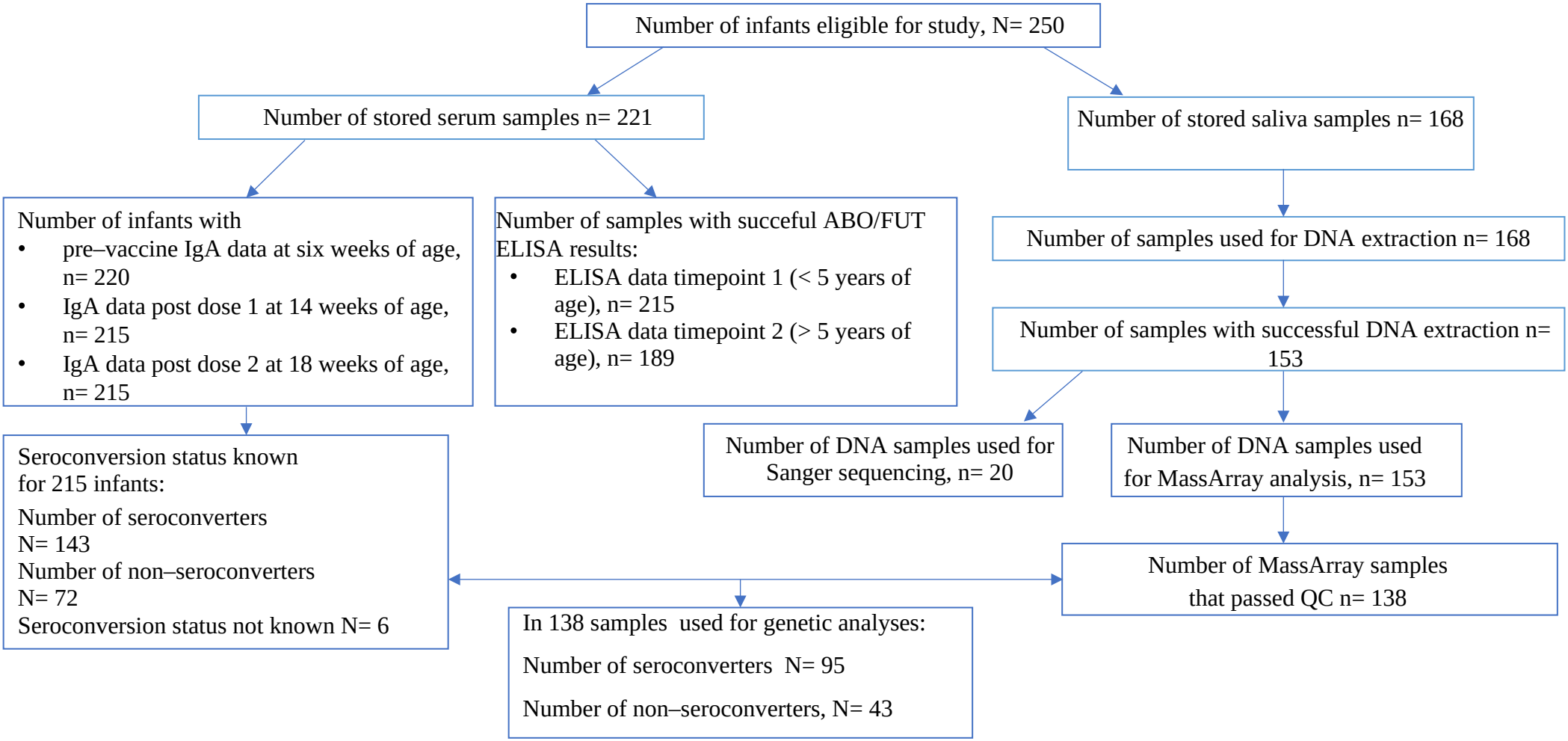



**Figure 2.1** Different materials sample types and sample sizes in the current study.

### 2.2 DNA extraction from serum or saliva

DNA was extracted from frozen stored serum using a Macherey–Nagel Nucleospin kit (Germany) and from frozen stored saliva using Norgen Biotek (Canada) saliva isolation kit (following the manufacturers' instructions). DNA quantity and purity assessment (A260/280 and A260/230) were carried out using NanoDrop® ND–1000 spectrophotometer (Thermofisher, Wilmington DE, USA). Twenty cohort samples with highest DNA concentrations were selected for subsequent Polymerase Chain Reaction (PCR) and Sanger sequencing. All DNA samples were used for MassArray analysis.

### 2.3 Polymerase Chain Reaction (PCR)

Seven PCR reactions were used to specifically amplify *FUT2* and *FUT3* coding regions which contain known SNPs of interest and not all coding regions were amplified in both genes. Thus, the assay was optimized prior to use on the cohort DNA samples. Various sections of the assay were changed accordingly until optimal conditions were obtained. PCR was deemed optimized when clear and distinct single bands were seen after the PCR products were ran on an agarose gel. Complete details of how PCR optimization process was carried out follows in the next section.

#### 2.3.1 PCR Optimization

PCR optimization was carried out using *FUT2* and *FUT3* primers (**Table 2.1**) from Nordgren et al. (2014) and Ferrer et al. (2009) and were synthesized by Inqaba Biotech (Pretoria). The optimization process started with a protocol from NEB OneTag® 2X master mix (NEB, Inc) as recommended by the manufacturer. For primers that were not optimized when the NEB OneTag® protocol was used, an alternative Touch Down PCR protocol from Nordgren et al. (2014) was employed. A total of seven primer pairs were used to cover specific exons containing SNPs of interest in these two candidate genes. Three amplicons were used to cover the coding region of exon 2 of the *FUT2* gene whereas, four amplicons were used to cover exon 3 of the *FUT3* gene, per sample. Thus, a total of seven primer pairs were optimized in this study (**Table 2.1**). Six practice DNA samples from the sensory neuropathy cohort were used during the optimization process of the primers involved.

**Table 2.1** Seven *FUT2* and *FUT3* primer sets sequences and their respective base pair lengths used for PCR optimization. From Nordgren et al. (2014) and Ferrer et al. (2009).

| **Forward** | **Reverse** | **bp length** |
|---|---|---|
| FUT2_CDS_1 ACTTGAGATACATGCCTGTGC | CAGGGGATCCTGCTGGC | 545 |
| FUT2_CDS_2 CTGGCCAAGATGAACGGG | GGTGTCAATGTTCTCCCGAC | 513 |
| FUT2_2009 ACACACCCACACTATGCCTGCAC | ACTTGCAGCCCAACGCATCTT | 530 |
| **Forward** | **Reverse** | **bp length** |
| FUT3_Ex3_1 GGAGCTTTGGTAAGCAGGAG | AAGTTGAACCAGATCCAGCG | 484 |
| FUT3_Ex3_2[1] AGTGGGTCCTCCCGACAGGACACCACTCCC (EL–3s) | GCTGAGTCCGGCTTCCAGTT (III–12as) | 470 |
| FUT3_Ex3_3 TCCGACATCTTCACGCC | GATGAAGGCGTCGGGTG | 387 |
| FUT3_Ex3_4 ATCACCGAGAAGCTGTGGA | AAAGGACTCCAGCAGGTGAG | 389 |

**FUT2_2009 was used to cover CDS_3 region*

PCR preparation process involved a master mix composed of the forward and reverse primers for each gene of interest and NEB OneTag® 2X master mix (NEB, Inc) in a 2 ml Eppendorf tube. The PCR mixture was prepared according to NEB OneTaq® 2X master mix recommendations (**Table 2.2**).

**Table 2.2** PCR master mix components for a 25 µl reaction according to NEB OneTaq® recommendations.

| **Components** | **25 µl reaction** | **Final volume (µl) (x7)** |
|---|---|---|
| OneTaq 2X master mix with standard buffer | 12.5 | 87.5 |
| Forward primer | 0.3 (0.5 µM) | 2.1 |
| Reverse primer | 0.3 (0.5 µM) | 2.1 |
| $dH_2O$ | 6.9 | 48.3 |

**Five µl (20µM) of DNA were added into each PCR tube to make a total of 25 µl reaction. $dH_2O$= distilled water*

The mixture was vortexed and briefly microfuged, followed by distribution into 0.2 ml PCR tubes. Subsequently, 5 µl (20 µM) of various DNA samples were added into each PCR tube.

Touch Down PCR protocol from Nordgren et al. (2014) and standard PCR protocol recommended by NEB OneTaq® were utilized as previously mentioned. **Table 2.3** below indicates calculated annealing temperatures that were used for each targeted amplicon in *FUT2/FUT3* genes.

**Table 2.3** *FUT2* and *FUT3* targeted amplicons and calculated annealing temperatures that were used as references for each of the primer sets during PCR optimization. The NEB online annealing temperature tool was used to carry out these calculations.

| Targeted amplicon | Ta (°C) |
|---|---|
| FUT2_CDS_1 | 53 |
| FUT2_CDS_2 | 54 |
| FUT2_CDS_3/FUT2_2009 | 58 |
| FUT3_Ex3_1 | 53 |
| FUT3_Ex3_2[1] | 58 |
| FUT3_Ex3_3 | 53 |
| FUT3_Ex3_4 | 54 |

**FUT2_CDS_3 amplicon was amplified using Ferrer et al. (2009) primer pair. The rest were amplified using Nordgren et al. (2014) primers.*

NEB OneTaq® and Touch Down PCR protocol from Nordgren et al. (2014) used during optimization of *FUT2/FUT3* primers are shown in **Table 2.4** and **Table 2.5** below, respectively.

**Table 2.4** NEB OneTaq® PCR protocol used during PCR optimization.

| Step | Temperature (°C) | Duration | Cycles |
|---|---|---|---|
| Initial denaturation | 94 | 30 seconds | 1 |
| Denaturation | 94 | 30 seconds | 28 |
| Annealing | $T_{m-5}$ | 30 seconds | |
| Extension | 68 | 1 minute | |
| Final Extension | 68 | 1 minute | 1 |
| Hold | 4 | ∞ | 1 |

**Table 2.5** Touch Down PCR protocol used during optimization. From Nordgren et al. (2014).

| Step | Temperature (ºC) | Duration | Cycles |
|---|---|---|---|
| Initial denaturation | 94 | 5 minutes | 1 |
| Denaturation | 94 | 15 seconds | 25 |
| Annealing | $T_m$ | 30 seconds | |
| Extension | 72 | 1 Minute | |
| Denaturation | 94 | 15 seconds | 20 |
| Annealing | 55 | 30 seconds | |
| Extension | 72 | 1 minute | |
| Final Extension | 68 | 7 minute | 1 |
| Hold | 4 | ∞ | 1 |

#### 2.3.1.1 FUT2_CDS_1 primer set optimization

Initially, gradient PCR was used for optimization of this primer set using the calculated annealing temperature as reference. Calculated gradient PCR ranges during optimization were 50.0 – 55.0°C, 57.0 – 68.0°C, 60 – 68.0°C, 63 – 68°C. A standard PCR protocol from NEB OneTag® at 67°C was used (**Table 2.4**). This was followed by the Touch Down PCR at 65°C (**Table 2.5**). The primers were optimized at an annealing temperature of 67°C using the Touch Down PCR protocol.

#### 2.3.1.2 FUT2_CDS_2 primer set optimization

Optimization process for this primer set started off with Touch Down PCR protocol at 54°C followed by 60°C, 67°C, 50°C. Then standard NEB PCR was employed at 54°C and then at 45°C. A gradient PCR range of 50 – 60°C was carried out, followed by NEB OneTag® protocol at 59°C. The primers were finally optimized at an annealing temperature of 59°C using Touch Down PCR protocol**.**

#### 2.3.1.3 FUT2_CDS_3/FUT2_2009 primer set optimization

For this primer set, gradient PCR range of 50 – 60°C was used. This set was optimized at an annealing temperature of 60°C using the standard NEB OneTag® protocol after only two gels were ran.

#### 2.3.1.4 FUT2_Ex3_1 primer set optimization

Optimization started with a 47 – 55°C gradient PCR range which was followed by 55 – 65 range. NEB OneTag® protocol at 59°C was carried out, followed by 57°C. Finally, the primers were optimized at 59°C annealing temperature using the Touch Down PCR protocol.

#### 2.3.1.5 FUT2_Ex3_2[1] primer set optimization

For this primer set, a gradient PCR range of 55 – 60°C was carried out and this was followed by NEB protocol at 58°C. Finally, the primers were optimized at 56°C annealing temperature.

#### 2.3.1.6 FUT2_CDS_3 primer set optimization

Optimization for this primer set began with a gradient PCR at 50 – 58°C range which was followed by a standard NEB protocol at 55°C. This was followed by two additional NEB OneTag® PCR protocols at 52°C and 50°C, respectively. The primers were finally optimized at 55°C using the Touch Down PCR protocol.

#### 2.3.1.7 FUT2_CDS_4 primer set optimization

A gradient PCR range of 50 – 60°C was used for this primer set, followed by a 60 – 68°C range. The 50 – 60°C range was employed again but this time 5%, 7%, and 10% DMSO was added into the reaction mixture, respectively. However, this primer set was optimized at 58°C annealing temperature using standard NEB OneTag® protocol. Optimal annealing temperatures for the seven primer sets are summarized in **Table 2.6** below.

**Table 2.6** Optimal annealing temperatures and PCR protocols used for *FUT2* and *FUT3*.

| Gene | Primer Set | Annealing Temperature Ranges (°C) | PCR protocol | Optimal Annealing Temperature (°C) |
|---|---|---|---|---|
| ***FUT2*** | FUT2_CDS_1 | 63 – 68 | Touch Down | 67 |
| ***FUT2*** | FUT2_CDS_2 | 50 – 60 | Touch Down | 59 |
| ***FUT2*** | FUT2_CDS_3/FUT2_2009 | 50 – 60 | NEB OneTag® | 60 |
| ***FUT3*** | FUT3_Ex3_1 | 55 – 65 | Touch Down | 59 |
| ***FUT3*** | FUT3_Ex3_2[1] | 55 – 60 | NEB OneTag® | 56 |
| ***FUT3*** | FUT3_Ex3_3 | 50 – 58 | Touch Down | 55 |
| ***FUT3*** | FUT3_Ex3_4 | 50 – 60 | NEB OneTag® | 58 |

PCR was followed by agarose gel electrophoresis to check whether the PCR worked or not and to also confirm that fragment sizes are of the expected sizes. Once all primer sets were optimized, a total of 20 cohort samples with the highest DNA concentrations were amplified using each of the seven optimized primers. This was carried out after DNA extraction from saliva was completed. Thus, a total of 140 PCR amplicons were produced. These amplicons were subsequently sent for Sanger sequencing.

### 2.4 Agarose gel electrophoresis

All PCR fragments were resolved on a 1% agarose gel using agarose gel electrophoresis. To do this, 50 ml of Tris/Borate/EDTA (TBE) buffer plus Seakem LE agarose powder (Lonza) were prepared. Ez–vision® (Amresco) dye or ethidium bromide (Bio–rad) were used as running dye. KAPA loading dye (KAPA Biosystems) was used for loading the DNA samples into the gel wells. NEB DNA ladders (NEB, Inc) were utilized in all gels depending on the expected fragment sizes of each primer set. All PCR products were electrophoresed at 80V for 45 minutes at room temperature (RT). The gels were viewed and photographed using Chemi–doc™ system (Bio–rad). Samples with bright and distinct bands were subsequently sent over to Inqaba Biotech (Pretoria) for Sanger sequencing.

### 2.5 Sanger sequencing of *FUT3* and *FUT2* amplicons

Sanger sequencing was carried out on *FUT*2/*FUT3* amplicons to identify known (published) and novel (new) SNPs in our cohort. The sequencing protocol was as follows: PCR products were purified before another round of PCR (cycle sequencing) was carried out using both the forward and reverse primers. This PCR was performed in a PCR reaction containing four fluorescently labelled dideoxyribonucleotide triphosphates dideoxynucleosides (ddNTPs) and conventional dNTPs. These ddNTPs lack the –OH group and thus acting as chain terminators. As such dNTPs and ddNTPs compete for incorporation into the elongated chain and the DNA becomes labelled fragments with varying lengths. These second amplicons were further purified and denature using formamide. The amplicons containing different labelled bases were analysed using capillary gel electrophoresis on an automated DNA Analyzer instrument – Inqaba routinely utilize Applied Biosystems 3100™ model. Generated results were provided in a form of an electropherogram, with each peak representing a specific base. Furthermore, each base is represented by a specific fluorescent colour. Importantly, Sanger sequencing cannot distinguish between maternal and paternal chromosomes when individuals are heterozygous. Chromas [v.2.6.5] (Technelysium DNA sequencing software, South Brisbane,

Australia) was used to view and analyze the electropherograms once the results became availale. Known and novel SNPs in our cohort were manually identified by printing out the Sanger sequencing alignments and comparing them to the reference sequence (GRCh38.p13) from Ensembl genome browser (http://www.ensembl.org/index.html).

### 2.6 SNP selection

A rigorous literature search was carried out to select SNPs of interest in various candidate genes. The following criteria was used to select and prioritize SNPs of interest:

- First, SNPs of interest were primarily selected and prioritize based on research articles (literature review) associating them with RV susceptibility and/or infection or RV vaccine efficacy on other populations, or immune function.
- Second, SNPs that were reported to have minor allele frequency (MAF) ≥ 5% in the Luhya Webuye in Kenya (LWK) population were selected.
- Third, Linkage disequilibrium (LD) scores ($r^2$) > 0.5 for tag–SNPs were prioritized. LD scores are important because they determine the probability of two SNPs or more being inherited or existing together and thus leading to formation haplotypes.
- Last, SNPs that could be multiplexed for iPLEX® MassARRAY® panel were prioritized in this study.

Complete details of each of the criterion aforementioned follows below.

#### 2.6.1 Literature review of candidate genes and SNPs of interest

Candidate genes were particularly (**1**) those that are associated with RV infection and/or susceptibility and (**2**) trigger innate immune response after RV infection. Thus, SNPs of interest from these candidate genes were those with effects on the expression of the candidate genes and effects on the overall function of the candidate genes. For literature review, the following databases were utilized:

- Pubmed (NCBI) (https://www.ncbi.nlm.nih.gov/pubmed),
- SNPedia database (https://www.snpedia.com/)
- Ensembl genome browser (http://www.ensembl.org/index.html)

Ensembl was particularly used to assess PolyPhen–2 and Sorting Intolerant from Tolerant (SIFT) scores of SNPs of interest. These scores are normally used as indicators of severity of the consequence of changes made to the protein coding sequences of a gene of interest.

Thus, PolyPhen–2 scores range from 0.0 (benign/ tolerated) to 1.0 (deleterious) whilst SIFT scores range from 0.0 (deleterious) to 1.0 (tolerated). PolyPhen–2 scores > 0.5 and SIFT scores < 0.5 were prioritised in this study.

### 2.6.2 Minor allele frequency (MAF)

SNPs with MAF ≥ 5% in the LWK population were used as a proxy for this study since there is no SA MAF data currently available. LWK was particularly chosen as a proxy because of the shared genomic similarities with SA populations as described in the Bantu migration theory (Chami, 1999). As such, the following databases were used to assess SNP frequencies in the LWK population:

- Variation viewer (NCBI) (http://www.ncbi.nlm.nih.gov/variation/view/)
- dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)
- 1000 Genomes (http://browser.1000genomes.org/index.html)
- Entire Genome Interface for Exploring SNPs (ENGINES) (http://spsmart.cesga.es/about.php?dataSet=engines)

Additional MAF data from the Xhosa cohort (SA) based on the Liu et al. (1998) and Pang et al. (1998) research articles were also used during the SNP selection process.

### 2.6.3 Tag–SNPs selection

LD scores ($r^2$) from the 1 000 Genomes Project (via Ensembl) (http://www.ensembl.org/index.html) based on the LWK population were also used to check for any tag–SNPs present in the selected SNPs. LD score ($r^2$) of 10 indicate a strong association between SNPs, suggesting that they are inherited together and thus, form a haplotype. Tag–SNPs were prioritised in this study during the SNP selection process.

### 2.6.4 SNPs selected for iPLEX® MassARRAY®

Once the SNP selection processes explained above were completed, the SNPs were further filtered down during the primer design procedure for iPLEX® MassARRAY®. Frequently some prioritized SNPs are located too close to each other, thus inhibiting optimal primer design. Such SNPs are usually discarded (see **Section 2.7** below for detailed iPLEX® MassARRAY® procedure and SNP filtering process).

## 2.7 SNP genotyping

Sequenom's iPLEX® Gold assay combined with MassARRAY® was also carried out at Inqaba Biotech (Pretoria). This assay was used to genotype the selected candidate SNPs as mentioned above for 141 cohort DNA samples. Sequenom's iPLEX® Gold assay combined with MassARRAY® was selected over other assays because it is cost effective and fast since it allows multiplexing of up to 40 SNPs in one reaction. Genomic DNA samples (20 ng/µl) (25 µl volume of sample) together with reference SNP identity numbers (rs) for all genes of interest were sent to Inqaba Biotech (Pretoria) for genotyping. A list of all SNPs (with associated phenotypes from literature) prioritized in the current study are in **Appendix C**.

However, the list was subsequently trimmed down due to the fact that some of the SNPs were in close proximity which made it difficult to design optimal primers for all the SNPs in the panel, as previously mentioned. A final list of 50 SNPs was created as shown in **Appendix D**. SNP genotyping process was as follows (see **Figure 2.2**): iPLEX® MassARRAY® System Designer software was used to design PCR primer pairs and MassEXTEND® primers for each SNP during multiplexing process. SNP–specific PCR amplicons were generated using the iPLEX® PCR recipe. Cycling conditions used were as follows (iPLEX® Reagents User Guide):

**Table 2.7** Reaction components for PCR of target regions containing the selected SNPs of interest.

| Component | Amount per PCR reaction (µl) |
|---|---|
| NanoPure Water | 0.8 |
| 10× PCR Buffer | 0.5 |
| 25 mM $MgCl_2$ | 0.4 |
| 25 mM dNTP mix | 0.1 |
| 0.5 µM Primer Pool | 1.0 |
| Polymerase, 5 U/µl | 0.2 |
| DNA template (25 ng/µl) | 2.0 |

**Table 2.8** PCR protocol for amplification of target regions using thermo cycler.

| Step | Temperature | Incubation Time | Number of Cycles |
|---|---|---|---|
| 1 | 94°C | 2 minutes | 1 |
| 2 | 94°C | 30 seconds | 46 |
| | 56°C | 30 seconds | |
| | 72°C | 60 seconds | |
| 3 | 72°C | 5 minutes | 1 |
| 4 | 4°C | ∞ | 1 |

Shrimp alkaline phosphatase (SAP) at 37°C for 40 minutes was used to dephosphorylate unincorporated dNTPs. SAP was inactivated by incubating the samples for minutes at 85°C. The samples then underwent single base extension reaction using the single extension primers and mass–modified ddNTPs (refer to the tables below for reaction components and cycling conditions; iPLEX® Reagents User Guide).

Single base extension reaction process involved the use of an extension primer which directly binds next to the SNP of interest. ddNTPs were used during the extension so that the primer is extended by only one nucleotide, complementary to the nucleotide at the SNP site. ddNTPs were also mass–modified for the DNA molecules produced to differ by mass based on the allele (Gabriel et al., 2009; Agena Bioscience, iPLEX® Reagents User Guide). The mass difference allows for differentiation between SNP alleles by data analysis software.

**Table 2.9** iPLEX® gold reaction components for single base extension reaction.

| Component | Amount per PCR reaction (µl) |
|---|---|
| NanoPure Water | 0.6182 |
| 10× iPLEX Buffer Plus | 0.2 |
| iPLEX dNTP Termination mix, 45× | 0.2 |
| Probe Pool (11 µM each) | 0.9409 |
| iPLEX Polymerase, 220× | 0.0409 |
| DNA template (PCR+SAP) | 7.0 |

**Table 2.10** Thermal cycler conditions for single base extension reaction using iPLEX® gold reaction components.

| Step | Temperature | Incubation Time | Number of Cycles |
|---|---|---|---|
| 1 | 94°C | 30 seconds | 1 |
| 2 | 95°C | 5 seconds | 60 |
| | 52°C | 3 seconds | 5 |
| | 80°C | 3 seconds | |
| 3 | 72°C | 3 minutes | 1 |
| 4 | 15°C | HOLD | 1 |

Extension reaction products were diluted with NanoPure water (20 µl) and resin was added into wells to desalt the extension products. The samples were transferred to a SpectroCHIP® array using a Nanodispenser to be crystalized with the MALDI matrix on the chip.

The chip underwent a Matrix–assisted laser desorption/ionization–time–of–flight (MALDI–TOF) mass spectrometry during which, a laser was used to produce ionized products. These products accelerate into a tube and get detected. TOF of these ionized products vary and depend on mass of the molecule measured by mass spectrometer. MassARRAY® Analyzer can differentiate between molecules that differ by 16 Da and therefore, mass of PCR products determine which alleles are present. The arrays were processed in approximately 45 to 60 minutes. The results were added into a database which allows for easy data analysis (iPLEX® Reagents User Guide).

The MassARRAY® Typer software created analysis reports which identify SNP alleles (homozygous/ heterozygous) in each sample. SNP genotypes were then exported into a worksheet. This assay is automated and easy to use and therefore, allows for rapid turn–around time and produces highly precise quantitative results (Sequenom, 2009).

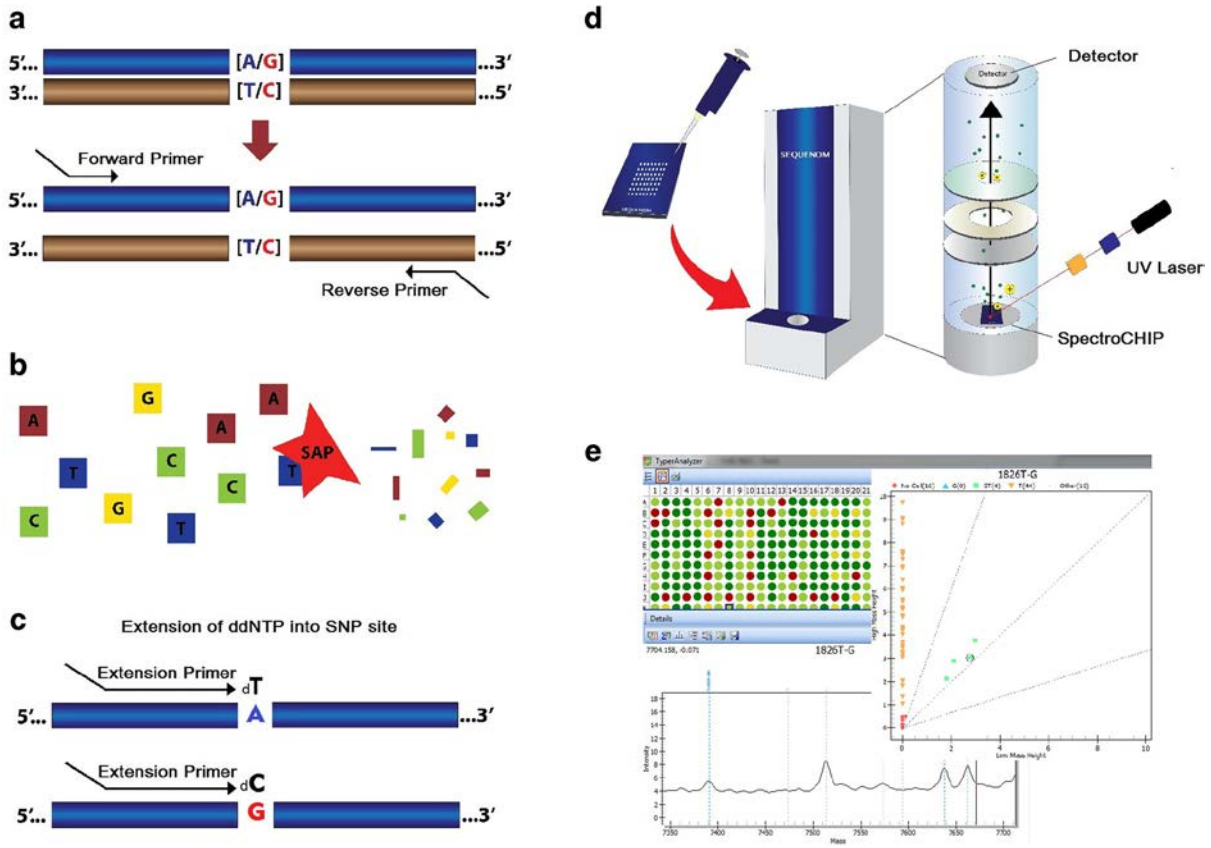



**Figure 2.2** A schematic representation of SNP genotyping using iPLEX® MassARRAY® technology. (**a**) Amplification of specific target region containing the SNP of interest. (**b**) Washing with SAP enzyme to remove any unbound dNTPs. (**c**) SNP–specific primer extension reaction step whereby the target DNA and the primer are incubated with mass–modified dNTP terminators. (**d**) Products of the extension reaction are spotted onto a spectroCHIP which is then read using MALDI–TOF mass spectrophotometry method. Sample molecules are then transferred into a TOF mass spectrophotometry to detect primer extension mass. (**e**) Finally, a Sequenom software automatically interprets the detected primer mass into specific SNP genotypes for each DNA sample. From Svidnicki et al. (2015).

### 2.8 Saliva phenotyping assay

To analyze the ABO, Lewis and Secretor phenotypes in the children's saliva samples, the ELISA assay, as described by Nordgren et al. (2014), was modified and it was carried out at two time–points i.e., (**1**) at six to eight weeks of age, when the infants were breastfeeding and (**2**) at 15–18 weeks of age, when the infants were no longer breastfeeding or exclusively breastfeeding. Two time–points were carried out because there is a possibility that maternally acquired IgA antibodies might have contaminated the first time–point infant saliva samples collected when the infants were still breastfeeding, which could give false results. In addition,

HBGAs are age–regulated and therefore, they change as the infants grow (Magwira et al., 2020). Therefore, two time–points were carried out to check whether difference in age of the infants has any significant effect on the expression of these HBGAs and ultimately their association with seroconversion. Time–point 1 was when the infants were six weeks of age and exclusively breastfeeding, whereas time–point 2 was when the infants were 15–18 weeks of age and no longer breastfeeding or exclusively breastfeeding.

This ELISA assay was adapted and optimized by Dr. M. Groome and Prof. J. Dorfman at RMPRU, Soweto, SA. All ELISA reactions included positive and negative control samples with known ABO/Le and Se statuses. These controls were SDAL (Le B+, Se+) and QCTE (Le A+, Se–).

ELISA maps used for all Le/ABO and UEA–I lectin reactions can be found in **Appendix B**.

**(1) First time–point ELISA:**

Optimized Le/ABO ELISA assay was carried out on 215 cohort saliva samples as follows: ELISA plates (NUNC 96F Maxisorp; Thermo Fisher Scientific, Roskilde Denmark) were coated with saliva (22 samples plus two controls per day in duplicates) diluted 1:500 in coating buffer (0.1 M carbonate–bicarbonate buffer, pH 9.6). This was followed by incubation at 37°C for two hours and then overnight at 4°C. Next day, plates were washed four times with 200 µl of washing buffer (0.9% phosphate–buffered saline (PBS), 0.05% Tween 20 in PBS) per well. Washing was followed by one and half hours of incubation at 37°C with α–A (ABO1) (Diagast, France), α–B (ABO2) (Diagast, France), α–$Le^a$ (Seraclone®, LEA1) (LEA2 (IgM) (Bio–Rad, Germany), and α–$Le^b$ (DiaClon, Monoclonal IgM antibody) (Bio–Rad, Cressier FR, Switzerland). These antibodies were individually diluted to 1:50000 with a dilution buffer (0.05% Tween 20 in PBS with 5% fetal bovine serum (FBS) (Invitrogen) (100 µl per well). The plates were once again washed four times with wash buffer (200 µl in each well). This was followed by incubation with goat anti–mouse IgG (heavy plus light chain) horseradish peroxidase conjugate (Invitrogen, USA) diluted to 1:1000 in a dilution buffer (0.05% Tween 20 in PBS with 5% FBS) (100µl per well) for one and half hours at 37°C. The plates were once again washed four times with wash buffer (200 µl per well for each wash; last wash was PBS only). The reaction was developed using 100 µl of O–Phenylenediamine (OPD) (Sigma® Life Science) for 15 minutes and stopped by adding 100 µl of Hydrochloric acid (HCl) before the plates were read. The plates were read at 492 nm using Multiskan plate reader (Labsystems Multiskan RC) and 690 nm was used as the reference wavelength.

Detection of Fucα1–2Gal (Se phenotype) using *Ulex europaeus* agglutinin (UEA–I) lectin ELISA assay was carried out as follows: The ELISA plates (NUNC 96F Maxisorp; Thermo Fisher Scientific) were once again coated with saliva diluted to 1:500 with a coating buffer (0.1 M carbonate–bicarbonate buffer, pH 9.5–9.6) (100 µl per well). This was followed by incubation at 37°C for two hours, followed by 4°C overnight. The following day, plates were washed with wash buffer (200 µl per well for each wash). Then HRP–conjugated UEA–I (Sigma® Life Science, USA) (starting conc. 1 mg/ml) diluted to 1:1600 in 0.05% Tween 20 in PBS plus 3% bovine serum albumin (BSA) was added into the plates (100 µl per well). Subsequently the plates were incubated at 37°C for one and half hours and washed four times with wash buffer (200 µl per well for each wash; last wash was PBS only). The reaction was once again developed using 100 µl of OPD (Sigma® Life Science) for 15 minutes and stopped using 100 µl of HCl before the plates were read. The plates were once again read at 492 nm using Multiskan plate reader (Labsystems Multiskan RC) and 690 nm was used as the reference wavelength.

OPD substrate for both ABO/Le detection was prepared as follows: 20 mg of OPD (Sigma® Life Science) tablet was dissolved in 50 ml of buffer (12.8 ml of 0.2M $Na_2HPO_4$, 12.2 ml of 0.1M Citric acid, and 25 ml $dH_2O$). Then 50 µl of 30% Hydrogen peroxide was added into the final mixture.

First time–point ELISA cohort samples and control samples were processed in duplicates and then an average was calculated as a QC measure.

The Se– or Se+ phenotype cut–off values were determined by multiplying the negative controls 3X for both time–points, samples with OD values > 3 but less or equals to 15 were declared weak Secretors. ABO phenotype cut–off value was determined by multiplying the negative controls 3X for both time–points.

The following commands were applied to determine the Le phenotypes in our cohort:

- **Le negative** if Le A OD < 2x neg Le A control AND Le B OD < neg Le B control (Participants were declared Lewis negative if their Le A OD values were two times less than negative Le A control, Le B, and negative Le B control OD values)
- **Le A+** if Le A OD > 2x neg LeA control AND ratio Le B OD/LeA OD < 0.75 (Participants were declared Le A+ if their Le A OD values were two times greater than negative and LeA control OD values, and the Le B OD/Le A OD ratio was less than 0.75).

- **Le B+** if LeB OD > neg LeB control AND ratio Le B OD/Le A OD >1.5
  (Participants were declared Lewis B+ if their Le B OD values were greater than negative Le B control OD values and the Le B OD/Le A OD ratio was greater than 1.5).

**(2) Second time–point ELISA:**

The same ELISA protocol used in the first time–point was also applied in the second time–point. However, few changes were made to the previous protocol, namely, 10 µl of the secondary antibody was used instead of 20 µl and BSA was dissolved in PBS only instead of PBS Tween 20 buffer during the detection of UEA–I lectin. In the previous protocol, OPD pills were not incubated at RT before they can be dissolved. This was also changed, and the pills were allowed to reach RT before they were utilized. The ELISA reactions were developed for 30 minutes instead of the previously used 15 minutes. Lastly, a total of 215 samples were processed in the first time–point, however 189 samples with corresponding immune response data were successfully processed in the second–timepoint ELISA.

Second time–point ELISA QC and data analysis were the same as those carried out in the first time–point ELISA.

### 2.9 Rotavirus–specific antibodies – Serology testing

RV–specific IgA titres in our cohort were already determined using ELISA as described by Groome et al. (2014) and Moon et al. (2016). A two–fold serial dilution was made with a total of 11 dilutions used in ELISA testing, from undiluted, 1:20 to 1:20480. IgA titre was reported as reciprocal of highest dilution with positive result (OD 3X larger than the negative control), therefore titres were reported with the following possible values: 1, 20, 40, 80, 160, 320, 640, and 1280. These were not continuous data because the titres were restricted to the eight values mentioned above and thus, these data could not be analyzed as true quantitative data.

Seroconverters (cases) and non–seroconverters (controls) determined from the IgA titres data were used in the current study to determine association between *FUT2*/*FUT3* and immune response genes (DDX58, IFIH1, TLR3, TLR2, TLR7, and TLR8) genotypes/phenotypes and immune response (seroconversion) after RV1 vaccine dosage. Seroconversion (immune response/vaccine–take) in this study was defined as a four–fold increase in IgA titres two months post–vaccination relative to pre–vaccination IgA titres at six weeks of age.

## 2.10 Statistical analysis

Data analysis were carried out using the following statistical software: gPLINK [v.2.050] (Chang et al., 2015) (http://zzz.bwh.harvard.edu/plink), HaploView [v.4.2] (Cambridge, Massachusetts) (Barret et al., 2005) and GraphPad Prism [v.7.00] (GraphPad Software, San Diego, California) (www.graphpad.com/scientific–software/prism/), and SPSS [v.25] (IBM Corp, 2017).

### 2.10.1 Comparison of seroconverters and non–seroconverters to detect other variables influencing antibody response.

Comparison of seroconverters and non–seroconverters to detect other variables influencing antibody response was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. Infant baseline IgA and IgG titres were used as covariates during multivariable analysis in the current study.

### 2.10.2 Analysis of association between demographic or clinical data and ABO/Se/Le phenotypes

Demographic and clinical data between cases (seroconverters) and controls (non seroconverters) were all categorical data and thus, Chi–square test ($\chi^2$) and Fisher's exact test were used. The association between categorical data (ABO/Se/Le phenotypes, and gender) and seroconversion (vaccine–take) was analyzed using Chi–square test ($\chi^2$) and Fisher's exact test. Differences between categories were declared statistically significant if the level of two–tailed significance was less than 0.05. Categorical data that were found to be significantly associated with seroconversion were used as covariates for multivariable analyses (see **subsection 3.1.6.5** below).

### 2.10.3 SNP data quality control (QC)

Once the raw SNP data was obtained from Inqaba Biotech (Pretoria) after genotyping, the following quality control (QC) steps were carried:

1) The data was checked and collated into one excel spreadsheet.
2) The data was then compared to the Sanger sequencing results of a subset of cohort samples (n= 40) to make sure that data match. Several samples were repeated during genotyping. For samples that were genotyped twice, results were only used if both samples had identical results.

Once the raw SNP data was manually checked for discrepancies and genotyping errors, gPLINK [v.2.050] was used for further QC analyses. To do this, PED and MAP files were created in separate Excel spreadsheets and saved as text files ( i.e., PED.txt and MAP.txt). This was followed by creation of binary filesets (namely, BED, BIM, and FAM). Once binary filesets were created, individuals or loci with too much missing or discrepant data were removed using gPLINK [v.2.050] before the association analyses commenced. To do this, the following parameters were checked: missing data, allele frequencies, and Hardy Weinberg equilibrium (HWE) as part of data QC. Thresholds of these parameters were as follows:

1) **Missing data**:
   Maximum SNP missingness rate (GENO) = 0.2 (This means SNPs that had more than 20% of missing SNP data we excluded from the analysis)
   Maximum individual missingness rate (MIND) = 0.2 (This means individuals in the cohort with more than 20% of genotyping data were excluded from the analysis)
2) **Allele frequencies**:
   MAF = 1.0 (This means all (100%) the MAF were used during the analysis)
   Maximum minor allele frequency (MAX–MAF) = 0.01
3) HWE = 0.01 (This means samples with HWE above 1% were excluded from the analysis)

Once the parameters were checked, ‘validate’ command (with thresholds applied) was used on the data in gPLINK [v.2.050], and then final count of individuals and SNPs after QC was obtained as a text file.

### 2.10.4 Allele, genotype, and haplotype frequency analysis of the cohort

Allele and genotype frequencies from Sequenom iPLEX® MassARRAY® data were determined using gPLINK [v.2.050] and compared to MAF data in LWK and Yoruba in Ibadan (YRI) (Nigeria) populations from the 1000 Genomes Project (http://www.internationalgenome.org/) in Ensembl (https://www.ensembl.org/index.html) because these populations share genetical similarities with SA populations. LD between SNPs of interest and haplotype frequencies were analyzed using Haploview software [v4.2] (Cambridge, Massachusetts) (Barret et al., 2005). Default Haploview threshold parameters were used namely, HWE P value= 0.01, minimum genotype= 75%, maximum number of Mendelian errors= 1, and minimum MAF= 0.001.

Haplotypes per locus per individual were phased in gPLINK [v.2.050]. The generated *FUT2* haplotype phasing data was used to predict Se phenotypes in our cohort by creating a 12 SNP Se haplotype reference/wild type sequence with known phenotype. The generated *FUT3* haplotype phasing data was also used to predict presence/absence of Le antigen in our cohort by creating a 9 SNP Le haplotype reference/wild type sequence with known phenotype. Se phenotype data containing weak Se phenotype was then correlated with Le phenotype data for each participant to determine their complete Le phenotypes. Thus, individuals who are Le+ and Se+ have Le (a–b+) phenotype, whereas weak Secretors have a Le (a–b+) or Le (a+b+) phenotype. On the other hand, those that are Le+ and Se– have Le (a+b–) phenotype. Individuals who are Le– and Se+ have Le (a–b–) phenotype. Individuals who are Le– and Se– have Le (a–b–) phenotype.

The estimated phenotypes based on genetic data were compared to phenotypes measured by ELISA by matching both genetic and phenotypic data for those who have both data in our cohort. Once that was completed, percentage concordance between genetic and phenotypic data was calculated by diving the number of concordant pairs with the total number of pairs.

### 2.10.5 Genotype–phenotype association analysis

gPLINK [v.2.050] was used to investigate the association between polymorphisms in HBGAs genes (*ABO*, *FUT2/FUT3)* and innate immune RNA–sensor genes (*DDX58*, *IFIH1*, *TLR3*, *TLR7*, *TLR8)*, and immune response (seroconversion).

Univariate and multivariable analyses were carried out following a model (**Figure 2.3**) proposed by Henry et al. (2015) which follows guidelines stipulated by gPLINK [v.2.050] for SNP data analyses. All SNPs of interest including their respective alleles, genotypes, and haplotypes were tested for association with ABO, Le, and Se phenotypes using gPLINK [v.2.050]. However, all SNPs were tested in multivariable analysis regardless of their P values. Multivariable analysis considers any demographic or clinical data that is significantly associated with seroconversion phenotype. P values less than 0.05 were considered statistically significant in both univariate and multivariable analyses.

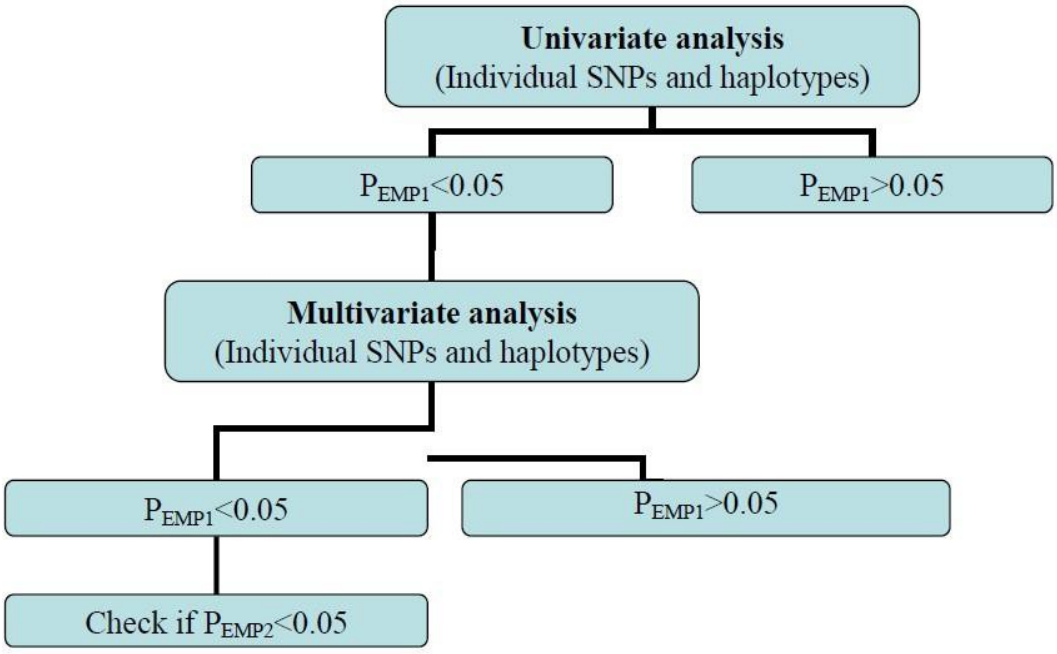



**Figure 2.3** A model for genotype–phenotype association analysis in gPLINK. From Henry et al. (2015).

### 2.10.6 Univariate analysis

Univariate analysis was carried out to test the association between alleles, genotypes, and haplotypes of SNPs of genes of interest and seroconversion (vaccine–take) (phenotype). Since seroconversion is a categorical variable, Chi–square test ($\chi^2$) or Fisher's exact test were used. Univariate analysis used a Chi–square test ($\chi^2$) with one or two degrees of freedom and runs Cohran–Armitage trend test. Fisher's exact test was employed whenever five or less observations were presented in one of the above allelic dominance models because Chi–square test ($\chi^2$) does not provide accurate results in such cases.

In the current study, four models of allelic dominance were used: Allelic, Recessive, Dominance, and Codominant (also called Genotypic). The dominant and recessive models were used to test for minor alleles, i.e., if D is the minor allele and d is the major allele, Purcell et al. (2007) computed the following models:

**Table 2.11** Three different models of allelic dominance. Adapted from Purcell et al. (2007).

| **Allelic** | **D** | **versus** | **d** | | |
|---|---|---|---|---|---|
| Dominant | (DD + Dd) | versus | dd | | |
| Recessive | DD | versus | (Dd + dd) | | |
| Genotypic (codominant) | DD | versus | Dd | versus | dd |

Bonferroni correction was used to adjust significance values to correct for multiple SNP testing. This correction was carried out as follows: The P value was multiplied by the number of tests performed. Bonferroni correction is used to reduce the probability of obtaining a false–positive (i.e., type I errors) when multiple pair–wise tests are carried out on one set of data.

In addition to P values, empirical P values ($P_{EMPs}$) were also generated during permutation testing in gPLINK [v.2.050], which is a Monte Carlo–based approach for testing significance of case–control association studies (http://pngu.mgh.harvard.edu/~purcell/plink/). The generated $P_{EMPs}$ after 100 000 permutations are: (**1**) Empirical P value ($P_{EMP1}$) (which was uncorrected) and (**2**) Max(T) empirical P value ($P_{EMP2}$) (which was corrected for multivariable analysis). $P_{EMP2}$ computes the number of permutations in which any of the test statistics surpassed the observed statistic. Furthermore, it does not assume that the tests are independent, however, it controls the chance of observing at least one false positive per experiment.

Therefore, this makes $P_{EMP2}$ stricter than uncorrected $P_{EMP1}$. P values less than 0.05 were considered statistically significant in both $P_{EMP1}$ and $P_{EMP2}$ analyses (Hider et al., 2008).

### 2.10.7 Multivariable analysis

Multivariable analysis takes covariates into account and it was carried out on all SNPs regardless of whether they were significantly associated with the phenotype of interest or not during univariate analyses. Since seroconversion is a categorical variable, logistic regression was used to investigate the multivariable association between alleles, genotypes, haplotypes of SNPs of genes of interest and seroconversion (phenotype) taking covariates into account using gPLINK [v.2.050].

OR values in this study were calculated with the major allele as the reference allele and thus, OR > 1 (positive) means that minor allele increases risk in relation to the major allele. On the other hand, OR < 1 (negative) means that minor allele decreases risk/ increases the protective effect in relation to the major allele.

### 2.10.8 Epistasis

Epistasis is the interaction between SNPs, whereby the effect of one SNP depends on the presence or the absence of another SNP in one gene or multiple genes. In the current study, the production of the Le A and Le B antigens is a clear example of the results of epistatic interactions between the *FUT2* and *FUT3* genes. Epistasis was tested between 21 *FUT2* and *FUT3* SNPs in gPLINK [v.2.050] which runs all pairwise SNP combinations. This was carried out by creating PED and MAP files carrying genetic data of all 21 SNPs and then allele association command was used. However, "–assoc" command was replaced with "–epistasis –epi11" command. P values < 0.05 were considered significant during epistasis analysis in the current study.

# 3. RESULTS

# 3.1 Secretor and Lewis phenotypes and genotypes, and their association with antibody responses to rotavirus vaccination in a South African cohort

### 3.1.1 Antibody responses to Rotarix vaccination

IgA was used in the current study as a measure of seroconversion (vaccine–take) after exposure of SA infants to RV1 vaccine at six and 14 weeks of age. These IgA data and cohort were generated by Groome et al. (2014) to investigate the effect of breastfeeding on immune response after exoposurexposure of the infants to two doses of RV1. We used a subset of these samples and data for the current study. **Figure 3.1** and **Table 3.1** below shows the data that were generated for the full cohort n= 215 in Groome et al. (2104; Moon et al. 2014). **Table 3.2** shows the same data, but only for the subset of (n= 138) samples used in the current study.

Median and range of IgA titre pre– and post–vaccination of the full cohort (n= 215) are shown in **Figure 3.1** below. IgA median and range before vaccination were 20 and 1–1280, respectively, while a month after the second dose of vaccine IgA median and range had increased significantly to 160 and 1–20480, respectively (**Figure 3.1**).

Seroconversion was defined as a four–fold increase in RV–specific IgA titres post vaccine exposure amongst infants as defined in Groome et al. (2014). Of the 215 infants, 141 (65.6%) were found to be seroconverters and 74 (34.4%) were non–seroconverters (**Table 3.1**).

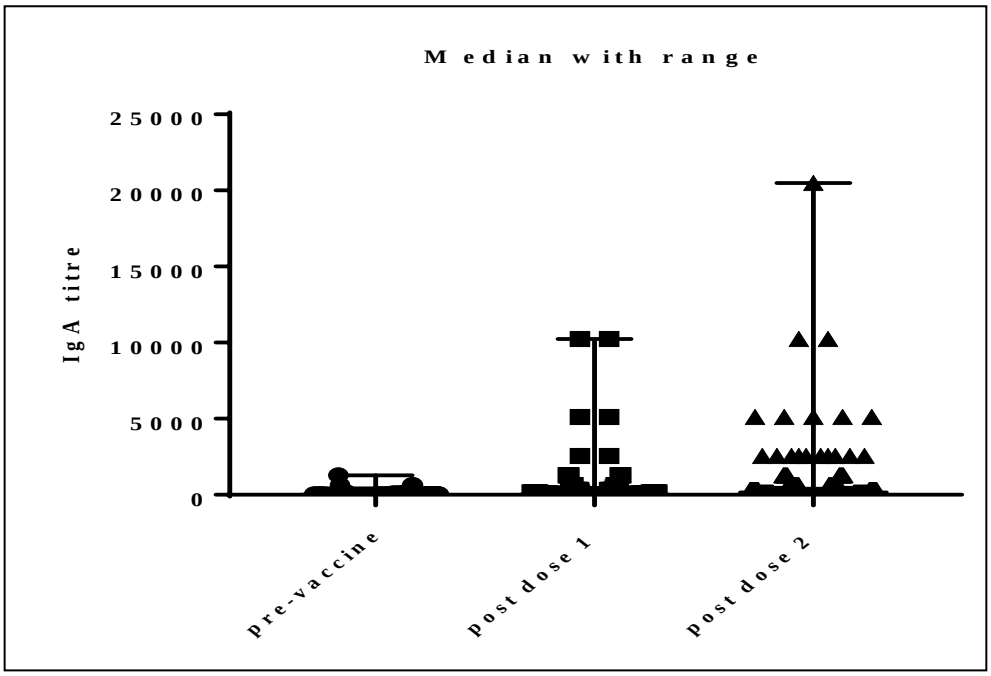



**Figure 3.1** A scatter plot showing IgA median and range at six weeks (pre–vaccine), 14 weeks (post dose 1), and 18 weeks (post dose 2). These data are from Groome et al. (2014), which is study based on the current study cohort.

**Table 3.1** below shows the results of statistical comparison of seroconverters and non–seroconverters amongst the infants in the Groome et al. (2014) cohort (n= 215). There was no significant difference in gender between seroconverters and non–seroconverters in this cohort (Fisher's exact test, P= 1.000) (**Table 3.1**). Thus, gender does not impact seroconversion after exposure to RV1 in this cohort. All children in this cohort were Black South Africans and of the same age at vaccination and at sampling, therefore ethnicity and age were not variables affecting antibody responses in this cohort. However, there was a significant difference between seroconverters and non–seroconverters in terms of the IgA at six weeks (pre–vaccine) (Mann–Whitney test; P= 0.0001). High pre–vaccine IgA titre in non–seroconverters suggests the possibility that trans–placentally acquired maternal IgA affect RV response in infants as discussed by Groome et al. (2014) and Moon et al. (2016).

**Table 3.1** Comparison of seroconverters to non–seroconverters before and after two doses of RV1 vaccine in the current study consisting of 215 infants.

| | **Seroconverters** | **Non–seroconverters** | **P value** | **Total** |
|---|---|---|---|---|
| **N** | 141 | 74 | – | 215 |
| **Number of males** | 77 | 40 | 1.000 [a] | 117 |
| **Number of females** | 64 | 34 | | 98 |
| **Geometric mean IgA at six weeks (pre–vaccine)** | 9.4 | 29 | 0.0001 [b] | – |
| **Geometric mean IgA at six weeks (post first vaccine dose)** | 72.1 | 26.4 | 0.0007 [b] | – |
| **Geometric mean IgA at 18 weeks (post second vaccine dose)** | 291.5 | 29.4 | <0.0001 [b] | – |

[a] *Fisher's exact test*

[b] *Mann–Whitney test*

Statistical comparisons carried out in **Table 3.1** above were repeated, however, this time only samples with MassArray SNP data were analysed (n= 189) (**Table 3.2**). There was no significant difference in gender between seroconverters and non–seroconverters (Fisher's exact test; P= 0.8772). However, we observed a significant difference between seroconverters and non–seroconverters in terms of the IgA at six weeks (pre–vaccine) (Mann–Whitney test; P= 0.0002) (**Table 3.2**). For genetic analyses, results from only 138 individuals were retained after QC processes, and the data shown in **Table 3.2** were re–examined in the 138, with the same results found (results not shown).

**Table 3.2** Comparison of seroconverters to non–seroconverters before and after two doses of RV1 vaccine in a South African cohort of 138 infants with MassArray SNP data.

| | **Seroconverters** | **Non–seroconverters** | **P value** | **Total** |
|---|---|---|---|---|
| **N** | 123 | 66 | – | 189 |
| **Number of males** | 73 | 38 | 0.8772 [a] | 111 |
| **Number of females** | 50 | 28 | | 78 |
| **Geometric mean IgA at six weeks (pre–vaccine)** | 8.5 | 28.4 | 0.0002 [b] | – |
| **Geometric mean IgA at 14 weeks (post first vaccine dose)** | 69.5 | 26.2 | 0.0016 [b] | – |
| **Geometric mean IgA at 18 weeks (post second vaccine dose)** | 289.1 | 29.2 | <0.0001 [b] | – |

[a] *Fisher's exact test*

[b] *Mann–Whitney tes*

### 3.1.2 Secretor and Lewis phenotyping by ELISA in saliva samples

ELISA was carried out to determine the Se and Le phenotypes of the study participants as well as their frequencies at two time–points (i.e., at six weeks and 18 months of age). Two time–points were used because Jiang et al. (2017) previously reported that HGBA (including lectin and Le antigens) biosynthesis is developmentally regulated. This ultimately result in the production of HBGAs which are age specific. Therefore, it was critically important to determine the Se and Le phenotypes of the infants at the two time–points to see if their age difference has any significant effect on the expression of these HBGAs and ultimately their association with seroconversion.

#### 3.1.2.1 Secretor phenotype at first time–point (six weeks of age) vs. second time–point (18 months of age)

Se statuses (Se+/ Se–) of the participants in this SA cohort for the first time–point (n= 215) were determined using ELISA as previously described. There were 93 (43.3%) Se+ individuals, 69 (32%) weak Se+, and 53 (24.7%) Se– individuals (**Table 3.3**).

Se status was measured by ELISA in 189 individuals at 18 months of age at time–point 2. The number of participants for the second time–point was lower (n= 189) compared to the first time–point (n= 215), the reason for the participant dropout was due to the fact that some of the participants did not show up for second round of sample collection. There were 86 Se+ individuals (45.5%), 53 (28%) weak Se+, and 50 (26.5%) Se– individuals for the second time–point Se results (**Table 3.3**).

**Table 3.3** Se phenotype frequency counts for time–point 1 and time–point 2.

| Secretor phenotype | Time–point 1 Count (%) | Time–point 2 Count (%) |
|---|---|---|
| **Se+** | 93 (43.3) | 86 (45.5) |
| **Weak or intermediate Se+** | 69 (32) | 53 (28) |
| **Se–** | 53 (24.7) | 50 (26.5) |
| **Total** | 215 (100) | 189 (100) |

There was no difference in the same individuals from time–point 1 to time–point 2 in terms of the Se phenotype.

### 3.1.2.2 Lewis phenotype at first time–point (six weeks of age) vs. second time–point (18 months of age)

Le phenotypes (Le A– B+, LeA+ B–, and LeA– B–)were determined using ELISA as previously described and statistically compared between seroconverters and non–seroconverters.

At time–point 1, most individuals were Le A–B+ (n= 136), followed by Le A+B– (n= 54), and lastly Le A–B– (n= 25) (**Table 3.4**).

At time point 2, there were Le A–B+ (n= 104), followed by Le A–B– (n= 44), and lastly Le A+B– (n= 38) (**Table 3.4**).

There was no difference in the same individuals from time–point 1 to time–point 2 in terms of the Le phenotype. Differences were only due to the different number of individuals sampled at each time point.

**Table 3.4** Le phenotype frequency counts at time–point 1 and time–point 2.

| **Le status** | **Time–point 1 Count (%)** | **Time–point 2 Count (%)** |
|---|---|---|
| **Le A–B+** | 136 (63.3) | 107 (56.6) |
| **Le A+B–** | 54 (25.1) | 38 (20.1) |
| **Le A–B–** | 25 (11.6) | 44 (23.3) |
| **Total** | 215 (100) | 189 (100) |

### 3.1.3 Associations between Secretor and Lewis phenotypes and IgA response to Rotarix vaccination for the two time–points at six weeks and 18 months of age, respectively

Se status at timepoint 1 (six weeks of age) was statistically compared between seroconverters and non–seroconverters (**Figure 3.2**). The frequency table can be found in **Appendix F**. There was no significant difference between Se phenotypes in seroconverters vs. non–seroconverters in this cohort (Chi–2 test; P= 0.263) (**Figure 3.2**). Percentage of seroconverters that are Se+ was 42.5% (60/141), 35.5% (50/141) for intermediate or weak, and 22% (31/141) for the Se– phenotype.

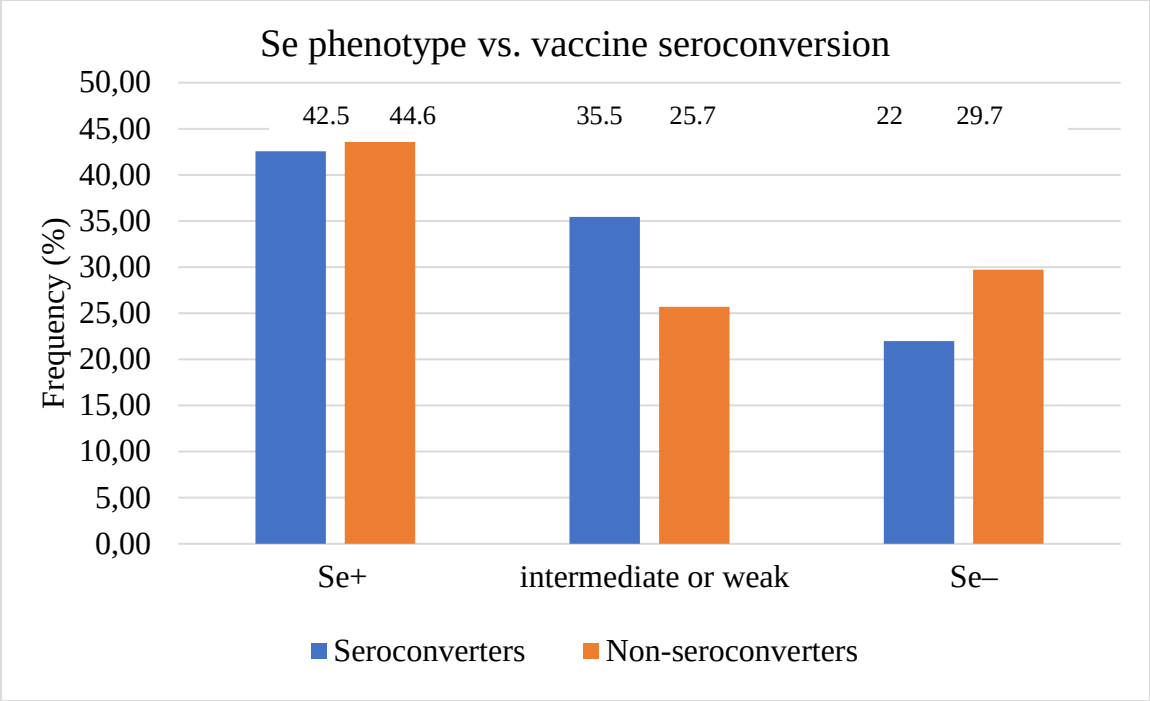



**Figure 3.2** Se phenotype frequencies between seroconverters and non–seroconverters for the first time–point. (Chi–2 test; P= 0.263).

Se status at time–point 2 (18 months of age) was statistically compared between seroconverters and non–seroconverters. There was no significant difference between Se statuses in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.242) (**Figure 3.3**). The frequency table can be found in **Appendix F**. Percentage of seroconverters that are Se+ was 44.7% (60/123), 31.7% (39/123) for intermediate or weak, and 23.6% (29/123) for the Se– phenotype.

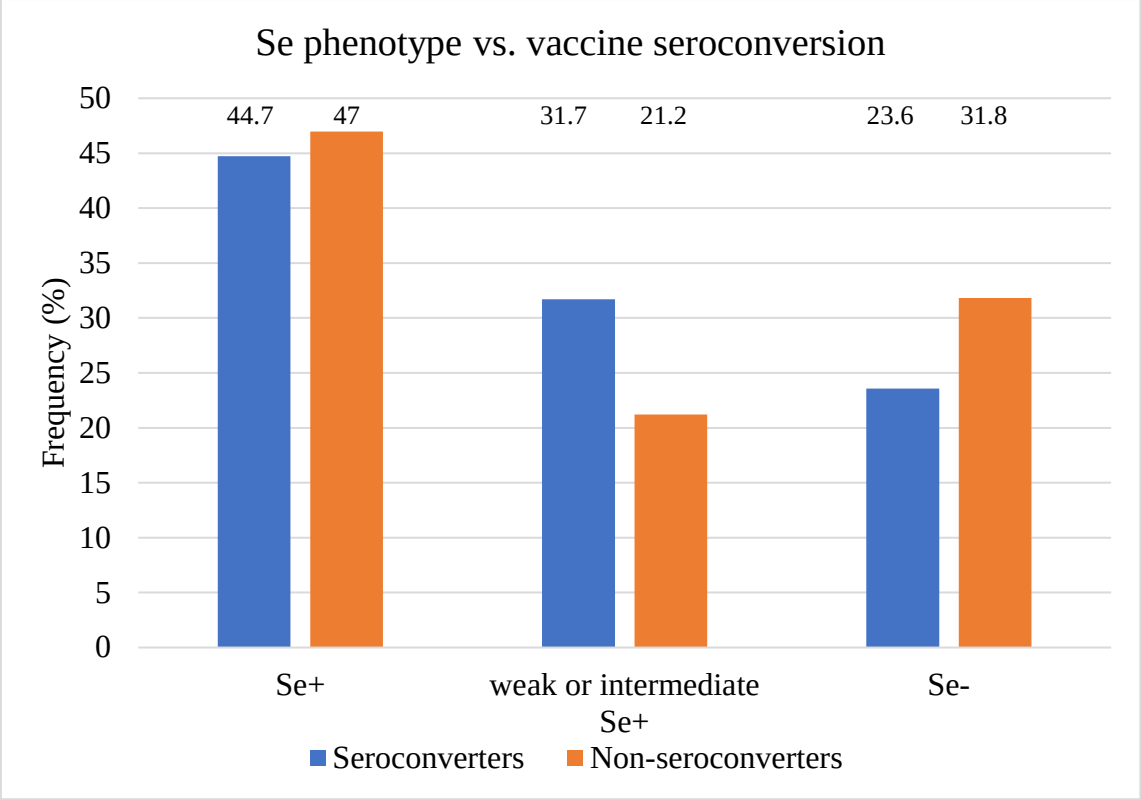



**Figure 3.3** Se phenotype frequencies between seroconverters and non–seroconverters for the second time–point at 18 months of age. There was no statistical difference between Se+, weak or intermediate, and Se– in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.242).

### 3.1.3.1 Analysis of combined Secretor/ Lewis phenotypes vs. seroconversion

There was no significant difference between combined Se/ Le phenotypes in seroconverters vs. non–seroconverters at time–point (six weeks of age) (Chi–2 test; P= 0.683) (**Figure 3.4**). The frequency table can be found in **Appendix F**. Percentage of seroconverters that are Le A–B+ was 65.2% (92/141), 23.4% (33/141) for Le A+B–, and 11.3% (16/141) for the Le A–B– phenotype.

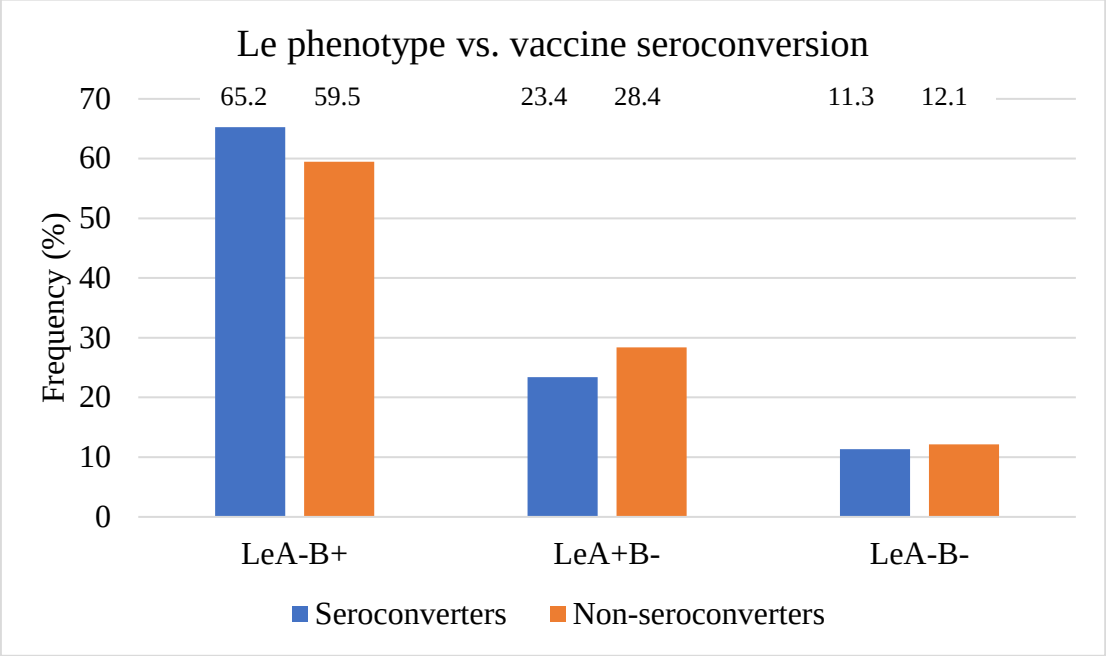



**Figure 3.4** Le phenotype frequencies compared between seroconverters and non–seroconverters. There was no statistical difference between Le phenotype frequencies in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.683).

There was no significant difference between combined Secretor/Lewis phenotypes in seroconverters vs. non–seroconverters at time point 2 / age There was no significant difference between Le phenotypes in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.0895) (**Figure 3.5**). The frequency table can be found in **Appendix F**. Percentage of seroconverters that are Le A–B+ was 59% (73/123), 15% (19/123) for Le A+B–, and 25% (31/123) for the Le A–B– phenotype.

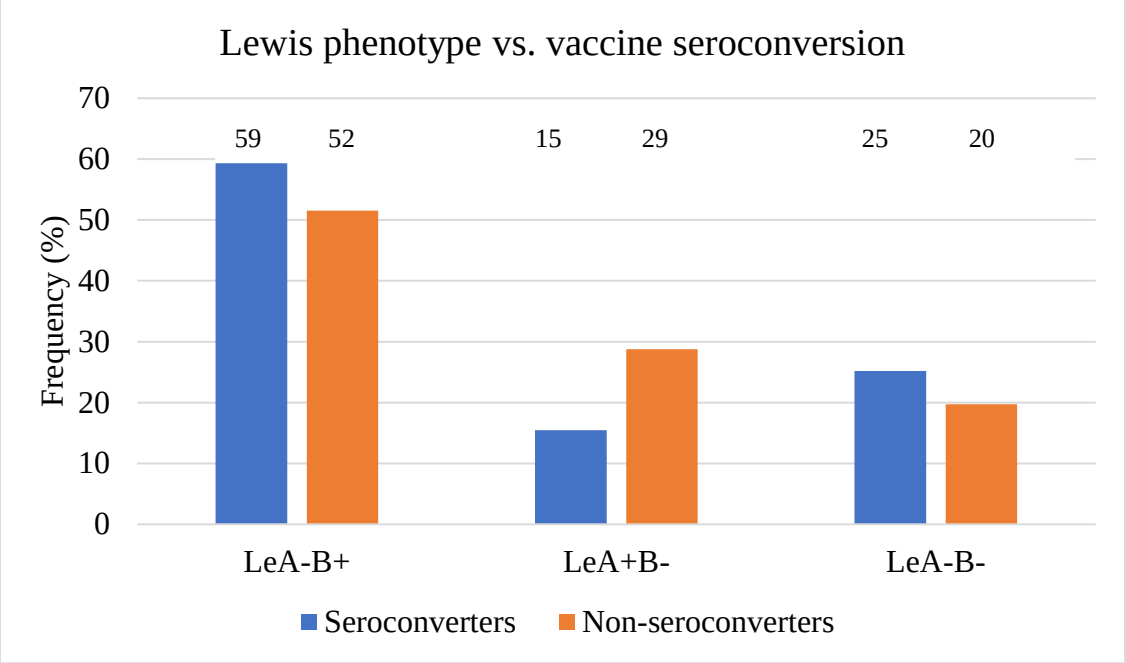



**Figure 3.5** Le phenotype frequencies compared between seroconverters and non–seroconverters. There was no statistical difference between Le phenotype frequencies in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.0895).

### 3.1.4 Results of DNA extraction

DNA extraction was initially attempted from stored frozen serum samples using Macherey–Nagel Nucleospin (Germany) DNA extraction kit, however, DNA yield was extremely low (**Figure 3.6**). DNA was subsequently extracted from 168 cohort saliva samples using Macherey–Nagel kit (Germany) and Norgen Biotek (Canada) saliva isolation kit according to the manufactures' instructions (**Figure 3.6**). Overall, DNA extraction from saliva using either the Norgen Biotek or Macherey–Nagel Nucleospin DNA extraction kit gave better results as compared to blood serum. DNA concentrations from saliva samples varied between 0 – 125 ng/µl, mean DNA concentration was 28 ng/µl and median DNA concentration was 17 ng/µl (**Figure 3.6**).

Optimal DNA quality is when the A260/280 ratio is between 1.8 and 2. Most cohort samples fell within this range. The few samples with A260/280 < 1.8 may have protein contamination as is common in saliva samples, while the few samples with high A260/280 ratio values may reflect DNA degradation, spectrophotometer inaccuracy or RNA contamination.

Twenty samples with the highest DNA concentrations after Nanodrop® spectrophotometry analysis were selected for PCR and Sanger sequencing experiments. The rest of the samples were used at 20 ng/µl for Sequenom iPLEX® MassARRAY® genotyping assay.

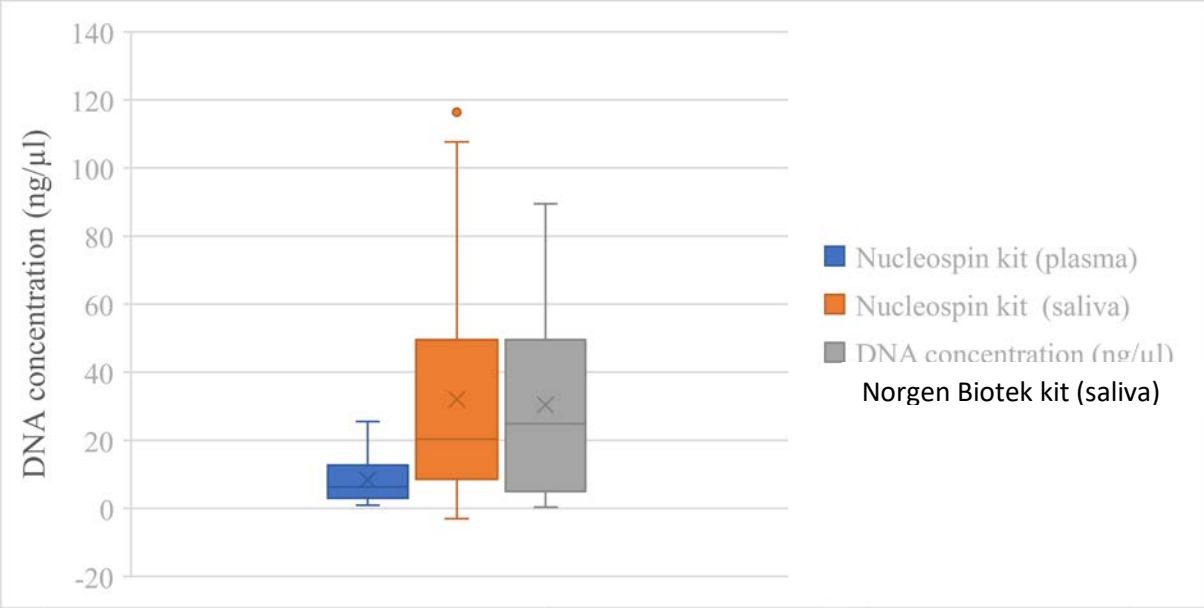



**Figure 3.6** A bar graph showing significant difference in DNA yield from serum samples (n= 40) vs. saliva samples (n= 78). No significant difference in DNA yield from saliva samples

using Macherey–Nagel Nucleospin kit (n= 78) vs. Norgen Biotek DNA extraction kits (n= 40) ($P > 0.05$). Student's t–test was used to calculate the P value.

### 3.1.5 Characterisation of *FUT2* and *FUT3* genetic variation in a SA cohort by Sanger sequencing and by MassArray analysis

#### 3.1.5.1 PCR optimisation

To generate DNA fragments for Sanger sequencing to identify known and novel SNPs from *FUT2* and *FUT3* candidate genes, PCR was carried out using seven primer pairs from Nordgren et al. (2014) and Ferrer et al. (2009). The optimized primer pairs, optimal annealing temperatures, PCR protocls used for each primer pair, and their respective expected band sizes are shown in **Table 3.5** below. The Chemi–doc$^{TM}$ gel image examples of the seven optimized PCR reactions for *FUT2* and *FUT3* genes are in **Appendix E**.

**Table 3.5** Optimal temperatures, PCR protocols used, expected band sizes for the seven *FUT2* and *FUT3* primers used during amplification of cohort DNA samples.

| Primer set | Ta (°C) | PCR protocol | Expected band size (bp) |
|---|---|---|---|
| FUT2_CDS_1 | 67 | Nordgren et al. (2014) Touch Down PCR protocol | 545 |
| FUT2_CDS_2 | 59 | Nordgren et al. (2014) Touch Down PCR protocol | 513 |
| FUT2_CDS_3/FUT2_2009 | 59 | OneTaq® PCR protocol | 1220 |
| FUT3_Ex3_1 | 59 | Nordgren et al. (2014) Touch Down PCR protocol | 484 |
| FUT3_Ex3_2[1] | 56 | OneTaq® PCR protocol | 470 |
| FUT3_Ex3_3 | 55 | Norgren et al. (2014) Touch Down PCR protocol | 387 |
| FUT3_Ex3_4 | 58 | Norgren et al. (2014) Touch Down PCR protocol | 460 |

Overall, all seven *FUT2*/*FUT3* PCR reactions were successfully optimized. Once PCR optimization for the seven primers was completed, 20 of the cohort samples with high DNA concentrations were amplified by PCR with optimized conditions. Subsequently the PCR fragments for all seven primers were sent to Inqaba Biotech (Pretoria) for Sanger sequencing and the results are recorded in the following section.

#### 3.1.5.2 Sequencing of the *FUT2* and *FUT3* genes

Twenty cohort DNA samples with the highest DNA concentrations were sequenced by Sanger sequencing of the second exon in the *FUT2* gene and exon three in the *FUT3* genes using four overlapping amplicons. SNPs in heterozygous and homozygous state were identified.

##### 3.1.5.2.1 *FUT2* sequencing results

Sequence data from four primer sets were analysed. The first primer set worked well whereas the FUT2_CDS_2 primer set did not amplify strongly in our cohort samples and gave disappointing sequencing results. Due to very limited DNA, we could not perform further sequencing in the same samples. The lack of sequence data from *FUT2* using primer set 2 was partially overcome by the use of sequences from the other overlapping primer sets, but many sequence results for the last 500 bp of the region of interest were not successful. For this reason, any allele frequencies arising from these data are from very small sample size and should not be used, rather refer to MassArray data for allele frequencies.

A total of 19 *FUT2* SNPs in 20 cohort samples were identified as shown in **Table 3.6** below. No new SNPs were found in the *FUT2* sequencing results in our cohort. Seven of the 19 SNPs are associated with non–Secretor phenotype: rs1800021 (se40), rs114018037 (se113), rs681343 (se216), rs601338 (se428), rs1800027 (se480), rs602662 (se739), and rs1799761 (se779). SNP rs601388 (se428) is the most common inactivating mutation seen in African populations but was not successfully genotyped in many of our sequences. In this small sample we found the se216 allele to be very frequent (13/40 chromosomes = 0.325) but only found in heterozygote state meaning that these individuals were potentially still Se+. The rs1799761 (779del) frameshift non–Secretor mutation commonly found in Caucasians and also reported in Xhosa by Liu et al. (1998) was found in one individual in the current cohort during sequencing.

Some of these SNPs may only partially ameliorate FUT2 enzymatic activity. SNP rs1800021 (Se40) at allele frequency has been reported to have a weak Secretor activity (47% of normal enzyme activity, Liu et al., 1998) rather than conferring complete non–Secretor phenotype. Similarly, rs1800025 (Se48) has previously been shown to have enzyme activity that is 50% of wild type (Liu et al., 1998). We observed rs1800021 (Se40) at allele frequency of 30%. Furthermore, there were also two other missense mutations which we are not sure if they cause non–Secretor status or not, namely, rs1800025 (Se481) and rs1373056780 (536). The enzyme activity and subsequent Secretor phenotype caused by missense mutation rs1373056780 (536) is unknown.

The remaining SNPs were either intron or 3'UTR variants or were synonymous and were assumed to encode functional Se+ phenotypes. In order to help us to prioritise which of these SNPs to include in the as MassArray panel, the associations of these SNPs in the literature with any functional phenotypes was noted as shown in **Table 3.6**, together with their potential use as tag–SNPs in the proxy LWK population. Interesting, many of the SNPs that were not associated with non–Secretor phenotype were still found to be associated with various phenotypes in the literature, suggesting possible functional roles, for example via TFB sites, miRNAs or by haplotype associations.

**Table 3.6** *FUT2* SNPs found through sequencing of 19 samples in our cohort.

| Sequencing primer | SNP | Alternative SNP name | Major allele | Minor allele | SA MAF (2n)* | variant type | Associations in literature review | SIFT/ Polyphen– | Tag SNP** |
|---|---|---|---|---|---|---|---|---|---|
| 1F/1R | rs679574 | –139 | G | C | 0.318 (12) | intron variant | Asssociated wtih susceptibity to type 1 diabetes (Ram et al., 2016). | – | Yes (19) |
| 1F/1R | rs516316 | –102 | C | G | 0.318 (12) | intron variant | Associated with plasma B12, B6, and folate leveles in an ischemic stroke population (Keene et al., 2014) | – | Yes (1) |
| 1F/1R | rs516246 | –75 | C | T | 0.318 (12) | intron variant | Influences concentration of liver enzymes in plasma (Chambers et al., 2011) | – | Yes (21) |
| 1F/1R | rs370871822 | Se30 | G | A | 0.725 (12) | synonymous variant | – | – | – |
| 1F/1R, 2009F | rs1800021 | se40 | A | G | 0.273 (20) | missense variant (A/V) | Associated with non–Secretor or weak Secretor status (50% activity of wildtype alone and 17%activivin in combinatin with Se481) (Liu 1998) | 0.4 SIFT (Tolerated) | No |
| 1F/1R | rs114018037 | se113 | C | T | 0.045 (4) | missense variant (A/V) | Associated with non–Secretor status (Ferrer et al., 2009) | 0.34 SIFT (Tolerated) | Yes (2) |
| 1F/1R, 2009F | rs492602 | Se171 | A | G | 0.318 (12) | synonymous variant | Associated with serum vitamin B12 concentration, Crohn's and type 1 diabetes (Koda et al., 2001; Velkova et al., 2017; McGovern et al., 2010; Smyth., 2011; Hazra., 2009) | – | Yes (3) |
| 1F/1R, 2009F | rs681343 | se216 | C | T | 0.325 (40) | stop | Confer susceptibity to familial otitis media (Santos–Cortez., 2018) | – | Yes (4) |
| 2009F | rs28362836 | 315 | C | T | 0.045 (2) | synonymous variant | – | – | Yes (4) |

**2n= number of chromosomes successfully sequenced*

***Number of other FUT2 SNPs tagged in proxy LWK on Ensembl*

**Table 3.6** *FUT2* SNPs found through sequencing of 19 samples in the SA cohort. /Continued

| Sequencing primer | SNP | Alternative SNP name | Major allele | Minor allele | SA MAF (2n)* | variant type | Associations in literature revi | SIFT/ Polyphen– | Tag SNP** |
|---|---|---|---|---|---|---|---|---|---|
| 2009F | rs281377 | Se357 | C | T | 0.136 (6) | synonymous variant | Associated with ulcerative colitis (Hu et al., 2016) | – | Yes (5) |
| 2009F/ 2009R | rs601338 | se428 | G | A | 0.091 (10) | Stop gained (W/*) | Associated with non–Secretor status; Associated with sussceptibility to Norwalk, RV, and E.coli, serum vitamin B12 concentration, oral cancer risl (Mottram et al., 2017; Bucardo et al., 2018; Gunaydin et al., 2016; Lindesmith et al., 2003; Kelly et al., 1995; Velkova et al., 2017; Su et al., 2016; Currier et al., 2015; Payne et al., 2015; Nordgren et al., 2014) | – | Yes (5) |
| 2009F | rs180027 | se480 | C | T | 0.045 (2) | missense variant (H/G) | Associated with non–Secretor status (Ferrer et al., 2009) | 0.07 SIFT (deleterious) | Yes (1) |
| 2009R | rs180025 | Se481 | G | A | 0.045 (2) | missense variant (D/N) | Assogiaaed wiah non–Segreaor status 50% activity (Liu et al., 1998) | 0.07 SIFT (tolerated) | Yes (4) |
| 2009R | rs1373056780 | 536 | G | A | 0.05 (2) | missense variant (S/N) | – | 0.16 SIT (tolerated) | – |
| 2009R | rs602662 | se739 | G | A | 0.045 (2) | missense variant | Associated with non–Secretor status. Associated with altered susceptibility to E.coli, serumvitamin B12 concentration, Crohn's disease (Mottram et al., 2017) | | Yes (6) |
| 2009R | rs1799761 | se779 | C | C | 0.045 (2) | frameshift variant (P/X) | – | 0.982 PolyPhen–2 (probably damaging) | Yes (2) |
| 2009R | rs485186 | Se960 | A | G | 0.045 (2) | synonymous variant | Associated with Crohn's disease, type 1 diabetes susceptibility, and resistance to infection (Smyth et al., 2011; McGovern et al., 2010) | – | Yes (7) |
| 2009R | rs485073 | 1009 | A | G | 0.045 (2) | 3 prime UTR variant | Associated with alteration of microRNA seed sites in GWAS (Richardson et al., 2011) | – | Yes (8) |
| 2009R | rs603985 | 1011 | T | C | 0.045 (2) | 3 prime UTR variant | Associated with alteration of microRNA seed sites in GWAS (Richardson et al., 2011) | – | Yes (29) |

#### 3.1.5.2.1.1 Haplotype analysis of *FUT2* sequences

At least six and up to eleven SNPs out of the 19 SNPs noted in the sequencing data always appeared together in the same samples. This implied the existence of a long haplotype across exon 2 of *FUT2* in SA (**Figure 3.7**). This was easier to see in the six SNPs in exon 2 upstream region where most sequences worked, than in the extra four SNPs in the downstream region where Primer 2F/R often failed. This potential eleven–SNP long haplotype inferred in the exon 2 region was formed by the following SNPs:

- rs679574 (c. –139) C–G (intron 1)
- rs516316 (c. –102) G–C (intron 1)
- rs516246 (c. –75) C–T (intron 1)
- rs1800021 (c.73) A–G (missense)
- rs492602 (c.171) A–G (synonymous in exon 2)
- rs681343 (c.216 ) C–T (stop in exon 2)
- rs601338 ( stop)
- rs602662
- rs485186
- rs485073
- rs603985

Furthermore, a shorter length of this haplotype has previously been reported in the literature (Koda et al., 2001).

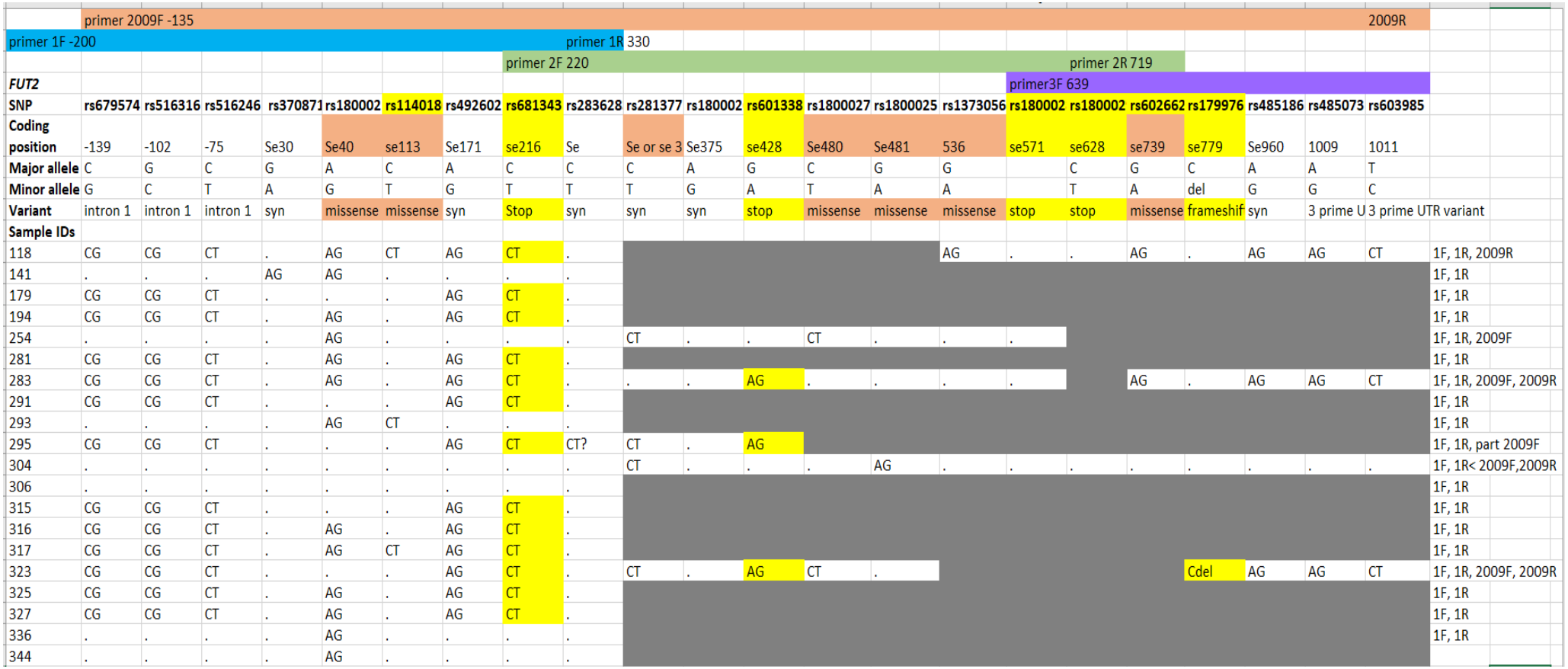



**Figure 3.7** *FUT2* sequence alignment of the current study. Brown colour represents regions amplified by Primer set 2009F/R; Blue colour represents the region amplified by Primer set 1F/R; Green colour represents the region amplified by Primer set 2F/R; Purple colour represents the region amplified by Primer set 3F/R. Yellow highlights represent non–Secretor SNPs while orange highlights represent other missense mutations with functional Secretor status. Grey represent failed sequencing.

#### 3.1.5.2.1.2 Final choice of *FUT2* SNPs to include in MassArray panel

More than 20 SNPs in *FUT2* were initially considered for inclusion in the MassArray panel. These included SNPs from the literature review that were shown to associate with various Se/Le phenotypes and with other clinical or immune–related phenotypes, as well as the 19 SNPs found from our own sequencing data. Our full panel showing all possible SNPs are shown in **Appendix C**.

After SNP prioritisation, a total of 13 SNPs in *FUT2* were chosen for the MassArray panel (**Table 3.7**), including five missense SNPs associated with non–Secretor phenotype: rs114018037 (se113); rs6013388 (se428); rs1800027 (se480); rs1799761 (se778) and rs602662 (se739) with minor allele frequencies of 0.045; 0.091; 0.045; 0.045; 0.045, respectively. Two of the SNPs chosen do not inactivate the FUT2 enzyme but were common in our cohort, namely, rs1800025 (Se481) and rs281377 (Se357). On the other hand, rs370871822 (Se30) SNP has unknown function but was common in our cohort. The remaining SNPs were somewhat redundant but formed an interesting haplotype with rs601388 (se428): rs679574, rs516316, rs516246, rs492602, rs681343, rs601338, rs602662, rs485186, rs485073, rs603985.

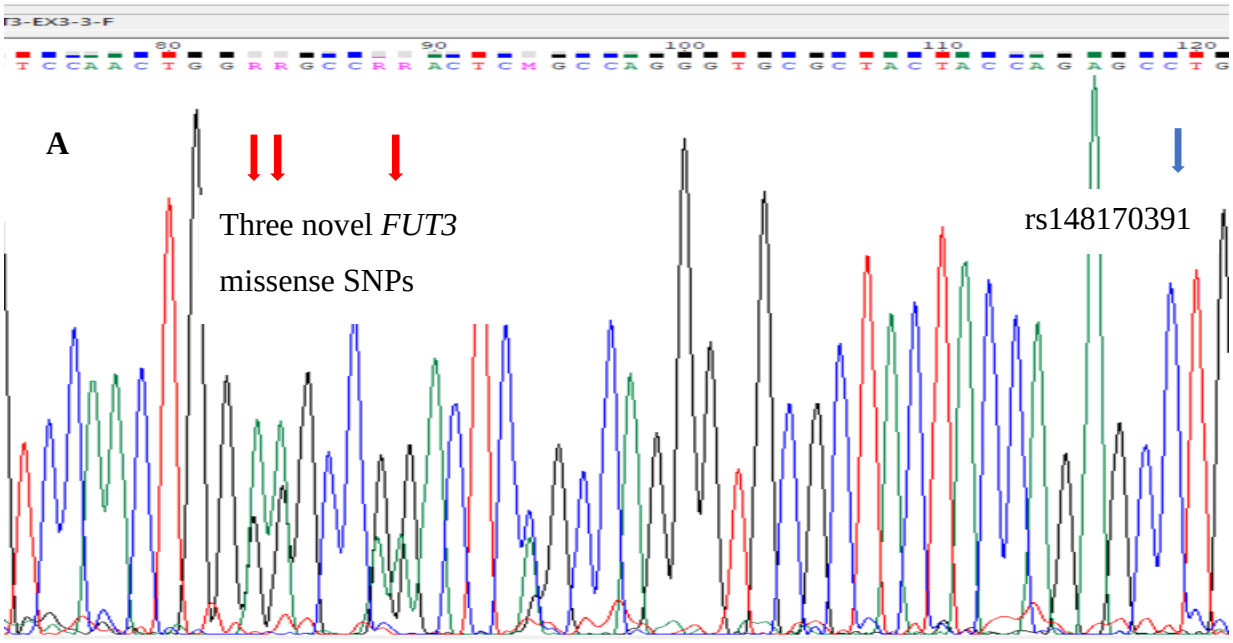
**Table 3.7** List of 13 *FUT2* SNPs selected for genotyping using iPLEX® MassARRAY® for the current study.

| SNP ID | Alternative SNP name |
|---|---|
| rs679574 | –139 |
| rs516316 | –102 |
| rs370871822 | Se30 |
| rs1800021 | Se40 |
| rs114018037 | se113 |
| rs492602 | Se171 |
| rs281377 | Se357 |
| rs601338 | se428 |
| rs1800027 | se480 |
| rs1800025 | Se481 |
| rs602662 | se739 |
| rs1799761 | se778 |
| rs485186 | Se960 |

### 3.1.5.2.2 *FUT3* sequencing results

Sequencing with the first primer set (FUT3_Ex3_1) and the fourth primer set (FUT3_Ex3_4) was extremely successful and worked in both directions in all 20 samples. However, PCR amplification and sequencing with primer sets 2 and 3 (FUT3_Ex3_2[1] and FUT3_Ex3_3) were not very successful and gave us very few useful results. Due to very limited DNA, we could not perform further sequencing in the same samples. We suspect that there may be SNPs present in the primer binding region of primer FUT3_Ex3_2[1]F which have not previously been reported in the literature. These were in a few cases revealed by reverse sequencing extending into that region using the FUT3_Ex3_3R primer (discussed below and indicated by red arrows in **Figure 3.8** and by * in the **Table 3.8**).

A total of 25 SNPs in *FUT3* were identified in the 20 cohort samples as shown in **Table 3.8** below. We noted the presence of 16 missense mutations in our dataset. Missense mutations in *FUT3* are very often associated with Le (A–B–) phenotypes. Some of these missense mutations such as c.508 and c.1067 are found in the catalytic domain of the enzyme and therefore fully inactivate the Le enzyme (Pang et al., 1998), while others such as the c.59 SNP, occur in the trans–membranous domain of this protein, and therefore only reduces FUT3 enzymatic activity rather than eliminating it altogether (Pang et al., 1998). Three of the observed missense SNPs were novel and occurred in region of the FUT3_Ex3_2[1]F primer binding site in heterozygous state in six individuals: c.601 A–G, c. 602 A–G and c.607 G–A (see **Figure 3.8** below).



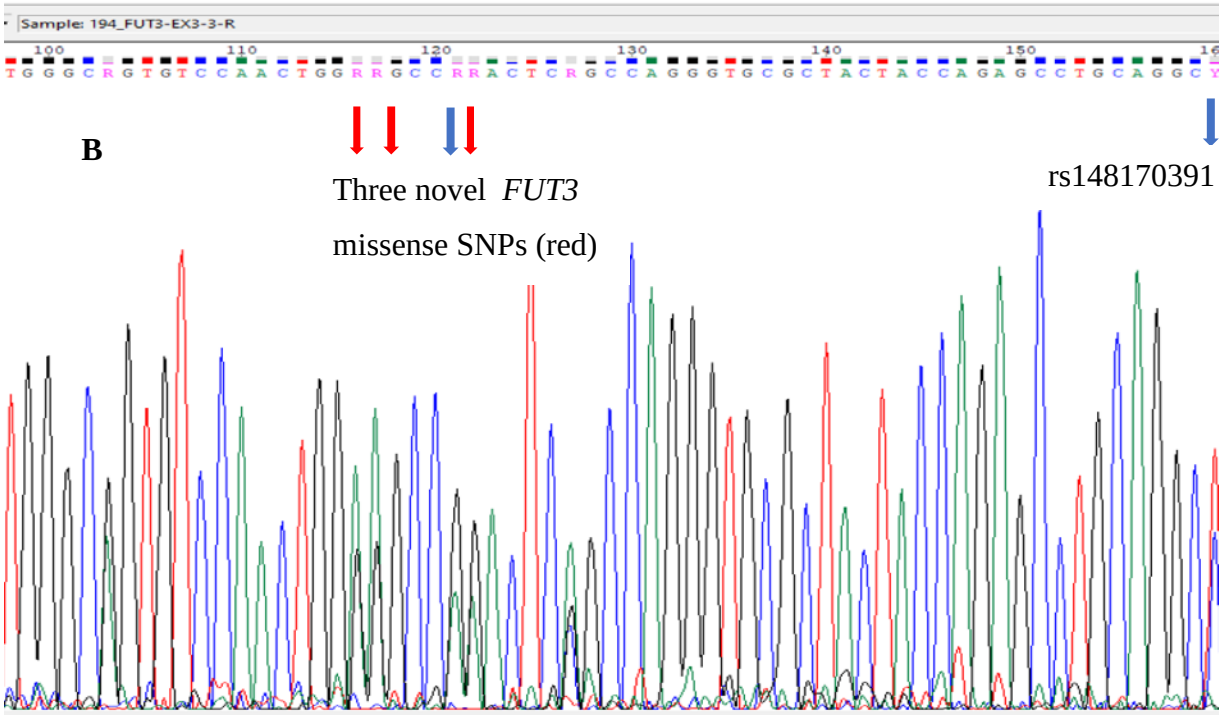



**Figure 3.8** Examples of electropherograms showing the three novel *FUT3* missense SNPs at positions c.601, c602 and c.607 which occurred in the region of the FUT3_Ex3_21F primer binding site in heterozygous state in our cohort (labelled with red arrows). Two unusual mutations (not novel but rare in Africans) are highlighted in blue. (**A**) Forward sequence. (**B**). Reverse complement of the reverse sequence. The blue arrow shows the position of synonymous rs148170391 *FUT3* SNP which is close to the three novel SNPs.

Two further missense mutations (c.304 and c.370) reported in the Xhosa cohort (Pang et al., 1998) at frequencies of 0.02 and 0.005 respectively, were not seen after sequencing of 20 cohort individuals.

For two SNPs in the primer set 1, the minor allele was present in the Ensembl reference sequence: rs812936 (c.202 C present in Ensembl sequence whereas mostly T alleles were found in our 20 individuals ) and rs778986 (c.314 T present in the Ensembl sequence whereas mostly C alleles were found in our 20 individuals). However, Ensembl also confirmed that the 202 T and the 314 C were ancestral alleles, and that the 202 C and 314 T should be considered as minor alleles (Hunt et al., 2018).

In order to help us to prioritise which of these SNPs to include in the MassArray panel, the associations of these SNPs in the literature with any functional phenotypes were noted as shown in **Table 3.8**, together with their potential use as tag–SNPs in the LWK population. Most of the missense SNPs in *FUT3* were associated with differences in Lewis enzyme activity, and few other clinical phenotypes, such as ulcerative colitis and cystic fibrosis. However, many of the synonymous SNPs were not genotyped in 1000 Genomes LWK, neither have they been reported as associated with any particular phenotypes in the literature.

**Table 3.8** *FUT3* SNPs in exon 3 found through sequencing in a Black SA cohort (n= 20).

| Sequence primer | SNP | Alternative SNPs name | Major allele | Minor allele | SA MAF (2n)* | Variant type | Associations in literature review | SIFT/Polyphen–2 | TagSNP** |
|---|---|---|---|---|---|---|---|---|---|
| 1R/1F | rs28362458 | c.13 | G | A | 0.075 (4) | missense (G/S) | Associated with Le status and RV infection (Nordgren et al., 2014). | 0.864 (possibly damaging) PolyPhen–2 | Yes (5) |
| 1R/1F | rs28362459 | c.59 | T | G | 0.225 (10) | missense variant (L/R) | Associated with Le A phenotype, ulcerative colitis, cyctic firbrosis lung disease, and susceptibility to RV (Gunaydin et al., 2016). | 0.02 (deleterious) SIFT | Yes (3) |
| 1R/1F | rs28362460 | c.61 | C | T (or A) | 0.05 (2) | missense variant (L/V) | – | 0 (deleterious) SIFT | Yes (8) |
| 1R/1F | rs28362461 | c.72 | G | C | 0.025 (2) | synonymous variant | – | – | No |
| 1R/1F | rs812936 | c.202 | T (or G) | C | 0.1 (4) | missense variant (R/W) | Associated with Le A phenotype, ulcerative colitis, cyctic firbrosis lung disease, and susceptibility to RV (Gunaydin et al., 2016). | 0.01 (deleterious) SIFT | Yes (1) |
| 1R/1F | rs59796499 | c.304 | C | A (or T) | 0.00 (0) | stop gained (Q/K) | – | 0.66 (tolerated) SIFT | No |
| 1R/1F | rs778986 | c.314 | C | T | 0.1 (4) | missense variant (M/T) | Associated with Le A phenotype, ulcerative colitis, cyctic firbrosis lung disease, and susceptibility to RV (Gunaydin et al., 2016). | 0.01 (deleterious) SIFT | No |
| 1R/1F | rs3745635 | c.508 | G | A | 0.025 (2) | missense variant (G/S) | Associated with ulcerative colitis in Southeast China patients (Hu et al., 2016). | – | Yes (3) |

**2n= number of chromosomes successfully sequenced*

***Number of other FUT2 SNPs tagged in proxy LWK on Ensembl*

**Table 3.8** *FUT3* SNPs in exon 3 found through sequencing of 20 cohort samples /Continued.

| Sequence primer | SNP | Alternative SNPs name | Major allele | Minor allele | SA MAF (2n)* | Variant type | Associations in literature review | SIFT/Polyphen–2 | TagSNP** |
|---|---|---|---|---|---|---|---|---|---|
| 2F/2R | rs28362463 | c.484 | G | A | 0.025 (2) | missense (D/N) | – | 0 (deleterious) SIFT | Yes (18) |
| 3F/3R | rs867576796 | c.589 | G | A | 0.15 (6) | missense variant (V/M) | – | 0 (deleterious) SIFT | – |
| 3F/3R | New R* | – | A | G | 0.15 (6) | missense variant (-/-) | – | – | – |
| 3F/3R | New R* | – | A | G | 0.15 (6) | missense variant (-/-) | – | – | – |
| 3F/3R | rs372597863 | c.606 | G | A | 0.15 (6) | synonymous variant | – | – | – |
| 3F/3R | New R* | – | G | A | 0.15 (6) | missense variant (-/-) | – | – | – |
| 3F/3R | rs28362465 | c.612 | A | C | 0.15 (6) | synonymous variant | – | 0 (deleterious) SIFT | – |
| 3F/3R | rs148170391 | c.645 | T | C | 0.15 (6) | synonymous variant | – | – | – |
| 3F/3R | rs28362466 | c.667 | G | A | 0.025 (2) | missense variant (G/R) | – | 0 (deleterious) SIFT | – |
| 3F/3R | rs750177892 | c.691 | A | C | 0.15 (6) | missense variant (K/E) | – | 0.49 (tolerated) SIFT | – |
| 3F/3R | rs765224654 | c.696 | G | A | 0.15 (6) | synonymous variant | – | – | – |
| 3F/3R | rs565590532 | c.732 | C | T | 0.15 (6) | synonymous variant | – | – | – |
| 4F/4R | rs28381968 | c.808 | G | A | 0.15 (6) | missense variant (V/M) | – | 0 (deleterious) SIFT | Yes (16) |
| 4F/4R | rs761121538 | c.862 | G | A | 0.05 (2) | missense variant (D/Y) | – | 0 (deleterious) SIFT | – |
| 4F/4R | rs3894326 | c.1067 | T | A | 0.05 (2) | missense variant (D/Y) | Associated with ulcerative colitis in Southeast China patients (Hu et al., 2016) | | No |
| 4F/4R | rs150146742 | c,*14 | T | G | 0.05 (2) | 3 prime UTR variant | – | – | Yes (7) |

#### 3.1.5.2.2.1 Haplotype analysis of *FUT3* sequences of the cohort

Haplotypes were inferred from sequence alignments shown in **Figure 3.9** in order to characterise the genetic variation in Black South Africans and also to enable the choice of tag–SNPs for the MassArray panel. A two–SNP haplotype in part of exon 3 was inferred from the *FUT3* primer set 1 sequencing results. Four individuals were heterozygous at both rs812936 (c.202T) and rs778986 (c.314C), suggesting a possible haplotype across these two SNPs (**Figure 3.9**). This haplotype did not extend upstream to rs8362459 (c.59) which had a different minor allele frequency of 0.26.

A long ten–SNP haplotype was inferred in the primer 2/ primer 3 region. In the six individuals where sequencing with primer set 2 and set 3 was successful, all six were heterozygous for a set of ten SNPs in the region c.500 – c.800, including three putative new mutations within the primer set 2 binding site:

- rs867576796 (c.589) G–A (missense)
- New mutation (c.601) A–G (missense)
- New mutation (c.602) A–G (missense)
- rs372597863 (c.606) G–A (synonymous)
- New mutation (c.607) G–A (missense)
- rs28362465 (c.612) A–C (synonymous)
- rs148170391 (c.645) T–C (synonymous)
- rs750177892 (c.691) A–C (missense)
- rs765224654 (c.696) G–A (synonymous)
- rs565590532 (c.732) C–T (synonymous)

This suggested the presence of two different haplotypes in this region (all major alleles linked, and all the minor alleles linked). Interestingly two of the known SNPs (c.589 and c.691) in this proposed ten–SNP haplotype, were missense mutations and the three putative new mutations were all also missense mutations. One of the six samples had a further three heterozygous mutations in this region (rs28362463, c.484; rs3745635, c.508 and rs28362466, c.667) which were also missense mutations that mostly failed typing in the other five samples. This suggested

that ten–SNP haplotype might be even longer, but this remains to be verified. Only one of these six individuals in the current study with the ten–SNP haplotype, also had the primer set 1 haplotype c.202T/c.314 C described above, so it appeared this that ten–SNP haplotype was not linked to the upstream region haplotype.

Genetic variation further downstream, in the region c.800 – c.1100, represented by sequences using the fourth primer set did not appear to be linked in any clear haplotypes.

Alleles with linked mutations at c.484 and 667, as well as between le c.484, 667 and c.808 have previously been reported in a Xhosa cohort (Pang et al., 1998). These haplotypes were also observed in our *FUT3* sequence data (**Figure 3.9**).

| FUT3 ex | primer set 1 | | | | | | up to ~400 | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 145-174 | primer set 2 | | | | | | | | | | | TO 614 | | | | | | | | | | |
| | | | | | | | | | | 488-505 | primer set 3 | | | | | | | | | | | | | | to 873 | | |
| sequenced on coding strand / reverse strand. | | | | | | | | | | | | | | | | | | | | | | from 750 | primer set 4 | | | | to 1200 |
| SNP | rs283624 | rs283624 | rs283624 | rs283624 | rs812936 | rs597964 | rs778986 | rs117540 | rs2836246 | rs374563 | rs8675767 | new | new | rs372597 | new | rs28362465 | rs148170 | rs2836246 | rs7501778 | rs7652246 | rs565590 | rs2838196 | rs13932685 | rs761121538 | rs389432 | rs150146742 | |
| c. position | 13 | 59 | 61 | 72 | 202 | 304 | 314 | 370 | 484 | 508 | 589 | 601 | 602 | 606 | 607 | 612 | 645 | 667 | 691 | 696 | 732 | 808 | 858 | 862 | 1067 | 1099 | |
| reverse / | G | T | C | G | T (or G) | C | C | T | G | G | G | A | A | G | G | A | T | G | A | G | C | G | A | G | T | T | |
| reverse / | A | G | T (or A) | C | C | A (or T) | T | G | A | A | A | G | G | A | A | C | C | A | C | A | T | A | G | A | A | G | |
| protein eff | missense | missense | syn | syn | missense | misssense | missense | missense | missense | missense | missense | missense | missense | syn | missense | syn | syn | missense | missense | syn | syn | missense | syn | missense | missense | 3'UTR | |
| 118 | . | . | . | . | . | . | . | . | . | . | . | . | - | - | - | - | - | - | - | - | - | . | . | . | . | . | |
| 141 | AG | . | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 179 | . | TG | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 194 | . | . | . | . | . | . | . | - | - | - | AG | AG | AG | AG | AG | AC | CT | . | AC | AG | CT | . | . | - | - | - | |
| 254 | . | . | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | . | . | |
| 281 | . | . | . | . | CT | . | CT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 283 | . | TG | CT | . | CT | . | CT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | AG | . | TG | |
| 291 | . | . | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 293 | . | . | . | . | . | . | . | - | - | - | AG | AG | AG | AG | AG | AC | CT | . | AC | AG | CT | . | . | . | . | . | |
| 295 | . | . | . | . | . | . | . | - | - | - | AG | AG | AG | AG | AG | AC | CT | . | AC | AG | CT | . | . | . | . | . | |
| 304 | . | TG | . | . | . | . | . | - | - | - | AG | AG | AG | AG | AG | AC | CT | . | AC | AG | CT | . | . | . | AT | . | |
| 306 | . | TG | . | . | CT | . | CT | - | . | . | AG | AG | AG | AG | AG | AC | CT | . | AC | AG | CT | . | . | . | . | . | |
| 315 | . | TG | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | . | . | |
| 316 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | AG | . | . | |
| 317 | AG | TG | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | AG | . | . | . | . | |
| 323 | AG | TG | . | . | . | . | . | . | AG | AG | AG | AG | AG | AG | AG | AC | CT | AG | AC | AG | CT | AG | . | . | . | . | |
| 325 | . | . | . | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | . | . | |
| 327 | . | GG | . | GC | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | AT | . | |
| 336 | . | TG | CT | . | . | . | . | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | . | TG | |
| 344 | . | . | . | . | CT | . | CT | - | - | - | - | - | - | - | - | - | - | - | - | - | - | . | . | . | . | . | |

**Figure 3.9** *FUT3* sequence alignment of the current study showing different haplotypes formed by common SNPs found in our cohort. Yellow colour represents the region amplified by Primer set 1; Orange colour represents the region amplified by Primer set 2; Green colour represents the region amplified by Primer set 3; Brown colour represents the region amplified by Primer set 4; Blue colour highlights novel SNPs. Grey represent failed sequencing.

#### 3.1.5.2.2.2 Final choice of *FUT3* SNPs to include in MassArray panel

A total of 24 SNPs in *FUT3* were considered for inclusion in the MassArray panel. These included SNPs from the literature review that were shown to associate with various Se/Le phenotypes and with other clinical or immune–related phenotypes, as well as 25 SNPs found from our own sequencing data. Our full possible panel showing all possible SNPs and all of the above criteria are shown in **Appendix C**.

After SNP prioritisation, a total of 11 SNPs in *FUT3* were chosen for the MassArray panel **Table 3.9**. Of these 11 SNPs, nine were missense mutations causing Le negative phenotypes, and two were synonymous mutations. These two synonymous mutations were chosen according to their ability to tag the ten–SNP haplotype described above, given MassArray design constraints on missense SNPs in this region which would have been preferred.

**Table 3.9** List of 11 *FUT3* SNPs selected for genotyping using iPLEX® MassARRAY® for the current study.

| SNP ID | SNP alternative name | SA MAF (this study) | SA MAF (Pang et al.) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | References |
|---|---|---|---|---|---|---|---|---|
| rs28362458 | le13 | 0.08 | 0.10 | 0.268 | Yes (5) | missense | – | – |
| rs812936 | le202 | 0.11 | | 0.071 | Yes (1) | missense | Associated with modification of infection and cystic fibrosis lung disease severity. | Taylor–Cousar, 2009 |
| rs778986 | le314 | 0.11 | 0.90 | 0.066 | No | missense | Associated with modification of infection and cystic fibrosis lung disease severity. | Taylor–Cousar, 2009 |
| rs28362463 | le484 | 0.025 | – | 0.268 | Yes (4) | missense | – | – |
| rs3745635 | le508 | 0.025 | – | 0.338 | Yes (3) | missense | Associated with Le– status, CA19–9 antigen expression, and ulcerative colitis. | Koda et al., 1993; Hu et al., 2016; Wannhoff et al., 2016; Guo et al., 2017; |
| rs372597863 | Le606 | 0.15 | 0.50 | – | – | synonymous | – | – |
| rs2836246 | Le612 | 0.15 | – | | | synonymous | Associated with type 2 diabetes mellitus in Asian patients. | Okamoto et al., 2010 |
| rs28362466 | le667 | 0.025 | – | 0.253 | Yes (17) | missense | – | – |
| rs28381968 | le808 | 0.05 | – | 0.131 | No | missense | – | – |
| rs761121538 | le862 | 0.05 | 0.13 | – | – | missense | – | – |
| rs3894326 | le1067 | 0.05 | – | 0.081 | No | missense | Associated with Le A phenotype and RV infection, cystic fibrosis lung disease, and ulcerative colitis. | Günaydin et al., 2016; Bucardo et al., 2018; Hu et al., 2016; Taylor – Cousar, 2009 |

Although the possible haplotype across SNPs rs812936 (c.202) and rs778986 (c.314) suggested that only one of these two was strictly necessary to be included in the MassArray genotyping panel, however both of these SNPs were included in the final panel.

Regarding the ten–SNP haplotype in the region c.500 – c.800, SNP choice was highly constrained by MassArray multiplex design parameters in this region due to the large amount of genetic variation in this small region. Two synonymous SNPs, rs372597863 (c. 606) and rs28362465 (c.612) were originally chosen to represent this haplotype and this region. However, upon receipt of initial MassArray results, we found that one of these SNPs (rs372597863, c.606) appeared to be monomorphic in MassArray results, and another (rs28362465, c.612) failed genotyping completely. Those two SNPs are very close, and they were amplified by the same initial PCR primer set and extended by two different extension primers. It is not clear if there are errors in the sequencing results (i.e., if these SNPs are not actually common in SA) or if there could be an issue with MassArray primer design since there are so many SNPs in the region. To make sure that we were adequately representing this region of *FUT3* in our MassArray SNP panel, we then chose three further SNPs that were genotyped in a separate (third) MassArray plex. These three SNPs included rs28362466 (c.667 missense), and rs28381968 (c.808 missense) as well as rs3745635 (c.508 missense), the most common Le– causing allele in African populations (Pang et al., 1998) which was previously missing from the panel due to SNP primer design incompatibility. As shown in the subsequent section, we still had less than 100% call rate for these three SNPs, again suggesting the existence of variation in this region influencing primer binding success.

### 3.1.5.3 Characterisation of *FUT2* and *FUT3* genetic variation in our cohort by MassArray analysis

#### 3.1.5.3.1 *FUT2* and *FUT3* allele frequency distribution in the current study cohort compared to other African populations

**Table 3.10** below shows the MAF of *FUT2* and *FUT3* SNPs in the SA cohort compared to LWK and YRI from the 1000 Genomes project. SA cohort data from the current study was compared to that of LWK and YRI because there is no previously published data based on SA cohorts. LWK and YRI share genetical similarities to SA according to the theory of recent common ancestry of Bantu–speaking populations (Chami, 1999), hence we used these two cohorts for comparisons. A total of 12 *FUT2* and nine *FUT3* SNPs were genotyped in 138 individuals in our cohort. *FUT2* rs370871822 SNP was the only rare SNP in our cohort and it was also rare in LWK and YRI populations. Ten SNPs were found to be significantly different between the three groups as marked with (*) (Chi–2 test; $P < 0.05$) (**Table 3.10**). Furthermore, eight SNPs were not significantly different between the three groups (Chi–2 test; $P > 0.05$). Lastly, rs761121538 (*FUT3*), rs370871822 (*FUT2*), and rs370871822 (*FUT2*) SNPs did not have MAF data from the 1000 Genomes project database (**Table 3.10**).

**Table 3.10** Allele frequencies of *FUT2* and *FUT3* in the current study cohort (SA) (n= 138) compared to LWK from Kenya and YRI from Nigeria.

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***FUT3*** | rs3894326 | A | 269 | 182 | 212 | 0,0004* |
| | | T | 5 | 16 | 4 | |
| | | | | | | |
| ***FUT3*** | rs761121538 | – | 249 | – | – | – |
| | | | 7 | – | – | |
| | | | | | | |
| ***FUT3*** | rs28381968 | C | 181 | 172 | 204 | 0,0259* |
| | | T | 17 | 26 | 12 | |
| | | | | | | |
| ***FUT3*** | rs28362466 | C | 237 | 148 | 177 | <0,0001* |
| | | T | 3 | 50 | 39 | |
| | | | | | | |
| ***FUT3*** | rs3745635 | C | 147 | 131 | 159 | 0,2504 |
| | | T | 61 | 67 | 57 | |
| | | | | | | |
| ***FUT3*** | rs28362463 | C | 245 | 145 | 176 | <0,0001* |
| | | T | 27 | 53 | 40 | |
| | | | | | | |
| ***FUT3*** | rs778986 | G | 252 | 185 | 193 | 0,3381 |
| | | A | 24 | 13 | 23 | |
| | | | | | | |
| ***FUT3*** | rs812936 | A | 254 | 184 | 193 | 0,3890 |
| | | G | 22 | 14 | 23 | |

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***FUT3*** | rs28362458 | C | 199 | 145 | 176 | <0,0001* |
| | | T | 9 | 53 | 40 | |
| | | | | | | |
| ***FUT2*** | rs679574 | C | 150 | 111 | 99 | 0,0159* |
| | | G | 106 | 87 | 117 | |
| | | | | | | |
| ***FUT2*** | rs516316 | G | 152 | 111 | 99 | 0,0205* |
| | | C | 110 | 87 | 117 | |
| | | | | | | |
| ***FUT2*** | rs370871822 | – | 269 | – | – | – |
| | | | 1 | | | |
| | | | | | | |
| ***FUT2*** | rs114018037 | C | 263 | 196 | 210 | 0,0664 |
| | | T | 13 | 2 | 6 | |
| | | | | | | |
| ***FUT2*** | rs492602 | A | 150 | 111 | 99 | 0,0403 |
| | | G | 116 | 87 | 117 | |
| | | | | | | |
| ***FUT2*** | rs281377 | C | 229 | 137 | 170 | <0,0001* |
| | | T | 35 | 61 | 46 | |
| | | | | | | |
| ***FUT2*** | rs601338 | G | 158 | 111 | 99 | 0,0154* |
| | | A | 112 | 87 | 117 | |
| | | | | | | |
| ***FUT2*** | rs1800027 | C | 267 | 195 | 210 | 0,4921 |
| | | T | 9 | 3 | 6 | |

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***FUT2*** | rs1800025 | G | 234 | 181 | 193 | 0,2676 |
| | | A | 36 | 17 | 27 | |
| | | | | | | |
| ***FUT2*** | rs602662 | G | 134 | 111 | 99 | 0,0743 |
| | | A | 112 | 87 | 117 | |
| | | | | | | |
| ***FUT2*** | rs1799761 | – | 272 | – | – | – |
| | | | 4 | | | |
| | | | | | | |
| ***FUT2*** | rs485186 | A | 159 | 110 | 97 | 0,0113* |
| | | G | 115 | 88 | 119 | |

**Significant P value (Chi–2 test)*

### 3.1.5.3.2 *FUT2* and *FUT3* haplotype frequencies in the current cohort

We next examined the *FUT2*/*FUT3* haplotypes present across chromosome 19 haplotypes in our cohort. We analysed LD in *FUT2* and *FUT3* genes using Haploview [v4.2] (**Figure 3.10**). Six *FUT2* and two SNPs in *FUT3* were in strong LD as indicated by dark red colour which represent high D' and $r^2$ values (**Figure 3.10**). However, these two haplotype blocks only covered small portions of both genes because only 21 SNPS were assessed.

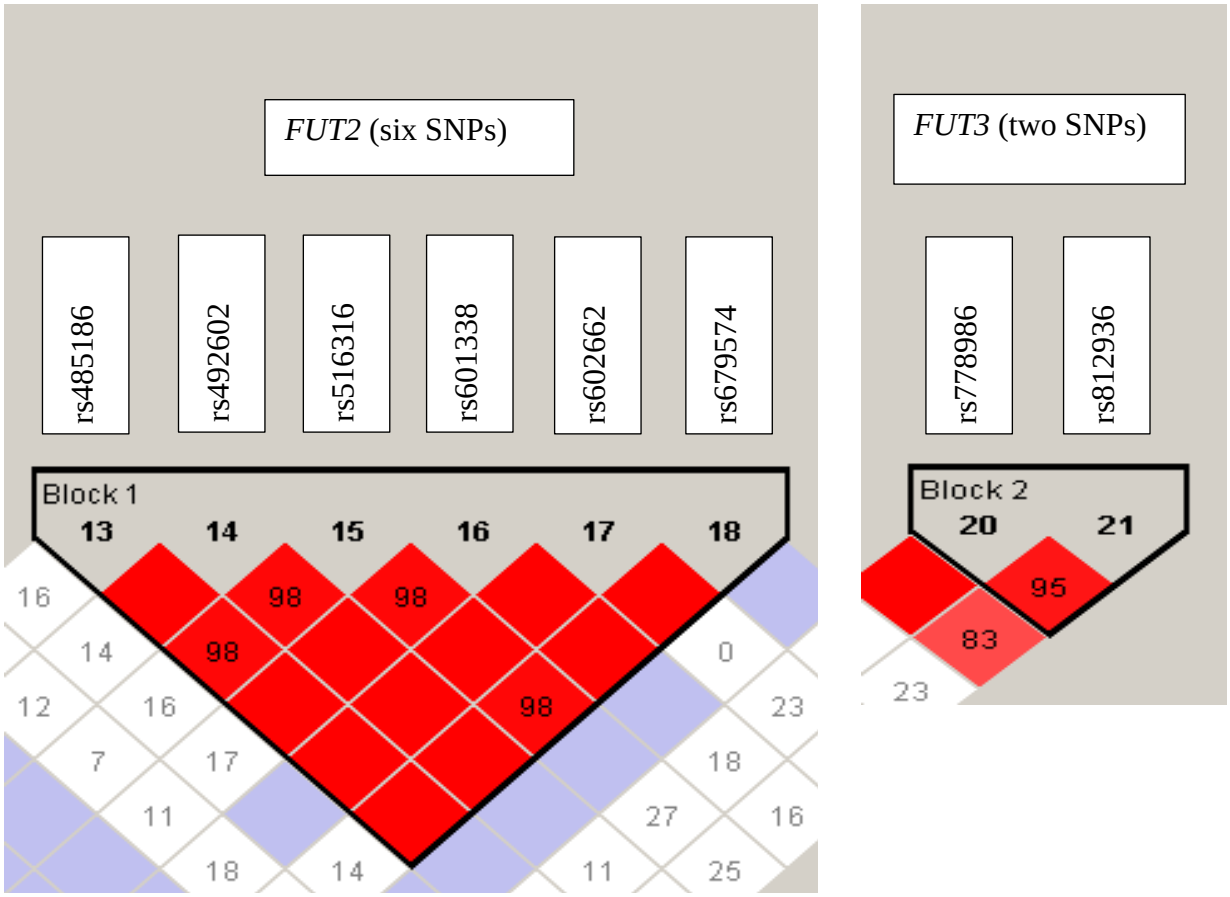



**Figure 3.10** LD plot of *FUT2* and *FUT3* markers (Pairwise LD). **Block 1 has six markers whereas Block 2 has two markers. Colours represent D' values (dark red=high inter–SNP D'; blue=statistically ambiguous D'; white=low inter–SNP D'). R–squared values written within the LD blocks*.

#### 3.1.5.3.2.1 *FUT2* haplotype phasing results

*FUT2*/*FUT3* haplotype frequencies were determined by phasing the MassArray sequencing data of each gene separately in gPLINK [v.2.050]. Haplotype phasing analysis was subsequently used to predict Se (*FUT2*) and Le (*FUT3*) statuses of our cohort and also to determine the association between Secretor phenotype and vaccine seroconversion. The results have been recorded below. **Table 3.11** below is a haplotype table of 12 *FUT2* SNPs that were analysed.

A total of 11 *FUT2* haplotypes were observed using 12 SNPs in 276 chromosomes. The wild type Se+ haplotype was found in 70/276 chromosomes (25%). The total frequency of Se+ haplotypes was 116/276 chromosomes (42%). Se481 associated with intermediate Se expression was present in 37/276 chromosomes (13.4%). Four non–Secretor haplotypes seen, two associated with 428 stop mutation and two associated with 778del frameshift mutation, overall frequency of non–Secretor haplotypes was 123/276 chromosomes (44.6%).

Haplotype phasing results were used to predict the Se status/phenotype of participants in our study. The following Se phenotype counts were inferred from *FUT2* haplotype phasing results (n= 138): Se+ = 96 (70%); Weak Se = 19 (14%). Even though a high frequency of non–Secretor haplotypes was found, they were very often found in genotypic combination with a Se haplotype. Therefore only 23/138 (17%) of individuals in our cohort predicted to be non–Secretors based on genetic data.

We compared the predicted Se phenotypes measured from genetic data to ELISA Se phenotypic data, and the genetic/phenotypic concordance was 83% (114/128). Fourteen samples did not have matching genetic and phenotypic and thus, genetic/phenotypic discordance was 10.9% (14/128). Se+ and Se– are the phenotypes that did not match in this case, however, all weak Se+ matched.

**Table 3.11** Haplotype table of 12 *FUT2* SNPs.

| | -139 | -102 | 30 | 113 | 171 | 375 | 428 | 480 | 481 | 739 | 778 | 960 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | intron | intron | syn | missense | syn | syn | stop | missense | missense | missense | frameshift | syn | | | | |
| **Secretor phenotype** | **rs679574** | **rs516316** | **rs370871822** | **rs114018037** | **rs492602** | **rs281377** | **rs601338** | **rs1800027** | **rs1800025** | **rs602662** | **rs1799761** | **rs485186** | **alt.name** | **Haplotype** | **Count** | **Frequency** |
| **Se+ (REF)** | C | G | G | C | A | C | G | C | G | G | C | A | wt | CGGCACGCGGCA | 70 | 25,4 |
| **non.1** | G | C | . | . | G | . | A | . | . | A | . | G | se428.1 | GCGCGCACGACG | 118 | 42,8 |
| **Se+.1** | . | . | . | T | . | . | . | . | . | . | . | . | Se113 | CGGTACGCGGCA | 13 | 4,7 |
| **non.2** | . | . | . | . | . | T | . | T | . | A | DEL | . | se480,739,778 | CGGCATGTGADELA | 1 | 0,4 |
| **Se+.2** | . | . | . | . | . | T | . | T | . | . | . | . | Se480 | CGGCATGTGGCA | 4 | 1,4 |
| **weak** | . | . | . | . | . | . | . | | A | . | . | . | Se481 | CGGCACGCAGCA | 37 | 13,4 |
| **Se+.3** | . | . | . | . | . | T | . | . | . | . | . | . | Se357 | CGGCATGCGGCA | 27 | 9,8 |
| **Se+.4** | . | . | A | . | . | . | . | . | . | . | . | . | Se30 | CGACACGCGGCA | 1 | 0,4 |
| **non.3** | . | . | . | . | . | T | . | T | . | . | DEL | . | se480,778 | CGGCATGTGGDELA | 3 | 1,1 |
| **Se+.5** | . | C | . | . | . | T | . | T | . | A | . | . | Se480,739 | CCGCATGTGACA | 1 | 0,4 |
| **non.5** | G | . | . | . | G | . | A | . | . | A | . | G | se428.2 | GGGCGCACGACG | 1 | 0,4 |
| | | | | | | | | | | | | | | **Total** | **276** | **100,0** |

#### 3.1.5.3.2.2 *FUT3* haplotype phasing results

**Table 3.12** below is a haplotype table of nine *FUT3* SNPs that were analysed. A total of 15 *FUT3* haplotypes were seen in 276 chromosomes, and the wild type haplotype was the most common, (50.4%). Only one other Le+ haplotype was observed, Le314 in one individual. The rest of the haplotypes (49.5%) were Le– due to a variety of inactivating missense mutations. The le508 haplotype was the most common Le– haplotype (29%), and there were two other phylogenetic clusters of other Le– haplotypes, including haplotypes with mutations 13/484/667/808 accounting for 10% of haplotypes, and haplotypes with mutations 202/314/862/1067 accounting for another 10% of haplotypes.

Haplotype phasing results were used to predict the Le status/phenotype of participants in our study. Overall, from genotypic data it was predicted that 79.7% of the cohort would have a Le+ phenotype, while 20.3% would have a Le– phenotype.

In the 138 samples with both genetic and ELISA data, we found 75% (97/129) concordance between predicted and measured phenotypes for Le status. On the other hand, discordance was 24.8% (32/129) whereby all four phenotypes: Le (a–b+); Le (a+b–); Le (a–b+) or Le (a+b+); Le (a–b–) did not match.

**Table 3.12** Haplotype table of nine *FUT3* SNPs.

| | le1067 | le862 | le808 | le667 | le508 | le484 | Le314 | le202 | le13 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | missense | missense | missense | missense | missense | missense | missense | missense | missense | | | | |
| **Le phenotype** | **rs3894326** | **rs761121538** | **rs28381968** | **rs28362466** | **rs3745635** | **rs28362463** | **rs778986** | **rs812936** | **rs28362458** | **Haplotype** | **Alt.name** | **Count** | **Frequency** |
| **Le+ (REF)** | A | C | C | C | C | C | G | A | C | ACCCCCGAC | Wt (ref) | 139 | 50,4 |
| **Le-.1** | . | . | . | . | . | . | . | . | T | ACCCCCGAT | le13 | 12 | 4,3 |
| **Le-.2** | . | . | . | . | T | . | . | . | . | ACCCTCGAC | le508 | 80 | 29,0 |
| **Le-.3** | . | . | . | . | . | . | A | G | . | ACCCCCAGC | le202,314 | 14 | 5,1 |
| **Le-.4** | . | . | . | T | . | T | . | . | T | ACCTCTGAT | le13,484, 667 | 5 | 1,8 |
| **Le+.5** | . | . | . | . | . | . | A | . | . | ACCCCCAAC | Le314 | 1 | 0,4 |
| **Le-.6** | . | T | . | . | . | . | A | . | . | ATCCCCAAC | le314, 862 | 1 | 0,4 |
| **Le-.7** | . | T | . | . | . | . | A | G | . | ATCCCCAGC | le202,314,862 | 6 | 2,2 |
| **Le-.8** | . | . | . | T | . | . | . | . | T | ACCTCCGAT | le13,667 | 3 | 1,1 |
| **Le-.9** | . | . | T | T | . | . | . | . | T | ACTTCCGAT | le13, 667, 808 | 6 | 2,2 |
| **Le-.10** | . | . | . | . | . | . | . | G | . | ACCCCCGGC | le202 | 1 | 0,4 |
| **Le-.11** | T | . | . | . | . | . | . | . | . | TCCCCCGAC | le1067 | 3 | 1,1 |
| **Le-.12** | T | . | . | . | . | . | A | G | . | TCCCCCAGC | le202,314,1067 | 2 | 0,7 |
| **Le-.13** | . | . | T | T | . | . | . | . | . | ACTTCCGAC | le667, 808 | 2 | 0,7 |
| **Le-.14** | . | . | T | T | . | T | . | . | T | ACTTCTGAT | le13,484, 667, 808 | 1 | 0,4 |
| | | | | | | | | | | | **Total** | **276** | **100,0** |

#### 3.1.5.3.2.3 *FUT2*/*FUT3* haplotype frequency data in the current study cohort compared to Erythrogene blood group genes database

The *FUT2* and *FUT3* haplotype frequency data above were compared to Erythrogene blood group genes database [v0.8] (Möller et al., 2016). However, only common *FUT2*/*3* haplotypes in our cohort were compared to the blood group genes database and the results are recorded in **Table 3.13** and **Table 3.14**, respectively.

**Table 3.13** *FUT2* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database.

| Haplotype | Current study cohort | Africa | America | East Asia | Europe | South Asia | Global |
|---|---|---|---|---|---|---|---|
| **Wild type** | 25.36 | 27.46 | 58.79 | 52.58 | 51.49 | 45.91 | 45.29 |
| **se428.1** | 42.75 | 48.79 | 34.15 | 0.40 | 43.84 | 26.28 | 31.63 |
| **Se481** | 13.41 | 8.85 | 0.14 | 0 | 0.10 | 0 | 2.38 |
| **se428.2** | 0.36 | 48.79 | 34.15 | 0.40 | 43.84 | 26.28 | 31.63 |

**Se – Secretor; *se – non–Secretor*

Chi–2 test was carried out to investigate whether there is significant difference between *FUT2* haplotype frequencies for the current study cohort and other countries (including global data). There was a significant difference between the frequencies (Chi–test; $P < 0.0001$).

**Table 3.14** *FUT3* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database.

| Haplotype | Current study cohort | Africa | America | East Asia | Europe | South Asia | Global |
|---|---|---|---|---|---|---|---|
| **Wild type** | 50.36 | 34.80 | 56.63 | 65.28 | 70.68 | 60.22 | 56.13 |
| **le202,314** | 5.07 | 7.72 | 11.96 | 3.08 | 16.20 | 11.15 | 9.74 |
| **le13,484, 667** | 1.81 | 8.70 | 0.86 | 0 | 0 | 0 | 2.42 |
| **le1067** | 1.09 | 0.08 | 0 | 0.60 | 0 | 0.20 | 0.18 |
| **le13,484, 667, 808** | 0.36 | 3.78 | 0.43 | 0 | 0 | 0 | 1.06 |

**Se – Secretor; *se – non–Secretor*

Chi–2 test was again carried out to investigate whether there is significant difference between the *FUT3* haplotype frequencies in the current study cohort and other countries (including global data). There was a significant difference between the frequencies (Chi–test; $P < 0.0001$).

### 3.1.6 Associations between *FUT2* and *FUT3* genetic variation and antibody responses to Rotarix vaccination

#### 3.1.6.1 Univariate analysis of association between *FUT2* and *FUT3* allele and seroconversion

Association between alleles in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. The results of the statistical analysis are recorded in **Table 3.15** below. One SNP, rs761121538 ( le862 in *FUT3*) was significantly associated with seroconversion (vaccine–take) ($P < 0.05$; PEMP1 < 0.05). The rs761121538 (*FUT3*) SNP remained significantly associated with seroconversion when the P value was adjusted using Bonferroni correction ($P = 0.042$) (**Table 3.15**).

**Table 3.15** Statistical results of univariate analysis of association between *FUT2* and *FUT3* allele frequencies and seroconversion using Chi–square test ($\chi^2$).

| Gene | SNP ID | MAF in cases (seroconverters) | MAF in controls (non–seroconverters) | P value (Chi–square test) | *P value* (Bonferroni correction) | $P_{EMP1}$ | $P_{EMP2}$ | OR |
|---|---|---|---|---|---|---|---|---|
| ***FUT3*** | rs761121538 | 0.005682 | 0.075 | 0.002* | 0.002*21= 0.042* | 0.003* | 0.051** | 0.070 |

*Significant P values indicated by (*)*

*Borderline significant P values indicated by (**)*

$P_{EMP1}$ *– Empirical P value*

$P_{EMP2}$ *– Max(T) Empirical P value*

*OR – Odds Ratio*

### 3.1.6.2 Univariate analysis of association between *FUT2* and *FUT3* genotypes and seroconversion

Genotypes associations were considered using dominant, recessive, and genotypic (codominant) models, as well as the Cochran–Armitage trend test. Association between genotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) or Fisher's exact test and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. None of the generated genotypes (14) with their respective SNPs were significantly associated with seroconversion when comparing cases vs. controls in the current study ($P > 0.05$; $P_{EMP1} > 0.05$).

### 3.1.6.3 Univariate analysis of association between *FUT2* and *FUT3* haplotypes and seroconversion

Association between haplotypes (all 21 SNPs in chromosome 19) in in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050].

One haplotype was significantly associated with seroconversion when comparing cases vs. controls in the current study (P= 0.038) (**Table 3.16**).

rs3894326, rs761121538, rs28381968, rs28362466, rs3745635, rs28362463, rs778986, rs812936, rs28362458, rs679574, rs516316, rs370871822, rs114018037, rs492602, rs281377, rs601338, rs1800027, rs1800025, rs602662, rs1799761, rs485186

Haplotype: (ATCCCCAGCGCGCGCACGACG) (*FUT3 – FUT2*) (**Table 3.16**).

**Table 3.16** Statistical results of univariate analysis of association between *FUT2* and *FUT3* haplotype (21 SNPs) frequencies and seroconversion using Chi–square test ($\chi^2$).

| Locus | Haplotype | Cases (seroconverters) | Controls (non–seroconverters) | P value (Chi–2 test) | *P value* (Bonferroni correction) | SNPs |
|---|---|---|---|---|---|---|
| ***FUT3 – FUT2*** | ATCCCCAGCGCGCGCACGACG | 0.009 | 0.052 | 0.038 | 0.038*21= 0.798 | rs3894326, rs761121538, rs28381968, rs28362466, rs3745635, rs28362463, rs778986, rs812936, rs28362458, rs679574, rs516316, rs370871822, rs114018037, rs492602, rs281377, rs601338, rs1800027, rs1800025, rs602662, rs1799761, rs485186 |

### 3.1.6.4 Univariate analysis of associations between Secretor or Lewis phenotypes predicted from genetic data, and seroconversion

*FUT2* phased genetic data were used to predict Se phenotype for each individual, and these Se phenotype counts were compared in seroconverters vs non–seroconverters. There was no significant relationship between Se phenotype in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.2885) (**Figure 3.11**). The frequency table can be found in **Appendix F**. This was similar to the result obtained in the same cohort using ELISA data only.

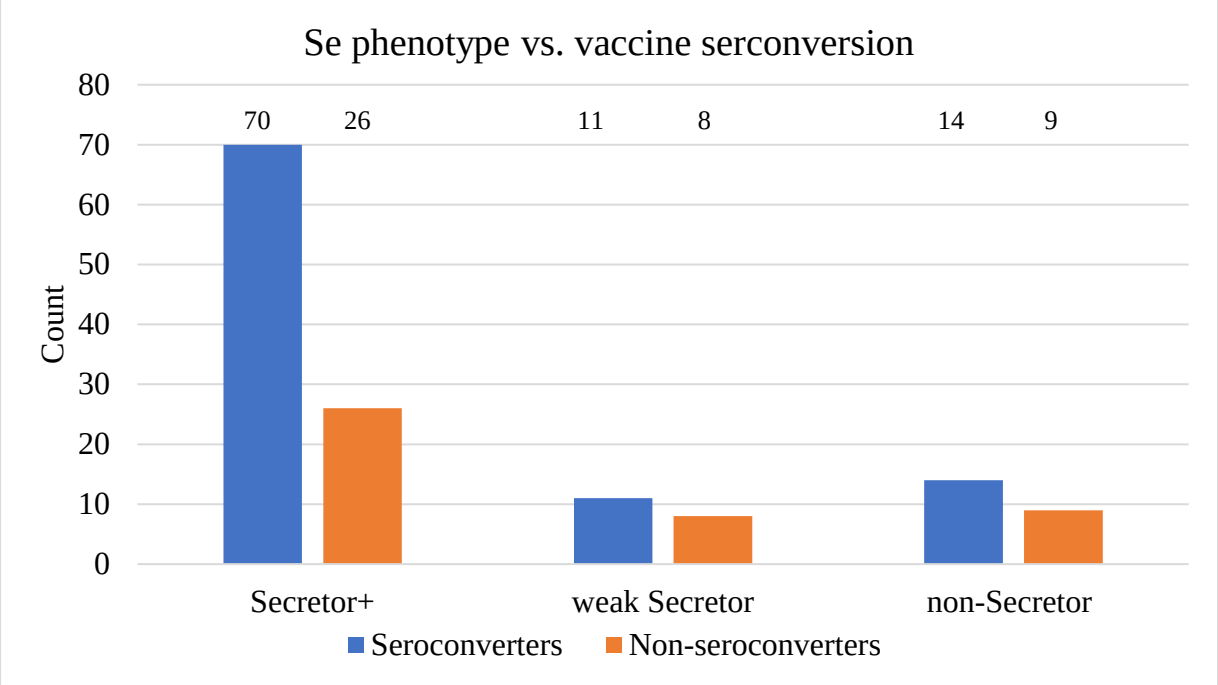



**Figure 3.11** Predicted Se phenotype counts compared between seroconverters and non–seroconverters (Chi–2 test; P= 0.2885).

*FUT3* phased genetic data were used to predict presence or absence of Le antigen for each individual, and these were compared in seroconverters vs non–seroconverters. **Figure 3.12** show Le antigen counts (Le+ vs Le–), and there was a significant difference between Le antigen presence in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.039). The frequency table can be found in **Appendix E**. This was different from the result obtained in the same cohort using ELISA data only.

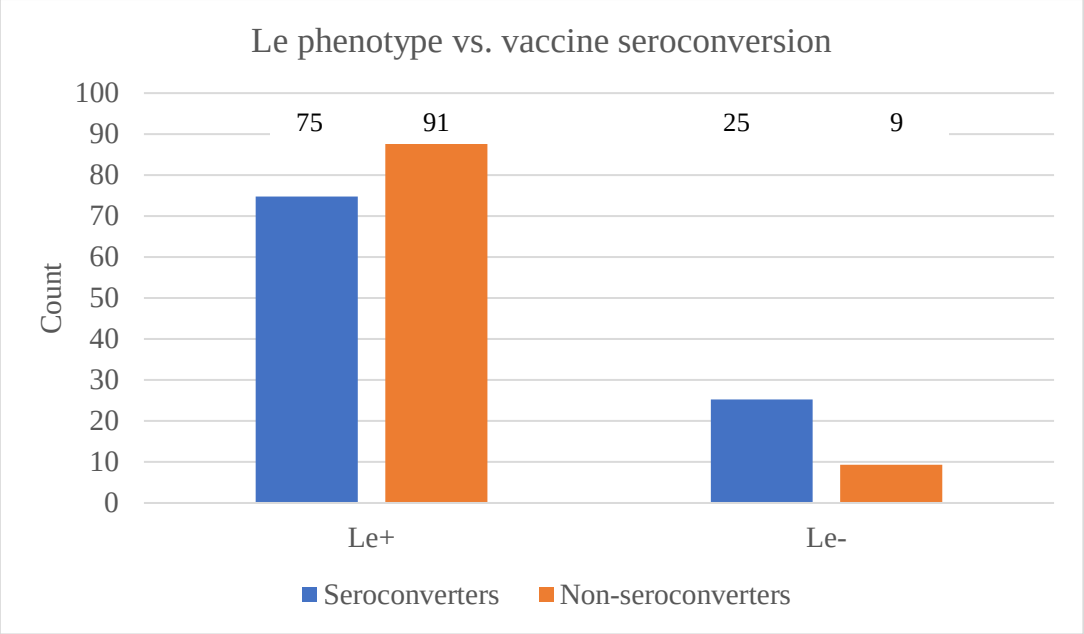



**Figure 3.12** Predicted Le phenotype counts compared between seroconverters and non–seroconverters. There was a statistical difference between positive and negative Le antigen counts in seroconverters vs. non–seroconverters (Chi–test; P= 0.039).

Se phenotype data generated from *FUT2* phasing was then combined with Le antigen phenotypes predicted from *FUT3* genetic data, to predict complete Le phenotypes and the counts were as follows (**Table 3.17**): Le (a+b–) = 15 (11%); Le (a–b–)= 28 (20%); Le (a–b+)= 95 (69%). The Le (a–b+) category included 17 weak Secretors who could also present as Le(a+b+).

**Table 3.17** The predicted Le antigens and their respective counts in the current study.

| | **Se+ (including 17 weak Se+)** | **Se–** | **Total** |
|---|---|---|---|
| **Le+ (including weak Le+)** | Le (a–b+): 95 | Le (a+b–): 15 | 110 |
| **Le–** | Le (a–b–): 20 | Le (a–b–): 8 | 28 |
| **Total** | 115 | 23 | 138 |

We then compared Le phenotype counts predicted from combined *FUT2* and *FUT3* genetic data vs. seroconversion status. **Figure 3.13** shows this analysis if the weak Secretors are analysed as a separate category (since they could cause either Le (a–b+) or Le (a+b+) i.e., four Le phenotypes), whilst **Figure 3.14** shows the analysis where the weak Secretors are included with the normal Secretors as Le (a–b+) i.e., three Le phenotypes.

In **Figure 3.13**, there was a significant difference between Le phenotypes in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.026). This was different from the result obtained in the same cohort using ELISA data only. The frequency table can be found in **Appendix E**. Similarly, there was a significant difference between predicted Le phenotypes defined in **Figure 3.14** between seroconverters vs. non–seroconverters (Chi–2 test; P= 0.026).
The frequency table can be found in **Appendix F**.

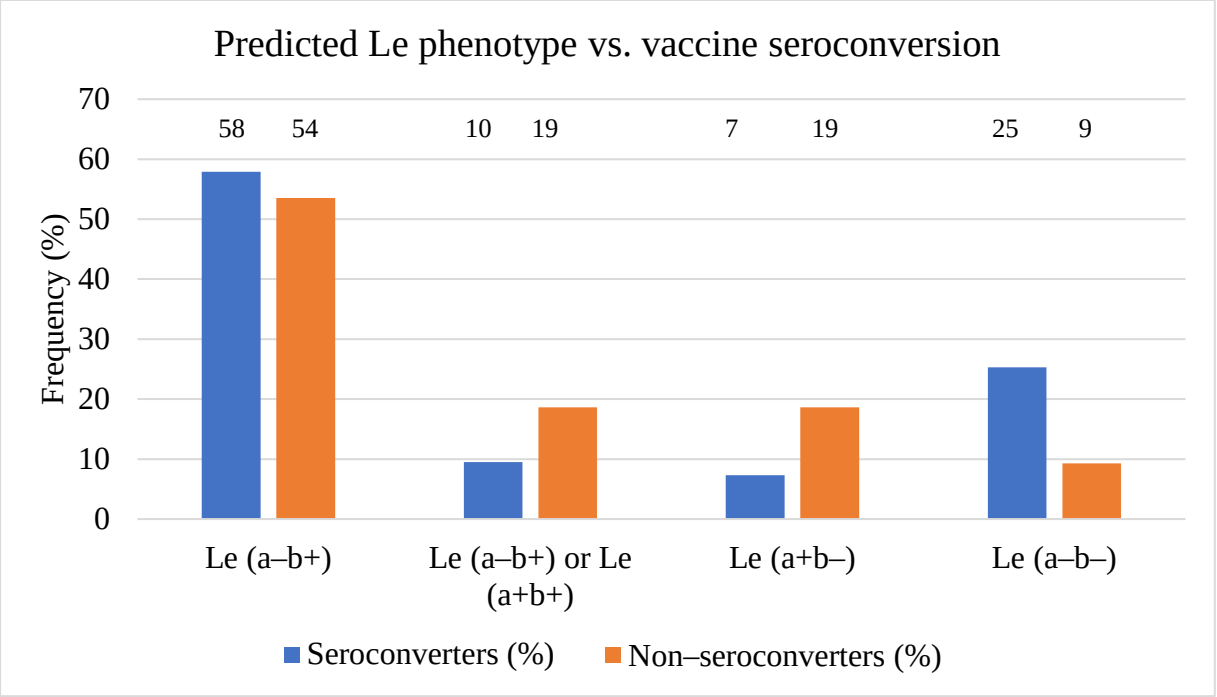



**Figure 3.13** Predicted Le phenotype counts between seroconverters and non–seroconverters. There was a significant statistical difference between Le phenotype frequencies in seroconverters vs. non–seroconverters (Chi–test; P= 0.026).

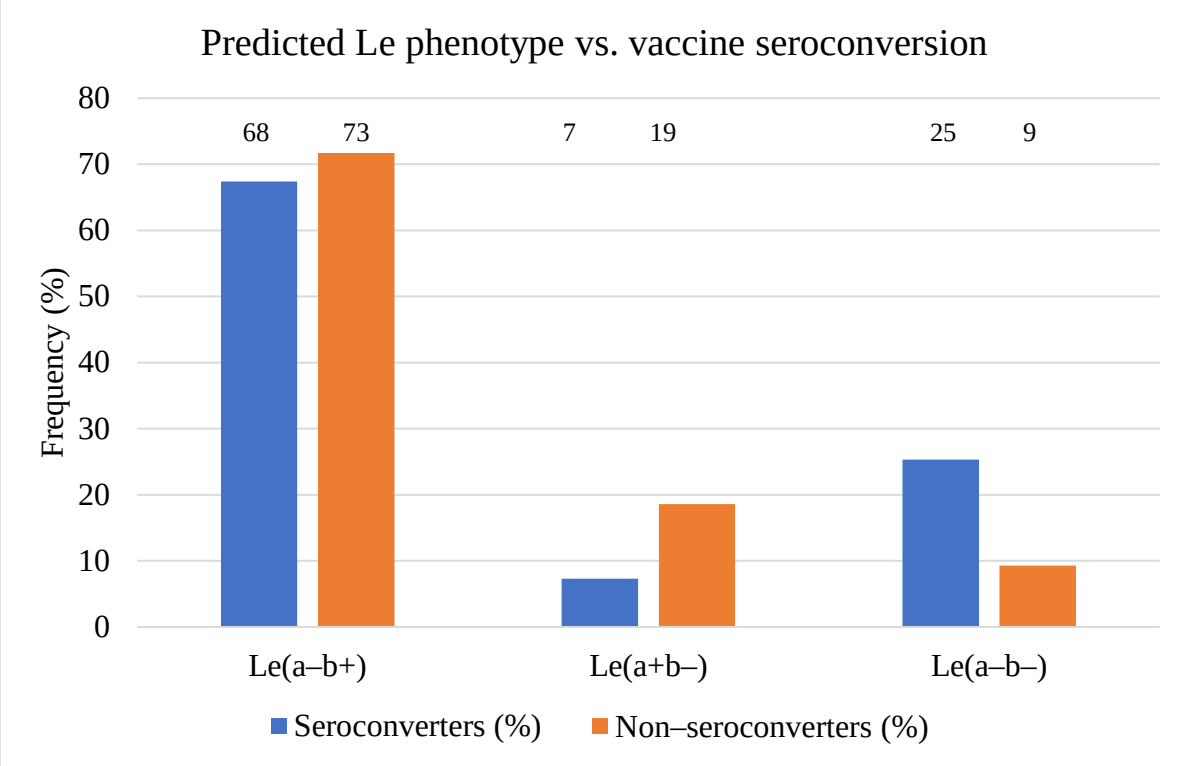



**Figure 3.14** Combined Le phenotype counts between seroconverters and non–seroconverters. There was a statistical difference between Le phenotype frequencies in seroconverters vs. non–seroconverters (Chi–test; P= 0.026).

### 3.1.6.5 Multivariable analysis of association between *FUT2* and *FUT3* genetic variation and seroconversion

Association between *FUT2* and *FUT3* genetic variation (alleles, genotypes, and haplotypes) in cases (seroconverters) vs. controls (non–seroconverters) including covariates (infant baseline IgA and IgG titres), was investigated using logistic regression in gPLINK [v.2.050].
The generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050].

None of the *FUT2* and *FUT3* alleles, genotypes or haplotypes remained significantly associated with seroconversion when the above–mentioned covariates were included in the multivariable model ($P > 0.05$ ).

### 3.1.7 Epistasis

Epistasis analysis was carried out in gPLINK [v.2.050] to determine the interaction between *FUT2* and *FUT3* SNPs of interest. This analysis tests all pairwise combination of SNPs in case–control data. A significant association was observed between *FUT3* SNP rs28362466 and two *FUT2* SNPs rs516316 at position –102 and rs485186 at position 960 after epistasis analysis (P= 0.047; P= 0.048) (**Table 3.18)**. The same *FUT3* SNP showed a trend towards significant epistatic relationship with two other *FUT2* SNPs, rs601338 (se429) and rs602662 (se739),

P= 0.06 for each (**Table 3.18**). These four *FUT2* SNPs are all part of the common non–Secretor–causing se428 haplotype. These results suggest that vaccine seroconversion is significantly associated with epistatic interactions between SNPs on *FUT2* se428 haplotype and *FUT3* le667 SNP (which is usually on 13/484/667 haplotypes). This combination of non–Secretor + absence of Le antigen would create the Le (a–b–) phenotype. OR was strongly positive (11–15) suggesting this combination is more common in seroconverters.

**Table 3.18** Epistasis analysis results between *FUT2* and *FUT3* SNPs.

| ***FUT3*** **SNP** | ***FUT2*** **SNP** | **P value** | **OR** |
|---|---|---|---|
| rs28362466 | rs516316 | 0.047 | 13.06 |
| rs28362466 | rs601338 | 0.066 | 11.26 |
| rs28362466 | rs602662 | 0.07 | 15 |
| rs28362466 | rs485186 | 0.048 | 12.74 |

# 3.2 ABO phenotypes and genotypes, and their association with antibody responses to rotavirus vaccination in a South African cohort

### 3.2.1 ABO phenotypes by ELISA in Secretors, compared at two different time–points

ABO phenotypes were tested in saliva samples by ELISA, and only Secretors would be expected to display these antigens in their saliva. Therefore, only results from the 138 participants deemed to be Secretors by previous experiments were considered and their ABO frequencies are shown in **Table 3.19**.

ABO blood types (A, B, AB, and O) of the participants in the current study cohort at 14 weeks of age were compared to frequencies at time–point 2 (at 18 months of age)

The A, B, O, and AB blood type distribution for the first time–point ELISA was 28.4% for A, 1.9% for B, 69.7% for O, and 0% for AB. Blood type AB was completely absent in this cohort for the first time–point ELISA and blood type B was extremely low (**Table 3.19**). ABO phenotyping by ELISA was repeated at a later time–point (18 months of age) to check whether difference in age of the infants has any significant effect on the expression of these HBGAs and ultimately their association with seroconversion. ABO distribution data in the current study was compared between time–point 1 and time–point 2, and there was no significant difference between the two data (Chi–2 test; P= 0,517).

The A, B, O, and AB blood type distribution for the second time–point ELISA was 26.6% for A, 11.5% for B, 59.7% for O, and 4.35% for AB blood type (**Table 3.19**). There was no difference in the same individuals from time–point 1 to time–point 2 in terms of the ABO blood types.

**Table 3.19** Current study cohort first time–point 1 and time–point 2 ABO blood type distribution in Secretors, compared to other SA Bantu speaking cohorts from Brain (1966) and Shapiro (1951) publications.

| | n= 162 | | n= 189 | n= 867 | n= 4820 |
|---|---|---|---|---|---|
| **Time–point 1 blood type** | **Current study (%)** | **Time–point 2 blood type** | **Current study (%)** | **Brain (1966) (%)** | **Shapiro (1951) (%)** |
| **A** | 33.33 | **A** | 26.6 | 30.70 | 29.50 |
| **B** | 12.32 | **B** | 11.5 | 19.40 | 19.60 |
| **O** | 50.00 | **O** | 59.7 | 43.80 | 46.10 |
| **AB** | 4.35 | **AB** | 2.2 | 6.10 | 4.77 |
| **Total** | 100 | **Total** | 100 | 100 | 99.97 |

### 3.2.2 Characterisation of *ABO* genetic variation by MassArray analysis

These analyses allowed us to consider ABO blood types in the entire cohort and not just in the Secretors.

#### 3.2.2.1 *ABO* allele frequency distribution in the current study cohort compared to other African populations

A total of nine SNPs in the *ABO* gene were used in MassArray, chosen to represent various ABO blood antigen types as shown in **Table 3.20** below. These SNPs were observed in the cohort at frequencies ranging from 3.3 to 43.2%.**Table 3.20** below shows MAF of ABO SNPs in the SA cohort compared to LWK and YRI cohorts from the 1000 Genomes project. As previously mentioned, there is no previously published SA data, so we used LWK and YRI as proxy cohorts in the current study.

Four ABO SNPs were found to be significantly different between the three groups as marked with (*) (Chi–2 test; $P < 0.05$) (**Table 3.20**). Furthermore, four SNPs were not significantly different between the three groups (Chi–2 test; $P > 0.05$). Lastly, rs782023144 SNP did not have MAF data from the 1000 Genomes project database (**Table 3.20**).

**Table 3.20** *ABO* allele frequencies in the current study cohort (SA) vs. other two African populations i.e., LWK and YRI.

| Gene | SNP ID | Alleles | ABO blood antigen that this SNP is associated with | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|---|
| | | | | MAF | MAF | MAF | |
| ***ABO*** | rs56392308 | GGG | A2 | 231 | 188 | 210 | 0,0052* |
| | | GG | | 25 | 10 | 6 | |
| | | | | | | | |
| ***ABO*** | rs8176748 | C | O | 219 | 148 | 174 | 0,2826 |
| | | T | | 55 | 50 | 42 | |
| | | | | | | | |
| ***ABO*** | rs8176747 | C | B | 247 | 166 | 179 | 0,0091* |
| | | G | | 23 | 32 | 37 | |
| | | | | | | | |
| ***ABO*** | rs8176746 | G | B | 250 | 166 | 32 | <0,0001* |
| | | T | | 26 | 179 | 37 | |
| | | | | | | | |
| ***ABO*** | rs8176741 | G | Aint/B | 233 | 165 | 179 | 0,0628 |
| | | A | | 27 | 33 | 37 | |
| | | | | | | | |
| ***ABO*** | rs1053878 | G | A2, ABantu, cisAB | 160 | 158 | 166 | 0,0166* |
| | | A | | 74 | 40 | 50 | |
| | | | | | | | |
| ***ABO*** | rs8176720 | T | Aint/B/O1v/O2 | 153 | 99 | 116 | 0,3147 |
| | | C | | 115 | 99 | 100 | |

| Gene | SNP ID | Alleles | ABO blood antigen that this SNP is associated with | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|---|
| | | | | MAF | MAF | MAF | |
| ***ABO*** | rs8176719 | – | O1 | 159 | 141 | 166 | 0,1373 |
| | | C | | 73 | 57 | 50 | |
| | | | | | | | |
| ***ABO*** | rs782023144 | – | ABantu, OBantu | 267 | – | – | – |
| | | | | 9 | | | |

**Significant P value (Chi–2 test)*

### 3.2.2.2 ABO haplotype frequencies in the current cohort

ABO haplotypes were determined by phasing the MassArray sequencing data of all the *ABO* SNPs in gPLINK [v.2.050] (n= 138). Haplotype phasing analysis was subsequently used to predict ABO blood types of our cohort and also to determine the association between ABO phenotypes and vaccine seroconversion. The results have been recorded below.

**Table 3.21** below is a haplotype table of nine *ABO* SNPs that were analysed. The following haplotypes and their respective blood antigens were the most common in our SA cohort: GCCGGATDELC (O1), GTCGGGCDELC (O1v.1), and GCCGGGTDELC (O1.1) (**Table 3.21**). In contrast, the following haplotypes and their respective blood antigens were the least common in our cohort: GCCGAGTDELC (O1.2), GCCGAGTCC (A1.4), and GCCGAGCDELDEL (Obantu.2) (**Table 3.21**).

**Table 3.21** Haplotype table of nine *ABO* SNPs.

| **SNP position** | 1-1054 | 276 | 800 | 793 | 686 | 464 | 326 | 258-259 | 1-203 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variant type** | frameshift variant | missense variant | missense variant | missense variant | synonymous variant | missense variant | synonymous variant | frameshift variant | splice donor variant | | | |
| **ABO blood typ** | **rs56392308** | **rs8176748** | **rs8176747** | **rs8176746** | **rs8176741** | **rs1053878** | **rs8176720** | **rs8176719** | **rs782023144** | **Haplotype** | **Count** | **Frequency** |
| **A1 (ref)** | G | C | C | G | G | G | T | C | C | GCCGGGTCC | 12 | 4,4 |
| **A2** | DEL | . | . | . | . | A | . | . | . | DELCCGGATCC | 24 | 8,8 |
| **A1.1** | . | . | . | . | . | A | . | . | . | GCCGGATCC | 6 | 2,2 |
| **A1.2** | . | . | . | . | . | . | C | . | . | GCCGGGCCC | 12 | 4,4 |
| **Abantu.1** | DEL | . | . | . | . | A | . | . | DEL | DELCCGGATCDEL | 3 | 1,1 |
| **A1.3** | . | T | . | . | . | . | C | . | . | GTCGGGCCC | 3 | 1,1 |
| **A1.4** | . | . | . | . | A | . | . | . | . | GCCGAGTCC | 1 | 0,4 |
| **B1.1** | . | . | G | T | A | . | C | . | . | GCGTAGCCC | 22 | 8,1 |
| **O1** | . | . | . | . | . | A | . | DEL | . | GCCGGATDELC | 41 | 15,1 |
| **O1v.1** | . | T | . | . | . | . | C | DEL | . | GTCGGGCDELC | 46 | 16,9 |
| **O1.1** | . | . | . | . | . | . | . | DEL | . | GCCGGGTDELC | 64 | 23,5 |
| **O1v.2** | . | . | G | T | A | . | C | DEL | . | GCGTAGCDELC | 2 | 0,7 |
| **O1v.3** | . | . | . | . | . | . | C | DEL | . | GCCGGGCDELC | 16 | 5,9 |
| **Obantu.1** | . | . | . | . | . | . | C | DEL | DEL | GCCGGGCDELDEL | 5 | 1,8 |
| **O1.2** | . | . | . | . | A | . | . | DEL | . | GCCGAGTDELC | 1 | 0,4 |
| **O1v.4** | . | . | . | | . | A | C | DEL | . | GCCGGACDELC | 7 | 2,6 |
| **O1.3** | . | T | . | . | . | . | . | DEL | . | GTCGGGTDELC | 4 | 1,5 |
| **O1v.4** | . | T | . | . | A | . | C | DEL | . | GTCGAGCDELC | 2 | 0,7 |
| **Obantu.2** | . | . | . | . | A | . | C | DEL | DEL | GCCGAGCDELDEL | 1 | 0,4 |
| | | | | | | | | | | **Total** | **272** | **100,0** |

After ABO haplotypes were phased, this allowed us to interpret the genotype data per individual in order to calculate blood type frequencies. The following ABO blood type frequencies counts were determined (n= 138): A= 46 (33%); B= 17 (12%); O= 69 (50%); AB= 6 (4%) (**Table 3.21)**

As previously mentioned, the Bantu speaking cohort (n= 171) from SA had the following ABO blood type frequencies: A (63/171= 36.8%); B (31/171= 18.1%); O (67/171= 39.2%); AB (8/171= 4.7%) (Moores and Brain, 1968). There was no significant difference between our cohort and the Bantu speaking cohort (Chi–2 test; P= 0.266) (**Figure 3.15**). This was expected as both cohorts are from SA.

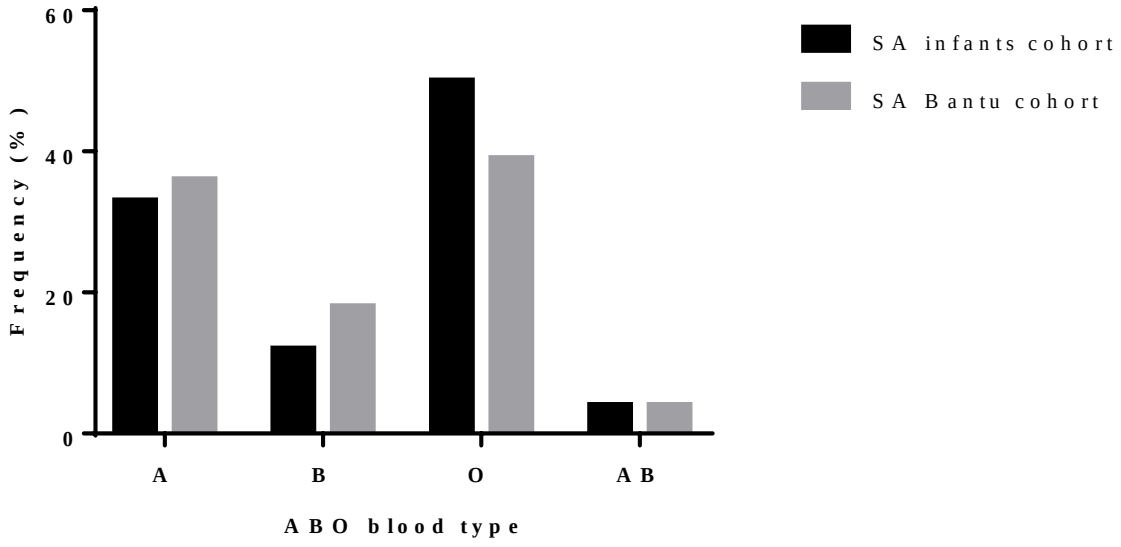



**Figure 3.15** ABO blood type frequency comparison between current study cohort and the SA Bantu speaking cohort (Chi–2 test; P= 0.266).

We compared thee predicted ABO blood types measured from genetic data to ELISA ABO phenotypic data at time–point 2, and the concordance was 102/130= 83%. This poor concordance may be due to lack of sensitivity of our ELISA assay.

#### 3.2.2.2.1 ABO haplotype frequency data compared to Erythrogene blood group genes database

The *ABO* haplotype frequency data above were compared to Erythrogene blood group genes database [v0.8] (Möller et al., 2016). However, only common *ABO* haplotypes in our cohort were compared to the blood group genes database and the results are recorded in **Table 3.22** below.

**Table 3.22** *ABO* haplotype frequencies for the current study cohort compared to Erythrogene blood group genes database.

| Haplotype | Current study cohort | Africa | America | East Asia | Europe | South Asia | Global |
|---|---|---|---|---|---|---|---|
| **A1 (ref)** | 4.41 | 4.39 | 11.53 | 3.47 | 18.09 | 12.58 | 9.54 |
| **A2** | 8.82 | 5.82 | 4.32 | 0 | 9.15 | 1.43 | 4.25 |
| **B1.1** | 8.09 | 13.39 | 4.76 | 18.85 | 8.45 | 22.90 | 14.16 |

**A1(ref) – A blood group (reference); A2 – A blood group; B1.1 – B1 blood group*

Chi–2 test was carried out to investigate whether there is significant difference between the current study cohort *ABO* haplotype frequencies and other countries (including global data). There was a significant difference between the frequencies (Chi–test; $P < 0.0001$).

### 3.2.2.3 Associations between ABO genetic variation and antibody responses to Rotarix vaccination

Association between alleles in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050]. None of ABO SNPs of interest in this study were significantly associated with seroconversion ($P > 0.05$; $P_{EMP1} > 0.05$). Association between genotypes in dominant, recessive, and genotypic (co–dominant) models were compared in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) or Fisher's exact test and the generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050]. None of the *ABO* genotypes with their respective SNPs were significantly associated with seroconversion when comparing cases vs. controls in the current study ($P > 0.05$; $P_{EMP1} > 0.05$).

Association between ABO haplotypes in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) in gPLINK [v.2.050]. Bonferroni correction was not applied to the generated P values. One haplotype, the OBantu haplotype, was significantly different in frequency between cases and controls (Chi–2 test; $P= 0.016$). ABO blood types were compared in seroconverters vs non–seroconverters. There was no significant relationship between ABO blood types in seroconverters vs. non–seroconverters (Chi–2 test; $P= 0.8076$) (**Figure 3.16**). The frequency table can be found in **Appendix F**.

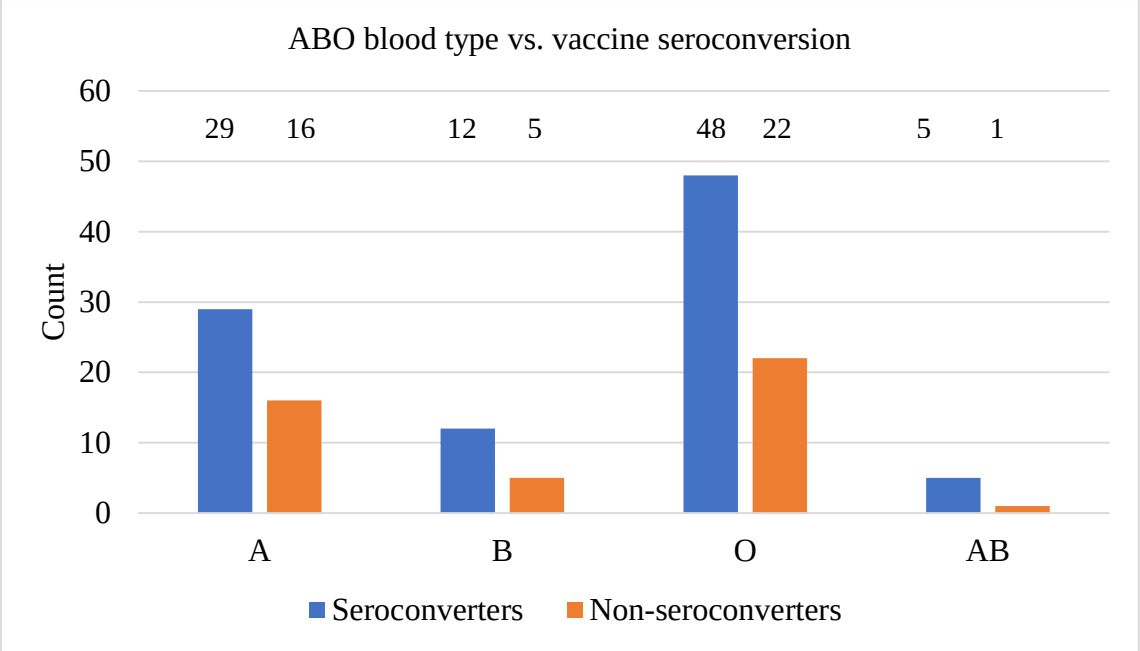



**Figure 3.16** ABO blood type counts between seroconverters and non–seroconverters. There was no statistical difference between Le phenotype counts in seroconverters vs. non–seroconverters (Chi–2 test; P= 0.8076).

Association between ABO alleles in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. None of the *ABO* genotypes were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

Association between *ABO* genotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. None of the *ABO* genotypes were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

Association between *ABO* haplotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was not investigated since there were no *ABO* haplotypes significantly associated with seroconversion during univariate analysis. None of the *ABO* haplotypes were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

# 3.3 Genetic variation in innate receptor genes, and their association with antibody responses to rotavirus vaccination in a South African cohort

### 3.3.1 Characterisation of innate receptor genetic variation by MassArray analysis

#### 3.3.1.1 Allele frequencies of SNPs in innate immune response genes in the current study cohort compared to other African populations

**Table 3.23** below shows the MAF distribution of innate immune response genes in the SA cohort compared to LWK and YRI populations from the 1000 Genomes Project. The SA innate immune receptor genes MAF range from 0.063 (minimum) to 0.463 (maximum). Seven SNPs were found to be significantly different between the three groups as marked with (*) (Chi–2 test; $P < 0.05$). Furthermore, 12 SNPs were not significantly different between the three groups (Chi–2 test; $P > 0.05$) (**Table 3.23**).

**Table 3.23** Allele frequencies of SNPs in innate immune response genes in the current study cohort (SA) vs. other two African populations LWK and YRI.

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***IFIH1*** | rs1990760 | C | 257 | 178 | 203 | 0,1903 |
| | | T | 17 | 20 | 13 | |
| | | | | | | |
| ***IFIH1*** | rs3747517 | T | 110 | 92 | 83 | 0,2341 |
| | | C | 160 | 106 | 133 | |
| | | | | | | |
| ***IFIH1*** | rs10930046 | T | 126 | 114 | 96 | 0,0150* |
| | | C | 146 | 84 | 120 | |
| | | | | | | |
| ***IFIH1*** | rs984971 | A | 26 | 40 | 42 | 0,0011* |
| | | G | 250 | 158 | 174 | |
| | | | | | | |
| ***TLR3*** | rs5743305 | T | 191 | 135 | 156 | 0,6602 |
| | | A | 83 | 63 | 60 | |
| | | | | | | |
| ***TLR3*** | rs13126816 | G | 257 | 184 | 195 | 0,4555 |
| | | A | 19 | 14 | 21 | |
| | | | | | | |
| ***TLR3*** | rs3775296 | C | 235 | 172 | 189 | 0,9644 |
| | | A | 33 | 26 | 27 | |
| | | | | | | |
| ***TLR3*** | rs5743312 | C | 230 | 182 | 199 | 0,9393 |
| | | T | 22 | 16 | 17 | |

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***TLR3*** | rs3775292 | C | 24 | 18 | 28 | 0,2602 |
| | | G | 250 | 180 | 188 | |
| | | | | | | |
| ***TLR3*** | rs10025405 | A | 216 | 143 | 174 | 0,0861 |
| | | G | 56 | 55 | 42 | |
| | | | | | | |
| ***DDX58*** | rs55789327 | G | 253 | 174 | 195 | 0,3929 |
| | | A | 23 | 24 | 21 | |
| | | | | | | |
| ***DDX58*** | rs669260 | T | 210 | 161 | 189 | 0,0091* |
| | | C | 64 | 37 | 27 | |
| | | | | | | |
| ***DDX58*** | rs10813831 | G | 212 | 166 | 174 | 0,2813 |
| | | A | 60 | 32 | 42 | |
| | | | | | | |
| ***TLR7*** | rs179008 | A | 178 | 140 | 137 | 0,0658 |
| | | T | 19 | 14 | 27 | |
| | | | | | | |
| ***TLR7*** | rs864058 | G | 123 | 111 | 129 | 0,0024* |
| | | A | 75 | 43 | 35 | |
| | | | | | | |
| ***TLR7*** | rs3853839 | C | 175 | 121 | 134 | 0,0191* |
| | | G | 21 | 33 | 30 | |

| Gene | SNP ID | Alleles | SA (current study) | LWK | YRI | P value (Chi–2 test) |
|---|---|---|---|---|---|---|
| | | | MAF | MAF | MAF | |
| ***TLR7*** | rs179007 | G | 49 | 16 | 25 | 0,0011* |
| | | A | 147 | 138 | 139 | |
| | | | | | | |
| ***TLR7*** | rs5935438 | G | 133 | 90 | 96 | 0,0884 |
| | | C | 62 | 64 | 68 | |
| | | | | | | |
| ***TLR8*** | rs2159377 | C | 175 | 110 | 137 | 0,0001* |
| | | T | 22 | 44 | 27 | |

**Significant P value (Chi–2 test)*

### 3.3.1.2 Haplotype frequencies of innate immune response genes in the current cohort

Haplotypes of SNPs in innate immune response genes (*DDX58*, *IF1H1*, *TLR3*, *TLR7*, and *TLR8*) were determined by phasing the MassArray sequencing data of each gene separately in gPLINK [v.2.050]

#### 3.3.1.2.1 *DDX58* haplotype phasing results

**Table 3.24** below is a haplotype table of three *DDX58* SNPs that were analysed.

Four *DDX58* haplotypes were found, and they had frequencies shown in **Table 3.24**. Wild type haplotype GTG was the most common (46%), whereas ATG was the least common in our cohort (8%) (**Table 3.24**).

**Table 3.24** Haplotype table of three *DDX58* SNPs.

| **SNP position** | **296** | **–** | **19** | | | |
|---|---|---|---|---|---|---|
| **Variant type** | missense variant | intron variant | missense variant | | | |
| **SNP** | **rs55789327 (G>A)** | **rs669260 (T>C)** | **rs10813831 (G>A)** | **Haplotype** | **Count** | **Frequency (%)** |
| **Reference** | G | T | G | GTG | 128 | 46,38 |
| **hap1** | . | C | . | GCG | 65 | 23,55 |
| **hap2** | . | . | A | GTA | 60 | 21,74 |
| **hap3** | A | . | . | ATG | 23 | 8,33 |
| | | | | **Total** | 276 | 100,00 |

#### 3.3.1.2.2 *IF1H1* haplotype phasing results

**Table 3.25** below is a haplotype table of four *IF1H1* SNPs that were analysed.

Five *IF1H1* haplotypes were found, and their frequencies were as follows: 39.13, 7.25, 5.07 1.45, and 0.72, respectively (**Table 3.25**). Haplotype CTTG was the most common in our cohort, whereas CTTT was the least common (**Table 3.25**).

**Table 3.25** Haplotype table of five *IF1H1* SNPs.

| **SNP position** | 2524 | 2411 | 1379 | – | | | |
|---|---|---|---|---|---|---|---|
| **Variant type** | missense variant | missense variant | missense variant | intron variant | | | |
| **SNP** | **rs1990760 (C>T)** | **rs3747517 (C>T)** | **rs10930046 (C>T)** | **rs984971 G>A)** | **Haplotype** | **Count** | **Frequency** |
| **Reference** | C | C | C | G | CCCG | 128 | 46,38 |
| **hap1** | . | T | T | . | CTTG | 108 | 39,13 |
| **hap2** | . | . | . | A | CCCA | 20 | 7,25 |
| **hap3** | T | . | T | . | TCTG | 14 | 5,07 |
| **hap4** | T | . | T | A | TCTA | 4 | 1,45 |
| **hap5** | . | T | T | A | CTTA | 2 | 0,72 |
| | | | | | **Total** | 276 | 100,00 |

#### 3.3.1.2.3 *TLR3* haplotype phasing results

**Table 3.26** below is a haplotype table of six SNPs that were analysed.

Ten *TLR3* haplotypes we found, and their respective frequencies are recorded in **Table 3.26** below. Wild type haplotype TGCCGA was the most common in our cohort (42%), whereas AACCGA was the least common (**Table 3.26**).

**Table 3.26** Haplotype table of six *TLR3* SNPs.

| **SNP postion** | – | – | – | – | – | – | – | – |
|---|---|---|---|---|---|---|---|---|
| **Variant type** | TF binding site | intron variant | splice region variant | intron variant | non coding transcript exon variant | 3 prime UTR variant | | |
| **SNP** | **rs5743305 (T>A)** | **rs13126816 (G>A)** | **rs3775296 (C>A)** | **rs5743312 (C>T)** | **rs37592 (G>C)** | **rs10025405 (A>G)** | **Haplotype** | **Frequency** |
| **Ref** | T | G | C | C | G | A | TGCCGA | 41,7 |
| **hap1** | A | . | . | . | . | . | AGCCGA | 17,9 |
| **hap2** | . | . | . | . | . | G | TGCCGG | 11,6 |
| **hap3** | . | . | . | . | C | . | TGCCCA | 6,8 |
| **hap4** | . | . | A | T | . | G | TGATGG | 5,6 |
| **hap5** | . | A | . | . | . | . | TACCGA | 3,2 |
| **hap6** | . | . | A | T | . | . | TGATGA | 2,9 |
| **hap7** | . | . | A | . | . | . | TGACGA | 3,8 |
| **hap8** | . | . | . | . | . | G | TGCCGG | 3,5 |
| **hap9** | A | A | . | | . | . | AACCGA | 3 |
| | | | | | | | Total | 100 |

#### 3.3.1.2.4 *TLR7* and *TLR8* haplotype phasing results

Both *TLR7* and *TLR8* genes are on the same chromosome (X chromosome) and thus, they make haplotypes together. A total of 15 haplotypes were formed by six SNPs in these two genes and their respective haplotypes frequencies are recorded in **Table 3.27** below. Interestingly the wild type haplotype was not the most common, found in only 17.5% of the cohort. Haplotype AACGGC was the most common in our cohort (20.5%), whereas TGCGGC was the least common (**Table 3.27**).

**Table 3.27** Haplotype table of six SNPs on chromosome X from *TLR7* and *TLR8*.

| **SNP postion** | – | – | – | – | – | – | – | – |
|---|---|---|---|---|---|---|---|---|
| **Variant type** | missense variant | synonymous variant | 3' UTR variant | intergenic variant | TF binding site | synonymous variant | | |
| **Gene** | *TLR7* | *TLR7* | *TLR7* | *TLR7/8* | *TLR7/8* | *TLR8* | | |
| **SNP** | **rs179008 (A>T)** | **rs864058 (G>A** | **rs3853839 (C>G)** | **rs179007 (A>G)** | **RS5935438 (G>C)** | **rs2159377 (C>T)** | **Haplotype** | **Frequency** |
| **Ref** | T | G | C | A | G | C | TGCAGC | 3,4 |
| **hap1** | A | A | . | G | . | . | AACGGC | 20,5 |
| **hap2** | A | . | . | . | . | . | AGCACC | 18,8 |
| **hap3** | A | . | . | . | . | . | AGCAGC | 17,5 |
| **hap4** | A | . | . | . | . | T | AGCAGT | 6,4 |
| **hap5** | A | A | G | . | . | . | AAGAGC | 6,3 |
| **hap6** | A | A | . | . | C | . | AACACC | 6,2 |
| **hap7** | A | A | . | . | . | . | AACAGC | 5,7 |
| **hap8** | . | . | . | . | . | T | TGCAGT | 3,4 |
| **hap9** | A | . | . | G | . | . | AGCGGC | 2,9 |
| **hap10** | A | A | . | . | C | T | AACACT | 2,6 |
| **hap11** | A | . | G | . | . | T | AGGAGT | 1,7 |
| **hap12** | . | A | . | . | C | . | TGCACC | 1,3 |
| **hap13** | . | A | . | . | . | . | TACAGC | 1,1 |
| **hap14** | A | A | . | G | C | . | AACGCC | 1,2 |
| **hap15** | T | . | . | G | . | . | TGCGGC | 1 |
| | | | | | | | **Total** | 100 |

### 3.3.2 Associations between genetic variation of innate immune response genes and antibody responses to Rotarix vaccination

#### 3.3.2.1 Univariate analysis of association between innate immune response genes allele frequencies and seroconversion

Association between allele frequencies in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050]. rs2159377 (*TLR8*) was found to be significantly associated with seroconversion ($P < 0.0146$; $P_{EMP1} < 0.01954$) in the current study (**Table 3.28**). However, rs2159377 (*TLR8*) became non–significantly associates with seroconversion when the P value was adjusted using Bonferroni correction (P= 0.2774) (**Table 3.28**).

**Table 3.28** Statistical results of univariate analysis of association between innate immune response genes allele frequencies and seroconversion using Chi–square test ($\chi^2$).

| Gene | SNP ID | MAF in cases (seroconverters) | MAF in controls (non–seroconverters) | P value (Chi–square test) | P value (Bonferroni correction) | $P_{EMP1}$ | $P_{EMP2}$ | OR |
|---|---|---|---|---|---|---|---|---|
| ***TLR8*** | rs2159377 | 0.1493 | 0.03175 | 0.0146* | 0.0146*19= 0.2774 | 0.01954 | 0.3771 | 5.351 |

*$P_{EMP1}$ – Empirical P value*

*$P_{EMP2}$ – Max(T) Empirical P value*

*OR – Odds Ratio*

#### 3.3.2.2 Univariate analysis of association between innate immune response genes genotype frequencies and seroconversion

Genotypes associations were considered using dominant, recessive, and genotypic (codominant) models, as well as the Cochran–Armitage trend test. Association between genotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) or Fisher's exact test and the generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050].

None of the genotypes from the innate immune response genes were significantly associated with seroconversion when using dominant, recessive, or codominant models comparing cases vs. controls in the current study ($P > 0.05$; $P_{EMP1} > 0.05$).

#### 3.3.2.3 Univariate analysis of association between innate immune response genes haplotype frequencies and seroconversion

Association between allele frequencies in cases (seroconverters) vs. controls (non–seroconverters) was investigated using Chi–square test ($\chi^2$) in gPLINK [v.2.050].

The following six SNP haplotype on the X chromosome was significantly associated with seroconversion (P= 0.042):

rs179008, rs864058, rs3853839, rs179007, rs5935438, rs2159377 (TGACACC).

This haplotype did not contain rs2159377 minor allele that was independently associated with seroconversion; instead, this haplotype (seen as haplotype number 12 on **Table 5.5**) included the minor alleles for rs5935438, affecting a transcription factor binding site.

Bonferroni correction was calculated by multiplying the number of haplotypes found on chromosome X by 15. The adjusted P value was not significant.

#### 3.3.2.4 Multivariable analysis of association between innate immune response genes allele frequencies and seroconversion

Association between innate immune response genes allele frequencies in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050]. None

of the alleles were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

#### 3.3.2.5 Multivariable analysis of association between innate immune response genes genotype frequencies and seroconversion

Association between innate immune response genes genotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction and also adjusted for multivariable analysis in gPLINK [v.2.050].

None of the genotypes were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

#### 3.3.2.6 Multivariable analysis of association between innate immune response genes haplotype frequencies and seroconversion

Association between innate immune response genes haplotype frequencies in cases (seroconverters) vs. controls (non–seroconverters) with covariates (infant baseline IgA and IgG titres) was investigated using Chi–square test ($\chi^2$) and the generated P values were adjusted using Bonferroni correction in gPLINK [v.2.050]. No significant association was found (P > 0.05).

None of the haplotypes were significantly associated with seroconversion when the above–mentioned covariates were considered (P > 0.05).

# 4. DISCUSSION AND CONCLUSION

The introduction of WHO pre–approved RV vaccines (RV1 and RV5) into the national childhood immunization programs of many countries across the world has resulted in a significant reduction of RV–related mortalities and hospitalizations of infants (Bucardo et al., 2017). However, efficacy of these vaccines varies markedly between high–income countries and low– and middle–income countries (Jiang et al., 2010). For examples, RV1 and RV5 vaccine trials in various Sub–Saharan Africa countries demonstrated efficacies ranging from 39 to 67% (Armah et al., 2010; Madhi et al., 2010; Isanaka et al., 2017). In contrast, vaccine trials in the US and Finland demonstrated efficacy of 98% (Vesikari et al., 2006), whilst Europe recorded 85% efficacy (Vesikari et al., 2007). Numerous reasons for these varying observations have been proposed. For examples, it has been hypothesized that breastfed–acquired or trans–placentally–acquired antibodies may inhibit immune response to RV vaccines (Becker–Drepps et al., 2015; Chilengi et al., 2016; Groome et al., 2014; Moon et al., 2016). Other proposed reasons for low RV vaccine efficacies in Sub–Saharan Africa include interference of oral polio vaccine (OPV) with RV vaccines, lack of nutrients (e.g., zinc deficiency), co–infections, environmental enteropathy or variations in intestinal microbiome composition, and host genetic variation (Patel et al., 2012; Kandasamy et al., 2014; Naylor et al., 2015; Moon et al., 2016; Taniuchi et al., 2016; Harries et al., 2017).

HBGAs have recently been shown to play an important role in RV vaccine–take (seroconversion) in various populations in a P genotype–dependent manner (Nordgren et al., 2014). A total of 37 RV P genotypes have been identified to date (Agocs et al., 2014). Among these, P[8] and P[4] genotypes are the most common strains worldwide (Agocs et al., 2014), whereas P[6] is mostly common in Sub–Saharan Africa. Importantly, both RV1 and RV5 vaccines are composed of the P[8] genotype strain (Bernstein et al., 1999; Matthijnssens et al., 2010). Previous studies have demonstrated that Se– and Le– individuals might be resistant to RV P[8] and P[4] strains (Imbert–Marcille et al., 2014; Nordgren et al., 2014; Van Trang et al., 2014; Ayouni et al., 2015; Payne et al., 2015; Kambhampati et al., 2016; Sun et al., 2016; Zhang et al., 2016; Lee et al., 2018; Pollock et al., 2018), however Le– individuals have a higher risk of P[6] infection (Nordgren et al., 2014). Since Le– phenotype is highly common in Africa, this might explain the high frequency of P[6] infections in this continent (Santos and Hoshino, 2005; Nordgren et al., 2014).

As previously mentioned, both RV1 and RV5 vaccines contain P[8] strains and thus, it has been proposed that HBGA–mediated resistance to P[8] RVs could lead to resistance to vaccinations and subsequent vaccine failure (Lee et al., 2018). There is a limited number of

studies based on sub–Saharan Africa cohorts investigating the contribution of HBGAs to low efficacy in of RV vaccines in this region (Kazi et al., 2017; Bucardo et al., 2018; Lee et al., 2018; Pollock et al., 2018). Thus, the aim of the current study was to determine whether there was an association between polymorphisms in HBGAs (*ABO*, *FUT2/FUT3)* and ABO, Se, and Le phenotypes in saliva, with immune response to RV1 vaccination in SA children. We also investigated the association between SNPs from innate immune response genes (*DDX58*, *IFIH1*, *TLR3*, *TLR7*, and *TLR8*) with seroconversion after exposure to RV1.

### 4.1 Non–genetic factors influencing seroconversion in our cohort

In our cohort (N= 250, n= 138), we found (95/138= 69%) seroconverters and 43/138= 31%) non–seroconverters/ poor immune response to the vaccine. In contrast, Kazi et al. (2017) reported (56/181= 31%) seroconverters and (125/181= 69%) non–seroconverters in Pakistani cohort of infants. Pollock et al. (2018) reported (47/196= 24%) and (149/196= 76%) non–seroconverters in a Malawian cohort of infants.

We did not observe a significant difference in gender between seroconverters and non–seroconverters (P= 1) in our cohort. In a study by Bucardo et al. (2018), there was no significant difference between males and females in terms of seroconversion among Nicaraguan infants who were vaccinated with RV1 and RV5 (P= 0.60). Payne et al. (2015) also reported no significant difference in gender between AGE patients and healthy controls in a study involving USA infants (White, Black, Hispanic, and non–Hispanic). However, seroconversion was not assessed in that study. These findings suggests that seroconversion (vaccine–take) is not influenced or determined by gender. Indeed, gender has not been previous implicated as one of the factors contributing to RV1 and RV5 varying efficacy in different ethnic groups across the world. Therefore, RV infects hosts regardless of their gender.

All children in this cohort were Black South Africans and of the same age at vaccination and at sampling, therefore ethnicity and age were not variables affecting antibody responses in this cohort.

However, there was a significant difference between seroconverters and non–seroconverters in terms of the IgA at six weeks (pre–vaccination) (P= 0.0001). High pre–vaccine IgA titre in non–seroconverters suggest that the infants were infected with RV at some point after birth. This is supported by Gurwith et al. (1981) and Velázquez et al. (1996) who reported that virtually all children below five years of age are infected with RV at least once regardless of

whether they come from high– or low–income setting. Chilengi et al. (2016) also reported a lower seroconversion possibility for infants seropositive pre–RV vaccination compared to those who were seronegative pre–vaccination (P= 0.004) in study involving Zimbabwean mother–child paired cohort. In the same study, Chilengi et al. (2016) reported an association between higher RV–specific IgA in breastmilk with a reduced seroconversion frequency after RV vaccination (P< 0.01).

A study by Berker–Drepps et al. (2018) involving Nicaraguan infants exposed to RV5 vaccine also reported a significant difference in RV–specific IgA seroconversion between pre–vaccination seropositive and seronegative infants (P= 0.005). Furthermore, infants who were non–Secretors had mothers with higher serum IgG GMT compared to mothers of infants who were Secretors (P= 0.02). However, there was no significant difference in breastmilk IgA GMT for moms of seroconverters vs. non–seroconverters (P= 0.80) (Berker–Dreps et al., 2018). Again, these observations provide more evidence that pre–vaccination RV–specific IgA and IgG GMT may negatively affect seroconversion after exposure to live oral RV vaccines.

However, Bucardo et al. (2018) reported contradicting results. In that study, there was no significant difference in seroconversion between seropositive and seronegative groups in terms of pre–vaccination RV–specific IgA GMT (P= 0.26). However, this study did not include IgG GMT readings. The authors concluded that pre–vaccination IgA GMT does not have a significant influence on seroconversion post exposure to both RV1 and RV5.

Asymptomatic maternal wildtype RV infections which are usually not accounted for might also be a contributing factor. Chilengi and colleagues carried out an analysis to investigate the link between seasonally and immune response as factors inducing RV vaccine failure. They discovered that dry cold seasons were significantly associated with higher IgA GMT in maternal breastmilk (Chilengi et al., 2016). The contradicting results between our study and the Bucardo et al. (2018) study might be due variation in the seasonality of RV infections in the moms, as well as differences in cohort sizes (n= 60 for Bucardo et al. (2018) vs. n= 215 for our study).

### 4.2 Genetic variation in the current study cohort

We identified a total of 19 SNPs from the *FUT2* Sanger sequencing results of our cohort saliva samples. *FUT2* SNPs such as rs679574, rs516316, rs516246, rs485073, and rs603985 were not observed in the European cohort as compared the SA Xhosa, LWK, and sub–Saharan Africa cohorts (Koda et al., 2001). This suggests that some *FUT2* SNPs are specific to certain ethnic groups. Ferrer and colleagues postulated that natural section is a force governing phylogenetic variation and geographical distribution of *FUT2* variants (Ferrer et al., 2009; Liu et al., 1998; Peng 1999). There were no novel *FUT2* SNPs identified in our current study cohort. This may be attributed to the small cohort size (n= 20). However, available SNP data allowed for subsequent genotyping of the SNPs of interest in our cohort.

We identified 16 SNPs from the *FUT3* Sanger sequencing results. These SNPs were a similar set reported by Peng et al. (1999) in SA Xhosa cohort. Furthermore, we observed an unusual set of mutations seen with primer set 3 at position 589, 601, 602, 606, 607, 612, 645, 691, 696. These mutations are very rare in published data (e.g., 1000Genomes, Gnomad, Pang Xhosa paper). In addition, 601, 602, 607 are novel in this study – mutations at these sites have never been reported before and thus, these data need to be verified. In MassArray genotyping we included 606, 612, and 667 but 606 was invariant and 612 failed typing. Again, this suggests either possible novel variation in this region or sequencing assay problems.

Majority of MAF of *FUT2* and *FUT3* SNPs were significantly different between the current study cohort, LWK and YRI cohorts. This may be attributed to variation in cohort sizes used during genotyping. Similarly, majority of MAF of *ABO* SNPs were also significantly different between the current study cohort, LWK and YRI cohorts. This may be attributed to variation in cohort sizes used during genotyping. Lastly, majority of MAF of innate immune response genes SNPs were also significantly different between the current study cohort, LWK and YRI cohorts. This may be attributed to variation in cohort sizes used during genotyping.

### 4.3 Do *FUT2* variation and Secretor phenotypes associate with seroconversion to RV1 in our current study cohort?

For the Se first time–point ELISA, we recorded 43% Secretors, 32% weak Secretors, and 25% non–Secretors. For the Se second time–point ELISA, we recorded 46% Secretors, 28% weak Secretors, and 27% non–Secretors in the current study. These Se phenotype frequencies were not different between the two time–points.

Predicted Se frequencies from genetic data in our cohort were as follows (n= 138): 70% for Se+ phenotype, 14% for weak Se phenotype, and 16% Se– phenotype. In comparison to other African data: McDonald et al. (2020) recorded 74% Se+ individuals and 26% S– individuals in a study involving SA infants (n= 250) below five years of age who were hospitalized for diarrhoea. Similarly, Moores and Brain (1968) recorded 73% Se+ individuals and 27% Se– individuals in a Bantu cohort (n= 171) from SA. Pollock et al. (2018) reported the following Se phenotypes: 91% for Se+ and 9% for Se– phenotype in a study involving a cohort of Malawian infants (n= 293) who were infected by P[8] and P[4], and P[6] RV strains. In the same study, there were 28% of Se– individuals and 72% of Se+ individuals in the community controls (Pollock et al., 2018). Armah et al. (2019) reported the following Se phenotypes: 81% for Se+ and 19% for Se– in study involving healthy Ghanaian infants (n= 166) who were subsequently exposed to RV1 vaccine. Lastly, Nordgren et al. (2014) reported the following Se phenotypes: 79% for Se+ and 21% for Se– in a study involving Burkina Faso infants (n= 208) who were suffering from diarrhoea. Our results were similar to the Se frequencies of (range= 9–58) noted above in other studies of African populations.

In non–African populations, variable Se frequencies are found. Lee et al. (2018) reported 68% Se+ and 32% Se– phenotypes in Bangladeshi cohort of infants. Zhang et al. (2016) reported 100% Se+ and 0% Se– phenotypes in a study involving RV–infected Chinese infants. Lastly, Norgdren et al. (2014) reported 93% Se+ and 7% Se– phenotypes in a study involving Nicaraguan infants. The Se phenotype frequency in our study is relatively low compared to other studies, therefore Se phenotype might be a factor for RV vaccine response in our SA cohort.

In the current study, percentage concordance between *FUT2* genetic and ELISA data was 83%. The reason for not getting 100% concordance may be attributed to lack of sensitivity of our ELISA assay. ELISA is a less reliable method compared to Sanger sequencing (Norgren et al., 2014). This is because ELISA can be easily contaminated (e.g., maternal IgG antibodies) which

could lead to cross reactions and false results. In addition, primary/secondary antibodies used in ELISA might lack target specificity, which could give false positives. Pollock et al. (2018) reported 90% genetic/phenotypic concordance, whilst Imbert–Marcille et al. (2014) and Bucardo et al. (2018) both reported 100% genetic/phenotypic concordance. Lastly, Armah et al. (2019) reported 87% (73/84) genetic/phenotypic concordance.

Regarding effect of *FUT2* on seroconversion to RV1 vaccine, we did not observe any significant difference between Secretors vs. non–Secretors and seroconverters vs. non–seroconverters (P= 0.5184) and (P= 0.2582) in the second time–point ELISA. Similarly, genetic variation in *FUT2* did not significantly associate with seroconversion P> 0.05). Therefore, there is not enough evidence to support our hypothesis that genetic variation in *FUT2* may affected Le B expression ultimately resulting in low seroconversion after exposure to RV1. Lee et al. (2018) also reported similar results to ours. In that study, Se phenotype did not have an effect on RV–induced AGE in vaccinated infants (P= 0.35). When other variables (e.g., IgA seroconversion, breastfeeding, and serum zinc concentration) were included in a multivariable analysis models, Se phenotype was still not significant in vaccinated infants (P= 0.5). There was no significant difference in vaccine failure according to the Se phenotype of the participants (P > 0.5) (Lee et al., 2018). Pollock et al. (2018) also reported similar results to ours. In that study, there was no significant difference in RV–specific IgA GMT stratified according to the Se (P= 0.52) of the participants in a study involving Malawian infants (**Table 4.1**). No significant difference was observed in terms of seroconversion between Secretors and non–Secretors (P= 0.08).

In contrast to our results, Armah et al. (2019) reported a significant difference in seroconversion between Se+ and Se– individuals in a Ghanaian cohort of infants (**Table 4.1**). According to the authors, 41% Se+ individuals seroconverted compared to 13% Se– individuals (P= 0.016). Bucardo et al. (2018) also reported an association between Se/Le phenotypes and seroconversion after exposure to RV1 and RV5 vaccines in a study based on a cohort of Nicaraguan infants (**Table 4.1**). A significant difference was also observed in IgA seroconversion between Secretors and non–Secretors in the RV1 cohort (P < 0.001) in Bucardo et al. (2018). Gunaydin et al. (2016) investigated the association between HBGA phenotypes and seroconversion in terms of both IgA and IgG titres in a Swedish cohort of infants. RV–specific IgG titres in Secretors were reported to be significantly higher in non–Secretors, independent of IgA titres (P< 0.001). A study by Kazi et al. (2017) based on Pakistani cohort of infants reported an association between seroconversion and Se status (P= 0.002) (**Table 4.1**).

**Table 4.1** Various studies investigating the association between seroconversion (vaccine–take) and HBGAs. Adapted from Sharma et al. (2020).

| Vaccine | Country | ABO | Lewis | Secretor Status | Method | | Reference |
|---|---|---|---|---|---|---|---|
| | | | | | Time Point | Measurement | |
| **Rotarix** | Nicaragua | B less seroconversion | Lewis A no seroconversion | Non–secretors less seroconversion | After 1 dose | Seroconversion | Bucardo et al. 2018 |
| | Pakistan | Non–O less seroconversion compared to O | No association | Non–secretors less seroconversion | After 3 doses | | Kazi et a., 2017 |
| | Ghana | O less seroconversion compared to B | No association | Non–secretors less seroconversion | After 2–3 doses | | Armah et al., 2019 |
| | Malawi | No association | No association | Non–secretors less seroconversion | After 2 doses | | Pollock et al., 2018 |
| **RotaTeq** | Nicaragua | A had most seroconversion | Lewis A no seroconversion, low power | No association | After 1 dose | | Bucardo et al. 2018 |
| **RV3–BB** | New Zealand | Not investigated | No association, low power | No association | Cumulative | | Boniface et al., 2019 |
| | | | | | | | |
| **Rotarix** | Nicaragua | B no shedding, low power | Lewis A no shedding, low power | Non–secretors no shedding | After 1 dose | Vaccine shedding | Bucardo et al., 2019 |
| | Malawi | No association | No association | Non–secretors less shedding | After 1 dose | | Pollock et al., 2018 |
| | Malawi | No association | No association | No association | After 2 doses | | Pollock et al., 2018 |
| **RotaTeq** | Nicaragua | No association | Lewis A no shedding, low power | No association | After 1 dose | | Bucardo et al., 2019 |
| **RV3–BB** | New Zealand | Not investigated | No association, low power | No association | Cumulative | | Boniface et al., 2019 |

### 4.4 Do *FUT3* variation and Lewis phenotypes associate with seroconversion to RV1 in a SA cohort?

For the Le first time–point ELISA, we recorded 88% Le+ and 12% Le– phenotypes. For the Le second time–point ELISA, we recorded 77% Le+ and 23% Le– in the current study. These Le phenotype frequencies were not different between the two time–points, thus no changes in the same individual over time.

Predicted Le frequencies from genetic data in our cohort were as follows (n= 138): 80% for Le+ phenotype and 20% Le– phenotype. In comparison to other African data: Moores and Brain (1968) recorded 77% Le+ individuals and 23% Le– individuals in a SA Bantu speaking cohort (n= 171) from SA. Pollock et al. (2018) reported 26% of Le– individuals and 74% of Le+ individuals in Malawian controls (Pollock et al., 2018). Armah et al. (2019) reported the following Le phenotypes: 57% for Le+ and 43% for Le– in study involving healthy Ghanaian infants (n= 166), Nordgren et al. (2014) reported the following Le phenotypes: 68% for Le+ and 32% for Le– in a study involving Burkina Faso infants (n= 208) who were suffering from Diarrhoea. Our results were within the range reported for other African cohorts.

In comparison to other non–African populations, Indians have been reported to have 82% Le+ and 18% Le– phenotypes, respectively (Owaidah et al., 2020). Chinese have been reported to have 77% Le+ and 23% Le– phenotypes (Owaidah et al., 2020).

In the current study, percentage concordance between *FUT3* genetic and ELISA data was 75%. This low percentage concordance may be attributed to lack of sensitivity of our ELISA assay. Furthermore, we only selected SNPs typed in exon 3. Our ELISA assay can be improved by systemically changing concentration of primary and secondary antibodies or by changing the format from indirect to sandwich ELISA. We can also systemically increase variables such as incubation period and temperature to improve sensitivity of our assay. In comparison, Bucardo et al. (2018) reported a 62% (8/13) genetic/phenotypic concordance for Le status. HBGA ELISA phenotyping results might still be used to predict genotyping results or vice versa. Most researchers genotype a fraction of their cohort to confirm their ELISA results. According to Kazi et al., 2017 genotyping is not superior to HBGA phenotyping. This is because some genetic Secretors express the H antigen modestly to a point where their phenotype and epidemiologic risk can be the same as that of non–Secretors (Dotz and Wuhrer, 2016).

No significant association between Le phenotype and seroconversion status was found using ELISA in saliva samples. We did not observe significant difference between Le A–B+ and Le A+B– phenotypes in seroconverters vs. non–seroconverters (P= 0.2563). Therefore, based on ELISA phenotyping, we did not find substantial evidence to conclude that variation in Se/Le phenotypes contribute to RV–specific IgA seroconversion (vaccine–take) in this cohort after exposure to two doses of RV1.

When univariate analyses were performed, SNP rs761121538 in *FUT3* were significantly associated with seroconversion, (P= 0.02 and OR= 0.070). However, other *FUT3* SNPs with similar functional effects were not significantly associated with seroconversion, leading to the question of why only this particular SNP could have been association with seroconversion. This could have just been a false positive – since after Bonferroni correction, this SNPs did not remain significant. In addition, multivariable analysis were not significant. Overall, these finding suggest that Le phenotype did not associate with seroconversion in this cohort. Lee et al, (2018) also reported there was no significant difference in vaccine failure according to the Le status of the Bangladeshi participants (P > 0.5). Lee et al. (2018) also reported that Le phenotype did not affect RV–induced AGE in both vaccinated and unvaccinated Bangladeshi infants. Pollock et al., (2018) reported similar results in a study involving Malawian infants (**Table 4.1**). where no significant difference was observed in terms of seroconversion between Le+ and Le– individuals (P= 0.77). Armah et al. (2019) reported that seroconversion was not significantly different between Le+ and Le– individuals in a study involving Ghanaian infants (Table 4.1). Bucardo et al. (2018) reported that Nicaraguan children with Le A phenotype did not seroconvert after exposure to RV1 or RV5 vaccines as compared to Le B individuals (**Table 4.1**).

Contrastingly, when combined *FUT2* and *FUT3* genotype data were used to predict Le phenotypes, we did find a significant association between predicted Le phenotype and seroconversion (P= 0.026). Epistasis analysis also suggested that the interaction between *FUT2* non–Secretor haplotype and *FUT3* might be significant (P= 0.047, OR= 13.06 and P= 0.048, OR= 12.74) although it became non–significant after Bonferroni correction. These findings support our hypothesis that genetic variations in *FUT2* and *FUT3* genes may affect Le B expression leading to the reduction of seroconversion after exposure to RV1. However, since the P value became non–significant after Bonferroni correction, we believe that our findings need to be verified. Nordgren et al. (2014) reported that P[8] strain only infected Se+ (Le+) individuals, whereas P[6] strain only infected Le– (Le A) individuals in Nicaragua – proving

that indeed Le and Se status does affect seroconversion. In another study, Nordgren et al. (2018) observed no seroconversion in Nicaraguan children with Le A phenotype, suggesting that Le A individuals have less robust immune response towards RV1 compared to individuals with Le B phenotype.

We observed contrasting signals from phenotypes predicted from genetic data (significant association with seroconversion) vs. Le ELISA data (not significantly associated with seroconversion). This contrast might be related to the low percentage concordance (75%) between genetic and Le ELISA phenotypes. These findings suggests that there might still be a role for Le phenotypes (defined by a combination of FUT2 and FUT activity) in seroconversion after RV1 vaccination. Importantly, the direction of effect that we observed in Le phenotypes associated with seroconversion was somewhat different from that reported in the literature. We found that vaccine non–responders had significantly higher frequency of Le antigen (presence) and a lower frequency of Le– phenotype (P= 0.039). We also observed significant difference in distribution of Le phenotypes, which included both a relative increase in Le A phenotype in non–responders and a relative increase in Le (a–b–) in vaccine responders (P= 0.026). Unlike literature (Nordgren et al., 2018) where Le A alone was significantly associated with seroconversion, in this study Le A alone was not significantly associated with seroconversion (P= 0.07). Instead, in our study the overall distribution (different ratios of all the Le phenotypes) was implicated. It is not clear how Le– phenotype could cause an increased vaccine response as RV strain P[8] is thought to bind to the Le B antigen.

### 4.5 Is there an association between ABO phenotypes or genotypes in our current study cohort with seroconversion?

ABO phenotypic distribution in Secretors measured by ELISA was no different between time–point 1 vs time–point 2 in the current study, thus no changes in the same individual over time.

ABO blood types were similar between time–point 2 ELISA data and genetic data, and similar to published SA data. Blood type O was the most common phenotype among the ABO blood types in both the first time–point (67%) and second time–point ELISA (65%) in our cohort. As seen globally, type O was most common and AB least common (Armah et al., 2019; Bucardo et al., 2018; Kazi et al., 2018). Variations in ABO distribution globally may be attributed to different cohort ethnicities in various studies. However, these observations suggest that blood type O is the most common blood phenotype in most regions of the world.

We did not observe any significant difference between ABO blood types in seroconverters vs. non–seroconverters (P= 0.3781) and Secretors vs. non–Secretors (P= 0.3112) in the first time–point ELISA of our study. Seroconversion after the P value was adjusted for multivariate analysis ($P_{EMP2}$= 0.6276). None of the *ABO* SNPs had significant $P_{EMP1}$ and $P_{EMP2}$ values. Thus, it can be concluded that genetic variation in *ABO* might not be associated with immune response (seroconversion) after exposure to RV1 in our current study cohort. Therefore, these findings do not support our hypothesis that genetic variation in the *ABO* gene might be associated with low seroconversion after exposure to RV1. This might be attributed to the small cohort size in this study which was a limiting factor because of less genetic diversity. In Ghana, individuals with blood type O were reported to seroconvert less compared to those with blood type B (Armah et al., 2019). Kazi et al. (2017) also reported that blood type O individuals seroconverted more when compared to non–blood type O individuals in a Pakistan study (P= 0.002). Bucardo et al. (2018) reported that seroconversion rate was significantly lower in children with blood type B compared to those with blood type O and A in Nicaragua (opposite results to Kazi et al., 2018).

### 4.6 Is there an association between genetic variation in innate immune receptors in our current study cohort and seroconversion?

We studied 19 SNPs in five genes influencing innate immune response to RV in the current study. Only the rs2159377 (*TLR8*) SNP was significantly associated with seroconversion after univariate allele frequency analysis (P= 0.0146; $P_{EMP1}$= 0.01954). Pertaining OR value of the rs2159377 SNP, the results suggest that the minor allele from this SNP increase the likelihood of seroconversion relative to the major alleles since its OR value was greater than one. However, this SNP did not remain significantly associated with seroconversion after the P value was adjusted for multiple testing or in multivariable analysis ($P_{EMP2}$= 0.6276). The six SNP haplotype (rs179008, rs864058, rs3853839, rs179007, rs5935438, rs2159377) found on the X chromosome was significantly associated with seroconversion in the current study (P= 0.042). Overall, these results suggest that the innate variation in the current study is not associated with seroconversion.

Although innate immune response SNPs in the current study have not been associated with seroconversions after RV vaccination, SNPs such rs55789327 and rs669260 from *DDX58* have been reported to be associated with antiviral immune response (Pothlichet et al., 2009) and with response to rubella and measles vaccination (Ovsyannikova et al., 2014), respectively.

## 4.7 Strengths and limitations of the current study

Our study has several strengths and limitations as discussed in detail below.

### 4.7.1 Strengths

To the best of our knowledge, this is a first SA based study investigating the association between variation in HBGA phenotypes/genotypes and RV vaccine efficacy. The study participants were all of Black SA ethnicity and matched for age, which removed sources of variation from the study. It is also a unique study that has started the process of investigating the role of candidate genes other than HGBA, by considering innate genetic variation. We have used both genetic data and ELISA phenotyping to study the role of HBGA in response to RV vaccination, whereas many studies in this field of research are focused on HBGA phenotyping using ELISA or sequencing only, but not both. We also addressed the interaction (epistasis) between genes which subsequently had an effect on response to RV1 vaccine – this is novel and important.

### 4.7.2 Limitations

The largest limitation was the small cohort size; our current study cohort consisted of 215 participants for the first time–point ELISA and 189 for the second time–point ELISA because not all participants reported for the second round of samples collection (i.e., participant drop out). These cohort sizes may not be adequate to draw any concrete conclusions regarding the general population of SA in terms of RV1 efficacy, which is a limiting power. For the MassArray analysis, case–control studies are much better powered to detect associations if larger cohorts are used.

Only 20 cohort samples were genotypically analysed for the *FUT2* and *FUT3* SNPs due to poor DNA quality during extraction from saliva and also financial constraints. This a limiting power to detect most of the known alleles or novel alleles that have never been reported before in SA cohorts. the novel variant discovery part of this project only interrogated two genes of interest (*FUT2* and *FUT3*) and only one exon from each gene. A genome–wide association studies (GWAS), or at least a more expansive candidate gene association study might well have yielded more significant results. Epistatic interaction of HBGA genes with innate genes was not carried out in this study. However, this would be a theoretical analysis only as there is no evidence that the products of these genes directly interact with each other.

Another limitation was the use of ELISA and saliva samples for HBGA phenotyping in this study. ELISA assays can experience possible cross reactions from maternal antibodies and non–specificity (lack of sensitivity) of the immunoglobulins used as previously mentioned and also possible contamination of saliva with maternal breastmilk. Thus, some study participants may have been misclassified, however genotyping data was used to confirm some of the phenotyping results. The contradiction between our ELISA and genotyping results also indicate the disadvantage of solely relying on ELISA for phenotyping.

Saliva samples of the participants were not matched with stool samples. Thus, we could not determine RV strains (if any) from each study participant. By including stool samples, we were going to be able to supplement the existing knowledge in terms of how various RV strains infect individuals based on their HBGA phenotypes and *FUT2* and *FUT3* genotypes from the SA perspective. Seroconversion in some of the study participants may be due to wild–type infection and not immune response due to RV1 vaccine exposure.

Our study participants were only exposed to live–attenuated RV1 oral vaccine, thus our deductions cannot be generalized to other currently available RV vaccines. Lastly, we did not include additional factors during multivariate analysis such as maternal antibodies concentration and co–infection data. These factors are important because they have previously been implicated in the observed disparity in immune response of various ethnic groups after exposure to live–attenuated oral RV vaccines (Nordgren et al., 2014).

### 4.8 Conclusion

In summary, we demonstrated in a cohort of SA infants that there is no strong association between variation in Se, Le, ABO phenotypes and genes affecting innate immune response with development of IgA antibodies after RV vaccination when these factors were considered individually. Since rs761121538 (Allelic) (*FUT3*), rs2159377 (Allelic) (*TLR8*) and single *FUT3* and X chromosome *TLR7/8* haplotypes had significant $P_{EMP1}$ and $P_{EMP2}$ values but non–significant P values after Bonferroni correction, it can be concluded that these variants were unlikely to contribute significantly towards seroconversion in this study in a non–epistatic fashion.

However, our data using *FUT2* and *FUT3* combined genetic data to predict Le phenotypes showed significant association with seroconversion. Epistasis may play an important role in predicting vaccine response. Importantly, these results supplement and support phenotypic

literature results that implicated variation in Le and Se phenotypes as factors responsible for the discrepancy of RV1 efficacy observed in Africa and Asia (Nordgren et al., 2014 and Armah et al., 2019). This is a similar result to previous studies which demonstrated that RV strains infect individuals in a P genotype–depended manner (Imbert–Marcille et al., 2014; Nordgren et al., 2014, Van Trang et al., 2014; Ayouni et al., 2015; Payne et al., 2015; Kambhampati et al., 2016; Zhang et al., 2016; Elnady et al., 2017; Kazi et al., 2017; Armah et al., 2019; Bucardo et al., 2018; Lee et al., 2018; Pollock et al., 2018; Pérez–Ortín et al., 2019) and thus, Se/Le phenotypes should be considered during live oral RV vaccine trials in the future.

Seroconversion was also affected by pre–vaccination antibodies which may have been due to early infection of the infants with RV. Other factors influencing outcomes of RV vaccination cannot be excluded such as enteric dysfunction, gut microbiota, nutritional status, and concomitant infections (Sharma et al., 2015; Steele et al., 2019).

### 4.9 Implications for future vaccine design and research

It was useful to use both genotyping techniques and phenotyping of the ABO, Le, and Se antigens. Results from the current study need to be replicated in another larger SA cohort. While candidate gene studies are interesting, GWAS should be used as it has not been done before for either RV infection outcomes or vaccination outcomes.

Future research as measure of expanding the current study could include other factors that have been implicated in the reported low efficacy of RV vaccines, particularly in Africa and Asia. For example, factors such as maternal breastmilk as well as maternal HBGA status of each of the infants should also be investigated. In fact, we already have stored maternal saliva samples that were matched with those of the infants ready to be phenotyped. Future investigations should also include stool sample so that RV genotypes can also be determined especially in the context of SA because that has never been done before. Lastly, the exact role of SNPs in *FUT2*, *FUT3*, and *TLR8* towards immune response (seroconversion) in individuals exposed to RV1 needs to be further investigated. Functional immune responses such as interferon production or measures of other immune intermediates could be assessed.

RV–induced AGE is still a major health burden in low–income countries such as those in Sub–Saharan Africa and Asia. This emphasises the need for development of new effective and affordable alternative RV vaccines. Host genetic variation continues to be an important factor that affects vaccine efficacy and should be taken into consideration during future vaccine

design. Advancements in this regard promise a better future pertaining the fight against RV infections, globally.

# 5. REFERENCES

Agena Bioscience (2016). iPLEX®: Pro reagents user guide. United States. Source: http://agenabio.com/resources/

Agocs, M.M., et al. (2014). WHO global rotavirus surveillance network: a strategic review of the first 5 years, 2008–2012. *MMWR Morb Mortal Wkly Rep*, **63**: 634–637.

Akhtar, T., Cargill, J., Gerrard, C., Shaw, F., Cunliffe, N.A., Cooke, R.D., and Pizer, B (2018). Detection of rotavirus in paediatric oncology patients with diarrhoea: the impact of rotavirus vaccine. *J Hosp Infect*, 99(**2**): 185–187.

Ameno, S., et al. (2001). Lewis and Secretor gene effects on Lewis antigen and postnatal development of Lewis blood type. *Biol Neonate*, **79**: 91–96.

Angel, J., Franco, M.A., and Greenberg, H.B (2012). Rotavirus immune responses and correlates of protection. *Curr. Opin. Virol*, **2**: 419–425.

Armah, G.E., Cortese, M.M., Dennis, F.E., Yu, Y., Morrow, A.L., et al. (2019). Rotavirus vaccine take in infants is associated with secretor status. *The Journal of Infectious Diseases*, **219**: 746–9.

Armah, G.E., Sow, S.O., Breiman, R.F., Dallas, M.J, Tapia, M.D., Feikin, D.R., et al. (2010). Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub–Saharan Africa: a randomised, double–blind, placebo–controlled trial. *Lancet*, **376**: 606–14.

Ayouni, S., Sdiri–Loulizi, K., de Rougemont, A., Estienney, M., Ambert–Balay, K., Aho, S., et al. (2015). Rotavirus P[8] Infections in Persons with Secretor and Nonsecretor Phenotypes, Tunisia. *Emerging Infectious Diseases*, 21 (**11**): 2055–2058.

Bahl, R., Ray, P., Subodh, S., Shambharkar, P., Saxena, M., Parashar, U., et al. (2005) Incidence of severe rotavirus diarrhea in New Delhi, India, and G and P types of the infecting rotavirus strains. *J Infect Dis*, **192**: 114–9.

Bailey, J.N., Kinzy, T.G., and Igo, R.P (2019). Genetic risk scores. *Current Protocols in Human Genetics*, **104**: e95.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*.

Bernstein, D.I., et al. (1999). Efficacy of live, attenuated, human rotavirus vaccine 89–12 in infants: a randomised placebo–controlled trial. *Lancet*, **354**: 287–290.

Besch, R., Poeck, H., Hohenauer, T., Senft, D., Hacker, G., et al. (2009) Proapoptotic signaling induced by RIG–I and MDA–5 results in type I interferon independent apoptosis in human melanoma cells. *J Clin Invest*, **119**: 2399–2411.

Bhandari, N., Rongsen–Chandola, T., Bavdekar, A., et al. (2014) (A). India Rotavirus Vaccine Group. Efficacy of a monovalent human–bovine (116E) rotavirus vaccine in Indian infants: a randomised, double–blind, placebo–controlled trial. *Lancet*, **383**: 2136–2143.

Bhandari, N., Rongsen–Chandola, T., Bavdekar, A., et al. (2014) (B). India Rotavirus Vaccine Group. Efficacy of a monovalent human–bovine (116E) rotavirus vaccine in Indian children in the second year of life. *Vaccine*, 32(**Suppl 1**): A110–A116.

Bines, J.E, At Thobari, J., Satria, C.D., et al. (2018). Human neonatal rotavirus vaccine(RV3–BB) to target rotavirus from birth. *N Engl J Med*, **378**: 719–730.

Bishop, R.F (1996). Natural history of human rotavirus infection. *Arch Virol Suppl*, **12**:119–28.

Black, R.E., Cousens, S., Johnson, H.L., et al. (2010). Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet*, **375**: 1969–87.

Blutt, S.E and Conner, M.E (2007). Rotavirus: to the gut and beyond! *Curr Opin Gastroenterol*, **23**: 39–43.

Blutt, S.E, Kirkwood, C.D., Parreno, V., Warfield, K.L., Ciarlet, M., Estes, M.K., et al. (2003). Rotavirus antigenaemia and viraemia: a common event? *Lancet*, **362**: 1445–9.

Brain, P (1966). Subgroups of A in the South African Bantu, *Vox Sanguinis* 11(**6**): 686–698.

Broquet, A.H., Hirata, Y.., McAllister, C.S., and Kagnoff, M.F (2011). RIG–I/MDA5/MAVS are required to signal a protective IFN response in rotavirus–infected intestinal epithelium. J Immunol, **186**: 1618–1626.

Bruns, A.M., Leser, G.P., Lamb, R.A., Horvath, C.M (2014). The innate immune sensor LGP2 activates antiviral signaling by regulating MDA5–RNA interaction and filament assembly. *Mol Cell*, **55**: 771–781.

Bucardo, F., Nordgren, J., Reyes, Y., Gonzalez, F., Sharma, S., and Svensson, L (2018). The Lewis A phenotype is a restriction factor for Rotateq and Rotarix vaccine–take in Nicaraguan children. *Scientific Reports*, **8**: 1502.

Burke, R.M., Tate, J.E., Kirkwood, C.D., Steele, A.D., and Parashar, U.D (2019). Current and new rotavirus vaccines. *Curr Opin Infect Dis*, **32**: 435–444.

Burnett, E., Jonesteller, C.L., Tate, J.E., Yen, C., and Parashar, U.D (2017). Global Impact of Rotavirus Vaccination on Childhood Hospitalizations and Mortality from Diarrhea. *J Infect Dis*, 215(**11**): 1666–1672.

Cantelli, C.P., Velloso, A.J., Santos de Assis, R.M., et al. (2020). Rotavirus A shedding and HBGA host genetic susceptibility in a birth community–cohort, Rio de Janeiro, Brazil, 2014–2018. *Scientific Reports*, 10: 6965.

Casrouge, A., Zhang, S. Y., Eidenschenk, C., Jouanguy, E., Puel, A., Yang, K., Alcais, A., Picard, C., Mahfoufi, N., et al. (2006). Herpes simplex virus encephalitis in human UNC–93B deficiency. *Science*, **314**: 308–312.

Chami, F.A. "Graeco–Roman Trade Link and the Bantu Migration Theory." Anthropos, vol. 94, no. 1/3, 1999, pp. 205–215., www.jstor.org/stable/40465703. [Accessed 19 Apr. 2021].

Chandola, T.R., Taneja, S., Goyal, N., et al. (2017). ROTAVAC(1) does not interfere with the immune response to childhood vaccines in Indian infants: a randomized placebo–controlled trial. *Heliyon*, **3**: e00302.

Chandrasekaran E.V., Chawda, R., Locke, R.D., Piskorz, C.F., and Matta, K.L (2002). Biosynthesis of the carbohydrate antigenic determinants, Globo H, blood group H, and Lewis b: a role for prostate cancer cell alpha1,2–L–fucosyltransferase. *Glycobiology*, 12(**3**): 153–62.

Chandrasekaran, E.V, Jain, R.K., Rhodes, J.M., Srnka, C.A., Larsen, R.D., and Matta K.L (1995). Expression of blood group Lewis b determinant from Lewis a: association of this novel alpha (1,2)–L–fucosylating activity with the Lewis type alpha (1,3/4)–L–fucosyltransferase. *Biochemistry*, 34(**14**): 4748–56.

Chang C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J (2015) Second–generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4. Package : PLINK [v2.0]. Authors: Shaun Purcell, Christopher Chang. URL: www.cog–genomics.org/plink/2.0/

Chen, N., Xia, P., Li, S., Zhang, T., Wang, T.T., and Zhu, J (2017). RNA Sensors of the Innate Immune System and Their Detection of Pathogens. *IUBMB Life*, 69 (**5**): 297–304.

Coldiron, M.E., Guindo, O., Makarimi, R., et al. (2018). Safety of a heat–stable rotavirus vaccine among children in Niger: Data from a phase 3, randomized, double–blind, placebo–controlled trial. *Vaccine*, **36**: 3674–3680.

Corvelo, T.O., Aguiar, D.F., and Sagica, F.S (2002). The expression of ABH and Lewis antigens in Brazilian semi–isolated black communities. *Genet Mol Biol*, **25**: 259–63.

Cozzi, G.D., Levinson, R.T., Toole, H., Snyder, M.R., Deng, A., Crispens, M.A., Khabele, D., and Beeghly–Fadiel, A (2017). Blood type, ABO genetic variants, and ovarian cancer survival. *PLoS One*, 12(**4**).

Cunliffe, N.A., Gondwe, J.S., Kirkwood, C.D., Graham, S.M., Nhlane, N.M, Thindwa, B.D., et al. (2001). Effect of concomitant HIV infection on presentation and outcome of rotavirus gastroenteritis in Malawian children. *Lancet*, **358**: 550–5.

Cunliffe, N.A., Kilgore, P.E., Bresee, J.S, Steele, A.D., Luo, N., Hart, C.A., et al. (1998). Epidemiology of rotavirus diarrhoea in Africa: a review to assess the need for rotavirus immunization. *Bull World Health Organ*, **76**: 525–37.

Dang, D.A., Nguyen, V.T., Vu, D.T., et al. (2012). A dose–escalation safety and immunogenicity study of a new live attenuated human rotavirus vaccine (Rotavin–M1) in Vietnamese children. *Vaccine*, 30(**Suppl 1**): A114–A121.

Daniels, G (2005). The molecular genetics of blood group polymorphism. *Transplant Immunology*, 14(**3**): 143–153.

Dean, L (2005a). Chapter 5: The ABO blood group, in: blood groups and red cell antigens. National Center for Biotechnology Information (US), Bethesda.

Dennehy, P.H (2008). Rotavirus Vaccines: An Overview. *Clinical Microbiology Reviews*, 21(**1**): 198–208.

Desai, S., Rathi, N., Kawade, A., et al. (2018). Noninterference of Bovine–Human reassortant pentavalent rotavirus vaccine ROTASIIL1 with the immunogenicity of infant vaccines in comparison with a licensed rotavirus vaccine. *Vaccine*, **36**: 5519–5523.

Desselberger, U (2017). Differences of rotavirus vaccine effectiveness by country: likely causes and contributing factors. *Pathogens*, **6**: 65.

Dickey, M., Jamison, L., Michaud, L., Care, M., Bernstein, D.I., and Staat, M.A (2009). Rotavirus meningoencephalitis in a previously healthy child and a review of the literature. *Pediatr Infect Dis J*, **28**: 318–21.

Dotz, V and Wuhrer, M (2016). Histo–blood group glycans in the context of personalized medicine. *Biochim Biophys Acta*, 1860: 1596–607.

Dou, Y., Yim, C.H., Kirkwood, C.D., Williams, B.R., and Sadler, A.J (2017). The innate immune receptor MDA5 limits rotavirus infection but promotes cell death and pancreatic inflammation. *The EMBO Journal*, **36**: 2742–2757.

Elnady, H.G., Abdel Samie, O.M., Tawhid Saleh, M., Sherif, L.S., Abdalmoneam, N., et al. (2017). ABO blood grouping in Egyptian children with rotavirus gastroenteritis. *Gastroenterology Rev*, 12(**3**): 175–180.

Estes, M.K and Kapikian, A.Z (2007). Rotaviruses. In: Knipe, D.M, Howley, P.M, Eds. Field's Virology. Philadelphia, Pennsylvania: Lippincott, Willliams & Wilkins.

Estes, M.K and Morris, A.P (1999). A viral enterotoxin. A new mechanism of virus–induced pathogenesis. *Adv. Exp. Med. Biol*, **473**: 73–82.

Feng, N., Kim, B., Fenaux, M., Nguyen, H., Vo, P., Omary, M.B., Greenberg, H.B (2008). Role of interferon in homologous and heterologous rotavirus infection in the intestines and extraintestinal organs of suckling mice. *J Virol*, **82**: 7578–7590.

Ferrer–Admetlla, A., Sikora, M., Laayouni, H., Esteve, A., Roubinet, F., Blancher, A., Calafell, F., Bertranpetit, J., and Casals, F (2009). A Natural History of *FUT2* Polymorphism in Humans. *Mol. Biol. Evol*, 26(**9**): 1993–2003.

Fischer, T.K., Ashley, D., Kerin, T., Reynolds–Hedmann, E., Gentsch, J., Widdowson, M.A., et al. (2005). Rotavirus antigenemia in patients with acute gastroenteritis. *J Infect Dis*, **192**: 913–9.

Fischer–Walker, C.L, Rudan, I., Liu, L., et al. (2013). Global burden of childhood pneumonia and diarrhoea. *Lancet*, **381**: 1405–16.

Fu, C., Dong, Z., Shen, J., et al. (2018). Rotavirus gastroenteritis infection among children vaccinated and unvaccinated with rotavirus vaccine in Southern China: a population–based assessment. *JAMA Netw Open*, **1**: e181382.

Fu, C., Wang, M., Liang, J., et al. (2007). Effectiveness of Lanzhou lamb rotavirus vaccine against rotavirus gastroenteritis requiring hospitalization: a matched case–control study. *Vaccine*, **25**: 8756–8761.

Gabriel, S., Ziaugra, L., and Tabbaa, D (2009). SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*, 2.12.1–2.12.18.

GBD 2016 (2017). Causes of Death Collaborators. Global, regional, and national age–sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the global burden of disease study 2016. *Lancet*, **390**: 1151–1210.

GBD Diarrhoeal Diseases Collaborators (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis*, **17**: 909–48.

GBD, 2013. Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all–cause and cause–specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **385**: 117–71.

Gilger, M.A., Matson. D.O., Conner, M.E., Rosenblatt, H.M., Finegold, M.J., and Estes, M.K (1992). Extraintestinal rotavirus infections in children with immunodeficiency. *J Pediatr*, **120**: 912–7.

Glass, R.I., Jiang, B., and Parashar, U (2018). The future control of rotavirus disease: Can live oral vaccines alone solve the rotavirus problem? *Vaccine*, **36**: 2233–2236.

Greenberg, H.B., and Estes, M.K (2009). Rotaviruses: from pathogenesis to vaccination. Gastroenterology, **136**: 1939–51.

Groome, M.J, Fairlie, L., Morrison, J., et al. (2018). Safety and immunogenicity of a parenteral P2–VP8 subunit rotavirus vaccine. International Rotavirus Symposium. Minsk, Belarus.

Groome, M.J, Koen, A., Fix, A., et al. (2018). Safety and immunogenicity of a parenteral P2–VP8–P[8] subunit rotavirus vaccine in toddlers and infants in South Africa: a randomised, double–blind, placebo–controlled trial. *Lancet Infect Dis*, **17**: 843–853.

Groome, M.J., Koen, A., Fix, A., Page, N., Jose, L., Madhi, S.A., et al. (2017). Safety and immunogenicity of a parenteral P2–VP8–P[8] subunit rotavirus vaccine in toddlers and infants in South Africa: a randomised, double–blind, placebo–controlled trial. *Lancet Infect Dis*, **17**: 843–53.

Groome, M.J., Moon, S., Velasquez, D.E., Jones, S., Koen, A., van Niekerk, A., Jiang, B., Parashar, U.D., and Madhi, S.A (2014). Effect of breastfeeding on immunogenicity of oral live–attenuated human rotavirus vaccine: a randomized trial in HIV–uninfected infants in Soweto, South Africa. *Bull World Health Organ*, **92**: 238–245.

Grundbacher, F.J and Summerlin, D.C (1971). Inherited Differences in blood group A subtypes in Caucasians and Negroes. *Human Heredity*, 21(**1**).

Günaydin, G., Nordgren, J., Svensson, L., and Hammarström, L (2014). Mutations in Toll–Like Receptor 3 Are associated with elevated levels of rotavirus–specific IgG antibodies in IgA–deficient but not IgA enough individuals. *Clinical and Vaccine Immunology*, 21(**3**): 298–301.

Guo, Y., Audry, M., Ciancanelli, M., Alsina, L., Azevedo, J., Herman, M., Anguiano, E., Sancho–Shimizu, V., Lorenzo, L., et al. (2011). Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *J Exp Med*, **208**: 2083–2098.

Gurwith, M., Wenman, W., Hinde, D., Feltham, S., and Greenberg, H (1981). A prospective study of rotavirus infection in infants and young children. *J Infect Dis*, **144**: 218–24.

Harris, V.C., et al. (2017). Significant correlation between the infant gut microbiome and rotavirus vaccine response in rural Ghana. *J Infect Dis*, **215**: 34–41.

Haselhorst, T., Fleming, F.E., Dyason, J.C., et al. (2009). Sialic acid dependence in rotavirus host cell invasion. *Nat Chem Biol*, **5**: 91–93.

He, J.F., Jia, W.H., Fan, Q., Zhou, X.X., Qin, H.D., Shugart, Y.Y., et al. (2007). Genetic polymorphisms of TLR3 are associated with nasopharyngeal carcinoma risk in Cantonese population. *BMC Cancer*, **7**: 194.

Hendry, L.M. (2015). Genetics of HIV–associated sensory neuropathy in black Southern Africans, MSc dissertation. Faculty of Health Sciences, University of the Witwatersrand.

Hider, S.L., Thomson, W., Mack, L.F., Armstrong, D.J., Shadforth, M., and Bruce, I.N (2008). Polymorphisms within the adenosine receptor 2a gene are associated with adverse events in RA patients treated with MTX. *Rheumatology (Oxford, England)*, 47(**8**): 1156–1159.

Hober, D., and Sauter, P (2010). Pathogenesis of type 1 diabetes mellitus: interplay between enterovirus and host. *Nat Rev Endocrinol*, **6**: 279–289.

Holloway, G and Coulson, B.S (2013). Innate cellular responses to rotavirus infection. *Journal of General Virology*, **94**: 1151–1160.

Holloway, G., Truong, T.T., and Coulson, B.S (2009). Rotavirus antagonizes cellular antiviral responses by inhibiting the nuclear accumulation of STAT1, STAT2, and NF–κB. *J Virol*, **83**: 4942–4951.

Honeyman, M.C., Coulson, B.S., Stone, N.L., Gellert, S.A., Goldwater, P.N., et al. (2000). Association between rotavirus infection and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes. *Diabetes*, **49**: 1319–1324.

Hosseini–Maaf, B., Smart, E., Chester, M.A., and Olsson, M.L (2005). The $A_{bantu}$ phenotype in the ABO blood group system is due to a splice–site mutation in a hybrid between a new $O_1$–like allelic lineage and the $A_2$ allele. *Vox Sanguinis*, 88(**4**): 256–264.

Hu, L., Crawford, S.E., Czako, R., Cortes–Penfield, N.W., Smith, D.F., Le Pendu, J., Estes, M.K., and Prasad, B.V (2012). Cell attachment protein VP8* of a human rotavirus specifically interacts with A–type Histo–Blood Group Antigen. *Nature*, **485:** 256–259.

Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., et al. (2018). Ensembl variation resources, *Database Volume*.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Imbert–Marcille, B.M., Barbe, L., Dupe, M., et al. (2013). A *FUT2* gene common polymorphism determines resistance to rotavirus A of the P[8] genotype. *J Infect Dis*, **209**: 1227–30.

International Vaccine Access Center (IVAC). Johns Hopkins Bloomberg School of Public Health. Available at: www.view–hub.org. [Accessed 14 March 2019]

Isanaka, S., Guindo, O., Langendorf, C., et al. (2017). Efficacy of a low–cost, heat–stable oral rotavirus vaccine in Niger. *N Engl J Med*, **376**: 1121–1130.

Jensen, S and Thomsen, A.R (2012). Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J Virol*, **86**: 2900–2910.

Jere, C., Cunliffe, N.A., Hoffman, I.F., Stewart P.W., Kilaru, R., Broadhead, R.L, et al. (2001). Plasma HIV burden in Malawian children co–infected with rotavirus. *AIDS*, **15**: 1439–42.

Jiang, B., Wang, Y., and Glass, R.I (2013). Does a monovalent inactivated human rotavirus vaccine induce heterotypic immunity? Evidence from animal studies. *Hum Vaccin Immunother*, **9**: 1634–1637.

Jiang, V., Jiang, B., Tate, J., Parashar, U.D., and Patel, M.M (2010). Performance of rotavirus vaccines in developed and developing countries. *Hum Vaccin*, **6**: 532–542.

Jiang, X., Liu, Y., and Tan, M (2017). Histo–blood group antigens as receptors for rotavirus, new understanding on rotavirus epidemiology and vaccine strategy. *Emerging Microbes & Infections*, **6**: e22.

Jonesteller, C.L., Burnett, E., Yen, C., Tate, J.E., and Parashar, U.D (2017). Effectiveness of Rotavirus Vaccination: A Systematic Review of the First Decade of Global Postlicensure Data, 2006–2016. *Clinical Infectious Diseases*, 65(**5**): 840–50.

Kandasamy, S., Chattha, K.S., Vlasova, A.N., and Saif, L.J. (2014). Prenatal vitamin A deficiency impairs adaptive immune responses to pentavalent rotavirus vaccine (RotaTeq(R)) in a neonatal gnotobiotic pig model. *Vaccine*, **32**: 816–824.

Kazi, A.M., Cortese, M.M., Yu, Y., Lopman, B., Morrow, A.L., Fleming, J.A., et al. (2017). Secretor and salivary ABO blood group antigen status predict rotavirus vaccine take in infants. *JID*, **215**: 786–9

Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G., and Lowe, J.B (1995). Sequence and expression of a candidate for the human secretor blood group α (1,2) fucosyltransferase gene (*FUT2*). Homozygosity for an enzyme–inactivating nonsense mutation commonly correlates with the non–secretor phenotype. *J Biol Chem*, **270**: 4640–4649.

Kindberg, E., Vene, S., Mickiene, A., Lundkvist, A., Lindquist, L., and Svensson, L (2011). A Functional Toll–Like Receptor 3 Gene (*TLR3*) may be a risk factor for Tick–borne encephalitis virus (TBEV) infection. *The Journal of Infectious Diseases*, **203**: 523–528.

Kitano, T., Blancher, A., and Saitou, N (2012). The functional A Allele was resurrected via recombination in the human ABO blood group gene. *Molecular Biology and Evolution*, 29(**7**): 1791–1796.

Koda, Y., Soejima, M., and Kimura, H (2001). The polymorphisms of fucosyltransferases. *Legal Medicine*, **3**: 2–14.

Koda, Y., Soejima, M., Wang, B.J., and Kimura, H (1997). Structure and expression of the gene encoding secretor–type galactoside 2–α–l–fucosyltranserase (*FUT2*). *Eur J Biochem*, **246**: 750–755.

Kramarz, P., France, E.K., Destefano, F., Black, S.B., Shinefield, H., Ward, J.I., et al. (2001). Population–based study of rotavirus vaccination and intussusception. *Pediatr Infect Dis J*, **20**: 410–6.

Kulkarni, P.S., Desai, S., Tewari, T., et al. (2017). SII BRV–PV author group. A randomized Phase III clinical trial to assess the efficacy of a bovine–human reassortant pentavalent rotavirus vaccine in Indian infants. *Vaccine*, **35**: 6228–6237.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, **35**: 1547–1549.

Lanata, C.F., Fischer–Walker, C.L., Olascoaga, A.C., et al. (2013). Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. *PLoS One*, **8**: e72788.

Lappalainen, S., Pastor, A.R., Malm, M., et al. (2015). Protection against live rotavirus challenge in mice induced by parenteral and mucosal delivery of VP6 subunit rotavirus vaccine. *Arch Virol*, **160**: 2075–2078.

Larsson, M.M., et al. (2006). Antibody prevalence and titer to norovirus (genogroup II) correlate with secretor (*FUT2*) but not with ABO phenotype or Lewis (*FUT3*) genotype. *J Infect Dis*, **194**: 1422–1427.

Lee, B., Dickson, D.M., deCamp, A.C., Colgate, E.R., Diehl, S.A., et al. (2018). Histo–blood group antigen phenotype determines susceptibility to genotype–specific rotavirus infections and impacts measures of rotavirus vaccine efficacy. *The Journal of Infectious Diseases*, **217**: 1399–407.

Levy, K., Hubbardm A.E., and Eisenberg, J.N (2009). Seasonality of rotavirus disease in the tropics: a systematic review and meta–analysis. *Int J Epidemiol*, **38**: 1487–6.

Li, D., Xu, Z., Xie, G., et al. (2015). Genotype of Rotavirus vaccine strain LLR in China is G10P[15]. *Bing Du Xue Bao*, **31**: 170–173.

Li, J.S., Cao, B., Gao, H.C., et al. (2018). Faecal shedding of rotavirus vaccine in Chinese children after vaccination with Lanzhou lamb rotavirus vaccine. *Sci Rep*, **8**: 1001.

Li, Y., Xue, M., Yu, L., et al. (2018). Expression and characterization of a novel truncated rotavirus VP4 for the development of a recombinant rotavirus vaccine. *Vaccine*, **36**: 2086–2092.

Liu, Y, Ramelot, T.A, Huang, P., et al. (2016). Glycan specificity of P[19] rotavirus and comparison with those of related P genotypes. *J Virol*, **90**: 9983–9996.

Liu, Y., Huang, P., Tan, M., Liu, Y., Biesiada, J., Meller, J., Castello, A.A., Jiang, B., Xi Jiang., et al. (2012). Rotavirus VP8*: phylogeny, host range, and interaction with histo–blood group antigens. *J. Virol*, **86**: 9899–9910.

Liu, Y.H., Koda, Y., Soejima, M., Pang, H., Schlaphoff, T., du Toit, E.D., and Kimura, H (1998). Extensive polymorphism of the *FUT2* gene in an African (Xhosa) population of South Africa. *Hum Genet*, **103**: 204–210.

Lopez, S and Arias, C.F (2003). Attachment and post–attachment receptors for rotavirus. In Viral Gastroenteritis (Desselberger, U. and Gray, J., eds), *Elsevier*, pp. 143–163.

Lozano, R., Naghavi, M., Foreman, K., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, **380**: 2095–128.

Luan le T., Trang, N.V., Phuong, N.M., et al. (2009). Development and characterization of candidate rotavirus vaccine strains derived from children with diarrhoea in Vietnam. *Vaccine*, 27(**Suppl 5**): F130–F138.

MacDonald, J., Groome, M.J., Mans, J., and Page, P (2020). *FUT2* Secretor Status Influences Susceptibility to VP4 Strain–Specific Rotavirus Infections in South African Children. *Pathogens*, 9: 795.

Madhi, S.A., Cunliffe, N.A., Steele, D., Witte, D., Kirsten, M., Louw, C., et al. (2010). Effect of human rotavirus vaccine on severe diarrhea in African infants. *N Engl J Med*, **362**: 289–98.

Magwira, C.A., Kgosana, L.P., Esona, M.D., and Seheri, M.L (2020). Low fecal rotavirus vaccine virus shedding is significantly associated with non–secretor histo–blood group antigen phenotype among infants in northern Pretoria, South Africa. *Vaccine*.

Matson, D.O (2003). Rotaviruses. In: Long SS, Pickering LK, Prober CG, Eds. Principles and Practice of Pediatric Infectious Diseases. New York, NY: Churchill Livingstone.

Matthijnssens, J., et al. (2010). Molecular and biological characterization of the 5 human–bovine rotavirus (WC3)–based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq. *Virology*, **403**: 111–127.

Misevic, G., 2018. ABO blood group system. *Asia–Pacific Journal of Blood Types and Genes*, 2(**2**): 71–84.

Möller, M., Jöud, M., Storry, J. R., and Olsson, M. L (2016). Erythrogene: a database for in–depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. *Blood Advances*, 1(**3**): 240–249.

Moller, M., Joud, M., Storry, J.R., and Olsson, M.L (2016). Erythrogene: a database for in–depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. *Blood Advances*, 1(**3**), 240–249.

Moon, S., Groome, M.J., Velasquez, D.E., Parashar, U.D., Jones, S., Koen, A., van Niekerk, N., Jiang, B., and Madhi, S.A (2016). Prevaccination rotavirus serum IgG and IgA are associated with lower immunogenicity of live, oral human rotavirus vaccine in South African Infants. *Clinical Infectious Diseases*, 62(**2**): 157–65.

Moon, S.S., et al. (2016). Prevaccination rotavirus serum IgG and IgA are associated with lower immunogenicity of live, oral human rotavirus vaccine in South African infants. *Clin Infect Dis*, **62**: 157–165.

Moores, P and Brain, P (1968). Lewis Groups and Secretor Status in Natal Bantu. *Transfusion*, 8(**5**): 283–288.

Moores, P.P (1991). Human blood groups and antibodies (PhD). University of Natal, Durban, South Africa.

Morelli, M., Dennis, A.F., and Patton, J.T (2015). Putative E3 ubiquitin ligase of human rotavirus inhibits NF–kB activation by using molecular mimicry to target β–TrCP. *mBio*, 6(**1**): e02490–14.

Morelli, M., Ogden, K.M., and Patton, J.T (2015). Silencing the alarms: Innate immune antagonism by rotavirus NSP1 and VP3. Virology, **479**–**480**:75–84.

Murphy, T.V., Gargiullo, P.M., Massoudi, M.S., Nelson, D.B., Jumaan, A.O., Okoro, C.A., et al. (2001). Intussusception among infants given an oral rotavirus vaccine. *N Engl J Med*, **344**: 564–72.

Naylor, C., et al. (2015). Environmental enteropathy, oral vaccine failure and growth faltering in infants in Bangladesh. *EBioMedicine*, **2**: 1759–1766.

Nordgren J, Nitiema L.W, Ouermi, D, Simpore J, and Svensson, L (2013). Host genetic factors affect susceptibility to norovirus infections in Burkina Faso. *PLoS One*, **8**: e69557.

Nordgren, J., Sharma, S., Bucardo, F., Nasir, W., Günaydin, G., Ouermi, D., et al. (2014). Both Lewis and Secretor Status Mediate Susceptibility to Rotavirus Infections in a Rotavirus Genotype–Dependent Manner. *Clinical Infectious Diseases*, 59 (**11**): 1567–73.

Owaidah, A.I., Noor M Naffaa, N.M., Alumran, A., and Alzahrani, F (2020). Phenotype frequencies of major blood group systems (Rh, Kell, Kidd, Duffy, MNS, P, Lewis, and Lutheran) among blood donors in the Eastern region of Saudi Arabia. *Journal of Blood Medicine*, **11** 59–65.

Page, P., Kruger, T., Seheri, M., Peenze, I., Quan, V., Groome, M., and Madhi, S. (2018). Rotavirus surveillance report, South Africa, 2014–2015: A comparison with previous rotavirus seasons. *Communicable Diseases Surveillance Bulletin*, 41(**4**): 119–125.

Pane, J.A., Webster, N.L, and Coulson, B.S (2014). Rotavirus activates lymphocytes from non–obese diabetic mice by triggering toll–like receptor 7 signaling and interferon production in plasmacytoid dendritic cells. *PLoS Pathog*, 10(**3**): e1003998.

Pang, H., Liu, Y., Koda, Y., Soejima, M., Jia, J., Schlaphoff, T., du Toit, E.D., and Kimura, H (1998). Five novel missense mutations of the Lewis gene (*FUT3*) in African (Xhosa) and Caucasian populations in South Africa. *Hum Genet*, **102**: 675–680.

Parashar, U.D., Burton, A., Lanata, C., Boschi–Pinto, C., Shibuya, K., Steele, D., et al. (2009). Global mortality associated with rotavirus disease among children in 2004. *J Infect Dis*, **200**: 9–15.

Parashar, U.D., Hummelman, E.G., Bresee, J.S., Miller M.A, and Glass, R.I (2003). Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis*, **9**: 565–72.

Patel, M., et al. (2009). Association between pentavalent rotavirus vaccine and severe rotavirus diarrhea among children in Nicaragua. *JAMA*, **301**: 2243–2251.

Patel, M., Glass, R.I., Jiang, B., Santosham, M., Lopman, B., and Parashar, U (2013). A systematic review of anti–rotavirus serum IgA antibody titer as a potential correlate of rotavirus vaccine efficacy. *J. Infect. Dis*, **208**: 284–294.

Patel, M., Shane, A.L., Parashar, U.D., Jiang B, Gentsch J.R., and Glass, R.I (2009). Oral rotavirus vaccines: how well will they work where they are needed most? *J Infect Dis*, 200(**Suppl 1**): S39–48.

Patel, M., Steele, A.D., and Parashar, U.D. (2012). Influence of oral polio vaccines on performance of the monovalent and pentavalent rotavirus vaccines. *Vaccine*, 30(**Suppl 1**), A30–35.

Payne, D.C., Currier, R.L., Staat, M.A., et al. (2015). Epidemiologic association between *FUT2* secretor status and severe rotavirus gastroenteritis in children in the United States. *JAMA Pediatr*, 169(**11**): 1040–1045.

Perez de Diego, R., Sancho–Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., Herman, M., Cardon, A., Durandy, A., et al. (2010). Human TRAF3 adaptor molecule deficiency leads to impaired Toll–like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity*, **33**: 400–411.

Pérez–Ortín, R., Vila–Vicent, S., Carmona–Vicente, N., and Santiso–Bellón, C (2019). *Viruses*, **339**: 1–11.

Perrett, K.P., Jachno, K., Nolan, T.M., and Harrison, L.C (2019). Association of rotavirus vaccination with the incidence of type 1 diabetes in children. *JAMA Pediatr*, 173(**3**): 280–282.

Peter, G., and Myers, M.G (2002). Intussusception, rotavirus, and oral vaccines: summary of a workshop. *Pediatrics*, **110**: 67.

Pollock, L., Bennett, A., Jere, K.C., Dube, Q., Mandolo, J., et al. (2018). Non–secretor histo–blood group antigen phenotype is associated with reduced risk of clinical rotavirus vaccine failure in Malawian infants. *Infectious Diseases Society of America*, **1**: 1–9.

Pott, J., Stockinger, S., Torow, N., Smoczek, A., Lindner, C., et al. (2012). Age–dependent TLR3 expression of the intestinal epithelium contributes to rotavirus susceptibility. *PLoS Pathog*, 8(**5**): e1002670.

Purcell, S., Neale, B., Todd–Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J (2007). PLINK: a tool set for whole–genome association and population–based linkage analyses. *The American Journal of Human Genetics*, 81(**3**): 559–575.

Ramani, S., Cortes–Penfield, N.W., Hu, L., et al. (2013). The VP8* domain of neonatal rotavirus strain G10P[11] binds to type II precursor glycans. *J. Virol*, **87**: 7255–7264.

Randall, R.E and Goodbourn, S (2008). Interferons and viruses: an interplay between induction, signalling, antiviral responses, and virus countermeasures. *J Gen Virol*, **89**: 1–47.

Ravn, V and Dabelsteen, E (2000). Tissue distribution of Histo–Blood Group Antigens. *AMIS*, **108**: 1–28.

Resch, T.K., Wang, Y., Moon, S.S., et al. (2018). Inactivated rotavirus vaccine by parenteral administration induces mucosal immunity in mice. *Sci Rep*, **8**: 561.

Rodríguez–Díaz, J and Buesa, J (2019). Histo–Blood Group Antigens in Children with Symptomatic Rotavirus Infection.

Ruiz–Palacios, G.M., Perez–Schael, I., Velazquez, F.R., Abate, H., Breuer, T., Clemens, S.C., et al. (2006). Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *N Engl J Med*, **354**: 11–22.

Santos, N and Hoshino, Y (2005). Global distribution of rotavirus serotypes/ genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev Med Virol*, **15**: 29–56.

Santos, N and Hoshino, Y (2005). Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev. Med. Virol*. **15**: 29–56.

Seheri, L.M., Dewar, J.B., van der Merwe, L., Geyer, A., John Tumbo, J., et al. (2010). Prospective Hospital–Based Surveillance to Estimate Rotavirus Disease Burden in the Gauteng and North West Province of South Africa during 2003–2005. *The Journal of Infectious Diseases*, 202(**S1**): S131–S138.

Sequenom (2009). Sequenom applications overview. *Sequenom, Inc*, V**1.1**: 1–36.

Serpa, J., et al. (2003). Lewis enzyme (alpha1–3/4 fucosyltransferase) polymorphisms do not explain the Lewis phenotype in the gastric mucosa of a Portuguese population. *J Hum Genet*, **48**: 183–189.

Settembre, E.C. Chen, J.Z., Dormitzer, P.R., Grigorieff, N., and Harrison, S.C (2011). Atomic model of an infectious rotavirus particle. *Embo. J*. **30**: 408–416.

Shapiro, M (1951). The ABO, MN, P and Rh blood group systems in the South African Bantu* A genetic study. S.A Medical Journal, 187–192.

Silvestri, L.S., Taraporewala, Z.F., and Patton, J.T (2004). Rotavirus replication: plus–sense templates for double–stranded RNA synthesis are made in viroplasms. *J Virol*, 78(**14**): 7763–74.

Sironi, M., Biasin, M.., Cagliani, R., Forni, D., De Luca, M., et al. (2012). A common polymorphism in *TLR3* confers natural resistance to HIV–1 infection. *J Immunol*, **188**: 818–823.

Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Tirgoviste, C., et al. (2006). A genome–wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon–induced helicase (IFIH1) region. *Nat Genet*, **38**: 617–619.

Soejima, M., Pang, H., and Koda, Y (2007). Genetic variation of *FUT2* in a Ghanaian population: identification of four novel mutations and inference of balancing selection. *Ann Hematol*, **86**: 199–204.

Soloff, A. C and Barratt–Boyes, S. M (2010). Enemy at the gates: dendritic cells and immunity to mucosal pathogens. *Cell Res*, **20**, 872–885.

Stachan, T and Read, A (2011). Human Molecular Genetics (4$^{th}$ edition). *Garland Science, Taylor & Francis Group, LLC*, New York.

Steele, A.D., Cunliffe, N., Tumbo, J., Madhi, SA., De Vos, B., and Bouckenooghe, A (2009). A review of rotavirus infection in and vaccination of human immunodeficiency virus–infected children. *J Infect Dis*, **200**: 57–62.

Steele, A.D., Peenze, M.C., de Beer, C.T., Pager, J., and Yeats, N (2003). Anticipating rotavirus vaccines: epidemiology and surveillance of rotavirus in South Africa, *Vaccine*, **21**: 354–60.

Sun, X., Guo, N., Li, J., et al. (2016). Rotavirus infection and histo–blood group antigens in the children hospitalized with diarrhoea in China. *Clinical Microbiology and Infection*, **22**: 740.e1–740.e3.

Sun, X., Wang, L., Qi, J., Li, D., Wang, M., et al. (2018). Human group C rotavirus VP8*s recognize type A histo–blood group antigens as ligands. *J Virol*, 92(**11**): e00442–18.

Svensson, A., Tunback, P., Nordstrom, I., Padyukov, L., and Eriksson, K (2012). Polymorphisms in Toll–like receptor 3 confer natural resistance to human herpes simplex virus type 2 infection. *Journal of General Virology*, **93**: 1717–1724.

Svidnicki, M.C., Silva–Costa, S.M., Ramos, P.Z., Pereira dos Santos, N.Z., Martins, F.T., et al. (2015). Screening of genetic alterations related to non–syndromic hearing loss using MassARRAY iPLEX® technology. *BMC Medical Genetics*, **16**: 85.

Tailford, L.E., Crost, E.H., Kavanaugh, D., et al. (2015). Mucin glycan foraging in the human gut microbiome. *Front Genet*, **6**: 81.

Tamminen, K., Lappalainen, S., Huhti, L., et al. (2013). Trivalent combination vaccine induces broad heterologous immune responses to norovirus and rotavirus in mice. *PLoS One*, **8**: e70409.

Taniuchi, M., et al. (2016). Impact of enterovirus and other enteric pathogens on oral polio and rotavirus vaccine performance in Bangladeshi infants. *Vaccine*, **34**: 3068–3075.

Tate, J.E., Burton, A.H., Boschi–Pinto, C., and U.D. Parashar (2016). Global, regional, and national estimates of rotavirus mortality in children <5 Years of Age, 2000–2013. Clinical Infectious Diseases, 62(**S2**): S96–105.

Tate, J.E., Burton, A.H., Boschi–Pinto. C., and Parashar, U.D (2016). Global, regional, and national estimates of rotavirus mortality in children; 5 years of age, 2000–2013. *Clin Infect Dis*, 62(**S2**): S96–105.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., et al. (2007) Robust associations of four new chromosome regions from genome–wide analyses of type 1 diabetes. *Nat Genet*, **39**: 857–864.

Todd, S., Nicola, A., Page, A., Steele, D., Peenze, I., Cunliffe, NA., et al. (2010). Rotavirus strain types circulating in Africa: review of studies published during 1997–2006. *J. Infect. Dis*, **202** (Suppl), S34–S42.

Torrado, J., et al. (1997). Lewis, secretor, and ABO phenotypes, and sulfomucin expression in gastric intestinal metaplasia. *Cancer Epidemiol Biomarkers Prev*, **6**: 287–289.

Troeger, C., Khalil, I.A., Rao, P.C., Cao, S., Blacker, B.F., Ahmed, T., Armah, G., Bines, J.E., Brewer, T.G., Colombara, D.V., et al. (2018). The changing epidemiology of rotavirus and its role as the leading cause of diarrhea burden among children under 5 years old. *JAMA Pediatr*. Available at doi:10.1001/jamapediatrics.2018.1960 [Accessed 13 August 2018]

Uchiyama, R., Chassaing, B., Zhang, B., and Gewirtz, A.T (2015). MyD88–mediated TLR signaling protects against acute rotavirus infection while inflammasome cytokines direct Ab response. *Innate Immunity*, 21(**4**): 416–428.

van der Werf, N., Kroese, F.M., Rozing, J., and Hillebrands, J.L (2007) Viral infections as potential triggers of type 1 diabetes. *Diabetes Metab Res Rev*, **23**: 169–183.

Van Trang, N., Vu, H.V., Le, N.T., Huang, P., Xi Jiang, X., and Anh, D.D (2014) Association between norovirus and rotavirus infection and histo–blood group antigen types in Vietnamese children. *J. Clin. Microbiol*, 52(**5**): 1366.

Velasquez, D.E, Parashar, U., and Jiang, B (2018). Decreased performance of live attenuated, oral rotavirus vaccines in low–income settings: causes and contributing factors. *Expert Rev Vaccines*, **17**: 145–161

Velazquez, F.R., Matson, D.O., Calva, J.J., Guerrero, L., Morrow, A.L., Carter–Campbell, S., et al. (1996). Rotavirus infections in infants as protection against subsequent infections. *N Engl J Med*, **335**: 1022–8.

Velazquez, F.R., Matson, D.O., Guerrero, L., Shults, J. J., Calva, J.J., Morrow, A.L., Glass, L. K., and Ruiz–Palacios, G. M (2000). Serum antibody as a marker of protection against natural rotavirus infection and disease. *J. Infect. Dis*, **182**: 1602–1609.

Vesikari, T., Matson, D.O., Dennehy, P., Van Damme, P., Santosham, M., Rodriguez, Z, et al. (2006). Safety and efficacy of a pentavalent human–bovine (WC3) reas–sortant rotavirus vaccine. *N Engl J Med*, **354**: 23–33.

Vesikari, T., Karvonen, A., Prymula, R., Schuster, V, Tejedor, J.C, Cohen, R., et al. (2007). Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, doubleblind controlled study. *Lancet*, **370**: 1757–63.

Wang, Y., Azevedo, M., Saif, L.J., et al. (2010) (A). Inactivated rotavirus vaccine induces protective immunity in gnotobiotic piglets. *Vaccine*, **28**: 5432–5436.

Wang, Y., Vlasova, A., Velasquez, D.E., et al. (2016) (B). Skin vaccination against rotavirus using microneedles: proof of concept in gnotobiotic piglets. *PLoS One*, **11**: e0166038.

Wellcome Trust Case Control C (2007). Genome–wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**: 66–678.

Wolpin, B.M., Kraft, P., Xu, M., et al. (2010). Variant ABO blood group alleles, secretor status, and risk of pancreatic cancer: results from the pancreatic cancer cohort consortium. *Cancer Epidemiol Biomarkers Prev*, 19(**12**): 3140–9.

World Health Organization (2013). Child cause of death estimates 2000–2013. Available at: http://www.who.int/entity/healthinfo/global_burden_disease/childCOD_estimates_2000_2013.xls?ua=1 [Accessed 12 January 2015].

World Health Organization. Meeting of the Global Advisory Committee on Vaccine Safety, December 2010 (2011). *Wkly Epidemiol Rec*, **86**: 38–43.

Wu, J.Y., Zhou, Y., Zhang, G.M., et al. (2018). Isolation and characterization of a new candidate human inactivated rotavirus vaccine strain from hospitalized children in Yunnan, China. *World J Clin Cases*, **6**: 426–440.

Xu, J., Yang, J., Sun, J., Ding, Y., Su, L., Shao, C., and Jiang, B (2006). Expression of Toll–like receptors and their association with cytokine responses in peripheral blood mononuclear cells of children with acute rotavirus diarrhoea. *Clinical and Experimental Immunology*, **144**: 376–381.

Xu, J., Yang, J., Sun, J., Ding, Y., Su, L., Shao, C., and Jiang, B (2006). Expression of Toll–like receptors and their association with cytokine responses in peripheral blood mononuclear cells of children with acute rotavirus diarrhoea. Clin Exp Immunol, 144(**3**): 376–81.

Yamamoto, F., Cid, E., Yamamoto, M., and Blancher, A (2012). ABO research in the modern era of genomics. *Transfusion Medicine Reviews*, 26(**2**): 103–118.

Yen, C., Tate, J.E., Patel, M.M., Cortese, M.M., Lopman, B., Fleming, J., et al. (2011). Update on global impact and future priorities. *Human Vaccines*, 7(**12**): 1282–1290.

Yip, S.P (2002). Sequence variation at the human ABO locus. *Annals of Human Genetics*, 66(**1**): 1–27.

Yoshida, A., Davé, V., and Prchal, J (1985). Uncertainty in identification of blood group A subtypes by agglutination test. *Human Heredity*, 35(**1**): 1–6.

Zaman, K., Dang, D.A., Victor, J.C., Shin, S., Yunus, M., Dallas, M.J., et al. (2010). Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double–blind, placebo–controlled trial. *Lancet*, **376**: 615–23.

Zhang, S. Y., Jouanguy, E., Ugolini, S., Smahi, A., Elain, G., Romero, P., Segal, D., Sancho–Shimizu, V., Lorenzo, L., et al. (2007). TLR3 deficiency in patients with herpes simplex encephalitis. *Science*, **317**: 1522–1527.

Zhang, X., Long, Y., Tan, M., et al. (2016). P[8] and P[4] Rotavirus Infection Associated with Secretor Phenotypes Among Children in South China. *Sci. Rep*, **6**: 34591.

# 6. APPENDICES

## Appendix A

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

**HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)**

08 August 2017

**Dr Michelle Groome**
Respiratory and Meningeal Pathogens Research Unit
11th Floor, Central West Wing
Nurses Residence
Chris Hani Baragwanath Academic Hospital
PO Box 90753
Bertsham
2013

Sent by email to: groomem@rmpru.co.za

Dear Dr Groome

**Re:** **Protocol Ref no: M090824**
**Protocol Title:** Immunogenicity of Pentaxim, Prevnar, Rotarix and other Childhood Vaccines Included in the new Immunization Schedule for South African Infants
**Principal Investigator:** Prof Shabir Madhi et al
**Protocol Amendment: Storage Amendment**

This letter serves to confirm that the Chairman of the Human Research Ethics Committee (Medical) has approved the protocol amendment for the abovementioned protocol, as detailed in your letter dated 07 July 2017.

The following documents were received:

- Cover Letter dated 07 July 2017.
- Storage Amendment Application Form dated 07 July 2017.

Thank you for keeping us informed and updated.

Yours Sincerely,

L.moeng

**Mr Lebohang Moeng**
**Administrative Assistant**
**Human Research Ethics Committee (Medical)**

UNIVERSITY OF THE WITWATERSRAND
OFFICE OF THE DEPUTY VICE CHANCELLOR RESEARCH
08 AUG 2017
RESEARCH OFFICE HREC (Medical)
JOHANNESBURG

Research Office Secretariat: Faculty of Health Sciences, Phillip Tobias Building, 3rd Floor, Office 302, Corner York Road and 29 Princess of Wales Terrace, Parktown, 2193 Private Bag 3, Wits 2050 I **T**+27 (0)11-717-1234/2656/2700/1252 **E**: Lebo.Moeng@wits.ac.za I **Office E** HREC-Medical.ResearchOffice@wits.ac.za IWebsite: www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

**Appendix B**

**Map:** UEA–1 lectin (secretor) – two wells per sample.

| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | | | | | | | | | | | | |
| **B** | | | | | | | | | | | | |
| **C** | | | | | | | | | | | | |
| **D** | | | | | | | | | | | | |
| **E** | | | | | | | | | | | | |
| **F** | | | | | | | | | | | | |
| **G** | | | | | | | | | | | | |
| **H** | | | | | | | | | | | | |

**Map:** ABO / Lewis – eight wells per sample.

| | ID | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | **A** | | | | | | | | | | | | |
| **A** | **B** | | | | | | | | | | | | |
| **B** | **C** | | | | | | | | | | | | |
| **B** | **D** | | | | | | | | | | | | |
| **Le A** | **E** | | | | | | | | | | | | |
| **Le A** | **F** | | | | | | | | | | | | |
| **Le B** | **G** | | | | | | | | | | | | |
| **Le B** | **H** | | | | | | | | | | | | |

**Appendix C**

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT3*** | rs3894326 | A/T | 0.10 (T) | 0.081 | No | missense variant | Associated with Le A phenotype and RV infection, cystic fibrosis lung disease, and ulcerative colitis. | 0 (deleterious) SIFT | Günaydin et al., 2016; Bucardo et al., 2018; Hu et al., 2016; Taylor –Cousar, 2009 |
| ***FUT3*** | rs761121538 | C/A/T | – | – | – | missense variant | – | 0 (deleterious) SIFT | – |
| ***FUT3*** | rs28381968 | C/T | 0.02 (T) | 0.131 | No | missense variant | – | 0 (deleterious) SIFT | – |
| ***FUT3*** | rs28362466 | C/T | 0.05 (T) | 0.253 | Yes (17) | missense variant | – | 1 (deleterious) PolyPhen–2 | – |
| ***FUT3*** | rs372597863 | C/T | <0.01 (T) | – | – | synonymous variant | – | – | – |
| ***FUT3*** | rs3745635 | C/T | 0.15 (T) | 0.338 | Yes (3) | missense variant | Associated with Le– status, CA19–9 antigen expression, and ulcerative colitis. | 0.04 (deleterious) SIFT | Koda et al., 1993; Hu et al., 2016; Wannhoff et al., 2016; Guo et al., 2017; |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT3*** | rs28362463 | C/T | 0.05 (T) | 0.268 | Yes (4) | missense variant | Associated with Le status and RV infection. | 0 (deleterious) SIFT | Nordgren et al., 2014 |
| ***FUT3*** | rs778986 | A/G | 0.11 (A) | 0.066 | No | missense | Associated with Le A phenotype and RV infection, and cystic fibrosis lung disease. | 0.01 (deleterious) SIFT | Günaydin et al., 2016; Bucardo et al., 2018; Taylor–Cousar, 2009 |
| ***FUT3*** | rs812936 | G/A | 0.12 (G) | 0.071 | Yes (1) | missense variant | Associated with Le A phenotype and RV infection, and cystic fibrosis lung disease. | 1 (tolerated) SIFT | Günaydin et al., 2016; Bucardo et al., 2018; Taylor–Cousar, 2009 |
| ***FUT3*** | rs28362458 | C/T | 0.05 (T) | 0.268 | Yes (5) | missense variant | Associated with Le status and RV infection. | 0.26 (tolerated) SIFT | Nordgren et al., 2014 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***TLR7*** | rs179008 | A/T | 0.12 (T) | 0.091 | Yes (5) | missense variant | Modulate natural hepatitis C outcomes and liver disease progression, susceptibility to HIV and HCV co–infection etc. | 0.25 (tolerated) SIFT | Fakhir et al., 2018; Valverde–Villegas et al., 2017 etc. |
| | | | | | | | | | |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***TLR7*** | rs864058 | G/A | 0.11 (A) | 0.279 | No | synonymous variant | Associated with symptoms and severity of rift valley fever, allergic rhinitis etc. | – | Hise et al., 2015; Nilsson et al., 2015 etc. |
| ***TLR7*** | rs3853839 | C/G/T | 0.40 (G) | 0.214 | Yes (8) | 3 prime UTR variant | Associated with risk of systemic lupus erythematosus, susceptibility to chikungunya virus infection etc. | – | Raafat et al., 2018; Dutta et al., 2017 etc. |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***TLR7*** | rs179007 | G/A | 0.13 (G) | 0.896 | No | intergenic variant | Associated with allergic rhinitis. | – | Nilsson et al., 2012 |
| ***TLR7*** | rs5935438 | G/C | 0.36 (C) | 0.416 | Yes (9) | regulatory region variant | Associated with allergic rhinitis. | – | Nilsson et al., 2012 |
| ***TLR8*** | rs2159377 | C/T | 0.37 (T) | 0.286 | | synonymous variant | Systemic lupus erythematosus. | – | Sabado et al., 2015 |
| ***DDX58*** | rs55789327 | G/A | 0.05 (A) | 0.121 | Yes (5) | missense variant | Associated with antiviral immune response. | – | Pothlichet et al., 2009 |
| ***DDX58*** | rs669260 | T/A/C | 0.10 (C) | 0.187 | Yes (5) | intron variant | Associated with response to rubella and measles vaccination, and dengue virus etc. | – | Ovsyannikova et al., 2014; Alagarasu et al., 2015 etc. |
| ***DDX58*** | rs10813831 | G/A | 0.18 (A) | 0.162 | Yes (1) | missense variant | Associated with various cancers, various vaccine responses, | 0.65 (possibly damaging) PolyPhen–2 | Ovsyannikova et al., 2010; Kennedy et al., 2012; Biggins et al., 2013; |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | HVC severity etc. | | Moumad et al., 2003 etc. |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT2*** | rs679574 | C/G | 0.32 (G) | 0.439 | Yes (19) | intron variant | Associated with type 1 diabetes susceptibility. | – | Ram et al., 2016 |
| ***FUT2*** | rs516316 | G/C | 0.32 (C) | 0.439 | Yes (19) | intron variant | Associated with plasma B12, B6, and folate levels in an ischemic stroke population. | – | Keene et al., 2014 |
| ***FUT2*** | rs370871822 | G/A/C | < 0.01 (A/C) | – | – | synonymous variant | – | – | – |
| ***FUT2*** | rs114018037 | C/T | 0.01 (T) | 0.010 | Yes (2) | missense variant | Associated with non–Secretor status | 0.34 (tolerated) SIFT | Ferrer et al., 2009 |
| ***FUT2*** | rs492602 | A/G | 0.32 (G) | 0.439 | Yes (3) | synonymous variant | Associated with serum vitamin B12 concentration, | – | Koda et al., 2001; Velkova et al., 2017., |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Crohn's, and type 1 diabetes. | | McGovern et al., 2010 Smyth., 2011; Hazra et al., 2009 |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT2*** | rs281377 | C/T | 0.50 (T) | 0.308 | Yes (5) | synonymous variant | Associated with ulcerative colitis. | – | Hu et al., 2016 |
| ***FUT2*** | rs601338 | G/A | 0.32 (A) | 0.439 | Yes (5) | Stop gained | Associated with susceptibility to Norwalk, RV, and *E. coli*, serum vitamin B12 concentration, oral cancer risk. | – | Mottram et al., 2017; Bucardo., 2018; Günaydin et al., 2016; Lindesmith et al., 2003; Kelly et al., 1995; Velkova et al., 2017 Su et al., 2016; |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Currier et al., 2015; Payne t al., 2015; Nordgren et al., 2014 |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT2*** | rs1800027 | C/G/T | 0.04 (T) | 0.015 | Yes (1) | missense variant | Associated with non–Secretor status. | 0.06 (tolerated) SIFT | Ferrer et al., 2009 |
| ***FUT2*** | rs1800025 | G/A | 0.02 (A) | 0.086 | Yes (4) | missense variant | Associated with non–Secretor status. | 0.07 (tolerated) SIFT | Liu et al., 1998 |
| ***FUT2*** | rs602662 | G/A | 0.33 (A) | 0.439 | Yes (6) | missense variant | Associated with altered susceptibility *E. coli*, serum vitamin B12 concentration, Crohn's disease. | 0.982 (probably damaging) Polyphen–2 | Mottram et al., 2017; Koda et al., 2001; Velkova et al., 2017; McGovern et al., 2010; Smyth et al., 2011; |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Hazra et al., 2008; Nordgren et al., 2014 |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***FUT2*** | rs1799761 | CC/C | 0.04 (C) | 0.010 | Yes (2) | frameshift variant | – | – | – |
| ***FUT2*** | rs485186 | A/G | 0.33 (G) | 0.444 | Yes (7) | synonymous variant | Associated with Crohn's disease, type 1 diabetes susceptibility, and resistance to infection. | – | McGovern et al., 201; Smyth et al., 2011 |
| ***IFIH1*** | rs1990760 | C/T | 0.36 (T) | 0.101 | No | missense variant | Associated with type 1 diabetes mellitus, Autoimmune disease, and thyroid diseases etc. | 0.76 (tolerated) SIFT | Jermendy et al., 2018; Zhang et al., 2018; Yun et al., 2017 etc. |
| ***IFIH1*** | rs3747517 | T/C | 0.41 (T) | 0.535 | Yes (13) | missense variant | Associated with | 1 (tolerated) | Zhang et al., 2018; |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Autoimmune disease, protective for psoriatic arthritis, type 1 diabetes etc. | SIFT | Budu–Aggrey et al., 2017; RodrÃguez et al., 2015 etc. |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***IFIH1*** | rs10930046 | T/C | 0.19 (C) | 0.424 | Yes (22) | missense variant | Autoimmune disease, type 1 diabetes etc. | 1 (tolerated) SIFT | Zhang et al., 2018; RodrÃguez et al., 2015 etc. |
| ***IFIH1*** | rs984971 | A/C/G | 0.41 (A) | 0.798 | No | intron variant | Protective for psoriatic arthritis, defence against Candida infection etc. | – | Budu–Aggrey et al., 2017 |

| Gene | SNP ID (Alternative ID) | Alleles | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|
| ***TLR3*** | rs5743305 | T/A | 0.32 (A) | 0.318 | Yes (1) | regulatory region variant | Associated with the Risk of HCV Infection and HBV–related Diseases etc. | – | Geng et al., 2016 etc. |
| ***TLR3*** | rs13126816 | G/A | 0.21 (A) | 0.071 | No | intron variant | Associated with age–related macular degeneration, severe liver disease in HIV/HCV coinfected patients etc. | – | Ma et al., 2016; GuzmÃn–Fulgencio et al., 2015 etc. |
| ***TLR3*** | rs3775296 | C/A | 0.18 (A) | 0.131 | Yes (10) | splice region variant | Associated with eosinophilic, Hepatitis C virus infection esophagitis, risk of HCV infection and HBV–related diseases etc. | – | Ávila–Castellano et al., 2018; Gen et al., 2016 |

| Gene | SNP ID (Alternative ID) | Alleles | Blood group | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|---|
| ***TLR3*** | rs5743312 | C/T | 0.15 (T) | 0.081 | Yes (1) | intron variant | Associated with eosinophilic, Hepatitis C virus infection esophagitis, hepatitis B virus infection in Saudi Arabian patients etc. | – | Ávila–Castellano et al., 2018; Al–Anazi et al., 2017; Al–Qahtani et al., 2012 | |
| ***TLR3*** | rs3775292 | C/G | 0.17 (C) | 0.909 | Yes (7) | non – coding transcript exon variant | Associated with eosinophilic esophagitis etc. | – | Ávila–Castellano et al., 2018 | |
| ***TLR3*** | rs10025405 | A/G | 0.40 (G) | 0.278 | No | 3 prime UTR variant | Associated with resistance to HIV among highly exposed Caucasian intravenous drug users. | – | Huik et al., 2013 | |

| Gene | SNP ID (Alternative ID) | Alleles | Blood group | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen –2 score | References |
|---|---|---|---|---|---|---|---|---|---|---|
| ***ABO*** | rs56392308 | GGG/GG | A | 0.12 (T) | – | Yes (7) | frameshift variant | Associated with venous thrombosis, cardiovascular disease. | – | Heit et al., 2012; Zhang et al., 2012; Long et al., 2017 |
| ***ABO*** | rs8176748 | C/A/T | – | 0.27 (T) | 0.253 | Yes (49) | missense variant | – | 0 (deleterious) SIFT | – |
| ***ABO*** | rs8176747 | C/A/G | O | 0.26 (G) | 0.162 | Yes (46) | missense variant | Risk of advanced gastric precancerous lesions in *Helicobacter pylori* infected subjects. risk of gastric cancer, atrophic gastritis, and *Helicobacter pylori* infection. associated with susceptibility to severe *Plasmodium falciparum* malaria. | 0 (deleterious) SIFT | Rizzato et al., 2013; Nakao et al., 2011; Fry et al., 2008 |

| Gene | SNP ID (Alternative ID) | Alleles | Blood group | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|---|
| ***ABO*** | rs8176746 | G/A/T | B | 0.15 (T) | 0.162 | Yes (10) | missense variant | Associated with susceptibility to severe *Plasmodium falciparum* malaria. Pancreatic cancer risk: results from the pancreatic cancer cohort consortium. Risk of gastric cancer, atrophic gastritis, and *Helicobacter pylori* infection. Risk of pancreatic cancer in a | 0.39 (tolerated) SIFT | Fry et al., 2008; Wolpin et al., 2010; Nakao et al., 2011; Nakao et al., 2011 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Japanese population. | | |

| Gene | SNP ID (Alternative ID) | Alleles | Blood group | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|---|
| ***ABO*** | rs8176741 | G/A | – | 0.15 (A) | 0.167 | Yes (12) | synonymous variant | Associated with cystic fibrosis, advanced gastric precancerous lesions, ovarian cancer survival, intraocular pressure. | – | Taylor–Cousar et al., 2009; Rizzato et al., 2013; Cozzi et al., 2017, Gao et al., 2018 |
| ***ABO*** | rs1053878 | G/A | AB | 0.13 (A) | 0.798 | – | missense variant | Associated with the risk of advanced gastric precancerous lesions in *Helicobacter pylori* infected subjects. | 0.05 (deleterious) SIFT | Rizzato et al., 2013 |
| ***ABO*** | rs8176720 | T/A/C/G | – | 0.45 (C) | 0.500 | Yes (74) | synonymous variant | Risk of advanced gastric | – | Rizzato et al., 2013; Taylor– |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | precancerous lesions in *Helicobacter pylori* infected subjects. Is one of potential genetic modifiers of infection and cystic fibrosis lung disease severity. | | Cousar et al., 2009 |

| Gene | SNP ID (Alternative ID) | Alleles | Blood group | Minor allele (Risk) | MAF (LWK) | Tag–SNP (number of other SNPs tagged) | Variant type | Functional effect | SIFT/ PolyPhen–2 score | References |
|---|---|---|---|---|---|---|---|---|---|---|
| ***ABO*** | rs8176719 | –/C | O | 0.34 (C) | 0.288 | Yes (59) | frameshift variant | Associated with thrombosis risk in patients with paroxysmal nocturnal hemoglobinuria. Gastric cancer, atrophic gastritis, and *Helicobacter pylori* infection. | – | Long et al., 2017; Nakao et al., 2011 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***ABO*** | rs782023144 | CC/C | | – | – | – | splice donor variant | – | – | – |

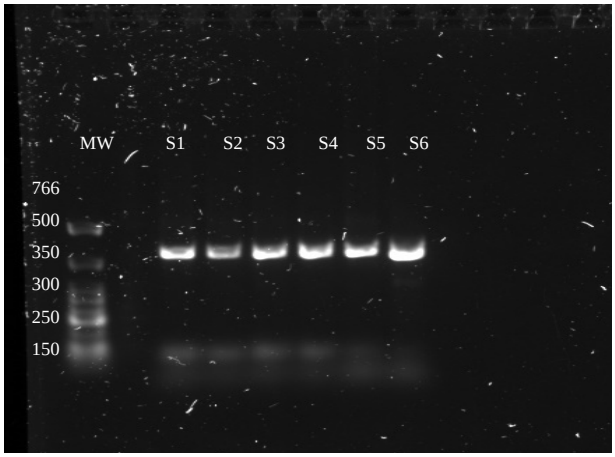
## Appendix D

Finalised list of 50 SNPs selected for genotyping cohort samples using iPLEX® MasssARRAY®.

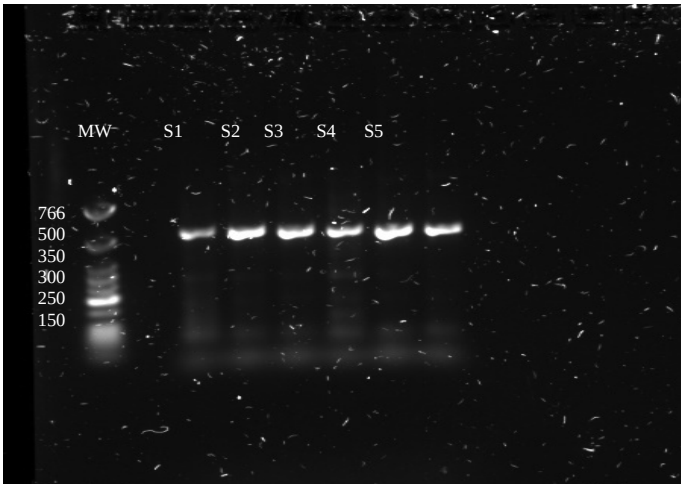
| Gene | SNP ID | Plex number |
|---|---|---|
| *FUT3* | rs3894326 | 1 |
| *FUT3* | rs28381968 | 1 |
| *FUT3* | rs28362466 | 1 |
| *FUT3* | rs3745635 | 1 |
| *ABO* | rs8176719 | 1 |
| *ABO* | rs8176748 | 1 |
| *FUT3* | rs28362458 | 1 |
| *FUT2* | rs1800025 | 1 |
| *ABO* | rs8176747 | 1 |
| *FUT3* | rs778986 | 1 |
| *FUT3* | rs28362463 | 1 |
| *FUT2* | rs1800027 | 1 |
| *FUT3* | rs761121538 | 1 |
| *FUT2* | rs1799761 | 1 |
| *ABO* | rs782023144 | 1 |
| *FUT2* | rs485186 | 1 |
| *FUT3* | rs812936 | 1 |
| *DDX58* | rs55789327 | 1 |
| *ABO* | rs563992308 | 1 |
| *TLR3* | rs3775292 | 1 |
| *FUT2* | rs492602 | 1 |
| *TLR3* | rs13126816 | 1 |
| *TLR3* | rs5743312 | 1 |
| *IFIH1* | rs1990760 | 1 |
| *TLR8* | rs2159377 | 1 |
| *DDX58* | rs669260 | 1 |
| *TLR3* | rs3775296 | 1 |
| *IFIH1* | rs984971 | 1 |
| *FUT2* | rs679574 | 1 |
| *TLR7* | rs179008 | 1 |
| *TLR7* | rs5935438 | 1 |

| Gene | SNP ID | Plex number |
|---|---|---|
| *TLR3* | rs10025405 | 2 |
| *FUT2* | rs516316 | 2 |
| *FUT3* | rs372597863 | 2 |
| *FUT2* | rs601338 | 2 |
| *ABO* | rs8176720 | 2 |
| *ABO* | rs1053878 | 2 |
| *FUT2* | rs281377 | 2 |
| *ABO* | rs8176746 | 2 |
| *FUT2* | rs602662 | 2 |
| *FUT2* | rs370871822 | 2 |
| *DDX58* | rs10813831 | 2 |
| *FUT2* | rs114018037 | 2 |
| *TLR7* | rs179007 | 2 |
| *ABO* | rs8176741 | 2 |
| *IFIH1* | rs10930046 | 2 |
| *TLR7* | rs3853839 | 2 |
| *IFIH1* | rs3747517 | 2 |
| *TLR3* | rs5743305 | 2 |
| *TLR7* | rs864058 | 2 |

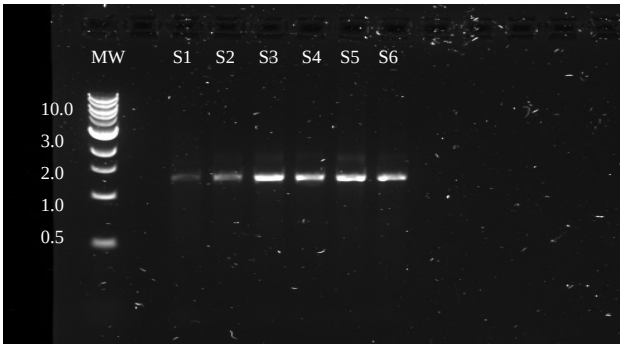
**Appendix E**



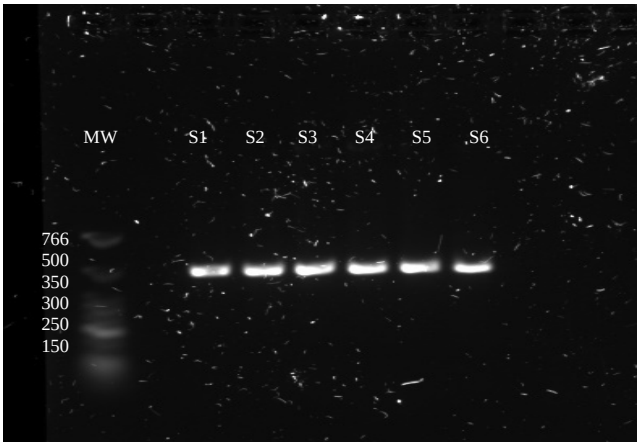

Optimization results of FUT2_CDS_1 primer set from Nordgren et al. (2014). Lane 1 is the Low molecular weight (LMW) DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 545 bp.



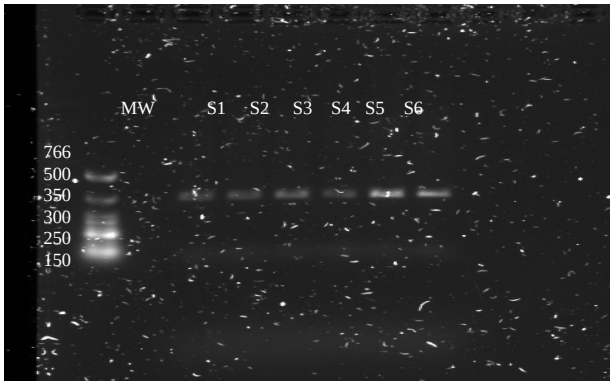

Optimization results of FUT2_CDS_2 primer set from Nordgren et al. (2014) Lane 1 is the LMW DNA ladder (KAPA, Inc) and lanes 3–8 are DNA samples (S1–S6). Expected band size was 513 bp.



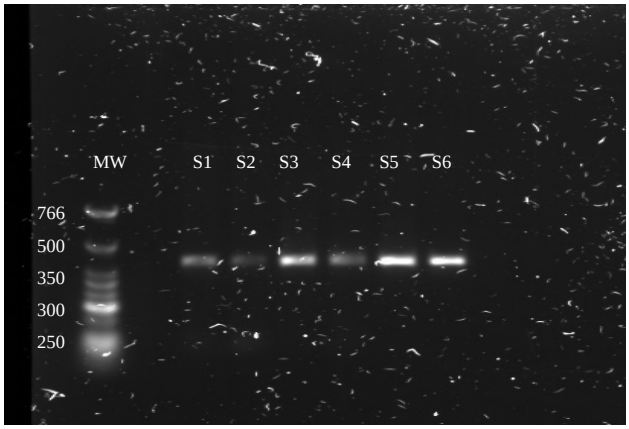

Optimization results of FUT2_2009 primer set from Ferrer et al. (2009). Lane 1 is the 1kb molecular weight DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 1220 bp.



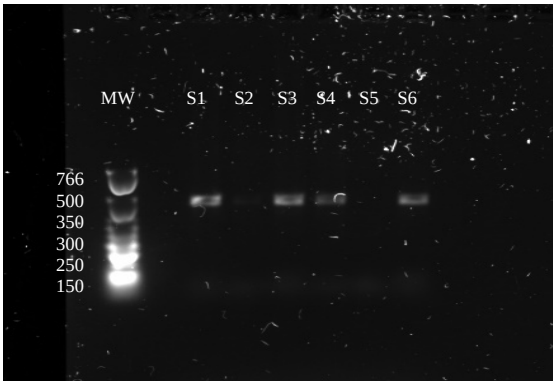

Optimization results of FUT3_Ex3_1 primer set from Nordgren et al. (2014) Lane 1 is the LMW DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 484 bp.




Optimization results of FUT3_Ex3_2[1] primer set from Nordgren et al. (2014) at 56°C annealing temperature optimized using OneTaq® PCR protocol. Lane 1 is the LMW DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 470 bp.




Optimization results of FUT3_Ex3_3 primer set from Nordgren et al. (2014). Lane 1 is the LMW DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 387 bp.




Optimization results of FUT3_Ex3_4[2] primer set from Nordgren et al. (2014). Lane 1 is the Low molecular weight DNA ladder (KAPA, Inc) and lanes 3–8 are the DNA samples (S1–S6). Expected band size was 460 bp.

**Appendix F**

Se phenotype frequency counts between seroconverters and non–seroconverters for the first time–point.

| Phenotype | Seroconverters (%) | Non–seroconverters (%) | Total cohort |
|---|---|---|---|
| **Se+** | 60 (42.5) | 33 (44.6) | 93 |
| **Weak or intermediate Se+** | 50 (35.5) | 19 (25.7) | 69 |
| **Se–** | 31 (22) | 22 (29.7) | 53 |
| **Total** | 141 (100) | 74 (100) | 215 |

Se phenotype frequency counts between seroconverters and non–seroconverters for the second time–point.

| Se status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Se+** | 55 (44.7) | 31 (47) | 86 |
| **Weak or intermediate Se+** | 39 (31.7) | 14 (21.2) | 53 |
| **Se–** | 29 (23.6) | 21 (31.8) | 50 |
| **Total** | 123 (100) | 66 (100) | 189 |

Le phenotype frequency counts between seroconverters and non–seroconverters.

| Le status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Le A–B+** | 92 (65.2) | 44 (59.5) | 136 |
| **Le A+B–** | 33 (23.4) | 21 (28.4) | 54 |
| **Le A–B–** | 16 (11.3) | 9 (12.1) | 25 |
| **Total** | 141 (100) | 74 | 215 |

Le phenotype frequency counts between seroconverters and non–seroconverters.

| Le status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Le A–B+** | 73 (59.3) | 34 (51.5) | 107 |
| **Le A+B–** | 19 (15.4) | 19 (28.8) | 38 |
| **Le A–B–** | 31 (25.2) | 13 (19.7) | 44 |
| **Total** | 123 (100) | 66 | 189 |

Se phenotype counts between seroconverters and non–seroconverters.

| Secretor status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Secretor+** | 70 (73.7) | 26 (60.5) | 96 |
| **weak Secretor** | 11 (11.6) | 8 (18.6) | 19 |
| **non–Secretor** | 14 (14.7) | 9 (20.9) | 23 |
| **Total** | 95 (100) | 43 (100) | 138 |

Le antigen counts between seroconverters and non–seroconverters.

| Lewis status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Le+** | 71 (74.7) | 39 (90.7) | 110 |
| **Le–** | 24 (25.3) | 4 (9.3) | 28 |
| **Total** | 95 (100) | 43 (100) | 138 |

Le phenotype counts (predicted from combined *FUT2* and *FUT3* genetic data). Compared between seroconverters and non–seroconverters.

| Lewis status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Le (a–b+)** | 55 (57.9) | 23 (53.5) | 78 |
| **Le (a–b+) or Le (a+b+)** | 9 (9.5) | 8 (18.6) | 17 |
| **Le (a+b–)** | 7 (7.3) | 8 (18.6) | 15 |
| **Le (a–b–)** | 24 (25.3) | 4 (9.3) | 28 |
| **Total** | 95 (100) | 43 (100) | 138 |

Combined Le phenotype counts (predicted from combined *FUT2* and *FUT3* genetic data). Compared between seroconverters and non–seroconverters.

| Lewis status | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **Le(a–b+)** | 64 (67.4) | 31 (72.1) | 95 |
| **Le(a+b–)** | 7 (7.3) | 8 (18.6) | 15 |
| **Le(a–b–)** | 24 (25.3) | 4 (9.3) | 28 |
| **Total** | 95 (100) | 43 (100) | 138 |

ABO blood type counts between seroconverters and non–seroconverters.

| Blood type | Seroconverters (%) | Non–seroconverters (%) | Total |
|---|---|---|---|
| **A** | 29 (30.9) | 16 (36.4) | 45 |
| **B** | 12 (12.8) | 5 (11.4) | 17 |
| **O** | 48 (51) | 22 (50) | 70 |
| **AB** | 5 (5.3) | 1 (2.2) | 6 |
| **Total** | 94 (100) | 44 (100) | 138 |

# 488744:Thesis_(Plegiarism_3).docx

ORIGINALITY REPORT

| 15% | 7% | 8% | 7% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to University of Witwatersrand<br>Student Paper | 5% |
| 2 | www.mdpi.com<br>Internet Source | 1% |
| 3 | www.ncbi.nlm.nih.gov<br>Internet Source | <1% |
| 4 | d-nb.info<br>Internet Source | <1% |
| 5 | Catherine Yen, Jacqueline E. Tate, Manish M. Patel, Margaret M. Cortese et al. "Rotavirus vaccines", Human Vaccines, 2014<br>Publication | <1% |
| 6 | hdl.handle.net<br>Internet Source | <1% |
| 7 | Filemón Bucardo, Johan Nordgren, Yaoska Reyes, Fredman Gonzalez, Sumit Sharma, Lennart Svensson. "The Lewis A phenotype is a restriction factor for Rotateq and Rotarix vaccine-take in Nicaraguan children", Scientific Reports, 2018 | <1% |

Publication

| | | |
|---|---|---|
| 8 | Benjamin Lee, Dorothy M Dickson, Allan C deCamp, E Ross Colgate et al. "Histoblood Group Antigen Phenotype Determines Susceptibility to Genotype-specific Rotavirus Infections and Impacts Measures of Rotavirus Vaccine Efficacy", The Journal of Infectious Diseases, 2018<br>Publication | <1 % |
| 9 | www.embopress.org<br>Internet Source | <1 % |
| 10 | Pawapol Kongchum, Eran Sandel, Sigal Lutzky, Eric M. Hallerman, Gideon Hulata, Lior David, Yniv Palti. "Association between IL-10a single nucleotide polymorphisms and resistance to cyprinid herpesvirus-3 infection in common carp (Cyprinus carpio)", Aquaculture, 2011<br>Publication | <1 % |
| 11 | Sumit Sharma, Marie Hagbom, Lennart Svensson, Johan Nordgren. "The Impact of Human Genetic Polymorphisms on Rotavirus Susceptibility, Epidemiology, and Vaccine Take", Viruses, 2020<br>Publication | <1 % |
| 12 | Xiaoman Sun, Lei Dang, Dandi Li, Jianxun Qi et al. "Structural Basis of Glycan Recognition in Globally Predominant Human P[8] Rotavirus", Virologica Sinica, 2019 | <1 % |

| | | |
|---|---|---|
| | Publication | |
| 13 | link.springer.com<br>Internet Source | <1 % |
| 14 | researchspace.ukzn.ac.za<br>Internet Source | <1 % |
| 15 | apps.who.int<br>Internet Source | <1 % |
| 16 | wwwnc.cdc.gov<br>Internet Source | <1 % |
| 17 | www.tandfonline.com<br>Internet Source | <1 % |
| 18 | Robin Uchiyama, Benoit Chassaing, Benyue Zhang, Andrew T Gewirtz. "MyD88-mediated TLR signaling protects against acute rotavirus infection while inflammasome cytokines direct Ab response", Innate Immunity, 2014<br>Publication | <1 % |
| 19 | Parisa Mohseni, Davood Khalili, Shirin Djalalinia, Farshad Farzadfar, Yadollah Mehrabi. "Factors Related to Serum Lipids Using Multilevel Quantile Mixed Model: Analysis of Nationwide STEPs Survey 2016", Research Square Platform LLC, 2021<br>Publication | <1 % |
| 20 | Jacques Le Pendu, Nathalie Ruvoën-Clouet. "Fondness for sugars of enteric viruses | <1 % |

| | | |
|---|---|---|
| | confronts them with human glycans genetic diversity", Human Genetics, 2019<br>Publication | |
| 21 | gupea.ub.gu.se<br>Internet Source | <1 % |
| 22 | Michelle J Groome, Lee Fairlie, Julie Morrison, Alan Fix et al. "Safety and immunogenicity of a parenteral trivalent P2-VP8 subunit rotavirus vaccine: a multisite, randomised, double-blind, placebo-controlled trial", The Lancet Infectious Diseases, 2020<br>Publication | <1 % |
| 23 | rheumatology.oxfordjournals.org<br>Internet Source | <1 % |
| 24 | Roma Chilengi, Michelo Simuyandi, Lauren Beach, Katayi Mwila et al. "Association of Maternal Immunity with Rotavirus Vaccine Immunogenicity in Zambian Infants", PLOS ONE, 2016<br>Publication | <1 % |
| 25 | files.www.quotientbd.com<br>Internet Source | <1 % |
| 26 | research.one.surgery<br>Internet Source | <1 % |
| 27 | academic.oup.com<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 28 | Günaydın, Gökçe, Johan Nordgren, Sumit Sharma, and Lennart Hammarström. "Association of elevated rotavirus-specific antibody titers with HBGA secretor status in Swedish individuals: The FUT2 gene as a putative susceptibility determinant for infection", Virus Research, 2016.<br>Publication | <1 % |
| 29 | Yun Ji Hong, Sang Mee Hwang, Taek Soo Kim, Eun Young Song, Kyoung Un Park, Junghan Song, Kyou-Sup Han. " Significance of Lewis Phenotyping Using Saliva and Gastric Tissue: Comparison with the Lewis Phenotype Inferred from and Genotypes ", BioMed Research International, 2014<br>Publication | <1 % |
| 30 | etd.aau.edu.et<br>Internet Source | <1 % |
| 31 | Submitted to Kenyatta University<br>Student Paper | <1 % |
| 32 | Submitted to Roehampton University<br>Student Paper | <1 % |
| 33 | M. Kathleen Glynn, Cindy R. Friedman, Benjamin D. Gold, Bhawna Khanna et al. " Seroincidence of Infection in a Cohort of Rural Bolivian Children: Acquisition and Analysis of | <1 % |

Possible Risk Factors ", Clinical Infectious Diseases, 2002
Publication

| | | |
|---|---|---|
| 34 | Mim A. Bower, Beatrice A. McGivney, Michael G. Campana, Jingjing Gu et al. "The genetic origin and history of speed in the Thoroughbred racehorse", Nature Communications, 2012<br>Publication | <1 % |
| 35 | Lung-Chih Yu. "Polymorphism and distribution of the Secretor alpha(1,2)-fucosyltransferase gene in various Taiwanese populations", Transfusion, 10/2001<br>Publication | <1 % |
| 36 | Stephen Henry. "Lewis Histo-Blood Group System and Associated Secretory Phenotypes", Vox Sanguinis, 10/1995<br>Publication | <1 % |
| 37 | bmcpublichealth.biomedcentral.com<br>Internet Source | <1 % |
| 38 | Francesco Addabbo, Valentina Baglioni, Anette Schrag, Markus J Schwarz et al. "Anti - dopamine D2 receptor antibodies in chronic tic disorders", Developmental Medicine & Child Neurology, 2020<br>Publication | <1 % |
| 39 | centrodinoferrari.com<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 40 | "Abstract", Pharmacoepidemiology and Drug Safety, 2017<br>Publication | <1 % |
| 41 | Daniels, . "ABO, H, and Lewis Systems", Human Blood Groups Daniels/Human Blood Groups, 2013.<br>Publication | <1 % |
| 42 | Submitted to University of Newcastle upon Tyne<br>Student Paper | <1 % |
| 43 | Abigail W. Bigham. "Host Genetic Risk Factors for West Nile Virus Infection and Disease Progression", PLoS ONE, 09/15/2011<br>Publication | <1 % |
| 44 | Cigdem S. Bayrak, Aayushee Jain, David Stein, Kumardeep Chaudhary et al. "Identification of Discriminative Gene-level and Protein-level Features Associated with Gain-of-Function and Loss-of-Function Mutations", Cold Spring Harbor Laboratory, 2021<br>Publication | <1 % |
| 45 | Submitted to De Montfort University<br>Student Paper | <1 % |
| 46 | Jbiomedsci.Biomedcentral.Com<br>Internet Source | <1 % |
| 47 | corescholar.libraries.wright.edu<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 48 | Laura Settier Ramírez. "Envases activos portadores de microorganismos para la bioconservación de alimentos", Universitat Politecnica de Valencia, 2021<br>Publication | <1 % |
| 49 | Seheri Luyanda Mapaseka, John Barr Dewar, Lize van der Merwe, Annelise Geyer et al. "Prospective Hospital ‐ Based Surveillance to Estimate Rotavirus Disease Burden in the Gauteng and North West Province of South Africa during 2003–2005", The Journal of Infectious Diseases, 2010<br>Publication | <1 % |
| 50 | Submitted to University of Mustansiriyah<br>Student Paper | <1 % |
| 51 | discovery.ucl.ac.uk<br>Internet Source | <1 % |
| 52 | patents.google.com<br>Internet Source | <1 % |
| 53 | studentsrepo.um.edu.my<br>Internet Source | <1 % |
| 54 | www.allindianpatents.com<br>Internet Source | <1 % |
| 55 | www.nature.com<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 56 | Michelle Goldsammler, Sangita K. Jindal, Amanda Kallen, Natu Mmbaga, Lubna Pal. "Blood type predicts live birth in the infertile population", Journal of Assisted Reproduction and Genetics, 2015<br>Publication | <1 % |
| 57 | Submitted to University of Edinburgh<br>Student Paper | <1 % |
| 58 | Xiaoman Sun, Dandi Li, Zhaojun Duan. "Structural Basis of Glycan Recognition of Rotavirus", Frontiers in Molecular Biosciences, 2021<br>Publication | <1 % |
| 59 | www.frontiersin.org<br>Internet Source | <1 % |
| 60 | "Scientific Section", Transfusion, 09/2008<br>Publication | <1 % |
| 61 | Ariane F Busso-Lopes, Cesar Rivera, Leandro X Neves, Daniela C Granato et al. "Wiring multiple microenvironment proteomes uncovers the biology in head and neck cancer", Cold Spring Harbor Laboratory, 2021<br>Publication | <1 % |
| 62 | Marco Morelli, Kristen M. Ogden, John T. Patton. "Silencing the alarms: Innate immune antagonism by rotavirus NSP1 and VP3", Virology, 2015 | <1 % |

| | | |
|---|---|---|
| | Publication | |
| 63 | v3r.esp.org<br>Internet Source | <1% |
| 64 | www.bmj.com<br>Internet Source | <1% |
| 65 | zzz.bwh.harvard.edu<br>Internet Source | <1% |
| 66 | Akiko Iwasaki, Malik Peiris. "Innate immunity", Wiley, 2013<br>Publication | <1% |
| 67 | Martin T. Schechter, Steffanie A. Strathdee, Peter G.A. Cornelisse, Sue Currie et al. "Do needle exchange programmes increase the spread of HIV among injection drug users?: an investigation of the Vancouver outbreak", AIDS, 1999<br>Publication | <1% |
| 68 | Nordgren, J., S. Sharma, F. Bucardo, W. Nasir, G. Gunayd n, D. Ouermi, L. W. Nitiema, S. Becker-Dreps, J. Simpore, L. Hammarstrom, G. Larson, and L. Svensson. "Both Lewis and Secretor Status Mediate Susceptibility to Rotavirus Infections in a Rotavirus Genotype-Dependent Manner", Clinical Infectious Diseases, 2014.<br>Publication | <1% |

| | | |
|---|---|---|
| 69 | R.E. Broadberry, M.L. Lin-Chu. "The Lewis Blood Group System among Chinese in Taiwan", Human Heredity, 1991<br>Publication | <1 % |
| 70 | www.cdc.gov<br>Internet Source | <1 % |
| 71 | "Immunity Against Mucosal Pathogens", Springer Science and Business Media LLC, 2008<br>Publication | <1 % |
| 72 | rupress.org<br>Internet Source | <1 % |
| 73 | www.jbc.org<br>Internet Source | <1 % |
| 74 | Submitted to Bolton Institute of Higher Education<br>Student Paper | <1 % |
| 75 | Rosella Mollicone, Annem Dalix, Anita Jacobsson, Bo E Samuelsson et al. "Red cell H-deficient, salivary ABH secretor phenotype of Reunion island. Genetic control of the expression of H antigen in the skin", Glycoconjugate Journal, 1988<br>Publication | <1 % |
| 76 | Submitted to Savitribai Phule Pune University<br>Student Paper | <1 % |

| | | |
|---|---|---|
| 77 | Submitted to University of San Francisco<br>Student Paper | <1% |
| 78 | mafiadoc.com<br>Internet Source | <1% |
| 79 | pcm.amegroups.com<br>Internet Source | <1% |
| 80 | www.scielo.br<br>Internet Source | <1% |
| 81 | M. Hagbom, L. Svensson. "Rotavirus Disease Mechanisms", Elsevier BV, 2016<br>Publication | <1% |
| 82 | Submitted to Pwani University<br>Student Paper | <1% |
| 83 | Wen, Xiaobo, Dianjun Cao, Ronald W Jones, Yasutaka Hoshino, and Lijuan Yuan. "Tandem truncated rotavirus VP8* subunit protein with T cell epitope as non-replicating parenteral vaccine is highly immunogenic", Human Vaccines & Immunotherapeutics, 2015.<br>Publication | <1% |
| 84 | Xiaoning Liu, Lei Meng, Juansheng Li, Xinfeng Liu et al. "Etiological epidemiology of viral diarrhea on the basis of sentinel surveillance in children younger than 5 years in Gansu, northwest China, 2009-2013", Journal of Medical Virology, 2015 | <1% |

Publication

| | | |
|---|---|---|
| 85 | Yan Cheng, Stephane Guindon, Allen Rodrigo, Seng Gee Lim. "Increased viral quasispecies evolution in HBeAg seroconverter patients treated with oral nucleoside therapy", Journal of Hepatology, 2013<br>Publication | <1 % |
| 86 | etheses.whiterose.ac.uk<br>Internet Source | <1 % |
| 87 | tesis.ipn.mx<br>Internet Source | <1 % |
| 88 | B-S Sheu. "Host gastric Lewis expression determines the bacterial density of Helicobacter pylori in babA2 genopositive infection", Gut, 2003<br>Publication | <1 % |
| 89 | M Luisetto, Ilman Ahnaf, Khan Farhan Ahmad, Edbey Khaled et al. "COVID-19 immunologic and toxicological implication: Innate immune sensor and immune escape", Archives of Pharmacy and Pharmaceutical Sciences, 2021<br>Publication | <1 % |
| 90 | Ting Zou, Jiye Yin, Wei Zheng, Ling Xiao et al. "Rho GTPases: RAC1 polymorphisms affected platinum-based chemotherapy toxicity in lung cancer patients", Cancer Chemotherapy and Pharmacology, 2016<br>Publication | <1 % |

| | | |
|---|---|---|
| 91 | file.scirp.org<br>Internet Source | <1 % |
| 92 | iai.asm.org<br>Internet Source | <1 % |
| 93 | Glycoscience Biology and Medicine, 2015.<br>Publication | <1 % |
| 94 | Jarvi, S.I.. "Protective efficacy of a recombinant subunit West Nile virus vaccine in domestic geese (Anser anser)", Vaccine, 20081003<br>Publication | <1 % |
| 95 | Ramani, Sasirekha, and Robert L. Atmar. "Acute Gastroenteritis Viruses", Mucosal Immunology, 2015.<br>Publication | <1 % |
| 96 | Tereza Cristina de Oliveira Corvelo, Rosane do Socorro Pompeu de Loiola, Délia Cristina Figueira Aguiar, Gyselly de Cássia Bastos de Matos et al. "The Lewis Histo-Blood Group System: Molecular Analysis of the 59T>G, 508G>A, and 1067T>A Polymorphisms in an Amazonian Population", PLoS ONE, 2013<br>Publication | <1 % |
| 97 | dmauldin.gdxbase.org<br>Internet Source | <1 % |
| 98 | pub.uni-bielefeld.de | |

| | | |
|---|---|---|
| | Internet Source | <1% |
| 99 | pure.southwales.ac.uk<br>Internet Source | <1% |
| 100 | s3.amazonaws.com<br>Internet Source | <1% |
| 101 | theses.gla.ac.uk<br>Internet Source | <1% |
| 102 | Alessandra Bridi. "Extracellular vesicles secreted by bovine embryos produced >i/i< and >i/i<: miRNAs content and molecular effects in endometrial and luteal tissues", Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2021<br>Publication | <1% |
| 103 | Andres Garcia-Campos, Alessandra Ravidà, D. Linh Nguyen, Krystyna Cwiklinski et al. "Tegument Glycoproteins and Cathepsins of Newly Excysted Juvenile Fasciola hepatica Carry Mannosidic and Paucimannosidic N-glycans", PLOS Neglected Tropical Diseases, 2016<br>Publication | <1% |
| 104 | Erwin Andreas Scharberg, Coral Olsen, Peter Bugert. "The H blood group system", Immunohematology, 2019 | <1% |

Publication

| | | |
|---|---|---|
| 105 | Joyce Yongxu Huang, Renwei Wang, Yu-Tang Gao, Jian-Min Yuan. "ABO blood type and the risk of cancer – Findings from the Shanghai Cohort Study", PLOS ONE, 2017<br>Publication | <1 % |
| 106 | Martina Durner. "Can We Increase the Likelihood of Success for Future Association Studies in Epilepsy?", Epilepsia, 10/2006<br>Publication | <1 % |
| 107 | Michelle M. Arnold. "Rotavirus Vaccines: Why Continued Investment in Research Is Necessary", Current Clinical Microbiology Reports, 2018<br>Publication | <1 % |
| 108 | Ng, C. S., H. Kato, and T. Fujita. "Recognition of viruses in the cytoplasm by RLRs and other helicases--how conformational changes, mitochondrial dynamics and ubiquitination control innate immune responses", International Immunology, 2012.<br>Publication | <1 % |
| 109 | openaccess.leidenuniv.nl<br>Internet Source | <1 % |
| 110 | repositorio.ufmg.br<br>Internet Source | <1 % |
| 111 | twitter.com | |

Internet Source

<1%

112 www.neb.sg
Internet Source

<1%

113 "Next Generation Sequencing and Whole Genome Selection in Aquaculture", Wiley, 2011
Publication

<1%

114 "The Molecular Immunology of Complex Carbohydrates-3", Springer Science and Business Media LLC, 2011
Publication

<1%

115 A. Said Ahmed, S. Rahmatalla, R. Bortfeldt, D. Arends, M. Reissmann, G. A. Brockmann. "Milk protein polymorphisms and casein haplotypes in Butana cattle", Journal of Applied Genetics, 2016
Publication

<1%

116 Burrell, Andrew S., Todd R. Disotell, and Christina M. Bergey. "The use of museum specimens with high-throughput DNA sequencers", Journal of Human Evolution, 2015.
Publication

<1%

117 Cristina Pérez-Ramírez, Marisa Cañadas-Garre, Ahmed Alnatsha, Eduardo Villar et al. "Impact of DNA repair, folate and glutathione

<1%

gene polymorphisms on risk of non small cell lung cancer", Pathology - Research and Practice, 2018
Publication

---

118 Desselberger, Ulrich. "Rotaviruses", Virus Research, 2014. <1%
Publication

---

119 E Clarke, U Desselberger. "Correlates of protection against human rotavirus disease and the factors influencing protection in low-income settings", Mucosal Immunology, 2014 <1%
Publication

---

120 Jenny Svärd. "Nuclear Receptor-Mediated Induction of CYP450 by Antiretrovirals: Functional Consequences of NR1I2 (PXR) Polymorphisms and Differential Prevalence in Whites and Sub-Saharan Africans :", JAIDS Journal of Acquired Immune Deficiency Syndromes, 12/2010 <1%
Publication

---

121 Submitted to Nottingham Trent University <1%
Student Paper

---

122 S. Metkar, S. Awasthi, E. Denamur, K. S. Kim, S. C. Gangloff, S. Teichberg, A. Haziot, J. Silver, S. M. Goyert. "Role of CD14 in Responses to Clinical Isolates of Escherichia coli: Effects of K1 Capsule Expression", Infection and Immunity, 2007 <1%

Publication

| | | |
|---|---|---|
| 123 | T.S. Rao, Rashmi Arora, Ajay Khera, Jacqueline E. Tate, Umesh Parashar, Gagandeep Kang. "Insights from global data for use of rotavirus vaccines in India", Vaccine, 2014<br>Publication | <1 % |
| 124 | Yue Qiu, Yuan Hu, Zuo-Yang Zhang, Lei Ye et al. "Genetic association of osteopontin (OPN) and its receptor CD44 genes with susceptibility to Chinese gastric cancer patients", Journal of Cancer Research and Clinical Oncology, 2014<br>Publication | <1 % |
| 125 | digitalcommons.du.edu<br>Internet Source | <1 % |
| 126 | dokumen.pub<br>Internet Source | <1 % |
| 127 | egrove.olemiss.edu<br>Internet Source | <1 % |
| 128 | escholarship.umassmed.edu<br>Internet Source | <1 % |
| 129 | idoc.pub<br>Internet Source | <1 % |
| 130 | molgenis26.gcc.rug.nl<br>Internet Source | <1 % |
| | open.metu.edu.tr | |

| | | |
|---|---|---|
| 131 | Internet Source | <1% |
| 132 | repository.lib.cuhk.edu.hk<br>Internet Source | <1% |
| 133 | uknowledge.uky.edu<br>Internet Source | <1% |
| 134 | uwspace.uwaterloo.ca<br>Internet Source | <1% |
| 135 | www.medrxiv.org<br>Internet Source | <1% |
| 136 | www.nellis.af.mil<br>Internet Source | <1% |
| 137 | B SHAZ. "Lewis, I and P Blood Group Systems", Transfusion Medicine and Hemostasis, 2009<br>Publication | <1% |
| 138 | Daisuke Nishizawa, Masako Iseki, Hideko Arita, Kazuo Hanaoka et al. "Genome-wide association study identifies candidate loci associated with chronic pain and postherpetic neuralgia", Molecular Pain, 2021<br>Publication | <1% |
| 139 | G. Cilla, M. Montes, M. Gomariz, L. Piñeiro, E. Pérez-Trallero. "Rotavirus genotypes in children in the Basque Country (northern Spain) over a 13-year period (July 1996–June | <1% |

2009)", European Journal of Clinical Microbiology & Infectious Diseases, 2010
Publication

---

140 Kubota, T., A. Kumagai, H. Ito, S. Furukawa, Y. Someya, N. Takeda, K. Ishii, T. Wakita, H. Narimatsu, and H. Shirato. "Structural basis for the recognition of Lewis antigens by genogroup I norovirus", Journal of Virology, 2012.
Publication

<1 %

---

141 Lynda Mottram, Gudrun Wiklund, Göran Larson, Firdausi Qadri, Ann-Mari Svennerholm. "FUT2 non-secretor status is associated with altered susceptibility to symptomatic enterotoxigenic Escherichia coli infection in Bangladeshis", Scientific Reports, 2017
Publication

<1 %

---

142 Setsuko Ameno, Hiroko Kimura, Kiyoshi Ameno, Xia Zhang, Hiroshi Kinoshita, Takako Kubota, Iwao Ijiri. "Lewis and Secretor Gene Effects on Lewis Antigen and Postnatal Development of Lewis Blood Type", Neonatology, 2001
Publication

<1 %

---

143 Yujia Zhang, Rachael Garner, Sana Salehi, Marianna La Rocca, Dominique Duncan. "Association between ABO blood types and

<1 %

coronavirus disease 2019 (COVID-19), genetic associations, and underlying molecular mechanisms: a literature review of 23 studies", Annals of Hematology, 2021
Publication

| | | |
|---|---|---|
| 144 | dspace.stir.ac.uk<br>Internet Source | <1% |
| 145 | Blood Cell Biochemistry, 1995.<br>Publication | <1% |
| 146 | David. J. Anstee. "The relationship between blood groups and disease", Blood, 2010<br>Publication | <1% |
| 147 | Helmut Schenkel-Brunner. "Human Blood Groups", Springer Science and Business Media LLC, 2000<br>Publication | <1% |
| 148 | Sun, Xiaoman, Nijun Guo, Dandi Li, Miao Jin, Yongkang Zhou, Guangcheng Xie, Lili Pang, Qing Zhang, Youde Cao, and Zhao-Jun Duan. "Binding specificity of P[8] VP8* proteins of rotavirus vaccine strains with histo-blood group antigens", Virology, 2016.<br>Publication | <1% |
| 149 | kclpure.kcl.ac.uk<br>Internet Source | <1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | Off |
| Exclude bibliography | On | | |