



Strathprints Institutional Repository

Gale, P. and Hill, A. and Kelly, L. and Bassett, J. and McClure, P. and Le Mark, Y. and Soumpasis, I. (2014) Applications of omics approaches to the development of microbiological risk assessment using RNA virus dose–response models as a case study. Journal of Applied Microbiology, 117 (6). pp. 1537-1548. ISSN 1364-5072 , <http://dx.doi.org/10.1111/jam.12656>

This version is available at <http://strathprints.strath.ac.uk/50158/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: strathprints@strath.ac.uk

Review: Applications of omics approaches to the development of microbiological risk assessment using RNA virus dose response models as a case study

Paul Gale¹, Andrew Hill¹, Louise Kelly¹, John Bassett², Peter McClure³, Yvan Le Marc³ and Ilias Soumpasis³

¹Animal Health and Veterinary Laboratories Agency, Weybridge, New Haw,
Addlestone, Surrey, KT15 3NB, UK

²John Bassett Consulting Ltd, Bedford, MK40 3DJ, UK

³Unilever Safety and Environmental Assurance Centre, Colworth Science Park,
Sharnbrook, Bedford, MK44 1LQ, UK

Abbreviated Title: Applying omics to dose response **relationships**

Correspondence: Dr Paul Gale, Animal Health and Veterinary Laboratories Agency,
Weybridge, New Haw, Addlestone, Surrey, KT15 3NB, UK. E-mail:

paul.gale@ahvla.gsi.gov.uk

Keywords: omics, norovirus, genetic susceptibility, dose response, infection

SUMMARY

The last decade has seen a huge increase in the amount of “omics” data available and in our ability to interpret those data. The aim of this paper is to consider how omics techniques can be used to improve and refine microbiological risk assessment, using dose response models for RNA viruses, with particular reference to norovirus through the oral route as the case study. The dose response model for initial infection in the gastrointestinal tract is broken down into the component steps at the molecular level and the feasibility of assigning probabilities to each step assessed. The molecular mechanisms are not sufficiently well understood at present to enable quantitative estimation of probabilities on the basis of omics data. At present, the great strength of gene sequence data appears to be in giving information on the distribution and proportion of susceptible genotypes (for example due to the presence of the appropriate pathogen-binding receptor) in the host population rather than in predicting specificities from the amino acid sequences concurrently obtained. The nature of the mutant spectrum in RNA viruses greatly complicates the application of omics approaches to development of mechanistic dose response models and prevents prediction of risks of disease progression (given infection has occurred) at the level of the individual host. However, molecular markers in the host and virus may enable more broad predictions to be made about the **consequences of exposure** in a population. In an alternative approach, **comparing the results of deep sequencing of RNA viruses in the faeces/vomit from donor humans with those from their** infected recipients may enable direct estimates of the average probability of infection per virion to be made.

INTRODUCTION

Classical microbiological risk assessments (MRAs) consist of four steps: hazard identification, hazard characterisation, exposure assessment and risk characterisation (CAC 1999). There is now a wealth of literature on this subject and MRAs have been developed for many pathogen/commodity combinations spanning a wide range of food types from ready-to-eat produce to animal products cooked in the home (e.g. Cassin et al., 1998; Gale, 2005; Nauta et al., 2007). Despite their potential contribution to MRA, the use of “omics” data in MRA has to date been limited. Therefore, the aims of this paper are to elucidate how omics data and the techniques used to generate and analyse such data can be used to improve and/or complement MRA, to identify where barriers may exist to the incorporation of omics data into MRA, and to consider where future omics data may prove useful. The focus is on the human oral dose **response relationship (DRR)** not only because some of the molecular processes governing the pathogen/host interaction are established, but also because there are known data gaps in the DRR for many pathogen/commodity combinations. The DRR is used to translate the exposure through the oral route into a risk of infection/disease. In classical MRAs, the DRR is based on human studies where different doses are fed to groups of volunteers (Ward et al., 1986) and the proportion infected (or ill) at each dose is plotted against the dose. Mathematical models are then fitted to those data and used to predict the probability of infection for doses not included in the volunteer study (Teunis et al., 2008; Haas et al., 1993). Over the last decade there has been a huge increase in the amount of omics data available and in our ability to interpret those data using bioinformatics techniques. For genomics, the introduction of high-throughput next generation sequencing (NGS) technologies in 2005 greatly increased the capability to sequence whole genomes

from individual living organisms rapidly and cheaply, providing information on variation across the host population (Pareek et al., 2011). NGS has also enabled the “deep-sequencing” of viral genomes and effectively gives the sequence of every virus in the **exposure** dose in terms of a mutant spectrum (Bull et al., 2012). Proteomics allows the identification of those proteins important to the cell in any given situation (Neidhardt, 2011). The relative proportions of proteins represent the protein signature which changes in response to a challenge and could be used to identify proteins expressed in a host cell in response to infection by a virus, for example. The great potential of proteomics is to be able to detect previously unknown proteins and assign their cellular location and their contacts with other proteins (Marchadier et al., 2011). Another approach which is considered here is “glycoproteomics”. Many proteins, particularly those on cell or pathogen surfaces and in the host immune system, have sugar units, called glycans, added to them. Indeed these glycans serve as receptors on cell surfaces for a number of pathogens including norovirus (NoV). Glycoproteomics characterises proteins containing glycans. Glycosylation (the synthesis of glycans) still remains practically unexplored at the proteome scale (Doerr 2012).

There are relatively few pathogens for which human volunteer dose response data are available, and in general such data are absent for newly emerged pathogens. While animal models have been used together with calibration from human outbreak data (Teunis et al., 2004), other techniques are needed to obtain DRR parameters.

Furthermore, account needs to be taken of the more susceptible members of the human population. The **focus of the first part of this paper is on the development of a mechanistic dose response model for estimating the probability of initial infection through the oral route.** The second part of this paper considers the exciting possibility that NGS data of virus isolated from donor and recipient human cases may facilitate a

direct estimation of the probability of initial infection. The final part addresses the problems in applying omics approaches to predicting disease progression and spread to other tissues once infection has established in the gastrointestinal tract. NoV is selected as an example to demonstrate the concepts because there are dose response data from human volunteer studies (Teunis et al., 2008) together with a wealth of molecular information on receptor binding for this pathogen.

AN OMICS-BASED APPROACH FOR ESTIMATING THE PROBABILITY OF INITIAL INFECTION BY VIRUS IN ORAL EXPOSURE

Infection by an oral pathogen is defined as the multiplication of organisms within the host, followed by excretion (Haas et al 1999). Biologically, the infection process for a virus such as NoV is comprised of a number of steps as set out in Table 1. Each pathogen particle of genomic form i , has a probability p_{ij} of initiating infection in a host individual of genomic form, j . The probability, p_{ij} , is expressed as:-

Equation 1: $p_{ij} = f(p_{1ij}, p_{2ij}, p_{3ij}, p_{4ij})$

where each p_{xij} is the probability of a virus particle i completing the step x of the four steps (Table 1) in human j . Calculating p_{ij} for every j in the host population and every i in the virus population would allow definition of the DRR for initial infection if the relative proportions of each genomic form in the virus and human populations were known. With NGS, it is conceivable that gene sequences for at least some j s and most i s could be available in the future, thus giving the relative proportions of genomic forms. The data required and feasibility of assigning a quantitative probability to p_{xij} for steps 1 to 4 in Table 1 are now discussed using NoV as the primary case study.

Probability of pathogen overcoming the immediate host defences to reach its cellular receptor

There are a number of immediate host defences in the mouth and gastrointestinal tract including the mucus barrier, microbial interference, decoy receptors and the innate immune system. The histoblood group antigens (HBGAs) are glycans to which certain viruses, including NoV and rotavirus, selectively bind. HBGAs are present on both decoy receptors in the saliva and on mucin, the main protein component of mucus (Boshuizen et al., 2005; Huang et al., 2009; McGuckin et al., 2011). Shanker et al. (2011) propose that local flexibility in the viral capsid protein could allow the NoV to disassociate from salivary HBGA due to the pH change during its passage through the gastrointestinal tract. However, the molecular mechanism is not understood, and the problem for estimation of p_1 is that the effect of pH on protein structure is going to be very difficult to predict from gene sequences. A further complication is that while mucin production is increased during rotavirus infection through a mechanism involving the innate immune system, other pathogens such as *Helicobacter pylori* suppress synthesis of gastric mucin (McGuckin et al., 2011). Thus information on p_1 from one pathogen cannot be applied to another. Mucin glycosylation patterns change with age of the host altering the antiviral capabilities (McGuckin et al., 2011; Boshuizen et al., 2005) but providing a molecular basis for understanding of variation in p_1 with age of the host.

The gastrointestinal tract microbiota prevent enteric disease by pathogens, including rotavirus, a trait referred to as colonisation resistance or microbial interference (Wardwell et al., 2011). However, the relationship between gastrointestinal viruses and commensal bacteria or how gut microbial populations affect infection by enteric viruses is not well understood (Wardwell et al., 2011). Interestingly, the microbiota modulate mucin production and glycosylation (McGuckin et al., 2011) and therefore indirectly affect p_1 . Sequencing the human microbiome will provide a “deep” genetic

perspective on individual bacteria (HMP, 2012) thus shedding light on the host-pathogen-microbiome interaction, although it will be difficult to translate this into an estimate of p_1 .

The innate immune system provides an immediate defense for the host against pathogens in a non-specific and generic manner. Central to the innate immune system is the recognition of specific, repeating structures (pathogen-associated molecular patterns or PAMPs) on the pathogen by the pathogen recognition receptors (PRRs) of the host (de Koning et al., 2012). A goal towards developing a mechanistic DRR would be to predict whether PAMPs on a given pathogen could be detected by the PRRs in a given host j . This requires an understanding of the molecular structure of the PRR in combination with its PAMP ligand. Crystal structures of PRR proteins complexed with pathogen PAMP ligands have been published (Takeuchi and Akira 2010) and in combination with genomics data could assist in predicting the strength of binding in different individuals. However, predicting effects of changes in gene sequence on binding may not be straightforward because although polymorphisms in host PRR genes attenuate recognition of pathogen PAMP ligands, the amino acids corresponding to these polymorphisms do not seem to directly contact the ligand molecules themselves in the case of *Salmonella* flagellin at least (Uenishi et al., 2012). Indeed, the differences may be subtle because it is not yet understood how the innate immune system distinguishes commensal bacteria from pathogenic bacteria, i.e. friend or foe (de Koning et al., 2012). Furthermore, there are numerous PRRs with each recognising different parts of the virus (Triantafilou et al., 2011) giving alternative pathways which will greatly complicate the estimation of the probability of a virus escaping the innate immune system prior to binding to its cellular receptor in step 2.

Probability of binding of pathogen to its cellular receptor

If the virus fails to bind to a receptor on a cell in the lumen of the gastrointestinal tract it will pass harmlessly through. The binding of virus to its receptor(s) on the host cell surface is a highly specific molecular interaction and controls host cell susceptibility and hence defines which cell types and tissues are infected (tissue tropism). The binding of NoV capsid protein to its HBGA receptor on the intestinal cells is well understood at the molecular level through X-ray crystallography studies (Choi et al., 2008; Bu et al., 2008). The ultimate goal would be to predict pathogen binding on the basis of nucleotide sequences of the viral capsid protein and the known (or ultimately predicted) structures of the HBGA glycan receptors. However, despite high conservation of the amino acid residues of the viral capsid protein that interact with HBGA, sequence similarity is not a predictor of the HBGA binding pattern (Bull and White 2011). Indeed closely related capsids can display distinct HBGA binding patterns, whereas genetically unrelated capsids from separate genogroups can display comparable HBGA binding patterns (Bull and White 2011). This will complicate the application of omics approaches to predicting pathogen specificity from viral protein sequences, and hence estimating p_2 . Further complications are that structures of NoV from genogroups GI and GII in complex with HBGAs reveal different modes of interaction (Hansman et al., 2011) and NoVs recognise HBGAs in a distinct strain-dependent manner (Shanker et al., 2011). The NoVs are constantly evolving with antigenic drift in the capsid protein giving different HBGA binding specificities (Shanker et al., 2011; Donaldson et al., 2008). Genomics data, together with crystal structure data, could be used to identify changes in surface moieties of viral capsids although interpretation of impact on function may be more difficult. An example of this has been observed by NGS of FMDV in laboratory-infected cattle (Wright et al.,

2011). The laboratory strain of virus had several positively charged amino acids in its coat protein selected for by negatively-charged heparin sulphate in cell culture. In the cow those positive charges reverted. While rotavirus targets the intestinal epithelium, it is not known exactly which cells in the intestine NoV binds to (Chan et al., 2011). To develop a mechanistic DRR, it is a requirement to know which cells are targeted initially by the virus as different cells express different surface proteins. It is concluded that even for the best understood systems, prediction of p_2 for individual virus i x human j interactions will be extremely difficult.

Using genomics data to assess the proportion of host population which is resistant to virus on the basis of receptor binding in step 2. The genetics of resistance to NoV infection provides an excellent example of how genetic markers could be used for calibrating a mechanistic DRR. Lindesmith et al. (2003) reported that 29% of humans studied in one population were homozygous recessive for the FUT2 gene and were unable to synthesise a functional enzyme to make the glycan part of the HBGA receptor to which the NoV binds. These populations do not express the receptor required for NoV binding and are thus highly protected against NoV infection, i.e. $p_2 \approx 0$. Indeed, none of those individuals developed an infection after **exposure**, however large the dose of virus ingested. About 20% of Europeans are in this highly protected category (Kindberg and Svensson, 2009). De Rougemont et al. (2011) could not accurately calculate the dissociation constant for NoV and its HBGA receptor, although the association is likely to be strong. Thus, in the case of NoV, it may be possible to simplify step 2 into assuming that $p_2 \rightarrow 1$ for those individuals who are genetically susceptible in having a functional FUT2 gene, while $p_2 = 0$ for those who are genetically resistant through having a non-functional FUT2 gene. This simplifies the calibration of p_2 in the DRR into determining the proportion of the population

which is resistant, an application to which genomics approaches are particularly well suited. However, the possibility that the pathogen could use alternative receptors would need to be eliminated, as in the case of the bacterium *H. pylori*, which is known to use other carbohydrate receptors in addition to HBGA (de Mattos et al., 2012) such that p_2 is the sum of probabilities of parallel routes each with a different p value.

The breakdown of volunteer dose response data of Ward et al. (1986) into a two component model has been proposed previously for rotavirus (Gale, 2005). However, unlike the case for NoV, there are limited molecular data on binding of rotavirus to its receptor or indeed on the genetic basis for variation between populations in susceptibility. Also, the binding process for rotavirus may be more complicated than for NoV with several cell-surface molecules implicated (Isa et al., 2008), and evidence that rotavirus surface protein may open up tight junctions between cells in the intestinal lumen allowing access for binding to the basal side of the cell (Nava et al., 2004; Boshuizen et al., 2005) in addition to the apical side of the cell used by NoV. Predicting such sequences of events would be difficult even if proteomics approaches identified the hubs of protein-protein interactions involved.

Probability of cell entry by the pathogen

The probability of host cell entry (p_3) by the virus (given it has bound to the cellular receptor in step 2) depends, for some viruses at least, on the efficiency of cleavage of a viral protein by a host cell proteolytic enzyme. This specific cleavage process typically occurs after the newly synthesised virions are released from the cells in the previous round of infection. The ability of the host cells' proteolytic enzymes to cleave the viral proteins may control entry to the cell in the next round of infection, as in the case of highly pathogenic avian influenza virus, for which the tissue tropism

and pathogenicity are determined by whether the host cells' proteolytic enzymes can cleave the viral haemagglutinin protein. This explains the ability of highly pathogenic H5 and H7 avian influenza strains to replicate in many organs outside the respiratory tract thus defining their highly pathogenic phenotype (Chaipan et al., 2009).

Predicting p_3 depends in part on being able to predict whether the viral protein is not only specifically cleaved (i.e. in the right place) but also efficiently cleaved. This is more of a traditional biochemistry/bioinformatics problem dependent on understanding the structure of the virus protein and the mechanism/specificity of the host proteolytic enzymes. Thus, for example, Misasi et al. (2012) found that specific differences in the requirements of filoviruses for a host cell proteolytic enzyme (called cathepsin B) are correlated with sequence polymorphisms at residues 47 and 584 in the viral glycoproteins. The use of cathepsin proteases as host factors for virus entry is a general property of members of the filovirus family. Those authors were able to predict that a newly identified ebolavirus is cathepsin B dependent. This approach may be applicable to assessing cell entry requirements for other viruses in the future and would contribute to our understanding of p_3 in different host cell types. In this respect the rate of cleavage in different tissues would affect the magnitude of p_3 . Cell entry by SARS coronavirus also requires cleavage of the virus spike protein by cellular cathepsin (Bosch et al., 2008). The mechanism of NoV entry is not clear although there is some evidence that a proteolytic process could be necessary for replication (Tan et al., 2006).

Probability of replication within the host cell

Isolation of viruses by cell culture relies on replication of the virus within the cells and using several types of cell culture provides a suitable environment for many viruses (Leland and Ginocchio 2007). Indeed, transgenic expression of the appropriate

receptor (step 2) may make a previously nonsusceptible cell susceptible (e.g. as for poliovirus, (PV)) (Leland and Ginocchio 2007). Thus the default position for step 4 would be to assume that $p_4 \rightarrow 1$, and that viral RNA replication, capsid synthesis and assembly can occur in the host cell given the virus nucleoprotein core has gained entry through steps 2 and 3. Supporting this, replication of the viral RNA is performed by RNA dependent RNA polymerase (RdRp) present in the virus core and synthesis of new virus capsid proteins takes place on the host cell ribosomes. Indeed, viral proteins seize control of cellular translation factors and host signalling pathways not only ensuring viral proteins are produced but also stifling the innate host defences that limit the capacity of infected cells to produce virus (Walsh and Mohr 2011).

Understanding the regulation of these pathways may enable refinement of the estimate of p_4 . Thus, viruses commonly use host cell survival mechanisms to their own advantage and in particular the PI3K/Akt pathway which suppresses apoptosis (programmed cell death) so giving time for completion of viral replication (Eden et al., 2011). Phosphorylation of NoV RdRp by the host Akt pathway decreases the polymerase activity (Eden et al., 2011), and may be an antiviral strategy which should be considered in estimating p_4 . Central to understanding virus replication in step 4 is the application of proteomics to determine protein-protein contacts in situ in the infected cell and thus how the individual virus proteins associate with host cell proteins and hence work together. It is envisaged that proteomic techniques could be used to identify hubs in virus-infected host cells in a similar way to those in the bacterium *Bacillus subtilis* (Marchadier et al., 2011). Of importance for assessing the infectivity of a virus in terms of the DRR is how each viral protein interacts with the host proteins in the different hubs, many of which correspond to regulatory pathways (Marchadier et al., 2011). A proteomic analysis (using mass spectrometry) of human

host proteins associated with the RdRp of H5N1 avian influenza virus identified 400 proteins which may have a role in induction of apoptosis and in innate antiviral signalling or are cellular RNA polymerase accessory factors i.e. involved in regulation of gene expression in the cell (Bradel-Tretheway et al., 2011). Such proteomics data by identifying host cell proteins bound to viral polymerases and so defining protein hubs may enhance our understanding of virus replication and assembly, and would enable qualitative refinement of the estimation of the p_4 .

USING DEEP SEQUENCING TO DIRECTLY DETERMINE THE PROBABILITY OF INFECTION FOR AN RNA VIRUS

NoV is an example of an RNA virus. Indeed many viruses which infect humans, including PV, hepatitis C virus (HCV) and human immunodeficiency virus (HIV), have RNA genomes and differ from DNA pathogens such as bacteria and protozoa in that the replication of the viral RNA during infection of a cell (step 4) is prone to error (Smidansky et al. 2008), so generating a cloud of mutants (i.e. the progeny virions differ in sequence, albeit by only one or two bases in 10,000). This is called a mutant spectrum and is central to quasispecies theory (Lauring and Andino 2010). The key point is that the individual viruses excreted by an infected person in faeces or in vomitus, and therefore present in an oral **exposure**, are slightly different in sequence, albeit related. NGS (or deep sequencing) provides sequences for all the viral variants in an **exposure** together with their relative frequencies (i.e. the mutant spectrum) and has been used to detect minority sequence variants for a number of human viruses (Wright et al., 2011). This is clearly an important first step for making predictions on the DRR. Indeed NGS of RNA viruses recovered from an infected person may give an indication of the number of virions that initiated the infection and has shown that in the case of HCV and HIV, infection arises from a single variant often representing

initial infection by one or a few viral particles (Wang et al. 2010; Bull et al. 2012). Moreover, the infection process may represent a bottleneck such that only one or two viral variants in the **exposure** dose successfully establish infection (Bull et al 2011). Thus in the case of NoV, only minor variants at frequencies as low as 0.01% were successfully transmitted to establish a new infection (Bull et al., 2012). This suggests the risk of infection from a NoV virion, on average, is ~0.0001 (assuming the **exposure** dose was high, which is likely through faeces or vomitus, such that the complete spectrum of variants was present). If every NoV had a high probability of infection, then Bull et al. (2012) would have seen a range of NoV sequences in the infected persons corresponding to most of the individual virions in the **exposure** dose. Indeed, the two major variants in the donor person accounted for 99% of the variants but were not present in either of the two recipient persons. Similarly, for HCV, only one or two viral variants successfully established infection (Wang et al., 2010; Bull et al., 2011). Wang et al. (2010) write that, “nothing in our data rules out the possibility that many additional particles were in the initial inoculum but did not replicate in the new host”. This raises the question of how many viruses were in the dose in the original exposure event but had very low, or even negligible, probabilities of infection due to the bottleneck. The number of foot-and-mouth disease virus (FMDV) RNA molecules/ml of serum measured in infected pigs by RT-PCR are three to four orders of magnitude higher than the numbers of plaque-forming units (pfu) measured by conventional methods (Rodriguez-Calvo et al., 2011). Thus, each pfu “dosage” in the human rotavirus volunteer study (Ward et al., 1986) may also comprise many individual virions, again consistent with the ID₅₀ being high in terms of virions. Thus, although Haas et al., (1993) estimated the risk of infection from one pfu of rotavirus to be 27%, that pfu may comprise thousands of virions. This is not inconsistent with

the results of Bull et al. (2012) for NoV which suggest the ID₅₀ is in the order of 7,000 virions since only 0.01% can get through the bottleneck and initiate infection. Thus, if the probability of infection by a virion is 0.0001, then according to the negative exponential dose response model about 7,000 virions would be needed for a 50% chance of infection.

NGS data on the viral mutant spectrum give complementary information on some of the steps of the mechanistic infection process in Table 1. Thus, Bull et al. (2012) speculate that only a few NoV variants can bind to HBGA in step 2 effectively filtering out the majority of NoV variants in **an exposure** (representing a bottleneck). Indeed, the minor variant transmitted differed from the dominant donor variant in an amino acid change directly adjacent to the primary HBGA binding site in the capsid. For HIV, the variants responsible for establishing infection in a new host have unique phenotypic properties that contribute to the establishment of infection with specific amino acids that increase viral entry efficiency (step 3) as well as unique glycosylation patterns that aid attachment to host receptors (step 2) (Bull et al., 2012). Thus a challenge for developing a mechanistic DRR for initial infection is determining which variants are important for a given step (Table 1).

APPLYING OMICS TO THE MODELLING OF DISEASE PROGRESSION ONCE INITIAL INFECTION HAS OCCURRED

Central to disease progression and viral pathogenesis by RNA viruses is the fidelity of replication by RdRp (Domingo et al., 2008; Smidansky et al., 2008) which controls the complexity of the mutant spectrum produced from step 4. The spreading of PV, for example, to the nervous system is secondary to the initial establishment of infection in the gastrointestinal tract. Thus PV titres peak in the small intestine and brain at day 1 and 5, respectively (Vignuzzi et al. 2006). Different individual progeny

viruses in the mutant spectrum have different cell and tissue tropisms reflecting their specific gene sequences compared to the original infecting virus. Specifically, a PV mutant with a high fidelity RdRp (i.e. produces a reduced range of variants) replicated in the small intestine, kidney and spleen but failed to reach the brain and spinal cord and to produce neuropathology in mice (Vignuzzi et al. 2006). However, restoration of the standard mutant spectrum complexity by subjecting the mutant PV to chemical mutagenesis led to a neuropathogenic mutant spectrum with the virus reaching the brain and spinal cord (Vignuzzi et al. 2006; Smidansky et al., 2008). Thus, enhanced opportunities for co-operation between more diverse clades of mutant spectra (arising from the high mutation rates) facilitate the virus's reaching specific target organs, thereby increasing viral loads and chances of transmission. However, too much genotypic variability may retard the progress of the virus infection by producing too many defective variants (Smidansky et al., 2008). A conclusion of Smidansky et al. (2008) is that RdRp fidelity is tuned by natural selection to achieve the optimal genotypic diversity in the mutant spectrum and, consequently, viral competitiveness in the dynamic and hostile environment of the host cell. Furthermore, the level of diversity and type of diversity needed may change in different tissues as different host challenges are encountered. It may also change during the infection cycle and may vary depending on the route of transmission (Wright et al., 2011; Bull et al., 2012). Bull et al., (2012) speculate that deep sequencing may yet identify differences between NoV transmitted through faeces compared to those in vomitus, for example. The huge complexity of the interactions together with the lack of understanding of how RdRp controls fidelity at the molecular level (Smidansky et al., 2008) severely limit prediction of disease progression. A further complication is that the host response to pathogen is also dynamic with the host HBGA specificity to NoV

changing over time (Lindesmith et al., 2008) together with changes in host mucin glycosylation during rotavirus infection such that the mucins become more potent in inhibiting the virus (Boshuizen et al., 2005).

The element of chance in disease progression

There is also evidence that disease progression could to a certain extent depend on chance (i.e. it is probabilistic). Thus deep-sequencing has shown that FMDV may take different evolutionary trajectories even within the same cow, albeit in different isolated compartments, namely the front left foot vs. the back right foot (Wright et al., 2011). This specifically confounds prediction for the individual host.

Viral markers of virulence

Although it is not possible at present to interpret pathogenic potential from viral gene sequences, there are examples of markers in the virus sequence known to be associated with disease-related traits. Thus, the amino acid threonine at residue position 33 in RdRp is associated with pandemic potential in NoV GII (Eden et al., 2011) and in the case of FMDV in cattle, an amino acid change in one of the four capsid proteins is associated with persistent infection (Horsington and Zhang 2007). Identification of such markers may enable more broad predictions to be made about disease potential in a population.

Host markers of resistance

Genetic markers on the host may give an indication of resistance to disease progression through being able to mount an effective immune response. For example, understanding polymorphisms in the PRR genes (Wang et al., 2011) and human leukocyte antigens (HLA) genes (Hoof et al., 2012) together with interferon coding regions (Thomas et al., 2009) may indicate the proportion of the population able to mount more effective responses to an RNA virus such as HCV. The HLA proteins

present viral peptides to the T cells in the acquired immune response and it is well documented that individual persons differ in their HLA proteins and that some are more protective than others against viruses including HIV (Wang et al., 2012), HCV, hepatitis B virus and herpes simplex type 1 virus. Hoof et al. (2012) found that protective HLA molecules show a preference to present peptide fragments from conserved HCV proteins (e.g. Core and NS5B), while non-protective HLA molecules preferentially target HCV proteins that are significantly less conserved (e.g. NS5A). Taken together, their analysis suggests that by targeting the most constrained, and thereby conserved, parts of the HCV genome, “protective” HLA molecules reduce the potential of HCV to escape the cytotoxic T cell response of the infected host. Thus it may be possible to predict viral clearance from the type of viral protein which is presented by the HLA. Carlsson et al. (2009) identified capsid residues in NoV that are evolutionarily conserved. X-ray crystallography structures are available for human HLA molecules bound to viral peptides including SARS coronavirus (Roder et al., 2008a) and other studies for swine leukocyte antigen (the porcine HLA equivalent) complexed with peptides from 2009 pandemic H1N1 influenza virus or Ebola virus have recently been published (Zhang et al., 2011). However, predicting whether conserved or less conserved peptide fragments are presented by the HLA and the effectiveness of viral peptide presentation for the purpose of estimating disease progression will be difficult. An additional problem is that glycans play a role in antigen presentation by the HLA, with glycosylation perhaps fine-tuning the binding properties (Ryan and Cobb 2012).

Viral countermeasures

Some viral proteins target the HLA proteins and interfere with antigen presentation (Roder et al., 2008b). Other viral countermeasures in disease progression include the

ability of the NoV capsid to evolve to evade the memory immune response and escape from herd immunity while retaining its ability to bind any of several HBGAs (Donaldson et al., 2008). Thus, many individuals are susceptible to reinfection by NoV (Lindesmith et al., 2010). NGS of viral sequences may allow identification of immune escape mutations which impair the ability of both T cell responses and neutralizing antibodies to maintain immune control as shown for HIV-1 (Henn et al., 2012). These rapid, low frequency viral variants are not detected by conventional sequencing approaches and demonstrate the application of genomics in understanding the dynamic interplay in immune control. Some viruses, e.g. porcine reproductive and respiratory syndrome virus (PRRSV), prevent the protective immune response by suppressing of certain host genes. As an example of the use of transcriptomics, Wysocki et al. (2012) compared differences in RNAs from lungs of high and low PRRSV burden pigs and found that in high burden pigs, expression of cellular genes associated with host protection was delayed.

DISCUSSION

This review summarises some of the potential applications of omics approaches to developing DRRs for infection through the oral route by RNA viruses, with a view to better defining the role of omics data in MRA. While some specific issues of the DRR can be addressed by omics data (e.g. estimating resistant/susceptible ratios in the host), there are clearly other areas where substantial progress in our understanding of the molecular interactions is required before omics data can be applied to MRA (if they can ever be applied).

Although viruses have far fewer genes than bacterial pathogens, the nature of the mutant spectrum in the case of RNA viruses greatly complicates the application of omics approaches to the development of mechanistic dose response models. This is

compounded by the fact that the frequencies of some important variants that initiate infection are as low as 0.01% as in the case of NoV and HCV (Bull et al., 2012; Wang et al., 2010), raising the question of which sequence to look at in terms of specific interactions with the host. Here, the initial infection process in the gastrointestinal tract has been broken down into four mechanistic steps (Table 1). The breadth of the fields considered in these steps demonstrates the need for a multidisciplinary approach including not only omics data but also conventional biochemistry and molecular biology inputs which provide detail on the molecular components of each step. Thus, currently protein structure determination by crystallography studies underpins our understanding of the specific molecular interactions (e.g. PRR with PAMP ligands in step 1 or NoV capsid binding to HBGA in step 2). The overall conclusion is that the molecular mechanisms of each step in Table 1 are not sufficiently well understood at present to enable quantitative estimation of probabilities on the basis of omics data. However, omics approaches will make a major contribution to specific areas of the initial infection process in particular in accommodating genetic variation both in the pathogen and the host population through extensive sequencing. The human microbiome should be included here too because the gastrointestinal microbiota play a major role in protection against infection in step 1, although the molecular mechanisms are not established. Genome sequencing for individual humans enables identification of genetic differences between susceptible and non-susceptible persons and it may be possible in future to tailor risk assessment to different host populations in terms of genetic susceptibility to initial infection. A promising application of such genetic markers is to define the proportion of the host population which is more susceptible to initial infection by a specific pathogen. Current examples include the FUT2 gene which affects NoV

binding in the gastrointestinal tract in step 2 and potentially polymorphisms in the host PRR genes in step 1 (Wang et al., 2011). Thus the great strength of genome sequencing is in giving information on the distribution and proportion of susceptible genotypes in the host population rather than in predicting specificities of interactions from the amino acid sequences concurrently obtained. Indeed, even for the best characterised step in the infection process, namely NoV capsid/HBGA binding for step 2, changes in amino acid sequence cannot be used to predict changes in binding specificity and affinity. Furthermore, not only are there many gaps in our understanding of how the few viral proteins actually function and interact with host cell proteins but also there are often multiple mechanisms and alternative molecular pathways used by the virus, e.g. inhibition of host gene expression and translation by SARS coronavirus (Lokugamage et al., 2012). The potential of proteomics approaches is in contributing to our understanding of how the few viral proteins interact with a multitude of cellular proteins within the host cell particularly in step 4. Host glycans play a central role in the pathogen infection process including binding of virus to specific receptors in steps 1 and 2 and also in the immune system. The example with HBGA glycans in NoV binding (step 2) is simple in the sense that a single mutation in a gene (the FUT2 gene coding for the enzyme that synthesizes HBGA) results in the absence of a functional receptor. However, not all relationships involving glycans and omics data are as simple as this and the link between a single mutation in a protein-coding gene and susceptibility to infection may be the exception rather than the rule. Glycans have complex structures which may be changed during infection as the host tries to evade the pathogen, and also during development of the host (McGuckin et al., 2011). Prediction of glycan structures from DNA sequences is not possible currently, and unlike gene expression, it is not clear how cells control

glycan structures. Thus our current understanding of glycoproteomics and glycan function is a major limitation in developing a mechanistic DRR for viruses.

The ultimate aim in the application of omics and molecular biology approaches would be in taking the DRR beyond the classic volunteer-derived model for infection in a broad group of (typically) healthy individuals into the realms of predicting disease progression and morbidity at the individual host level, based on the genetic information of the host and virus. However, although genomics approaches may enable consideration of the virus as a dynamic and evolving system, as opposed to a static single entity within the infected host, predicting the progression of an RNA virus in terms of its ability to spread to new tissues (e.g. nervous system) or to evade the host acquired immune system and so persist, resulting in chronic infection is fraught with difficulties. This is due to both the complexity of the pathways and the nature of how variants in the RNA virus mutant spectrum interact. Thus the variants collectively and cooperatively contribute to the characteristics of the viral population and can express different phenotypic traits such that our ability to predict the outcome of an infection is limited (Lauring and Andino 2010). However, interaction of mutant spectra could help elucidate in the future the mechanisms by which higher exposure doses increase the probability of illness, given infection has occurred. Thus for NoV, Teunis et al. (2008) reported that infected subjects had a dose-dependent probability of becoming ill, ranging from 0.1 (at a dose of 10^3 genomes) to 0.7 (at 10^8 genomes). Genomics data should increase our repertoire of genetic markers both in the host and virus giving better understanding of those factors affecting disease progression.

As an alternative to attempting to use omics data to develop a mechanistic DRR, NGS may provide empirical data for calibrating a DRR. Thus, the data obtained by NGS of virus in donor and recipient hosts (e.g. that obtained for NoV by Bull et al. 2012)

show great promise for directly estimating the probability of infection by a virion while avoiding the uncertainties associated with the role of viral proteins, glycan interactions and mutant spectra. NGS of virus in donor and recipient provides a unique insight into the infection process in terms of the frequency of the variant which initiated infection and demonstrated in the case of NoV that 99.99% of virions, including the major variants, do not initiate infection (Bull et al., 2012). This suggests most of the virions in **an exposure** have a very low probability of infection. NGS has shown that different virions in a **high dose exposure** have different sequences and some are better adapted at initiating infection than others (Bull et al., 2012).

In conclusion, it is anticipated that omics approaches can make a contribution to specific areas of the initial infection process. The emphasis will be on omics approaches, together with traditional biochemistry techniques, as a tool to understand more fully the mechanisms of each step in the initial infection process, although it seems certain that it will be an extremely difficult task to interpret this information quantitatively. Genomics data will continue to give information on the frequency of genes affecting susceptibility to initial infection within the host population. Lack of knowledge on glycans together with difficulties in predicting the functional effects of amino acid changes in key proteins pose challenges. It is concluded that predicting disease progression (given initial infection has occurred) for RNA viruses at the level of the individual host is not possible. This reflects the lack of understanding in variability of the virus, evolution of the virus during progression of the disease, co-operation between different clades of mutant spectra and the countermeasures by both the host and virus, together with chance factors. However, molecular markers in the host and virus may provide useful information on disease progression for risk assessment. The potential use of NGS to compare mutant spectra of viruses in the

exposure dose presented by the infected donor and in the infected recipient as a method to give direct estimates of probability of infection should be explored further.

ACKNOWLEDGEMENTS

We thank Prof. Trevor Drew of AHVLA for his helpful comments.

CONFLICT OF INTEREST

No conflict of interest declared.

REFERENCES

- Boshuizen, J.A., Reimerink, J.H.J., Korteland-van Male, A.M., van Ham, V.J.J., Bourma, J., Gerwig, G.J., Koopmans, M.P.G., Buller, H.A. et al. (2005) Homeostasis and function of goblet cells during rotavirus infection in mice. *Virology* **337**, 210-221.
- Bradel-Tretheway, B.G., Mattiaco, J.L., Krasnoselsky, A., Stevenson, C., Purdy, D., Dewhurst, S. and Katze, M.G. (2011) Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *Journal of Virology* **85**, 6569-8581.
- Bosch, B.J., Bartelink, W. and Rottier, P.J.M. (2008) Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class 1 fusion protein upstream of rather than adjacent to the fusion peptide. *Journal of Virology* **82**, 8887-8890.
- Bu, W., Mamedova, A., Tan, M., Xia, M., Jiang, X. and Hegde, R.S. (2008) Structural basis for the receptor binding specificity of Norwalk virus. *Journal of Virology* **82**, 5340-5347.
- Bull, R.A. and White, P.A. (2011) Mechanisms of GII.4 norovirus evolution. *Trends in Microbiology* **19**, 233-240.

Bull, R.A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S.T., Chopra, A., Cameron, B., Maher, L. et al. (2011) Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog* **7**, e1002243.

Bull, R.A., Eden, J.-S., Luciani, F., McElroy, K., Rawlinson, W.D. and White, P.A. (2012) Contribution of intra- and interhost dynamics to norovirus evolution. *Journal of Virology* **86**, 3219-3229.

CAC (1999) Codex Alimentarius Commission. Principles and guidelines for the conduct of a microbiological risk assessment. FAO, Rome. CAC/GL-30.

Carlsson, B., Lindberg, A.M., Rodriguez-Diaz, J., Hedlund, K.O., Persson, B. and Svensson, L. (2009) Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection. *Journal of General Virology* **90**, 432-441.

Cassin, M.H., Lammerding, A.M., Todd, E.C.D., Ross, W. and McColl, R.S. (1998) Quantitative risk assessment for *Escherichia coli* O157 : H7 in ground beef hamburgers. *International Journal of Food Microbiology* **41**, 21-44.

Chaipan, C., Kobasa, D., Bertram, S., Glowacka, I., Steffen, I., Tsegaye, T.S., Takeda, M., Bugge, T.H. et al. (2009) Proteolytic activation of the 1918 influenza virus hemagglutinin. *Journal of Virology* **83**, 3200-3211.

Chan, M.C.-W., Ho, W.-S. and Sung, J.J.-Y. (2011) In vitro whole-virus binding of a norovirus geogroup II genotype 4 strain to cells of the lamina propria and Brunner's glands in the human duodenum. *Journal of Virology* **85**, 8427-8430.

Choi, J.-M., Hutson, A.M., Estes, M. and Prasad, B.V.V. (2008) Atomic resolution structural characterisation of recognition of histo-blood group antigens by Norwalk virus. *Proceedings of the National Academy of Sciences USA* **105**, 9175-9180.

De Koning, H.D., Simon, A., Zeeuwen, P.L.J.M. and Schalkwijk, J. (2012) Pattern recognition receptors in infectious skin diseases. *Microbes and Infection* **14**, 881-893.

De Mattos, L.C. (2012) Histo-blood group carbohydrates and *Helicobacter pylori* infection. *Journal Bacteriology and Parasitology* **3**, 1000e104.

De Rougemont, A., Ruvoen-Clouet, N., Estienney, S.M., Elie-Caille C, Aho, S., Pothier, P., Le Pendu, J., Boireaum W. and Belliot, G. (2011) Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. *Journal of Virology* **85**, 4057-4070.

Doerr, A. (2012) Glycoproteomics. *Nature Methods* **9**, 36.

Domingo, E., Escarmis, C., Menendez-Arias, L., Perales, C., Herrera, M., Novella, I.S. and Holland, J.J. (2008) Viral quasispecies: Dynamics, interactions, and pathogenesis. In *Origin and Evolution of Viruses* ed. Domingo, E., Parrish, C.R. and Holland, J.J. pp 87-118. Amsterdam, Netherlands, Elsevier.

Donaldson, E.F., Lindesmith, L.C., Lobue, A.D. and Baric, R.S. (2008) Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. *Immunological Reviews* **225**, 190-211.

Eden, J.-S., Sharpe, L.J., White, P.A. and Brown, A.J. (2011) Norovirus RNA-dependent RNA polymerase is phosphorylated by an important survival kinase, Akt. *Journal of Virology* **85**, 10894-10898.

Gale, P. (2005) Land application of treated sewage sludge: quantifying pathogen risks from consumption of crops. *Journal of Applied Microbiology* **98**, 380-396.

Haas, C.N., Rose, J.B., Gerba, C. and Regli, S. (1993) Risk assessment of virus in drinking water. *Risk Analysis* **13**, 545-552.

Haas, C.N., Rose, J.B. and Gerba, C.P. (1999) Conducting the Dose-response assessment. In Quantitative microbial risk assessment ed. Haas, C.N., Rose, J.B. and Gerba, C.P. pp 260-319. John Wiley: New York.

Hansman, G.S., Biertumpfel, C., Georgiev, I., McLellan, J.S., Chen, L., Zhou, T., Katayama, K. and Kwong, P.D. (2011) Crystal structures of GII.10 and GII.12 norovirus protruding domains in complex with histo-blood group antigens reveal details for a potential site of vulnerability. *Journal of Virology* **85**, 6687-6701.

Henn, M.R., Boutwell, C.L., Charlebois, P., Lennon, N.J., Power, K.A., Macalalad, A.R., Berlin, A.M., Malboeuf, C.M., et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* **8**, e1002529.

Hoof, I., Rao, X.Y., van Baarle, D. and Kesmir, C. (2012) Protective HLA molecules determine infection outcome in hepatitis C virus infection by preferential preservation of peptides from conserved viral proteins. *Tissue antigens* **79**, 411-412.

Horsington, J. and Zhang, Z. (2007) Consistent change in the B-C loop of VP2 observed in foot-and-mouth disease virus from persistently infected cattle: implications for association with persistence. *Virus Research* **125**, 114-118.

Huang, P.W., Morrow, A.L. and Jiang, X. (2009) The carbohydrate moiety and high molecular weight carrier of histo-blood group antigens are both required for norovirus-receptor recognition. *Glycoconjugate Journal* **26**, 1085-1096.

HMP (2012) Human Microbiome Project – Science publications

<http://commonfund.nih.gov/hmp/publications.aspx>.

Isa, P., Gutierrez, M., Arias, C.F. and Lopez, S. (2008) Rotavirus cell entry. *Future Virology* **3**, 135-146.

- Kindberg, E. and Svensson, L. (2009) Genetic basis of host resistance to norovirus infection. *Future Virology* **4**, 369-382.
- Lauring, A.S. and Andino, R. (2010) Quasispecies theory and the behaviour of RNA viruses. *PLoS Pathogens* **6**, e1001005.
- Leland, D.S. and Ginocchio, C.C. (2007) Role of cell culture for virus detection in the age of technology. *Clinical Microbiology Reviews* **20**, 49-78.
- Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindbald, L., Stewart, P., LePendou, J. and Baric, R. (2003) Human susceptibility and resistance to Norwalk virus infection. *Nature Medicine* **9**, 548–553.
- Lindesmith, L.C., Donaldson, E.F., Lobue, A.D., Cannon, J.L., Zheng, D.P., Vinje, J. and Baric, R.S. (2008) Mechanisms of GII.4 norovirus persistence in human pathogens. *PLoS Med* **5**, e31.
- Lindesmith, L.C., Donaldson, E., Leon, J., Moe, C., Frelinger, J.A., Johnston, R.E., Weber, D.J. and Baric, R.S. (2010) Heterotrophic humoral and cellular immune responses following Norwalk virus infection. *Journal of Virology* **84**, 1800-1815.
- Lokugamage, K.G., Narayanan, K., Huang, C. and Makino, S. (2012) Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *Journal of Virology* **86**, 13598-13608.
- Marchadier, E., Carballido-Lopez, R., Brinster, S., Fabret, C., Mervelet, P., Bessieres, P., Noirot-Gros, M.-F., Fromion, V. and Noirot, P. (2011) An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach. *Proteomics* **11**, 2981-2991.
- McGuckin, M.A., Linden, S.K., Sutton, P. and Florin, T.H. (2011) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* **9**, 265-278.

Misasi, J., Chandaran, K., Yang, Y.-Y., Considine, B., Filone, C.M., Cote, M., Sullivan, N., Fabozzi, G., Hensley, L. and Cunningham, J. (2012) Filoviruses require endosomal cysteine proteases for entry but exhibit distinct protease preferences. *Journal of Virology* **86**, 3284-3292.

Nauta, M.J., Jacobs-Reitsma, W.F., Havelaar, A.H. (2007) A risk assessment model for *Campylobacter* in broiler meat. *Risk Analysis* **27**, 845–861.

Nava, P., Lopez, S., Arias, C.F. and Gonzalez-Mariscal, L. (2004) The rotavirus surface protein VP8 modulates the gate and fence function of tight junctions in epithelial cells. *Journal of Cell Science* **117**, 5509-5519.

Neidhardt, F.C. (2011) How microbial proteomics got started. *Proteomics* **11**, 2943-2946.

Pareek, C.S., Smoczynski, R. and Tretyn, A. (2011) Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**, 413-435.

Roder, G., Kristensen, O., Kastrup, J.S., Buus, S. and Gajhede, M. (2008a) Structure of a SARS coronavirus-derived peptide bound to the human major histocompatibility complex class 1 molecule HLA-B*1501. *Acta Crystallographica Section F – Structural biology and Crystallisation Communications* **64**, 459-462.

Roder, G., Geironsen, L., Bressendorff, I. and Paulsson, K. (2008b) Viral proteins interfering with antigen presentation target the major histocompatibility complex class 1 peptide-loading complex. *Journal of Virology* **82**, 8246-8252.

Rodriguez-Calvo, T., Diaz-San Segundo, F., Sanz-Ramos, M. and Sevilla, N. (2011) A replication analysis of foot-and-mouth disease virus in swine lymphoid tissue might indicate a putative carrier state in pigs. *Veterinary Research* **42**, 22.

Ryan, S.O. and Cobb, B.A. (2012) Host glycans and antigen presentation. *Microbes and Infection* **14**, 894-903.

Shanker, S., Choi, J.-M., Sankaran, B., Atmar, R.L., Estes, M.K. and Prasad, B.V.V.

(2011) Structural analysis of histo-blood group antigen binding specificity in a Norovirus GII.4 epidemic variant: Implications for epochal evolution. *Journal of Virology* **85**, 8635-8545.

Smidansky, E.D., Arnold, J.J. and Cameron, C.E. (2008) Nucleic acid polymerase fidelity and viral population fitness. In *Origin and Evolution of Viruses* ed. Domingo, E., Parrish, C.R. and Holland, J.J. pp 135-160. Amsterdam, Netherlands, Elsevier.

Takeuchi, O. and Akira, S. (2010) Pattern recognition receptors and inflammation. *Cell* **140**, 805-820.

Tan, M., Meller, J. and Jiang, X. (2006) C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein. *Journal of Virology* **80**, 7322-7331.

Teunis, P., Takumi, K. and Shinagawa K. (2004) Dose response for infection by *Escherichia coli* O157:H7 from outbreak data. *Risk Analysis* **24**, 401-407.

Teunis, P.F.M., Moe, C.L., Liu, P., Miller, S.E., Lindesmith, L., Baric, R.S., Le Pendu, J. and Calderon, R.L. (2008) Norwalk virus: how infectious is it? *Journal of Medical Virology* **80**, 1468-1476.

Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., O’Huigin, C., Kidd, J., Kidd, K., et al. (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798-801.

Triantafilou, K., Vakakis, E., Richer, E.A.J., Evans, G.L., Villiers, J.P. and Triantafilou, M. (2011) Human rhinovirus recognition in non-immune cells is mediated by Toll-like receptors and MDA-5, which trigger a synergetic pro-inflammatory immune response. *Virulence* **2**, 22-29.

Uenishi, H., Skinkai, H., Morozumi, T. and Muneta, Y. (2012) Genomic survey of polymorphisms in pattern recognition receptors and their possible relationship to infections in pigs. *Veterinary Immunology and Immunopathology* **148**, 69-73.

Vignuzzi, A., Stone, J.K., Arnold, J.J., Cameron, C.E. and Andino, R. (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344-348.

Walsh, D. and Mohr, I. (2011) Viral subversion of the host protein synthesis machinery. *Nature Reviews Microbiology* **9**, 860-975.

Wang, G.P., Sherrill-Mix, S.A., Chang, K.-M., Quince, C. and Bushman, F.D. (2010) Hepatitis C virus transmission bottlenecks analysed by deep sequencing. *Journal of Virology* **84**, 6218-6228.

Wang, C.H., Eng, H.L., Lin, K.H., Chang, C.H., Hsieh, C.A., Lin, Y.L. and Lin, T.M. (2011) TLR7 and TLR8 gene variations and susceptibility to hepatitis C virus infection. *PLOS ONE* **6**, e26235.

Wang, C., Krishnakumar, S., Wilhelmy, J., Babrzadeh, F., Stepanyan, L., Su, L.F., Levinson, D., Fernandez-Vina, M.A. et al. (2012) High-throughput, high fidelity HLA genotyping with deep sequencing. *Proceedings of the National Academy of Sciences USA* **109**, 8676-8681.

Ward, R.L., Bernstein, D.I., Young, E.C., Sherwood, J.R., Knowlton, D.R. and Schiff, G.M. (1986) Human rotavirus studies in volunteers: Determination of infectious dose and serological response to infection. *Journal of Infectious Diseases* **154**, 871-880.

Wardwell, L.H., Huttenhower, C. and Garrett, W.S. (2011) Current concepts of the intestinal microbiota and the pathogenesis of infection. *Current Infectious Disease Reports* **13**, 28-34.

Wright, C.F., Morelli, M.J., Thebaud, G., Knowles, N.J., Harzyk, P., Paton, D., Haydon, D.T. and King, D.P. (2011) Beyond the consensus: Dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *Journal of Virology* **85**, 2266-2275.

Wysocki, M., Chen, H., Steibel, J.P., Kuhar, D., Petry, D., Bates, J., Johnson, R., Ernst, C.W. and Lunney, J.K. (2012) Identifying putative candidate genes and pathways involved in immune responses to porcine reproductive and respiratory syndrome virus (PRRSV) infection. *Animal Genetics* **43**, 328-332.

Zhang, N., Qi, J., Feng, S., Gao, F., Liu, J., Pan, X., Chen, R., Li, Q., et al. (2011) Crystal structure of swine major histocompatibility complex class 1 SLA-1*0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. *Journal of Virology* **85**, 11709-11724.

Table 1: Breakdown of the initial infection process into four steps for building a mechanistic dose response relationship for RNA viruses through the oral route: Information needs.

Step	Description	Current information gaps and understanding requirements
1	Overcoming initial host defences including mucus barrier, the microbiome, decoy receptors and the innate immune system to reach the cellular receptor in the gastrointestinal tract*	Mechanism of disociation of virus from the decoy receptor and mucus, changes in glycosylation. Mechanisms of microbial interference.
2	Binding of virion to its cellular receptor in the gastrointestinal tract	Predicting specificity of glycan binding by viruses. Genetic markers of resistance and frequency of specific alleles in host population
3	Entry of virion to the host cell	Better understanding on proteolytic cleavage of viral proteins by host cell proteases
4	Replication of the virus in the infected cells (to give a mutant cloud) and capsid assembly	Proteomics may facilitate identification of host protein/virus protein interactions that control the regulatory pathways within the infected cell

*Microbiome includes the indigenous bacteria in the gastrointestinal tract and decoy receptors are host molecules which mop up viruses before they reach their target cells.