Deep sequencing of influenza A viruses tempered by antibodies against the M2 ectodomain

Silvie Van den Hoecke

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Science: Biochemistry and Biotechnology

Deep sequencing of influenza A viruses tempered by antibodies against the M2 ectodomain

Silvie Van den Hoecke

Molecular Virology Unit
VIB - Medical Biotechnology Center
Ghent University - Department of Biomedical Molecular Biology
Technologiepark 927
9052 Zwijnaarde (Ghent)
Belgium

Academic year: 2016-2017

Examination committee:

Chairman:	Prof. Dr. Johan Grooten[1]
Promotor:	Prof. Dr. Xavier Saelens[2,3]
Reading committee:	Prof. Dr. Paul Digard[4]
	Prof. Dr. Ir. Martin Guilliams[2,5]
	Prof. Dr. Philippe Lemey[6]
	Prof. Dr. Kristien Van Reeth[7]

[1] Laboratory Molecular Immunology, Department of Biomedical Molecular Biology, Ghent University, Belgium

[2] Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

[3] Medical Biotechnology Center, VIB, Ghent, Belgium

[4] Division of Infection and Immunity, The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, UK

[5] Laboratory of Immunoregulation and Mucosal Immunology, Inflammation Research Center, VIB, Ghent, Belgium

[6] Department of Microbiology and Immunology, Rega Institute, KU Leuven, University of Leuven, Leuven, Belgium

[7] Laboratory of Virology, Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, Belgium

Summary

Influenza is one of the most contagious respiratory diseases and is associated with a high annual disease burden. The disease is caused by influenza viruses and can be prevented through vaccination. However, the effectiveness of the licensed influenza vaccines is rather limited since they mainly elicit virus neutralizing antibodies, which are directed against the highly variable head domain of the hemagglutinin (HA), the major membrane protein of the virus. Effective protection is usually obtained when the virus strains present in the vaccine antigenically matches the circulating influenza strains. Since the replication of influenza viruses is error-prone, they carry a high genetic flexibility with continuous and gradual selection of mutations in the antigenic sites of HA. As a result, neutralizing antibodies, *e.g.* induced by natural infection or by vaccination, can no longer bind to antigenically drifted influenza viruses. This antigenic drift necessitates annual vaccination with an updated vaccine composition.

Consequently, there is an urgent need for an influenza vaccine that elicits broad-spectrum and long-lasting protection to further reduce the yearly economical and social disease burden associated with influenza virus infections. Several experimental approaches based on more conserved antigenic sites in influenza viruses are being developed in order to obtain a broadly protective influenza vaccine. One such approach was developed in our lab at Ghent University and VIB, and is based on the conserved ectodomain of the viral membrane protein 'matrix protein 2' (M2e) [1]. Preclinical data demonstrated the antigenic potential and heterosubtypic protective effect of M2e-based vaccines [2]. Some of these vaccines are being evaluated in clinical trials, and their safety and immunogenicity has been verified [3-6]. It is known that protection by M2e-based vaccines is mediated by non-neutralizing antibodies, which require Fcγ Receptor (FcγR) expressing immune cells to exert their protective effect [7, 8]. However, which antibody isotypes exactly cooperate with which FcγR *in vivo* is not entirely clear and the contribution of FcγRIV in protection by M2e-specific antibodies has not yet been studied. It is also important to investigate if, and how, influenza viruses will be able to escape from M2e-based immune pressure.

The genetic flexibility of influenza viruses which enables them to escape from natural immune selection pressure, is the result of their error-prone replication, together with their fast replication cycle, large population size and the high inherent tolerance of HA for mutations in its antigenic sites [9]. To study this intrinsic genetic diversity, a sensitive sequencing technique is required with a high sequencing throughput. The progress in the field of next-generation sequencing (NGS), where millions of DNA fragments are sequenced in parallel, makes it possible to study the genetic diversity in a viral population with great sensitivity.

At the beginning of this PhD project is was unclear which sequencing platform was the most accurate to detect genetic variation in rapidly evolving RNA viruses. Therefore, we first compared the suitability of two benchtop next-generation sequencers to determine the whole-genome genetic

diversity in an influenza A virus population: the Illumina MiSeq sequencing-by-synthesis and the Ion Torrent PGM semiconductor sequencing technique [10]. Since it is often unclear how NGS data is processed or analyzed, we also designed an NGS data analysis pipeline to determine the variants present in an RNA virus population, although this workflow is also more broadly applicable. The sensitivity and accuracy of both sequencers were first determined using plasmids carrying a wild type or mutant influenza M segment, derived from the plasmid-based reverse genetics system for PR8 virus, a well characterized influenza A virus lab strain. This comparison revealed that the read length after processing of the sequence reads was comparable for both sequencers. However, the sequence reads obtained on the Illumina MiSeq were one and a half times more accurate than the ones sequenced on the Ion Torrent PGM. Next, recombinant wild type and mutant PR8 viruses were created and an influenza-specific RT-PCR was designed, based on the conserved sequence at the influenza genome segment ends, which resulted in efficient amplification of all eight genomic segments. The viral sequencing reads obtained with both sequencers could successfully be assembled *de novo* into the segmented influenza virus genome. After mapping of the reads to the reference genome, we found that the detection limit for reliable recognition of variants in the viral genome required a frequency of 0.5% or higher. Most of the variants in the PR8 virus genome were present in HA, and these mutations were detected by both sequencers. Based on its lower total error rate and higher sequencing output, we concluded that the Illumina MiSeq platform is better suited to detect variant sequences, whereas the Ion Torrent PGM platform has a shorter turnaround time, which can be important in viral diagnosis. The designed influenza-specific RT-PCR protocol and NGS data analysis pipeline can be implemented in several applications *e.g.* in viral surveillance, influenza resistance testing and vaccine control.

An additional research application for which the NGS approach can be instrumental, is assessment of the genetic stability of recombinant viruses. In our lab, we designed a GFP-expressing influenza reporter virus, based on the PR8 backbone, and used the designed RT-PCR protocol and NGS data analysis pipeline to evaluate the genetic stability of this virus [11]. The introduced reporter gene has no selective advantage for the virus. Therefore, it is important to study its retention within the virus genome. Although GFP remained stably expressed after passaging the virus in permissive cells for several rounds of replication, a two-fold reduction in sequencing coverage was observed in the GFP sequence of the PR8-GFP viral stock. A sequence bias introduced during preparation of the sequencing library could be excluded, since the bias was present when either Nextera XT transposase or mechanical Covaris shearing was used for DNA fragmentation [12]. Likewise, deletion of this foreign sequence by the virus itself was unlikely, since the sequencing drop was also present when the parental plasmid DNA, used to create the GFP-expressing virus through reverse genetics, was sequenced. Based on these results, and in agreement with the findings of Ekblom *et al.*, we surmised that the 'CCCGCC' sequence motif in the GFP coding sequence is disfavoured by Illumina sequencing [13]. This assumption was verified by mutating this motif, which indeed resolved the reduction in sequence coverage [12]. This study shows that potential sequence biases should be taken into

account before making a conclusion based on the number of sequence reads of a particular target sequence when performing NGS coverage analysis.

The third objective of this PhD project was to investigate the protective mechanism of M2e-based vaccines. We compared the functional engagement of FcγR family members by two mouse monoclonal antibodies (mAbs) that bind to M2e with similar affinity, but are either of the IgG1 or IgG2a antibody isotype. We first compared the potency of this antibody pair to activate individual FcγRs in the context of a viral infection using an *in vitro* FcγR activation assay. Subsequently, the importance of the FcγRs for protection *in vivo* was investigated by passive transfer of the M2e-specific monoclonal antibodies to wild type mice or mice with different deficiencies in their FcɣR compartment, followed by a lethal viral challenge. From these experiments, we could conclude that M2e-specific IgG1 requires FcγRIII to accomplish protection upon infection, whereas M2e-specific IgG2a isotype antibodies can protect against influenza A virus challenge via any of the three activating FcγRs. These results demonstrate the higher protective potential of M2e-specific antibodies from the IgG2a isotype compared to IgG1. The protective potential of M2e-based vaccines can thus be increased by designing them to elicit a robust Th1-biased immune response.

In a last part of this PhD thesis we investigated the potential evasion strategies of influenza A viruses under M2e-based humoral immune pressure. The M2e sequence of influenza A viruses is highly conserved. It is thus important to examine if and how influenza viruses will escape once the M2e-based vaccines will be used to vaccinate the human population. To examine this, we infected SCID mice with influenza A virus in the presence of anti-M2e immune pressure, in the form of passively transferred M2e-specific mAbs. These mAbs recognize either an internal M2e epitope or the highly conserved first eight amino acids at the N-terminus, which are encoded in-frame with the M1 protein. We found that all M2e-specific antibodies significantly prolonged survival of challenged SCID mice compared to isotype antibody control treatment. In addition, and in agreement with the aforementioned study, M2e-specific IgG2a protected significantly better than IgG1 and even resulted in complete virus-clearance in some of the treated SCID mice. Subsequently, the viral diversity of virus released in the mouse lung was determined using the optimized influenza-specific RT-PCR and NGS data analysis pipeline. Escape in M2e was only detected when mice were treated with antibodies which recognize an internal M2e epitope. However, the genetic diversity was limited to a proline to histidine or leucine mutation at position 10 or an isoleucine to threonine mutation at position 11 in M2e. These changes in M2e abolish recognition by the antibodies, explaining how viruses with these mutations emerge. Surprisingly, in half of the samples from mice that had been treated with the anti-M2e mAbs that recognize an internal epitope, and in all samples from mice that were treated with an antibody that recognizes the N-terminal epitope, no mutation in the M2e sequence was detected. Instead, non-synonymous mutations were detected in other viral proteins, mainly the polymerases and/or HA. Some of these mutations, when combined, were associated with a delayed M2 expression compared to the other structural proteins in infected cells. This delayed expression of M2e may represent an alternative escape route of the virus to circumvent M2e-based

immune pressure in an immune compromised host. From this study we can conclude that only limited variation in M2e is tolerated. Moreover, these M2e variants can still be recognized by polyclonal anti-M2e immune serum. However, further research is required to investigate if influenza viruses with alternative escape routes will also emerge in immunocompetent mice.

Taken together, the work presented in this thesis shows that the genetic diversity in an influenza virus population can be determined by deep sequencing using an influenza-specific RT-PCR and NGS analysis pipeline [10]. Validating this workflow on a GFP-expressing reporter virus demonstrated that sequence bias should be taken into account when analysing sequence coverage [12]. In addition, the intrinsic genetic flexibility of influenza viruses enables escape to M2e-based immune selection pressure. However, the tolerated genetic diversity in M2e is limited and it will probably be harder for the virus to escape from a polyclonal anti-M2e immune response. Further research is required to investigate if influenza viruses that follow alternative escape routes can also emerge in immunocompetent individuals. Moreover, we showed that M2e-specific IgG2a antibodies have a higher protective potential than their IgG1 counterpart, which can be attributed to their ability to trigger all three activating Fcγ Receptors. These results suggest that M2e vaccine formulations should be used that induce high levels of M2e-specific IgG2a antibodies.

Samenvatting

De griep is één van de meest besmettelijke luchtweginfecties en is geassocieerd met een hoge jaarlijkse ziektelast. De ziekte wordt veroorzaakt door infectie met het influenzavirus en kan voorkomen worden door vaccinatie. De doeltreffendheid van de huidige griepvaccins is echter beperkt aangezien deze voornamelijk antilichamen opwekken tegen het zeer variabele hoofddomein van hemagglutinine (HA), het voornaamste membraaneiwit van het virus. Effectieve bescherming wordt gewoonlijk bekomen wanneer de antigene eigenschappen van de virusstammen aanwezig in het vaccin overeenkomen met de circulerende influenzavirusstammen. Aangezien de replicatie van influenzavirussen mutatiegevoelig is, kennen deze een grote genetische flexibiliteit met continue en graduele selectie van mutaties in de antigene posities van HA. Neutraliserende antilichamen, bijvoorbeeld deze opgewekt tijdens een natuurlijke infectie of na vaccinatie, kunnen bijgevolg niet langer binden op virussen met gewijzigde antigene eigenschappen. Deze antigene drift maakt jaarlijkse vaccinatie met een aangepaste vaccinsamenstelling noodzakelijk.

Er is dus dringend nood aan een griepvaccin dat langdurige bescherming biedt aan een breed spectrum van influenzavirussen om zo de jaarlijkse economische en sociale ziektelast, die gepaard gaat met influenzavirusinfecties, verder te reduceren. Momenteel worden er verschillende experimentele strategieën ontwikkeld die gericht zijn tegen de meer geconserveerde antigene posities van influenzavirussen, om zo een griepvaccin te bekomen dat een brede bescherming biedt. In ons laboratorium aan de Universiteit Gent en het VIB werd een dergelijk vaccin ontwikkeld dat gebaseerd is op het geconserveerde ectodomein van het virale membraaneiwit 'matrix eiwit 2' (M2e) [1]. Preklinische resultaten hebben het antigene potentieel en het beschermend effect van deze op M2e gebaseerde vaccins tegen verschillende subtypes van influenzavirussen aangetoond [2]. Sommige van deze vaccins worden momenteel geëvalueerd in klinische testen, waarbij hun veiligheid en immunogeniciteit reeds aangetoond is [3-6]. De bescherming door deze op M2e gebaseerde vaccins is te wijten aan niet-neutraliserende antilichamen die hiervoor afhankelijk zijn van immuuncellen die Fcγ Receptoren (FcγRs) op hun celoppervlak tot expressie brengen [7, 8]. Het is echter niet helemaal duidelijk welke antilichaamisotypes welke FcγRs vereisen om hun *in vivo* beschermend effect uit te kunnen oefenen. Daarnaast is de rol van FcγRIV in bescherming door M2e-specifieke antilichamen nog niet onderzocht. Het is ook belangrijk om te onderzoeken of, en hoe, influenzavirussen zullen ontsnappen aan deze op M2e-gebaseerde immuundruk.

De genetische flexibiliteit van influenzavirussen zorgt ervoor dat ze kunnen ontsnappen aan de natuurlijke immuunselectiedruk en is het resultaat van hun mutatiegevoelige replicatie, in combinatie met hun snelle replicatiecyclus, hun omvangrijke populatiegrootte en de inherente tolerantie van HA voor mutaties in zijn antigene posities [9]. Om deze intrinsieke genetische diversiteit te bestuderen is een gevoelige sequeneringstechniek vereist met een grote sequentiecapaciteit. De vooruitgang in het veld van 'next-generation' DNA sequentiebepalingstechnieken (NGS), waar verschillende miljoenen DNA fragmenten in parallel

gesequeneerd worden, maakt het mogelijk om de genetische diversiteit in een viruspopulatie met een grote gevoeligheid te bepalen.

Aan het begin van dit doctoraatsproject was het onduidelijk welk sequeneringsplatform het meest accuraat was om de genetische variatie in snel evoluerende RNA-virussen te bepalen. Daarom hebben we eerst de geschiktheid van twee verschillende NGS platformen vergeleken om de genetische diversiteit over de volledige genoomsequentie in een influenza A viruspopulatie te bepalen: de Illumina MiSeq sequenering-door-synthese en de Ion Torrent PGM halfgeleider-sequeneringstechniek [10]. Aangezien het vaak onduidelijk is hoe NGS data verwerkt of geanalyseerd wordt, hebben we ook een werkschema voor NGS data-analyse opgesteld om de variatie in een RNA-viruspopulatie te bepalen. Dit werkschema kan echter ook voor andere toepassingen gebruikt worden. De gevoeligheid en accuraatheid van beide sequeneringstechnieken werd eerst bepaald op basis van plasmiden die een wild type of mutant influenza M segment bevatten, afkomstig van het 'reverse genetics' plasmidesysteem van het PR8 virus, wat een goed gekarakteriseerde influenza A labostam is. Deze vergelijking toonde aan dat de bekomen sequentielengte na het verwerken van de sequenties vergelijkbaar was voor beide sequeneringstechnieken. De sequenties bekomen op de Illumina MiSeq waren echter anderhalve keer meer accuraat dan deze gesequeneerd op de Ion Torrent PGM. Daarna werden wild type en mutant PR8 virus recombinant aangemaakt en werd een influenza-specifieke RT-PCR ontwikkeld op basis van de geconserveerde sequenties aanwezig aan de uiteindes van de influenzagenoomsegmenten. Dit resulteerde in efficiënte amplificatie van elk van de acht genoomsegmenten. De virussequenties bekomen op de beide sequencers konden succesvol *de novo* geassembleerd worden tot het gesegmenteerde influenzavirusgenoom. Na alignneren van de sequenties aan het referentiegenoom, hebben we bepaald dat een detectielimiet van 0.5% gehanteerd moet worden om betrouwbare detectie van varianten in een virusstaal toe te laten. De meeste varianten in het PR8 virus werden gedetecteerd in HA en werden door beide sequeneringstechnieken gedetecteerd. Op basis van zijn lagere totale foutenfrequentie en hogere sequeneringscapaciteit, kunnen we besluiten dat het Illumina MiSeq platform meer geschikt is om mutaties op te sporen in een viruspopulatie, terwijl het Ion Torrent PGM platform een kortere doorlooptijd kent, wat belangrijk kan zijn in virusdiagnose. Het ontwikkelde influenza-specifieke RT-PCR protocol en het werkschema voor NGS data-analyse kunnen voor verschillende toepassingen gebruikt worden, bijvoorbeeld in virus surveillance, in het testen voor influenzaresistentie en vaccincontrole.

Een bijkomende onderzoekstoepassing waar de NGS benadering een grote rol kan spelen is in het bepalen van de genetische stabiliteit van recombinant aangemaakte virussen. In ons labo hebben we een rapporteervirus ontwikkeld dat GFP tot expressie brengt en waarbij we gebruik gemaakt hebben van het PR8 virusgenoom [11]. Vervolgens hebben we het ontwikkelde influenza-specifiek RT-PCR protocol en het werkschema voor NGS data-analyse gebruikt om de genetische stabiliteit van dit virus te bepalen [11]. Aangezien het geïntroduceerde rapporteergen geen selectief voordeel biedt voor het virus, is het daarom belangrijk om na te gaan of het gen in het virusgenoom behouden blijft

na virusreplicatie. Fenotypische analyse leerde ons dat het GFP eiwit stabiel tot expressie gebracht werd na verschillende rondes van replicatie in permissieve cellen. Daarentegen werd er echter een tweevoudige reductie van de sequentiedekking geobserveerd in de GFP sequentie na sequeneren van de PR8-GFP virusstock. Deze sequentiebias was aanwezig wanneer zowel de Nextera XT transposase methode of de mechanische Covaris methode gebruikt werd voor DNA fragmentatie, waardoor uitgesloten kon worden dat deze sequentiebias geïntroduceerd was tijdens het bereiden van de sequentiebibliotheek [12]. Bovendien was het onwaarschijnlijk dat het virus deze sequentie van vreemde oorsprong zelf verwijderd had, aangezien deze reductie in sequentiedekking ook geobserveerd werd indien het parentale plasmide DNA, dat gebruikt was om het virus dat GFP tot expressie brengt te produceren via 'reverse genetics', gesequeneerd werd. Op basis van deze resultaten, en in overeenstemming met de bevindingen van Ekblom *et al.*, vermoedden we dat het 'CCCGCC' sequentiemotief in de GFP coderende sequentie benadeeld werd tijdens het Illumina sequeneren [13]. Deze veronderstelling werd bevestigd door dit motief te muteren, aangezien dit de afname in sequentiedekking tenietdeed [12]. Deze studie toont aan dat rekening gehouden moet worden met een potentiële sequentiebias vooraleer een besluit genomen wordt op basis van het aantal sequenties dat aligneert aan een bepaalde doelsequentie bij NGS dekkingsanalyse.

In een derde luik van dit doctoraatsproject hebben we het beschermingsmechanisme van de op M2e-gebasseerde vaccins onderzocht. Hiervoor hebben we de functionele betrokkenheid van de verschillende FcγRs bepaald voor twee monoklonale antilichamen van muis origine (mAbs) die M2e binden met een gelijkaardige affiniteit, maar die ofwel van het IgG1 of IgG2a antilichaam isotype zijn. We hebben eerst het potentieel van beide antilichamen onderzocht om de individuele FcγRs te activeren in de context van een virale infectie door gebruik te maken van een *in vitro* FcγR activatietest. Vervolgens werd de *in vivo* rol van de FcγRs onderzocht door via passieve transfer M2e-specifieke mAbs toe te dienen aan wild type muizen of aan muizen met verschillende tekortkomingen in hun FcγR compartiment, gevolgd door infectie met een letale virusdosis. Op basis van deze experimenten konden we besluiten dat de M2e-specifieke antilichamen van het IgG1 isotype FcγRIII vereisen om tijdens een infectie bescherming te kunnen bieden, terwijl de M2e-specifieke antilichamen van het IgG2a isotype bescherming kunnen bieden tegen een influenza A virus infectie via alle drie de activerende FcγRs. Deze resultaten tonen dus aan dat M2e-specifieke antilichamen van het IgG2a isotype meer bescherming kunnen bieden tijdens een infectie. Het beschermend effect van op M2e-gebaseerde vaccins kan dus verhoogd worden door hen zodanig te ontwikkelen dat ze een robuuste Th1 immuunrespons opwekken.

In een laatste luik van deze doctoraatsthesis hebben we de mogelijke strategieën onderzocht die influenza A virussen kunnen volgen om aan de op M2e-gebaseerde humorale immuundruk te ontsnappen. De sequentie van M2e in influenza A virussen is zeer geconserveerd. Het is daarom belangrijk om na te gaan of, en op welke manier, influenzavirussen kunnen ontsnappen eens de bevolking met deze op M2e-gebaseerde vaccins gevaccineerd zal worden. Om dit te onderzoeken hebben we immunodeficiënte (SCID) muizen geïnfecteerd met influenza A virussen in de

aanwezigheid van anti-M2e immuundruk, onder de vorm van M2e-specifieke mAbs toegediend via passieve transfer. Deze mAbs herkennen ofwel een intern M2e epitoop of de zeer geconserveerde acht aminozuren aan de N-terminus, welke gecodeerd worden in hetzelfde leesraam als het M1 eiwit. Alle M2e-specifieke antilichamen resulteerden in een significante verlenging van de overleving van de geïnfecteerde SCID muizen, wanneer vergeleken werd met muizen die de isotype controle behandeling toegediend kregen. Daarnaast, en overeenkomstig met de bovengenoemde studie, resulteerde de behandeling met M2e-specifieke IgG2a antilichamen in een significant betere bescherming dan die van het IgG1 isotype, met zelfs complete virusklaring in sommige van de behandelde SCID muizen tot gevolg. Vervolgens werd de diversiteit van het in de muislong vrijgestelde virus bepaald door middel van de geoptimaliseerde influenza-specifieke RT-PCR en het werkschema voor NGS data-analyse. Mutaties in M2e werden enkel gedetecteerd indien de muizen behandeld werden met antilichamen die een intern M2e epitoop herkennen. De genetische diversiteit was echter wel beperkt tot een proline naar histidine of leucine mutatie op positie 10 of een isoleucine naar threonine mutatie op positie 11 in M2e. Deze mutaties verhinderen de binding van de antilichamen aan M2e, wat verklaart waarom virussen met deze mutaties geselecteerd worden. Het is opmerkelijk dat in de helft van de muizen die behandeld werden met anti-M2e mAbs die een intern epitoop herkennen en in alle stalen van muizen die behandeld werden met mAbs die een N-terminaal epitoop binden, geen mutaties in de M2e sequentie gedetecteerd werden. Daarentegen werden er wel verschillende niet-synonieme mutaties in andere viruseiwitten gedetecteerd, voornamelijk in de polymerasen en/of HA. Een combinatie van enkele van deze mutaties resulteert in vertraagde M2 expressie ten opzichte van de overige structurele eiwitten in de geïnfecteerde cellen. Dit kan een alternatieve manier zijn om de op M2e-gebasseerde immuundruk in een immunodeficiënte gastheer te omzeilen. Op basis van deze resultaten kunnen we besluiten dat enkel beperkte variatie in M2e toegelaten wordt. Bovendien hebben we aangetoond dat deze M2e varianten nog steeds herkend worden door een polyklonaal anti-M2e immuunserum. Verder onderzoek is echter vereist om na te gaan of influenzavirussen die aan de op M2e-gebaseerde immuundruk ontsnappen via een alternatieve manier ook zullen doorbreken in immunocompetente muizen.

Samengevat toont deze doctoraatsthesis aan dat de genetische diversiteit in een influenza viruspopulatie bepaald kan worden door middel van NGS indien gebruik gemaakt wordt van de influenza specifieke RT-PCR en een werkschema voor NGS data-analyse [10]. Valideren van dit werkschema op een rapporteervirus dat GFP tot expressie brengt toonde aan dat rekening gehouden moet worden met een eventuele sequentiebias indien de sequentiedekking bestudeerd wordt [12]. Daarnaast hebben we aangetoond dat de intrinsieke genetische flexibiliteit van influenzavirussen ervoor zorgt dat deze virussen kunnen ontsnappen aan de op M2e-gebasseerde immuunselectiedruk. De toegestane genetische diversiteit in M2e is echter beperkt waardoor het waarschijnlijk moeilijk zal zijn voor het virus om te ontsnappen aan een polyklonale anti-M2e immuunrespons. Verder onderzoek is echter vereist om te bepalen of de influenzavirussen die

alternatieve ontsnappingsroutes volgen ook geselecteerd zullen worden in immunocompetente gastheren. Daarnaast hebben we ook aangetoond dat M2e-specifieke antilichamen van het IgG2a isotype een betere bescherming bieden dan hun IgG1 equivalent, wat een gevolg is van hun mogelijkheid om alle drie de activerende Fcγ receptoren te stimuleren. Deze resultaten suggereren dat er voor de op M2e-gebaseerde vaccins een vaccinformulatie gekozen moet worden die hoge niveaus van M2e-specifieke IgG2a antilichamen zal opwekken.

References

1. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.
2. Deng L, Cho KJ, Fiers W, Saelens X: M2e-Based Universal Influenza A Vaccines. *Vaccines* 2015, 3(1):105-136.
3. Safety Study of Recombinant M2e Influenza-A Vaccine in Healthy Adults (FLU-A) [https://clinicaltrials.gov/ct2/show/results/NCT00819013?sect=Xa0156#outcome4]
4. Turley CB, Rupp RE, Johnson C, Taylor DN, Wolfson J, Tussey L, Kavita U, Stanberry L, Shaw A: Safety and immunogenicity of a recombinant M2e-flagellin influenza vaccine (STF2.4xM2e) in healthy adults. *Vaccine* 2011, 29(32):5145-5152.
5. Talbot HK, Rock MT, Johnson C, Tussey L, Kavita U, Shanker A, Shaw AR, Taylor DN: Immunopotentiation of trivalent influenza vaccine when given with VAX102, a recombinant influenza M2e vaccine fused to the TLR5 ligand flagellin. *PloS one* 2010, 5(12):e14442.
6. Dynavax Reports New Phase 1a and Phase 1b Data for Universal Flu Vaccine Candidate [http://investors.dynavax.com/releasedetail.cfm?ReleaseID=551606]
7. El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.
8. Lee YN, Lee YT, Kim MC, Hwang HS, Lee JS, Kim KH, Kang SM: Fc receptor is not required for inducing antibodies but plays a critical role in conferring protection after influenza M2 vaccination. *Immunology* 2014, 143(2):300-309.
9. Thyagarajan B, Bloom JD: The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 2014, 3.
10. Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X: Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC genomics* 2015, 16:79.
11. De Baets S, Verhelst J, Van den Hoecke S, Smet A, Schotsaert M, Job ER, Roose K, Schepens B, Fiers W, Saelens X: A GFP expressing influenza A virus to report in vivo tropism and protection by a matrix protein 2 ectodomain-specific monoclonal antibody. *PloS one* 2015, 10(3):e0121491.
12. Van den Hoecke S, Verhelst J, Saelens X: Illumina MiSeq sequencing disfavours a sequence motif in the GFP reporter gene. *Scientific reports* 2016, 6:26314.
13. Ekblom R, Smeds L, Ellegren H: Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC genomics* 2014, 15:467.

x

Table of contents

Abbreviations

A	aa	Amino acid
	Ab	Antibody
	ADCC	Antibody-dependent cellular cytotoxicity
	ADCP	Antibody-dependent cellular phagocytosis
	APC	Antigen presenting cell
	APS	Ammonium Persulfate
B	BAL	Bronchoalveolar lavage
	bp	Base pair
	BSA	Bovine serum albumine
C	CCD	Charge coupled device
	CDC	Complement-dependent cytotoxicity
	cDNA	Copy DNA
	CDR	Complementarity-determining regions
	cRNA	Copy RNA
D	ddNTP	Dideoxynucleotidetriphosphate
	DIP	Defective interfering particle
	DMEM	Dulbecco's Modified Eagle medium
	DNA	Deoxyribonucleic acid
	dNTP	Deoxynucleotidetriphosphate
	Dpi	Days post infection
	dsRNA	Double-stranded RNA
E	ELISA	Enzyme linked immunosorbent assay
	emPCR	Emulsion PCR
F	FCS	Fetal calf serum
	FcγR	Fc receptors for IgG
G	GFP	Green fluorescent protein
H	HA	Hemagglutinin
	HBc	Hepatitis B virus core
	HCMV	Human cytomegalovirus
	HEK	Human embryonic kidney

	HPAIV	Highly pathogenic avian influenza virus
	HRP	Horseradish peroxidase
I	IAV	Influenza A virus
	Indel	Insertion or deletion
	i.n.	Intranasal
	i.p.	Intraperitoneal
K	Kd	Dissociation constant
	kDa	Kilo Dalton
L	LD_{50}	Lethal dose for 50% of subjects
	LPAIV	Low pathogenic avian influenza virus
M	M1	Matrix protein 1
	M2	Matrix protein 2
	M2e	Ectodomain of matrix protein 2
	ma	Mouse adapted
	mAb	Monoclonal antibody
	MAP	Multiple antigenic peptide
	MDCK	Madin-Darby Canine Kidney
	MFI	Median fluorescence intensity
	Moi	Multiplicity of infection
	mRNA	Messenger RNA
	MW	Molecular weight
N	NA	Neuraminidase
	NCBI	National Center for Biotechnology Information
	NEP	Nuclear export protein
	NES	Nuclear export signal
	NCR	Non-coding region
	NGS	Next-generation sequencing
	NK	Natural killer
	NLS	Nuclear localization sequence
	NP	Nucleoprotein
	NS	Non-structural
O	ORF	Open reading frame

P	PA	Polymerase acidic
	PAMP	Pathogen-associated molecular pattern
	PB	Polymerase basic
	PBS	Phosphate buffered saline
	PFA	Paraformaldehyde
	PFU	Plaque forming unit
	PGM	Personal Genome Machine
	PPi	Inorganic pyrophosphate
	PR8	A/Puerto Rico/8/34
R	rAd	Recombinant adenoviral vector
	RBD	Receptor binding domain
	RdRpol	RNA-dependent-RNA polymerase
	RNA	Ribonucleic acid
	RNP	Ribonucleoprotein
	RT	Room temperature
	RT-PCR	Reverse transcription - polymerase chain reaction
S	SA	Sialic acid
	SCID	Severe combined immunodeficient
	SD	Standard deviation
	SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
	SMS	Single-molecule sequencers
	SNP	Single nucleotide polymorphism
	SOLiD	Sequencing by oligoligation detection
	ssRNA	Single-stranded RNA
T	TEMED	N, N, N, N'-tetramethyl-ethylenediamine
	TPCK	L-1-tosylamide-2-phenylethyl chloromethyl ketone
	TLR	Toll-like receptor
V	VLP	Virus like particle
	vRNA	Viral RNA
	vRNP	Viral ribonucleoprotein
W	WHO	World Health Organization
	wt	Wild type
Z	ZMW	Zero-mode waveguide

Part I: Introduction

Chapter 1

Influenza viruses

1.1. Social and economic impact of influenza viruses

Influenza epidemics have a widespread impact, infecting 250-500 million people each year. Of these, 3 to 5 million result in severe illnesses requiring hospitalization and an estimated 250.000-500.000 people die each year due to influenza [1]. Influenza pandemics are rarer, but usually have a more severe global impact. In the last century, the world population was confronted with four influenza pandemics ('Spanish flu' in 1918, 'Asian flu' in 1957, 'Hong Kong' flu in 1968 and 'Mexican flu' in 2009). The Spanish flu was the most devastating pandemic, with an estimated 500 million persons infected, resulting in fifty millions deaths worldwide [2, 3].

The symptoms that can be associated with influenza virus infection are multiple: fever, cough, sore throat, runny nose, muscle or body aches, headaches, fatigue, vomiting and diarrhea. The elderly (over 65 years), the very young (under 2 years), pregnant women, patients with chronic medical conditions and immuno-compromised patients are susceptible to more severe respiratory disease and death due to influenza [4].

Although influenza is a vaccine preventable disease, the yearly medical cost for the US is estimated between three and five billion dollars [5]. However, the indirect cost due to lost or reduced productivity is estimated to be ten times higher. The percentage of hospitalizations that is prevented by using the current vaccination strategies is estimated to be between 9% and 22% [6]. Seasonal influenza vaccines mainly target the highly antigenically variable viral surface protein hemagglutinin (HA) by eliciting virus neutralizing antibodies. However, influenza viruses are intrinsically genetically variable and can escape immune pressure by mutating the antigenic sites in HA, rendering the vaccine ineffective. In addition, seasonal influenza vaccines contain the viral strains that are predicted to cause the next epidemic. A mismatch between the vaccine strain and the circulating strain, results in suboptimal protection. Taking into account the social and economic impact of influenza infections, there is thus a need for a more successful vaccination strategy, *e.g.* based on more conserved influenza antigens.

1.2. Classification, nomenclature and host range of influenza viruses

Influenza, one of the most highly contagious viral infectious diseases of the upper respiratory tract in humans, is caused by the influenza A, B and occasionally C viruses. The addition of a fourth influenza genus, influenza D, has been suggested recently based on a virus that has been isolated from pigs and cattle [7]. Influenza viruses represent four of the seven genera from the family of *Orthomyxoviridae,* which are characterized by a enveloped, segmented, negative-sense RNA genome. The influenza viruses share the same genetic ancestor and the classification in influenza A, B and C viruses is based on the antigenic differences in two of their internal proteins: matrix protein 1 (M1) and nucleoprotein (NP) [8, 9]. The influenza A viruses are further subdivided based on the antigenic nature of their major surface glycoproteins: HA and neuraminidase (NA) [10]. There are 18

serologically distinguishable HA (H1 - H18) and 11 NA (N1 - N11) subtypes [11, 12]. The HA proteins can be further subdivided into two groups based on their phylogeny: group 1 (H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17 and H18) and group 2 (H3, H4, H7, H10, H14, and H15) [13-16]. The same subdivision into two phylogenetic groups holds true for the NA proteins: group 1 (containing N1, N4, N5 and N8) and group 2 (containing N2, N3, N6, N7 and N9) [17]. Influenza B viruses are more antigenically stable and since the late 1970's diverged into two antigenically distinguishable lineages: B/Yamagata and B/Victoria [18, 19]. Influenza C viruses have a high antigenic stability and are not further subdivided [20, 21].

Influenza viruses follow an internationally accepted naming convention: the antigenic type, followed by the host of origin (omitted if isolated from human), geographical origin, strain number and year of isolation [22]. For influenza A viruses, the HA and NA subtype are added in parentheses. *E.g.* A/Puerto Rico/8/1934 (H1N1), strain number eight of an influenza A virus of the H1N1 subtype isolated in 1934 in Puerto Rico from a human patient.

Wild birds are the natural reservoir for almost all influenza A virus subtypes. In these species, the virus usually replicates asymptomatically in the intestines. However, influenza A infection causes disease in a wide range of vertebrate species, ranging from domestic poultry, horses, pigs to humans. In humans, the influenza A (H3N2 and H1N1) and B viruses are of epidemiological interest, since they cause the recurrent seasonal influenza epidemics, as such they are also the constituents of the annual influenza vaccine [23]. Avian influenza A viruses that cause disease in domestic poultry are divided into two groups based on the elicited disease symptoms: highly pathogenic avian influenza viruses (HPAIV), which can replicate systemically, and low pathogenic avian influenza viruses (LPAIV), which mostly infect the epithelial cells of the intestinal tract. On rare occasions avian influenza viruses from the H5 and H7 subtype have been transmitted to humans, resulting in severe illness. Until now, only a few human-to-human transmissions between close contacts have been reported [24, 25]. However, these viruses are of epidemiological interest since no pre-existing immunity to avian influenza strains is present in the human population. Consequently, there is the fear that these viruses will result in a next influenza pandemic once they obtain sustained human-to-human transmission.

1.3. Influenza A virion

Influenza A virions are pleiomorphic particles ranging from 80 to 120 nm in diameter, with filamentous particles reaching microns in length (Figure 1, schematic representation of a spheric influenza A virion) [10, 27, 28]. Clinical isolates of influenza A viruses frequently produce long filamentous particles, while laboratory-adapted strains are mainly spherical [29, 30]. Each virion contains only one influenza A genome of approximately 13 kb, divided over eight single-stranded, negative-sense RNA segments of varying length [31, 32]. Only a small portion of all influenza A virus particles released from an infected cell is infectious [33]. The remaining part are defective interfering

particles (DIPs) containing at least one viral segment (mainly the polymerases) with a large internal deletion retaining the packaging signals at the segment ends [34, 35]. The viral genome can encode for at least 16 proteins, with new protein products still being discovered, demonstrating the high coding capacity of the rather small viral genome (Table 1).
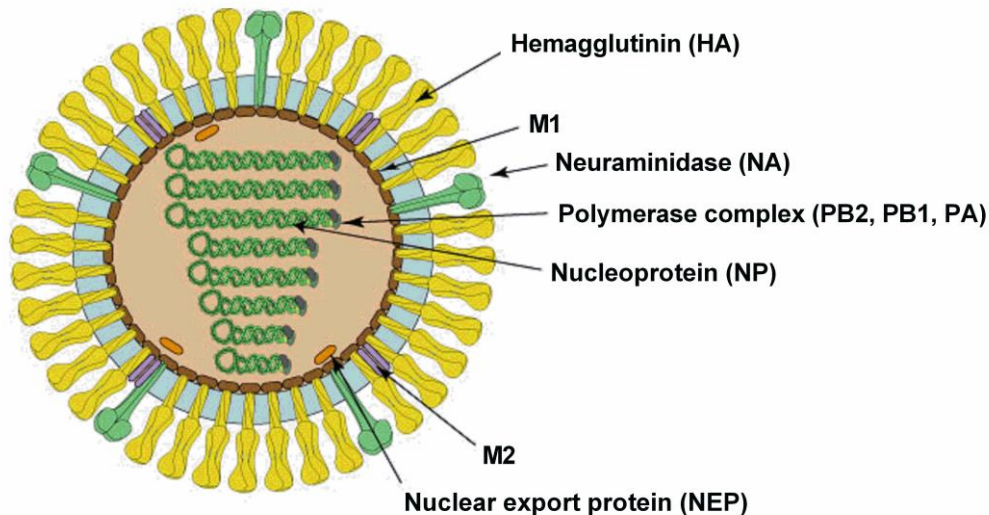
Figure 1: Schematic representation of an influenza A virion. Figure adapted from [26].

The viral membrane contains three membrane proteins: the two protruding glycoproteins HA (~400 HA trimers per virion) and NA (~100 NA tetramers per virion), and the less abundant and much smaller matrix protein M2 (10-50 M2 tetramers per virion) [27, 36, 37]. HA and NA have both a large protruding ectodomain composed of a globular head and a stalk domain, while the M2 ectodomain is small and consists of only 23 amino acids.

HA is a type I transmembrane protein with a bilobed peanut shape and is composed of three identical subunits (HA0) of 550 amino acids (aa) [27]. HA0 has to be cleaved by a host cell protease into two polypeptides, HA1 and HA2, to become functionally active. The HA cleavage site of human and LPAIV contains a monobasic motif, susceptible only to trypsin-like proteases, which limits infection to the respiratory, respectively gastrointestinal tract. In contrast, HPAIV have a polybasic HA cleavage site which can be cleaved by the ubiquitously expressed furin protease, which can result in systemic viral replication. The HA head domain (HA1) plays an important role in viral attachment by binding to sialic acid (SA) on carbohydrate side chains of cell-surface glycoproteins and glycolipids, initiating endocytosis of the virus [49]. The HA receptor specificity is a key determinant of influenza A virus tropism. Human influenza viruses preferentially bind to SA linked to galactose in an α2,6 linkage (SAα2,6Gal) whereas avian influenza viruses prefer the SA in an α2,3 linkage (SAα2,3Gal). This correlates with the tissue tropism of the human and avian influenza viruses in their host: SAα2,6Gal are mainly present on the epithelial cells of the human trachea and SAα2,3Gal are predominantly present on epithelial cells of the avian intestine [65, 66]. However, SAα2,3Gal are also present in the

human lower respiratory tract, which explains the severe disease outcome after infection with avian influenza. The receptor binding specificity of HA can be altered by single amino acid substitutions [67]. The N-terminus of the stalk domain (HA2) is composed of hydrophobic amino acids, the fusion peptide, which is exposed on low pH to trigger membrane fusion [50]. As a result of this, the viral ribonucleoprotein complexes are released into the cell-cytoplasm. The humoral immune response after infection or influenza vaccination mainly targets the antigenic sites of HA expressed on the HA globular head part.

The second abundant surface glycoprotein, **NA**, is a type II transmembrane glycoprotein with a mushroom shape and composed of four identical subunits of 470 aa. Each subunit is composed of an N-terminal cytoplasmic tail, a transmembrane domain, a thin stalk of variable length and a large globular head domain [53]. The length of the NA stalk affects the host range of influenza A viruses. It has been reported, *e.g.*, that a deletion of 20 aa in the NA stalk of H5N1 viruses is favourable for adaptation to chicken [68]. However, this deletion compromises transmission between ferrets and replication on human airway epithelial cells [69]. The NA exosialidase acitivity cleaves α-ketosidic linkages between the SA and an adjacent sugar residue and is important in several stages in the viral replication cycle [70]. Upon viral entry in the lung, the sialidase acitivity of NA is required to cross the mucus layer lining the respiratory tract which contains heavily sialylated glycoproteins [71]. During viral budding, NA cleaves the SA from sialylated viral proteins, hereby preventing the aggregation of viruses. Finally, NA cleaves the SA from glycans on the host cell, to release newly produced virus [53].

The third membrane protein, matrix protein **M2**, is a homotetrameric type III membrane protein which contains a small 23 aa N-terminal ectodomain, a transmembrane α-helical domain and a cytoplasmic domain consisting of a membrane proximal amphiphatic helix and cytoplasmic tail [72]. The ectodomain of M2 (M2e) is highly conserved in nature and can adopt multiple conformations [73-75]. M2 forms a pH-dependent ion channel which is essential for viral uncoating through acidification of the virion and correct maturation of HA in the trans-Golgi network [76, 77]. The ion channel activity of newly synthesized M2 also results in activation of the inflammasome [56]. The cytoplasmic domain of M2 contributes in virion assembly, budding and release [58, 59]. Furthermore, this domain is also involved in subverting the autophagy machinery from the host cell [57]. The M segment in some viral strains also encodes for an alternative splice variant, **M42**, which has an antigenically distinct ectodomain compared to M2 and can functionally replace M2 despite being mainly present in the perinuclear region of the infected cell [60]. We note that a more extended outline on M2 and M2e-based vaccines will be provided in Chapter 2.

Table 1: Coding capacity influenza A virus genome.

Segment	Protein	Length (aa)	Function	Ref.
1	PB2	759	Subunit viral polymerase complex Binding 5' cap of host pre-mRNA in cap-snatching	[38]
	PB2-S1	508	Inhibition RIG-I signaling, interferes with viral polymerase activity	[39]
2	PB1	757	Subunit viral polymerase complex RNA polymerase activity	[40]
	PB1-F2	87	Induction of host-cell apoptosis Influences the polymerase activity (interaction PB1)	[41-44]
	PB1-N40	718	Maintaining balance between PB1 and PB1-F2 expression, exact function unknown	[45]
3	PA	716	Subunit viral polymerase complex Endonuclease activity in cap-snatching	[46]
	PA-X	61	Inhibition of host protein synthesis	[47]
	PA-N155	568	Unknown	[48]
	PA-N182	535	Unknown	[48]
4	HA	550	Binding receptor Membrane fusion Virus budding	[49, 50]
5	NP	498	Binding and protecting vRNA Supporting polymerase complex	[51, 52]
6	NA	454	Host entry Viral budding Virion release	[53]
7	M1	252	Nuclear export vRNPs Virus budding and virion structure	[54, 55]
	M2	97	Virus entry, assembly and budding Autophagy Inflammasome activation	[56-59]
	M42*	99	Can functionally replace M2	[60]
8	NS1	230	Evasion of the host immune system	[61, 62]
	NEP/NS2	121	Export viral RNPs from nucleus	[63]
	NS3*	174	Adaptation to mammalian host	[64]

Protein length refers to A/Puerto Rico/8/34(H1N1). *Expression depends on viral isolate

Lying beneath the lipid envelope is a matrix layer formed by matrix protein **M1**, which is the most abundant protein present in the virion [78]. M1 plays a role in virus budding and surrounds the rod-shaped viral ribonucleoprotein (vRNPs) complexes [55]. A vRNP is a genomic RNA segment wrapped by nucleoprotein (**NP**) oligomers and has a single polymerase complex bound to the 5' and 3' end of the vRNA, which form a partially double-stranded structure through base-pairing [79-81]. NP binds RNA through its highly conserved RNA binding groove and forms oligomers by inserting a flexible tail loop into a neighbouring molecule [51]. The viral polymerase complex is an RNA-dependent-RNA polymerase (RdRpol) and consists of three different viral proteins: Polymerase Basic 1 and 2 (**PB1** and **PB2**) and Polymerase Acidic (**PA**). The viral RdRpol has a relatively high error rate (2.3×10^{-5}) due to absence of 3'→5' exonuclease activity and thus proof-reading activity [82, 83]. Additionally, M1 also interacts in the virion with the nuclear export protein (**NEP**), also known as non-structural protein 2 (NS2). NEP is responsible, together with M1, for nuclear export of newly produced vRNPs in infected cells [63]. The presence of "non-structural" protein 1 (**NS1**), in low abundance, in the virion has recently been described [78]. NS1 is a multifunctional protein with a major role in evasion of the host immune system by inhibiting cellular gene expression and antagonising the effector functions of IFN [61, 62]. Its role in the virion is unclear, but it could potentially play a role in virion assembly or enhance infection when introduced into a new host cell [78].

Several influenza proteins are also expressed in the infected cell, but are absent in the virion [78]. **PB2-S1** has recently been described to be encoded by a splice variant of PB2 mRNA. PB2-S1 localizes to mitochondria, inhibits the RIG-I dependent interferon signaling pathway and interferes with the viral polymerase activity [39]. **PB1-F2** is encoded by an alternate reading frame in the PB1 gene segment and targets the mitochondria, leading to apoptosis [41-43]. PB1-F2 contributes to pathogenicity in avian hosts but has a minimal effect on virulence on humans [84]. The PB1 segment also encodes for an N-terminally truncated form, **PB1-N40**, with a so far unknown function but lacks polymerase activity and interacts with PB2, PB1, NP and PB1-F2 in the nucleus [45, 85]. **PA-X** is expressed as a result of ribosomal frameshifting and contains the N-terminal endonuclease domain and an alternative C-terminus [86]. PA-X inhibits host protein synthesis in the infected cell, hereby suppressing the antiviral response [47]. In addition, two N-terminally truncated forms of PA have been described recently, **PA-N155** and **PA-N182**, with a suggested role for enhanced influenza A virus replication and pathogenicity [48]. **NS3**, a splice-variant of NS1 mRNA, is present in some viral strains and is associated with host-adaptation from avian to mammalian hosts [64].
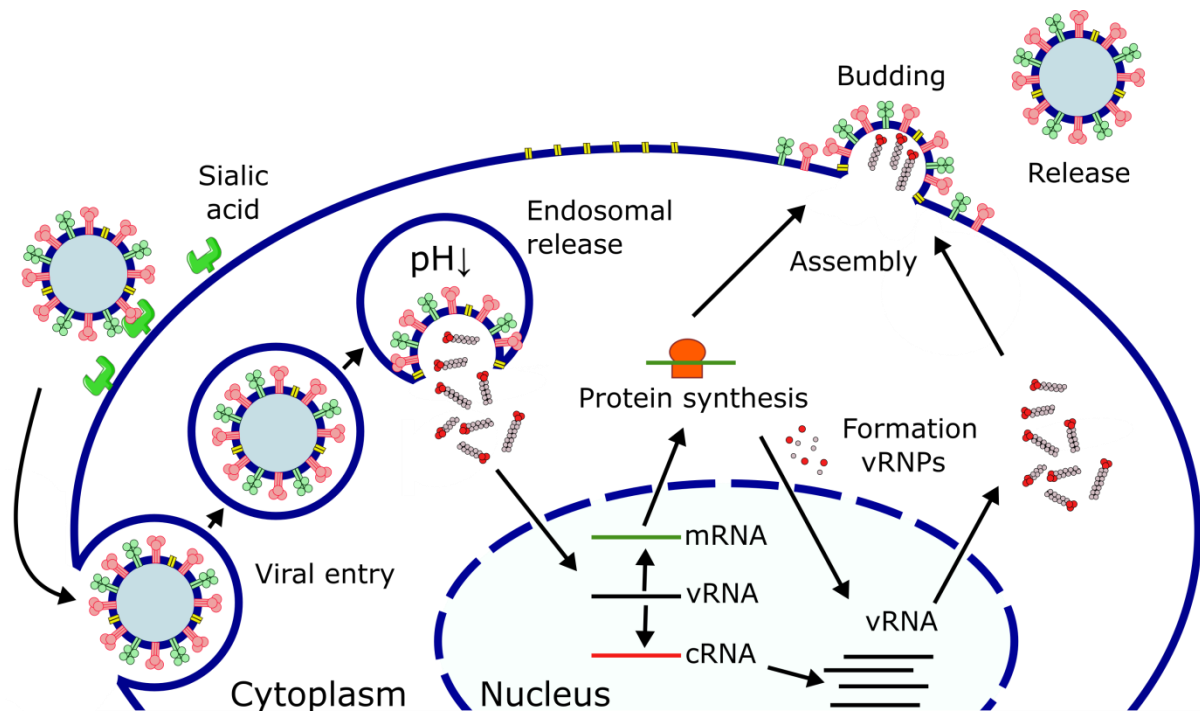
Figure 2: Schematic representation of the influenza A virus replication cycle.

1.4. Virus replication

A schematic representation of the influenza A virus replication cycle is shown in Figure 2. Initiation of virus infection involves binding of HA to SA on carbohydrate side chains of cell-surface glycoproteins and glycolipids [49]. Upon binding, the virion is engulfed by either clathrin-dependent-receptor-mediated endocytosis or macropinocytosis and trafficked to an endosomal compartment [87-89]. The low endosomal pH induces a conformational change in HA, exposing the HA2 fusion peptide. This hydrophobic fusion peptide inserts itself in the endosomal membrane, resulting in fusion of the viral and endosomal membrane [50]. In parallel, the ion channel activity of M2 is also activated by the acidic environment in the endosome, resulting in acidification of the viral core and release from the vRNPs in the cytoplasm. These vRNPs are then imported to the nucleus for viral transcription and replication [90, 91]. Their nuclear import is mediated by nuclear localization sequences (NLS) on the vRNPs, which are exposed after they dissociate from M1 [92]. Since influenza viruses have a negative-sense RNA genome, a positive-sense copy is synthesized for both viral transcription and viral replication. The synthesis of viral mRNA is initiated by a process called 'cap snatching', in which a 5' cap of a host cell derived pre-mRNA is bound by PB2 and cleaved-off by the endonuclease activity of PA at approximately 10 to 13 nucleotides from the cap structure [38, 46]. This 5' cap is subsequently used as an RNA primer and elongated by the RNA polymerase activity of PB1 [40]. Transcription is performed by the polymerase complex initially bound to the vRNP and supported by NP [52, 93]. Transcriptional elongation is terminated when the poly-U stretch near the 5'-terminus of the vRNA is transcribed by reiterative stuttering [94, 95]. This results in an mRNA containing a 5' cap

structure and 3' poly(A) tail, resembling the cellular mRNAs. Afterwards, the synthesized viral mRNA is exported to the cytoplasm for translation to viral proteins by the host cell machinery. The newly produced proteins are then transported to the cell surface or to the nucleus to form new vRNPs. Viral transcription dominates early in infection, with replication becoming more abundant with progression of the infection [96]. In contrast to transcription, replication requires a newly produced polymerase complex and NP proteins to synthesize a positive-sense copy RNA (cRNA) from which new full-length vRNAs will be synthesized [97]. Newly synthesized vRNAs are co-transcriptionally encapsidated by NP and a polymerase complex, forming vRNPs. Nuclear export of these progeny vRNPs is mediated by M1 and NEP, which both contain nuclear export signals (NES) [98, 99]. Afterwards, the vRNPs are transported across the cytoplasm to the apical cell membrane, where assembly of progeny virions takes place in lipid raft domains. The M1 protein plays an important role in viral assembly and budding since it interacts both with the viral membrane proteins and the vRNPs. The presence of all eight genomic segments in a single virion is ensured by the packaging signals at their 5' and 3' termini [100]. Influenza viruses obtain their viral envelope during budding from the infected cell. The virions are finally released by cleavage of sialic acid from the infected cell by NA.

1.5. Influenza quasispecies, antigenic drift and shift.

The huge diversity in an influenza A virus population is the consequence of the relatively high error-rate of its RNA polymerase (2.3×10^{-5}), in combination with the large population size and rapid replication kinetics [82, 83]. Influenza viruses, as many other RNA viruses, exists as a large population of closely related genotypes linked through mutation: the influenza quasispecies [101, 102]. As a consequence, a given consensus sequence for an influenza virus is thus the average sequence of all viruses present in that sample [103]. The immense variation present in the viral quasispecies helps the virus to easily adapt to a changing environment. In addition, it is suggested that selection takes places at the population level, rather than on the individual genomes [104, 105].

The humoral immune response induced during seasonal vaccination or infection mainly generates neutralizing antibodies against the influenza virus major surface proteins HA and NA. Upon infection and due to the intrinsicly high genetic variability of influenza viruses, these antibodies select for viruses with amino acid substitutions in HA and/or NA that retain their structure and function, but to which the antibodies can no longer bind. This is called antigenic drift and is a continuous and unpredictable process, since each introduced amino acid substitution changes the mutation landscape [106]. Once a virus emerges that is no longer inhibited by the host humoral immune response elicited by the seasonal influenza vaccine, re-infection can occur and the virus can spread more rapidly among the population, resulting in the yearly recurrence of influenza epidemics. Antigenic drift is the main reason why the current influenza vaccines are updated annually since the vaccines are only effective when the vaccine antigenically matches the circulating virus strains. The

process of antigenic drift can be recapitulated *in vitro* and has been extensively studied [106-108]. Escape to neutralizing HA antibodies has also been shown to correlate with increased receptor avidity, suggesting a critical role of receptor binding avidity in antigenic drift [109]. The head domain of HA contains at least four different, surface-exposed highly-variable antigenic sites: Sa, Sb, Ca and Cb to which virus neutralizing antibodies can bind [107]. HA is inherently highly tolerant for mutations in its antigenic sites, contributing to influenza's rapid antigenic evolution [110]. NA is also subject to antigenic drift and carries three surface-exposed antigenic sites: A, B and C [111, 112]. However, the sequence evolution of NA is slower compared to HA1 [113].

Antigenic shift is rather rare compared to antigenic drift and results in the introduction of a new influenza virus subtype from one species into another species, to which the new host has no pre-existing immunity. Influenza viruses can follow three different routes leading to antigenic shift [114]. The first route is a consequence of the segmented nature of influenza viruses and takes place when an avian and human influenza virus from a different subtype infect a single cell in an intermediate host, *e.g.* pig. During assembly, exchange of viral segments between the two viruses can take place, resulting in the release of a virus from a different subtype that can infect a human host. Antigenic shift can also take place when an avian influenza virus first infects an intermediate host, followed by infection of a human host or directly, when an avian influenza virus crosses the species border and directly infects a human host. When such a antigenic shifted virus gains efficient human-to-human transmission, antigenic shift will introduce a viral subtype into the human population to which that population is immunologically (largely) naive. This can result in an influenza pandemic. Four human influenza pandemics in the last century were the result of antigenic shift by reassortment: The 'Spanish Flu' (H1N1) in 1918, the 'Asian Flu' (H2N2) in 1957, the 'Hong Kong Flu' (H3N2) in 1968 and the recent H1N1 pandemic in 2009 [115-118].

A third mechanism introducing genetic diversity in the viral population is recombination. Recombination can take place when two different influenza viruses infect a single cell and RNA segments of both viruses are coupled by strand replacement during replication. In contrast to antigenic drift and shift, recombination in influenza viruses is less well studied and its importance in the evolution of influenza viruses remains controversial [119-121].

1.6. Influenza control

1.6.1. Vaccination

Vaccination is the primary means to protect humans against influenza virus infection and control virus spread. The WHO recommends seasonal influenza vaccination for pregnant women, children between six and 59 months, elderly, individuals with specific chronic medical conditions and health-care workers.

Seasonal influenza vaccine

Protection induced by the current influenza vaccines correlates primarily with the induction of strain-specific neutralizing antibodies mainly targeting the antigenic sites in the head domains of HA, and to a lesser extent NA. Antigenic drift in HA and NA can lead to escape from antibody recognition and loss of vaccine effectiveness. As a consequence, the vaccine is continually assessed for cross-reactivity with currently circulating strains and its composition is re-evaluated twice a year. The WHO makes recommendations on the composition of the influenza vaccines in February and September based on the results of surveillance, laboratory (*e.g.* testing for antiviral resistance) and clinical studies (*e.g.* human serology studies) by the WHO Global Influenza Surveillance and Response System.

Three different types of seasonal influenza vaccines are approved for use in Europe: inactivated 'split' virus vaccines, subunit vaccines or live-attenuated virus vaccines. The traditional vaccines, 'the trivalent influenza vaccine', contains three different viral strains (or parts thereof): an influenza A H1N1 virus, an influenza A H3N2 virus and an influenza B virus. Since the two influenza B lineages are co-circulating, a 'quadrivalent vaccine', in which both the B/Victoria and B/Yamagata virus strain are included, was recently developed to give broader protection. The vaccine efficacy thus depends on how well the vaccine strains matches the viral strains that will circulate in the upcoming season. In general, the vaccine effectiveness is estimated to be ~40-60% for the three different viruses in the vaccine [122]. If the strains in the vaccine match the circulating ones, the vaccine efficacy in healthy young adults ranges from 70 to 90 % [123]. The vaccine viruses are created using classical reassortment or reverse genetics. The majority of vaccine doses are produced by amplifying the selected vaccine strains in embryonated chicken eggs, with a time-span of five to six months starting from isolation of the vaccine strain to bringing the vaccine on the market. This is an important disadvantage in pandemic preparedness: when suddenly an influenza virus emerges with pandemic potential, the new vaccine doses will only be on the market when the virus has already spread through the population. This was the case for the vaccine against the pandemic H1N1 virus in 2009.

Despite being a critical tool in combating influenza infection worldwide, current available influenza vaccines are limited by their subtype-specificity and the ease of influenza viruses to evade the elicited humoral immune response. There is thus a need for a single vaccine that provides long-lasting immunity against a broad spectrum of influenza viruses: the so-called 'universal influenza vaccine'.

Universal influenza vaccine

The 'universal influenza vaccine' candidates target epitopes present in influenza viral proteins that are conserved between influenza subtypes, with the aim to provide heterosubtypic protection. In addition, the broadness of protection of the vaccine can be increased if both the humoral and the

14

cell-mediated immune system are activated upon vaccination. Several 'universal influenza vaccine' candidates have been proposed eliciting a cross-reactive antibody-inducing and/or a cross-reactive T-cell response. The cross-reactive antibody-inducing response is directed against conserved, extracellular parts of influenza membrane proteins, with the HA2 stem domain and the M2 ectodomain (discussed in Chapter 2) being the major targets [124, 125]. In contrast to the seasonal vaccines eliciting mainly neutralizing antibodies, these vaccines are infection permissive, with protection mainly mediated through antibody-dependent cellular cytotoxicity and/or phagocytosis [126]. The cross-reactive T cell-inducing antigens are typically directed against conserved internal structural proteins, with the influenza M1 and NP being the major targets [127]. Several 'universal influenza vaccine' candidates are currently being evaluated in clinical trials [128].

1.6.2. Antivirals

Influenza antivirals are used to prevent viral infection and treat infected individuals. Although vaccination is the method of choice for influenza prophylaxis, influenza antivirals are the first line of defense against emerging, antigenically different influenza strains and are thus stockpiled for pandemic preparedness. Two classes of antiviral drugs have been approved by the European Centre for Disease Prevention and Control: M2 inhibitors (the adamantanes: amantadine (Symmetrel®) and rimantadine (Flumadine®)) and neuraminidase inhibitors (oseltamivir (Tamiflu®), zanamivir (Relenza®)). The NA inhibitors Peramivir (Rapivab®) and Laninamivir (Inavir®) are currently in Phase III clinical trials in Europe.

The M2 inhibitors, the adamantanes, exert their antiviral activity by blocking the ion channel activity of M2 and hence preventing the uncoating of the virus in infected cells. The adamantanes are only effective against influenza A viruses and their use is limited due to several toxic effects and the rapid emergence of drug-resistant variants [129]. A single amino acid substitution can result in resistance, with the predominant mutation being the M2-S31N. The resistance to adamantanes remains high among circulating influenza A isolates, with resistance detected among all tested influenza A(H3N2) and pandemic 2009 H1N1 viruses [130]. Furthermore, resistance was also detected in highly pathogenic H5N1 viruses isolated from poultry recently, limiting their use in pandemic preparedness [131-134]. Due to the high levels of antiviral resistance among circulating influenza viruses, these antivirals are no longer recommended for use by the WHO.

The second class of influenza antivirals, the NA inhibitors, inhibit the enzymatic activity of neuraminidase by mimicking its substrate sialic acid. As a consequence, the release and spread of influenza viruses to other cells is prevented. The NA inhibitors are effective against influenza A and B viruses. The most frequently detected resistance mutation to Oseltamivir is H275Y, however, viral surveillance and resistance data indicate that >99% of currently circulating influenza virus strains are sensitive to the NA inhibitors [130].

The use of currently available influenza antivirals are limited by the emergence of drug resistance. There is thus a demanding need to develop new antivirals with reduced drug resistance potential and novel mechanisms of action. A newly developed influenza antiviral is Favipiravir (T-705 or Avigan), which is currently being evaluated in clinical trials. Favipiravir selectively targets the viral RNA polymerase complex, thereby inhibiting viral replication [135]. Favipiravir is converted by the cellular kinases to its active form, ribofuranosyl triphosphate, which block the RNA polymerase complex. Favipiravir is active against influenza A, B and C viruses, and against viruses resistant to the currently used antivirals. Recently, the broad-spectrum antiviral activity of favipiravir against a variety of RNA viruses, including Ebola virus, has been demonstrated [136, 137]. The availability of crystal structures results in the development of structure-based drug design, with several candidates being evaluated for their effectiveness as influenza antiviral, *e.g.* HB36.6 binding with high affinity to the conserved HA stem domain [138]. The possibility of specifically targeting the host cellular mechanisms activated upon influenza infection are also currently being explored as an antiviral strategy [139].

References

1. Influenza (Seasonal) [http://www.who.int/mediacentre/factsheets/fs211/en/]
2. Patterson KD, Pyle GF: The geography and mortality of the 1918 influenza pandemic. *Bulletin of the history of medicine* 1991, 65(1):4-21.
3. Taubenberger JK, Morens DM: 1918 Influenza: the mother of all pandemics. *Emerging infectious diseases* 2006, 12(1):15-22.
4. Prevention CfDCa: 2009 H1N1: Overview of a Pandemic - Background on influenza. 2010.
5. Patriarca PA: New options for prevention and control of influenza. *Jama* 1999, 282(1):75-77.
6. Estimated Influenza Illnesses and Hospitalizations Averted by Vaccination - United States, 2014-15 Influenza Season [http://www.cdc.gov/flu/about/disease/2014-15.htm#references]
7. Hause BM, Collin EA, Liu R, Huang B, Sheng Z, Lu W, Wang D, Nelson EA, Li F: Characterization of a novel influenza virus in cattle and Swine: proposal for a new genus in the Orthomyxoviridae family. *mBio* 2014, 5(2):e00031-00014.
8. Tille PM: Bailey and Scott's Diagnostic Microbiology. 2014, 13th Edition:836.
9. Bouvier NM, Palese P: The biology of influenza viruses. *Vaccine* 2008, 26 Suppl 4:D49-53.
10. Lamb RA, Krug RM: Orthomyxoviridae: the viruses and their replication. *Fields Virology* 2001, 4th edition:1487-1531.
11. Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA *et al*: A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(11):4269-4274.
12. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J *et al*: New world bats harbor diverse influenza A viruses. *PLoS pathogens* 2013, 9(10):e1003657.
13. Air GM: Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America* 1981, 78(12):7639-7643.
14. Nobusawa E, Aoyama T, Kato H, Suzuki Y, Tateno Y, Nakajima K: Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* 1991, 182(2):475-485.
15. Laursen NS, Wilson IA: Broadly neutralizing antibodies against influenza viruses. *Antiviral research* 2013, 98(3):476-483.
16. Wiersma LC, Rimmelzwaan GF, de Vries RD: Developing Universal Influenza Vaccines: Hitting the Nail, Not Just on the Head. *Vaccines* 2015, 3(2):239-262.
17. Air GM, Laver WG: The neuraminidase of influenza virus. *Proteins* 1989, 6(4):341-356.
18. Hay AJ, Gregory V, Douglas AR, Lin YP: The evolution of human influenza viruses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2001, 356(1416):1861-1870.
19. Rota PA, Wallis TR, Harmon MW, Rota JS, Kendal AP, Nerome K: Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology* 1990, 175(1):59-68.
20. Meier-Ewert H, Petri T, Bishop DH: Oligonucleotide fingerprint analyses of influenza C virion RNA recovered from five different isolates. *Archives of virology* 1981, 67(2):141-147.
21. Sugawara K, Nakamura K, Homma M: Analyses of structural polypeptides of seven different isolates of influenza C virus. *The Journal of general virology* 1983, 64 Pt 3:579-587.
22. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bulletin of the World Health Organization* 1980, 58(4):585-591.
23. Krug RM, Aramini JM: Emerging antiviral targets for influenza A virus. *Trends in pharmacological sciences* 2009, 30(6):269-277.
24. Human cases of avian influenza A (H5N1) in North-West Frontier Province, Pakistan, October-November 2007. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations* 2008, 83(40):359-364.
25. Wang H, Feng Z, Shu Y, Yu H, Zhou L, Zu R, Huai Y, Dong J, Bao C, Wen L *et al*: Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. *Lancet* 2008, 371(9622):1427-1434.
26. Influenzavirus A [http://viralzone.expasy.org/all_by_species/6.html]

27. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC: Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103(50):19123-19127.

28. Roberts PC, Lamb RA, Compans RW: The M1 and M2 proteins of influenza A virus are important determinants in filamentous particle formation. *Virology* 1998, 240(1):127-137.

29. Nakajima N, Hata S, Sato Y, Tobiume M, Katano H, Kaneko K, Nagata N, Kataoka M, Ainai A, Hasegawa H *et al*: The first autopsy case of pandemic influenza (A/H1N1pdm) virus infection in Japan: detection of a high copy number of the virus in type II alveolar epithelial cells by pathological and virological examination. *Japanese journal of infectious diseases* 2010, 63(1):67-71.

30. Yamaguchi M, Danev R, Nishiyama K, Sugawara K, Nagayama K: Zernike phase contrast electron microscopy of ice-embedded influenza A virus. *Journal of structural biology* 2008, 162(2):271-276.

31. Noda T, Sagara H, Yen A, Takada A, Kida H, Cheng RH, Kawaoka Y: Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* 2006, 439(7075):490-492.

32. Chou YY, Vafabakhsh R, Doganay S, Gao Q, Ha T, Palese P: One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(23):9101-9106.

33. Brooke CB, Ince WL, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW: Most influenza a virions fail to express at least one essential viral protein. *Journal of virology* 2013, 87(6):3155-3162.

34. Davis AR, Hiti AL, Nayak DP: Influenza defective interfering viral RNA is formed by internal deletion of genomic RNA. *Proceedings of the National Academy of Sciences of the United States of America* 1980, 77(1):215-219.

35. Saira K, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V *et al*: Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *Journal of virology* 2013, 87(14):8064-8074.

36. Ruigrok RW, Andree PJ, Hooft van Huysduynen RA, Mellema JE: Characterization of three highly purified influenza virus strains by electron microscopy. *The Journal of general virology* 1984, 65 (Pt 4):799-802.

37. Zebedee SL, Lamb RA: Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of virology* 1988, 62(8):2762-2772.

38. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ *et al*: The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature structural & molecular biology* 2008, 15(5):500-506.

39. Yamayoshi S, Watanabe M, Goto H, Kawaoka Y: Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *Journal of virology* 2016, 90(1):444-456.

40. Kobayashi M, Toyoda T, Ishihama A: Influenza virus PB1 protein is the minimal and essential subunit of RNA polymerase. *Archives of virology* 1996, 141(3-4):525-539.

41. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P *et al*: A novel influenza A virus mitochondrial protein that induces cell death. *Nature medicine* 2001, 7(12):1306-1312.

42. Yamada H, Chounan R, Higashi Y, Kurihara N, Kido H: Mitochondrial targeting sequence of the influenza A virus PB1-F2 protein and its function in mitochondria. *FEBS letters* 2004, 578(3):331-336.

43. Zamarin D, Garcia-Sastre A, Xiao X, Wang R, Palese P: Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS pathogens* 2005, 1(1):e4.

44. Zamarin D, Ortigoza MB, Palese P: Influenza A virus PB1-F2 protein contributes to viral pathogenesis in mice. *Journal of virology* 2006, 80(16):7976-7983.

45. Wise HM, Foeglein A, Sun J, Dalton RM, Patel S, Howard W, Anderson EC, Barclay WS, Digard P: A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *Journal of virology* 2009, 83(16):8021-8031.

46. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW: The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 2009, 458(7240):914-918.

47. Hayashi T, MacDonald LA, Takimoto T: Influenza A Virus Protein PA-X Contributes to Viral Growth and Suppression of the Host Antiviral and Immune Responses. *Journal of virology* 2015, 89(12):6442-6452.

48. Muramoto Y, Noda T, Kawakami E, Akkina R, Kawaoka Y: Identification of novel influenza A virus proteins translated from PA mRNA. *Journal of virology* 2013, 87(5):2455-2462.

18

49.   Gambaryan AS, Tuzikov AB, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS, Bovin NV, Matrosovich MN: Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyllactosamine). *Virology* 1997, 232(2):345-350.

50.   Cross KJ, Langley WA, Russell RJ, Skehel JJ, Steinhauer DA: Composition and functions of the influenza fusion peptide. *Protein and peptide letters* 2009, 16(7):766-778.

51.   Ye Q, Krug RM, Tao YJ: The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* 2006, 444(7122):1078-1082.

52.   Marklund JK, Ye Q, Dong J, Tao YJ, Krug RM: Sequence in the influenza A virus nucleoprotein required for viral polymerase binding and RNA synthesis. *Journal of virology* 2012, 86(13):7292-7297.

53.   Air GM: Influenza neuraminidase. *Influenza and other respiratory viruses* 2012, 6(4):245-256.

54.   Bui M, Wills EG, Helenius A, Whittaker GR: Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *Journal of virology* 2000, 74(4):1781-1786.

55.   Gomez-Puertas P, Albo C, Perez-Pastrana E, Vivo A, Portela A: Influenza virus matrix protein is the major driving force in virus budding. *Journal of virology* 2000, 74(24):11538-11547.

56.   Ichinohe T, Pang IK, Iwasaki A: Influenza virus activates inflammasomes via its intracellular M2 ion channel. *Nature immunology* 2010, 11(5):404-410.

57.   Beale R, Wise H, Stuart A, Ravenhill BJ, Digard P, Randow F: A LC3-interacting motif in the influenza A virus M2 protein is required to subvert autophagy and maintain virion stability. *Cell host & microbe* 2014, 15(2):239-247.

58.   Rossman JS, Jing X, Leser GP, Lamb RA: Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* 2010, 142(6):902-913.

59.   Rossman JS, Lamb RA: Influenza virus assembly and budding. *Virology* 2011, 411(2):229-236.

60.   Wise HM, Hutchinson EC, Jagger BW, Stuart AD, Kang ZH, Robb N, Schwartzman LM, Kash JC, Fodor E, Firth AE *et al*: Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS pathogens* 2012, 8(11):e1002998.

61.   Krug RM: Functions of the influenza A virus NS1 protein in antiviral defense. *Current opinion in virology* 2015, 12:1-6.

62.   Marc D: Influenza virus non-structural protein NS1: interferon antagonism and beyond. *The Journal of general virology* 2014, 95(Pt 12):2594-2611.

63.   Shimizu T, Takizawa N, Watanabe K, Nagata K, Kobayashi N: Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. *FEBS letters* 2011, 585(1):41-46.

64.   Selman M, Dankar SK, Forbes NE, Jia JJ, Brown EG: Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerging microbes & infections* 2012, 1(11):e42.

65.   Couceiro JN, Paulson JC, Baum LG: Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus research* 1993, 29(2):155-165.

66.   Rogers GN, Paulson JC: Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* 1983, 127(2):361-373.

67.   Glaser L, Stevens J, Zamarin D, Wilson IA, Garcia-Sastre A, Tumpey TM, Basler CF, Taubenberger JK, Palese P: A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *Journal of virology* 2005, 79(17):11533-11536.

68.   Li Y, Chen S, Zhang X, Fu Q, Zhang Z, Shi S, Zhu Y, Gu M, Peng D, Liu X: A 20-amino-acid deletion in the neuraminidase stalk and a five-amino-acid deletion in the NS1 protein both contribute to the pathogenicity of H5N1 avian influenza viruses in mallard ducks. *PloS one* 2014, 9(4):e95539.

69.   Blumenkrantz D, Roberts KL, Shelton H, Lycett S, Barclay WS: The short stalk length of highly pathogenic avian influenza H5N1 virus neuraminidase limits transmission of pandemic H1N1 virus in ferrets. *Journal of virology* 2013, 87(19):10539-10551.

70.   Gottschalk A: Neuraminidase: the specific enzyme of influenza virus and Vibrio cholerae. *Biochimica et biophysica acta* 1957, 23(3):645-646.

71.   Cohen M, Zhang XQ, Senaati HP, Chen HW, Varki NM, Schooley RT, Gagneux P: Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology journal* 2013, 10:321.

72. Schnell JR, Chou JJ: Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 2008, 451(7178):591-595.

73. Deng L, Cho KJ, Fiers W, Saelens X: M2e-Based Universal Influenza A Vaccines. *Vaccines* 2015, 3(1):105-136.

74. Cho KJ, Schepens B, Moonens K, Deng L, Fiers W, Remaut H, Saelens X: Crystal Structure of the Conserved Amino Terminus of the Extracellular Domain of Matrix Protein 2 of Influenza A Virus Gripped by an Antibody. *Journal of virology* 2016, 90(1):611-615.

75. Cho KJ, Schepens B, Seok JH, Kim S, Roose K, Lee JH, Gallardo R, Van Hamme E, Schymkowitz J, Rousseau F *et al*: Structure of the extracellular domain of matrix protein 2 of influenza A virus in complex with a protective monoclonal antibody. *Journal of virology* 2015, 89(7):3700-3711.

76. Wharton SA, Belshe RB, Skehel JJ, Hay AJ: Role of virion M2 protein in influenza virus uncoating: specific reduction in the rate of membrane fusion between virus and liposomes by amantadine. *The Journal of general virology* 1994, 75 (Pt 4):945-948.

77. Ivanovic T, Rozendaal R, Floyd DL, Popovic M, van Oijen AM, Harrison SC: Kinetics of proton transport into influenza virions by the viral M2 channel. *PloS one* 2012, 7(3):e31566.

78. Hutchinson EC, Charles PD, Hester SS, Thomas B, Trudgian D, Martinez-Alonso M, Fodor E: Conserved and host-specific features of influenza virion architecture. *Nature communications* 2014, 5:4816.

79. Moeller A, Kirchdoerfer RN, Potter CS, Carragher B, Wilson IA: Organization of the influenza virus replication machinery. *Science* 2012, 338(6114):1631-1634.

80. Arranz R, Coloma R, Chichon FJ, Conesa JJ, Carrascosa JL, Valpuesta JM, Ortin J, Martin-Benito J: The structure of native influenza virion ribonucleoproteins. *Science* 2012, 338(6114):1634-1637.

81. Desselberger U, Racaniello VR, Zazra JJ, Palese P: The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* 1980, 8(3):315-328.

82. Nobusawa E, Sato K: Comparison of the mutation rates of human influenza A and B viruses. *Journal of virology* 2006, 80(7):3675-3678.

83. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R: Viral mutation rates. *Journal of virology* 2010, 84(19):9733-9748.

84. Schmolke M, Manicassamy B, Pena L, Sutton T, Hai R, Varga ZT, Hale BG, Steel J, Perez DR, Garcia-Sastre A: Differential contribution of PB1-F2 to the virulence of highly pathogenic H5N1 influenza A virus in mammalian and avian species. *PLoS pathogens* 2011, 7(8):e1002186.

85. Tauber S, Ligertwood Y, Quigg-Nicol M, Dutia BM, Elliott RM: Behaviour of influenza A viruses differentially expressing segment 2 gene products in vitro and in vivo. *The Journal of general virology* 2012, 93(Pt 4):840-849.

86. Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL *et al*: An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 2012, 337(6091):199-204.

87. Rossman JS, Leser GP, Lamb RA: Filamentous influenza virus enters cells via macropinocytosis. *Journal of virology* 2012, 86(20):10950-10960.

88. Chen C, Zhuang X: Epsin 1 is a cargo-specific adaptor for the clathrin-mediated endocytosis of the influenza virus. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105(33):11790-11795.

89. Lakadamyali M, Rust MJ, Zhuang X: Ligands for clathrin-mediated endocytosis are differentially sorted into distinct populations of early endosomes. *Cell* 2006, 124(5):997-1009.

90. Herz C, Stavnezer E, Krug R, Gurney T, Jr.: Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. *Cell* 1981, 26(3 Pt 1):391-400.

91. Jackson DA, Caton AJ, McCready SJ, Cook PR: Influenza virus RNA is synthesized at fixed sites in the nucleus. *Nature* 1982, 296(5855):366-368.

92. Wu WW, Sun YH, Pante N: Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virology journal* 2007, 4:49.

93. Jorba N, Coloma R, Ortin J: Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS pathogens* 2009, 5(5):e1000462.

94. Li X, Palese P: Characterization of the polyadenylation signal of influenza virus RNA. *Journal of virology* 1994, 68(2):1245-1249.

95.   Poon LL, Pritlove DC, Fodor E, Brownlee GG: Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *Journal of virology* 1999, 73(4):3473-3476.

96.   Shapiro GI, Gurney T, Jr., Krug RM: Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of virology* 1987, 61(3):764-773.

97.   Vreede FT, Jung TE, Brownlee GG: Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *Journal of virology* 2004, 78(17):9568-9572.

98.   Martin K, Helenius A: Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell* 1991, 67(1):117-130.

99.   O'Neill RE, Talon J, Palese P: The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *The EMBO journal* 1998, 17(1):288-296.

100.   Hutchinson EC, von Kirchbach JC, Gog JR, Digard P: Genome packaging in influenza A virus. *The Journal of general virology* 2010, 91(Pt 2):313-328.

101.   Lauring AS, Andino R: Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens* 2010, 6(7):e1001005.

102.   Eigen M, Schuster P: The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Die Naturwissenschaften* 1977, 64(11):541-565.

103.   Eigen M: Viral quasispecies. *Scientific American* 1993, 269(1):42-49.

104.   Borderia AV, Isakov O, Moratorio G, Henningsson R, Aguera-Gonzalez S, Organtini L, Gnadig NF, Blanc H, Alcover A, Hafenstein S *et al*: Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS pathogens* 2015, 11(5):e1004838.

105.   Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R: Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006, 439(7074):344-348.

106.   Das SR, Hensley SE, Ince WL, Brooke CB, Subba A, Delboy MG, Russ G, Gibbs JS, Bennink JR, Yewdell JW: Defining influenza A virus hemagglutinin antigenic drift by sequential monoclonal antibody selection. *Cell host & microbe* 2013, 13(3):314-323.

107.   Caton AJ, Brownlee GG, Yewdell JW, Gerhard W: The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* 1982, 31(2 Pt 1):417-427.

108.   Li C, Hatta M, Burke DF, Ping J, Zhang Y, Ozawa M, Taft AS, Das SC, Hanson AP, Song J *et al*: Selection of antigenically advanced variants of seasonal influenza viruses. *Nature microbiology* 2016, 1(6):16058.

109.   Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR *et al*: Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* 2009, 326(5953):734-736.

110.   Thyagarajan B, Bloom JD: The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 2014, 3.

111.   Luther P, Bergmann KC, Oxford JS: An investigation of antigenic drift of neuraminidases of influenza A (H1N1) viruses. *The Journal of hygiene* 1984, 92(2):223-229.

112.   Air GM, Els MC, Brown LE, Laver WG, Webster RG: Location of antigenic sites on the three-dimensional structure of the influenza N2 virus neuraminidase. *Virology* 1985, 145(2):237-248.

113.   Westgeest KB, de Graaf M, Fourment M, Bestebroer TM, van Beek R, Spronken MI, de Jong JC, Rimmelzwaan GF, Russell CA, Osterhaus AD *et al*: Genetic evolution of the neuraminidase of influenza A (H3N2) viruses from 1968 to 2009 and its correspondence to haemagglutinin evolution. *The Journal of general virology* 2012, 93(Pt 9):1996-2007.

114.   Antigenic   shift [https://www.niaid.nih.gov/topics/Flu/Research/basic/Pages/AntigenicShiftIllustration.aspx]

115.   Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V *et al*: Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 2009, 325(5937):197-201.

116.   Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S *et al*: Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009, 459(7250):1122-1125.

117.   Lindstrom SE, Cox NJ, Klimov A: Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. *Virology* 2004, 328(1):101-119.

118.   Belshe RB: The origins of pandemic influenza--lessons from the 1918 virus. *The New England journal of medicine* 2005, 353(21):2209-2211.

119.   Boni MF, de Jong MD, van Doorn HR, Holmes EC: Guidelines for identifying homologous recombination events in influenza A virus. *PloS one* 2010, 5(5):e10434.

120.   De A, Sarkar T, Nandy A: Bioinformatics studies of Influenza A hemagglutinin sequence data indicate recombination-like events leading to segment exchanges. *BMC research notes* 2016, 9(1):222.

121.   Boni MF, Zhou Y, Taubenberger JK, Holmes EC: Homologous recombination is very rare or absent in human influenza A virus. *Journal of virology* 2008, 82(10):4807-4811.

122.   Influenza   vaccination [http://ecdc.europa.eu/en/healthtopics/seasonal_influenza/vaccines/Pages/influenza_vaccination.aspx#types]

123.   Regoes RR, Bonhoeffer S: Emergence of drug-resistant influenza virus: population dynamical considerations. *Science* 2006, 312(5772):389-391.

124.   Hashem AM: Prospects of HA-based universal influenza vaccine. *BioMed research international* 2015, 2015:414637.

125.   Schotsaert M, De Filette M, Fiers W, Saelens X: Universal M2 ectodomain-based influenza A vaccines: preclinical and clinical developments. *Expert review of vaccines* 2009, 8(4):499-508.

126.   DiLillo DJ, Tan GS, Palese P, Ravetch JV: Broadly neutralizing hemagglutinin stalk-specific antibodies require FcgammaR interactions for protection against influenza virus in vivo. *Nature medicine* 2014, 20(2):143-151.

127.   Lillie PJ, Berthoud TK, Powell TJ, Lambe T, Mullarkey C, Spencer AJ, Hamill M, Peng Y, Blais ME, Duncan CJ *et al*: Preliminary assessment of the efficacy of a T-cell-based influenza vaccine, MVA-NP+M1, in humans. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012, 55(1):19-25.

128.   Status of Vaccine Research and Development of Universal Influenza Vaccine Prepared for WHO PD-VAC [http://who.int/immunization/research/meetings_workshops/Universal_Influenza_VaccineRD_Sept2014.pdf]

129.   Dong G, Peng C, Luo J, Wang C, Han L, Wu B, Ji G, He H: Adamantane-resistant influenza a viruses in the world (1902-2013): frequency and distribution of M2 gene mutations. *PloS one* 2015, 10(3):e0119115.

130.   Influenza   Antiviral   Medications: Summary   for   Clinicians [http://www.cdc.gov/flu/professionals/antivirals/summary-clinicians.htm]

131.   Jacob A, Sood R, Chanu Kh V, Bhatia S, Khandia R, Pateriya AK, Nagarajan S, Dimri U, Kulkarni DD: Amantadine resistance among highly pathogenic avian influenza viruses (H5N1) isolated from India. *Microbial pathogenesis* 2016, 91:35-40.

132.   Tosh C, Murugkar HV, Nagarajan S, Tripathi S, Katare M, Jain R, Khandia R, Syed Z, Behera P, Patil S *et al*: Emergence of amantadine-resistant avian influenza H5N1 virus in India. *Virus genes* 2011, 42(1):10-15.

133.   Hurt AC, Selleck P, Komadina N, Shaw R, Brown L, Barr IG: Susceptibility of highly pathogenic A(H5N1) avian influenza viruses to the neuraminidase inhibitors and adamantanes. *Antiviral research* 2007, 73(3):228-231.

134.   He G, Qiao J, Dong C, He C, Zhao L, Tian Y: Amantadine-resistance among H5N1 avian influenza viruses isolated in Northern China. *Antiviral research* 2008, 77(1):72-76.

135.   Furuta Y, Takahashi K, Fukuda Y, Kuno M, Kamiyama T, Kozaki K, Nomura N, Egawa H, Minami S, Watanabe Y *et al*: In vitro and in vivo activities of anti-influenza virus compound T-705. *Antimicrobial agents and chemotherapy* 2002, 46(4):977-981.

136.   Sissoko D, Laouenan C, Folkesson E, M'Lebing AB, Beavogui AH, Baize S, Camara AM, Maes P, Shepherd S, Danel C *et al*: Experimental Treatment with Favipiravir for Ebola Virus Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea. *PLoS medicine* 2016, 13(3):e1001967.

22

137.   Furuta Y, Gowen BB, Takahashi K, Shiraki K, Smee DF, Barnard DL: Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral research* 2013, 100(2):446-454.

138.   Koday MT, Nelson J, Chevalier A, Koday M, Kalinoski H, Stewart L, Carter L, Nieusma T, Lee PS, Ward AB *et al*: A Computationally Designed Hemagglutinin Stem-Binding Protein Provides In Vivo Protection from Influenza Independent of a Host Immune Response. *PLoS pathogens* 2016, 12(2):e1005409.

139.   Lee SM, Yen HL: Targeting the host or the virus: current and novel concepts for antiviral approaches against influenza virus infection. *Antiviral research* 2012, 96(3):391-404.

24

Chapter 2

Structure and function of influenza M2 and its development as a broadly protective vaccine

2.1. Structure and biological function of the influenza A M2 protein

M2 is an influenza A virus membrane protein of 97 aa encoded by genome segment seven, the M segment. Four different mRNAs are transcribed from the M segment: unspliced M1 mRNA and alternative splice variants M2 mRNA, mRNA 3 and M42 mRNA [1-4]. The three splice variants use a common 3'-splice acceptor site, but a different 5'-splice donor site in the M1 mRNA [1]. The splicing of M2 mRNA is regulated by the viral NS1 protein and polymerase complex, and the cellular SF2/ASF protein [5-8]. Recent data demonstrate that splicing of M1 mRNA occurs at nuclear speckles [9]. M1 and M2 use the same start codon for their translation and thus share the first nine amino acids before the 5' splice site of the mRNA. The C-terminal 89 aa of M2 are translated in the +1 reading frame relative to M1. M2 is the least abundant viral membrane protein, with only 14 to 68 molecules incorporated per virion [10, 11]. In contrast, it is highly expressed on infected cells [3, 12]. Interestingly, in a small number of influenza viruses, mRNA 4 can encode an M2-related protein, M42, which has an antigenically distinct ectodomain and can functionally replace M2 [2].
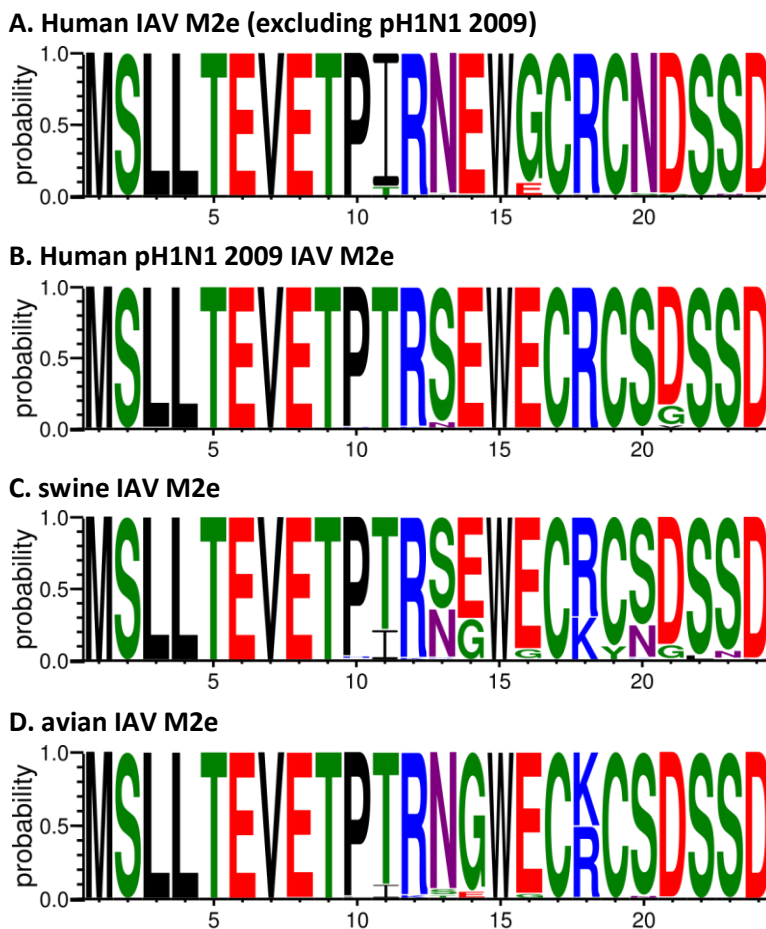
The influenza M2 protein is a homotetrameric type III membrane protein composed of two disulphide-linked dimers [13]. It is a multifunctional protein with three domains: an N-terminal ectodomain, a transmembrane domain and a C-terminal domain consisting of an amphiphatic helix and cytoplasmic tail.

2.1.1. The M2 ectodomain (M2e)

The ectodomain of M2 (M2e) comprises 23 amino acids and has a sequence which is highly conserved in all known human influenza A viruses (IAV) (Figure 1.A). However, the M2e sequence of viruses isolated during the influenza pandemic in 2009 (pH1N1 2009) deviates at four positions from the human consensus (Figure 1.A and B). The M segment in this reassortant virus is of avian-like swine origin, explaining the observed variation in the pH1N1 2009 M2e sequence (Figure 1.C) [14]. Comparably, there are minor variations in the M2e sequences derived from human and avian influenza A viruses (Figure 1). The cysteines at position 17 and 19 are highly conserved and are both equally competent to form the intermolecular disulfide bond in an M2 dimer [13]. Although they are not essential for tetramerization and *in vitro* and *in vivo* viral replication, they stabilize the M2 tetramer when present [13, 15]. Little is known about the functional role of M2e. By using chimeric proteins consisting of domain-swaps between M2 and Sendai virus F protein, it was concluded that M2e is implicated in the selective incorporation of the M2 protein in virions [16]. A role for M2e in regulating the conformation of the transmembrane domain of M2 has recently also been suggested [17].

The high conservation of the M2e sequence is partially due to the genetic overlap with the highly-conserved structural protein M1. The first nine amino acids are identical and show almost no variation in influenza viruses isolated from avian, human or swine, hinting toward a restriction in

genetic flexibility (Figure 1). An additional factor that might limit the genetic diversity in M2e is the low immune pressure present in the population for M2e: only poor anti-M2e antibody responses are elicited after influenza A virus infection or vaccination with conventional vaccines [21-23]. This can be the result of its small size and its low abundance in virions, where it is shielded by the large and more abundant surface proteins HA and NA. Interestingly, the seroprevalence of anti-M2 antibodies increases with age in humans, which possibly reflects the cumulative effect of multiple influenza A exposures [24]. This parallels findings in several inbred and outbred mice strains after repetitive influenza infections [25].

A. Human IAV M2e (excluding pH1N1 2009)

B. Human pH1N1 2009 IAV M2e

C. swine IAV M2e

D. avian IAV M2e

Figure 1: Conservation of the M2e sequence in human, swine and avian influenza A viruses (IAV). Logo based on alignment of all complete M2 protein sequences of influenza A viruses isolated from human (A: excluding all pH1N1 2009 sequences, n = 11230 ; B: only pH1N1 2009 sequences, n = 1924), swine (C, n = 7046) or avian (D, n = 8218) hosts, extracted from the Influenza Research Database (http://www.fludb.org/) on 9th of July, 2016 for human and avian influenza A viruses and on 10th of August, 2016 for swine influenza A viruses. Amino acid sequence logo created using WebLogo after aligning the M2 sequences with Mega7 [18-20].

The structure of M2e is highly flexible, making it very difficult to obtain a crystal structure of native M2e [17, 28]. Nevertheless, crystal structures of M2e in complex with Fab-fragments derived from M2e-specific mAbs have modeled M2e in two different conformations. M2e complexed with an

antibody (mAb148) binding to the highly conserved N-terminus (aa 2-9) resembles a fishing hook, with residues Ser2-Leu3-Leu4-Thr5-Glu6 forming a β-turn (Figure 2.A) [26]. In contrast, M2e in complex with an antibody (mAb65) with an epitope specificity similar to that of mAb 14C2 (aa 5-15) forms a remarkable compact structure with a β-turn from Thr5 to Glu8 and a 3_{10} helix from Ile11 to Trp15, with Trp15 stabilizing the M2e conformation (Figure 2.B) [27].
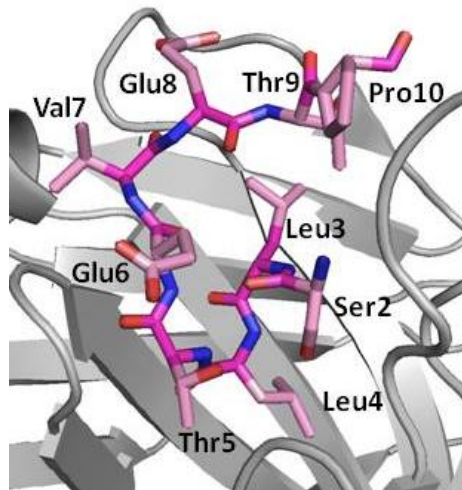
A. Fab148-M2eW15G   **B. Fab65-M2e**
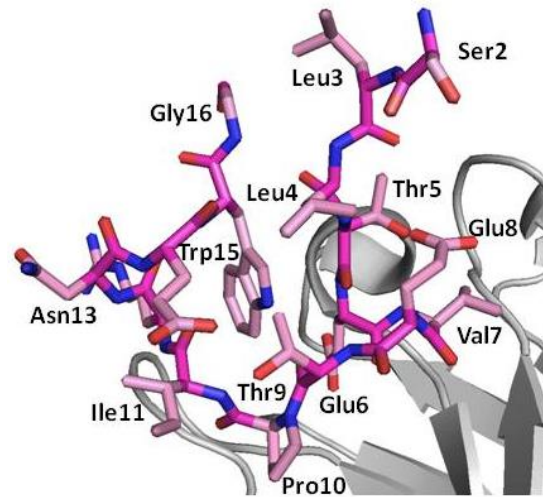
Figure 2: Crystal structure of M2e in complex with Fab-fragments of M2e-specific mAbs. Crystal structure of Fab148 in complex with M2e-W15G (PDB: 5DLM, [26]) (A) and Fab65 in complex with M2e (B) (PDB: 4N8C, [27]). M2e in complex with Fab148 or Fab65 was modeled in PyMol (Delano Scientific, http://www.pymol.org)).

2.1.2. M2 transmembrane domain

The transmembrane domain of M2 (aa 25-46) forms a homotetrameric proton channel, adopting an α-helical secondary structure [29-32]. The four helices assemble into a left-handed four-helix bundle with a twist angle of about 23°, somewhat depending on the membrane composition, and a well-defined pore [31].

M2 forms an ion channel that transfers protons selectively across membranes along the electrochemical proton gradient [33]. Its activity is regulated by protonation of His37 at low pH, which also participates in selectively shuttling protons across the channel [34-36]. Trp41 acts as a gate of the proton channel, helping to define the rate of proton flux [37]. Interestingly, M2 is an asymmetric conductor and conducts only protons from the N-terminal part to the C-terminal part when the pH at the N-terminus is low [32]. To cope with the buildup of a large electrical potential as a consequence of proton transport, M2 has antiporter-like activity and mediates a significant outward flux of K^+ ions [38].

Evidence for the ion channel function of the M2 transmembrane domain originates from studies investigating the method of action of the M2-targeting influenza antivirals amantadine and rimantadine, derivates of adamantane [39-42]. M2 ion channel activity has multiple functions during the viral life cycle. Upon virus entry in the host cell, the lower pH in the endosome activates the M2 channel, resulting in the influx of protons into the virion and disruption of the interactions between vRNPs, M1 and the viral membrane [43]. In parallel, this lower endosomal pH activates the fusion activity of HA, leading to fusion of the viral and endosomal membranes and subsequent release of the vRNPs into cytoplasm. Conversely, during protein trafficking of newly synthesized HA, M2 neutralizes the intravesicular pH in the trans-Golgi network, ensuring the correct maturation of HA by preventing its low-pH induced fusogenic form [44]. Moreover, the ion channel activity of M2 also plays a role in regulation of cell death by arresting autophagy through the prevention of fusion of autophagosomes with lysosomes [45, 46]. A role for the M2 transmembrane domain in viral budding by augmenting the membrane curvature has recently also been suggested but is still under debate [47-49]. In addition, the imbalance in ionic concentrations in the Golgi apparatus caused by the M2 channel activity of newly produced M2 is also required and sufficient for the stimulation of NLRP3 inflammasomes, which is important in the development of an adaptive immune response [50-52].

2.1.3. Cytoplasmic domain

The cytoplasmic domain (aa 47-97) of M2 is highly conserved among influenza A viruses and contains a membrane-surface associated amphipathic helix (aa 46-62) and an unstructured C-terminal tail [28, 53]. Although the cytoplasmic domain of M2 is not required for ion channel activity of the protein, it facilitates proton conduction through the transmembrane pore [54, 55]. An important role has been assigned to the cytoplasmic domain in the formation of filamentous viral particles [56, 57]. The amphiphatic helix is required for virus budding, with a suggested stabilizing role for cholesterol, and scission of the viral particles [48, 49, 58-61]. The cytoplasmic tail of M2 plays an essential role in its interactions with M1, promoting the efficient packaging of genome segments into influenza virus particles [62]. Furthermore, by mimicking a host short linear protein-protein interaction motif, the cytoplasmic tail of M2 also interacts directly with autophagy protein LC3, promoting its relocalization to the plasma membrane, which results in increased virion stability and budding of filamentous viruses [63].

2.2. Unraveling the antigenic potential of M2

The interest in M2 as antiviral target was raised after Zebedee and Lamb demonstrated an effect on viral growth by anti-M2e mAbs [11]. They observed inhibition of viral growth, but not of viral infectivity, for some of the viruses, when applying a monoclonal antibody recognizing the ectodomain of M2, mAb 14C2, in a plaque-reduction assay [11]. By generating viruses that can escape to the *in vitro* growth inhibitory effect of mAb 14C2, the growth restriction could be assigned

to single amino acid positions in the cytoplasmic domain of M2 (positions 71 and 78) or at the N terminus of the M1 protein (positions 31 and 41) [64]. Hughey *et al.* could further demonstrate that inhibition of virus replication by mAb 14C2 is coupled to reduced cell surface expression and redistribution of the M2 protein [65]. In addition to the *in vitro* growth inhibitory effect, mAb 14C2 also inhibited viral replication in influenza A infected mice [66]. The lung viral titer in 14C2 treated mice was 100 fold lower on day 3 and 4 post infection compared to control treated mice. In addition, a larger effect on virus replication was observed in the lungs than in the upper respiratory tract [66].

The immunogenic potential of M2 was further investigated by Black and his colleagues. They determined the antibody titers against M2 in serum of recently infected humans and ferrets, applying an in-house developed enzyme immunoassay and Western blot assay based on M2 expressing insect cells [22]. Low antibody levels against M2 were detected in the convalescent sera from individuals recently infected with influenza A (H3N2) virus and in serum samples from ferrets experimentally infected with influenza A H1N1, H2N2 and H3N2 viruses [22].

These observations, together with the conserved structure of M2, made researchers speculate on the potential of M2 as a broad-spectrum vaccine against influenza. Mice vaccinated with a partially purified M2 protein, derived from recombinant baculovirus expressing M2 in insect cells, were protected against lethal challenges with both homologous and heterologous influenza viruses [67]. These results demonstrated for the first time that vaccination with a conserved influenza protein can elicit heterosubtypic immunity to influenza A viruses. In addition, no protection against influenza B virus was obtained, suggesting that protection was mediated by a mechanism specific to M2 [67].

Inspired by the protective effect of anti-M2e mAb 14C2 in mice, Neirynck *et al*. were the first to focus on the conserved ectodomain of M2 (M2e) as candidate for a universal influenza vaccine [68]. Since M2e is poorly immunogenic in nature, the immunogenicity of the peptide was enhanced by linking it N-terminally to the hepatitis B virus core (HBc) protein [68]. The HBc protein self-assembles to form so-called virus-like particles (VLPs), displaying M2e on the outside in a manner similar to influenza virus particles or infected cells. The ordered presentation of M2e on VLPs promotes a strong immune response [69]. Mice vaccinated with M2e-HBcore demonstrated long-lasting, heterosubtypic immunity against influenza A viruses which was mediated by antibodies since protection was transferable by serum [68].

2.3. M2e based vaccines

The initial reports on the M2- and M2e-based influenza A vaccines, were quickly followed by multiple studies from different research groups confirming its potential as antigen for the development of broadly protective influenza vaccines, with homo- and heterosubtypic protection. The immunogenicity of M2e is hereby enhanced using several strategies, most of them by coupling M2e to a carrier or by the use of adjuvant formulations. Also a significant portion of research focuses on

evaluation of different routes for immunization, large-scale production systems and activation of the desired branch of the immune system.

The M2e peptide is poorly immunogenic

The short nature of the M2e peptide, together with the absence or low abundance of M2e-specific antibodies after influenza infection or seasonal vaccination, generally leads to the assumption that the M2e peptide itself is a poor immunogen [21-23]. However, a small number of studies could show induction of M2e-specific T cell responses and antibodies after vaccination with free synthetic M2e peptide in incomplete Freund's and aluminum adjuvant, but not in the absence of adjuvant [70, 71]. The immunogenicity of the M2e peptide can be attributed to the presence of a BALB/c-specific CD4[+] T-cell epitope in its sequence (discussed in paragraph 2.5.2) [72].

Increasing immunogenicity by coupling M2e to a vaccine carrier

As mentioned earlier, our group was the first to increase the immunogenicity of M2e by fusing its sequence to HBc, which assemble into VLPs. An important advantage of M2e-HBc is their ease to produce in *Escherichia coli* in large scale at low cost. M2e-HBc VLPs with M2e fused N-terminally or in the immunodominant loop of HBc are equally protective [73]. The protective capacity of monomeric M2e-HBc VLPs has been enhanced by synthesizing M2e with a terminal cysteine and chemically coupling it to a lysine in the immunodominant region of HBc, hereby avoiding the steric hindrance during capsid assembly [74]. Multiple different other successful strategies have been implemented to produce protective M2e expressing VLPs, differing in fusion partner, M2e expression levels, production and induced immune response and are presented in Table 1. Another strategy to increase the immunogenicity of M2e, is by fusing it either genetically or chemically to a highly immunogenic fusion protein, that exerts an adjuvant effect. Several successful examples of M2e-fusion proteins have been reported and are presented in Table 1. A similar strategy was used to design M2e lipopeptides, where the lipid Pam2Cys moiety was coupled to M2e [75].

DNA and adenoviral vector vaccines

DNA vaccines and recombinant adenoviral vectors (rAd) are also a promising design for M2e-based vaccines since they can directly induce both humoral and cellular immune responses, by endogenous expression of the antigen. This is in contrast with natural infection, where M2-specific cellular responses are poorly raised due to immunodominance of NP and M1 in eliciting cellular immune responses [113, 114]. DNA vaccines or adenoviral vectors often contain the full-length protein to include a high number of T cell epitopes. However, the presence of the transmembrane part of M2 should be avoided, since its proton channel activity induces toxicity in mammalian cells [115]. Some studies on the protective effect of M2 DNA or rAd vaccination are available, mostly in combination with NP, HA or M1; or as M2 DNA vaccination, followed by M2 rAd boost vaccination [101, 110-112, 116, 117]. A synergistic effect has been described using a DNA vaccine expressing a fusion protein of H1N1 HA and M2e, resulting in complete protection against an avian H5N2 influenza virus [112].

Adenoviruses have the advantage over DNA vaccination that no needles are required for administration since they can be administered using a nasal spray.

Table 1: Overview on successful M2e-based vaccination strategies

Vaccine type	Carrier	Origin M2e	Copies M2e	Animal model	Route	Ref
VLP		Human	1, 3	Mice	i.p. or i.n.	[68, 76, 77]
	HBc	Human, avian	4	Mice	i.m.	[78]
		Avian	1, 2 or 4	Mice	i.m.	[79]
	M1	Human, avian, swine	5	Mice, ferrets	i.m.	[80-82]
	tGCN4, M1	Human, swine, avian	5	Mice	m.n.	[83]
	tGCN4, tFliC, M1	Human	Tetramer	Mice	i.n.	[84]
	AuNPs	Human	1	Mice	i.n.	[85]
	OMV	Human, avian, swine	4	Mice	s.c.	[86]
	PMV	Human	1	Mice	s.c.	[87]
	MaMV	Dog	3	Mice	s.c.	[88]
	WHP	Avian-like	1	Mice	o	[89]
	CotB	Human	3	Mice	o	[90]
	T7	Human	1	Mice	s.c.	[91]
	f88	Human (aa 2-16)	1	Mice	i.p.	[92]
	Qβ	Human	1	Mice	i.n., s.c.	[93]
Protein	mHSP70c	Human	4	Mice	i.m.	[94, 95]
	HSA	Human	1	Mice	i.p.	[96]
	flagelin	Human, avian	4	Mice	i.n. , m.n.	[97-100]
	KLH	Human	1	Mice	i.p.	[101]
	BLS	Human	1,4	Mice	i.n., s.c., i.m.	[102]
	CTA1-DD	Human	1, 3	Mice	i.n.	[72]
	OMPC	Human	1	Ferrets, Rhesus Monkey	i.m.	[103]
	ASP-1	Avian	1,3	Mice	i.m.	[104]
	GST	Human	1	mice	i.p.	[105]
	HA2	Human	1	Mice	i.n.	[106]
	LTB	Avian	3	Mice	i.n.	[107]
	tGCN4	Human	Tetramer	Mice	i.p., i.n.	[108]
Peptide	Pam2Cys	Human	1	Mice	i.n., s.c.	[75]
	MAP	Human	1,2,4	Mice	i.n.	[109]
DNA	NP	Human	1	Mice	i.d., e.p.	[110]
	NP, HA	Avian	1	Mice, ferrets	i.m.	[111]
	HA	Human, avian	2	Mice	i.m.	[112]

i.p.= intraperitoneal, i.n. = intranasal, i.m.= intramuscular, m.n. = microneedle, s.c.= subcutaneous, o. = oral and e.p.= electroporation

Tetrameric M2e as immunogen

The M2 protein forms a tetramer which can contain structural epitopes. Consequently, some research groups focus on mimicking the native structure of M2 by producing tetrameric M2e as vaccine target. In addition, the protective potential of M2e-based vaccines has also been linked to the amount of tetramer specific M2e-antibodies in serum [118]. Oligomer-specific antibodies could be obtained after immunization with M2e coupled to a modified form of the leucine zipper of the yeast transcription factor GCN4 [108]. In another strategy, four M2e copies were coupled to a peptidyl core matrix resulting in tetra-branched M2e-MAPs (multiple antigenic peptide), from which the M2e peptides radially extend [119, 120]. To increase their immunogenicity, the four copies of M2e were fused to foreign T helper cell epitopes or tuftsin [109, 121, 122]. Tetrameric M2e has also been expressed in a membrane-anchored form onto VLPs, resulting in heterosubtypic protection when a truncated form of flagellin was co-expressed [84].

Enhancing immunogenicity by epitope density and vaccine broadness by inclusion of M2e variants

Several studies demonstrate enhanced immunogenicity and protective capacity of M2e-based vaccines by increasing the epitope density, possibly by facilitating cross-linking of the B-cell receptor on M2e-specific B cells [73, 79, 123-125]. Using M2e-GST fusion proteins carrying different amounts of M2e epitopes, it was shown that the serum M2e-specific antibodies increase with increasing M2e number [126]. In addition, higher M2e epitope densities resulted in higher survival rates and slower weight losses in mice [126]. Next to increasing the epitope density, the broadness of the vaccine can also be enhanced by including M2e sequences from human, avian and/or swine origin. Incorporation of human, swine and avian M2e sequences in M2e(5x)-VLPs resulted in enhanced cross-reactivity against human, swine and avian M2e peptides and enhanced protection to heterologous viruses in mice, when compared to wild type VLPs [82, 127]. In addition, an equal antibody response was raised against human and avian M2e when mice were vaccinated with a flagelin fusion protein to which two human and two avian M2e sequences were coupled, resulting in similar significant protection against lethal challenges with human or avian influenza viruses [128]. However, in a side-by-side analysis of particles carrying four human M2e, four avian M2e or two human and two avian M2e epitopes, a weaker protective effect was observed for VLPs carrying both human and avian M2e epitopes and enhanced cross-reactivity for VLPs carrying only human or avian M2e sequences [78].

Enhancing the immune response by the use of an adjuvant

The protective effect induced by M2e based vaccines is primarily by a non-neutralizing cell-mediated antibody response with an essential role for Fcγ Receptors (described in more detail in paragraph 2.5.1) [129]. Antibodies of the IgG2a isotype are the most potent in binding activatory Fcγ Receptors with high affinity in mice [130]. In addition, IgG2a antibodies have been associated with increased efficacy of influenza vaccination [131-134]. This holds also true for M2e based vaccines, where protection correlates closely with the presence of anti-M2e antibodies, in particular of the IgG2a subclass [68, 135, 136]. It is thus important to direct the evoked immune response by vaccination to the Th1 cellular response by the use of an adjuvant [133, 137]. Adjuvants in M2e-based vaccines are

either fused to M2e as carrier or separately added in the vaccine preparation. A potent, protective Th1 response can for example be obtained by including a pathogen-associated molecular pattern in the M2e vaccine, *e.g.* RNA (TLR7 ligand) or RC-529 (TLR4 ligand) [73, 136]. Incorporation of RNA in M2e-VLPs resulted in an increased Th1 response after vaccination and better protection against infection when compared to M2e-VLPs without associated RNA [138]. Another potent and safe mucosal adjuvant with proven protective effect for M2e is CTA1-DD, containing the A1 subunit of cholera toxin with a dimer of an IgG binding element from *Staphylococcus aureus* protein A, which augments both T and B cell responses following intranasal immunizations [72, 139].

Route of immunization

Mucosal surfaces lining the airways are the site of viral entry so high antibody titers are required at these sites. Intramuscular vaccination is the most common route of vaccine administration, as is the case for M2e-based vaccines, and results in a systemic and protective immune response. Oral or intranasal vaccine administration induce localized immunity which results in higher antibody titers at the site of infection (both IgG and IgA), decreased viral spread in the lungs and increased protection when compared to systemic immunizations, making them the preferred routes of vaccine administration [68, 118, 140, 141]. In addition, since needles are omitted, oral and intranasal vaccines can be administered in a safer manner then most of the seasonal influenza vaccines. Intranasal vaccination also results in a higher portion of IgG2a isotypes antibodies, compared to parenteral vaccination, independent of the adjuvant used [118]. The success of intranasal vaccine administration for use in the clinic is somewhat limited due to safety issues regarding retrograde transport of applied antigens to the central nervous system. An alternative is sublingual M2e vaccine administration, resulting in significant better protection against homologous and heterologous influenza viruses infections, when compared to systemic vaccination [140]. Coated microneedle vaccination is another painless vaccination method which induces strong humoral an mucosal antibody responses and conferred complete protection against homo- and heterosubtypic lethal virus challenges, significantly better than conventional intramuscular injection [100].

Co-immunization of M2e with the licensed influenza vaccines

Co-immunization of M2e based vaccines with seasonal influenza vaccines is also being investigated to overcome the strain-specific protection elicited by current influenza vaccines. Supplementing the inactivated A/PR8 influenza vaccine with M2-VLPs resulted in cross-protection against lethal challenges with heterologous and heterosubtypic influenza viruses by preventing disease symptoms [142]. In addition, co-immunization of the seasonal H3N2 split vaccine with a heterologous M2e(5x)-GCN4 vaccine resulted in higher M2e-specific IgG2a antibody titers, compared to both vaccines separately, and synergistic improved protection against heterologous H3N2 viruses [143]. Moreover, co-immunization resulted in improved cross-protection to different H1N1 and H5N1 viruses in mice [143]. Supplementation of the H1N1pdm09 split vaccine with heterologous M2e(5x)-VLPs also resulted in significantly reduced disease symptoms while conferring improved cross-protection in ferrets [80].

Viral escape from M2e-based immunity

An important concern regarding vaccines based on the conserved M2e epitope, is their potential to select for viral resistance. There are only two studies that describe escape mutations in M2 under M2e selection pressure, both using M2e-specific mAbs. *In vitro* selection of viruses that were no longer growth inhibited by M2e-specific mAb 14C2, resulted in the selection of resistant viruses carrying single amino acid mutations in the M2 cytoplasmic domain or the N terminus of M1 [64]. In a second study, chronic treatment of infected SCID mice with different M2e-specific mAbs, with the same epitope specificity as 14C2, resulted in a mutated M2e sequence in 65% of all treated mice [144]. Mutations were only detected at amino acid position 10 in M2e, with the variation limited to a proline to leucine or histidine and both mutations were silent in M1 [144]. However, no escape mutants emerged in SCID mice treated with a combination of mAbs specific for M2e of PR8 and the P10H and P10L escape mutants [145]. In addition, Wolf *et al*. showed that mice vaccinated with M2e-MAP, which carry four wild type M2e side chains, were fully protected against challenge with PR8-M2 P10L or PR8-M2 P10H [25]. Moreover, eleven consecutive passages of PR8 virus in M2e-vaccinated BALB/c mice did not result in selection of a single M2e escape virus [145]. In addition, no mutations inside the 'SLLTE' epitope or the M2 protein were detected using 454 pyrosequencing on the last positive nasal swab sample of experimentally infected patients treated with M2e-specific mAb TCN-032 (discussed in section 2.4) [146]. Consequently, the genetic constraints in M2e suggest that the chance for developing resistance to M2e based vaccination is rather low.

2.4. Clinical development of M2e based vaccines

Giving the long-standing record of protection by M2e-based immunogens in pre-clinical models, a number of companies have moved forward with a variety of designs to clinical testing.

Sanofi Pasteur, formerly known as Acambis, enrolled the M2e-HBc VLPs (ACAM-FLU-A[TM]) in a Phase I clinical trial. ACAM-FLU-A[TM] was developed in our research group and carries three human consensus M2e sequences at the N-terminus of HBc. Its immunogenicity was proven after intramuscular administration, alone or in combination with aluminium hydroxide or QS-21 adjuvants. The highest immune responses were recorded in 90% of participants after two vaccine doses of ACAM-FLU-A[TM] (50 µg) in combination with QS-21 adjuvant [147].

VaxInnate focuses on STF2.4xM2e (VAX102), a fusion protein where four copies of M2e are linked to flagellin of *Salmonella typhimurium*, a TLR5 ligand, as adjuvant. Two subsequent phase I studies were performed to determine the optimal vaccine dose for intramuscular administration. From these studies, it could be concluded that VAX102 was safe and high M2e-specific antibody titers were induced in 96% of participants when using two subsequent vaccine doses of 0.3 and 1 µg [148]. A follow-up study, applying co-immunization of VAX102 with the seasonal influenza vaccine, resulted in increased immune responses, including M2e immunity, which may lead to cross-protection [149]. VAX102 is currently undergoing Phase II testing [150].

In another TLR-triggering based vaccine design, Dynavax employs a fusion protein of M2e and NP covalently linked to a TLR9 agonist (N8295). Clinical phase 1a/b data show antibody responses to M2e and NP and cellular immune responses to NP for all evaluated doses [151, 152]. Addition of N8295 to an investigational H5N1 avian influenza vaccine, resulted in increased HA-specific responses [151].

Next to active immunization methods, passive immunization using human monoclonal antibodies is also being explored. TCN-032 is a monoclonal antibody isolated directly from M2e-seropositive individuals that binds the very conserved N-terminal 'SLLTE' epitope of M2e on infected cells and virions, but not to synthetic M2e peptide [153]. TCN-032 has successfully passed a clinical phase II study where the protective potential of the mAb was tested in a controlled, experimental influenza infection in healthy human participants [146]. The phase II study confirmed that TCN-032 is safe and well tolerated, and has a half-life of approximately 16 days [146]. Treatment with TCN-032 (40 mg/kg) twenty-four hours post infection resulted in reductions in clinical symptoms and viral shedding, with no detectable emergence of resistant virus [146]. The clinical phase II results thus prove the therapeutic potential of TCN-032 after influenza infection.

SEEK investigates a synthetic influenza vaccine (Flu-v) comprising equimolar amounts of conserved T cell epitopes of M1, M2, NP and PB1 [154]. A phase I clinical trial demonstrated the safety of Flu-v and the induction of vaccine-specific cellular immunity[154]. A follow-up phase Ib study involving live virus challenge in humans, resulted in increased cellular immunity and reduction in both symptom scores and virus loads after Flu-v vaccination [155]. Flu-v is now being evaluated in a Phase II study [150].

2.5. Mechanism of M2e-based protection

What is clear from preclinical studies using M2e-based active and passive vaccinations, is that M2e-based protection is not-neutralizing, *i.e.* it cannot prevent initial infection of host cells. However, despite extensive reports on M2e-based vaccines, the exact mechanism of the elicited immune response remains poorly understood. A better understanding of the protective mechanism of M2e based vaccines will result in the development of improved vaccination strategies. In addition, the efficacy of conventional influenza vaccines is currently evaluated based on the induction of HA inhibition antibody titers. Since M2e-specific antibodies are not virus neutralizing, understanding their mechanism of action will result in the development of a read-out system to test and define the efficacy of the M2e-based vaccines. A schematic representation of the proposed mode of action of M2e based vaccines is represented in Figure 3.
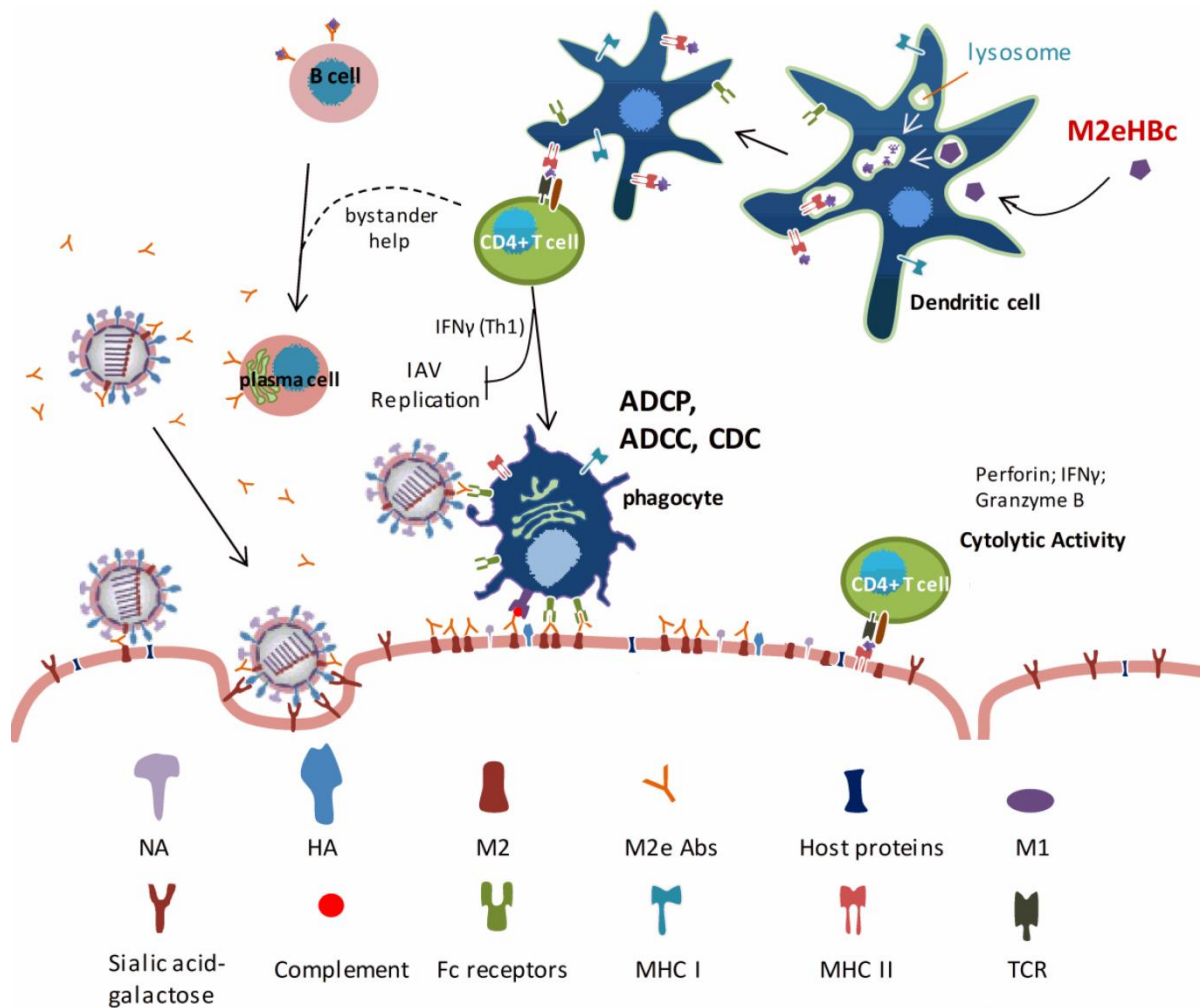
Figure 3: Schematic representation of mechanism of action of M2e based vaccines. The M2e based vaccine, in this example M2e-HBc, is processed by an antigen presenting cell and presented on MHCII to CD4$^+$ T cells. Upon activating, cytokines and chemokines are released as bystander help to generate M2e-specific mAbs by plasma cells. These mAbs can bind to M2e present on virus or highly expressed on infected cells. Phagocytes will bind to cells opsonised with M2e-specific antibodies and clear the infected cell through antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP) or complement-dependent cytotoxicity (CDC). NK cells also contribute to M2e-based protection through ADCC. A cytolytic activity for CD4$^+$ T-cells has also been suggested. Figure adapted from [158].

2.5.1. Antibody-mediated M2e-based protection

As mentioned, vaccination with M2e results in non-neutralizing protection. This protection is antibody-mediated, since the protective effect of vaccination can be transferred by serum and protection correlates with M2e-specific antibody titers [68, 71, 109, 156]. In addition, passive transfer of M2e-specific mAbs can provide protection against lethal viral challenges [157]. However, the M2e-specific serum antibodies do not, or only poorly, bind to virions and hence are not virus neutralizing [68, 156]. Nevertheless, they can easily bind to M2 that is expressed on infected cells and exert their protective effect through the effector functions of their Fc-tail [12]. These effector

38

functions can be subdivided into two groups: binding to cellular receptors that recognize the constant fragment (Fc tail) of IgG (FcγRs), resulting in antibody dependent cell-mediated cytotoxicity (ADCC) and antibody dependent cell-mediated phagocytosis (ADCP), or binding to complement, resulting in complement-dependent cytotoxicity (CDC) (Figure 3). Elimination of the virus-infected cell opsonized with M2e-specific antibodies through ADCC, ADCP or CDC will suppress virus production and release, and thus progression of the infection.

Role of complement

In general, complement-dependent cytotoxicity has a protective role in the early response to acute influenza infection [159]. To what extent complement assists in M2e-mediated protection however is somewhat controversial. A role for both CDC and ADCC in protection by a human anti-M2 mAb in mice was demonstrated by Wang *et al.*, who showed that the anti-M2e mAbs failed to reduce lung viral titers after a sublethal infection in the absence of complement or FcγRIII [157]. The CDC and ADCC potential of this antibody was also confirmed *in vitro* [157]. Heat inactivated immune serum from mice vaccinated with both inactivated PR8 and M2-VLPs could protect against a lethal infection [142]. However, these mice exhibited more body weight loss compared to mice receiving immune serum that was not heat inactivated, hinting towards a role of heat-sensitive serum components such as complement in providing cross-protection [142]. In contrast, a role for complement was excluded by Jegerlehner *et al.* based on the observation that C3-deficient mice were as protected as wild-type mice to a lethal influenza infection by passively transferred mouse anti-M2e hyperimmune serum [156]. In addition, El Bakkouri *et al.* also demonstrated that complement is not sufficient for protection, since mice that lack FcγRs were not protected after passive transfer of M2e immune serum [129]. Further studies are thus required to investigate the role of complement in M2e based protection.
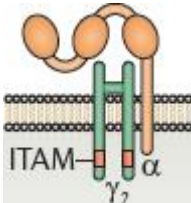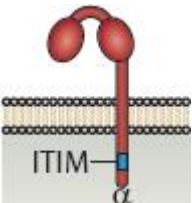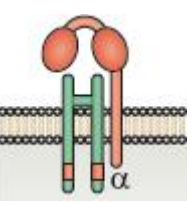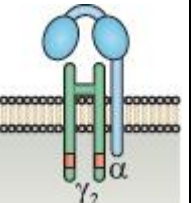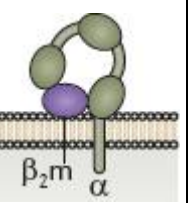
Role of Fcγ receptors

The availability of FcγR knock-out mice makes it possible to study their role in M2e-based protection. Mice have three activating Fcγ receptors (FcγRI, FcγRIII and FcγRIV), one inhibitory (FcγRIIb) and the neonatal Fc receptor (FcRn). The affinity of the different mouse Fcγ receptors for the different murine IgG subtypes is listed in Table 2; the expression pattern of the different Fcγ receptors on immune cells is shown in Figure 4.

Independent studies have shown that FcγRs are essential in M2e-mediated protection using *fcer1g* knock-out mice, lacking the Fc common γ chain, which is required for expression and function of all activating FcγRs [127, 129, 160]. M2e-based vaccination in *fcer1g*[-/-] mice elicits similar levels of IgG1 and IgG2a M2e-specific antibodies as wild type BALB/c mice [160]. Nevertheless, the elicited immune response failed to effectively clear the virus and protect *fcer1g*[-/-] mice. However, passive transfer of M2e serum elicited in *fcer1g*[-/-] mice resulted in similar protection of wild type BALB/c mice after
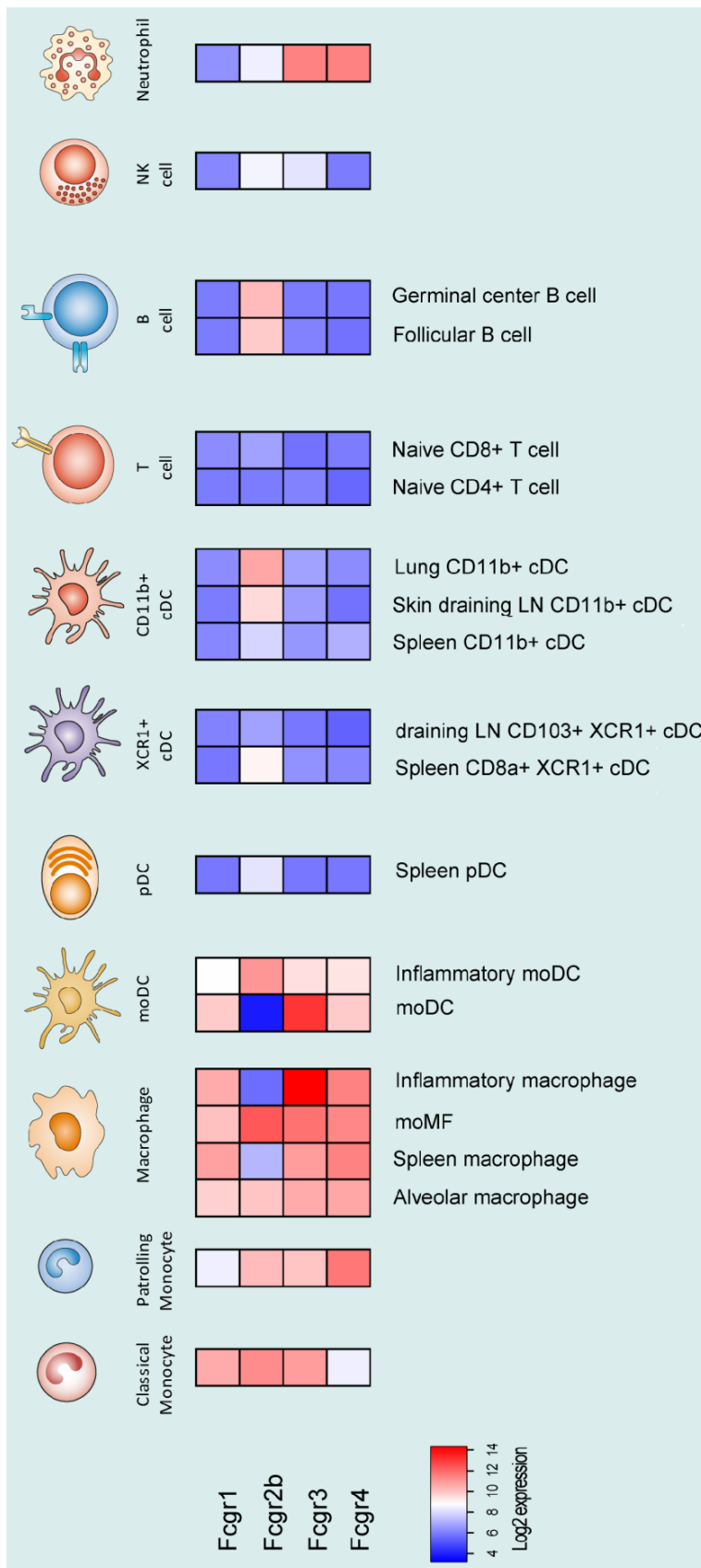
challenge with a lethal viral dose, compared to serum of WT vaccinated mice [160]. *Fcer1g*[-/-] mice are thus able to induce M2e-specific antibodies with similar protective potential as wild type mice, but FcγRs play a critical role in viral clearance and protection after M2e vaccination [160].

Different IgG isotypes have a distinct protective potential in cell-mediated M2e-based protection against influenza virus infection. This is a direct consequence of their selective binding to the activating FcγRs. Since IgG2a can bind to all three types of activating FcγRs, this antibody subtype possesses the strongest potential to stimulate effector functions (Table 2). This has been shown by passive transfer of the IgG2a switch variant of M2e-specifc mAb 14C2 (IgG1), which resulted in significant less morbidity and mortality when compared to its IgG1 and IgG2b counterparts [118]. In addition, IgG1 M2e-specific mAbs failed to protect *fcgr3*[-/-] mice, while IgG2a isotype Abs could restore the protective capacity [129]. Enhanced levels of IgG2a antibodies also correlate with better protection against influenza viruses, which holds true both for M2e-based and seasonal influenza vaccination strategies [132, 136].

Table 2: Mouse Fcγ receptors. Schematic representation of the different mouse Fcγ receptors together with their function and affinity for the different IgG isotype subclasses. NB: not binding, +: binds receptor but affinity unknown. Table adapted from [161] and [130].

	FcγRI	FcγRIIB	FcγRIII	FcγRIV	FcRn
	Function				
	Activation	Inhibition	Activation	Activation	IgG recycling Transport IgG Ag Uptake
	IgG subclass binding (M^{-1})				
IgG1	NB	3×10^6	3×10^5	NB	8×10^6
IgG2a	1×10^8	4×10^5	7×10^5	3×10^7	+
IgG2b	1×10^5	2×10^6	6×10^6	2×10^7	+
IgG3	+	NB	NB	NB	+

Figure 4: Mouse FcγR expression profiles based on publically available microarray data. Expression values are colour-coded, varying from blue (showing low expression), to white (showing medium expression) to red (showing high expression). Figure kindly provided by Martin Guilliams and adapted from [161].

Murine alveolar macrophages express high levels of all three activating FcγRs and form an important part of the innate immune system of the lung (Figure 4). Phagocytosis exerted by macrophages plays an important role in viral clearance during influenza infection, since they ingest opsonized viral particles using FcγR-mediated phagocytosis and remove influenza virus-infected cells by apoptosis-driven phagocytosis [132, 162]. Likewise, an essential role for alveolar macrophages in M2e-based protection by digesting M2e-antibody coated infected cells through ADCP has been suggested. Intratracheal delivery of clodronate-loaded liposomes resulted in elimination of all alveolar macrophages, while leaving the NK cells intact, and failure of anti-M2e immune serum to protect these mice [129, 163]. The role of alveolar macrophages was confirmed by transferring wild type alveolar macrophages into FcγRI and FcγRIII knock-out mice, depleted of alveolar macrophages, restoring the protection exerted by M2e-immune serum [129].

Murine natural killer (NK) cells are also part of the first line of defense against influenza infections by infiltrating the lungs to kill influenza infected cells through ADCC [164]. NK cells solely express the activating FcγRIII (Figure 4) and a role for NK-dependent ADCC in M2e-based protection has been suggested, although the exact part NK cells play in M2e-based protection is unclear. An essential role for NKs was suggested by Jegerlehner *et al.*, since depletion of NK cells resulted in significant diminished protection after lethal infection of M2e-HBc immunized mice [156]. However, these mice showed slightly enhanced protection compared to naive C57BL/6 mice, albeit not significant. In addition, two human M2e-specific IgG1 mAbs, Ab1-10 and C40G1, induce *in vitro* NK cell-mediated ADCC [157, 165]. To boot, M2e-specific human mAb Z3G1 (hIgG1) failed to reduce lung viral titers in FcγRIII knock-out mice, the only activating low-affinity FcγR on NK cells (Figure 4) [157]. In contrast, another study showed that M2e-vaccinated BALB/c mice depleted of NK cells were similarly protected to controls [101]. Similar findings were obtained in C57BL/6 mice treated with IgG1 or IgG2a M2e-specific mAbs and depleted of NK cells [166]. In addition, $fcgr3^{-/-}$ mice are equally protected as wild type mice after passive transfer of polyclonal M2e immune serum or the IgG2a isotype antibody fraction alone [129]. Altogether, these results indicate that NK cells can contribute in M2e-based protection, but that their role is probably not essential.

2.5.2. T-cell mediated M2e-based protection

Although the humoral immune response is essential for M2e-based protection, also the induced T-cell response has an important impact on protection [101, 167]. M2e contains two overlapping human CD8[+] T cell epitopes and one murine CD4[+] T cell epitope [72, 168-170]. The latter was discovered by our group, in collaboration with the group of Nils Lycke, after intranasal vaccination of mice with CTA1-M2e-DD and an *ex vivo* M2e-specific proliferation recall assay on isolated splenocytes [72]. Proliferation of the splenocytes was lost when T cells were depleted but not when B cells or CD8 T cells were depleted, indicating the presence of a CD4[+] T cell epitope in M2e, which was confirmed by a cell proliferation assay using an anti-MHCII mAb [72]. The induction of a murine

M2e-specific CD4[+] T cell response was further illustrated by Stepanova and coworkers after vaccination with Flg-2M2eh2M2ek, a flagellin-fusion protein bearing two human consensus M2e and two avian A/H5N1 M2e sequences [128].

This murine CD4[+] T cell epitope however is restricted to the H-2[d] MHC haplotype of BALB/c mice [25, 167]. Vaccination of mice with a different MHC haplotype with M2-DNA, followed by a boost vaccination with an M2 recombinant adenoviral vector, resulted in protection of BALB/c (H-2[d]), but not of C57BL/6 (H-2[b]) or CBA (H-2[k]) mice [167]. This was confirmed using MHC congenic mice in a BALB/c or C.B10 background, where also a role for the genetic background was demonstrated, since lower anti-M2 antibody responses together with a lower number of anti-M2 T-cells, secreting less IFN-γ, were registered for congenic C.B10 mice of the H-2[d] haplotype compared to BALB/c mice [167]. The absence of protection and induction of a T cell response in mice with MHC haplotype differing from H-2[d] was also confirmed by Wolf *et al*. using different inbred and outbred mice strains [25]. Although the M2e-specific T-cell response varies with different genetic background of the host, the protective response in low responding haplotypes can be increased by the use of carrier, adjuvants or the use of combined vaccination [167, 171]. This is illustrated by the fact that vaccination with M2 or NP alone failed to protect CBA mice from lethal infection, while all mice vaccinated with both M2 and NP survived the infection [167].

2.5.3. Infection permissive immunity

M2e-based vaccines are infection-permissive and, in contrast to the licensed influenza vaccines, do not block infection. This is an important advantage, since also a T-cell mediated immune response against the conserved internal proteins can be elicited during infection, resulting in cross-protective cellular immunity [81, 172]. The importance of cellular immunity in providing heterosubtypic protection and the use of infection-permissive vaccines in immunologically naive individuals, was demonstrated by Schotsaert and co-workers [172, 173]. The induction of cross-reactive T cells in M2e vaccinated mice was not hindered during primary infection with a homologous virus and resulted in increased protection against a secondary infection with a heterosubtypic virus. In contrast, WIV-vaccinated mice were protected during primary homologous infection, but failed to raise a T-cell response and succumbed to the secondary heterosubtypic challenge [172]. The induction and protective effect of heterosubtypic immunity by infection-permissive vaccination upon virus exposure, was later confirmed using combinations of M2e-VLPs, tetrameric NA and trimeric HA [173].

2.6. Concluding remark

The M2e sequence is highly conserved across all influenza A viruses and M2e-based vaccines result in broad-spectrum protection against both homo- and heterosubtypic viral challenges. In addition, alterations in M2e, which could contribute to escape from M2e-based immunity, are restricted due to genetic constraints because the sequence of M2e overlaps with the sequence coding for the highly conserved, structural protein M1. These findings make M2e-based vaccines appropriate as a 'universal influenza vaccine' candidate. Next to broad-spectrum protection, another advantage over licensed influenza vaccines, is the absence of annual revisions of the vaccine. Consequently, M2e-based vaccines could be stockpiled. In addition, several of the successful M2e-based vaccine approaches can be produced in bacteria, omitting the need for eggs or cell culture, which results in faster vaccine production and a lower cost per vaccine dose. However, further studies are required to unravel the mechanism of action of M2e-based vaccines. This will help in the optimization of vaccine design and the development of a read-out system to define the efficacy of M2e-based vaccines. In addition, reports on the evolution of the viral genome sequence under M2e-based immune pressure are limiting. It will thus be important to investigate if M2e immune escape will occur and how. After all, the hope remains that one day this vaccine will be introduced into the human population to mitigate disease caused by influenza A viruses.

References

1. Shih SR, Suen PC, Chen YS, Chang SC: A novel spliced transcript of influenza A/WSN/33 virus. *Virus genes* 1998, 17(2):179-183.
2. Wise HM, Hutchinson EC, Jagger BW, Stuart AD, Kang ZH, Robb N, Schwartzman LM, Kash JC, Fodor E, Firth AE *et al*: Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain. *Plos Pathog* 2012, 8(11).
3. Lamb RA, Zebedee SL, Richardson CD: Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* 1985, 40(3):627-633.
4. Lamb RA, Choppin PW: Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. *Virology* 1981, 112(2):729-737.
5. Bier K, York A, Fodor E: Cellular cap-binding proteins associate with influenza virus mRNAs. *The Journal of general virology* 2011, 92(Pt 7):1627-1634.
6. Robb NC, Fodor E: The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. *The Journal of general virology* 2012, 93(Pt 1):113-118.
7. Shih SR, Krug RM: Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *The EMBO journal* 1996, 15(19):5415-5427.
8. Shih SR, Nemeroff ME, Krug RM: The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. *Proceedings of the National Academy of Sciences of the United States of America* 1995, 92(14):6324-6328.
9. Mor A, White A, Zhang K, Thompson M, Esparza M, Munoz-Moreno R, Koide K, Lynch KW, Garcia-Sastre A, Fontoura BM: Influenza Virus mRNA Trafficking Through Host Nuclear Speckles. *Nature microbiology* 2016, 2016.
10. Jackson DC, Tang XL, Murti KG, Webster RG, Tregear GW, Bean WJ: Electron microscopic evidence for the association of M2 protein with the influenza virion. *Archives of virology* 1991, 118(3-4):199-207.
11. Zebedee SL, Lamb RA: Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of virology* 1988, 62(8):2762-2772.
12. Zebedee SL, Richardson CD, Lamb RA: Characterization of the influenza virus M2 integral membrane protein and expression at the infected-cell surface from cloned cDNA. *Journal of virology* 1985, 56(2):502-511.
13. Holsinger LJ, Lamb RA: Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. *Virology* 1991, 183(1):32-43.
14. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S *et al*: Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009, 459(7250):1122-1125.
15. Castrucci MR, Hughes M, Calzoletti L, Donatelli I, Wells K, Takada A, Kawaoka Y: The cysteine residues of the M2 protein are not required for influenza A virus replication. *Virology* 1997, 238(1):128-134.
16. Park EK, Castrucci MR, Portner A, Kawaoka Y: The M2 ectodomain is important for its incorporation into influenza A virions. *Journal of virology* 1998, 72(3):2449-2455.
17. Kwon B, Hong M: The Influenza M2 Ectodomain Regulates the Conformational Equilibria of the Transmembrane Proton Channel: Insights from Solid-State NMR. *Biochemistry* 2016.
18. Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: A sequence logo generator. *Genome Res* 2004, 14(6):1188-1190.
19. Schneider TD, Stephens RM: Sequence Logos - a New Way to Display Consensus Sequences. *Nucleic Acids Res* 1990, 18(20):6097-6100.
20. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016, 33(7):1870-1874.
21. Feng J, Zhang M, Mozdzanowska K, Zharikova D, Hoff H, Wunner W, Couch RB, Gerhard W: Influenza A virus infection engenders a poor antibody response against the ectodomain of matrix protein 2. *Virology journal* 2006, 3:102.
22. Black RA, Rota PA, Gorodkova N, Klenk HD, Kendal AP: Antibody response to the M2 protein of influenza A virus expressed in insect cells. *The Journal of general virology* 1993, 74 (Pt 1):143-146.

23.   Lambrecht B, Steensels M, Van Borm S, Meulemans G, van den Berg T: Development of an M2e-specific enzyme-linked immunosorbent assay for differentiating infected from vaccinated animals. *Avian diseases* 2007, 51(1 Suppl):221-226.

24.   Zhong W, Reed C, Blair PJ, Katz JM, Hancock K: Serum antibody response to matrix protein 2 following natural infection with 2009 pandemic influenza A(H1N1) virus in humans. *The Journal of infectious diseases* 2014, 209(7):986-994.

25.   Wolf AI, Mozdzanowska K, Williams KL, Singer D, Richter M, Hoffmann R, Caton AJ, Otvos L, Erikson J: Vaccination with M2e-based multiple antigenic peptides: characterization of the B cell response and protection efficacy in inbred and outbred mice. *PloS one* 2011, 6(12):e28445.

26.   Cho KJ, Schepens B, Moonens K, Deng L, Fiers W, Remaut H, Saelens X: Crystal Structure of the Conserved Amino Terminus of the Extracellular Domain of Matrix Protein 2 of Influenza A Virus Gripped by an Antibody. *Journal of virology* 2016, 90(1):611-615.

27.   Cho KJ, Schepens B, Seok JH, Kim S, Roose K, Lee JH, Gallardo R, Van Hamme E, Schymkowitz J, Rousseau F *et al*: Structure of the extracellular domain of matrix protein 2 of influenza A virus in complex with a protective monoclonal antibody. *Journal of virology* 2015, 89(7):3700-3711.

28.   Liao SY, Fritzsching KJ, Hong M: Conformational analysis of the full-length M2 protein of the influenza A virus using solid-state NMR. *Protein science : a publication of the Protein Society* 2013, 22(11):1623-1638.

29.   Sugrue RJ, Hay AJ: Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* 1991, 180(2):617-624.

30.   Pinto LH, Holsinger LJ, Lamb RA: Influenza virus M2 protein has ion channel activity. *Cell* 1992, 69(3):517-528.

31.   Schnell JR, Chou JJ: Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 2008, 451(7178):591-595.

32.   Thomaston JL, Alfonso-Prieto M, Woldeyes RA, Fraser JS, Klein ML, Fiorin G, DeGrado WF: High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(46):14260-14265.

33.   Chizhmakov IV, Geraghty FM, Ogden DC, Hayhurst A, Antoniou M, Hay AJ: Selective proton permeability and pH regulation of the influenza virus M2 channel expressed in mouse erythroleukaemia cells. *The Journal of physiology* 1996, 494 (Pt 2):329-336.

34.   Hu J, Fu R, Nishimura K, Zhang L, Zhou HX, Busath DD, Vijayvergiya V, Cross TA: Histidines, heart of the hydrogen ion channel from influenza A virus: toward an understanding of conductance and proton selectivity. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103(18):6865-6870.

35.   Wang C, Lamb RA, Pinto LH: Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. *Biophysical journal* 1995, 69(4):1363-1371.

36.   Venkataraman P, Lamb RA, Pinto LH: Chemical rescue of histidine selectivity filter mutants of the M2 ion channel of influenza A virus. *The Journal of biological chemistry* 2005, 280(22):21463-21472.

37.   Tang Y, Zaitseva F, Lamb RA, Pinto LH: The gate of the influenza virus M2 proton channel is formed by a single tryptophan residue. *The Journal of biological chemistry* 2002, 277(42):39880-39886.

38.   Leiding T, Wang J, Martinsson J, DeGrado WF, Arskold SP: Proton and cation transport activity of the M2 proton channel from influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(35):15409-15414.

39.   Hay AJ, Wolstenholme AJ, Skehel JJ, Smith MH: The molecular basis of the specific anti-influenza action of amantadine. *The EMBO journal* 1985, 4(11):3021-3024.

40.   Hay AJ, Zambon MC, Wolstenholme AJ, Skehel JJ, Smith MH: Molecular basis of resistance of influenza A viruses to amantadine. *The Journal of antimicrobial chemotherapy* 1986, 18 Suppl B:19-29.

41.   Belshe RB, Smith MH, Hall CB, Betts R, Hay AJ: Genetic basis of resistance to rimantadine emerging during treatment of influenza virus infection. *Journal of virology* 1988, 62(5):1508-1512.

42.   Kendal AP, Klenk HD: Amantadine inhibits an early, M2 protein-dependent event in the replication cycle of avian influenza (H7) viruses. *Archives of virology* 1991, 119(3-4):265-273.

43.   Stauffer S, Feng Y, Nebioglu F, Heilig R, Picotti P, Helenius A: Stepwise priming by acidic pH and a high K+ concentration is required for efficient uncoating of influenza A virus cores after penetration. *Journal of virology* 2014, 88(22):13029-13046.

46

44.   Ciampor F, Bayley PM, Nermut MV, Hirst EM, Sugrue RJ, Hay AJ: Evidence that the amantadine-induced, M2-mediated conversion of influenza A virus hemagglutinin to the low pH conformation occurs in an acidic trans Golgi compartment. *Virology* 1992, 188(1):14-24.

45.   Ren Y, Li C, Feng L, Pan W, Li L, Wang Q, Li J, Li N, Han L, Zheng X *et al*: Proton Channel Activity of Influenza A Virus Matrix Protein 2 Contributes to Autophagy Arrest. *Journal of virology* 2016, 90(1):591-598.

46.   Gannage M, Dormann D, Albrecht R, Dengjel J, Torossi T, Ramer PC, Lee M, Strowig T, Arrey F, Conenello G *et al*: Matrix protein 2 of influenza A virus blocks autophagosome fusion with lysosomes. *Cell host & microbe* 2009, 6(4):367-380.

47.   Ho CS, Khadka NK, She F, Cai J, Pan J: Influenza M2 Transmembrane Domain Senses Membrane Heterogeneity and Enhances Membrane Curvature. *Langmuir : the ACS journal of surfaces and colloids* 2016, 32(26):6730-6738.

48.   Wang T, Hong M: Investigation of the curvature induction and membrane localization of the influenza virus M2 protein using static and off-magic-angle spinning solid-state nuclear magnetic resonance of oriented bicelles. *Biochemistry* 2015, 54(13):2214-2226.

49.   Schmidt NW, Mishra A, Wang J, DeGrado WF, Wong GC: Influenza virus A M2 protein generates negative Gaussian membrane curvature necessary for budding and scission. *Journal of the American Chemical Society* 2013, 135(37):13710-13719.

50.   Ichinohe T, Pang IK, Iwasaki A: Influenza virus activates inflammasomes via its intracellular M2 ion channel. *Nature immunology* 2010, 11(5):404-410.

51.   Fernandez MV, Miller E, Krammer F, Gopal R, Greenbaum BD, Bhardwaj N: Ion efflux and influenza infection trigger NLRP3 inflammasome signaling in human dendritic cells. *Journal of leukocyte biology* 2016, 99(5):723-734.

52.   Ichinohe T, Lee HK, Ogura Y, Flavell R, Iwasaki A: Inflammasome recognition of influenza virus is essential for adaptive immune responses. *The Journal of experimental medicine* 2009, 206(1):79-87.

53.   Huang S, Green B, Thompson M, Chen R, Thomaston J, DeGrado WF, Howard KP: C-terminal juxtamembrane region of full-length M2 protein forms a membrane surface associated amphipathic helix. *Protein science : a publication of the Protein Society* 2015, 24(3):426-429.

54.   Tobler K, Kelly ML, Pinto LH, Lamb RA: Effect of cytoplasmic tail truncations on the activity of the M(2) ion channel of influenza A virus. *Journal of virology* 1999, 73(12):9695-9701.

55.   Liao SY, Yang Y, Tietze D, Hong M: The influenza m2 cytoplasmic tail changes the proton-exchange equilibria and the backbone conformation of the transmembrane histidine residue to facilitate proton conduction. *Journal of the American Chemical Society* 2015, 137(18):6067-6077.

56.   McCown MF, Pekosz A: Distinct domains of the influenza a virus M2 protein cytoplasmic tail mediate binding to the M1 protein and facilitate infectious virus production. *Journal of virology* 2006, 80(16):8178-8189.

57.   Rossman JS, Jing X, Leser GP, Balannik V, Pinto LH, Lamb RA: Influenza virus m2 ion channel protein is necessary for filamentous virion formation. *Journal of virology* 2010, 84(10):5078-5088.

58.   Roberts KL, Leser GP, Ma C, Lamb RA: The amphipathic helix of influenza A virus M2 protein is required for filamentous bud formation and scission of filamentous and spherical particles. *Journal of virology* 2013, 87(18):9973-9982.

59.   Rossman JS, Jing X, Leser GP, Lamb RA: Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* 2010, 142(6):902-913.

60.   Kim SS, Upshur MA, Saotome K, Sahu ID, McCarrick RM, Feix JB, Lorigan GA, Howard KP: Cholesterol-Dependent Conformational Exchange of the C-Terminal Domain of the Influenza A M2 Protein. *Biochemistry* 2015, 54(49):7157-7167.

61.   Ekanayake EV, Fu R, Cross TA: Structural Influences: Cholesterol, Drug, and Proton Binding to Full-Length Influenza A M2 Protein. *Biophysical journal* 2016, 110(6):1391-1399.

62.   Chen BJ, Leser GP, Jackson D, Lamb RA: The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *Journal of virology* 2008, 82(20):10059-10070.

63.   Beale R, Wise H, Stuart A, Ravenhill BJ, Digard P, Randow F: A LC3-interacting motif in the influenza A virus M2 protein is required to subvert autophagy and maintain virion stability. *Cell host & microbe* 2014, 15(2):239-247.

64. Zebedee SL, Lamb RA: Growth restriction of influenza A virus by M2 protein antibody is genetically linked to the M1 protein. *Proceedings of the National Academy of Sciences of the United States of America* 1989, 86(3):1061-1065.

65. Hughey PG, Roberts PC, Holsinger LJ, Zebedee SL, Lamb RA, Compans RW: Effects of antibody to the influenza A virus M2 protein on M2 surface expression and virus assembly. *Virology* 1995, 212(2):411-421.

66. Treanor JJ, Tierney EL, Zebedee SL, Lamb RA, Murphy BR: Passively transferred monoclonal antibody to the M2 protein inhibits influenza A virus replication in mice. *Journal of virology* 1990, 64(3):1375-1377.

67. Slepushkin VA, Katz JM, Black RA, Gamble WC, Rota PA, Cox NJ: Protection of mice against influenza A virus challenge by vaccination with baculovirus-expressed M2 protein. *Vaccine* 1995, 13(15):1399-1402.

68. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.

69. Bachmann MF, Rohrer UH, Kundig TM, Burki K, Hengartner H, Zinkernagel RM: The influence of antigen organization on B cell responsiveness. *Science* 1993, 262(5138):1448-1451.

70. Wu F, Huang JH, Yuan XY, Huang WS, Chen YH: Characterization of immunity induced by M2e of influenza virus. *Vaccine* 2007, 25(52):8868-8873.

71. Pejoski D, Zeng W, Rockman S, Brown LE, Jackson DC: A lipopeptide based on the M2 and HA proteins of influenza A viruses induces protective antibody. *Immunology and cell biology* 2010, 88(5):605-611.

72. Eliasson DG, El Bakkouri K, Schon K, Ramne A, Festjens E, Lowenadler B, Fiers W, Saelens X, Lycke N: CTA1-M2e-DD: a novel mucosal adjuvant targeted influenza vaccine. *Vaccine* 2008, 26(9):1243-1252.

73. De Filette M, Min Jou W, Birkett A, Lyons K, Schultz B, Tonkyro A, Resch S, Fiers W: Universal influenza A vaccine: optimization of M2-based constructs. *Virology* 2005, 337(1):149-161.

74. Jegerlehner A, Tissot A, Lechner F, Sebbel P, Erdmann I, Kundig T, Bachi T, Storni T, Jennings G, Pumpens P *et al*: A molecular assembly system that renders antigens of choice highly repetitive for induction of protective B cell responses. *Vaccine* 2002, 20(25-26):3104-3112.

75. Zeng W, Tan AC, Horrocks K, Jackson DC: A lipidated form of the extracellular domain of influenza M2 protein as a self-adjuvanting vaccine candidate. *Vaccine* 2015, 33(30):3526-3532.

76. De Filette M, Martens W, Smet A, Schotsaert M, Birkett A, Londono-Arcila P, Fiers W, Saelens X: Universal influenza A M2e-HBc vaccine protects against disease even in the presence of pre-existing anti-HBc antibodies. *Vaccine* 2008, 26(51):6503-6507.

77. Gao X, Wang W, Li Y, Zhang S, Duan Y, Xing L, Zhao Z, Zhang P, Li Z, Li R *et al*: Enhanced Influenza VLP vaccines comprising matrix-2 ectodomain and nucleoprotein epitopes protects mice from lethal challenge. *Antiviral research* 2013, 98(1):4-11.

78. Tsybalova LM, Stepanova LA, Kuprianov VV, Blokhina EA, Potapchuk MV, Korotkov AV, Gorshkov AN, Kasyanenko MA, Ravin NV, Kiselev OI: Development of a candidate influenza vaccine based on virus-like particles displaying influenza M2e peptide into the immunodominant region of hepatitis B core antigen: Broad protective efficacy of particles carrying four copies of M2e. *Vaccine* 2015, 33(29):3398-3406.

79. Ravin NV, Blokhina EA, Kuprianov VV, Stepanova LA, Shaldjan AA, Kovaleva AA, Tsybalova LM, Skryabin KG: Development of a candidate influenza vaccine based on virus-like particles displaying influenza M2e peptide into the immunodominant loop region of hepatitis B core antigen: Insertion of multiple copies of M2e increases immunogenicity and protective efficiency. *Vaccine* 2015, 33(29):3392-3397.

80. Music N, Reber AJ, Kim MC, York IA, Kang SM: Supplementation of H1N1pdm09 split vaccine with heterologous tandem repeat M2e5x virus-like particles confers improved cross-protection in ferrets. *Vaccine* 2016, 34(4):466-473.

81. Lee YN, Lee YT, Kim MC, Gewirtz AT, Kang SM: A Novel Vaccination Strategy Mediating the Induction of Lung-Resident Memory CD8 T Cells Confers Heterosubtypic Immunity against Future Pandemic Influenza Virus. *J Immunol* 2016, 196(6):2637-2645.

82. Kim MC, Song JM, O E, Kwon YM, Lee YJ, Compans RW, Kang SM: Virus-like particles containing multiple M2 extracellular domains confer improved cross-protection against various subtypes of influenza virus. *Molecular therapy : the journal of the American Society of Gene Therapy* 2013, 21(2):485-492.

83. Kim MC, Lee JW, Choi HJ, Lee YN, Hwang HS, Lee J, Kim C, Lee JS, Montemagno C, Prausnitz MR *et al*: Microneedle patch delivery to the skin of virus-like particles containing heterologous M2e extracellular domains of influenza virus induces broad heterosubtypic cross-protection. *Journal of controlled release : official journal of the Controlled Release Society* 2015, 210:208-216.

84. Wang L, Wang YC, Feng H, Ahmed T, Compans RW, Wang BZ: Virus-like particles containing the tetrameric ectodomain of influenza matrix protein 2 and flagellin induce heterosubtypic protection in mice. *BioMed research international* 2013, 2013:686549.

85. Tao W, Gill HS: M2e-immobilized gold nanoparticles as influenza A vaccine: Role of soluble M2e and longevity of protection. *Vaccine* 2015, 33(20):2307-2315.

86. Rappazzo CG, Watkins HC, Guarino CM, Chau A, Lopez JL, DeLisa MP, Leifer CA, Whittaker GR, Putnam D: Recombinant M2e outer membrane vesicle vaccines protect against lethal influenza A challenge in BALB/c mice. *Vaccine* 2016, 34(10):1252-1258.

87. Denis J, Acosta-Ramirez E, Zhao Y, Hamelin ME, Koukavica I, Baz M, Abed Y, Savard C, Pare C, Lopez Macias C *et al*: Development of a universal influenza A vaccine based on the M2e peptide fused to the papaya mosaic virus (PapMV) vaccine platform. *Vaccine* 2008, 26(27-28):3395-3403.

88. Leclerc D, Rivest M, Babin C, Lopez-Macias C, Savard P: A novel M2e based flu vaccine formulation for dogs. *PloS one* 2013, 8(10):e77084.

89. Ameiss K, Ashraf S, Kong W, Pekosz A, Wu WH, Milich D, Billaud JN, Curtiss R, 3rd: Delivery of woodchuck hepatitis virus-like particle presented influenza M2e by recombinant attenuated Salmonella displaying a delayed lysis phenotype. *Vaccine* 2010, 28(41):6704-6713.

90. Zhao G, Miao Y, Guo Y, Qiu H, Sun S, Kou Z, Yu H, Li J, Chen Y, Jiang S *et al*: Development of a heat-stable and orally delivered recombinant M2e-expressing B. subtilis spore-based influenza vaccine. *Human vaccines & immunotherapeutics* 2014, 10(12):3649-3658.

91. Hashemi H, Pouyanfard S, Bandehpour M, Noroozbabaei Z, Kazemi B, Saelens X, Mokhtari-Azad T: Immunization with M2e-displaying T7 bacteriophage nanoparticles protects against influenza A virus challenge. *PloS one* 2012, 7(9):e45765.

92. Deng L, Ibanez LI, Van den Bossche V, Roose K, Youssef SA, de Bruin A, Fiers W, Saelens X: Protection against Influenza A Virus Challenge with M2e-Displaying Filamentous Escherichia coli Phages. *PloS one* 2015, 10(5):e0126650.

93. Bessa J, Schmitz N, Hinton HJ, Schwarz K, Jegerlehner A, Bachmann MF: Efficient induction of mucosal and systemic immune responses by virus-like particles administered intranasally: implications for vaccine design. *European journal of immunology* 2008, 38(1):114-126.

94. Dabaghian M, Latifi AM, Tebianian M, Dabaghian F, Ebrahimi SM: A truncated C-terminal fragment of Mycobacterium tuberculosis HSP70 enhances cell-mediated immune response and longevity of the total IgG to influenza A virus M2e protein in mice. *Antiviral research* 2015, 120:23-31.

95. Ebrahimi SM, Dabaghian M, Tebianian M, Jazi MH: In contrast to conventional inactivated influenza vaccines, 4xM2e.HSP70c fusion protein fully protected mice against lethal dose of H1, H3 and H9 influenza A isolates circulating in Iran. *Virology* 2012, 430(1):63-72.

96. Mu X, Hu K, Shen M, Kong N, Fu C, Yan W, Wei A: Protection against influenza A virus by vaccination with a recombinant fusion protein linking influenza M2e to human serum albumin (HSA). *Journal of virological methods* 2016, 228:84-90.

97. Mardanova ES, Kotlyarov RY, Kuprianov VV, Stepanova LA, Tsybalova LM, Lomonossoff GP, Ravin NV: High immunogenicity of plant-produced candidate influenza vaccine based on the M2e peptide fused to flagellin. *Bioengineered* 2016, 7(1):28-32.

98. Mardanova ES, Kotlyarov RY, Kuprianov VV, Stepanova LA, Tsybalova LM, Lomonosoff GP, Ravin NV: Rapid high-yield expression of a candidate influenza vaccine based on the ectodomain of M2 protein linked to flagellin in plants using viral vectors. *BMC biotechnology* 2015, 15:42.

99. Huleatt JW, Nakaar V, Desai P, Huang Y, Hewitt D, Jacobs A, Tang J, McDonald W, Song L, Evans RK *et al*: Potent immunogenicity and efficacy of a universal influenza vaccine candidate comprising a recombinant fusion protein linking influenza M2e to the TLR5 ligand flagellin. *Vaccine* 2008, 26(2):201-214.

100. Wang BZ, Gill HS, He C, Ou C, Wang L, Wang YC, Feng H, Zhang H, Prausnitz MR, Compans RW: Microneedle delivery of an M2e-TLR5 ligand fusion protein to skin confers broadly cross-protective influenza immunity. *Journal of controlled release : official journal of the Controlled Release Society* 2014, 178:1-7.

101. Tompkins SM, Zhao ZS, Lo CY, Misplon JA, Liu T, Ye Z, Hogan RJ, Wu Z, Benton KA, Tumpey TM *et al*: Matrix protein 2 vaccination and protection against influenza viruses, including subtype H5N1. *Emerging infectious diseases* 2007, 13(3):426-435.

102. Alvarez P, Zylberman V, Ghersi G, Boado L, Palacios C, Goldbaum F, Mattion N: Tandem repeats of the extracellular domain of Matrix 2 influenza protein exposed in Brucella lumazine synthase decameric carrier molecule induce protection in mice. *Vaccine* 2013, 31(5):806-812.

103. Fan J, Liang X, Horton MS, Perry HC, Citron MP, Heidecker GJ, Fu TM, Joyce J, Przysiecki CT, Keller PM *et al*: Preclinical study of influenza virus A M2 peptide conjugate vaccines in mice, ferrets, and rhesus monkeys. *Vaccine* 2004, 22(23-24):2993-3003.

104. Zhao G, Du L, Xiao W, Sun S, Lin Y, Chen M, Kou Z, He Y, Lustigman S, Jiang S *et al*: Induction of protection against divergent H5N1 influenza viruses using a recombinant fusion protein linking influenza M2e to Onchocerca volvulus activation associated protein-1 (ASP-1) adjuvant. *Vaccine* 2010, 28(44):7233-7240.

105. Frace AM, Klimov AI, Rowe T, Black RA, Katz JM: Modified M2 proteins produce heterotypic immunity against influenza A virus. *Vaccine* 1999, 17(18):2237-2244.

106. Noh HJ, Chowdhury MY, Cho S, Kim JH, Park HS, Kim CJ, Poo H, Sung MH, Lee JS, Lim YT: Programming of Influenza Vaccine Broadness and Persistence by Mucoadhesive Polymer-Based Adjuvant Systems. *J Immunol* 2015, 195(5):2472-2482.

107. Zhang J, Fan HY, Zhang Z, Huang JN, Ye Y, Liao M: Recombinant baculovirus vaccine containing multiple M2e and adjuvant LTB induces T cell dependent, cross-clade protection against H5N1 influenza virus in mice. *Vaccine* 2016, 34(5):622-629.

108. De Filette M, Martens W, Roose K, Deroo T, Vervalle F, Bentahir M, Vandekerckhove J, Fiers W, Saelens X: An influenza A vaccine based on tetrameric ectodomain of matrix protein 2. *The Journal of biological chemistry* 2008, 283(17):11382-11387.

109. Mozdzanowska K, Feng J, Eid M, Kragol G, Cudic M, Otvos L, Jr., Gerhard W: Induction of influenza type A virus-specific resistance by immunization of mice with a synthetic multiple antigenic peptide vaccine that contains ectodomains of matrix protein 2. *Vaccine* 2003, 21(19-20):2616-2626.

110. Shen X, Soderholm J, Lin F, Kobinger G, Bello A, Gregg DA, Broderick KE, Sardesai NY: Influenza A vaccines using linear expression cassettes delivered via electroporation afford full protection against challenge in a mouse model. *Vaccine* 2012, 30(48):6946-6954.

111. Lalor PA, Webby RJ, Morrow J, Rusalov D, Kaslow DC, Rolland A, Smith LR: Plasmid DNA-based vaccines protect mice and ferrets against lethal challenge with A/Vietnam/1203/04 (H5N1) influenza virus. *The Journal of infectious diseases* 2008, 197(12):1643-1652.

112. Park KS, Seo YB, Lee JY, Im SJ, Seo SH, Song MS, Choi YK, Sung YC: Complete protection against a H5N2 avian influenza virus by a DNA vaccine expressing a fusion protein of H1N1 HA and M2e. *Vaccine* 2011, 29(33):5481-5487.

113. Chen L, Zanker D, Xiao K, Wu C, Zou Q, Chen W: Immunodominant CD4+ T-cell responses to influenza A virus in healthy individuals focus on matrix 1 and nucleoprotein. *Journal of virology* 2014, 88(20):11760-11773.

114. Grant E, Wu C, Chan KF, Eckle S, Bharadwaj M, Zou QM, Kedzierska K, Chen W: Nucleoprotein of influenza A virus is a major target of immunodominant CD8+ T-cell responses. *Immunology and cell biology* 2013, 91(2):184-194.

115. Ilyinskii PO, Gabai VL, Sunyaev SR, Thoidis G, Shneider AM: Toxicity of influenza A virus matrix protein 2 for mammalian cells is associated with its intrinsic proton-channeling activity. *Cell Cycle* 2007, 6(16):2043-2047.

116. Okuda K, Ihata A, Watabe S, Okada E, Yamakawa T, Hamajima K, Yang J, Ishii N, Nakazawa M, Ohnari K *et al*: Protective immunity against influenza A virus induced by immunization with DNA plasmid containing influenza M gene. *Vaccine* 2001, 19(27):3681-3691.

117. Price GE, Soboleski MR, Lo CY, Misplon JA, Pappas C, Houser KV, Tumpey TM, Epstein SL: Vaccination focusing immunity on conserved antigens protects mice and ferrets against virulent H1N1 and H5N1 influenza A viruses. *Vaccine* 2009, 27(47):6512-6521.

118. Mozdzanowska K, Zharikova D, Cudic M, Otvos L, Gerhard W: Roles of adjuvant and route of vaccination in antibody response and protection engendered by a synthetic matrix protein 2-based influenza A virus vaccine in the mouse. *Virology journal* 2007, 4:118.

119.	Zhao G, Lin Y, Du L, Guan J, Sun S, Sui H, Kou Z, Chan CC, Guo Y, Jiang S *et al*: An M2e-based multiple antigenic peptide vaccine protects mice from lethal challenge with divergent H5N1 influenza viruses. *Virology journal* 2010, 7:9.

120.	Zhao G, Sun S, Du L, Xiao W, Ru Z, Kou Z, Guo Y, Yu H, Jiang S, Lone Y *et al*: An H5N1 M2e-based multiple antigenic peptide vaccine confers heterosubtypic protection from lethal infection with pandemic 2009 H1N1 virus. *Virology journal* 2010, 7:151.

121.	Ma JH, Yang FR, Yu H, Zhou YJ, Li GX, Huang M, Wen F, Tong G: An M2e-based synthetic peptide vaccine for influenza A virus confers heterosubtypic protection from lethal virus challenge. *Virology journal* 2013, 10:227.

122.	Liu X, Guo J, Han S, Yao L, Chen A, Yang Q, Bo H, Xu P, Yin J, Zhang Z: Enhanced immune response induced by a potential influenza A vaccine based on branched M2e polypeptides linked to tuftsin. *Vaccine* 2012, 30(46):6527-6533.

123.	Liu WL, Chen YH: High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. *European journal of immunology* 2005, 35(2):505-514.

124.	Zhou C, Zhou L, Chen YH: Immunization with high epitope density of M2e derived from 2009 pandemic H1N1 elicits protective immunity in mice. *Vaccine* 2012, 30(23):3463-3469.

125.	Bachmann MF, Zinkernagel RM: Neutralizing antiviral B cell responses. *Annual review of immunology* 1997, 15:235-270.

126.	Liu W, Peng Z, Liu Z, Lu Y, Ding J, Chen YH: High epitope density in a single recombinant protein molecule of the extracellular domain of influenza A virus M2 protein significantly enhances protective immunity. *Vaccine* 2004, 23(3):366-371.

127.	Kim MC, Lee JS, Kwon YM, O E, Lee YJ, Choi JG, Wang BZ, Compans RW, Kang SM: Multiple heterologous M2 extracellular domains presented on virus-like particles confer broader and stronger M2 immunity than live influenza A virus infection. *Antiviral research* 2013, 99(3):328-335.

128.	Stepanova LA, Kotlyarov RY, Kovaleva AA, Potapchuk MV, Korotkov AV, Sergeeva MV, Kasianenko MA, Kuprianov VV, Ravin NV, Tsybalova LM *et al*: Protection against multiple influenza A virus strains induced by candidate recombinant vaccine based on heterologous M2e peptides linked to flagellin. *PloS one* 2015, 10(3):e0119520.

129.	El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.

130.	Bruhns P: Properties of mouse and human IgG receptors and their contribution to disease models. *Blood* 2012, 119(24):5640-5649.

131.	Hovden AO, Cox RJ, Haaheim LR: Whole influenza virus vaccine is more immunogenic than split influenza virus vaccine and induces primarily an IgG2a response in BALB/c mice. *Scandinavian journal of immunology* 2005, 62(1):36-44.

132.	Huber VC, Lynch JM, Bucher DJ, Le J, Metzger DW: Fc receptor-mediated phagocytosis makes a significant contribution to clearance of influenza virus infections. *J Immunol* 2001, 166(12):7381-7388.

133.	Moran TM, Park H, Fernandez-Sesma A, Schulman JL: Th2 responses to inactivated influenza virus can Be converted to Th1 responses and facilitate recovery from heterosubtypic virus infection. *The Journal of infectious diseases* 1999, 180(3):579-585.

134.	Huber VC, McKeon RM, Brackin MN, Miller LA, Keating R, Brown SA, Makarova N, Perez DR, Macdonald GH, McCullers JA: Distinct contributions of vaccine-induced immunoglobulin G1 (IgG1) and IgG2a antibodies to protective immunity against influenza. *Clinical and vaccine immunology : CVI* 2006, 13(9):981-990.

135.	Fiers W, De Filette M, Birkett A, Neirynck S, Jou WM: A "universal" human influenza A vaccine. *Virus research* 2004, 103(1-2):173-176.

136.	Schmitz N, Beerli RR, Bauer M, Jegerlehner A, Dietmeier K, Maudrich M, Pumpens P, Saudan P, Bachmann MF: Universal vaccine against influenza virus: linking TLR signaling to anti-viral protection. *European journal of immunology* 2012, 42(4):863-869.

137.	Stevens TL, Bossie A, Sanders VM, Fernandezbotran R, Coffman RL, Mosmann TR, Vitetta ES: Regulation of Antibody Isotype Secretion by Subsets of Antigen-Specific Helper T-Cells. *Nature* 1988, 334(6179):255-258.

138.	Ibanez LI, Roose K, De Filette M, Schotsaert M, De Sloovere J, Roels S, Pollard C, Schepens B, Grooten J, Fiers W *et al*: M2e-displaying virus-like particles with associated RNA promote T helper 1 type adaptive immunity against influenza A. *PloS one* 2013, 8(3):e59081.

139.	Eriksson AM, Schon KM, Lycke NY: The cholera toxin-derived CTA1-DD vaccine adjuvant administered intranasally does not cause inflammation or accumulate in the nervous tissues. *J Immunol* 2004, 173(5):3310-3319.

140.	Shim BS, Choi YK, Yun CH, Lee EG, Jeon YS, Park SM, Cheon IS, Joo DH, Cho CH, Song MS *et al*: Sublingual immunization with M2-based vaccine induces broad protective immunity against influenza. *PloS one* 2011, 6(11):e27953.

141.	Eliasson DG, Helgeby A, Schon K, Nygren C, El-Bakkouri K, Fiers W, Saelens X, Lovgren KB, Nystrom I, Lycke NY: A novel non-toxic combined CTA1-DD and ISCOMS adjuvant vector for effective mucosal immunization against influenza virus. *Vaccine* 2011, 29(23):3951-3961.

142.	Song JM, Van Rooijen N, Bozja J, Compans RW, Kang SM: Vaccination inducing broad and improved cross protection against multiple subtypes of influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108(2):757-761.

143.	Lee YN, Kim MC, Lee YT, Kim YJ, Lee J, Kim C, Ha SH, Kang SM: Co-immunization with tandem repeat heterologous M2 extracellular proteins overcomes strain-specific protection of split vaccine against influenza A virus. *Antiviral research* 2015, 122:82-90.

144.	Zharikova D, Mozdzanowska K, Feng J, Zhang M, Gerhard W: Influenza type A virus escape mutants emerge in vivo in the presence of antibodies to the ectodomain of matrix protein 2. *Journal of virology* 2005, 79(11):6644-6654.

145.	Gerhard W, Mozdzanowska K, Zharikova D: Prospects for universal influenza virus vaccine. *Emerging infectious diseases* 2006, 12(4):569-574.

146.	Ramos EL, Mitcham JL, Koller TD, Bonavia A, Usner DW, Balaratnam G, Fredlund P, Swiderek KM: Efficacy and safety of treatment with an anti-m2e monoclonal antibody in experimental human influenza. *The Journal of infectious diseases* 2015, 211(7):1038-1044.

147.	Safety Study of Recombinant M2e Influenza-A Vaccine in Healthy Adults (FLU-A) [https://clinicaltrials.gov/ct2/show/results/NCT00819013?sect=Xa0156#outcome4]

148.	Turley CB, Rupp RE, Johnson C, Taylor DN, Wolfson J, Tussey L, Kavita U, Stanberry L, Shaw A: Safety and immunogenicity of a recombinant M2e-flagellin influenza vaccine (STF2.4xM2e) in healthy adults. *Vaccine* 2011, 29(32):5145-5152.

149.	Talbot HK, Rock MT, Johnson C, Tussey L, Kavita U, Shanker A, Shaw AR, Taylor DN: Immunopotentiation of trivalent influenza vaccine when given with VAX102, a recombinant influenza M2e vaccine fused to the TLR5 ligand flagellin. *PloS one* 2010, 5(12):e14442.

150.	Status of Vaccine Research and Development of Universal Influenza Vaccine Prepared for WHO PD-VAC [http://who.int/immunization/research/meetings_workshops/Universal_Influenza_VaccineRD_Sept2014.pdf]

151.	Dynavax Reports New Phase 1a and Phase 1b Data for Universal Flu Vaccine Candidate [http://investors.dynavax.com/releasedetail.cfm?ReleaseID=551606]

152.	Clinical Evaluation of N8295 Influenza A Vaccine Containing M2e and NP Antigens Conjugated to an Oligonucleotide Immunostimulatory Sequence [http://www.who.int/immunization/diseases/193_janssen_abstract_line469.pdf]

153.	Grandea AG, 3rd, Olsen OA, Cox TC, Renshaw M, Hammond PW, Chan-Hui PY, Mitcham JL, Cieplak W, Stewart SM, Grantham ML *et al*: Human antibodies reveal a protective epitope that is highly conserved among human and nonhuman influenza A viruses. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(28):12658-12663.

154.	Pleguezuelos O, Robinson S, Stoloff GA, Caparros-Wanderley W: Synthetic Influenza vaccine (FLU-v) stimulates cell mediated immunity in a double-blind, randomised, placebo-controlled Phase I trial. *Vaccine* 2012, 30(31):4655-4660.

155.	Pleguezuelos O, Robinson S, Fernandez A, Stoloff GA, Mann A, Gilbert A, Balaratnam G, Wilkinson T, Lambkin-Williams R, Oxford J *et al*: A Synthetic Influenza Virus Vaccine Induces a Cellular Immune Response That Correlates with Reduction in Symptomatology and Virus Shedding in a Randomized Phase Ib Live-Virus Challenge in Humans. *Clinical and vaccine immunology : CVI* 2015, 22(7):828-835.

156.   Jegerlehner A, Schmitz N, Storni T, Bachmann MF: Influenza A vaccine based on the extracellular domain of M2: weak protection mediated via antibody-dependent NK cell activity. *J Immunol* 2004, 172(9):5598-5605.

157.   Wang R, Song A, Levin J, Dennis D, Zhang NJ, Yoshida H, Koriazova L, Madura L, Shapiro L, Matsumoto A *et al*: Therapeutic potential of a fully human monoclonal antibody against influenza A virus M2 protein. *Antiviral research* 2008, 80(2):168-177.

158.   Deng L, Cho KJ, Fiers W, Saelens X: M2e-Based Universal Influenza A Vaccines. *Vaccines* 2015, 3(1):105-136.

159.   Verbonitz MW, Ennis FA, Hicks JT, Albrecht P: Hemagglutinin-specific complement-dependent cytolytic antibody response to influenza infection. *The Journal of experimental medicine* 1978, 147(1):265-270.

160.   Lee YN, Lee YT, Kim MC, Hwang HS, Lee JS, Kim KH, Kang SM: Fc receptor is not required for inducing antibodies but plays a critical role in conferring protection after influenza M2 vaccination. *Immunology* 2014, 143(2):300-309.

161.   Guilliams M, Bruhns P, Saeys Y, Hammad H, Lambrecht BN: The function of Fcgamma receptors in dendritic cells and macrophages. *Nature reviews Immunology* 2014, 14(2):94-108.

162.   Hashimoto Y, Moki T, Takizawa T, Shiratsuchi A, Nakanishi Y: Evidence for phagocytosis of influenza virus-infected, apoptotic cells by neutrophils and macrophages in mice. *J Immunol* 2007, 178(4):2448-2457.

163.   Song JM, Wang BZ, Park KM, Van Rooijen N, Quan FS, Kim MC, Jin HT, Pekosz A, Compans RW, Kang SM: Influenza virus-like particles containing M2 induce broadly cross protective immunity. *PloS one* 2011, 6(1):e14538.

164.   Nogusa S, Ritz BW, Kassim SH, Jennings SR, Gardner EM: Characterization of age-related changes in natural killer cells during primary influenza infection in mice. *Mechanisms of ageing and development* 2008, 129(4):223-230.

165.   Simhadri VR, Dimitrova M, Mariano JL, Zenarruzabeitia O, Zhong W, Ozawa T, Muraguchi A, Kishi H, Eichelberger MC, Borrego F: A Human Anti-M2 Antibody Mediates Antibody-Dependent Cell-Mediated Cytotoxicity (ADCC) and Cytokine Secretion by Resting and Cytokine-Preactivated Natural Killer (NK) Cells. *PloS one* 2015, 10(4):e0124677.

166.   Fu TM, Freed DC, Horton MS, Fan J, Citron MP, Joyce JG, Garsky VM, Casimiro DR, Zhao Q, Shiver JW *et al*: Characterizations of four monoclonal antibodies against M2 protein ectodomain of influenza A virus. *Virology* 2009, 385(1):218-226.

167.   Misplon JA, Lo CY, Gabbard JD, Tompkins SM, Epstein SL: Genetic control of immune responses to influenza A matrix 2 protein (M2). *Vaccine* 2010, 28(36):5817-5827.

168.   Jameson J, Cruz J, Terajima M, Ennis FA: Human CD8+ and CD4+ T lymphocyte memory to influenza A viruses of swine and avian species. *J Immunol* 1999, 162(12):7578-7583.

169.   Jameson J, Cruz J, Ennis FA: Human cytotoxic T-lymphocyte repertoire to influenza A viruses. *Journal of virology* 1998, 72(11):8682-8689.

170.   Gianfrani C, Oseroff C, Sidney J, Chesnut RW, Sette A: Human memory CTL response specific for influenza A virus is broad and multispecific. *Human immunology* 2000, 61(5):438-452.

171.   Lee YN, Kim MC, Lee YT, Hwang HS, Cho MK, Lee JS, Ko EJ, Kwon YM, Kang SM: AS04-adjuvanted virus-like particles containing multiple M2 extracellular domains of influenza virus confer improved protection. *Vaccine* 2014, 32(35):4578-4585.

172.   Schotsaert M, Ysenbaert T, Neyt K, Ibanez LI, Bogaert P, Schepens B, Lambrecht BN, Fiers W, Saelens X: Natural and long-lasting cellular immune responses against influenza in the M2e-immune host. *Mucosal immunology* 2013, 6(2):276-287.

173.   Schotsaert M, Ysenbaert T, Smet A, Schepens B, Vanderschaeghe D, Stegalkina S, Vogel TU, Callewaert N, Fiers W, Saelens X: Long-Lasting Cross-Protection Against Influenza A by Neuraminidase and M2e-based immunization strategies. *Scientific reports* 2016, 6:24402.

Chapter 3

A brief history of DNA sequencing technologies

3.1. Introduction

The ability to determine the genome sequence of viruses had an important impact on the progress of viral research. The coat protein of bacteriophage MS2 was the first complete gene sequenced, an achievement that was accomplished using RNA sequencing by the research group of Walter Fiers in 1972 [1]. The full genome sequence of phage MS2 was reported four years later by the same group [2]. Sooner or later, however, Sanger sequencing became the 'Golden Standard' to determine the sequence of DNA for many years since its introduction in 1977 [3]. However, the introduction of next-generation sequencing technologies in 2005 revolutionized the sequencing field, resulting in an exponential increase in deposited viral genome sequences [4]. Up to now, more than 99,000 complete viral genome sequences are registered in the NCBI virus genome database [5]. We present here a brief history of DNA sequencing technologies and some successful applications of next-generation sequencing technologies in the virology field.

3.2. First generation of automated DNA sequencers

History of DNA sequencing

DNA sequencing, one of the most important technologies in life science today, has a relatively recent history, dating back to the late 1970's, twenty years after Francis Crick and James Watson, along with Rosalind Franklin and Maurice Wilkins, discovered the three-dimensional structure of DNA [6, 7]. The first techniques to determine the order of nucleotides in a stretch of DNA were simultaneously described by Frederick Sanger and by Walter Gilbert and Allan Maxam, for which the first two were awarded the Nobel Prize in chemistry in 1980 [8]. Besides the development of these DNA sequencing techniques, the introduction of reverse transcription was crucial for DNA sequencing of viral RNA genomes [9, 10].

The method of Maxam-Gilbert sequencing was the first widely adopted DNA sequencing technique, and was based on chemical reactions inducing nucleotide specific breaks in radiolabelled DNA [11]. The generated fragments were subsequently separated on a polyacrylamide gel and the nucleotide sequence determined by autoradiography [11]. An advantage of this technique is the fact that purified DNA could be used directly. The first full-length influenza genome segment, the HA of fowl plague influenza virus, was sequenced using the method of Maxam and Gilbert [12]. In addition, Maxam and Gilbert sequencing of the HA genome segment of A/duck/Ukraine/1/63 (H3N8), provided very strong molecular evidence for viral reassortment between a human and avian influenza virus as a step that can lead to a new pandemic [13].

In parallel with the chemical DNA sequencing method, Frederick Sanger worked on an alternative approach to determine the sequence of a DNA molecule: the 'plus and minus technique' [14]. In this technique, a radioactively labelled complement of the template is made by enzymatically extending

an annealed primer, which is followed by two different, separate reactions: The 'minus' and the 'plus' reaction [14]. In the minus reaction, one type of deoxynucleotide triphosphates (dNTPs) is missing, resulting in termination of the DNA polymerase at its 3' end at one position before the missing dNTP [14]. In contrast, in the 'plus' reaction only one type of dNTP and T4 DNA polymerase are present, so all fragments end with the same dNTP [14]. Afterwards, the fragments in the eight separate reactions were resolved on a polyacrylamide gel, followed by determining the nucleotide sequence based on the pattern of radioactive bands [14]. This technique was used to sequence the first DNA genome in history, the genome of bacteriophage PhiX 174 [15].

A breakthrough in the DNA sequencing field came when Frederick Sanger described the more rapid and more accurate 'chain termination' or 'dideoxy' sequencing technique, which was based on his 'plus and minus' technique [3]. This sequencing technique uses a mix of all four dNTPs (of which one is radioactively labelled), together with one type of dideoxynucleotidetriphosphates (ddNTPs) in the elongation reaction [3]. Since these ddNTPs lack a 3'-hydroxyl group, no extra nucleotides can be added once a ddNTP is incorporated by the DNA polymerase, resulting in termination of the DNA elongation reaction [3]. Since these ddNTPs are 100 times less abundant in the reaction, DNA fragments of varying lengths are produced from a single template [3]. Separating these fragments based on their size using polyacrylamide gel electrophoresis enables deduction of the sequence by autoradiography.

First generation of automated sequencing technologies

The Sanger sequencing technique, together with some important adaptations, resulted in the development of the first prototype of an automated sequencer by Lloyd Smith in 1985 [16]. An important adjustment was the labelling of the sequencing primers with a fluorescent dye, omitting radioactive labelling and hazardous chemicals in the reaction. Four different fluorophores were coupled to the sequencing primer, emitting light at a different wavelength when excited by a laser. Per sequencing primer, a different sequencing reaction was performed, each reaction with a mix of all four dNTPs and only one type of ddNTPs. Afterwards, the reactions were pooled and co-electrophoresed, increasing the throughput per gel four times, since all four reactions could be analysed in parallel. The fluorophores coupled to the sequencing primers were excited by a fixed laser while the fragments of different lengths migrate through the polyacrylamide gel. The emitted light was then recorded by a camera, resulting in a sequencing chromatogram. The DNA sequence could be directly determined based on the order of the detected colours. This prototype was brought to the market as the first automated DNA sequencer, the ABI 370, by Applied Biosystems in 1986 [17]. The throughput was later increased by fluorescent labelling of the ddNTPs instead of the sequencing primers. Consequently, the sequence of a DNA fragment could be determined in a single sequencing reaction.

A next important increase in throughput was reached when the polyacrylamide gel electrophoresis was replaced by capillary gel electrophoresis, where a semi-liquid polymer is injected before sequencing, omitting the need to pour gels. In addition, the introduction of thermal cycling of enzymatic DNA synthesis allowed to sequence minute amounts of DNA in a biological sample. In this method, the DNA is subjected to successive rounds of denaturation, annealing and extension with only one primer, resulting in linear amplification of the extension products [18]. In parallel, all the different steps in the sequencing process (*e.g.* more accurate enzymes and data processing) are still continuously being optimized. The capillary sequencers of Applied BioSystems currently on the market have up to 96 capillaries, can handle 384-well plates and can sequence up to 2304 samples per day (3730xl DNA Analyzer) [19].

The main advantages of automated capillary sequencing today are the high accuracy (99.999%) and the long read length (750 - 1000 bp). However, these 'first generation of automated sequence techniques' were limited in throughput, rather slow, have a high cost and limited sensitivity. In addition, the sequence reaction is sequence dependent, since the sequencing primer should anneal upstream of the sequence of interest. Despite the emergence of next- and third-generation sequencing techniques, automated Sanger sequencing technology remains a powerful technique and is still the mainstream sequencing method for many applications, *e.g.* to confirm the sequence of cloned genes of interest in a plasmid. In addition, Sanger sequencing is used to complement the NGS technologies, *e.g.* to confirm single-nucleotide polymorphisms detected by NGS or to aid in *de novo* genome assembly when repetitive sequence motifs are present [20, 21].

3.3. Next-generation sequencing (NGS) techniques: Massive parallel sequencing

Focused on the race to the $1000 genome, which should make it possible to sequence the full genome of a person for $1000, researchers searched for alternative sequencing methods that could reduce time and cost, whilst increasing the output [22]. This lead to the development of several 'next-generation sequencing' (NGS) techniques. In contrast to capillary Sanger sequencing, many DNA fragments are sequenced in parallel on these platforms, without the requirement of prior sequence knowledge on the sample. Different NGS techniques are on the market, which are quite diverse in biochemistry, with the result that each platform has its own characteristics concerning sequence read length, accuracy, output, cost and time (Table 1). However, they all share a conceptually similar workflow:

- Library preparation: DNA is randomly fragmented into fragments with an appropriate length. Different fragmentation techniques exist: enzymatic (*e.g.* enzymatic digestion and Nextera XT) or mechanic (*e.g.* nebulization and sonication). Subsequently, sequencing adaptors are ligated to the fragments to enable immobilization of the DNA fragment on the sequencing chip or beads. These adaptors also provide priming sequences for both amplification and

sequencing. When different samples are pooled in a single sequencing run, barcodes are added. A barcode is a small nucleotide stretch that is sequenced along with the DNA fragment and makes discrimination between the reads derived from different samples possible.

- Clonal amplification of DNA fragments: The limited sensitivity of the detectors currently used requires clonal amplification of DNA fragments, *e.g.* by emulsion PCR (emPCR) or bridge-amplification, resulting in clusters of identical DNA molecules.

- Sequencing: most of the NGS technologies use sequencing-by-synthesis (454, Solexa/Illumina, Ion Torrent), although also sequencing-by-ligation (AB SOLID) can be performed. At each sequencing cycle, the sequence data for all clusters is analysed in parallel.

- Bioinformatics data analysis: Millions to billions of sequencing reads are generated in a single sequencing run and exported in 'fastq'-format. In such files, all sequence reads are present, together with their read name and the quality score per nucleotide in the sequence read. The sequencing quality per base is expressed as a Phred score (Q), which is logarithmically related to the base-calling error probabilities (P): $Q = -10 \cdot \log_{10}(P)$. When a Phred score Q of 30 is assigned to a base, it means that the chance that this base is called incorrectly is 1 in 1000. Before analysis, the parts of the sequencing read with low sequencing quality should be trimmed, together with contamination of adaptor sequences. To prevent aspecific mapping, short reads are also filtered out. Multiple NGS software methods are available to process primary NGS sequencing reads.

Four different NGS technologies have been commercially released since 2005: 454 pyrosequencing (Roche), Illumina sequencing-by-synthesis (Illumina), SOLiD sequencing-by-ligation (Thermo Scientific) and Ion Torrent semiconductor sequencing (Thermo Scientific) (Table 1). In the course of 2016, both 454 (Roche) and SOLiD (Thermo Scientific) were commercially discontinued.

Table 1: Overview of next-generation sequencing platforms. The maximum read length, the maximum number of reads, the maximum output per sequencing run and the time for a sequencing run with the maximum read length are presented. The information provided is obtained from manufacturer's data (as on August 12, 2016), except the values for accuracy. The accuracy is determined as the percentage of errors per base within single reads of the maximum read length (2014 update of the 'Field guide to next-generation sequencers') [23, 24].

Sequencing platform	Read length (bp)	Number of reads	Output	Accuracy	Run time
454					
- GS Junior	400	100,000	35 Mb		10 hours
- GS Junior +	700	100,000	70 Mb	1%	18 hours
- GS FLX+	700	1,000,000	700 Mb		23 hours
Illumina					
- MiniSeq	2x150	25,000,000	7,5 Gb		24 hours
- MiSeq	2x300	25,000,000	15 Gb		55 hours
- NextSeq500	2x150	400,000,000	120 Gb	0.1%	29 hours
- HiSeq	2x150	5,000,000,000	1500 Gb		3.5 days
- HiSeq x	2x150	6,000,000,000	1800 Gb		3 days
SOLiD					
- 5500	75	700,000,000	120 Gb	≤0.1%	6 days
- 5500xl	75	1,400,000,00	240 Gb		10 days
Ion Torrent					
- PGM	400	5,500,000	2 Gb		7,3 hours
- Ion Proton	200	80,000,000	10 Gb	1%	4 hours
- Ion S5	400	20,000,000	8 Gb		4 hours

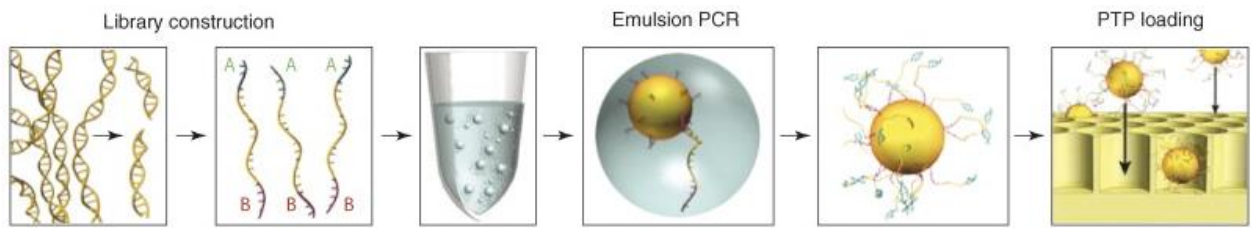### 3.3.1. 454 pyrosequencing (Roche)

The 454 GS20 sequencing machine was the first next-generation DNA sequencer launched on the market in 2005 by Roche Diagnostics. The most recent 454 sequencers were the GS Junior, GS Junior+ and GS FLX+, which were all discontinued in 2016.

In 454 sequencing (Figure 1), two different adaptors, adaptor A and B, are first ligated to the DNA fragments. Adaptor B is biotinylated and is used to capture the sequencing library on streptavidin-coated beads. The complementary strand is removed and the single stranded library is subsequently released from the beads. This library is then clonally amplified using emulsion PCR (emPCR), to increase the signal upon incorporation of a nucleotide during sequencing. To enable emPCR, the single stranded DNA fragments are first captured on beads containing covalently linked, complementary adaptors (Figure 1.A). The fragments are bound to the beads under conditions that favour the annealing of one fragment per bead. Subsequently, the beads are captured in droplets containing heat-stable PCR-reaction-mixture-in-oil emulsions. In these microreactors, the fragment is clonally amplified by emPCR [25-27]. This results in millions of identical sequences coupled to a bead. Subsequently, the beads carrying single-stranded DNA clones are deposited into wells of a fibre-optic

slide and the DNA sequence of each bead is individually determined using pyrosequencing (Figure 1.B). This sequencing technique uses natural nucleotides in a sequencing-by-synthesis protocol and finds its origin in the fact that DNA polymerization can be monitored by measuring phyrophosphate production, which can be detected by a reaction that gives light [28]. During pyrosequencing, the dNTPs are sequentially flown over the microtiterplate in which the sequencing primer is bound to its complementary sequence on the immobilized DNA fragments. When a nucleotide is incorporated by the DNA polymerase, inorganic pyrophosphate (PPi) and a hydrogen ion are released [29-31]. This released pyrophosphate is quantitatively converted into ATP by sulfurylase. ATP is a cofactor for the enzyme luciferase, which subsequently oxidizes luciferin to oxyluciferin along with the production of visible light (Figure 1.B). The amount of emitted light per well is detected by a CCD (charge-coupled device) sensor. The number of nucleotides added can be determind since the emitted light is proportional to the amount of pyrophosphate produced, which is directly proportional to the number of incorporated nucleotides [32]. Before the addition of the next dNTP, the unincorporated dNTPs and the produced ATP are degraded by an apyrase. The DNA sequence can be determined since dNTPs are sequentially added in a known order to the sequencing reaction [32].

The major advantage of 454 sequencing was the long reads that were obtained (up to 1000 bases on the GS FLX Titanium XL+). Nevertheless, its output was limited (700 Mb on the GS FLX Titanium XL+). In addition, the reagents were quite expensive, resulting in the highest cost per nucleotide of the NGS technologies [23, 33]. Since multiple identical nucleotides can be incorporated in a single reaction which generates a single light signal, the main sequencing errors were homopolymer errors.

A. Sequence library preparation

B. Pyrosequencing
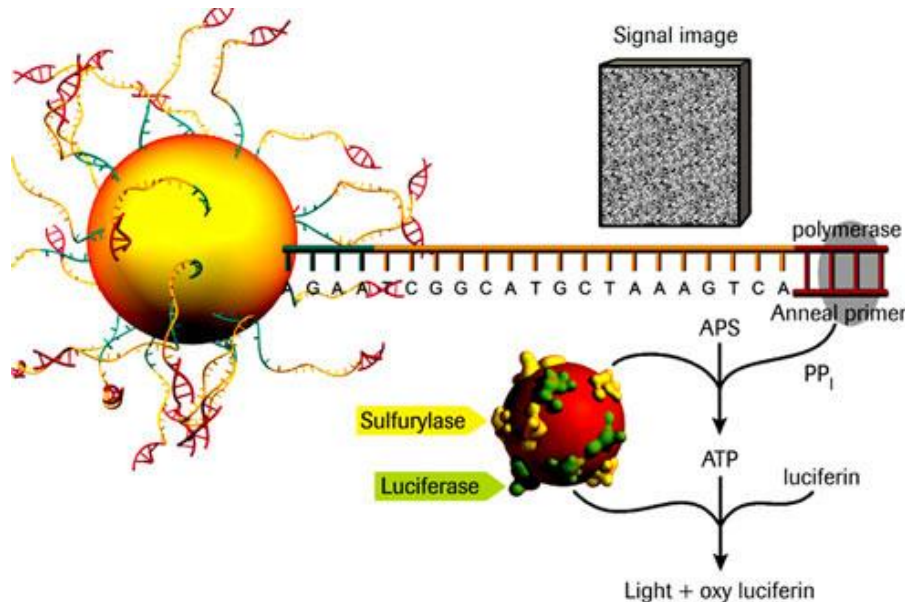
Figure 1: 454 pyrosequencing. (A) Sequence library preparation. Adaptors (A and B) are added to the DNA fragments, followed by clonally amplification of the DNA library on beads by emulsion PCR. The beads carrying multiple copies of the DNA fragment are deposited into wells of a pico titer plate (PTP). **(B) Pyrosequencing reaction.** After annealing of the sequencing primer, the different types of dNTPs are sequentially added. When a nucleotide is incorporated by the DNA polymerase, inorganic pyrophosphate (PPi) is released. The sulfurylase converts this PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). ATP is a cofactor for luciferase, which oxides luciferin to oxyluciferin and the production of visible light, which is quantitative for the amount of released PPi, and thus the number of incorporated nucleotides. Figure obtained from [34] and [35].

3.3.2. Solexa/Illumina sequencing

The first Solexa sequencer, the Genome Analyzer, was commercially launched in 2006. In 2007, Solexa was acquired by Illumina. There are currently five series of Illumina sequencers on the market: Illumina MiniSeq, Illumina MiSeq, NextSeq 500, HiSeq and HiSeq X, with maximum output ranging from 8 Gb to 1800 Gb and read lengths up to 2 x 300 bp. In the second half of 2017, Illumina announced that they will commercialize their 'Project Firefly', a sequencing instrument which employs the Illumina sequencing-by-synthesis chemistry on a semiconductor chip [36].

The DNA fragments that are to be sequenced by Illumina sequencing are first clonally amplified using "bridge amplification" (Figure 2.A) [37]. To enable this, different Illumina adaptors (P5 and P7) are

first ligated to each end of the DNA fragment. The DNA fragments are subsequently denatured and applied on an Illumina flow cell containing the two types of single-stranded adaptor sequences covalently coupled to the glass plate (Figure 2.A1). The DNA fragments hybridize to the adaptors on the flow cell (Figure 2.A2) and a polymerase extends the adaptor to synthesize the complementary strand (Figure 2.A3). This complementary strand is covalently coupled to the chip and the original DNA fragment is removed after denaturation (Figure 2.A4). The coupled fragment is clonally amplified by an isothermic reaction on the chip, called 'bridge amplification'. In this reaction, a single strand of DNA bends over to hybridize to an adjacent adaptor primer on the flow cell to form a bridge (Figure 2.A5). The hybridized primer is extended by polymerases, generating the complementary strand, forming a double stranded bridge (Figure 2.A6). The bridge is then denatured, resulting in two complementary copies of covalently bound single stranded templates (Figure 2.A7). These steps are repeated for millions of clusters in parallel, until multiple clonal copies of all the DNA fragment are generated (Figure 2.A8). Finally, all the reverse strands are removed by restriction digestion and the free 3' ends are blocked (using ddNTPs) to prevent unwanted DNA priming (Figure 2.A9).

After bridge-amplification, the first sequencing primer is hybridized to the P5 adaptor sequence and extended by a polymerase to create the sequencing read (Figure 2.B). During each sequencing cycle, all four fluorescently labelled dNTPs (all with a different, cleavable fluorophore) are flown over the flow cell. Due to the presence of a removable blocking group (3'-O-azidomethyl) on each dNTP, only one complementary dNTP will be incorporated per cycle in a cluster [38]. A laser excites the fluorophores coupled to the dNTPs per cluster and the characteristic emission spectrum, with specific wave length and signal intensity, is detected by a CCD camera and determines the base call. Millions of clusters are analyzed in parallel. Afterwards, the terminator and fluorescent dye are cleaved and the cycle repeated [39]. Since the clusters are fixed, the sequence of the DNA fragment is obtained by joining the detected fluorescent signals per position on the chip. After completion of the first read, all DNA fragments are denatured and the complementary sequencing product washed away. When multiple samples are pooled on a single flow cell, the barcode sequence is determined in a separate sequencing reaction. A second sequencing primer, recognizing a different sequence, binds to the template and the barcode sequence is determined in the same way as for the first sequencing read.

Illumina sequencing has the possibility to perform paired-end sequencing, *i.e.* sequencing of both the forward and reverse strand of a DNA fragment. To enable paired-end sequencing, the sequencing primer and its sequencing product are first removed through denaturation, followed by removal of the 3' blocking group from the template. The complementary strand is now synthesised by bridge amplification, followed by denaturation of the bridge and removal of the forward strand by restriction digest. After addition of the second sequencing primer, the sequencing protocol of the reverse strand is identical to that of the forward strand.

A. Bridge amplification
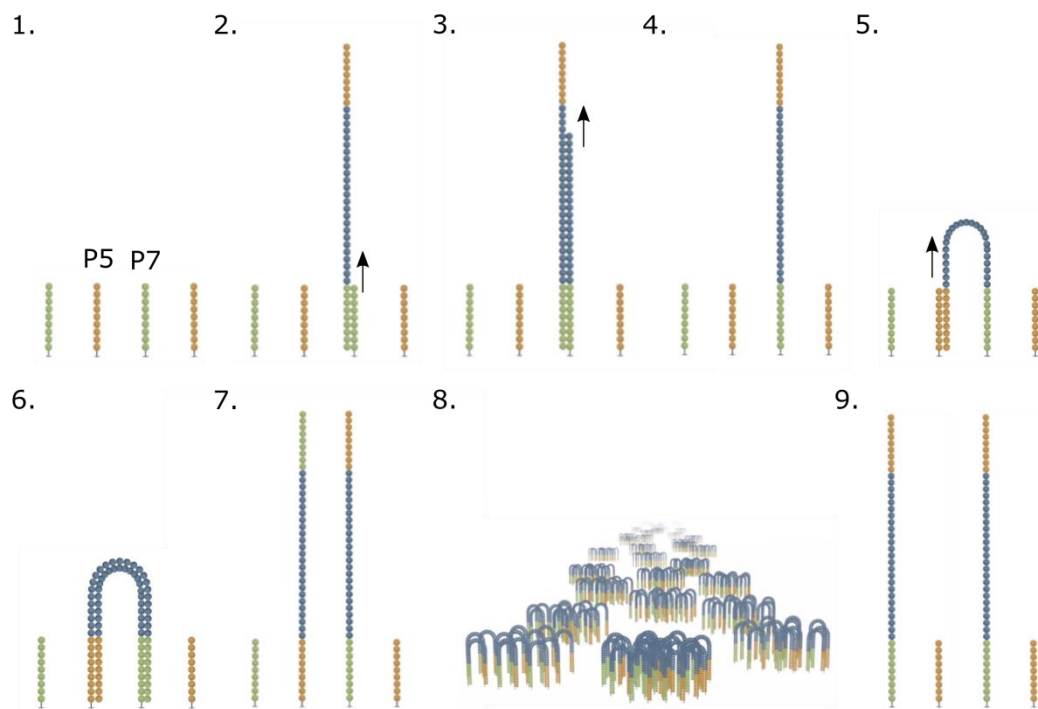
B. Sequencing-by-synthesis
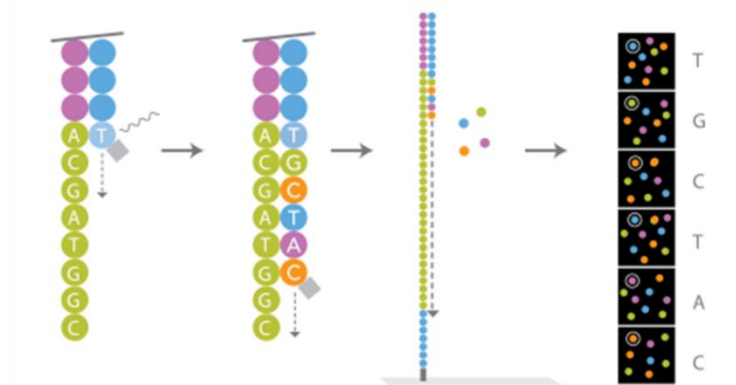
Figure 2: Illumina sequencing. (A) Bridge amplification. An Illumina flow cells contains billions of two different types of covalently coupled adaptors (P5 and P7) (1) and the DNA library anneals with its adaptor sequence to its complementary sequence on the flow cell (2), followed by elongation by a DNA polymerase (3). After denaturation, the original DNA fragment is washed away and its complement is now covalently attached to the flow cell (4). This DNA fragment bends to hybridize with an adjacent adaptor sequence (5), which is used as primer in an isothermal reaction to synthesize the reverse strand (6). After denaturation (7), this process is repeated several times, resulting in millions to billions (depending on the sequencing platform) of clusters of clonally amplified 'double-stranded DNA bridges' (8). In a final step, all 'double-stranded bridges' are denatured, and the reverse strand removed by restriction digest, leaving clusters of clonally amplified forward strands on the chip (9). **(B) Sequencing-by-synthesis**. A single nucleotide coupled to a fluorophoric group and a blocking group is incorporated during each sequencing cycle. The incorporated nucleotide is detected based on the emission spectra of its coupled fluorophore. Next, the blocking group and fluorophore are removed and the sequencing cycle repeated. The number of sequencing cycles determines the read length. The sequence of the DNA fragment is obtained after concatenating the fluorescent signals per cluster. Figures obtained adapted from [40] and [41].

An advantage of Illumina sequencing is that the clonal amplification is directly performed on the chip, which is directly followed by sequencing. This decreases the hands-on time and the chance for contamination. In addition, the base per base analysis in Illumina sequencing, decreases the change of detecting insertions or deletions at homopolymers. The possibility to perform paired-end sequencing allows sequencing of fragments in both directions resulting in overlapping reads. Additionally, it allows both ends of a larger fragment to be sequenced, facilitating *e.g.* the detection of genomic rearrangements. At present, the different Illumina platforms have the lowest total error rate of the commercially available NGS technologies [23, 33]. A disadvantage of Illumina sequencing is the requirement of a heterogeneous base composition across the sequence clusters that are imaged to obtain high-quality data. The nucleotide diversity can be increased by adding a sample with high-diversity to the sequencing run, such as the PhiX genome which is also included as sequencing control in an Illumina sequencing run. In addition, Illumina sequencing uses lasers, optics and fluorophores which may, in the future, limit reductions in sequencing costs.

3.3.3. SOLiD

The SOLiD system (Thermo Scientific) was commercially available since 2007. Thermo Scientific decided to discontinue their two SOLiD platforms, the 5500 and the 5500xl, in 2016.

In contrast to the other sequencers, the SOLiD (Sequencing-by-Oligo-Ligation Detection) technology uses two-base sequencing and sequencing-by-ligation [42]. In the first series of SOLiD sequencers, the DNA fragments were first clonally amplified on beads using emPCR, followed by covalently coupling of these beads to a glass slide. In the latest 5500 Series of Genetic Analysers, the library was directly amplified on the flowchip in an isothermal reaction. In the SOLiD sequencing technique, a universal sequencing primer first hybridizes to the template. Afterwards, the sequence of the DNA fragment is determined by ligating di-base probes to the primer (Figure 3.A). These eight nucleotide probes contain a ligation site, two specific bases (each dinucleotide corresponds to one of the four fluorescent dyes), followed by six universal bases binding to any of the four nucleotides and a fluorescent dye coupled to the last nucleotide [43]. Four fluorescent dyes are used and each dye is coupled to four out of 16 possible dinucleotide sequences (Figure 3.A). The complementary probe will hybridize to the template and a ligase will join the 5' of the growing strand with the 3' of the probe (Figure 3.B). A laser will excite the fluorescent dye and the emission spectrum is recorded. The dye and the three 5' universal bases are removed by a chemical cleaving agent, leaving the remaining five nucleotides bound to the template and a free 5' phosphate group, which allows the ligation of the next probe. This process is repeated several times, determining the read length. Finally, the synthesized strand is removed and new sequencing primers hybridizes to the template, offset by one base and the ligation cycles are repeated. The whole process is repeated in total five times, resulting in dual measurements of each base. Since a detected colour can correspond to four different dinucleotides, the detected colours should be decoded using the two-base colour code (Figure 3.A).

An advantage of SOLiD sequencing is the dual measurements of each base, which results in a low error rate (Table 1). The drawback of SOLiD sequencing are its inability to sequence through palindromic sequences and its relatively short read-length (75 nucleotides) [44].

A. SOLiD dibase probes

B. Sequencing-by-ligation

Decoding sequence using two-base colour code
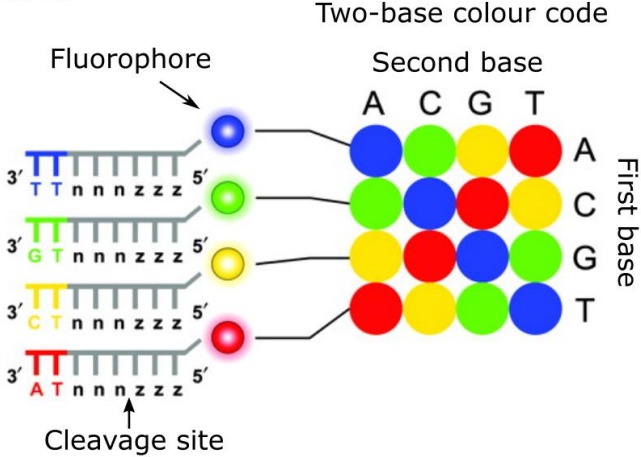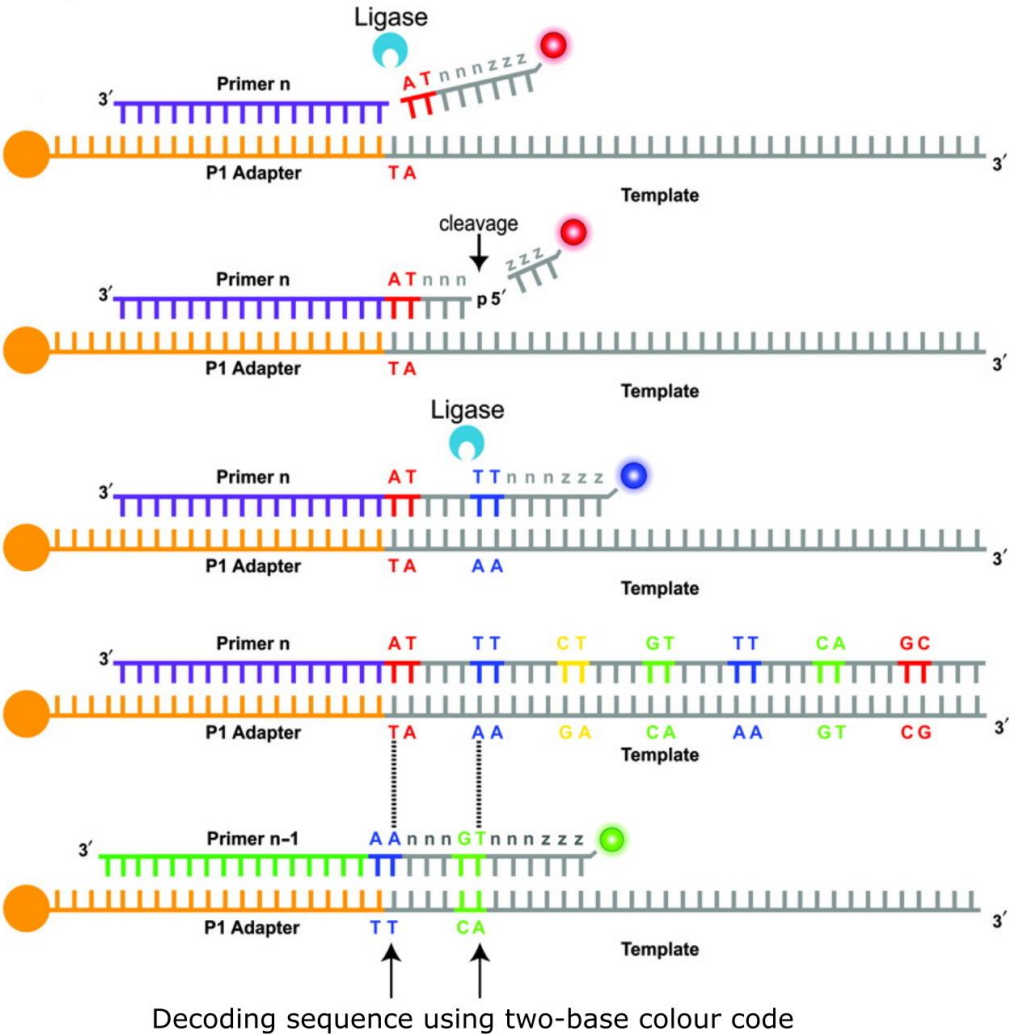
Figure 3 (previous page): SOLiD sequencing. (A) SOLiD dibase probes. Each dibase probe consists of a ligation site, two specific bases (each dinucleotide corresponds to one of the four fluorescent dyes), followed by six universal bases and a fluorescent dye coupled to the last nucleotide. **(B) Sequencing-by-ligation.** After binding of the sequencing primer, a complementary dibase probe hybridizes to the template and is subsequently ligated to the primer. Following detection of the incorporated dibases, the dye and the three 5' nucleotides are removed by a chemical cleaving agents. This process is repeated several times and the number of cycles determines the read length. After removal of the complementary strand, the sequencing reaction is repeated with a sequencing primer offset by one base. Finally, the DNA sequence can be decoded using the two-base colour code. Figure adapted from [45].

3.3.4. Ion Torrent

Ion Torrent (Thermo Scientific) sequencers have been commercially available since 2010. Currently, three different Ion Torrent sequencers are on the market: the Ion S5, the Ion Torrent Personal Genome Machine (PGM) and the Ion Proton, with sequencing output ranging from 30 Mb to 15 Gb and read lengths up to 400 bases (Table 1). The throughput per sequencer varies based on the sequencing chip used.
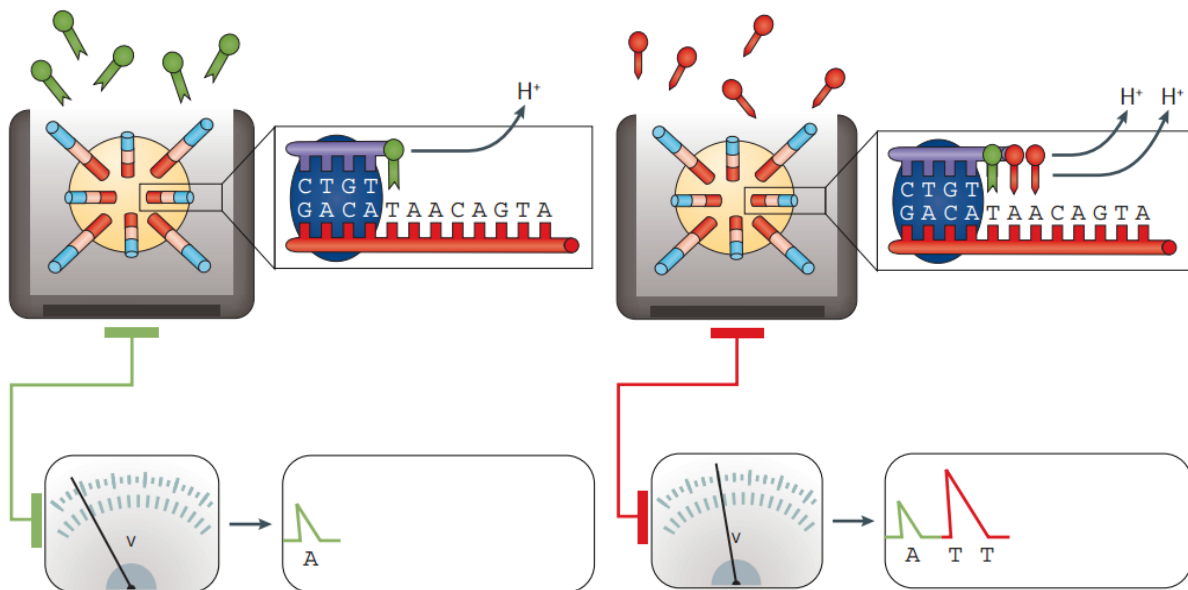
Figure 4: Ion Torrent semiconductor sequencing. Beads carrying clonally amplified DNA fragments are deposited into wells on a semiconductor chip. Each sequencing cycle, one type of nucleotide is flown over the chip (dATPs are marked in green; dTTPs are marked in red). Upon nucleotide incorporation, a proton is released, which results in a change in pH, and an accompanying change in voltage, which is detected by the ion sensor at the bottom of the well. When a homopolymer is sequenced (right panel, in this example two adenosines), a change in voltage is recorded proportional to the number of incorporated bases. Figure adapted from [47].

For Ion Torrent sequencing, the beads carrying the emPCR amplified single-stranded DNA clones, are flown across a semiconductor chip containing millions of wells, resulting in the deposition of one

bead per well. The DNA sequence of each fragment is determined by semi-conductor sequencing, a technique similar to pyrosequencing, but the incorporation of a given nucleotide is detected based on the release of hydrogen ions (Figure 4) [46]. To enable sequencing, the sequencing primers first hybridize to the adapter. Subsequently, the well is flooded with one of the four nucleotides. If the complementary nucleotide is added, the polymerase will incorporate this dNTP, and a hydrogen ion and PPi are released. The hydrogen ion changes the pH of the solution in the well, which is measured by an ion sensitive layer beneath the well as a change in voltage, which is recorded by the semiconductor sensor underneath. The incorporation of identical bases next to each other in the DNA sequence, results in a plural release of protons and the detected change in voltage is proportional to the number of nucleotides incorporated (Figure 4, right panel). The semiconductor wells thus capture chemical information and translate it into digital information. The DNA sequence is determined by adding the different types of nucleotides in a well-known order to the semiconductor chip, while the voltage is continuously monitored.

An important advantage of Ion Torrent sequencing is its shorter run time (two to seven hours) compared to the other sequencing technologies. In addition, native dNTPs and semiconductor chips are used, omitting the need for fluorescent labels, enzyme cascades, lasers and cameras. This will possibly result in a lower cost for Ion Torrent sequencing in the long run. The high error rate in homopolymers is an important disadvantage.

3.4. Third generation sequencing techniques: Single-Molecule-Sequencers

Although the characteristics of the next-generation of automated sequencers are constantly being improved, also newer technologies are being developed: the third generation of sequencing techniques, also called single-molecular sequencers (SMS). These recent techniques have a much higher sensitivity, making direct sequencing of single DNA molecules possible, abolishing the need for clonal amplification of the DNA library. However, the current SMS technologies are still suffering from high error rates [33, 48]. Although many different types of SMS are being developed, only three platforms have been commercially released:

- HeliScope - Helicos BioSciences Corporation: Helicos was launched as the first SMS on the market in 2008. Helicos sequencing used a similar sequencing-by-synthesis technique as Illumina using terminally blocked, fluorescently labelled nucleotides, omitting the need for bridge-amplification [49]. The DNA fragment was hybridized by its adaptor to its complementary sequence which was covalently coupled to the chip. The sequence was determined by adding each time a single fluorescently labelled nucleotide which could be incorporated by the DNA polymerase. Due to bankruptcy of the company in 2011, this sequencer is no longer sold.

- PacBio RS II and the Sequel System - Pacific BioSystems: The PacBio sequencer was commercially launched in 2010 and currently is the most established SMS. Before sequencing, the DNA is circularized using SMRTbell hairpin adaptors. Subsequently, the template-polymerase complex is immobilized at the bottom of a microwell, the so called zero-mode waveguide (ZMW) cell. Differentially fluorescently labelled phospholinked nucleotides are incorporated by the DNA polymerase. During incorporation, the nucleotide is detected by the very sensitive ZMW at the bottom of the microwell. Since the fluorophore is linked to the terminal phosphate, it is released upon incorporation [50]. In addition to DNA sequencing, the PacBio system can also be used to study epigenetic changes based on the kinetic variation measured during base incorporation. The very long read lengths (average > 10 kb, some reaching >60 kb) are an important advantage of the PacBio platforms [51]. The recently released 'Sequel System' has an almost seven times higher output (1,000,000 ZMWs per SMRT cell), then its prefacing PacBio RS II (150,000 ZMWs per SMRT cell).

- MinION - Oxford Nanopore Technologies: The MinION is commercially available since May 2015 and weights only 87g. The small size and weight of the MinION sequencer are a big advantage, since it improves the flexibility of sequencing, making it also possible to sequence outside of the lab. In this technology, single-stranded DNA is passed through a protein nanopore in an electrically resistant polymer membrane. Since an ionic current is constantly passing through the membrane, the current is changed when DNA passes through it. This change in current can be used to identify the nucleotide that passed through the nanopore [52, 53]. The MinION contains up to 512 sequencing channels, which measure 280 bases per second per pore, with a flow cell lifetime of 72 hours [54]. Oxford Nanopore Technologies also optimized the nanopore enzyme to be specific for RNA, enabling direct RNA sequencing [55]. A developers kit for RNA sequencing will be released by the end of 2016. The read length is determined by the fragment length, with the longest reported read length being 230-300 kb [54]. PromethION devices with an increased throughput of up to 144,000 nanopores will soon be available through the PromethION early access programme.

3.5. Successful applications of NGS in virology research

The small genome size of viruses, together with the high output, the possibility to multiplex samples and the sequence independence of the different NGS technologies, has resulted in an exponential increase in available viral sequence data. The implementation of NGS technologies in virology is very diverse and opens new opportunities for viral research. Consequently, NGS is becoming the mainstream technique for more and more virologic applications.

In a disease outbreak, it is important to quickly characterize the infectious agent and its evolutionary rate to determine the appropriate control measures. The importance of NGS in viral diagnosis was demonstrated by an outbreak of hemorrhagic fever in northern Uganda. Patients with suspected

hemorragic fevers tested negative for several suspected hemorrhagic fever viruses using traditional methods (PCR and ELISA) [56]. Eventually, the presence of yellow fever virus was detected by whole RNA sequencing on serum samples using 454 sequencing. Subsequently, a vaccination campaign was started, to limit the viral spread and disease burden [56]. The importance of real-time genomic surveillance using the portable MinION sequencer was demonstrated during the Ebola outbreak in 2015 in West-Africa, where sequence data were generated less than 24 hours after receiving an Ebola-positive sample and were used to monitor the transmission history and viral evolution during the outbreak [57]. Despite the instrument's high error rate, 25-fold coverage of genome positions was sufficient to accurately determine the genotype [57].

The intrinsically high variability of RNA viruses can result in the selection of mutant viruses under antiviral pressure. Consequently, the high sensitivity of NGS plays an important role in the detection of viral resistance. Genetic testing for resistance markers is routinely used in HIV patients to select the appropriate antiretroviral regimen before the start of an initial drug therapy or when the current drug therapy fails, before therapy switch. Selection of an antiviral therapy after genetic testing has shown to improve the disease outcome [58]. However, conventional tests are based on Sanger sequencing, which is unable to detect resistant variants present at low frequency. The high sensitivity of NGS makes early detection of clinically relevant low-frequency variants and the intra-host viral diversity possible [59]. Sentosa announced the commercial release of the first next-generation sequencing test for HIV drug resistance testing, 'the Sentosa SQ HIV-1 genotyping assay', in the near future [60]. The assay is developed for sequencing on the Ion Torrent PGM sequencer. In this assay, a sample is also screened for resistance mutations that are not included in conventional test, increasing the potential for optimization of HIV treatment.

Current licensing and quality control of live attenuated virus vaccines is based on the consensus sequence of the attenuated virus, along with phenotypic characterization. The development of sensitive NGS technologies, creates opportunities to develop new quality control criteria. This will be of utmost importance for live-attenuated vaccines of genetically variable RNA viruses to survey the minor variants present in the vaccine. The importance of the use of NGS for vaccine control has been shown for commercially available modified-live oral rabies vaccines. None of the investigated vaccines were genetically uniform, representing a more or less heterogeneous populations [61]. The utility of NGS for quality control of viral vaccines was also demonstrated for the oral Poliovirus vaccine [62]. Low frequency signature mutations reflected subtle differences in manufacturing conditions between different manufacturers and can be used to monitor the molecular consistency of viral vaccines [62].

The high sequencing output of NGS has revolutionized the field of metagenomics, since a mixture of genetic materials can now be sequenced with high sensitivity. However, the sequence-independent nature of NGS and the rather low abundance of viral sequences, in for *e.g.* clinical samples, required the optimization of virus enrichment protocols [63]. The use of NGS in the study of the human

virome, resulted in better characterization of this viral ecosystem and its function, and the human-virus interactions. Illumina MiSeq sequencing on viruses isolated from human stool of patients with Crohn's disease and ulcerative colitis assigned a role for the virome in intestinal inflammation and bacterial dysbiosis [64]. Both diseases were associated with a significant expansion of Caudovirales bacteriophages in the virome of patients [64]. Consequently, the virome is suggested to be a biomarker for human inflammatory bowel diseases [64]. The use of NGS in metagenomics resulted also in the discovery of several new viruses due to the sequence independent nature of the system. Bas-Congo virus, a novel strain of Rhabdovirus, was detected by performing 454 and Illumina sequencing on serum samples of a cluster of three human cases with acute hemorrhagic fever in Central Africa, which tested negative by TaqMan real-time PCR on all viruses known to cause acute hemorrhagic fever in Africa [65]. This was also the first successful demonstration of *de novo* assembly of a novel, highly divergent viral genome in the absence of a reference sequence [65]. Moreover, the use of 454 sequencing resulted in the discovery of a previously unknown polyomavirus, Merkel cell polyomavirus, as a contributing factor of to Merkel cell carcinoma [66].

3.6. Concluding remark

The different NGS techniques have contributed to significant progress in virology research. Their higher sensitivity and larger output makes that not only more sequencing results can be obtained faster, but opens also several new research opportunities (*e.g.* sequence independent viral discovery and intra-host evolution). In addition, the single-molecule sequencers are even accelerating the sequencing process while avoiding the sequence bias introduced during library preparation, although these platforms still suffer from a high error rate. However, the large amount of sequencing data generated, creates also challenges on how to handle and interpret this enormous amount of information. In addition, information on how sequencing data was processed before sequence analysis was performed, is often lacking or rather limited in literature.

References

1. Min Jou W, Haegeman G, Ysebaert M, Fiers W: Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 1972, 237(5350):82-88.
2. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A *et al*: Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 1976, 260(5551):500-507.
3. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 1977, 74(12):5463-5467.
4. Can we identify potential viral zoonoses before they cross the species barrier? [http://www.microbiologysociety.org/all-microsite-sections/new-mt/index.cfm/article/7D58FF51-48A8-475B-819FB6F5B9BD7313]
5. Accession list of all viral genomes [http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&cmd=download2]
6. Watson JD, Crick FH: Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953, 171(4356):737-738.
7. Franklin RE, Gosling RG: Molecular configuration in sodium thymonucleate. *Nature* 1953, 171(4356):740-741.
8. Kolata GB: The 1980 Nobel Prize in Chemistry. *Science* 1980, 210(4472):887-889.
9. Temin HM, Mizutani S: RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 1970, 226(5252):1211-1213.
10. Baltimore D: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970, 226(5252):1209-1211.
11. Maxam AM, Gilbert W: A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 1977, 74(2):560-564.
12. Porter AG, Barber C, Carey NH, Hallewell RA, Threlfall G, Emtage JS: Complete nucleotide sequence of an influenza virus haemagglutinin gene from cloned DNA. *Nature* 1979, 282(5738):471-477.
13. Fang R, Min Jou W, Huylebroeck D, Devos R, Fiers W: Complete structure of A/duck/Ukraine/63 influenza hemagglutinin gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza hemagglutinin. *Cell* 1981, 25(2):315-323.
14. Sanger F, Coulson AR: A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* 1975, 94(3):441-448.
15. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, Fiddes JC, Hutchison CA, 3rd, Slocombe PM, Smith M: The nucleotide sequence of bacteriophage phiX174. *Journal of molecular biology* 1978, 125(2):225-246.
16. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE: Fluorescence detection in automated DNA sequence analysis. *Nature* 1986, 321(6071):674-679.
17. Applied Biosystems: Celebrating 25 Years of Advancing Science [http://home.appliedbiosystems.com/about/presskit/pdfs/celebrating_25_years_aln_article.pdf]
18. Sears LE, Moran LS, Kissinger C, Creasey T, Perry-O'Keefe H, Roskey M, Sutherland E, Slatko BE: CircumVent thermal cycle sequencing and alternative manual and automated DNA sequencing protocols using the highly thermostable VentR (exo-) DNA polymerase. *BioTechniques* 1992, 13(4):626-633.
19. 3730xl DNA Analyzer [http://www.thermofisher.com/order/catalog/product/3730XL]
20. Diguistini S, Liao NY, Platt D, Robertson G, Seidel M, Chan SK, Docking TR, Birol I, Holt RA, Hirst M *et al*: De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome biology* 2009, 10(9):R94.
21. Keller A, Harz C, Matzas M, Meder B, Katus HA, Ludwig N, Fischer U, Meese E: Identification of novel SNPs in glioblastoma using targeted resequencing. *PloS one* 2011, 6(6):e18158.
22. Bennett ST, Barnes C, Cox A, Davies L, Brown C: Toward the 1,000 dollars human genome. *Pharmacogenomics* 2005, 6(4):373-382.
23. Glenn TC: Field guide to next-generation DNA sequencers. *Molecular ecology resources* 2011, 11(5):759-769.
24. 2014 NGS Field Guide – Table 3c – Error rates [http://www.molecularecologist.com/next-gen-table-3c-2014/]

25.   Tawfik DS, Griffiths AD: Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998, 16(7):652-656.

26.   Ghadessy FJ, Ong JL, Holliger P: Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98(8):4552-4557.

27.   Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B: Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America* 2003, 100(15):8817-8822.

28.   Nyren P: Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal Biochem* 1987, 167(2):235-238.

29.   Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P: Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 1996, 242(1):84-89.

30.   Ronaghi M, Uhlen M, Nyren P: A sequencing method based on real-time pyrophosphate. *Science* 1998, 281(5375):363, 365.

31.   Hyman ED: A new method of sequencing DNA. *Anal Biochem* 1988, 174(2):423-436.

32.   Harrington CT, Lin EI, Olson MT, Eshleman JR: Fundamentals of pyrosequencing. *Archives of pathology & laboratory medicine* 2013, 137(9):1296-1303.

33.   2016 NGS Field Guide: Overview [http://www.molecularecologist.com/next-gen-fieldguide-2016/]

34.   Mardis ER: The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG* 2008, 24(3):133-141.

35.   454 sequencing - The Technology [http://454.com/products/technology.asp]

36.   Illumina extends Genomics portfolio [http://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2128277]

37.   Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR *et al*: Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008, 456(7218):53-59.

38.   Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, Sano Marma M, Meng Q, Cao H, Li X *et al*: Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105(27):9145-9150.

39.   Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR *et al*: Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008, 456(7218):53-59.

40.   Intro to Sequencing by Synthesis: Industry-leading Data Quality [https://youtu.be/HMyCqWhwB8E]

41.   SBS Sequencing [http://www.illumina.com/company/news-center/multimedia-images.html]

42.   Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K *et al*: A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res* 2008, 18(7):1051-1063.

43.   McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu YT, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC *et al*: Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 2009, 19(9):1527-1541.

44.   Huang YF, Chen SC, Chiang YS, Chen TH, Chiu KP: Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC systems biology* 2012, 6 Suppl 2:S10.

45.   Voelkerding KV, Dames SA, Durtschi JD: Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry* 2009, 55(4):641-658.

46.   Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M *et al*: An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011, 475(7356):348-352.

47.   Goodwin S, McPherson JD, McCombie WR: Coming of age: ten years of next-generation sequencing technologies. *Nature reviews Genetics* 2016, 17(6):333-351.

48.   Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M: Improved data analysis for the MinION nanopore sequencer. *Nature methods* 2015, 12(4):351-356.

49.   Pushkarev D, Neff NF, Quake SR: Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 2009, 27(9):847-U101.

74

50. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 2009, 323(5910):133-138.
51. SMRT sequencing: read lengths [http://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/]
52. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H: Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology* 2009, 4(4):265-270.
53. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH: Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* 2012, 30(4):349-353.
54. Specifications MinION [https://www2.nanoporetech.com/products/specifications]
55. Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M, Lloyd J, Warland A, Pantic N, Admassu T, Ciccone J *et al*: Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* 2016.
56. McMullan LK, Frace M, Sammons SA, Shoemaker T, Balinandi S, Wamala JF, Lutwama JJ, Downing RG, Stroeher U, MacNeil A *et al*: Using next generation sequencing to identify yellow fever virus in Uganda. *Virology* 2012, 422(1):1-5.
57. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A *et al*: Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016, 530(7589):228-232.
58. Tural C, Ruiz L, Holtzer C, Schapiro J, Viciana P, Gonzalez J, Domingo P, Boucher C, Rey-Joly C, Clotet B: Clinical utility of HIV-1 genotyping and expert advice: the Havana trial. *AIDS* 2002, 16(2):209-218.
59. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS *et al*: Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *The Journal of infectious diseases* 2009, 199(5):693-701.
60. First next-generation sequencing test for HIV drug resistance could help combat AIDS worldwide [www.sciencedaily.com/releases/2016/08/160803104045.htm]
61. Hoper D, Freuling CM, Muller T, Hanke D, von Messling V, Duchow K, Beer M, Mettenleiter TC: High definition viral vaccine strain identity and stability testing using full-genome population data--The next generation of vaccine quality control. *Vaccine* 2015, 33(43):5829-5837.
62. Neverov A, Chumakov K: Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(46):20063-20068.
63. Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, Moore NE, Ren X, Huang QS, Carter PE *et al*: Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of virological methods* 2014, 195:194-204.
64. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P *et al*: Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015, 160(3):447-460.
65. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe JJ, Sittler T, Veeraraghavan N, Ruby JG, Wang C *et al*: A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS pathogens* 2012, 8(9):e1002924.
66. Feng H, Shuda M, Chang Y, Moore PS: Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008, 319(5866):1096-1100.

Chapter 4

**Successful applications of NGS
to study influenza viruses**

4.1. Introduction

The development of next-generation sequencing (NGS) methods has led to significant progress in all branches of life sciences, including the field of influenza biology. The high sequencing throughput of NGS, along with the relatively small genome size of influenza viruses, makes it possible to sequence multiple viral samples in parallel and with a very high coverage per nucleotide (so called sequencing depth). This high capacity and sensitivity of NGS compared to earlier sequencing methods has contributed to major advances in several domains of influenza research. Here we will highlight some of the successful applications of NGS to study influenza viruses.

4.2. Virus discovery

In contrast to Sanger sequencing, NGS sequencing does not require prior knowledge of the target sequence, which increases the speed and sensitivity for viral discovery. Recently, RT-PCR using random primers on total nucleic acids extracted from rectal swabs, followed by Illumina GAIIx or 454 GS-FLX sequencing was a key experimental step in the discovery of H17N10 and H18N11 influenza viruses in New World bats [1, 2]. In the genome segments coding for PB2, PB1, PA and NA, more genetic diversity was detected than present in all known mammalian and avian influenza virus species combined, suggesting that H17N10 and H18N11 have a long-standing association with their host [2]. Since bats represent approximately twenty percent of all mammalian species, it will be important to characterize these viruses and assess their potential risk for the human population.

The use of NGS also resulted in the detection of a suggested fourth influenza genus: the influenza D viruses. This new influenza genus was detected in pigs exhibiting influenza-like illness, after random RT-PCR amplification of RNA retrieved from virus present in nasal swabs amplified on swine testicle cells, followed by Ion Torrent PGM sequencing [3]. The *de novo* assembled, seven genome segments turned out to be only 50% identical at the protein level to human influenza C viruses and serum antibodies exhibited no cross-reactivity [3]. Sequencing the viral transcriptome upon infection of swine testicle cells, resulted in identification of spliced mRNA for both the NS and M segment [4]. While splicing of the NS segment occurs in the same way as for influenza C viruses, splicing of the M segment of influenza D viruses occurs differently [4]. Splicing of the colinear mRNA of segment 6 creates a stop codon at the splice junction for influenza C viruses, whereas four extra codons are added to the M1-ORF upon splicing of this mRNA in the case of influenza D viruses [4, 5].

The merit of NGS in viral discovery has also nicely been demonstrated by the determination of the genome sequence of the Spanish influenza H1N1 strain responsible for the 1918 pandemic. It took nine years to obtain the full genome sequence of this virus using traditional overlapping RT-PCR followed by Sanger sequencing [6]. In contrast, the full-genome sequence could be obtained in a single RNA sequencing run on a formalin-fixed and paraffin-embedded lung sample [7].

4.3. Viral surveillance

Viral surveillance is important to monitor emerging influenza viruses and viral evolution of the circulating influenza strains. The high throughput of the NGS platforms combined with their high sensitivity has refined the genetic analysis of influenza viruses because many more sequences are determined than was the case a decade ago. Consequently, NGS is used more and more in current reports on influenza genome evolution for virus surveillance in their human, swine and avian hosts [8-17]. The sequence information obtained during viral surveillance can also be used to guide vaccine design. This has been illustrated by Stucker *et al.*, who deep sequenced 154 nasopharyngeal swabs, which tested positive for the presence of H3N2 viruses [18]. The sequencing data showed the presence of co-circulating H3N2 clades and antigenic drift variants, which contained HA substitutions and alterations in the potential N-linked glycosylation sites of HA, during the 2012-2013 influenza epidemic. Such information could be used to guide the vaccine selection process for the next influenza season. In addition, this study also showed a role for intrasubtypic reassortment in the evolution of seasonal influenza viruses [18].

The ease and utility of performing whole-genome sequence analysis in viral surveillance, was demonstrated by the study of Zaraket *et al.*, where 100 influenza A H3N2 isolates collected during the 2012-2015 epidemics in four Asian countries were analysed by NGS [19]. Analysis of the phylogenetic tree of each viral genome segment revealed several viral reassortment events and the co-circulation of multiple viral clades within the same influenza season [19]. Moreover, the presence of a common ancestral PB1 gene in singleton reassortants suggests a fitness advantage for this gene, allowing it to persist into the following season [19]. In addition, the phylogenetic three of HA revealed temporal and geographical clustering of the samples, *e.g.* it appears that the H3N2 viruses circulating in Myanmar in 2013 later gave rise to the 2013-2014 epidemic in the northern hemisphere [19]. Furthermore, this study also showed that resistance to amantadine, due to the presence of the M2-S31N (most frequently) or -S31D mutation, was present in the majority of all samples [19]. Nevertheless, no mutations conferring resistance to neuraminidase inhibitors were detected [19]. Interestingly, five of the virus isolates contained novel polymorphisms in the PB1-F2 coding sequence leading to an early stop codon and shortening of the protein with 24 to 34 amino acids [19].

Viral surveillance can also be used to determine the likely origin of an outbreak of influenza infections or the relatedness between samples. This was demonstrated during an outbreak of highly-pathogenic H5N8 influenza A viruses in Dutch poultry farms in 2014 [20]. Whole genome sequencing on viral RNA isolated from cloacal and oropharyngeal samples of infected hens, suggested that four out of the five outbreaks were not the result of farm-to-farm spread, but of separate introduction of the virus in the different farms, which were located 16 to 112 km from each other [20]. The results of this study therefore support the conclusion that migratory birds are spreaders of avian influenza into

and possible between poultry farms and the importance of hygienic measures in these farms to keep potential pathogens out.

4.4. Antigenic drift

The humoral immune response elicited upon influenza infection or vaccination selects for mutations in the antigenic sites of HA, a process called antigenic drift. A study performed by Cushing *et al.*, demonstrated that antigenic drift is already present early in infection [21]. NGS on HA-specific RT-PCR products of virus isolated from human nasopharyngeal swabs, identified the emergence of genetic diversity in HA upon infection [21]. In addition, the intra-host evolution of HA showed unique mutations specific to an individual, with temporal genetic variation during infection [21]. The generation of HA mutants with antigenic variation, in both vaccinated and non-vaccinated individuals, was also demonstrated by Dinis *et al.* [22]. The group of Bloom followed a synthetic DNA approach to try to understand the inherent tolerance of influenza HA for mutations at its antigenic sites [23]. A mutant HA library was made in which mutations at practically every position within the HA protein were introduced. This library, theoretically consisting of 40.000 different HA sequences, was used to generate a library of recombinant A/WSN/1933(H1N1) viruses [23]. Passaging this virus library on cells was done to select for those virus clones that could replicate, leading to the positive selection of tolerated mutations within HA [23]. NGS sequencing of the passaged virus population revealed a strong purifying selection against stop-codon and many non-synonymous mutations. However, the HA antigenic sites appeared to be highly tolerant for mutations compared to the rest of the HA sequence, a feature which contributes to its fast antigenic evolution [23].

4.5. Viral diagnosis

Accurate and fast diagnosis of a newly emerging influenza strain, will accelerate the start, and thus the effectiveness, of protective measures. Influenza virus detection tests based on nucleic acid, *e.g.* PCR and microarray, are more and more preferred above traditional culture- or antigen based diagnostic procedures [24]. However, these tests are sequence dependent, which limits their use in the detection of newly emerging influenza viruses or of viruses which are mutated in the binding site of the PCR primer or micro-array probe. These limitations of traditional viral testing were demonstrated at the start of the influenza A (H1N1) pandemic in 2009, when no clinical or laboratory tests were initially available to identify this virus with high sensitivity and specificity. Whole RNA sequencing, on the other hand, of a nasopharyngeal swab sample using the Illumina GA IIx, resulted in *de novo* assembly of the genome of the 2009 pandemic influenza H1N1 virus, without the requirement of a reference sequence [25]. The suitability of NGS to accurately identify the virus subtype during an influenza outbreak was also demonstrated by Seong *et al.*, who applied RNA sequencing of nasopharyngeal swabs samples to investigate a nosocomial influenza outbreak [26]. Furthermore, direct NGS on virus in a clinical sample has an added advantage compared to Sanger

sequencing of amplified virus, as there is no selection for culture-induced mutations. This was demonstrated by Ren *et al.*, who showed that the full-length genome sequence of an H7N9 virus obtained after NGS on RNA directly isolated from a sputum specimen of a patient differed prominently from the Sanger sequencing results on virus amplified on eggs [27]. The latter virus carried mutations that were introduced during its *in vitro* propagation.

4.6. Viral host adaptation and transmission

Animal reservoirs of influenza viruses have been important contributors to the last four human influenza pandemics. Consequently, it is important to identify circulating viruses with an increased potential to infect humans and to understand how these viruses gain sustained human-to-human transmission. Novel mutations in the polymerase complex of a H5N1 virus that confer increased replication in mammalian cells and increased virulence in mice were detected based on a high throughput screen using reporter gene expression and randomly mutated polymerase genes, which was followed by Ion Torrent PGM sequencing to determine the mutations in the polymerases and NP responsible for increased replication [28].

Infection with highly pathogenic avian influenza H5N1 viruses can cause severe morbidity and mortality in humans, but these viruses do not efficiently transmit between humans. It is of interest for pandemic preparedness plans, to be able to forecast, at least with some confidence, what genetic changes would be required for these viruses to evolve into human-to-human transmissible pathogens. The potential of A/Indonesia/5/2005(H5N1) to acquire sustained human-to-human transmission, was investigated by performing serial passages of this virus in ferrets [29]. Four amino acid substitutions in HA and one in PB2 were consistently present in the airborne-transmitted viruses and resulted in increased receptor binding and enhanced replication in mammalian cells [29]. These mutations were already detected by the use of 454 sequencing after one or two passages of the avian influenza virus in ferret, and they became dominant starting from passage seven [30]. Similar results were obtained in an independently performed ferret-to-ferret transmission study in which an H5N1 reassortant virus, comprising H5 HA (from an H5N1 virus) and the other seven genomic segments from a prototype 2009 pandemic H1N1 virus was used as the starting inoculum [31]. Although the HA diversity increased during infection of the index animals, transmission of H5N1 viruses via respiratory droplets imposed a severe bottleneck and resulted in limited HA diversity in contact animals [32]. Interestingly, this study revealed that minor virus variants in index animals can become the dominant genotype in the contact animal, an insight that was revealed thanks to the high sensitivity of NGS, and likely would have been missed or much more difficult to find by traditional Sanger sequencing [32].

The impact of influenza A virus transmission on the genetic diversity present in a virus population has also been investigated using a set of 100 different recombinant A/California/04/2009(H1N1) viruses, each carrying a unique, neutral barcode in their viral NS segment which makes it possible to track the

viral diversity during transmission [33]. A viral inoculum containing equivalent levels of each barcoded virus was used to infect donor guinea pigs or ferrets. Although these viruses replicated as a highly diverse virus population in the inoculated animals, viral diversity was very limited in the recipient animal [33]. The difference between the barcode profiles in the different recipient animals originating from a single index animal, suggest that the observed bottleneck occurs at the level of the recipient, independent of virus genetics [33]. In addition, this study demonstrated that the strength of the bottleneck is influenced by the infection route, with a higher selection imposed by aerosol transmission than direct contact [33].

The 1918 Spanish influenza pandemic had a devastating effect on the human population at that time whereas the subsequent influenza pandemics claimed significantly less lives. There are concerns that a pandemic virus with a similar pathogenicity as the 1918 pandemic virus may emerge again in the future. Watanabe *et al.* investigated the chance for a virus similar to the 1918 pandemic virus to emerge from wild bird reservoirs, and their experimental set-up is schematically represented in figure 1 [34]. They first searched the publically available avian influenza virus sequence repository for sequences that are closely related to the 1918 pandemic virus [34]. Subsequently, an influenza virus was created by reverse genetics (named '1918-like avian virus') that contained avian viral genome segments with high homology to the 1918 virus, presumably to lend support to the hypothesis that such a constellation of naturally occurring influenza A virus genes, could in the future blend together by reassortment in nature [34]. However, this virus replicated poorly in index ferrets and did not result in virus transmission. Consequently, to enhance the chance to select for virus transmission, three described mammal-adapting substitutions were introduced in the 1918-like avian virus genome: E627K in PB2 and E190D together with G225D in HA (H3 numbering) [34]. Subsequently, this man-made virus (named '1918-like avian HA190D225D virus') was passaged into ferrets to select for a respiratory droplet transmissible virus. The virus strains that emerged in contact animals during ferret adaptation, suggest the association of three mutations in the polymerases (E627K and A684D in PB2 and V253M in PA), six mutations in HA (E89D, S113N, I187T, E190D, G225D and D265V) and a single mutation in NP (T232I), with efficient transmission of a potential pandemic 1918-like avian influenza strain [34].

In a follow up report NGS analysis was used to try to elucidate which evolutionary pathway the 1918-like avian influenza virus had followed to become transmissible between ferrets through the air [35]. This whole genome NGS analysis, performed on virus that was retrieved from upper respiratory tract of the ferret sampled at each transmission step, showed that the early stages of ferret adaptation of this avian virus are marked by diversification of HA during its replication in the index animals [35]. Transmission on the other hand first involved a loose genetic bottleneck, with maintenance of the HA diversity and the fixation of two mutations in the polymerases (V253M in PA and A684D in PB2), that did not confer a detectable replication advantage (Figure 1) [35]. These viruses transmitted better and gave rise to a new virus population that seemed to have undergone a strong genetic bottleneck, with fixation of the I187T and D265V or S113N mutations in HA1 [35]. This study shows that the

stringency of the transmission bottleneck can change during host adaptation [35]. In addition, the occurrence of mutations in multiple combinations in the transmitted viruses suggests that an 1918-like influenza virus can evolve to transmission via multiple genetic pathways [35].
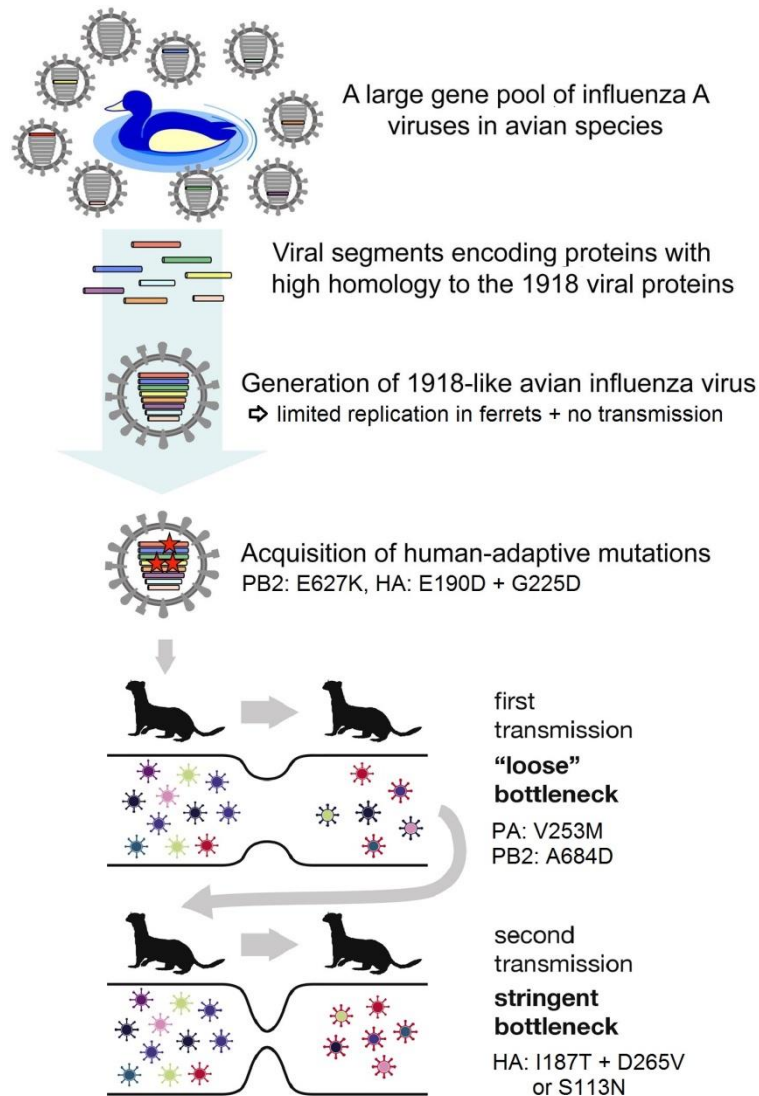
Figure 1: Avian influenza viruses with sequences similar to the 1918 pandemic influenza A virus may have pandemic potential. The groups of Kawaoka and Friedrich investigated the potential that a transmissible virus with a similar pathogenicity as the 1918 pandemic virus may emerge from the wild bird reservoir. Their experimental procedure is schematically represented. Figure adapted from [34] and [35].

The principal measure to limit adaptation of an avian influenza virus towards replication in humans, is by restricting exposure to the avian influenza reservoir. The importance of reducing human exposure to avian influenza viruses has been demonstrated during an H7N7 outbreak in the Netherlands in 2003, were a veterinarian got infected in a direct manner after visiting a virus-struck chicken farm and died subsequently [36]. Within the clinical samples isolated from the veterinarian 11 days post exposure and one day after hospitalization, the human adaptation mutation PB2 E627K

was present at 75% in the BAL and 93% in the sputum samples [36]. Moreover, the mutation was homogenously present in the lung tissue obtained during autopsy five days later. Importantly, this mutation could not be detected by NGS in the H7N7 positive chicken samples obtained on the day the veterinarian visited the farm, which suggests that the PB2 E627K mutation emerged in the infected veterinarian [36].

The above mentioned studies show that a small number of amino acid substitutions in the circulating avian influenza viruses is sufficient to enhance viral replication and transmission in humans, and as such, create a virus with pandemic potential. The possibility of minor variants to become dominant upon transmission, demonstrates the importance of using NGS for viral surveillance.

4.7. Vaccine control

Current recommendations to assure the quality, safety and efficacy of influenza vaccines require reporting of the consensus sequence of the virus population used for vaccine production [37]. However, due to the intrinsic genetic heterogeneity of influenza viruses, NGS is more suitable to survey the genetic composition of influenza vaccines. Laasri *et al.* demonstrated that deep sequencing can be used to monitor the consistency of influenza vaccines [38]. Illumina sequencing of the A/California/07/2009(H1N1) vaccine strains generated either through reverse genetics or traditional reassortment, resulted in effective and reproducible detection of low quantities of mutants in the entire genome of influenza viruses, including variations in antigenic sites of HA [38]. Remarkably, the heterogeneity in the virus stock derived by reverse genetics was higher than in the virus stocks created by conventional reassortment [38]. Due to the error-prone replication of influenza viruses, it is thus important to accurately determine the viral variants present in the vaccine strains, since vaccine antigenicity may be influenced by mutations in the antigenic sites of HA and NA. In addition, the high sensitivity of NGS is also ideally suited to monitor the genetic stability and safety of live-attenuated influenza vaccines.

4.8. Viral resistance and escape

The high variability of the influenza viruses, requires a sensitive method for viral resistance testing. Direct Sanger sequencing of amplified viral RT-PCR products has a limited sensitivity of 20 to 30 % and is as such unable to detect minor drug resistant virus populations. The detection of minor variants using NGS sequencing makes is possible to study the evolution of drug-resistant mutations, both intra-host and on a large population scale, and to predict the speed at which resistance will arise [39]. In addition, whole genome analysis by NGS enables analysis of all variants in the viral genome. Consequently, also variants that would likely remain undetected by traditional screening for resistance can be analyzed. NGS has proven to be successful in detecting and identifying antiviral

resistance in *e.g.* seasonal influenza A H1N1 and H3N2 virus and the pandemic 2009 H1N1 viruses [40-43].

In vitro antiviral selection pressure can also help in the understanding of the evolution of viral resistance. *In vitro* selection for oseltamivir resistance, followed by full-genome sequencing on an Illumina HiSeq 2000 platform, revealed selection of H274Y in NA and that this mutation alone is required for oseltamivir resistance [44]. In addition, the observed rise in oseltamivir resistance in seasonal H1N1 influenza A viruses can be explained by the fact that this mutation is neutral for the virus, since further passaging of this resistant virus in the absence of oseltamivir resulted in maintenance of the mutation [44]. The fitness of influenza B viruses resistant to neuraminidase inhibitors has also been determined using recombinant mutant influenza B/Yamanashi/166/1998 viruses containing single amino acid mutations responsible for resistance to oseltamivir or zanamivir [45]. Cells were co-infected with an equal amount of wild type virus and one of these mutants, followed by NGS analysis [45]. Interestingly, the H274Y mutation resulted in a fitness advantage over wild type virus, but these viruses retained sensitivity to zanamivir [45]. In addition, the NA-E119A mutant virus could replicate to high viral titers in the presence of both oseltamivir or zanamivir, however it is less fit compared to wild type virus [45].

NGS can also be used to analyze viral escape to antibody pressure. NGS has been applied to study the viral resistance after *in vitro* viral escape selection of the influenza viruses A/California/7/2009(H1N1) and A/Perth/16/2009(H3N2) for the broadly neutralizing pan-HA stalk-binding human mAb 39.29 [46, 47]. Eight rounds of virus passage in the presence of mAb 39.29 did not result in the selection for resistant A/California/7/2009(H1N1) viruses. In contrast, passaging of A/Perth/16/2009(H3N2) in the presence of mAb 39.29 resulted in the selection of three mutant viruses that had acquired resistance to the neutralizing activity of this mAb. These viruses each carried one mutation in the HA stalk that resulted in viral escape: either Gln387Lys, Asp391Tyr or Asp391Gly [48]. These mutations rendered the viruses completely resistant to the mAb by two different mechanisms: either by abolishing antibody binding (mutation Gln387Lys) or enhancing the fusion ability of HA (mutations Asp391Tyr and Asp391Gly) [48]. NGS has also been used to investigate the escape routes of a highly-pathogenic avian H5N1 influenza A virus to vaccine pressure, by *in vitro* passaging of the virus in the presence of immune serum of chickens that had been vaccinated with a commercial inactivated, clade-matching H5N2 vaccine only or followed by viral challenge with A/cygnus cygnus/Germany/R65/2006 (H5N1) [47]. Although no complete escape to the immune serum could be achieved even after 100 passages, several mutations scattered throughout the viral genome were spotted that contributed to partial resistance of the immunity provided by the H5N2 vaccine. These mutations mapped to regions that have been described to be antigenically active, *i.e.* HA and PA [47]. Interestingly, some of the *in vitro* selected escape mutations have been detected in natural isolates, demonstrating the *in vivo* relevance of *in vitro* escape selection and its potential to be used to predict the antigenic changes present in circulating influenza viruses [47]. Such studies not only help in screening for antiviral-

resistant viruses, but also help in understanding the mechanisms behind escape and the biology of influenza viruses as a whole.

4.9. Influenza virus-host biology

The high throughput of NGS platforms makes it possible to perform RNA sequencing of the transcriptome, which can result in new insights in influenza virus biology. As an example, an important role for I64 in the anti-inflammatory functions of NS1 has been demonstrated by comparing alterations in the transcriptome in MDCK cells upon infection with different seasonal human H3N2 influenza virus isolates [49]. Remarkably, upregulation of genes associated with the innate immune response, *e.g.* genes coding for IFN-β, RIG-1 and TNF, was most pronounced after infection of MDCK cells with just one of the twelve tested clinical isolates [49]. The observed defect in suppressing the innate antiviral response could be linked to a I64T mutation in the dsRNA-binding domain of NS1, which attenuates the virus [49]. This mutation decreases its interaction with cleavage and polyadenylation specificity factor 30 (CPSF30), resulting in decreased inhibition of host protein synthesis [49]. Interestingly, the virus with the I64T mutation in NS1 was isolated from a subject with a defect in induction of IFN responses [49].

4.10. Concluding remark

The diversity of applications where NGS contributed to progress in the influenza research field were highlighted here. It is clear that their higher sequencing output and sensitivity, together with the sequence independence of sequencing, are important advantages compared to traditional techniques. However, large data sets of millions to billions of sequencing reads are nowadays easily generated during a single sequencing run, which require skilled people to convert this huge amount of information to results with biological relevance. In addition, NGS continues to evolve. One development that may breakthrough in the future is the possibility to directly sequence RNA, which is being developed by Oxford Nanopore and is characterized by long read lengths and a fast turnaround time [50]. This method, once mature, will circumvent the sequence bias and errors that are introduced during the preparation of the sequencing library for RNA viruses such as influenza, leading to even more sensitive surveillance options.

References

1. Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA *et al*: A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(11):4269-4274.

2. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J *et al*: New world bats harbor diverse influenza A viruses. *PLoS pathogens* 2013, 9(10):e1003657.

3. Hause BM, Ducatez M, Collin EA, Ran Z, Liu R, Sheng Z, Armien A, Kaplan B, Chakravarty S, Hoppe AD *et al*: Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS pathogens* 2013, 9(2):e1003176.

4. Hause BM, Collin EA, Liu R, Huang B, Sheng Z, Lu W, Wang D, Nelson EA, Li F: Characterization of a novel influenza virus in cattle and Swine: proposal for a new genus in the Orthomyxoviridae family. *mBio* 2014, 5(2):e00031-00014.

5. Yamashita M, Krystal M, Palese P: Evidence That the Matrix Protein of Influenza C-Virus Is Coded for by a Spliced Messenger-Rna. *Journal of virology* 1988, 62(9):3348-3355.

6. Taubenberger JK, Hultin JV, Morens DM: Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antiviral therapy* 2007, 12(4 Pt B):581-591.

7. Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK: High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *The Journal of pathology* 2013, 229(4):535-545.

8. Van Borm S, Rosseel T, Marche S, Steensels M, Vangeluwe D, Linden A, van den Berg T, Lambrecht B: Complete Coding Sequences of One H9 and Three H7 Low-Pathogenic Influenza Viruses Circulating in Wild Birds in Belgium, 2009 to 2012. *Genome announcements* 2016, 4(3).

9. Farooqui A, Leon AJ, Huang L, Wu S, Cai Y, Lin P, Chen W, Fang X, Zeng T, Liu Y *et al*: Genetic diversity of the 2013-14 human isolates of influenza H7N9 in China. *BMC infectious diseases* 2015, 15:109.

10. Hurt AC, Su YC, Aban M, Peck H, Lau H, Baas C, Deng YM, Spirason N, Ellstrom P, Hernandez J *et al*: Evidence for the introduction, reassortment and persistence of diverse influenza A viruses in Antarctica. *Journal of virology* 2016.

11. Fang S, Bai T, Yang L, Wang X, Peng B, Liu H, Geng Y, Zhang R, Ma H, Zhu W *et al*: Sustained live poultry market surveillance contributes to early warnings for human infection with avian influenza viruses. *Emerging microbes & infections* 2016, 5(8):e79.

12. Fang S, Wang X, Dong F, Jin T, Liu G, Lu X, Peng B, Wu W, Liu H, Kong D *et al*: Genomic characterization of influenza A (H7N9) viruses isolated in Shenzhen, Southern China, during the second epidemic wave. *Archives of virology* 2016, 161(8):2117-2132.

13. Tassoni L, Fusaro A, Milani A, Lemey P, Awuni JA, Sedor VB, Dogbey O, Commey AN, Meseko C, Joannis T *et al*: Genetically Different Highly Pathogenic Avian Influenza A(H5N1) Viruses in West Africa, 2015. *Emerging infectious diseases* 2016, 22(12).

14. Nelson MI, Wentworth DE, Culhane MR, Vincent AL, Viboud C, LaPointe MP, Lin X, Holmes EC, Detmer SE: Introductions and evolution of human-origin seasonal influenza a viruses in multinational swine populations. *Journal of virology* 2014, 88(17):10110-10119.

15. Kitikoon P, Vincent AL, Gauger PC, Schlink SN, Bayles DO, Gramer MR, Darnell D, Webby RJ, Lager KM, Swenson SL *et al*: Pathogenicity and transmission in pigs of the novel A(H3N2)v influenza virus isolated from humans and characterization of swine H3N2 viruses isolated in 2010-2011. *Journal of virology* 2012, 86(12):6804-6814.

16. USDA Swine influenza virus surveilance system: computational evolutionary biology of whole genome sequences [http://portal.nifa.usda.gov/web/crisprojectpages/0423118-usda-swine-influenza-virus-surveillance-system-computational-evolutionary-biology-of-whole-genome-sequences.html]

17. Bahl J, Krauss S, Kuhnert D, Fourment M, Raven G, Pryor SP, Niles LJ, Danner A, Walker D, Mendenhall IH *et al*: Influenza a virus migration and persistence in North American wild birds. *PLoS pathogens* 2013, 9(8):e1003570.

18. Stucker KM, Schobel SA, Olsen RJ, Hodges HL, Lin X, Halpin RA, Fedorova N, Stockwell TB, Tovchigrechko A, Das SR *et al*: Haemagglutinin mutations and glycosylation changes shaped the 2012/13 influenza A(H3N2) epidemic, Houston, Texas. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2015, 20(18).

19. Zaraket H, Kondo H, Hibino A, Yagami R, Odagiri T, Takemae N, Tsunekuni R, Saito T, Myint YY, Kyaw Y *et al*: Full Genome Characterization of Human Influenza A/H3N2 Isolates from Asian Countries Reveals a Rare Amantadine Resistance-Conferring Mutation and Novel PB1-F2 Polymorphisms. *Frontiers in microbiology* 2016, 7:262.

20. Bouwstra RJ, Koch G, Heutink R, Harders F, van der Spek A, Elbers AR, Bossers A: Phylogenetic analysis of highly pathogenic avian influenza A(H5N8) virus outbreak strains provides evidence for four separate introductions and one between-poultry farm transmission in the Netherlands, November 2014. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2015, 20(26).

21. Cushing A, Kamali A, Winters M, Hopmans ES, Bell JM, Grimes SM, Xia LC, Zhang NR, Moss RB, Holodniy M *et al*: Emergence of Hemagglutinin Mutations During the Course of Influenza Infection. *Scientific reports* 2015, 5:16178.

22. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, Meece JK, Belongia EA, Friedrich TC: Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *Journal of virology* 2016, 90(7):3355-3365.

23. Thyagarajan B, Bloom JD: The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 2014, 3.

24. Kumar S, Henrickson KJ: Update on influenza diagnostics: lessons from the novel H1N1 influenza A pandemic. *Clinical microbiology reviews* 2012, 25(2):344-361.

25. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T *et al*: A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PloS one* 2010, 5(10):e13381.

26. Seong MW, Cho SI, Park H, Seo SH, Lee SJ, Kim EC, Park SS: Genotyping Influenza Virus by Next-Generation Deep Sequencing in Clinical Specimens. *Annals of laboratory medicine* 2016, 36(3):255-258.

27. Ren X, Yang F, Hu Y, Zhang T, Liu L, Dong J, Sun L, Zhu Y, Xiao Y, Li L *et al*: Full genome of influenza A (H7N9) virus derived by direct sequencing without culture. *Emerging infectious diseases* 2013, 19(11):1881-1884.

28. Taft AS, Ozawa M, Fitch A, Depasse JV, Halfmann PJ, Hill-Batorski L, Hatta M, Friedrich TC, Lopes TJ, Maher EA *et al*: Identification of mammalian-adapting mutations in the polymerase complex of an avian H5N1 influenza virus. *Nature communications* 2015, 6:7491.

29. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, Munster VJ, Sorrell EM, Bestebroer TM, Burke DF, Smith DJ *et al*: Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* 2012, 336(6088):1534-1541.

30. Linster M, van Boheemen S, de Graaf M, Schrauwen EJ, Lexmond P, Manz B, Bestebroer TM, Baumann J, van Riel D, Rimmelzwaan GF *et al*: Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell* 2014, 157(2):329-339.

31. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, Zhong G, Hanson A, Katsura H, Watanabe S *et al*: Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* 2012, 486(7403):420-428.

32. Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, O'Connor DH, Hughes AL, Neumann G, Kawaoka Y *et al*: Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nature communications* 2013, 4:2636.

33. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, Garcia-Sastre A, tenOever BR: Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell host & microbe* 2014, 16(5):691-700.

34. Watanabe T, Zhong G, Russell CA, Nakajima N, Hatta M, Hanson A, McBride R, Burke DF, Takahashi K, Fukuyama S *et al*: Circulating avian influenza viruses closely related to the 1918 virus have pandemic potential. *Cell host & microbe* 2014, 15(6):692-705.

35. Moncla LH, Zhong G, Nelson CW, Dinis JM, Mutschler J, Hughes AL, Watanabe T, Kawaoka Y, Friedrich TC: Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. *Cell host & microbe* 2016, 19(2):169-180.

36. Jonges M, Welkers MR, Jeeninga RE, Meijer A, Schneeberger P, Fouchier RA, de Jong MD, Koopmans M: Emergence of the virulence-associated PB2 E627K substitution in a fatal human case of highly

pathogenic avian influenza virus A(H7N7) infection as determined by Illumina ultra-deep sequencing. *Journal of virology* 2014, 88(3):1694-1702.

37. Annex 4: Recommendations to assure the quality, safety and efficacy of influenza vaccines (human, live attenuated) for intranasal administration [http://www.who.int/biologicals/areas/vaccines/influenza/TRS_977_Annex_4.pdf?ua=1]

38. Laassri M, Zagorodnyaya T, Plant EP, Petrovskaya S, Bidzhieva B, Ye Z, Simonyan V, Chumakov K: Deep Sequencing for Evaluation of Genetic Stability of Influenza A/California/07/2009 (H1N1) Vaccine Viruses. *PloS one* 2015, 10(9):e0138650.

39. Ghedin E, Holmes EC, DePasse JV, Pinilla LT, Fitch A, Hamelin ME, Papenburg J, Boivin G: Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. *The Journal of infectious diseases* 2012, 206(10):1504-1511.

40. Tellez-Sosa J, Rodriguez MH, Gomez-Barreto RE, Valdovinos-Torres H, Hidalgo AC, Cruz-Hervert P, Luna RS, Carrillo-Valenzo E, Ramos C, Garcia-Garcia L *et al*: Using high-throughput sequencing to leverage surveillance of genetic diversity and oseltamivir resistance: a pilot study during the 2009 influenza A(H1N1) pandemic. *PloS one* 2013, 8(7):e67010.

41. L'Huillier AG, Abed Y, Petty TJ, Cordey S, Thomas Y, Bouhy X, Schibler M, Simon A, Chalandon Y, van Delden C *et al*: E119D Neuraminidase Mutation Conferring Pan-Resistance to Neuraminidase Inhibitors in an A(H1N1)pdm09 Isolate From a Stem-Cell Transplant Recipient. *The Journal of infectious diseases* 2015, 212(11):1726-1734.

42. Pizzorno A, Abed Y, Plante PL, Carbonneau J, Baz M, Hamelin ME, Corbeil J, Boivin G: Evolution of oseltamivir resistance mutations in Influenza A(H1N1) and A(H3N2) viruses during selection in experimentally infected mice. *Antimicrobial agents and chemotherapy* 2014, 58(11):6398-6405.

43. Takashita E, Fujisaki S, Shirakura M, Nakamura K, Kishida N, Kuwahara T, Shimazu Y, Shimomura T, Watanabe S, Odagiri T: Influenza A(H1N1)pdm09 virus exhibiting enhanced cross-resistance to oseltamivir and peramivir due to a dual H275Y/G147R substitution, Japan, March 2016. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2016, 21(24).

44. Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, Porter AJ, Kurt-Jones EA, Bolon DN, Poh YP *et al*: Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. *Journal of virology* 2014, 88(1):272-281.

45. Burnham AJ, Armstrong J, Lowen AC, Webster RG, Govorkova EA: Competitive fitness of influenza B viruses with neuraminidase inhibitor-resistant substitutions in a coinfection model of the human airway epithelium. *Journal of virology* 2015, 89(8):4575-4587.

46. Nakamura G, Chai N, Park S, Chiang N, Lin Z, Chiu H, Fong R, Yan D, Kim J, Zhang J *et al*: An in vivo human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies. *Cell host & microbe* 2013, 14(1):93-103.

47. Hoper D, Kalthoff D, Hoffmann B, Beer M: Highly pathogenic avian influenza virus subtype H5N1 escaping neutralization: more than HA variation. *Journal of virology* 2012, 86(3):1394-1404.

48. Chai N, Swem LR, Reichelt M, Chen-Harris H, Luis E, Park S, Fouts A, Lupardus P, Wu TD, Li O *et al*: Two Escape Mechanisms of Influenza A Virus to a Broadly Neutralizing Stalk-Binding Antibody. *PLoS pathogens* 2016, 12(6):e1005702.

49. DeDiego ML, Nogales A, Lambert-Emo K, Martinez-Sobrido L, Topham DJ: An NS1 protein mutation (I64T) affects interferon responses and virulence of circulating H3N2 human influenza A viruses. *Journal of virology* 2016.

50. Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M, Lloyd J, Warland A, Pantic N, Admassu T, Ciccone J *et al*: Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* 2016.

Part II: Aims of the thesis

Aims of the thesis

Influenza is one of the most common infectious diseases, resulting in a high morbidity and mortality each year, although it can be prevented through effective vaccination. However, the current influenza vaccination strategies are mainly based on eliciting neutralizing antibodies against the antigenically, highly variable surface protein HA [1]. This profound antigenic variability in HA results from the relatively high error rate of the viral RNA polymerase, the humoral herd immunity directed against HA that is build up in the population during each influenza season, and the structural flexibility of HA, which tolerates almost any amino acid substitution in the most antigenic parts of this glycoprotein [2, 3]. In addition, optimal protection by current vaccination strategies is only achieved when the vaccine strain antigenically matches the circulating influenza strains. Therefore, there is an urgent need for 'a universal influenza vaccine' that, ideally, would give heterosubtypic and life-long immunity. A vaccine based on the conserved influenza M2e is a good candidate for such a vaccine [4]. Although immunity against M2e in response to influenza infection is rather low or undetectable, vaccination with M2e constructs induces a very strong anti-M2e antibody response which protects against homo- and heterosubtypic influenza strains in animal models of influenza [4-6]. However, the protective mechanisms behind M2e-based vaccines are still poorly understood. In addition, since influenza viruses are genetically diverse, it is important to investigate how these viruses could evolve when under the immune pressure elicited by M2e-based vaccines.

In this project, we first wanted to study the *in vitro* and *in vivo* viral diversity of influenza A viruses. The development of next-generation sequencing techniques (NGS), which enables sequencing of millions of DNA fragments in parallel, makes it possible to study the composition of variants in a viral population with high sensitivity. At the beginning of this project, it was unclear which NGS technique was the most suitable to study the genetic diversity present in an influenza virus population. Consequently, the first aim of this project was to compare the suitability of two NGS benchtop sequencers, the Illumina MiSeq and Ion Torrent PGM, to accurately identify the mutations and their prevalence present in an influenza A virus population (Chapter 5). Due to the segmented nature of the influenza RNA genome, an RT-PCR protocol had to be designed that would result in sufficient amplification of all eight genome segments. In addition, there was no standardized variant calling on NGS data derived from highly variable RNA viruses available at the start of this PhD project. Therefore, there was need to develop a sequencing analysis pipeline using a user-friendly bioinformatics platform to identify and determine the frequency of nucleotide variants in a viral sample.

The designed RT-PCR protocol and sequencing analysis pipeline were subsequently implemented to evaluate the genetic stability of a recombinant GFP virus, PR8-NS1(1–73)GFP, that was recently developed in our lab [7]. Recombinant influenza viruses expressing a reporter gene are very useful in a plethora of *in vitro* and *in vivo* virus applications (*e.g.* to study viral replication, spread and cell tropism). Although a reporter gene has no selective benefit for the virus, we could conclude, based

on phenotypic characterization of the virus, that this GFP-expressing virus is genetically very stable [7]. However, this conclusion seemed to be contradicted by the outcome of our NGS sequencing analysis pipeline for influenza A viruses. A remarkable drop in sequencing coverage in the GFP sequence was observed. Therefore, a second aim of this project was to investigate the origin of this observed phenomenon (Chapter 6).

Although the (pre)clinical results obtained with several M2e-based vaccines are promising to develop a 'universal influenza vaccine' based on M2e, the mechanism of action of these vaccines is still poorly understood. M2e-specific immune serum or monoclonal antibodies do not inhibit influenza A virus replication *in vitro*, or do so very rarely for some viral strains [8]. In contrast, passive transfer studies of serum derived from M2e-immunized mice demonstrated that M2e-specific IgG antibodies can confer protection and are able to reduce lung virus titers [4, 9]. Our lab, among others, has already established an essential role for FcγRs in immune-protection by M2e-specific IgGs [10]. In this project we investigated further the protective mechanism of M2e based vaccines and the role of FcγRs in M2e-based protection using two M2e-specific monoclonal antibodies of different antibody isotype and FcγRs knock-out mice (Chapter 7).

The sequence of M2e is highly conserved in nature, which likely results from the genetic constraint of this part of M2 due to sequence overlap with the gene encoding for the conserved structural protein M1, likely combined with the low immune selection pressure on this antigen induced by natural infections [5]. Therefore, it remains an open question if influenza viruses will find a mechanism to escape to the M2e immune pressure once, in the future, an M2e-based vaccine would be implemented on a large scale in the human population. Earlier in *vitro* and *in vivo* studies suggest that M2e escape viruses can emerge with mutations within or outside M2e [11, 12]. The last aim of this project was to use the established protocol and NGS data analysis pipeline to address the research question: 'How do influenza A viruses evolve under anti-M2e immune pressure *in vivo*?' (Chapter 8).

References

1. Yewdell JW, Webster RG, Gerhard WU: Antigenic variation in three distinct determinants of an influenza type A haemagglutinin molecule. *Nature* 1979, 279(5710):246-248.
2. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA: Mapping the antigenic and genetic evolution of influenza virus. *Science* 2004, 305(5682):371-376.
3. Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA *et al*: Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* 2013, 342(6161):976-979.
4. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.
5. Feng J, Zhang M, Mozdzanowska K, Zharikova D, Hoff H, Wunner W, Couch RB, Gerhard W: Influenza A virus infection engenders a poor antibody response against the ectodomain of matrix protein 2. *Virology journal* 2006, 3:102.
6. De Filette M, Ramne A, Birkett A, Lycke N, Lowenadler B, Min Jou W, Saelens X, Fiers W: The universal influenza vaccine M2e-HBc administered intranasally in combination with the adjuvant CTA1-DD provides complete protection. *Vaccine* 2006, 24(5):544-551.
7. De Baets S, Verhelst J, Van den Hoecke S, Smet A, Schotsaert M, Job ER, Roose K, Schepens B, Fiers W, Saelens X: A GFP expressing influenza A virus to report in vivo tropism and protection by a matrix protein 2 ectodomain-specific monoclonal antibody. *PloS one* 2015, 10(3):e0121491.
8. Zebedee SL, Lamb RA: Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of virology* 1988, 62(8):2762-2772.
9. Treanor JJ, Tierney EL, Zebedee SL, Lamb RA, Murphy BR: Passively transferred monoclonal antibody to the M2 protein inhibits influenza A virus replication in mice. *Journal of virology* 1990, 64(3):1375-1377.
10. El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.
11. Zebedee SL, Lamb RA: Growth restriction of influenza A virus by M2 protein antibody is genetically linked to the M1 protein. *Proc Natl Acad Sci U S A* 1989, 86(3):1061-1065.
12. Zharikova D, Mozdzanowska K, Feng J, Zhang M, Gerhard W: Influenza type A virus escape mutants emerge in vivo in the presence of antibodies to the ectodomain of matrix protein 2. *J Virol* 2005, 79(11):6644-6654.

Part III: Results

Chapter 5

**Analysis of the genetic diversity of influenza A viruses
using next-generation DNA sequencing**

Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing

Silvie Van den Hoecke[1,2], Judith Verhelst[1,2], Marnik Vuylsteke[3] and Xavier Saelens[1,2,*]

[1] Inflammation Research Center, VIB, B-9052 Ghent, Belgium

[2] Department of Biomedical Molecular Biology, Ghent University, B-9052 Ghent, Belgium

[3] Gnomixx, Onafhankelijkheidslaan 38, B-9000 Ghent, Belgium

* To whom correspondence should be addressed. Tel: +3293313620; Fax: +3292217673; Email: xavier.saelens@vib-ugent.be

Relative contributions of the authors:

SVDH performed the experiments and performed the data analysis. SVDH and XS designed the experiments. MV performed the statistical analysis. XS and JV carried out scientific supervision. XS carried out project design. SVDH, JV and XS co-wrote the manuscript. All authors read and approved the final manuscript.

ABSTRACT

Background:

Influenza viruses exist as a large group of closely related viral genomes, also called quasispecies. The composition of this influenza viral quasispecies can be determined by an accurate and sensitive sequencing technique and data analysis pipeline. We compared the suitability of two benchtop next-generation sequencers for whole genome influenza A quasispecies analysis: the Illumina MiSeq sequencing-by-synthesis and the Ion Torrent PGM semiconductor sequencing technique.

Results:

We first compared the accuracy and sensitivity of both sequencers using plasmid DNA and different ratios of wild type and mutant plasmid. Illumina MiSeq sequencing reads were one and a half times more accurate than those of the Ion Torrent PGM. The majority of sequencing errors were substitutions on the Illumina MiSeq and insertions and deletions, mostly in homopolymer regions, on the Ion Torrent PGM. To evaluate the suitability of the two techniques for determining the genome diversity of influenza A virus, we generated plasmid-derived PR8 virus and grew this virus *in vitro*. We also optimized an RT-PCR protocol to obtain uniform coverage of all eight genomic RNA segments. The sequencing reads obtained with both sequencers could successfully be assembled *de novo* into the segmented influenza virus genome. After mapping of the reads to the reference genome, we found that the detection limit for reliable recognition of variants in the viral genome required a frequency of 0.5% or higher. This threshold exceeds the background error rate resulting from the RT-PCR reaction and the sequencing method. Most of the variants in the PR8 virus genome were present in hemagglutinin, and these mutations were detected by both sequencers.

Conclusions:

Our approach underlines the power and limitations of two commonly used next-generation sequencers for the analysis of influenza virus gene diversity. We conclude that the Illumina MiSeq platform is better suited for detecting variant sequences whereas the Ion Torrent PGM platform has a shorter turnaround time. The data analysis pipeline that we propose here will also help to standardize variant calling in small RNA genomes based on next-generation sequencing data.

Keywords: influenza virus, quasispecies, next-generation sequencing, Illumina MiSeq, Ion Torrent PGM, RT-PCR

BACKGROUND

Viruses outnumber all other known life forms on earth. Furthermore, viruses in general and RNA viruses in particular have a huge genetic diversity, which is the driving force of their evolutionary success. Viral genomic diversity is well captured in the term 'quasispecies'. The term 'quasispecies theory' was first introduced by Manfred Eigen as a theoretical model to study molecular evolution by mutation and selection in self-reproducing macromolecules [1, 2]. Later, the term was also used to describe an RNA virus population consisting of a mixture of related genomes [3-6]. A viral quasispecies is defined as a proliferating population of non-identical but closely related viral genomes in a mutation-prone environment subjected to continuous competition and selection [5, 7]. Biologically, the quasispecies is the level at which selection takes place [8]. Human influenza viruses represent a prototypical example of rapid virus evolution facilitated by error-prone genome replication combined with the selection pressure imposed by host immune responses. This situation favors the emergence of fit mutant viruses that escape the herd immunity induced by infection with parental viruses or by vaccination [9, 10].

Influenza is an acute and highly contagious viral disease of the respiratory tract in humans. It is caused by influenza A and B viruses and occasionally by influenza C virus. These viruses represent three of the five genera of the *Orthomyxoviridae* family, which is characterized by enveloped viruses that have a segmented, single-stranded, negative sense RNA genome [11]. Replication of the RNA genome of influenza viruses is associated with a relatively high mutation rate (2.3×10^{-5}) because the viral RNA-dependent RNA polymerase lacks 3'-5'-exonuclease activity and therefore has no proof-reading function [12, 13]. Mutations that are introduced during replication are tolerated because they are neutral for virus fitness in a particular environment, rapidly lost because they reduce fitness, or expanded because they are advantageous [5].

The mutation rate of influenza A viruses has been traditionally determined by sequencing different cDNA clones obtained from multiple plaques descending from a plaque-purified influenza A virus [14]. In other words, viral genomes that are fit enough to generate plaques were sequenced. This approach revealed a mutation rate of approximately 1.5×10^{-5} per nucleotide per infectious cycle. Sequence analysis of multiple clones of cDNA fragments derived from one or more gene segments has also been used to study sequence variation of influenza virus derived from clinical samples [15, 16]. In addition, deep amplicon sequencing of one or two gene segments from avian H7N1 and equine H3N8 influenza viruses has been applied to study within and between host genetic variation [17, 18]. However, identification of the extent of genetic variation in a viral quasispecies under a given condition requires a highly accurate sequencing method that does not rely on molecular cloning, or a phenotypic selection method such as plaque generation. Next-generation sequencing (NGS) seems to fulfill this requirement [19-21]. However, experimental errors are introduced during the preparatory steps, *i.e.* reverse transcription and PCR amplification, and the NGS method itself is also an error-prone process [22].

102

NGS enables sequencing of multiple gigabases of DNA in a single run; the output size depends on the sequencing instrument [23]. Consequently, because the influenza genome consists of only 13,000 ribonucleotides, it is straightforward to sequence it at high coverage (*i.e.* the number of times the genome is sequenced). However, its segmented RNA genome makes it technically challenging to obtain full genome coverage. Stoichiometric RT-PCR amplification of each of the eight genomic RNA segments is difficult, in particular when starting from *ex vivo* samples such as nasal swabs or bronchoalveolar lavage from experimentally infected animals. NGS studies of influenza virus reported to date did not start from the amplification of all eight full-length genomic segments in sufficient amounts in a single reaction, and homogeneous coverage across all eight segments was not always obtained [24-29].

Here, we compared the suitability of two NGS methods to determine the influenza A virus quasispecies diversity. We deep-sequenced A/Puerto Rico/8/34 (PR8) influenza virus, which is used extensively in many research laboratories for *in vitro* and mouse experiments. In addition, PR8 virus is used as a donor to generate egg-grown reassortant viruses for seasonal influenza vaccine production. Importantly, we also took advantage of the available plasmid-based reverse genetics system for PR8 virus because it is a genetically stable equivalent of the virus [30]. We compared the quality of the primary sequence data, the read length, the coverage across the viral genome, the method-associated error rate, and the sensitivity of two modern NGS platforms: the Illumina MiSeq sequencing-by-synthesis and the Ion Torrent PGM semiconductor sequencing technique. For both sequencing platforms, we used the latest available software and the most recent chemistries available.
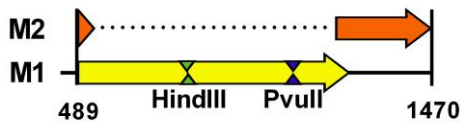
RESULTS

High-throughput sequencing of plasmid samples

Our aim was twofold: (1) to compare the performance of two high-throughput sequencing instruments; (2) to determine the complexity of an influenza A virus quasispecies (to count the number of nucleotide variants present in a swarm of genomes of that virus). We selected the Illumina MiSeq and the Ion Torrent PGM sequencing platforms because the accuracy of single nucleotide polymorphism (SNP) identification of these two popular NGS platforms is unclear. A study by Quail and colleagues concluded that the overall SNP calling rate is slightly higher for the data generated by Ion Torrent PGM than for Illumina MiSeq data [21], whereas Loman and colleagues reported a lower substitution error rate for the Illumina MiSeq [20].

We first compared the accuracy and sensitivity of these two sequencers. We used plasmid DNA to compare the accuracy of the sequencing output because it is genetically very stable. We also generated a plasmid with two tracer mutations, which allowed us to prepare mixtures with different, defined amounts of wild type and mutant plasmid before sequence analysis, in order to determine the sensitivity of the sequencers for picking out the occurrence of the introduced SNPs. For this comparison, we chose plasmids that also allowed us to generate PR8 virus with or without the introduced tracer mutations [30, 31].

We generated a mutated version of plasmid pHW197-M (pHW197-Mmut). This mutant has two silent mutations in the influenza virus M1 open reading frame (ORF) that served as tracers when mixing pHW197-Mmut and pHW197-M plasmids at different ratios. Because we intended to perform such mixing experiments with both plasmids and viruses generated from these plasmids, we carefully selected two silent mutations that most likely would not affect virus fitness. We chose these mutations based on their prevalence in human H1N1 virus isolates (see Materials and Methods). We selected two silent mutations in M1, which at the same time also added a restriction site to facilitate screening (Figure 1.A). These mutations were introduced in pHW197-M at positions 797 (C797T, pHW197-M numbering; C354T, segment 7 numbering) and 1088 (A1088T, pHW197-M numbering; A645T, segment 7 numbering). So the resulting plasmid, pHW197-Mmut, had additional HindIII and PvuII restriction sites. The presence of these mutations was verified by restriction analysis and conventional Sanger sequencing (Figure 1.B).
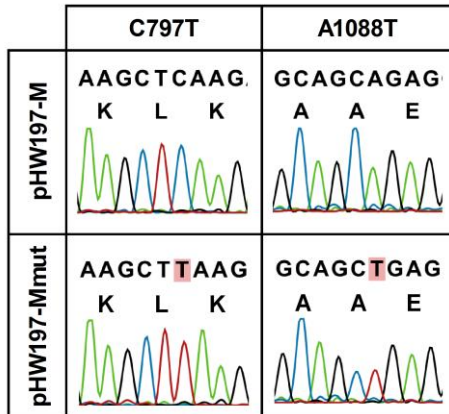
Figure 1: Introduction of synonymous tracer mutations in gene segment 7 of PR8 virus. (A) Schematic representation of the influenza M segment present in pHW197-Mmut. The open reading frames of M1 (yellow, starting at position 489, relative to the upstream CMV promoter (not depicted)) and M2 (orange, starting at position 489 and ending at position 1470) are indicated. The resulting HindIII and PvuII restriction sites are indicated. (B) Fluorograms showing the synonymous substitutions in pHW197-Mmut relative to pHW197-M at positions 797 (C to T) and 1088 (A to T). The predicted amino acid sequence is shown underneath the nucleotide sequence.

Sequence read length.

Assuming an equal error rate per base, longer read lengths are preferred for the *de novo* sequence assembly. In addition, longer read lengths increase the likelihood that one can conclude whether mutations observed in a genomic segment are linked or not. The two point mutations that we introduced in the M gene segment are 291 nucleotides apart. Therefore, to confirm the presence of these two mutations on the same DNA molecule, read lengths after processing should be at least 291 nucleotides long. Such a length should be obtained when using the Ion Torrent PGM 400 base-pair sequencing kits. The length distribution of the sequencing reads of the plasmid samples generated by both sequencers is shown in black in figure 2. Plasmid samples were fragmented with Nextera XT transposase for Illumina MiSeq and mechanically sheared by Covaris, followed by adaptor ligation before Ion Torrent PGM sequencing. Nearly 70% of the unprocessed reads obtained on the Illumina MiSeq (2x250 bp sequencing) have a length of 250 bp, and the mean read length is 233.70 bp ± 1.65 bp (Figure 2.A). The length of the unprocessed reads generated by the Ion Torrent PGM (400-bp sequencing on Ion 318 chip v2) follows a Gaussian distribution with a peak around 280 bp and a mean read length of 261.06 bp ± 2.51 bp (Figure 2.B). These values are lower than expected since the Ion PGM Template OT2 400 Kit, Ion PGM Sequencing 400 Kit and Ion 318 chip v2 (revision 2.0) that

we used should offer sequence reads of 400 bp according to their manuals. As analyzed on a High Sensitivity DNA Chip on the Agilent Bioanalyzer, the peak fragment size before emulsion PCR (emPCR) was situated around 450 bp (data not shown), indicating that Covaris shearing and subsequent size selection did not account for this relatively short average sequence length. We note that Junemann *et al.* also obtained fragments with the OT2 400 kit that were shorter than expected [19].
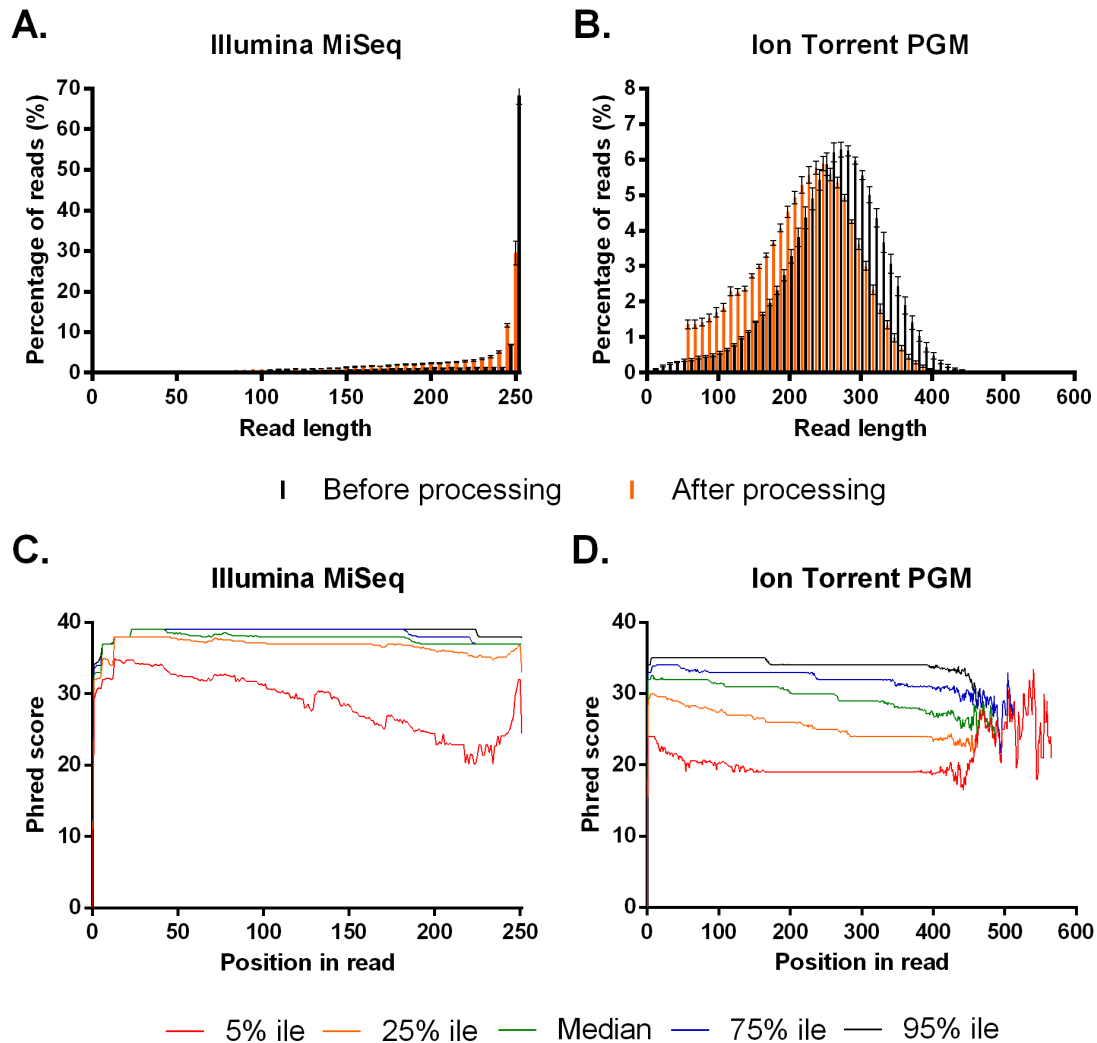
Figure 2: Quality of sequencing reads obtained on the Illumina MiSeq and Ion Torrent PGM platforms. The pHW197-M and pHW197-Mmut plasmids (= 7) were fragmented with the Nextera XT DNA sample preparation kit (Illumina MiSeq) or with Covaris mechanical shearing followed by adaptor ligation (Ion Torrent PGM). Distribution of the read lengths obtained on the Illumina MiSeq (A) and Ion Torrent PGM (B) before processing (in black, output files of sequencer) and after processing (in orange) the obtained sequencing reads. Processing implies removal of adaptor contamination, quality trimming (> Q20), the removal of ambiguous bases and removal of reads shorter than 50 bases. For the Illumina MiSeq reads, broken pairs after read processing were also removed during the processing. Error bars represent the standard deviation. (C, D) Per-base quality distribution of sequencing reads. The Phred score distribution (Y-axis) relative to the processed reads obtained after sequencing on the Illumina MiSeq (C) and Ion Torrent PGM (D). x% ile = xth percentile of quality scores observed at that position.
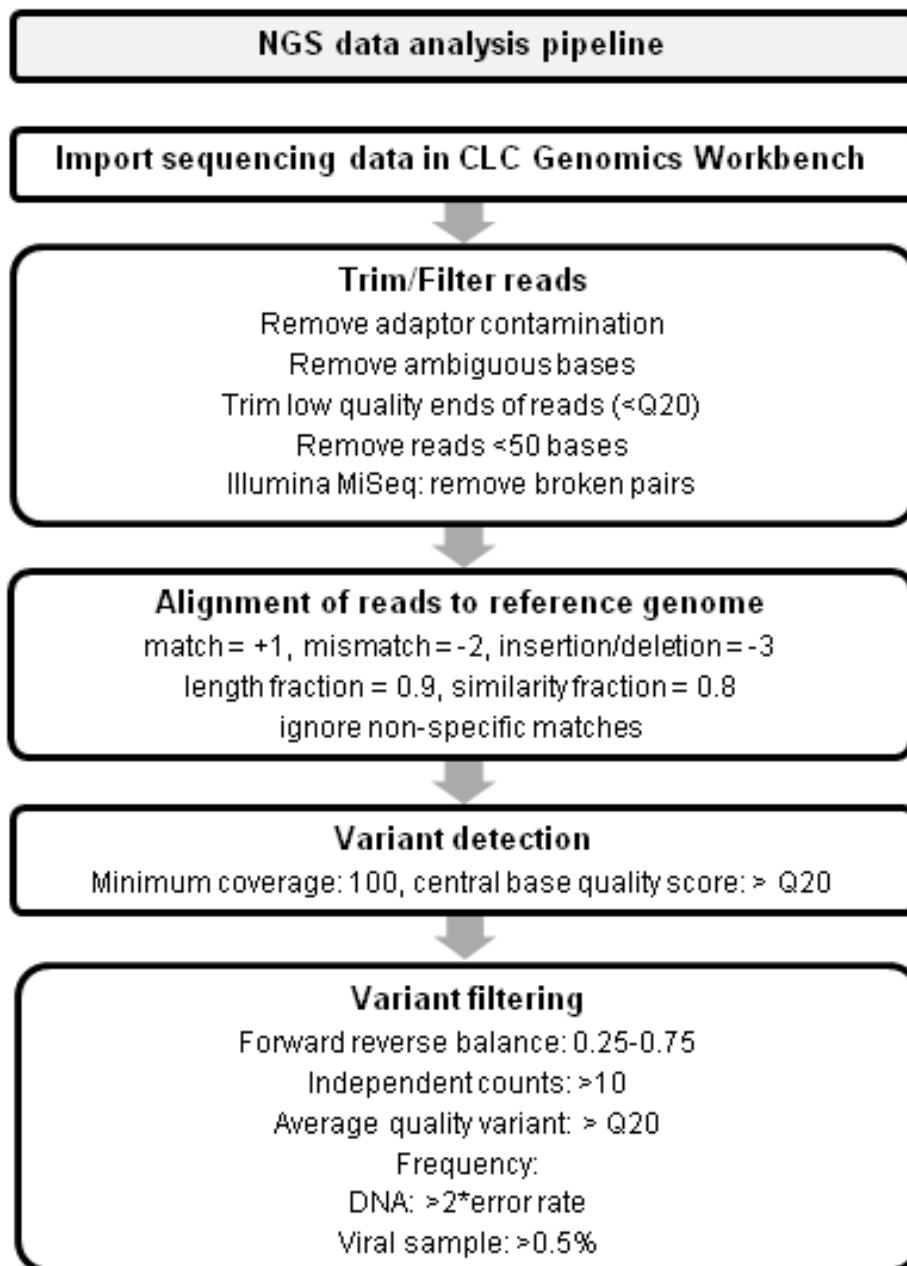
Figure 3: Next Generation Sequencing data analysis pipeline. Schematic representation of the analysis pipeline for in silico processing of next-generation sequencing data.

In silico processing of the sequencing reads.

Accurate analysis of viral quasispecies composition has to be based on high quality reads to ensure that SNPs and insertions and deletions (indels) can be confidently counted, because low quality reads could lead to over-interpretation of the number of mutations. In addition, high quality reads will lead to a higher accuracy of *de novo* sequence assembly. Therefore, we performed a quality control using the CLC Genomics Workbench software; we also propose a NGS data analysis pipeline that is generally applicable (Figure 3). First, we removed adaptor contamination and the low quality ends of the sequencing reads from the data generated by the two deep sequencing techniques. It was

recently reported that applying a Phred score of 20 or higher to filter Illumina MiSeq NGS data dramatically reduces the noise in SNP calling [32]. Hence, we applied this quality threshold to all our plasmid-derived sequencing reads. A Phred score is logarithmically related to the base-calling error probabilities. When a Phred score of 20 is assigned to a base, it means that the chance that this base is called incorrectly is 1 in 100. We also discarded ambiguous bases and read lengths below 50 bases, which further reduces the background because such short reads are often mapped inaccurately. This quality trimming and read length filtering retained 94.89% ± 0.55% of the Illumina MiSeq and 95.26% ± 0.57% of the Ion Torrent PGM reads. On the other hand, 85.99% ± 0.72% of the bases sequenced on the Illumina MiSeq and 78.99% ± 1.22% of the bases sequenced on the Ion Torrent PGM were retained. This indicates that the greatest loss of bases was due to quality trimming rather than read length filtering and that Illumina MiSeq sequencing provides higher sequencing quality than Ion Torrent PGM. The resulting read length distribution after this *in silico* filtering is shown in orange in figure 2, where the mean read length is 211.78 bp ± 2.18 bp on the Illumina MiSeq and 216.43 bp ± 1.15 bp on the Ion Torrent PGM after processing of the reads.

Quality of the sequencing reads.

The per-base quality distribution on both sequencers, using the plasmid samples as template, is shown in figure 2. Bases with a Phred score of 30 (chance of a wrong base call of 1 in 1000) are a measure of high quality data. For the raw reads obtained on the Illumina MiSeq, the 25th percentile of the Phred scores is ≥ 33 until position 251, and thus most of the sequencing reads are without sequencing error (Figure 2.C). For the reads obtained on the Ion Torrent PGM, the median of the Phred scores is ≥ 30 until position 266 (Figure 2.D). Therefore, we conclude that the overall sequencing quality of the reads obtained on the Illumina MiSeq is higher than that obtained on the Ion Torrent PGM.

Table 1: Alignment metrics for Illumina MiSeq and Ion Torrent PGM sequencing runs.

	Illumina MiSeq				Ion Torrent PGM			
	pHW197-M		pHW197-Mmut		pHW197-M		pHW197-Mmut	
	S1	S2	S1	S2	S1	S2	S1	S2
Minimum coverage	683	744	815	609	3532	4510	3995	4830
Maximum coverage	27389	28589	32802	26275	15716	17632	14664	18196
Average coverage	15369	16315	18236	14610	11525	13236	11118	13636
Standard deviation	6739	7120	7888	6315	3323	3499	2853	3562
Unmapped reads (%)	0.20	0.16	0.21	0.22	1.06	1.05	1.28	1.19
Unmapped bases (%)	0.17	0.14	0.19	0.19	1.07	1.05	1.26	1.18

Wild type (pHW197-M) and mutant (pHW197-Mmut) plasmids were sequenced in duplicate (S1 and S2) on both sequencers and the processed reads were mapped to the plasmid reference sequence.

Mapping of the sequencing reads.

To evaluate the accuracies of both sequencers, the processed reads were mapped to the plasmid reference sequence (Table 1). The percentage of unmapped bases was lower for the Illumina MiSeq (0.17% ± 0.02%) than for the Ion Torrent PGM (1.14% ± 0.10%). This is due to the lower quality of the Ion Torrent PGM sequencing reads, which reflects the intrinsic sequencing errors that lead to reduced alignment and a higher number of unmapped bases, particularly at the ends of the longer reads.
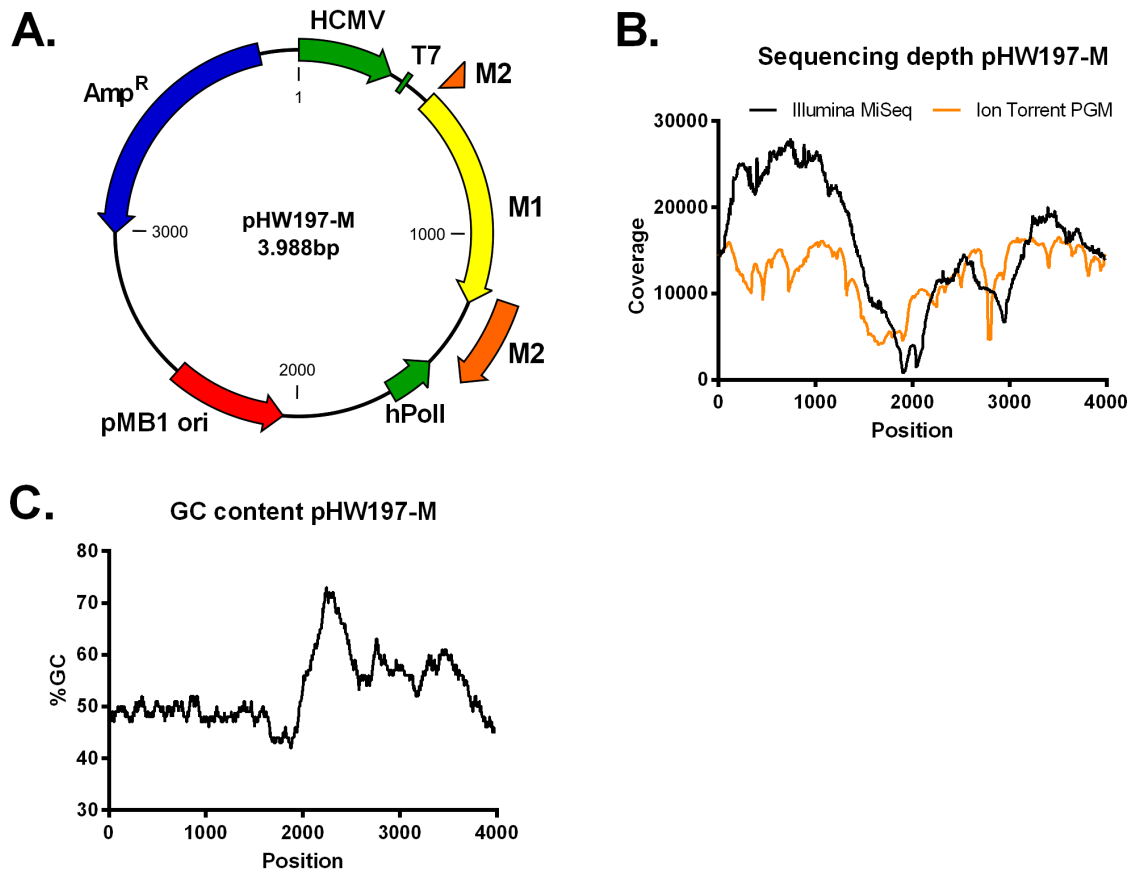
Figure 4: Next Generation Sequence analysis of pHW197-M. (A) Schematic representation of pHW197-M. HCMV: human cytomegalovirus promoter, T7: T7 RNA polymerase promoter, M1: matrix protein 1 open reading frame, M2: matrix protein 2 open reading frame (interrupted by an intron), hPolI: human RNA polymerase I promoter, pMB1 ori: origin of replication, AmpR: ampicillin resistance gene. (B) Mean sequencing depth after mapping the processed reads (n = 2) to the reference plasmid genome. The pHW197-M plasmid was fragmented with the Nextera XT DNA sample preparation kit before Illumina MiSeq sequence analysis or by Covaris mechanical shearing, followed by adaptor ligation before Ion Torrent PGM sequence analysis. (C) Percentage GC distribution in the pHW197-M plasmid reference sequence. The peak after position 2000 corresponds to the origin of replication.

For both sequencers, we observed a striking fluctuation in coverage depth (times a nucleotide is sequenced plotted against the position in the genome) (Figure 4). The largest fluctuation was seen

for the Illumina MiSeq (Figure 4.B). It is known that Illumina MiSeq and Ion Torrent PGM sequencers perform rather poorly when sequencing DNA with very low or very high GC content, which leads to low sequencing coverage of AT and GC rich regions [33, 34]. In addition, the Nextera transposon-based fragmentation that we used for the samples sequenced on the Illumina MiSeq has some sequence preference, which can lead to a fragmentation bias, particularly in small genomes [35].

Since the plasmid reference sequence is known, we were confident that any mismatching nucleotide variant could be reported as a sequencing error. The error rate per read position was 0.08% ± 0.01% for the Illumina MiSeq and 0.12% ± 0.01% for the Ion Torrent PGM. The error rate increases slightly with the read length for both sequencers, with a pronounced rise at the end of the reads on the Ion Torrent PGM (data not shown). For the Illumina MiSeq, substitutions are the dominant error type with A-to-C and T-to-G being the most prevalent (Figure 5.A), which is consistent with an earlier report [36]. In contrast, indels are dominant on the Ion Torrent PGM (Figure 5.B), and most of them are single nucleotide insertions or deletions (data not shown). Nearly all of these indels occur in homopolymeric regions. Since these regions require multiple incorporations of identical nucleotides, this increases the chance of non-linearity between the signal intensity and homopolymer length, explaining the higher indel error rate of the Ion Torrent PGM.
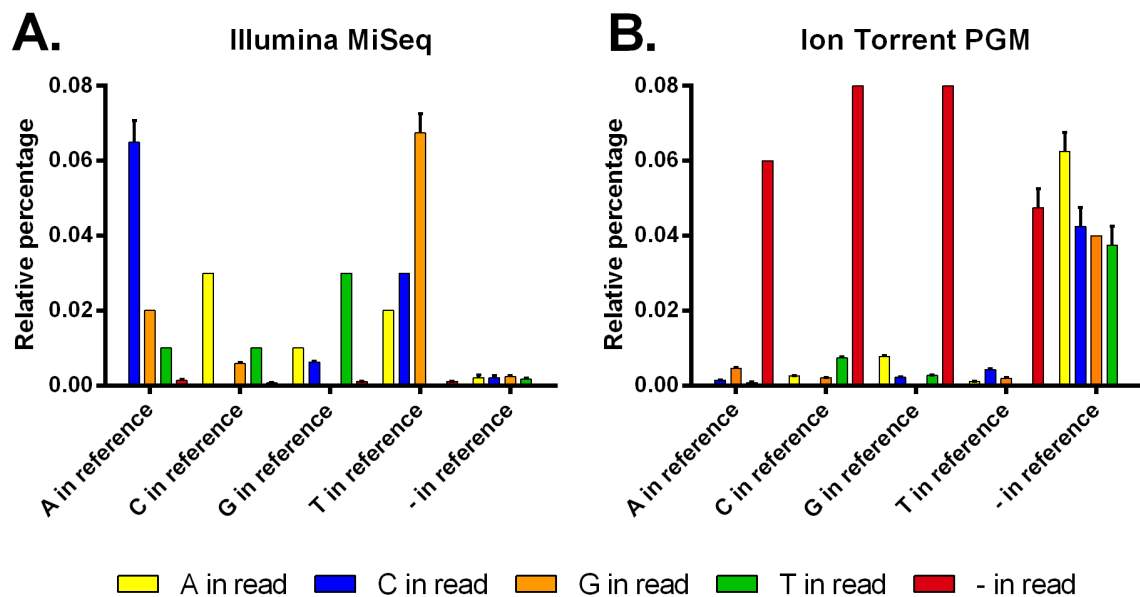
Figure 5: Comparison of nucleotide variants revealed by Illumina MiSeq and Ion Torrent PGM sequencing. The pHW197-M and pHW197-Mmut plasmids were fragmented with the Nextera XT DNA sample preparation kit (Illumina MiSeq) or by Covaris mechanical shearing, followed by adaptor ligation (Ion Torrent PGM). The samples were sequenced in duplicate and the sequence reads were processed (adaptor removal, Q20 trimming, removal of ambiguous bases and removal of reads shorter than 50 bases). For reads obtained on the Illumina MiSeq: broken pairs after read processing were also removed. The relative percentages of substitutions, insertions and deletions were determined after mapping the processed Illumina MiSeq (A) and Ion Torrent PGM (B) sequencing reads to the pHW197-M or pHW197-Mmut reference sequence. Bars represent averages from 4 samples and error bars represent the standard deviation.

Variant detection

We considered the frequency of a given nucleotide significant (a real mutation) when it was higher than twice the sequencing error background, *i.e.* above 0.16% for the Illumina MiSeq and above 0.24% for the Ion Torrent PGM. Since we are dealing with proportions very close to zero, the proportion of variants that could be miscalled at this threshold was estimated using the Agresti-Coull interval as an approximate binomial confidence interval [37]. Setting twice the background error rate as upper bound of the binomial confidence interval, only 0.0041% and 0.00002% of the variants are expected to be miscalled as true variant on the Illumina MiSeq and Ion Torrent PGM, respectively. Despite this stringent cut-off, false positive errors were still detected, mostly as a consequence of the sequence specific error profiles of both sequencers (Table 2, [21, 38, 39]). The largest number of variants was deduced from the Ion Torrent PGM data, and all of them were indels (Table 2). In contrast, the variant calls on the Illumina MiSeq were mainly SNPs (Table 2). To eliminate false positive variants, we applied extra *in silico* filtering parameters. We set the forward/reverse balance between 0.25 and 0.75, meaning that the minimum ratio between the number of forward and reverse reads that support the surmised variant should be at least 0.25. In addition, a nucleotide variant should be counted at least 10 times independently and should have an average Phred score of at least 20 (based on [40]) (Figure 3). Applying these variant filters removed most of the false positive variant calls and retained one variant from the Illumina MiSeq and six or five variants from the Ion Torrent PGM data (Table 2). So applying the variant filtering parameters has the largest impact on removing false positive variants detected in the Ion Torrent PGM data. Regardless of the sequencing method used, all false positive indels were present in homopolymer regions (at least two consecutive identical bases in the plasmid reference sequence). These variants can be excluded by using a homopolymer indel filter. However, homopolymeric regions are also the sites were the viral RNA polymerase may have the highest error rate. Therefore, applying this homopolymer indel filter to analyze viral RNA sequences (see below) could lead to underestimation of the number of variant genomes. Alternatively, the number of called variants based on the Ion Torrent PGM data can be reduced in order to exclude likely false positive variants, by increasing the average Phred score for a registered variant to 30. However, this also increased the number of false negative variant calls (data not shown).

Table 2: Number of detected variants in the pHW197-M sample before and after filtering.

	Illumina MiSeq				Ion Torrent PGM			
	Before		After[a]		Before		After[a]	
	S1[b]	S2[b]	S1[b]	S2[b]	S1[b]	S2[b]	S1[b]	S2[b]
SNP[c]	4	4	0	0	0	0	0	0
Insertion	0	0	0	0	14	12	3	1
Deletion	0	2	0	1	71	66	3	4

[a]The filtering parameters used were average quality threshold > Q20, forward/reverse balance > 0.25, and independent counts of variant > 10.

[b]Sequencing was performed in duplicate (S1 and S2).

[c]SNP = single nucleotide polymorphism.

Table 3: Sensitivity of Illumina MiSeq and Ion Torrent PGM.

| | | Illumina MiSeq | | | | Ion Torrent PGM | | | |
| | | 797 | | 1088 | | 797 | | 1088 | |
pHW197-M	pHW197-Mmut	C	T	A	T	C	T	A	T
0	100	< d.l.	99.97	< d.l.	99.96	< d.l.	99.56	< d.l.	99.89
0	100	< d.l.	99.95	< d.l.	99.94	< d.l.	99.62	< d.l.	99.96
95	5	94.84	5.14	95.40	4.58	95.22	4.75	95.02	4.96
99	1	98.78	1.19	98.93	1.06	98.97	1.02	98.96	1.02
99.9	0.1	99.80	< d.l.	99.85	< d.l.	99.93	< d.l.	99.87	< d.l.

The observed mutation frequencies (%) after mapping the reads to the reference sequence of pHW197-M are shown. < d.l. = mutation frequency falls below detection limit (< 2 * error rate, < 0.16% for Illumina MiSeq and < 0.24% for Ion Torrent PGM). The pHW197-Mmut plasmid contains the tracer mutations C797T and A1088T.

To determine the sensitivity for variant calling, we mixed pHW197-M and pHW197-Mmut plasmids in ratios of 95:5, 99:1 and 99.9:0.1 (v:v) and then sequenced the mixtures on both platforms. On both sequencers, the calculated frequency of pHW197-M or pHW197-Mmut based on the output data closely resembled the used ratios (Table 3). Nevertheless, the average quality (average Phred score) of the tracer mutations was higher on the Illumina MiSeq (37.97 ± 0.09) than on the Ion Torrent PGM (30.72 ± 1.07), making the detected variants on the Illumina MiSeq more reliable. Since the mutations are physically linked on one plasmid, both mutations should be present at similar frequencies in a single sample. This was indeed the case: the observed frequencies of the linked tracer mutations varied only slightly with on average 0.18% ± 0.26% on the mapped Illumina MiSeq reads and 0.22% ± 0.15% on the mapped Ion Torrent PGM reads. Finally, we found that the 99.9:0.1 plasmid input ratio could not be resolved because it is too close to the intrinsic error rate of both sequencers. Overall, the Illumina MiSeq is more accurate than the Ion Torrent PGM sequencer but they have similar sensitivities for detection of SNPs in plasmid DNA.

Sequencing of influenza virus samples

To compare the efficacy of the sequencers for detecting mutations in an influenza A virus sample, we generated influenza virus starting from eight plasmids, including pHW197-M or pHW197-Mmut. This resulted in wild type PR8 and mutant PR8 (PR8mut), respectively, the latter carrying two silent mutations in the M1 ORF (C354T and A645T, segment 7 numbering). These mutations did not seem to affect viral fitness because PR8 and PR8mut replicated equally well *in vitro* (Figure 6). In addition, Sanger sequencing and restriction analysis of the mutant M segment after RT-PCR revealed that the introduced tracer mutations in PR8mut were uniformly present in the stock preparation (data not shown). These viral samples were sequenced in duplicate (*i.e.* from each RT-PCR sample two libraries of DNA fragments were generated in parallel) to evaluate the consistency of the two NGS methods. In addition, wild type and mutant viruses were mixed at a ratio of 99:1 before RNA isolation to compare the accuracy of the two NGS sequencing methods to resolve this ratio. Finally, we also

wanted to quantify the number of differences, if any, between the plasmid encoded influenza virus information and the *in vitro* cultured virus samples. This quantification would reflect the baseline quasispecies diversity, in the absence of exogenous selection pressure.

A.

B.
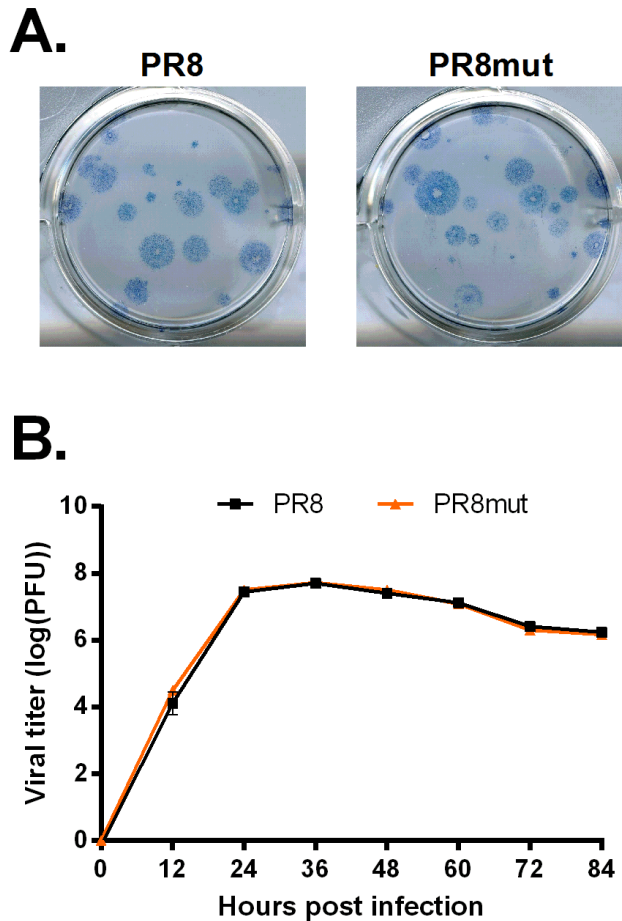
Figure 6: Comparison of the *in vitro* replication of PR8 and PR8mut influenza viruses. (A) Individual plaques of PR8 and PR8mut. Plaques were revealed by immunostaining with an anti-M2 ectodomain-specific monoclonal antibody. (B) Multi-cycle growth analysis of PR8 and PR8mut viruses. MDCK cells were infected in triplicate at a MOI of 0.01 of PR8 or PR8mut virus. Every twelve hours after infection, samples in the cell supernatant were analyzed for the presence of infectious virus by plaque assay. Error bars represent the standard deviation.

Amplification of the genomic influenza virus segments.

Ensuring sufficient coverage across all segments requires an RT-PCR protocol that amplifies all eight influenza genome segments with equal efficiency. We used an RT-PCR protocol based on the conserved termini of the influenza genome segments, which allowed us to amplify all eight segments in sufficient amounts (Figure 7.A) [41]. Surprisingly, next to the eight genomic segments, an unexpected band with a length of about 850 bp was also amplified. This band was identified by conventional Sanger sequencing after blunt-end cloning in pBlueScript and corresponded to the first 847 nucleotides of HA. Its amplification in the RT-PCR reaction was probably due to partial overlap of

the CommonUni12G primer with a nine-nucleotide perfect match in the coding region of HA (GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG, match underlined). By lowering the concentration of the CommonUni12G primer, we could avoid this extra band and obtained the eight amplicons of the expected size (Figure 7.B). Overall, these results show that this RT-PCR protocol based on the conserved termini of the influenza A genome segments is suitable for amplifying all eight segments simultaneously and efficiently.
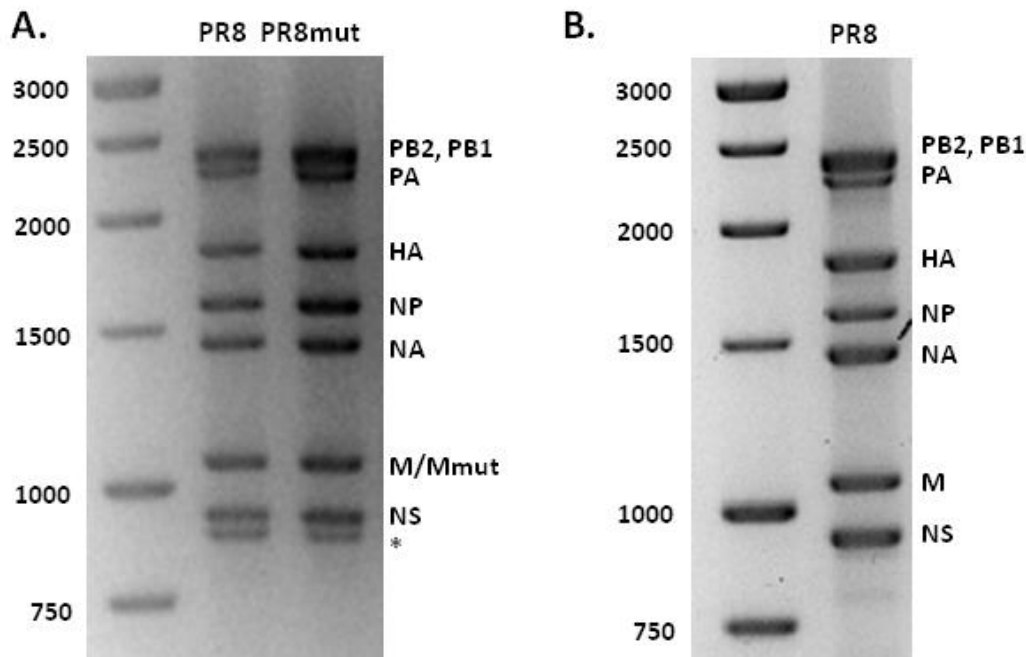
Figure 7: RT-PCR amplification of influenza A virus PR8 and PR8mut genomic RNA. (A) Electrophoretic analysis of RT-PCR products of PR8 and PR8mut separated on a 1.5% agarose gel and subsequently stained with Ethidium Bromide. PB1: polymerase basic 1, PB2: polymerase basic 2, PA: polymerase acidic, HA: hemagglutinin, NP: nucleoprotein, NA: neuraminidase, M: matrix, NS: non-structural. The amplified PB1 and PB2 RT-PCR products run at the same position in the gel. * = aspecific amplification product of 847 bp. (B) Optimized RT-PCR product resolved as in A.

De novo assembly of sequencing reads derived from viral RNA.

Accurate *de novo* nucleotide sequence assembly is essential to identify the viral quasispecies that is present in (clinical) samples. The viral RT-PCR products were purified and subjected to NGS on the Illumina MiSeq and the Ion Torrent PGM platforms. Before assembly, the reads were processed *in silico* as described above for the plasmid-derived sequences (Figure 3). Afterwards, the sequencing reads were assembled *de novo* using de Bruijn graphs [42]. This assembly method is ideally suited for high coverage next-generation sequencing data since the computational burden is lowered by first subdividing all sequencing reads in all possible subsequences with a certain short length (k), followed by looking for all neighbors with k-1 overlap. The consensus sequence is then constructed as being the alignment of k-mers that follows the shortest path connecting all overlap sequences [43]. In this

114

way, 99.90% ± 0.02% of the reads on the Illumina MiSeq and 99.65% ± 0.16% of the reads on the Ion Torrent PGM were assembled in eight contigs corresponding to the eight genome segments of the PR8 virus. These eight contigs had a mean coverage depth of 23020 ± 3504 on the Illumina MiSeq and 13768 ± 394 on the Ion Torrent PGM. All viral genome segments were almost completely covered by the consensus contigs (Table 4). Only the extreme 3' and 5' ends of each segment were not covered in all consensus sequences. This is partly due to the high sequence similarity and partial complementarity of the 5' and 3' ends of the influenza virus genome, making those reads more difficult to assemble *de novo*. In addition, the transposase-based fragmentation and tagging of the samples sequenced on the Illumina MiSeq disfavors coverage of free ends, making *de novo* assembly at these ends more difficult. For the Ion Torrent PGM samples, the adaptors were ligated to the DNA fragments that had been generated by Covaris shearing, with the free ends of the influenza genome DNA segments favoring adaptor ligation, resulting in higher coverage of the segment termini and making full-length *de novo* assembly easier. Nevertheless, in all viral contigs, the coding sequences were highly covered and entirely present. In summary, both sequencers are equally suited for *de novo* assembly of the influenza virus genome, and transposase based fragmentation should be avoided when high coverage of the influenza virus genome ends is desired.

**Table 4: Percent coverage of the influenza PR8 reference sequence after de novo assembly.**

Segment	Illumina MiSeq[a] (SD[b])	Ion Torrent PGM[a] (SD[b])
PB2	99.55 (0.30)	100.00 (0.00)
PB1	99.37 (0.52)	100.00 (0.00)
PA	99.35 (0.54)	99.30 (0.21)
HA	98.65 (0.52)	99.04 (0.50)
NP	98.79 (0.92)	98.07 (0.00)
NA	98.92 (0.82)	99.97 (0.07)
M - Mmut	98.20 (1.36)	99.55 (0.89)
NS	96.94 (0.76)	98.17 (2.12)

[a]Viral RT-PCR product sequencing reads obtained on Illumina MiSeq and Ion Torrent PGM were *de novo* assembled, followed by alignment of the obtained consensus sequence to the PR8 (n = 2) or PR8mut (n = 2) reference genome. For each segment, the percentage of the influenza reference sequence (based on the sequence from the plasmids from which the virus was produced) that is covered by the assembled contigs is given.

[b]SD = standard deviation.

Mapping of sequencing reads.

Mapping of the above-mentioned reads to the viral reference genome (based on the eight plasmids used to generate the recombinant PR8 virus) resulted in sufficient full-length coverage of the entire influenza genome (Figure 8 and Table 5). This allowed us to study the viral quasispecies, *i.e.* to determine the number of variable nucleotides at each position in the viral genome. When mapping was done with the Illumina MiSeq data, we noticed a significant coverage dip near the middle of the NP segment as well as a dip around position 600 of the PA segment, but this did not occur when the Ion Torrent PGM data were used (Figure 8). These parts of NP and PA are not particularly GC-rich or AT-rich, and these coverage dips therefore likely reflect a sequence dependency of the Nextera transposase [35, 44]. Indeed, when we used mechanical shearing to fragment the RT-PCR products before Illumina MiSeq sequencing, coverage of the NP and PA segments was high and consistent over the entire length of all PR8 genome segments (Figure 9, orange). For the viral samples sequenced on the Ion Torrent PGM, the sequencing depth is more homogenous across the segments, and the regions close to the ends of the viral segments are slightly overrepresented. This overrepresentation is probably due to mechanical shearing and subsequent adaptor ligation. The inadvertent RT-PCR amplification of the 847-bp HA fragment mentioned earlier was clearly reflected in the sequence read coverage of that segment, which showed a higher coverage for the 5' half of this segment (Figure 8). Moreover, the gradual *versus* steep drop of coverage near position 847 in the HA segment reflects the different chemistries of the Nextera transposase and the Covaris shearing/adapter ligation methods. Homogenous coverage across the HA segment was evident with the optimized RT-PCR method in which the extra partial HA-fragment was not present (Figure 9).
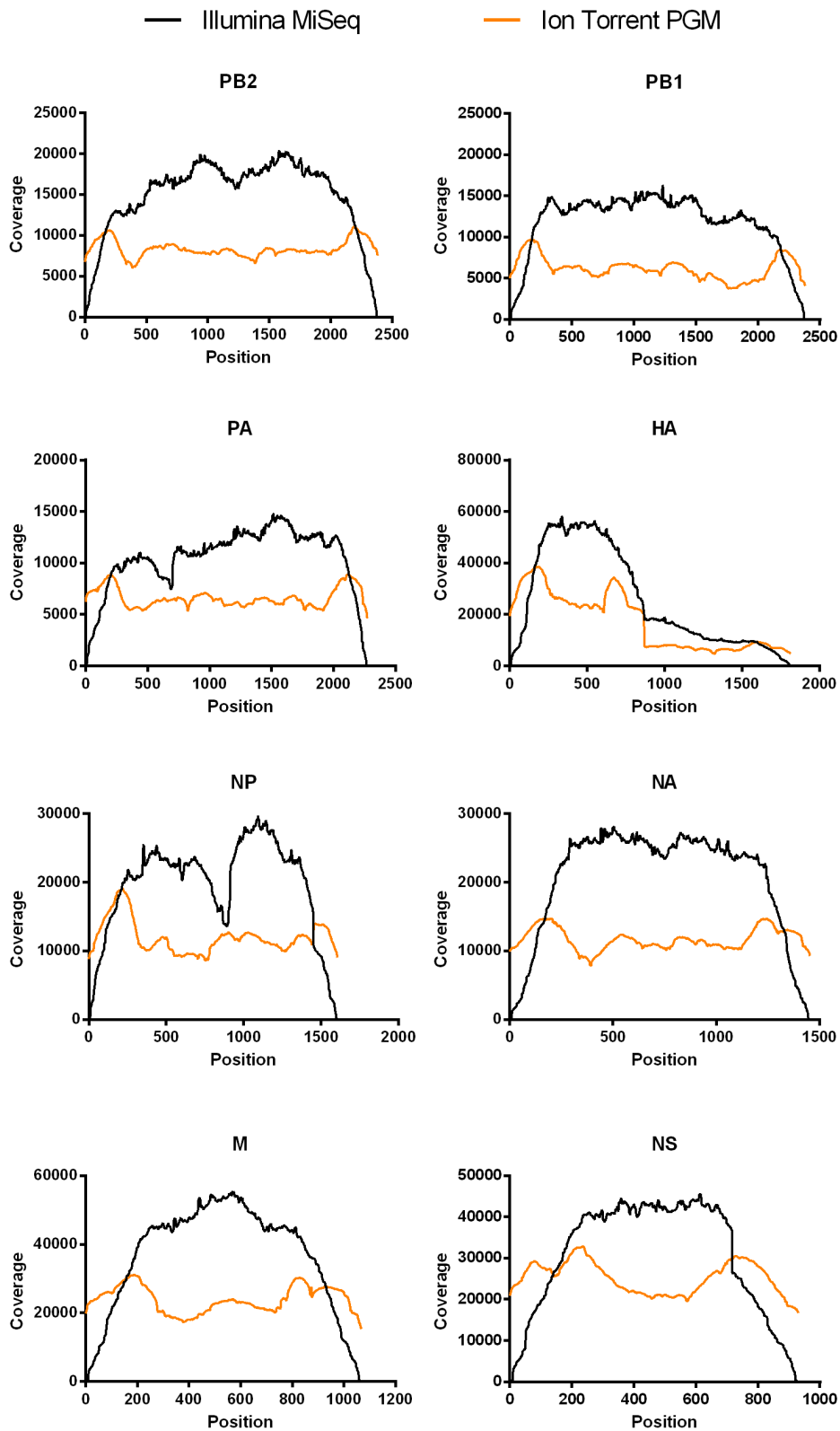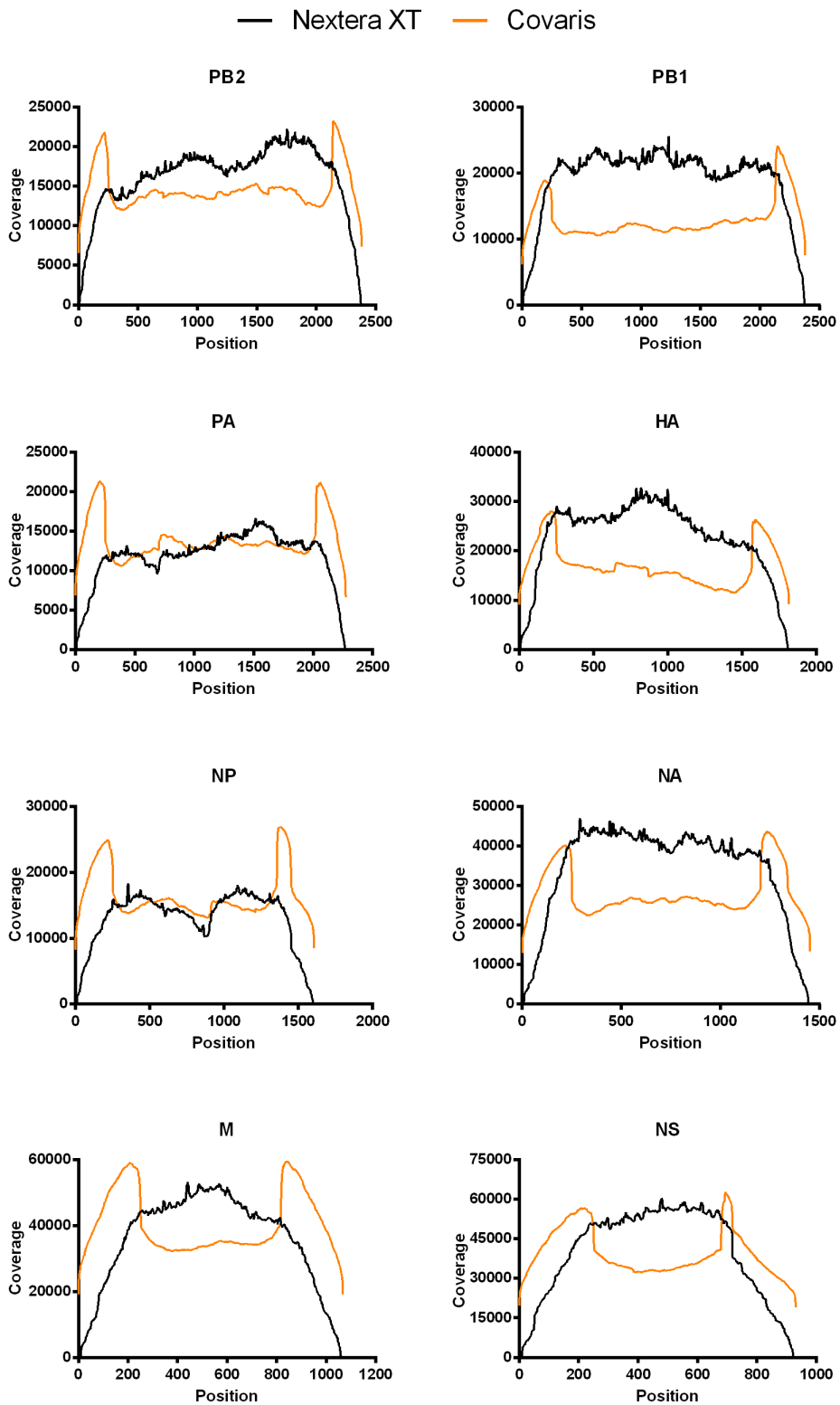
Figure 8: Sequence coverage of the influenza virus genome. Sequence coverage for the different genome segments of wild type PR8 virus sequenced on Illumina MiSeq (2x250 bp, black lines, n = 2) or Ion Torrent (Ion 318 chip v2, orange lines, n = 2). The obtained sequences were mapped to the reference genome (based on the pHW plasmids that were used to generate the virus, with addition of the extra 20 nucleotides present at the 5' site in the RT-PCR primers).

Table 5: Alignment metrics for Illumina MiSeq and Ion Torrent PGM sequencing runs.

Illumina MiSeq

PR8 S1

Segment	Length	Mapped reads	Min. coverage	Max. coverage	Average coverage
PB2	2381	159869	12	20525	15057
PB1	2381	126244	6	16513	11960
PA	2273	107490	7	14883	10533
HA	1815	213169	6	58709	25756
NP	1605	149599	9	29927	19883
NA	1453	139858	5	29353	21256
M	1067	180592	13	56656	37788
NS	930	140785	4	47293	31651

PR8 S2

Segment	Length	Mapped reads	Min. coverage	Max. coverage	Average coverage
PB2	2381	163969	14	20266	14923
PB1	2381	128791	9	16043	11750
PA	2273	110954	5	14733	10486
HA	1815	222513	5	57511	25860
NP	1605	150831	11	29497	19330
NA	1453	135597	14	27006	19834
M	1067	177520	13	54233	35854
NS	930	136505	12	44068	29591

Ion Torrent PGM

PR8 S1

Segment	Length	Mapped reads	Min. coverage	Max. coverage	Average coverage
PB2	2381	93676	6396	11399	8765
PB1	2381	72187	4016	10132	6471
PA	2273	70492	4735	9315	6940
HA	1815	148242	4613	39585	17544
NP	1605	94509	8518	19617	12324
NA	1453	77561	7918	14959	11904
M	1067	119301	15170	31854	24331
NS	930	112041	16425	33280	25428

PR8 S2

Segment	Length	Mapped reads	Min. coverage	Max. coverage	Average coverage
PB2	2381	84783	5612	10775	7947
PB1	2381	65635	3442	9253	5900
PA	2273	63994	4529	8607	6290
HA	1815	139240	4438	37662	16517
NP	1605	88966	8283	18590	11625
NA	1453	74318	7629	14494	11453
M	1067	115512	15395	30397	23553
NS	930	109661	16936	32481	24950

Wild type PR8 virus was sequenced in duplicate (S1 and S2) on both sequencers and the processed reads were mapped to the reference sequence (based on the sequence obtained from the plasmids from which the virus was produced).

Figure 9: Coverage of PR8 virus genome with the optimized RT-PCR protocol. Sequence coverage for the different genome segments of wild type PR8 virus sequenced on Illumina MiSeq (2x250 bp) using two different fragmentation methods: Nextera XT transposase-based fragmentation (black lines) and mechanical Covaris shearing followed by adaptor ligation (orange lines).The obtained sequences were mapped to the reference genome (based on the plasmids used to generate the virus).

Analysis of the viral quasispecies.

After mapping the reads to the reference genome, we called the variants using the optimal parameters described above (Figure 3). Since we started with viral RNA, we increased the background threshold for variant calling to 0.5%, what we believe is the biologically relevant frequency threshold. This value is above the estimated total error rate (including errors introduced by the virus itself) obtained after mapping all sequencing reads to the PR8 reference genome, which is 0.10% ± 0.01% on Illumina MiSeq and 0.12% ± 0.01% on Ion Torrent PGM.

Table 6: Sensitivity of Illumina MiSeq and Ion Torrent PGM to detect mutations in viral samples.

| | | Illumina MiSeq | | | | Ion Torrent PGM | | | |
| | | 354 | | 645 | | 354 | | 645 | |
PR8	PR8mut	C	T	A	T	C	T	A	T
0	100	< 0.5	99.96	< 0.5	99.95	< 0.5	99.61	< 0.5	99.95
0	100	< 0.5	99.96	< 0.5	99.95	< 0.5	99.62	< 0.5	99.95
99	1	98.21	1.77	98.31	1.64	98.90	0.76	98.95	0.90

The observed mutation frequencies (%) after mapping the reads of the PR8 and PR8mut viral samples to the wild type PR8 viral reference genome (based on the sequence from the plasmids from which the virus was produced) are shown. The PR8mut virus contains the tracer mutations C354T and A645T.

PR8, PR8mut and a mixture of PR8 and PR8mut (99% PR8:1% PR8mut, v:v, virus samples mixed before RNA isolation), were used to prepare RT-PCR products that were subsequently sequenced on both platforms (in duplicate, except for the mixed sample) (Figure 7.A). All obtained sequences were aligned to the PR8 reference genome. The output data of both sequencing platforms were processed *in silico* as described above and used to count the number of reads with C/T at position 354 and A/T at position 645 in the M segment. Illumina MiSeq slightly overestimated and Ion Torrent PGM slightly underestimated the expected percentage of tracer mutations in the PR8:PR8mut mix (Table 6). As the two introduced mutations are linked, we expected to retrieve them with the same frequency. This was indeed the case, and the observed frequencies of the linked tracer mutations differed on average by only 0.05% on the mapped Illumina MiSeq reads and by 0.27% on the mapped Ion Torrent PGM reads.

Next, we determined the number of variants at each nucleotide position in the virus-derived sequences, which would reflect the quasispecies diversity of *in vitro* grown PR8 and PR8mut virus. Sequencing each sample in duplicate and simultaneously on the same machine also allowed us to determine and compare the intrinsic variability of the two platforms. The number and types of nucleotide variants that were retained after applying the variant filter are presented in table 7. Most variants were present in both sequencing duplicates, with the highest proportion of shared variants on the Illumina MiSeq (Table 7). However, the variants that were identified in only one of the duplicates were actually also detectable in the duplicate sample, but just below one of the four variant filtering parameters. As for the plasmid samples, all of the indels in the samples sequenced on the Illumina MiSeq and most of the indels in the samples sequenced on the Ion Torrent PGM were

present in homopolymer regions. The frequencies of the sequencing variants detected by both sequencers in duplicate are presented in tables 8 (PR8) and 9 (PR8mut). This revealed 19 mutations (18 SNPs and 1 deletion) for wild type PR8 and 29 SNPs for PR8mut. Nearly all SNPs were detected with a higher average Phred score on the Illumina MiSeq (37.39 ± 0.43 for PR8) and were thus more reliable than on the Ion Torrent PGM (28.58 ± 2.44 for PR8).

Table 7: Number of variants detected in wild type and mutant PR8 quasispecies after filtering.

		Illumina MiSeq			Ion Torrent PGM			shared
		S1[a]	S2[a]	shared	S1[a]	S2[a]	Shared	
PR8	**SNP[b]**	25	26	24	19	21	18	18
	Insertion	0	0	0	1	1	0	0
	Deletion	6	5	4	9	9	3	1
PR8mut	**SNP[b]**	48	46	46	32	37	32	29
	Insertion	0	0	0	4	4	4	0
	Deletion	5	6	5	8	11	4	0

The filtering parameters were: average quality threshold > Q20, forward/reverse balance > 0.25, independent counts of variant > 10, and frequency > 0.5%.
[a]The wild type and mutant PR8 quasispecies were sequenced in duplicate (S1 and S2). [b]SNP = single nucleotide polymorphism.

Table 8: Wild type PR8 quasispecies sequenced in duplicate on both Illumina MiSeq and Ion Torrent PGM.

Segment	Position	Type	Ref	Allele	aa change	Frequency (in %)				Function/location
						Illumina MiSeq		Ion Torrent PGM		
PB1	1482	Deletion	A	-	frameshift	1.87	2.19	3.18	2.88	
PB1	**1486**	**SNP**	**A**	**G**	**Lys481Arg**	2.32	2.62	1.91	1.91	K481 crucial for polymerase function *in vivo*, not *in vitro* [46]
PA	539	SNP	A	G	silent	1.37	1.42	0.56	0.54	/
HA	**607**	**SNP**	**A**	**G**	**silent**	1.60	1.56	2.02	1.85	/
HA[#]	659	SNP	G	A	Glu203Lys	1.13	1.23	0.65	0.60	enhanced receptor binding activity [47]
HA	660	SNP	A	G	Glu203Gly	3.11	3.02	1.76	1.55	slightly increased α2-6 and decreased α2-3 binding [48]
HA	**747**	**SNP**	**A**	**G**	**Glu232Gly**	11.56	11.43	7.29	7.19	receptor specificity [49]
HA	764	SNP	G	A	Asp238Asn	0.83	0.80	0.65	0.60	enables binding to α2.3- and α2.6-linked sialic acids [50]
HA	**765**	**SNP**	**A**	**G**	**Asp238Gly**	39.73	39.43	35.33	35.00	enables binding to α2.3- and α2.6-linked sialic acids [51, 52]
HA	768	SNP	A	G	Gln239Arg	2.81	3.12	1.43	1.23	preferential binding to α-2,3-linked glycans [51]
HA	**823**	**SNP**	**A**	**G**	**Ile257Met**	1.76	1.54	0.72	0.74	located in head domain close to Sa antigenic site [53]
HA[#]	**1199**	**SNP**	**A**	**G**	**Ser383Gly**	1.41	1.15	1.14	1.29	located in stem domain
HA	**1330**	**SNP**	**A**	**G**	**silent**	1.59	1.50	1.02	1.20	/
HA	1424	SNP	G	A	Val458Met	95.25	95.67	97.85	97.77	located in stem domain
HA	1440	SNP	A	G	Glu463Gly	1.91	1.75	0.58	0.56	located in stem domain
HA	1451	SNP	A	G	Ser467Gly	0.70	0.81	0.62	0.63	located in stem domain
NP	212	SNP	C	T	silent	1.80	1.76	0.83	0.64	/
NP	1249	SNP	A	G	Asn395Ser	10.71	11.01	5.97	6.76	located in NP-NP and NP-PB2 interaction domain [54, 55]
NP	1324	SNP	T	G	Phe420Cys	3.43	3.41	1.31	1.15	located in the hypervariable NP418-426 CTL epitope [56]

[#] = not present in Genbank or Influenza Research Database, Bold = variant also present in PR8mut quasispecies. Ref: Reference nucleotide

HA segment = numbering of HA amino acid residues is based on the PR8 HA open reading frame with the starting methionine as position = 1.

Table 9: Mutant PR8 quasispecies sequenced in duplicate on both Illumina MiSeq and Ion Torrent PGM.

Segment	Position	Type	Ref	Allele	aa change	Illumina MiSeq		Ion Torrent PGM		Function/location
PB2	416	SNP	A	G	silent	1.55	1.30	0.59	0.57	/
PB1	**1486**	**SNP**	**A**	**G**	**Lys481Arg**	2.52	2.80	1.79	2.37	K481 crucial for polymerase function *in vivo*, not *in vitro* [46]
PA	212	SNP	G	T	Glu56Asp	5.72	4.85	2.09	2.18	located in endonuclease domain [57, 58]
PA	1139	SNP	G	T	Gln365His	2.50	2.40	1.00	1.05	located in PB1 interacting domain [59]
HA	524	SNP	A	C	Ser158Arg	13.17	12.93	9.70	9.63	Compensatory mutation in [60], located in Ca antigenic site [53]
HA	524	SNP	A	T	Ser158Cys	0.98	0.90	0.61	0.63	located in variable Ca antigenic site [53]
HA	**607**	**SNP**	**A**	**G**	**silent**	1.54	1.43	2.23	2.47	/
HA	**747**	**SNP**	**A**	**G**	**Glu232Gly**	39.95	40.14	36.74	36.02	receptor specificity [49]
HA	**765**	**SNP**	**A**	**G**	**Asp238Gly**	3.17	3.07	1.50	1.44	enables binding to α2,3- and α2,6-linked sialic acids [51, 52]
HA	**823**	**SNP**	**A**	**G**	**Ile257Met**	2.92	3.08	1.49	1.64	located in head domain close to Sa antigenic site [53]
HA	828	SNP	A	G	Glu259Gly	5.17	4.96	2.09	1.95	located on surface head domain close to Sa antigenic site [53]
HA	1088	SNP	T	A	Phe346Ile	6.99	6.76	3.75	4.18	located in fusion peptide [61]
HA	1090	SNP	T	G	Phe346Leu	1.28	1.09	0.59	0.69	located in fusion peptide [61, 62]
HA	1109	SNP	A	G	Ile353Val	59.69	60.02	62.48	61.74	described as fusion peptide pseudorevertant [61, 63]
HA[#]	**1199**	**SNP**	**A**	**G**	**Ser383Gly**	1.13	1.18	0.97	1.05	located in stem domain
HA	**1330**	**SNP**	**A**	**G**	**Silent**	1.59	1.67	1.94	1.14	/
HA	1424	SNP	G	T	Val458Leu	3.43	3.13	1.63	1.64	located in stem domain, not surface exposed
HA	1430	SNP	A	G	Asn460Asp	10.08	9.60	5.82	6.10	present in the PR8 quasispecies grown on MDCK cells [64]
HA	1431	SNP	A	G	Asn460Ser	14.79	14.62	10.15	10.29	located in stem domain
HA	1487	SNP	G	A	Gly479Arg	1.59	1.39	0.61	0.57	located in stem domain, not surface exposed
NP	635	SNP	G	A	silent	4.44	3.93	2.45	1.87	/
NP[#]	739	SNP	T	C	Ile225Thr	1.11	1.15	0.52	0.56	surface exposed, in NP-NP interaction domain [55]
NA	476	SNP	T	A	Cys146Ser	6.85	6.71	4.33	3.85	located in head domain, involved in coupling of subunits [65]
NA	994	SNP	C	T	silent	1.03	1.00	0.60	0.65	/

Table 9 (continued): Mutant PR8 quasispecies sequenced in duplicate on both Illumina MiSeq and Ion Torrent PGM.

Segment	Position	Type	Ref	Allele	aa change	Frequency (in %)				Function/location
						Illumina MiSeq		Ion Torrent PGM		
M	354	SNP	C	T	introduced	99.96	99.96	99.61	99.62	/
M	645	SNP	A	T	introduced	99.95	99.95	99.95	99.95	/
NS	409	SNP	G	T	NS1: Gln121His	40.37	39.54	31.96	32.61	situated next to the $NS1_{122-130}$ CTL epitope [66]
NS	549	SNP	G	A	NS1: Gly168Glu	1.27	1.20	0.71	0.59	NS1: located in effector domain [67]
					NS2: Asp11Asn					NS2: N-terminal domain
NS[#]	564	SNP	A	G	NS1: Asp173Gly	1.05	1.10	0.65	0.64	NS1: located in effector domain [67]
					NS2: Met16Val					NS2: Met16 is involved in nuclear export NP [68]

[#] = not present in Genbank or Influenza Research Database. Bold = variants also present in PR8 quasispecies. Ref: Reference nucleotide

HA segment = numbering of HA amino acid residues is based on the PR8 HA open reading frame with the starting methionine as position = 1.

The average difference between the frequencies of a variant in PR8 sequencing duplicates was only 0.17% ± 0.12% for the Illumina MiSeq and 0.16% ± 0.18% for the Ion Torrent PGM, again indicating that both sequencing platforms provide reproducible output (Table 8). However, the frequency of occurrence of the variants differed substantially between sequencers. For example, the mean variant frequency differed between 0.06% (position 1199 in the PR8 HA segment) and 4.5% (position 1249 in the PR8 NP segment) for the same viral sample sequenced on both sequencers. In addition, most detected variants were present at a lower frequency based on the Ion Torrent PGM output. Similar results were obtained for the PR8mut samples (Table 9). To determine whether this difference in frequencies is significant between the sequencing platforms, variant frequencies obtained in PR8 and PR8mut were analyzed using logistic regression, considering loci with low (< 15%) and high (> 15%) minor variant frequencies as separate classes. This analysis clearly indicates that when the minor variant is present at a low frequency, the Illumina MiSeq systematically detects the minor variants at significantly higher frequencies than the Ion Torrent PGM (Figure 10).
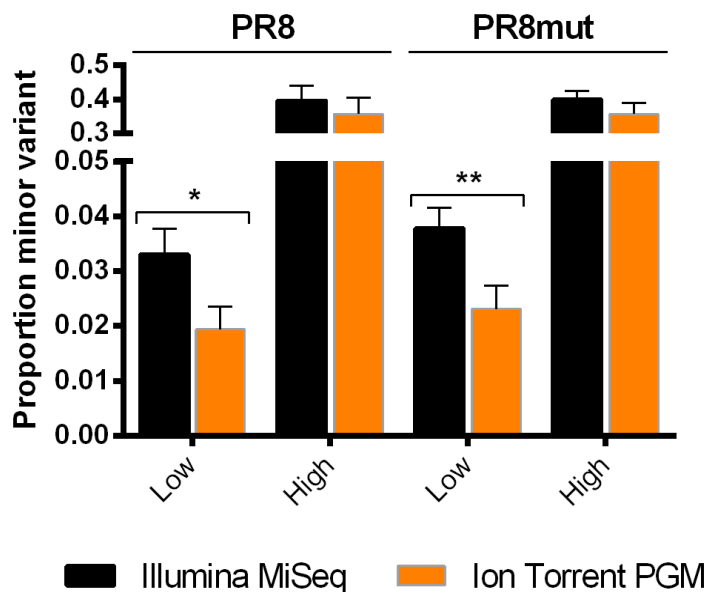
Figure 10: Low frequency minor alleles are detected at significantly higher frequencies by Illumina MiSeq compared to Ion Torrent PGM. Nucleotide variants were subdivided in two frequency classes: high (frequency minor allele > 15%, n = 4) and low (frequency minor allele: < 15%, n = 42). Mean proportions ± s.e. of the minor variants detected in PR8 and PR8mut viral samples by the Illumina MiSeq and Ion Torrent PGM are shown. Minor allele proportions were analyzed by logistic regression (link function = logit). Significance levels of pairwise comparisons were assessed by a Fisher's protected least significance difference test * = p<0.05, ** = p<0.01.

Almost all mutations detected in the wild type and/or mutant PR8 quasispecies are also present in H1N1 viral sequences retrieved from the Influenza Research Database and/or Genbank. The exceptions are indicated with a number sign (#) in tables 8 and 9. These sequence variants (Glu203Lys and Ser383Gly in HA, Ile225Thr in NP and Asp173Gly/Met16Val in NS1/NS2) might exist in

nature but have not been reported yet. Most of the detected mutations are present in the HA segment, which is also the most variable influenza protein in nature [45]. Most of the detected mutations were substitutions occurring at a frequency < 5%. However, three mutations in HA and one in NP of PR8 as well as four mutations in HA and one in NS of PR8mut were present at a frequency > 10% (based on Illumina MiSeq data) (Tables 8 and 9). Of all detected variants, only seven (five non-synonymous and two synonymous) were shared by both PR8 and PR8mut and present in all samples sequenced. These were all in the HA segment, except for one variant in PB1 (Tables 8 and 9, bold).

Taken together, these results show that both the wild type and mutant PR8 virus behave as a fairly heterogeneous virus populations even in the absence of external selection pressure.

DISCUSSION

Next generation sequencing (NGS) has become increasingly valuable to study virus diversity. NGS instruments have a very high sequencing capacity and therefore allow a very high coverage of the relatively small genome of RNA viruses. NGS analysis is thus in principle well suited for determining the genetic heterogeneity of RNA viruses. Unfortunately, in many research articles on viral quasispecies diversity there is little information on how the raw data were processed. Furthermore, the performance of different commercially available NGS platforms for quasispecies analysis has not been evaluated. Here, we compared the quality of the sequencing output obtained on the Illumina MiSeq and Ion Torrent PGM benchtop sequencers. We also propose an analysis pipeline for *in silico* processing of the sequencing data that allows identification and frequency determination of nucleotide variants in the influenza A virus (Figure 3). This analysis pipeline will help to standardize variant calling in small RNA genomes based on NGS data.

To determine the influenza genome diversity by NGS technology, different hurdles have to be overcome. First, it is technically challenging to obtain high quality full-length RT-PCR products that cover the complete segmented RNA genome of influenza viruses. We optimized an RT-PCR protocol with primers based on the conserved 3' (Uni12) and 5' ends (Uni13) of the eight genome segments [69-71]. Critical steps in this protocol are primer concentration and annealing and elongation times. Because the sequence of these segment ends is conserved, this RT-PCR should be applicable to different influenza A virus strains.

A second hurdle is to distinguish between mutations that truly represent the viral genome diversity from errors introduced by RT-PCR amplification and the NGS chemistry. The first step is to filter the output sequence data *in silico* to retain only high quality reads. However, the available software and filtering parameters vary and are not always clearly described in the literature, making comparison of results very difficult. To reduce false positive variant calls introduced by the sequencing method, we applied specific trimming, filtering and variant calling parameters in the CLC Genomics Workbench

126

software. We first applied this bioinformatics analysis pipeline to sequencing reads derived from plasmid DNA samples. We removed adaptor contamination, ambiguous nucleotides and trimmed low quality bases at the end of the reads by applying a Phred score of 20. Then, we excluded reads shorter than 50 bases to avoid unspecific mapping of these short reads. Trimming eliminated relatively more bases from the Ion Torrent PGM, meaning that the base quality of sequencing reads from the Ion Torrent PGM is lower than that from the Illumina Miseq. Subsequent filtering on read length also had a bigger effect on the Ion Torrent PGM reads. In other words, the potential advantage of longer read lengths obtained with the Ion Torrent machine was cancelled by their relatively low quality. Together this resulted in a higher relative loss of bases for the Ion Torrent PGM data than for the Illumina MiSeq data (21.01% versus 14.01% respectively). Furthermore, the Phred score distribution across the reads, a measure of the intrinsic sequencing quality, was higher for the Illumina MiSeq data than for the Ion Torrent PGM data, resulting in a lower error rate. After this quality control, the sequencing reads were mapped to the reference sequence, resulting in a higher percentage of mapped reads for the Illumina MiSeq. The total mapping error rate of the Illumina MiSeq (mainly nucleotide substitutions) was lower than that of the Ion Torrent PGM (mainly indels). This finding is in agreement with Loman and colleagues [20]. However, for plasmid DNA analysis the substitution error rate on the Ion Torrent PGM appeared to be lower than that of Illumina MiSeq (Figure 5). After variant calling, the resulting hits were filtered based on frequency, forward/reverse balance, average quality, and independent counts to remove false positive variants. After filtering, both sequencers detected the tracer mutations we had introduced with excellent accuracy and sensitivity. Nevertheless, the average quality (Phred score) of the detected variants was higher on the Illumina MiSeq than on the Ion Torrent PGM, making the variants detected on the Illumina MiSeq more reliable. The number of false positive variants can be further reduced by cross-platform replication, but the different biases of the sequencing platforms may cause many true variants to be overlooked when cross-platform replicates are compared [72, 73].

We then applied the analysis pipeline outlined in figure 3 to PR8 and PR8mut virus, which were generated by a plasmid-based reverse genetics system and amplified in MDCK cells. In our opinion, variants in the influenza virus genome that appear with a frequency below 0.5% are very difficult to distinguish from the background noise that is cumulatively introduced by RT-PCR and the inherent variation due to the chemistry of currently available Illumina and Ion Torrent sequencers. We propose that a similar threshold of 0.5% should be applied to interpret the genetic diversity of RNA viruses. Nevertheless, mutations with a frequency as low as 0.05−0.2% in Chikungunya virus have been reported in the literature as meaningful based on Illumina GAIIX sequencing [74]. Given the error rate of the influenza virus polymerase, resulting in approximately one mutation per 10.000 nucleotides, together with the errors introduced during RT-PCR and the technical background error rate of the NGS platforms applied in this study, it is not straightforward for both the Illumina MiSeq and the Ion Torrent PGM to identify each variant in the viral quasispecies. Nevertheless, even with

the threshold of 0.5% proposed here, NGS will enable studying of the viral diversity in much more detail than in the past.
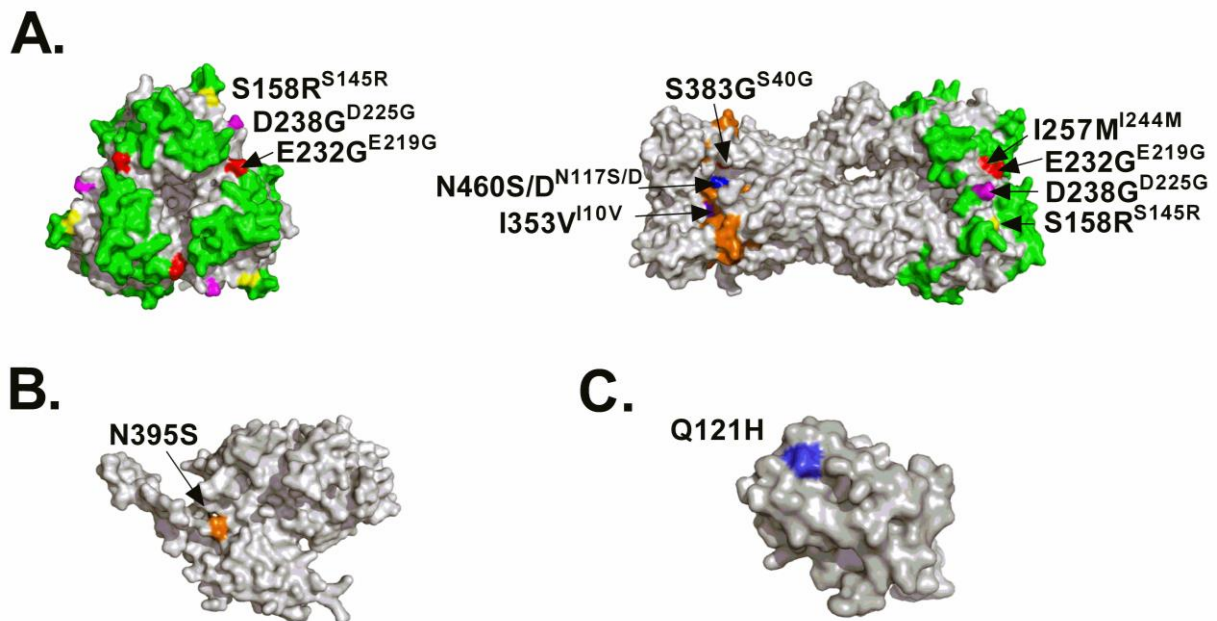
Figure 11. Position of variants present in PR8 and PR8mut quasispecies in the HA, NP and NS1. The variants in HA (hemagglutinin), NP (nucleoprotein) and NS1 (non-structural protein 1) detected in the PR8 and PR8mut quasispecies were modeled with PyMol (Delano Scientific, http://www.pymol.org), using the HA from A/Puerto Rico/8/1934 (H1N1) (PDB code: 1RVX), the NP from A/WSN/1933 (H1N1) (PDB code: 2IQH) and the effector domain of NS1 from A/Puerto Rico/8/1934 (H1N1) (PDB code: 3RVC). (A) Top (left) and lateral (right) view of the surface exposed amino acids of the HA trimer. The Cb, Ca, Sa and Sb antigenic sites are shown in green. The mutations that are present in both PR8 and PR8mut are shown in red or in magenta if they overlap with the antigenic sites. Mutations in PR8mut that are present at a frequency > 5% are shown in blue or in yellow when overlapping with the antigenic sites or in purple when overlapping with the fusion peptide (orange). The mature H3 amino acid numbering of the variants is provided in superscript. (B) Lateral view of the NP monomer with the N395S mutation present in PR8 shown in brown. (C) The effector domain of NS1 with the Q121H mutation in PR8mut shown in blue.

Our analysis showed that the *de novo* assembled PR8 and PR8mut sequences correspond very well to the plasmid-derived reference genome. We detected 19 mutations in PR8 and 29 mutations (including the two tracer mutations) in PR8mut with a frequency of 0.5% or higher. When a variant was present at low frequency (<15%), the Illumina MiSeq detected it with significantly higher frequency than the Ion Torrent PGM. Most of the detected mutations were transitions and appeared with a frequency below 5%. However, three mutations in HA and one in NP of PR8, as well as four mutations in HA and one in NS of PR8mut, were present at a frequency > 10% (based on Illumina MiSeq data) (Tables 8 and 9). We detected only one single nucleotide deletion in the PR8 virus. This deletion was in a homopolymer at position 1482 in PB1 but was detected with a frequency of 2−3% by both sequencers, in both duplicates of PR8 virus. In addition, this deletion was also detected with

a similar frequency in both PR8mut samples sequenced on the Illumina MiSeq and in one of the duplicate samples sequenced on the Ion Torrent PGM. This deletion disrupts the open reading frame, leading to premature termination of PB1. This detrimental mutation is in line with the finding of Brooke and colleagues, who showed that most of the infectious influenza A virions fail to express detectable levels of one or more viral proteins [75].

We focused on the mutations detected by both sequencers with a frequency > 5% and on the mutations that appeared in both wild type and mutant PR8 viruses. There are three such mutations in the HA head domain of PR8 and four in the HA head domain of PR8mut, and all of them are part of or close to the antigenic sites (Figure 11.A). The shared Asp238Gly mutation (Asp225Gly for H3 numbering) is associated with enhanced virion binding to the avian-type Sia(α2-3)Gal and was reported previously as a position that is selected by egg-adaptation of influenza viruses [76]. The Ser158Arg mutation (Ser145Arg for H3 numbering) in PR8mut has been described as a compensatory mutation in PR8 virus possessing the Lys165Glu mutation in HA (H3 numbering), which decreases the receptor binding avidity and replication kinetics of the virus [60]. The two mutations in the stem domain are relatively conservative (Ser383Gly and Val458Met; Ser40Gly and Val115Met for H3 numbering of HA2) and therefore might not affect virus replication. Remarkably, the G-to-A substitution at position 1424, leading to the Val458Met change in HA, had a frequency close to 100% in the PR8 HA segment but was absent in PR8mut (although a Val458Leu change is present in a small percentage of PR8mut). This mutation was probably fixed in the wild type virus genome at a very early step, *e.g.* during plaque purification of the PR8 seed virus we used to prepare stock virus. We also picked up two other codon changes in the HA stem region of PR8mut: Asn460Asp (5−10%) and Asn460Ser (10−15%) (Asn117Asp and Asn117Ser for H3 numbering of HA2). Based on pyrosequencing of the HA segment, the Asn460Asp mutation has been observed in 12.2% in PR8 virus grown on MDCK cells [64]. In addition, the PR8mut caries the Ile353Val (Ile10Val for H3 numbering of HA2) mutation in the HA fusion peptide at a frequency of about 60%. A valine at this position has been observed in a PR8 pseudo-revertant after introducing the Ile10Ala mutation. A valine at this position is compatible with the α-helical structure of the fusion peptide [63]. Both PR8 viruses also contain mutations in other segments. For example, both viruses share the conservative Lys481Arg mutation in PB1. This lysine at position 481 is crucial for the polymerase function of PB1 *in vivo* but mutating it to alanine was tolerated *in vitro* [46]. In wild type PR8, the Asn395Ser variant in NP is in a domain involved in NP−NP and NP−PB2 interactions (Figure 11.B) [77]. The Gln121His variant detected in NS1 of PR8mut is situated just before a human CTL epitope (Figure 11.C) [66]. Remarkably, none of the variants we observed correspond to the variants described in an earlier study, in which a PR8 strain (originally adapted for growth on embryonated chicken eggs) was adapted for growth on MDCK cells [78]. However, we used MDCK cells only to expand our virus stock, which corresponds to about six cycles of PR8 virus replication. Furthermore, we generated our PR8 virus starting from eight plasmids, indicating that the passaging history is a determinant of the variants detected in an influenza virus quasispecies.

Both sequencers are highly effective for accurate detection of low frequency mutations, but each one has its advantages and limitations. On the one hand, the Illumina MiSeq platform has about three times higher output capacity than the Ion Torrent PGM, enabling sequencing of more samples in parallel on the Illumina MiSeq. On the other hand, the Ion Torrent PGM is significantly faster: its time from sample preparation to data analysis is one day less than for the Illumina MiSeq. After the *in silico* quality control, the two sequencers produced reads of comparable lengths. The Illumina MiSeq had a higher intrinsic sequencing quality than the Ion torrent PGM, presumably because detecting incorporated bases based on a coupled fluorescent dye (Illumina) gives less noise than a change in pH caused by release of a proton after incorporation of a base (Ion Torrent). However, the Ion Torrent PGM had a lower false-positive rate for detecting SNPs. Another interesting observation is the lower coverage of the ends of the viral segments on the Illumina MiSeq due to the transposase-based fragmentation. Nextera transposase-based fragment library preparation is convenient and fast but results in low coverage of segment termini. We also noticed some sequence bias of this transposase-based fragmentation approach (Figures 8 and 9). Mechanical fragmentation followed by adaptor ligation enables comparable coverage of all bases of the influenza virus genome, and is therefore the preferred method for library preparation (Figures 8 and 9).

The proposed RT-PCR protocol and subsequent analysis pipeline for influenza viruses is widely applicable, *e.g.* to study vaccine composition, analyze virus evolution under selection pressure, monitor mutations associated with antiviral resistance, and assemble the reference genome of new viral isolates. For clinical samples, the shorter turnaround time of the Ion Torrent PGM (sample preparation, sequencing and analysis in about 2 days) is clearly advantageous to the Illumina MiSeq (about 3 days). In contrast, when analyzing many viral samples at high coverage, the greater output of the Illumina MiSeq is an important advantage.

CONCLUSION

Our study underlines the power and limitations of two commonly used next-generation sequencers for the analysis of influenza gene diversity. We propose an *in silico* pipeline for selecting high quality reads obtained by NGS platforms. This pipeline is also more widely applicable. Due to the lower total error rate and the higher sequencing quality of the reads, we conclude that the Illumina MiSeq platform is more suited than the Ion Torrent PGM for detecting variant sequences, whereas the Ion Torrent platform has a shorter turnaround time. In addition, we found that the detection limit for reliable recognition of variants in the viral genome required a frequency of 0.5% or higher.

MATERIAL AND METHODS

Cell lines.

MDCK and HEK293T cells were cultured in Dulbecco's Modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, non-essential amino acids, 2 mM L-glutamine, 0.4 mM sodium pyruvate, 100 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in 5% CO_2.

Generation and production of plasmids with tracer mutations.

Reverse genetics plasmids for PR8 virus were kindly provided by Dr. Robert G. Webster (St. Jude Children's Research Hospital, Memphis, USA) [31]. We introduced two silent mutations in the M coding gene, a C-to-T substitution at position 797 (numbering relative to the human cytomegalovirus promoter in the pHW197-M plasmid) and an A-to-T substitution at position 1088 in pHW197-M. These two positions were selected as follows. First, we generated a consensus sequence of the M-gene based on all full-length segment 7 sequences of human H1N1 viruses present in the Influenza Virus Resource Database (NCBI) on September 11th, 2011. Next, we aligned the consensus sequence to the M segment of PR8 (present in pHW197-M) and selected two synonymous mutations in the M1 open reading frame at positions C354T and A645T (segment 7 numbering). These two mutations were introduced by two consecutive rounds of quickchange site-directed mutagenesis (Stratagene) at positions C797T and A1088T in pHW197-M to generate pHW197-Mmut. The two mutations also introduced a HindIII and a PvuII restriction site, respectively. These plasmids and the plasmids encoding the other seven PR8 genome segments were transformed and amplified in *E. coli* DH5α. Plasmid DNA was isolated with the Plasmid Midi Kit (Qiagen) according to the manufacturer's instructions. The resulting air-dried pellet was dissolved in 50 µl of sterile ultrapure water. The presence of the introduced mutations in pHW197-Mmut was confirmed by restriction analysis and Sanger sequencing on a capillary sequencer (Applied Biosystems 3730XL DNA Analyzer).

Generation of recombinant PR8 and PR8mut viruses.

To generate recombinant wild type PR8 virus and PR8 virus with the two tracer mutations in the M gene (PR8mut), 1 µg of pHW191-PB2, pHW192-PB1, pHW193-PA, pHW194-HA, pHW195-NP, pHW196-NA and pHW198-NS, together with 1 µg of pHW197-M (wild type PR8) or pHW197-Mmut (PR8mut) was transfected using calcium phosphate co-precipitation into a HEK293T-MDCK cell co-culture in Opti-MEM (3 x 10^5 HEK293T and 2 x 10^5 MDCK cells in a 6-well plate). After 30 h, L-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin (Sigma) was added to a final concentration of 2 µg/ml. After 72 h, the culture medium was collected and the presence of virus was confirmed by hemagglutination of chicken red blood cells. Reverse genetics-generated PR8 and PR8mut viruses were plaque-purified on MDCK cells as follows. Confluent MDCK cells in a six-well plate were infected with a serial dilution series of virus. After 1 h, an overlay of low melting agarose (Type VII agarose, Sigma; final concentration 1%) in serum-free cell culture medium containing 2 µg/ml TPCK-treated trypsin (Sigma) was added. After 56 h, cytopathic effect was checked, agar overlaying viral plaques were selected with a pipette tip, and virus was allowed to diffuse from the

agar for 24 h at 4°C in serum-free medium. Afterwards, virus derived from one plaque was amplified on MDCK cells in serum-free cell culture medium in the presence of 2 µg/ml TPCK-treated trypsin (Sigma). After 96 h, the culture medium was collected, and cell debris was removed by centrifugation for 10 min at 2500 g at 4°C, and the virus was pelleted from the supernatants by overnight centrifugation at 16,000 g at 4°C. The pellet was dissolved in sterile 20% glycerol in PBS, aliquoted and stored at −80°C. The infectious titer of the obtained PR8 and PR8mut virus stocks was determined by plaque assay on MDCK cells, on three different aliquots each performed in triplicate. The presence of the introduced mutations in the M segment of PR8mut was confirmed by segment-7-specific RT-PCR followed by purification from 1% agarose gel (High Pure PCR Product Purification Kit, Roche) and conventional Sanger sequencing of the amplified PCR fragment.

Plaque assay.
MDCK cells were seeded in complete DMEM in 12-well plates at 3×10^5 cells per well. After 18 h, the cells were washed once with serum-free medium and incubated (in triplicate) with a two-fold dilution series of the virus (made in serum-free cell culture medium containing 0.1% BSA) in 500 µl medium. After 1 h incubation at 37°C, an overlay of 500 µl of 1.6% Avicel RC-591 (FMC Biopolymer) in serum-free medium with 4 µg/ml TPCK-treated trypsin (Sigma) was added. After incubation at 37°C for 48 h, the overlay was removed and the cells were fixed with 4% paraformaldehyde and permeabilized with 20 mM glycine and 0.5% (v/v) Triton X-100. Plaques were stained with an anti-M2e IgG1 mouse monoclonal antibody (final concentration 0.5 µg/ml) followed by a secondary anti-mouse IgG horseradish peroxidase (HRP)-linked antibody (GE Healthcare). After washing, TrueBlue peroxidase substrate (KPL) was used to visualize the plaques.

RNA isolation.
RNA was isolated with the High Pure RNA Isolation Kit (Roche) according to the manufacturer's instructions, excluding the DNase I digestion step. In brief, a 200-µl sample containing 1×10^7 PFU of stock virus in serum-free cell culture medium with 0.1% BSA was combined with 400 µl lysis-binding buffer and mixed by vortexing. The mixture was loaded on a two-layered glass fiber column. After binding to the column and washing, the RNA was eluted in 50 µl elution buffer (water, PCR grade).

RT-PCR.
Primers used for cDNA synthesis and PCR were designed based on the 5' and 3' conserved ends of the influenza A genomic segments and contain an additional sequence of 20 nucleotides at their 5' end necessary for PCR amplification [70]. cDNA was generated using the Transcriptor First Strand cDNA Synthesis Kit (Roche). Reverse transcription was performed with the Transcriptor Reverse Transcriptase (10 U, Roche), using 12.5 µl RNA, 2.5 µM CommonUni12G primer (GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG), 1x Transcriptor Reverse Transcriptase Reaction Buffer, 20 U Protector RNAse inhibitor and 4 mM dNTPs, in a total volume of 20 µl. The components were mixed, and the reaction was incubated for 15 min at 42°C, 15 min at 55°C, 5 min at 60°C, and finally 5 min at 85°C to inactivate the reverse transcriptase. Ten microliters of the resulting cDNA

132

sample was amplified in a 100-μl PCR reaction using 2 U Phusion High Fidelity polymerase (Thermo Scientific), 0.2 μM CommonUni12G and CommonUni13 (GCCGGAGCTCTGCAGATATCAGTAGAAACAAGG), 0.2 mM dNTPs, and 1x High-Fidelity buffer. Thermocycling was performed in a PTC-200 Thermal Cycler (MJ Research) with the following conditions: initial denaturation for 30 s at 98°C, 25 cycles of 10 s at 98°C followed by 7.5 min at 72°C, and a final elongation step of 7 min at 72°C. PCR products were purified using the High Pure PCR Product Purification kit (Roche) according to the manufacturer's instructions, and the product was eluted in 50 μl sterile ultrapure water (preheated to 65°C). One microgram of the product was analyzed by agarose gel electrophoresis (1.5% agarose gel) followed by ethidium bromide staining.

Illumina MiSeq sequence determination.
We used 0.5 ng of purified plasmid or RT-PCR sample and the Nextera XT DNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions to generate multiplexed paired-end sequencing libraries. Sequencing libraries were generated in duplicate, meaning that from each plasmid or RT-PCR sample two libraries were prepared in parallel and sequenced on the same Illumina MiSeq sequencing chip. In brief, DNA samples were fragmented and tagged with adapters by Nextera XT transposase. These adaptor ligated DNA fragments were amplified by a limited-cycle PCR program (12 cycles) to add the barcodes and sequences required for subsequent cluster formation. The resulting fragments were purified and simultaneously size-selected by using 0.6x AMpure beads. Fragments were analyzed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies) before loading on the sequencing chip. The fragment lengths showed a negatively skewed distribution with a peak at approximately 700−1000 bases. From the optimized RT-PCR products, also 500 ng was sheared with an M220 focused-ultrasonicator (Covaris) set to obtain peak fragment lengths of 300−400 bp. Next, the NEBNext Ultra DNA Library Preparation kit (New England Biolabs) was used to repair the ends and to add Illumina MiSeq-compatible barcode adapters to 100 ng of fragmented DNA. The resulting fragments were size-selected using Agencourt AMPure XP bead sizing (Beckman Coulter). Afterwards, indexes were added in a limited-cycle PCR (10 cycles), followed by purification on Agencourt AMpure XP beads. Fragments were analyzed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies) before loading on the sequencing chip. Equimolar amounts of normalized libraries were combined and diluted 25-fold in hybridization buffer. The multiplex sample was heat denatured for 2 min at 96°C before loading on the MiSeq chip. After the 2x250 bp MiSeq paired-end sequencing run, the data were base called and reads with the same barcode were collected and assigned to a sample on the instrument, which generated Illumina FASTQ files (Phred +64 encoding). These files were imported in the CLC Genomics Workbench software (CLC Bio, Qiagen). During import in CLC Genomics Workbench, the uncallable ends of the MiSeq reads (B in input file) were automatically trimmed and the failed reads (Y in header information for the quality score) were removed.

Ion Torrent PGM 318 chip sequence determination.

Samples for sequence analysis were generated in duplicate, meaning that from each plasmid or RT-PCR sample two libraries were prepared in parallel for sequencing on the same Ion Torrent PGM 318 sequencing chip. From each plasmid or RT-PCR product, 100 ng was sheared with an M220 focused-ultrasonicator (Covaris) set to obtain peak fragment lengths of 400–500 bp. After shearing, blunt ends were created using the end repair enzyme from the Ion Plus Fragment Library kit (Life Technologies). Next, the fragments were ligated to Ion Torrent PGM-compatible barcode adapters. Since the adaptors are not 5' phosphorylated, the nick repair polymerase in the kit repairs subsequently the nick on one strand at each ligation site, in order to minimize adaptor-dimer formation. We purified and simultaneously size-selected the adapter-ligated library using Agencourt AMPure XP bead sizing (Beckman Coulter). Fragments were analyzed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies); the fragment length peak was situated around 450 bp. Barcoded libraries were pooled in equimolar amounts. From the resulting diluted multiplexed library, 20 µl was loaded on an Ion OneTouch 2 instrument (Life Technologies) to perform emulsion PCR on Ion Sphere particles using the Ion PGM Template OT2 400 kit. We used the Ion PGM sequencing 400 kit (Life Technologies) to sequence templated ion sphere particles deposited in the Ion 318 chip v2 (revision 2.0, Life Technologies). The Ion Torrent Suite version 4.6 (Life Technologies) was used with the default parameters for base calling and assigning of the reads to a sample based on their barcode. The default settings in the Ion Torrent Suite already filter and trim the sequencing reads to some extent. These default trimming parameters are not stringent and remove only very low quality 3' ends (mean Phred score of at least 15 in a base window of 30) and adaptor contamination. The resulting FASTQ files were imported into CLC Genomics Workbench for further analysis.

Analysis of sequencing data.

CLC Genomics Workbench version 7.0.3 (CLC Bio, Qiagen) was used to analyze and process the sequencing reads of both the Ion Torrent PGM and the Illumina MiSeq. First, adaptor contamination was removed from the reads. Next, the sequencing reads were trimmed from both sides using the modified Mott trimming algorithm to reach a Q20 score, which means that the chance that a particular base in the sequence is called incorrectly by the sequencer is 1 in 100. Afterwards, all ambiguous (N) bases were trimmed from the reads. We also removed the reads with a read length below 50. For the Illumina MiSeq, the broken pairs resulting from trimming and filtering were also removed. The remaining reads were assembled using default settings for de novo assembly. In addition, the processed reads were also aligned with the pHW197-M plasmid reference sequence or the influenza PR8 reference genome (based on the sequences encoding the eight segments in the pHW vectors, determined by Sanger sequencing, with addition of the extra 20 nucleotides present at the 5' site in the RT-PCR primers) using local alignment. For this, the following default penalties were used: match = +1, mismatch = -2, insertion/deletion = -3, filtering threshold: length fraction = 0.9 and similarity fraction = 0.8. Non-specific matches, defined as reads aligning to more than one position with an equally good score, were ignored. Sequence variants were called using all available

134

sequencing data that covered each nucleotide at least 100 times and had a central base quality score of Q20 or greater. The A-to-G variant introduced by the primer at position 24 in the HA, NP, NA, M and NS segments was not taken into account during the influenza quasispecies variant analysis. All numerical data mentioned in the text are presented as averages with their standard deviations (± SD).

Statistical analysis.

Sequence variants with the lowest proportion were considered as minor alleles. Analysis of minor allele proportions was performed by fitting a logistic regression model of the form logit(p) = constant + PLATFORM*VIRUS*CLASS+error, where p indicates the minor allele proportion, PLATFORM refers to the sequencing platform, VIRUS refers to virus population, and CLASS refers to class of loci having either low (<15%) or high (>15%) minor variant frequencies. Significance of the fixed PLATFORM, VIRUS and CLASS effects was assessed by an F-test. Significance of pair-wise comparisons between mean proportions was assessed by a Fisher's protected least significance difference test. The logistic regression and assessment of significance was performed in Genstat v16.

Sequencing data.

The output sequencing reads obtained on the Illumina MiSeq and Ion Torrent PGM were submitted to NCBI's Sequence Read Archive and can be found under project numbers SRP052608 (plasmid samples) and SRP052225 (viral samples).

COMPETING INTERESTS

The authors declared that they have no competing interests.

ACKNOWLEDGEMENTS

REFERENCES

1.	Eigen M: Molecular self-organization and the early stages of evolution. *Experientia* 1971, 27(11):149-212.
2.	Eigen M, Schuster P: The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 1977, 64(11):541-565.
3.	Eigen M: Viral quasispecies. *Sci Am* 1993, 269(1):42-49.
4.	Eigen M: On the nature of virus quasispecies. *Trends Microbiol* 1996, 4(6):216-218.
5.	Lauring AS, Andino R: Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 2010, 6(7):e1001005.
6.	Domingo E, Martinez-Salas E, Sobrino F, de la Torre JC, Portela A, Ortin J, Lopez-Galindez C, Perez-Brena P, Villanueva N, Najera R *et al*: The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance--a review. *Gene* 1985, 40(1):1-8.
7.	Nowak MA: What is a quasispecies? *Trends Ecol Evol* 1992, 7(4):118-121.
8.	Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R: Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006, 439(7074):344-348.
9.	Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR *et al*: Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* 2009, 326(5953):734-736.
10.	Schmolke M, Garcia-Sastre A: Evasion of innate and adaptive immune responses by influenza A virus. *Cell Microbiol* 2010, 12(7):873-880.
11.	Wright P, Webster R. In: *Fields Virology* Edited by Knipe D, Howley P, Fourth edn: Lippincott Williams & Wilkins, Philadelphia; 2001: 1533-1579.
12.	Nobusawa E, Sato K: Comparison of the mutation rates of human influenza A and B viruses. *J Virol* 2006, 80(7):3675-3678.
13.	Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R: Viral mutation rates. *J Virol* 2010, 84(19):9733-9748.
14.	Parvin JD, Moscona A, Pan WT, Leider JM, Palese P: Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J Virol* 1986, 59(2):377-383.
15.	Baz M, Abed Y, McDonald J, Boivin G: Characterization of multidrug-resistant influenza A/H3N2 viruses shed during 1 year by an immunocompromised child. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006, 43(12):1555-1561.
16.	Robertson JS, Nicolson C, Bootman JS, Major D, Robertson EW, Wood JM: Sequence analysis of the haemagglutinin (HA) of influenza A (H1N1) viruses present in clinical material and comparison with the HA of laboratory-derived virus. *The Journal of general virology* 1991, 72 (Pt 11):2671-2677.
17.	Iqbal M, Xiao H, Baillie G, Warry A, Essen SC, Londt B, Brookes SM, Brown IH, McCauley JW: Within-host variation of avian influenza viruses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2009, 364(1530):2739-2747.
18.	Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, Newton R, Parrish CR, Hoelzer K, Dougan G *et al*: Intra- and interhost evolutionary dynamics of equine influenza virus. *Journal of virology* 2010, 84(14):6943-6954.
19.	Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J *et al*: Updating benchtop sequencing performance comparison. *Nat Biotechnol* 2013, 31(4):294-296.
20.	Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ: Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012, 30(5):434-439.
21.	Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012, 13:341.
22.	Robasky K, Lewis NE, Church GM: The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* 2014, 15(1):56-62.
23.	Glenn TC: Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 2011, 11(5):759-769.
24.	Croville G, Soubies SM, Barbieri J, Klopp C, Mariette J, Bouchez O, Camus-Bouclainville C, Guerin JL: Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *Journal of clinical microbiology* 2012, 50(9):2881-2887.

136

25.   Hoper D, Hoffmann B, Beer M: A comprehensive deep sequencing strategy for full-length genomes of influenza A. *PLoS One* 2011, 6(4):e19075.

26.   Lin Z, Farooqui A, Li G, Wong GK, Mason AL, Banner D, Kelvin AA, Kelvin DJ, Leon AJ: Next-generation sequencing and bioinformatic approaches to detect and analyze influenza virus in ferrets. *J Infect Dev Ctries* 2014, 8(4):498-509.

27.   Rutvisuttinunt W, Chinnawirotpisan P, Simasathien S, Shrestha SK, Yoon IK, Klungthong C, Fernandez S: Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform. *J Virol Methods* 2013, 193(2):394-404.

28.   Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, O'Connor DH, Hughes AL, Neumann G, Kawaoka Y *et al*: Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nat Commun* 2013, 4:2636.

29.   Zhou B, Lin X, Wang W, Halpin RA, Bera J, Stockwell TB, Barr IG, Wentworth DE: Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics. *Journal of clinical microbiology* 2014, 52(5):1330-1337.

30.   Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG: A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc Natl Acad Sci U S A* 2000, 97(11):6108-6113.

31.   Hoffmann E, Krauss S, Perez D, Webby R, Webster RG: Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine* 2002, 20(25-26):3165-3170.

32.   Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM: An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One* 2013, 8(12):e85024.

33.   Dohm JC, Lottaz C, Borodina T, Himmelbauer H: Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 2008, 36(16):e105.

34.   Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. *Genome Biol* 2013, 14(5):R51.

35.   Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS: Tn5/IS50 target recognition. *Proc Natl Acad Sci U S A* 1998, 95(18):10716-10721.

36.   Minoche AE, Dohm JC, Himmelbauer H: Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 2011, 12(11):R112.

37.   Agresti A, Coull BA: Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998, 52(2):119-126.

38.   Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H *et al*: Sequence-specific error profile of Illumina sequencers. *Nucleic acids research* 2011, 39(13):e90.

39.   Allhoff M, Schonhuth A, Martin M, Costa IG, Rahmann S, Marschall T: Discovering motifs that induce sequencing errors. *BMC bioinformatics* 2013, 14 Suppl 5:S1.

40.   Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000, 407(6803):513-516.

41.   Adeyefa CA, Quayle K, McCauley JW: A rapid method for the analysis of influenza virus genes: application to the reassortment of equine influenza virus genes. *Virus research* 1994, 32(3):391-399.

42.   Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008, 18(5):821-829.

43.   Compeau PE, Pevzner PA, Tesler G: How to apply de Bruijn graphs to genome assembly. *Nature biotechnology* 2011, 29(11):987-991.

44.   Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X *et al*: Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 2010, 11(12):R119.

45.   Ellebedy AH, Webby RJ: Influenza vaccines. *Vaccine* 2009, 27 Suppl 4:D65-68.

46.   Dreger M, Leung BW, Brownlee GG, Deng T: A quantitative strategy to detect changes in accessibility of protein regions to chemical modification on heterodimerization. *Protein Sci* 2009, 18(7):1448-1458.

47.   Gen F, Yamada S, Kato K, Akashi H, Kawaoka Y, Horimoto T: Attenuation of an influenza A virus due to alteration of its hemagglutinin-neuraminidase functional balance in mice. *Arch Virol* 2013, 158(5):1003-1011.

48. Chen LM, Blixt O, Stevens J, Lipatov AS, Davis CT, Collins BE, Cox NJ, Paulson JC, Donis RO: In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology* 2012, 422(1):105-113.

49. Nobusawa E, Ishihara H, Morishita T, Sato K, Nakajima K: Change in receptor-binding specificity of recent human influenza A viruses (H3N2): a single amino acid change in hemagglutinin altered its recognition of sialyloligosaccharides. *Virology* 2000, 278(2):587-596.

50. Kong W, Liu L, Wang Y, Gao H, Wei K, Sun H, Sun Y, Liu J, Ma G, Pu J: Hemagglutinin mutation D222N of the 2009 pandemic H1N1 influenza virus alters receptor specificity without affecting virulence in mice. *Virus Res* 2014, 189:79-86.

51. Mochalova L, Gambaryan A, Romanova J, Tuzikov A, Chinarev A, Katinger D, Katinger H, Egorov A, Bovin N: Receptor-binding properties of modern human influenza viruses primarily isolated in Vero and MDCK cells and chicken embryonated eggs. *Virology* 2003, 313(2):473-480.

52. Zhang W, Shi Y, Qi J, Gao F, Li Q, Fan Z, Yan J, Gao GF: Molecular basis of the receptor binding specificity switch of the hemagglutinins from both the 1918 and 2009 pandemic influenza A viruses by a D225G substitution. *J Virol* 2013, 87(10):5949-5958.

53. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W: The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* 1982, 31(2 Pt 1):417-427.

54. Biswas SK, Boutz PL, Nayak DP: Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J Virol* 1998, 72(7):5493-5501.

55. Elton D, Medcalf E, Bishop K, Digard P: Oligomerization of the influenza virus nucleoprotein: identification of positive and negative sequence elements. *Virology* 1999, 260(1):190-200.

56. Rimmelzwaan GF, Kreijtz JH, Bodewes R, Fouchier RA, Osterhaus AD: Influenza virus CTL epitopes, remarkably conserved and remarkably variable. *Vaccine* 2009, 27(45):6363-6365.

57. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW: The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 2009, 458(7240):914-918.

58. Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T *et al*: Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* 2009, 458(7240):909-913.

59. Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, Nagata K, Tame JR, Park SY: The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* 2008, 454(7208):1127-1131.

60. Myers JL, Wetzel KS, Linderman SL, Li Y, Sullivan CB, Hensley SE: Compensatory hemagglutinin mutations alter antigenic properties of influenza viruses. *J Virol* 2013, 87(20):11168-11172.

61. Skehel JJ, Waterfield MD: Studies on the primary structure of the influenza virus hemagglutinin. *Proc Natl Acad Sci U S A* 1975, 72(1):93-97.

62. Daniels RS, Downie JC, Hay AJ, Knossow M, Skehel JJ, Wang ML, Wiley DC: Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* 1985, 40(2):431-439.

63. Cross KJ, Wharton SA, Skehel JJ, Wiley DC, Steinhauer DA: Studies on influenza haemagglutinin fusion peptide mutants generated by reverse genetics. *EMBO J* 2001, 20(16):4432-4442.

64. Roedig JV, Rapp E, Hoper D, Genzel Y, Reichl U: Impact of host cell line adaptation on quasispecies composition and glycosylation of influenza A virus hemagglutinin. *PLoS One* 2011, 6(12):e27989.

65. Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA: Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J Virol* 2008, 82(21):10493-10501.

66. Man S, Newberg MH, Crotzer VL, Luckey CJ, Williams NS, Chen Y, Huczko EL, Ridge JP, Engelhard VH: Definition of a human T cell epitope from influenza A non-structural protein 1 using HLA-A2.1 transgenic mice. *Int Immunol* 1995, 7(4):597-605.

67. Qian XY, Alonso-Caplen F, Krug RM: Two functional domains of the influenza virus NS1 protein are required for regulation of nuclear export of mRNA. *J Virol* 1994, 68(4):2433-2441.

68. Neumann G, Hughes MT, Kawaoka Y: Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J* 2000, 19(24):6751-6758.

69. Baillie GJ, Galiano M, Agapow PM, Myers R, Chiam R, Gall A, Palser AL, Watson SJ, Hedge J, Underwood A *et al*: Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J Virol* 2012, 86(1):11-18.

70.   Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P: Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci* 2013, 368(1614):20120205.

71.   Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE: Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *J Virol* 2009, 83(19):10309-10313.

72.   Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB *et al*: Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 2012, 30(1):78-82.

73.   Ratan A, Miller W, Guillory J, Stinson J, Seshagiri S, Schuster SC: Comparison of sequencing platforms for single nucleotide variant calls in a human sample. *PLoS One* 2013, 8(2):e55089.

74.   Lee CY, Kam YW, Fric J, Malleret B, Koh EG, Prakash C, Huang W, Lee WW, Lin C, Lin RT *et al*: Chikungunya virus neutralization antigens and direct cell-to-cell transmission are revealed by human antibody-escape mutants. *PLoS Pathog* 2011, 7(12):e1002390.

75.   Brooke CB, Ince WL, Wrammert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW: Most influenza a virions fail to express at least one essential viral protein. *J Virol* 2013, 87(6):3155-3162.

76.   Gambaryan AS, Robertson JS, Matrosovich MN: Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. *Virology* 1999, 258(2):232-239.

77.   Li Z, Watanabe T, Hatta M, Watanabe S, Nanbo A, Ozawa M, Kakugawa S, Shimojima M, Yamada S, Neumann G *et al*: Mutational analysis of conserved amino acids in the influenza A virus nucleoprotein. *J Virol* 2009, 83(9):4153-4162.

78.   de Wit E, Spronken MI, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA: Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus Res* 2004, 103(1-2):155-161.

Chapter 6

Illumina MiSeq sequencing disfavours a sequence motif in the GFP reporter gene

Illumina MiSeq sequencing disfavours a sequence motif in the GFP reporter gene

Silvie Van den Hoecke[1,2], Judith Verhelst[1,2] and Xavier Saelens[1,2,*]

[1] Medical Biotechnology Center, VIB, Ghent, B-9052, Belgium
[2] Department of Biomedical Molecular Biology, Ghent University, Ghent, B-9052, Belgium
* Correspondence: xavier.saelens@vib-ugent.be

Relative contributions of the authors:
SVDH performed the experiments and performed the data analysis. SVDH and XS designed the experiments. XS and JV carried out scientific supervision. SVDH, JV and XS co-wrote the manuscript. All authors read and approved the final manuscript.

Abstract

Green fluorescent protein (GFP) is one of the most used reporter genes. We have used next-generation sequencing (NGS) to analyse the genetic diversity of a recombinant influenza A virus that expresses GFP and found a remarkable coverage dip in the GFP coding sequence. This coverage dip was present when virus-derived RT-PCR product or the parental plasmid DNA was used as starting material for NGS and regardless of whether Nextera XT transposase or Covaris shearing was used for DNA fragmentation. Therefore, the sequence coverage dip in the GFP coding sequence was not the result of emerging GFP mutant viruses or a bias introduced by Nextera XT fragmentation. Instead, we found that the Illumina MiSeq sequencing method disfavours the 'CCCGCC' motif in the GFP coding sequence.

Introduction

Influenza viruses are important human and animal pathogens that have evolved numerous mechanisms to subvert their host's innate and adaptive immune system. Recombinant influenza viruses that express a reporter gene are thus very useful to study viral replication, spread and cell tropism *in vitro* and *in vivo.* The use of such reporter viruses can also facilitate the discovery of new antivirals and vaccines [1-4]. However, adding a reporter gene to the relatively small influenza virus genome has no selective advantage for the virus. Instead, influenza viruses expressing a reporter gene are attenuated compared to their parental counterparts [5-9]. Influenza viruses that have lost (part of) the reporter gene can thus quickly outgrow the original reporter virus. Such reporter gene loss can *e.g.* lead to false negative hits in a compound screening experiment that is based on reporter gene expression as a read out. It is therefore important to study the genomic stability of the viral population derived from a recombinant influenza virus clone. Next-generation sequencing (NGS) is very well suited to determine the genomic stability of recombinant influenza viruses due to its high sequencing output (up to hundreds of Gigabases) [10, 11]. In addition, the small genomic size, approximately 14,000 bases of negative stranded RNA, of influenza viruses enables sequencing of viral samples at high coverage for each position in the genome. We recently optimized an influenza RT-PCR protocol and NGS data analysis pipeline to study the genomic composition of an influenza A virus population [12].

We previously reported the generation and characterization of a recombinant influenza A virus that expresses GFP [1]. In that virus, named PR8-NS1(1-73)GFP, the GFP transgene is encoded by the middle part of a tri-cistronic gene segment 8 [1]. The virus is phenotypically stable and appeared to be genetically stable based on lllumina MiSeq sequencing of full genome RT-PCR products of this virus [1]. However, we noticed a twofold drop in sequence coverage within the GFP coding sequence (Figure 1, orange line) [1]. This coverage dip could be the result of different processes. First, a proportion of the PR8-NS1(1-73)GFP progeny virus might have lost part of the GFP coding sequence. Second, sequence preference of the transposase-based Nextera XT fragmentation could account for

the sequence coverage dip [13-15]. A less favourable sequence motif in the GFP coding region could lead to a fragmentation bias and hence lower sequence coverage [13-16]. The 'Illumina Nextera XT DNA library preparation kit' fragments the DNA and adds the desired sequencing adaptors in a single step by using a transposition reaction, a process that is named tagmentation [13]. The transposase is target sequence based and, as a consequence, near-random [13-17]. Finally, sequencing bias by the Illumina MiSeq sequencer itself, due to a motif in the GFP sequence, could explain the drop in sequence coverage that we observed. It has been shown that the Illumina MiSeq exhibits sequencing biases for different sequence types, *e.g.* in regions with a low or high GC-content, long homopolymers or inverted repeats [17-19].

Here, we report that a sequence motif in the GFP coding sequence leads to a significant reduction in sequence coverage when using Illumina MiSeq sequencing. This finding is important for NGS analysis of small microbial genomes, in particular when GFP is included as a reporter in those genomes.

Results

An NGS coverage dip in the GFP sequence irrespective of the fragmentation method used
We previously used Nextera XT tagmentation followed by NGS on an Illumina MiSeq sequencing platform to study the genetic heterogeneity of a GFP-expressing influenza A virus [1]. Sequence coverage was high and homogenous for all eight virus genome segments, with the exception of the 5' and 3' termini [1]. The latter is expected when using a transposon-based fragmentation technique to make a library of fragments derived from a discrete set of relatively small linear double stranded DNA molecules [12, 13]. However, we noticed a twofold drop in sequence coverage from over 40,000 to almost 20,000 reads per position near the middle of the NS1(1-73)-GFP segment (Figure 1, orange line) [1]. The PR8-NS1(1-73)GFP virus retained GFP expression over multiple rounds of replication *in vitro*, suggesting that the coverage dip was unlikely the result of the rapid evolution of a subpopulation of viruses that had lost part of the GFP information [1]. We first explored the possibility that this apparent coverage dip could be the result of the sequence dependency of Nextera XT fragmention, which has a known target sequence bias [13-15]. Therefore, we repeated the Illumina MiSeq NGS analysis of the PR8-NS1(1-73)GFP virus using Covaris shearing for the library preparation. This is a mechanical shearing technique that is based on adaptive focused acoustics, and therefore more random. We used the same RT-PCR sample of the PR8-NS1(1-73)GFP virus which had been sequenced previously on the Illumina MiSeq after Nextera XT fragmentation [1]. Mapping of the reads to the PR8-NS1(1-73)GFP reference genome resulted in high coverage across all eight segments, which now also included the genome segment ends (Supplementary Figure S1). However, we again observed a decrease in coverage at the same position in the GFP coding sequence (nucleotide position: 452-1162, with the lowest coverage at position 952; Figure 1, black line), similar to the one observed after Nextera XT fragmentation. This indicates that this dip is not caused by the sequence dependency of the transposition reaction in the Nextera XT fragmentation. We note that

144

the ends of the viral fragments are overrepresented after Covaris fragmentation because adaptors are ligated to mechanically sheared DNA, a process that is favored at the free ends of the influenza genome [12].
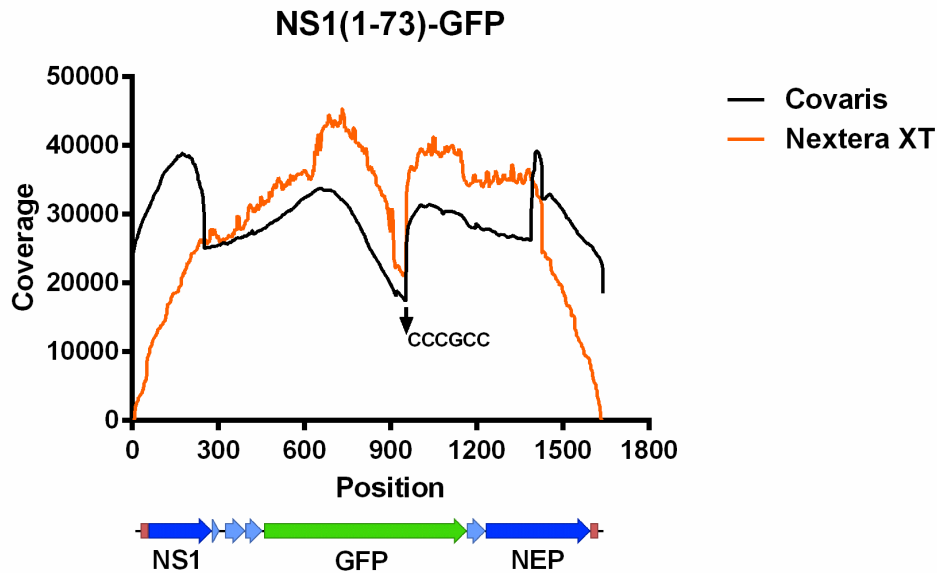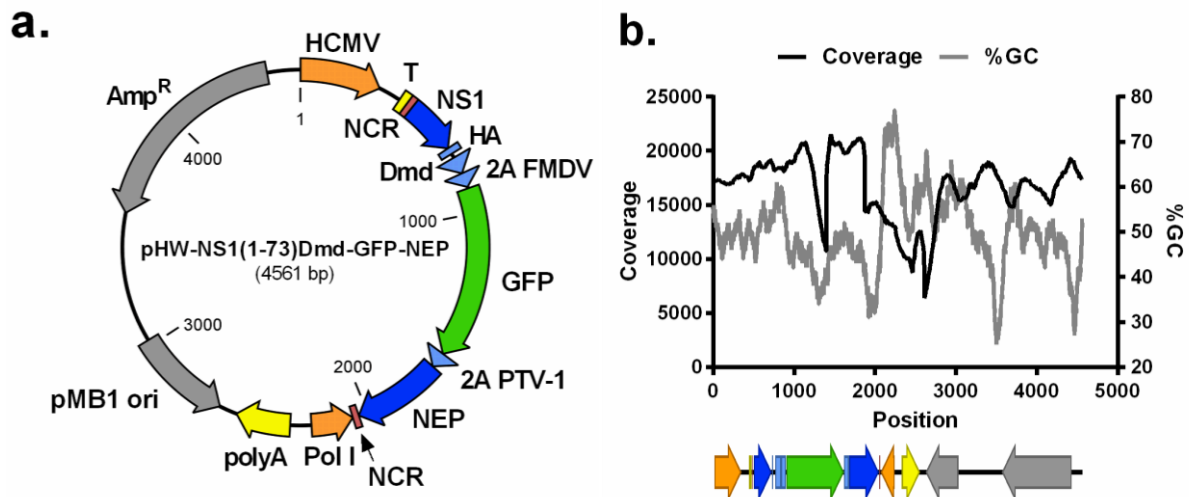

Figure 1: Sequence coverage of the viral NS1(1-73)-GFP segment. Sequence coverage as determined by Illumina MiSeq sequencing after Nextera XT (orange) or Covaris (black) fragmentation and CLC Genomics Workbench version 7.0.3 data processing. The obtained sequences were filtered, trimmed and mapped to the reference sequence of the NS1(1-73)-GFP segment (based on the plasmid used to generate the recombinant PR8-NS1(1-73)GFP virus, with addition of the extra 20 nucleotides present at the 5' site in the RT-PCR primers) [12]. Below sequencing coverage plot: schematic representation of NS1(1-73)-GFP segment. For an explanation of the different features of this segment, see Figure 2.

The above results do not exclude the possibility that viruses with a deletion in the GFP coding sequence were present in the virus population that we used as starting material. We investigated the presence of major deletions in the GFP sequence by using the 'CLC Genomics Workbench Large Gap Mapper', which aligns reads to the reference sequence, while allowing large gaps in the mapping. Based on the 'Large Gap Mapper' 0.22% more reads were aligned to the reference genome, compared to regular mapping. The distribution of these extra mapped reads over the eight segments ranged from 0.04% (M segment) to 0.60% (PA segment). An increase of 0.41% of mapped reads was recorded for the NS1(1-73)-GFP segment. Therefore, fragments with a large deletion were not substantially enriched for the NS1(1-73)-GFP segment, indicating that the dip in coverage was not caused by large deletions in the GFP sequence.

NGS sequencing of the pHW-NS1(1-73)Dmd-GFP-NEP plasmid also reveals a coverage dip
Based on the above analyses it is unlikely that the observed variability in the PR8-NS1(1-73)GFP virus population was responsible for the coverage dip in GFP. We therefore hypothesized that the Illumina MiSeq platform caused the coverage bias in the GFP coding sequence. To test this, we sequenced the pHW-NS1(1-73)Dmd-GFP-NEP plasmid that was used to generate the GFP expressing influenza A

virus. In this way, we could also assess a possible effect of RT-PCR efficacy on the sequencing coverage of the GFP sequence. A mean sequencing coverage of 16,655 (+/- 2,927) was obtained, with the coverage per position ranging from 6,376 (position 2,612) to 21,513 (position 1,451). We observed a twofold drop in nucleotide coverage in the GFP coding sequence of the plasmid, with the lowest coverage being 10,683 at position 1,395 (Figure 2).
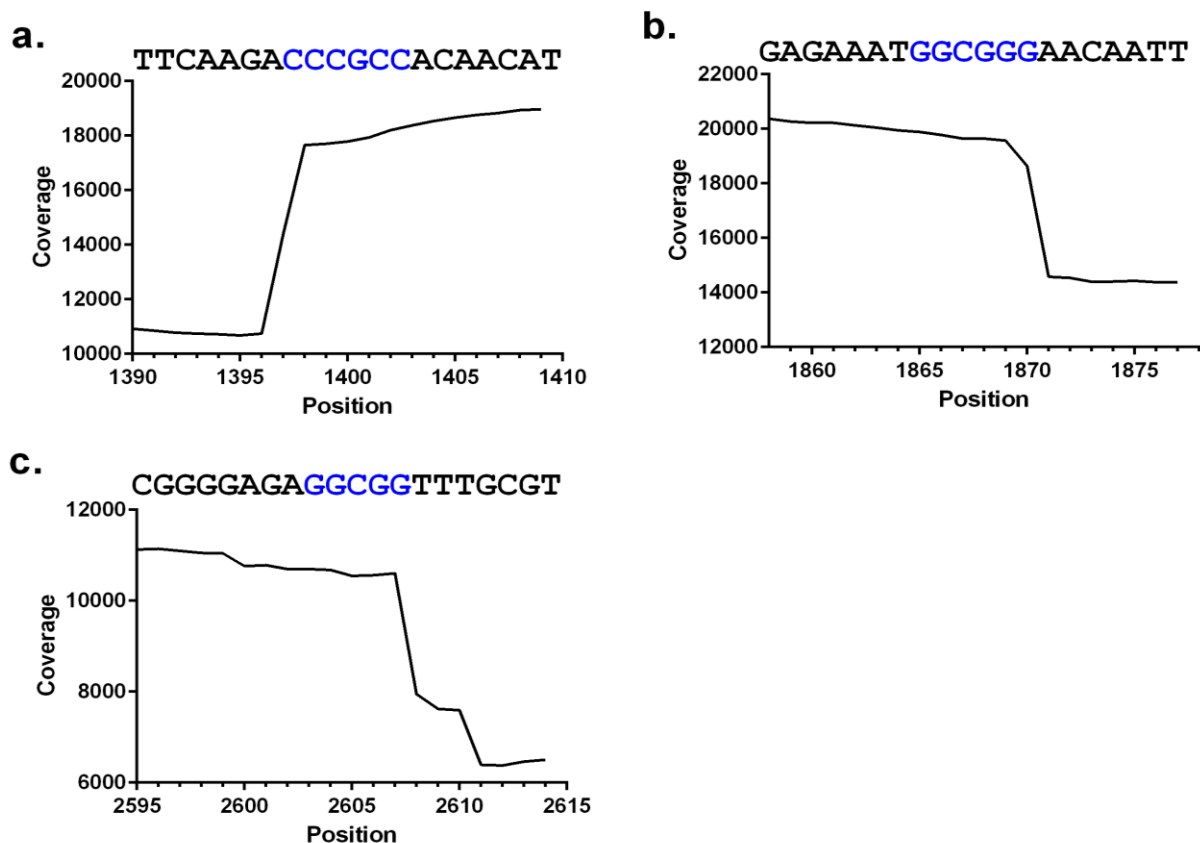

Figure 2: Sequence coverage of pHW-NS1(1-73)Dmd-GFP-NEP based on Illumina MiSeq sequencing. (a) Map of pHW-NS1(1-73)Dmd-GFP-NEP. (b) Sequence coverage (black line) and GC-percentage distribution (grey line, window size: 100) of the pHW-NS1(1-73)Dmd-GFP-NEP plasmid as determined by Illumina MiSeq sequencing after Covaris shearing and CLC Genomics Workbench version 7.0.3 data processing [12]. The diagram below the graph shows the organization of the different features from position 1 to position 4561 in the linearized pHW-NS1(1-73)Dmd-GFP-NEP plasmid. HCMV: human cytomegalovirus promoter, T: terminator sequence, NCR: non-coding region, NS1: non-structural protein 1, HA: hemagglutinin-tag, Dmd: dimerization domain (Dmd) of the *Drosophila melanogaster* Ncd protein, 2A FMDV: foot-and-mouth disease virus-2A auto processing site, 2A PTV-1: porcine teschovirus-1 2A cleavage site, NEP: nuclear export protein, Pol I: human RNA polymerase I promoter, polyA: polyA terminator, pMB1 ori: origin of replication, Amp[R]: ampicillin resistance gene.

It has been reported that the performance of Illumina MiSeq sequencing is reduced in regions that have a high or low GC-content [18, 20]. However, the GFP coding sequence is slightly GC-poor (average 43.18%), compared to the overall GC-percentage of the plasmid (49.55%) (Figure 2). Based on the relation between the GC-content and sequencing coverage at each position, we conclude that there was no strict correlation between GC-content and sequencing coverage.

Before sequencing on the Illumina MiSeq platform, the DNA fragments are ligated on the sequencing chip through their adaptors and subjected to bridge amplification PCR. Presuming that bridge amplification PCR occurs less efficiently when secondary structures are present in the template, the minimal energy to form secondary structures of fragments of 350 bp (approximately the mean DNA fragment length), with a sliding window of 50 bp was calculated using mFold [21]. This minimal energy is inversely correlated with the formation of secondary structures: the lower the minimal energy needed to form a secondary structure, the higher the chance that this structure will be

formed. The mFold calculation predicted that the minimal energy required to form secondary structures is not lower in the GFP coding sequence than the average minimal energy to form secondary structures in the pHW-NS1(1-73)Dmd-GFP-NEP sequence (data not shown). This suggests that the GFP sequence is not more prone to form secondary structures compared to the rest of the plasmid sequence. The dip in coverage in the GFP sequence can therefore not be explained by a less efficient bridge amplification of the fragments containing the GFP sequence.

Figure 3: Drop in sequencing coverage at the 'GGCNGG' or 'GGCNG' motifs. The sequencing coverage is plotted in function of the nucleotide position in the pHW-NS1(1-73)Dmd-GFP-NEP plasmid. The presence of the 'CCCGCC' motif (reverse complement of 'GGCGGG') at position 1397-1402 (a), the 'GGCGGG' motif at position 1865-1870 (b) and the 'GGCGG' motif at position 2603-2607 (c), result in a drop in sequencing coverage.

Sequence-specific errors were previously reported to be common in Illumina HiSeq reads, with the highest error rates seen at the 'GGC' motif, and in particular at 'GGCNG' [22]. In addition, Ekblom *et al.* observed a negative correlation between the site specific sequencing error rate and the sequencing coverage [19]. In particular, they observed a steep drop in coverage exactly at and upstream of the error prone motif 'CCNGCC' (or downstream of its reverse complement 'GGCNGG') [19]. This 'CCNGCC' motif occurs 12 times in pHW-NS1(1-73)Dmd-GFP-NEP. At two of these motifs (positions 1,397-1,402 and 1,865-1,870) a drop in sequencing coverage is observed. Position 1,397-1,402 is within the GFP coding sequence. At position 1,396 (one nucleotide upstream of the 1,397-1,402 'CCCGCC' motif), there was a drop in sequencing coverage from 14,384 (position 1,397) to

10,744 (position 1,396; Figure 3.a). The 'GGCGGG' motif at position 1,865-1,870 was associated with a similar steep sequencing coverage dip (coverage of 18,639 at position 1,870 to a coverage of 14,576 at position 1,871; Figure 3.b). Finally, the error-prone 'GGCGG' motif at position 2,603-2,607 was also associated with a drop in coverage from 10,604 at position 2,607 to 7,947 at position 2,608 (Figure 3.c). We manually inspected the quality trimmed reads with unaligned ends at the 'CCCGCC' motif and found that part of these reads contained unaligned (mainly single) nucleotides at this position. We also inspected the reads manually prior to quality control trimming, *i.e.* with only the adaptor removed. This analysis revealed that most of the reverse reads were of a too low quality at the nucleotides next to the 'CCCGCC' motif and were thus removed during trimming on base quality. Therefore, the steep coverage drop next to the 'CCCGCC' motif results from a combination of poor quality at the error prone motif and actual loss of coverage immediately after the motif.

Since multiple different GFP variants are used to generate reporter RNA viruses, we sequenced four plasmids that encode other GFP variants (Supplementary Figure S2) [23, 24]. A drop in coverage at 'CCNGCC' or the shorter 'CNGCC' motif can also be observed in some of the sequences encoding these GFP variants [23, 24]. Although the 'CCNGCC' motif is absent in the *Aequorea victoria* GFP coding sequence, the shorter 'CNGCC' motif occurs two times in this sequence (positions 820 to 824 and 975 to 979, plasmid numbering). Mapping the reads to the plasmid reference sequence did not reveal a steep drop in sequencing coverage. However, the GFP sequence is not homogenously covered, with the sequencing coverage ranging from 75,529 (minimum; position 1,048) to 108,654 (maximum; position 722). Nevertheless, we found two sharp drops in coverage outside the *A. victoria* GFP coding sequence: upstream of a 'CCNGCC' motif at positions 425 to 430 and downstream of a 'GGCNG' motif at positions 1,704 to 1,708. From the sequenced GFP variants, the largest loss in coverage in the GFP coding sequence is present in eGFP: a 'CCGCC' motif (positions 2,992 – 2,996) results in a coverage drop of 42,730 at position 2,992 to 32,146 at position 2,991. The MaxGFP/TurboGFP that was used in the NS1-GFP influenza virus reported by Manicassamy *et al.,* displays only a minor sequencing drop (Supplementary Figure S2.c) [5].

To provide evidence that the observed dip in coverage is a consequence of the presence of the 'CCCGCC' motif in the GFP coding region, two silent mutations (separate or in combination) were introduced in pHW-NS1(1-73)Dmd-GFP-NEP: C1398T (resulting in CTCGCC), C1401T (resulting in CCCGTC) and the double mutant C1398T - C1401T (resulting in CTCGTC). The two single mutations in the 'CCCGCC' motif largely abolished and the double mutant completely overcame the sequence coverage drop following Illumina MiSeq sequencing (Figure 4). We can thus conclude that the observed drop in sequencing coverage in the GFP coding region can be linked to the 'CCCGCC' motif and that this drop can be eliminated by mutating this motif.
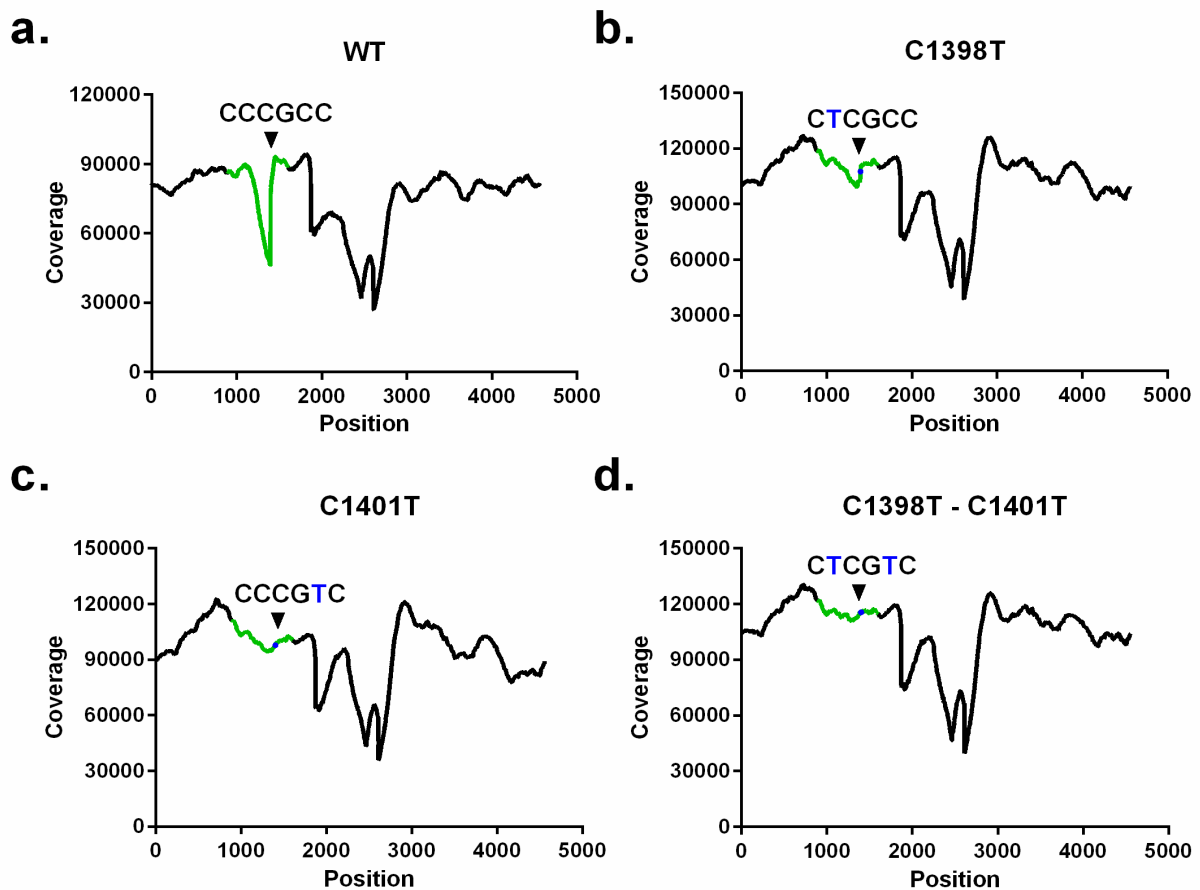
Figure 4: Introducing silent mutations (C1398T and/or C1401T) that interrupt the 'CCCGCC' motif in the GFP coding sequence abrogates the drop in sequence coverage. The sequencing coverage is plotted in function of the nucleotide position in pHW-NS1(1-73)Dmd-GFP-NEP (a), pHW-NS1(1-73)Dmd-GFP-NEP C1398T (b), pHW-NS1(1-73)Dmd-GFP-NEP C1401T (c) or pHW-NS1(1-73)Dmd-GFP-NEP C1398T-C1401T (d) plasmid. Illumina MiSeq sequencing was performed after Covaris shearing and followed by CLC Genomics Workbench version 7.0.3 data processing and mapping of the reads to the plasmid reference sequence [12]. The position of the CCCGCC (a), CTCGCC (b), CCCGTC (c) and CTCGTC (d) motif is marked with an arrow head and the introduced mutation is marked in blue in this motif. The position of the sequence that codes for GFP is marked on the coverage plots in green.

Discussion

NGS analysis is a powerful tool to study nucleotide sequence variation in biological samples. Ideally, such analysis should result in high and unbiased nucleotide coverage across the target region(s) to provide an accurate picture of the real ratio of sequences present. Uneven coverage of sequences can result in the false interpretation of data, *e.g.* as has been reported for transcriptomics analysis [20, 25, 26].

In general, RNA viruses have a relatively high mutation and recombination rate [27, 28]. NGS analysis of the genome diversity of certain RNA viruses, such as influenza A, is used to detect escape

mutations after antiviral treatment or host immunity and to study the viral population dynamics [29-32]. In addition, influenza A viruses with various reporter genes have been generated to facilitate the study of immune responses and cell tropism *in vivo* [1, 5, 6, 33]. Because these studies rely on monitoring the reporter gene products, it is very important to be able to rely on a genetically stable reporter virus. Because the reporter gene does not have a selective advantage for the virus its (partial) deletion in the progeny virus would in most cases offer a competitive advantage over the parental virus. NGS enables sequencing of many viral genomes in a viral population at once. Mapping of these sequencing reads to the reference genome results in a coverage plot, which provides information on the genomic stability of a viral population.

We previously reported on the genomic stability of a GFP expressing influenza A virus that we generated in our lab [1]. Nextera XT tagmentation and Illumina MiSeq sequence analysis of this PR8-NS1(1-73)GFP virus revealed a clear coverage dip in the GFP sequence, which was puzzling because we found that the virus was phenotypically stable over multiple generations [1]. Here, we identified the cause of this GFP-associated coverage drop. This dip was equally apparent when the same sample was analysed after Covaris fragmentation, so it could not be attributed to a sequence preference of the Nextera XT transposase. We also excluded that large deletions in the GFP sequence in the viral population were responsible for the reduced coverage, since NGS analysis of the parental pHW-NS1(1-73)Dmd-GFP-NEP plasmid revealed a similar coverage dip at the same position in the GFP coding sequence. We identified a 'CCNGCC' motif in the GFP coding sequence next to the steep drop in coverage. This motif was recently reported to be associated with more errors in the reads generated by Illumina sequencing [19]. The observed coverage dip in the NS1(1-73)-GFP segment is thus the result of a sequencing bias of the Illumina MiSeq for this 'CCNGCC' motif.

This work shows that caution is needed when analysing samples containing the GFP sequence by NGS. To avoid this sequencing bias a Quantum SuperGlo GFP coding sequence with silent mutations at positions C504T and/or C507T (GFP numbering) should be used. The 'CCNGCC' sequence motif is also present in other GFP versions, *e.g.* the Emerald and ZsGreen1 GFP, which also have been used to generate reporter RNA viruses [34-37]. The MaxGFP (also named TurboGFP) that was used in the NS1-GFP influenza virus reported by Manicassamy *et al.,* displays only a minor sequencing drop (Supplementary Figure S2.c) [5].

When designing reporter viruses it is thus important to take into account that there could be a sequencing bias against the reporter gene used. To prevent such an Illumina MiSeq sequencing bias, it is worthwhile to avoid the presence of the error prone 'CCNGCC' motif in the reporter gene. In the reported PR8-NS1(1-73)GFP virus, this sequencing bias could lead to the false conclusion that the reporter virus is genetically diverse at the GFP coding sequence.

Conclusion

We report a striking variation in coverage depth in the GFP sequence of the PR8-NS1(1-73)GFP virus, as analysed by Illumina MiSeq sequencing. We investigated the different sources that could be responsible for this reduced sequencing coverage and found that a 'CCNGCC' motif in the GFP coding sequence was the cause of the steep drop in sequencing coverage. Since Illumina MiSeq is the most popular NGS platform that is currently used and GFP is widely used as a reporter gene, we believe that this finding is of value for other researchers, in particular for those instances where genetic variability in concert with GFP reporter gene expression are studied.

Material and methods

Plasmids

The cloning strategy used to construct the pHW-NS1(1-73)Dmd-GFP-NEP plasmid has been described in De Baets, *et al*. [1] The C1398T and/or C1401T mutations were introduced by QuickChange site-directed mutagenesis (Stratagene).The plasmids were transformed and amplified in *Escherichia coli* DH5α. Plasmid DNA was isolated with the Plasmid Midi Kit (Qiagen) according to the manufacturer's instructions. The sequence of NS1(1-73)Dmd-GFP-NEP and the introduced C1398T or/and C1401T mutations were confirmed by Sanger sequencing on a capillary sequencer (Applied Biosystems 3730XL DNA Analyzer). Plasmids pBluAGFP [24], pEF6-turboGFP-MCS, pLVX-EF1a-IRES-ZsGreen1 (Clontech-BD Biosciences, Palo Alto, United States) and pDG2-hRIPK4-WT-EGFP-puro [23] were kindly provided by the BCCM/LMBP Plasmid Collection, Dr. Jens Staal and Giel Tanghe from our department.

Cell lines

MDCK, MDCK.PIV5V and HEK293T cells were cultured in DMEM supplemented with 10% FCS, non-essential amino acids, 2 mM L-glutamine, 0.4 mM sodium-pyruvate, 100 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in 5% CO2. MDCK cells stably expressing the type I IFN antagonist Paramyxovirus Simian Virus 5 V protein (MDCK.PIV5V) were kindly provided by Dr. Rick Randall (University of St. Andrews, United Kingdom) [38, 39]. These cell lines were used to rescue and grow PR8-NS1(1-73)GFP virus.

Production of recombinant viruses

Recombinant influenza virus PR8-NS1(1-73)-GFP was rescued using the A/Puerto Rico/8/34 based reverse genetics system [40]. To generate recombinant PR8-NS1(1-73)-GFP virus, 1 µg of each pHW-plasmid (pHW191-PB2, pHW192-PB1, pHW193-PA, pHW194-HA, pHW195-NP, pHW196-NA, pHW197-M and pHW-NS1(1-73)Dmd-GFP-NEP) was transfected in a HEK293T/MDCK coculture using calcium phosphate precipitation in Optimem. After 36 h, TPCK-treated trypsin (Sigma) was added to a final concentration of 2 µg/ml. After 72 h, the medium was collected. The virus in the medium was

amplified on MDCK.PIV5V cells in serum-free cell culture medium in the presence of 2 μg/ml TPCK-treated trypsin (Sigma).

RT-PCR on PR8-NS1(1-73)GFP virus

Total RNA was isolated from 2 x 10^5 PFU of PR8-NS1(1-73)GFP virus with the High Pure RNA isolation Kit (Roche), and cDNA was synthesized with the Transcriptor First Strand cDNA Synthesis kit (Roche), both according to the instructions of the manufacturer. cDNA synthesis was performed with the CommonUni12G (GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG) primer specific for influenza A vRNA. Next, all eight genomic segments were amplified in one reaction with Phusion High Fidelity polymerase (Thermo Scientific) using primers CommonUni12G and CommonUni13 (GCCGGAGCTCTGCAGATATCAGTAGAAACAAGG) [12, 32].

Illumina MiSeq library preparation and sequencing

500 ng of the PR8-NS1(1-73)GFP virus RT-PCR product or the pHW-NS1(1-73)Dmd-GFP-NEP plasmid was sheared with an M220 focused-ultrasonicator (Covaris) set to obtain peak fragment lengths of 300-400 bp. Next, the NEBNext Ultra DNA Library Preparation kit (New England Biolabs) was used to repair the ends and to add the Illumina MiSeq-compatible barcode adapters to 100 ng of fragmented DNA. The resulting fragments were size-selected using Agencourt AMPure XP bead sizing (Beckman Coulter). Afterwards, indexes were added in a limited-cycle PCR (10 cycles), followed by purification on Agencourt AMpure XP beads. Fragments were analysed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies). The multiplex sample was heat denatured for 2 min at 96°C before loading on the Illumina MiSeq chip. After the 2×250 bp Illumina MiSeq paired-end sequencing run, the data were base called and reads with the same barcode were collected and assigned to a sample on the instrument, which generated Illumina FASTQ files (Phred +64 encoding).

Data analysis

The downstream data analyses were performed on the resulting Illumina FASTQ files (Phred +64 encoding) using CLC Genomics Workbench (Version 7.0.3) following the analysis pipeline as described in Van den Hoecke, *et al. [12]* The trimmed and filtered reads were aligned to the PR8-NS1(1-73)GFP reference genome (based on the plasmids used to generate the recombinant PR8 virus, with addition of the extra 20 nucleotides present at the 5′ site in the RT-PCR primers) or the plasmid reference sequence using the following parameters: match = +1; mismatch = -2; insertion/deletion = -3; length fraction = 0.9; similarity fraction = 0.8; non-specific match handling = ignore [12]. For the 'Large Gap Mapper', the same mapping parameters were used, together with the default 'Large Gap Mapper' settings, allowing large gaps in the mapping.

152

Accession codes

The raw sequencing data can be found in the NCBI Sequence Read Archive with the accession numbers SRP052023 (virus sample) and SRP062322 (plasmid samples).

Competing interests

The authors declare no competing financial interests.

References

1. De Baets S, Verhelst J, Van den Hoecke S, Smet A, Schotsaert M, Job ER, Roose K, Schepens B, Fiers W, Saelens X: A GFP expressing influenza A virus to report in vivo tropism and protection by a matrix protein 2 ectodomain-specific monoclonal antibody. *PloS one* 2015, 10(3):e0121491.
2. Eckert N, Wrensch F, Gartner S, Palanisamy N, Goedecke U, Jager N, Pohlmann S, Winkler M: Influenza A virus encoding secreted Gaussia luciferase as useful tool to analyze viral replication and its inhibition by antiviral compounds and cellular proteins. *PloS one* 2014, 9(5):e97695.
3. Kim JI, Park S, Lee I, Lee S, Shin S, Won Y, Hwang MW, Bae JY, Heo J, Hyun HE *et al*: GFP-expressing influenza A virus for evaluation of the efficacy of antiviral agents. *J Microbiol* 2012, 50(2):359-362.
4. Lo MK, Nichol ST, Spiropoulou CF: Evaluation of luciferase and GFP-expressing Nipah viruses for rapid quantitative antiviral screening. *Antiviral research* 2014, 106:53-60.
5. Manicassamy B, Manicassamy S, Belicha-Villanueva A, Pisanelli G, Pulendran B, Garcia-Sastre A: Analysis of in vivo dynamics of influenza virus infection in mice using a GFP reporter virus. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(25):11531-11536.
6. Fukuyama S, Katsura H, Zhao D, Ozawa M, Ando T, Shoemaker JE, Ishikawa I, Yamada S, Neumann G, Watanabe S *et al*: Multi-spectral fluorescent reporter influenza viruses (Color-flu) as powerful tools for in vivo studies. *Nature communications* 2015, 6:6600.
7. Kittel C, Ferko B, Kurz M, Voglauer R, Sereinig S, Romanova J, Stiegler G, Katinger H, Egorov A: Generation of an influenza A virus vector expressing biologically active human interleukin-2 from the NS gene segment. *Journal of virology* 2005, 79(16):10672-10677.
8. Li F, Feng L, Pan W, Dong Z, Li C, Sun C, Chen L: Generation of replication-competent recombinant influenza A viruses carrying a reporter gene harbored in the neuraminidase segment. *Journal of virology* 2010, 84(22):12075-12081.
9. Pena L, Sutton T, Chockalingam A, Kumar S, Angel M, Shao HX, Chen HJ, Li WZ, Perez DR: Influenza Viruses with Rearranged Genomes as Live-Attenuated Vaccines. *Journal of virology* 2013, 87(9):5118-5127.
10. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C: Ten years of next-generation sequencing technology. *Trends Genet* 2014, 30(9):418-426.
11. Ansorge WJ: Next-generation DNA sequencing techniques. *N Biotechnol* 2009, 25(4):195-203.
12. Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X: Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC Genomics* 2015, 16:79.
13. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X *et al*: Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology* 2010, 11(12):R119.
14. Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS: Tn5/IS50 target recognition. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(18):10716-10721.
15. Green B, Bouchier C, Fairhead C, Craig NL, Cormack BP: Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob DNA* 2012, 3(1):3.
16. Marine R, Polson SW, Ravel J, Hatfull G, Russell D, Sullivan M, Syed F, Dumas M, Wommack KE: Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl Environ Microbiol* 2011, 77(22):8071-8079.
17. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012, 13:341.
18. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. *Genome biology* 2013, 14(5):R51.
19. Ekblom R, Smeds L, Ellegren H: Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC Genomics* 2014, 15:467.
20. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic acids research* 2008, 36(16):e105.
21. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 2003, 31(13):3406-3415.

22. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H *et al*: Sequence-specific error profile of Illumina sequencers. *Nucleic acids research* 2011, 39(13):e90.

23. De Groote P. GS, Lippens S., Eichperger C., Leurs K., Kahr I., Tanghe G., Bruggeman I., De Schamphelaire W., Urwyler C., Vandenabeele P., Haustraete J. and W. Declercq.: Generation of a new gateway-compatible, inducible lentiviral vector platform allowing easy derivation of co-transduced cells. *BioTechniques* 2016.

24. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ: Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* 1992, 111(2):229-233.

25. Zheng W, Chung LM, Zhao H: Bias detection and correction in RNA-Sequencing data. *BMC bioinformatics* 2011, 12:290.

26. Li J, Jiang H, Wong WH: Modeling non-uniformity in short-read rates in RNA-Seq data. *Genome biology* 2010, 11(5):R50.

27. Drake JW, Charlesworth B, Charlesworth D, Crow JF: Rates of spontaneous mutation. *Genetics* 1998, 148(4):1667-1686.

28. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R: Viral mutation rates. *Journal of virology* 2010, 84(19):9733-9748.

29. Borderia AV, Isakov O, Moratorio G, Henningsson R, Aguera-Gonzalez S, Organtini L, Gnadig NF, Blanc H, Alcover A, Hafenstein S *et al*: Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS pathogens* 2015, 11(5):e1004838.

30. Isakov O, Borderia AV, Golan D, Hamenahem A, Celniker G, Yoffe L, Blanc H, Vignuzzi M, Shomron N: Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics* 2015, 31(13):2141-2150.

31. Ghedin E, Laplante J, DePasse J, Wentworth DE, Santos RP, Lepow ML, Porter J, Stellrecht K, Lin X, Operario D *et al*: Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *The Journal of infectious diseases* 2011, 203(2):168-174.

32. Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P: Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci* 2013, 368(1614):20120205.

33. Tran V, Moser LA, Poole DS, Mehle A: Highly sensitive real-time in vivo imaging of an influenza reporter virus reveals dynamics of replication and spread. *Journal of virology* 2013, 87(24):13321-13329.

34. Chang Z, Pan J, Logg C, Kasahara N, Roy-Burman P: A replication-competent feline leukemia virus, subgroup A (FeLV-A), tagged with green fluorescent protein reporter exhibits in vitro biological properties similar to those of the parental FeLV-A. *Journal of virology* 2001, 75(18):8837-8841.

35. Teerawanichpan P, Hoffman T, Ashe P, Datla R, Selvaraj G: Investigations of combinations of mutations in the jellyfish green fluorescent protein (GFP) that afford brighter fluorescence, and use of a version (VisGreen) in plant, bacterial, and animal cells. *Biochimica et biophysica acta* 2007, 1770(9):1360-1368.

36. Henrik Gad H, Paulous S, Belarbi E, Diancourt L, Drosten C, Kummerer BM, Plate AE, Caro V, Despres P: The E2-E166K substitution restores Chikungunya virus growth in OAS3 expressing cells by acting on viral entry. *Virology* 2012, 434(1):27-37.

37. Voigt E, Inankur B, Baltes A, Yin J: A quantitative infection assay for human type I, II, and III interferon antiviral activities. *Virology journal* 2013, 10:224.

38. Didcock L, Young DF, Goodbourn S, Randall RE: The V protein of simian virus 5 inhibits interferon signalling by targeting STAT1 for proteasome-mediated degradation. *Journal of virology* 1999, 73(12):9928-9933.

39. Young DF, Andrejeva L, Livingstone A, Goodbourn S, Lamb RA, Collins PL, Elliott RM, Randall RE: Virus replication in engineered human cells that do not respond to interferons. *Journal of virology* 2003, 77(3):2174-2181.

40. Hoffmann E, Krauss S, Perez D, Webby R, Webster RG: Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine* 2002, 20(25-26):3165-3170.

Supplementary Figures
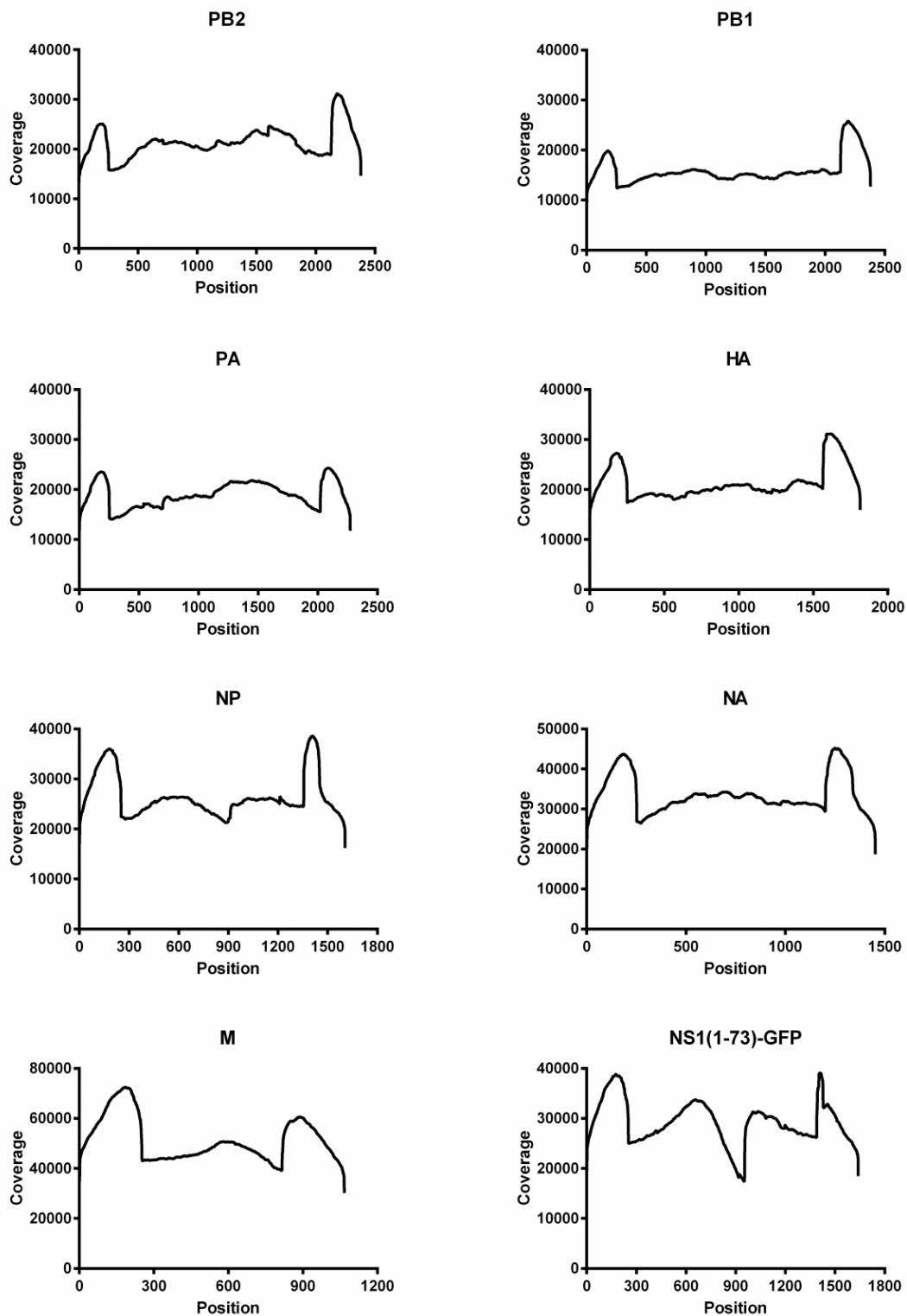
Supplementary Figure S1. Sequence coverage of the PR8-NS1(1-73)GFP virus. Sequence coverage for the different genome segments of the virus stock determined by Illumina MiSeq sequencing after Covaris shearing and CLC Genomics Workbench version 7.0.3 data processing. The obtained sequences were filtered, trimmed and mapped to the reference genome based on the eight plasmids used to generate the recombinant PR8 virus (with addition of the extra 20 nucleotides present at the 5' site in the RT-PCR primers) [12].
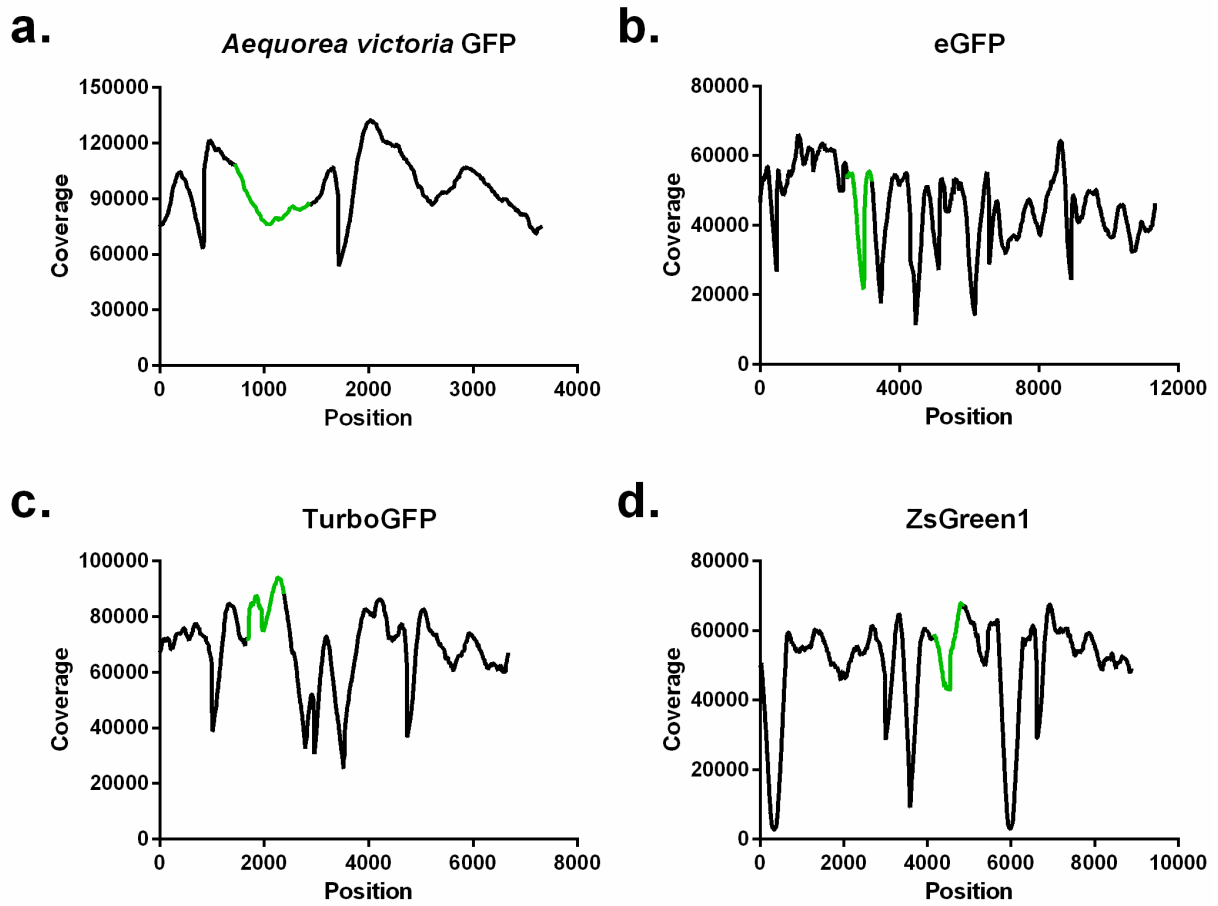
Supplementary Figure S2. Sequence coverage of plasmids expressing different GFP variants. The sequencing coverage is plotted in function of the nucleotide position in the plasmid containing the coding sequence for *Aequorea victoria* GFP (pBluAGFP [24]) (a), eGFP (pDG2-hRIPK4-WT-EGFP-puro [23]) (b), TurboGFP (pEF6-turboGFP-MCS) (c) and ZsGreen1 (pLVX-EF1a-IRES-ZsGreen1) (d). The position of the sequence coding for GFP in the different expression plasmids is marked in green on the coverage plot. Samples were sequenced on Illumina MiSeq (2*250bp) after Covaris shearing, followed by CLC Genomics Workbench version 7.0.3 data processing and mapping of the reads to the plasmid reference sequence [12].

Chapter 7

Hierarchical and redundant roles of activating FcγRs in protection against influenza disease by M2e-specific IgG1 and IgG2a antibodies

Hierarchical and redundant roles of activating FcγRs in protection against influenza disease by M2e-specific IgG1 and IgG2a antibodies

Running title: FcγR-anti-M2e IgG interactions

Silvie Van den Hoecke[a,b], Katrin Ehrhardt[c], Annasaheb Kolpe[a,b], Karim El Bakkouri[a,b,h], Lei Deng[a,b,i], Hendrik Grootaert[a,d], Steve Schoonooghe[e,f], Anouk Smet[a,b], Mostafa Bentahir[a,b,j], Kenny Roose[a,b], Michael Schotsaert[a,b,k], Bert Schepens[a,b], Nico Callewaert[a,d], Falk Nimmerjahn[g], Peter Staeheli[c], Hartmut Hengel[c] and Xavier Saelens[a,b]

[a] Medical Biotechnology Center, VIB, 9052 Ghent, Belgium

[b] Department of Biomedical Molecular Biology, Ghent University, 9052 Ghent, Belgium

[c] Institute of Virology, Medical Center University of Freiburg, 79106 Freiburg, Germany

[d] Department of Biochemistry and Microbiology, Ghent University, 9052 Ghent, Belgium

[e] Laboratory of Myeloid Cell Immunology, VIB, 1050 Brussels, Belgium

[f] Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, 1050 Brussels, Belgium

[g] Department of Biology, University of Erlangen-Nuremberg, 91054 Erlangen, Germany.

[h] Current affiliation: King Fahad Medical City, 11525 Riyadh, Kingdom of Saudi Arabia

[i] Current affiliation: Department of Microbiology and Immunology, School of Medicine, Emory University, GA 30322 Atlanta, USA

[j] Current affiliation: Center for Applied Molecular Technologies, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Institut de Recherche Expérimentale et Clinique, 1200 Brussels, Belgium

[k] Current affiliation: Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

160

Relative contributions of the authors:

SVDH (Table 1, Figure 2, Figure 5 and Figure 7), KE, AK, KEB, LD, HG, SS, AS and MB performed the experiments. SVDH, KE, AK, KEB, HG, SS, NC, FN, PS, HH and XS designed the experiments. KR, MS, BS, NC, FN, PS, HH and XS supervised the research. XS carried out project design. SVDH, KE, AK, KEB, LD, HG, KR, MS, BS, NC, FN, PS, HH and XS co-wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Corresponding author: Xavier Saelens

E-mail: xavier.saelens@vib-ugent.be

Telephone: +32 9 33 13 620, Fax: +32 9 221 76 73

ABSTRACT

The ectodomain of matrix protein 2 is a universal influenza A vaccine candidate that provides protection through antibody-dependent effector mechanisms. Here we compared the functional engagement of Fcγ Receptor family members by two M2e-specific monoclonal antibodies: mAb 37 (IgG1) and mAb 65 (IgG2a), which recognize the same epitope in M2e with similar affinity. Binding of mAb 65 to influenza A virus-infected cells triggered all three activating mouse Fcγ receptors *in vitro*, whereas mAb 37 only activated FcγRIII. Passive transfer of mAb 37 or mAb 65 in wild type, *Fcer1g*[-/-], *Fcgr3*[-/-] and *Fcgr1*[-/-] *Fcgr3*[-/-] BALB/c mice revealed the importance of these receptors for protection against influenza A virus challenge, with a clear requirement of FcγRIII for IgG1 mAb 37. We also report that FcγRIV contributes to protection by M2e-specific IgG2a antibodies.

IMPORTANCE

There is increased awareness that protection by antibodies directed against viral antigens is also mediated by the Fc domain of these antibodies. These Fc-mediated effector functions are often missed in clinical assays, which are used for example to define correlates of protection induced by vaccines. The use of antibodies to prevent and treat infectious diseases is on the rise, and has proven a promising approach in our battle against newly emerging viral infections. It is now also realized that broadly neutralizing antibodies directed against the conserved parts of the influenza virus hemagglutinin require the engagement of Fcγ receptors. We show here that two M2e-specific monoclonal antibodies with close to identical antigen- binding specificity and affinity have a very different *in vivo* protective potential that is controlled by their capacity to interact with activating Fcγ receptors.

Keywords: Influenza A, M2e, viral infection, Fcγ Receptors, mechanism of protection

INTRODUCTION

The ectodomain (M2e) of the influenza membrane protein M2 is an interesting candidate for a universal influenza vaccine. M2e vaccine-induced protection against influenza A viruses (IAV) is mainly conferred by antibodies [1-3]. Although only some M2 molecules are incorporated in the virion, M2 is abundantly expressed on the surface of infected cells [4, 5]. Hence, these cells are the most likely *in vivo* targets of M2e-based immune protection.

Influenza A virus infection elicits poor serum antibody responses against M2e [6, 7]. Immunization with M2e fused to a heterologous carrier, however, readily induces M2e-specific antibody responses in animal models [1, 4, 5, 8-12]. Some M2e vaccine candidates have reached early stage clinical testing, which showed their safety and immunogenicity [11, 13]. Despite these developments, there is still confusion about the mechanism of protection of M2e-specific responses. For example, a role for complement, natural killer cells and alveolar macrophages has been proposed [2, 10, 14, 15]. It is important to understand this *in vivo* mechanism in order to anticipate on potential immune evasion strategies of influenza viruses under selection pressure of M2e-based immunity, and to establish correlates of protection that are measurable by *in vitro* assays.

FcγR family members are crucial for protection by M2e-specific and broadly-neutralizing hemagglutinin-specific IgG [2, 14-16]. The mouse FcγR family comprises four members: three activating FcγRs (FcγRI, FcγRIII and FcγRIV) and one inhibitory FcγR (FcγRIIB) [14, 15]. We showed that polyclonal IgG1 isotype antibodies purified from mouse M2e immune serum required FcγRIII for immune protection and that protection of *Fcgr3*-deficient mice could be restored by an IgG fraction containing M2e-specific IgG1 and IgG2a antibodies [2]. That study, however, did not address the affinity of the purified IgG subclasses for M2e, nor did it address protection by IgG2a antibodies alone. Moreover, the possible role of mouse FcγRIV in protection by anti-M2e IgG was unknown.

Here we compared the protective potential of two mouse monoclonal antibodies (mAbs) with very similar affinities for M2e but different Fc domains: mAb 37 (IgG1) and mAb 65 (IgG2a). Our data shows that M2e-specific IgG1 requires FcγRIII, whereas IgG2a isotype antibodies protect against influenza A virus challenge via any of the three activating FcγRs.

RESULTS

M2e-specific monoclonal antibodies 37 and 65 have comparable target specificity and affinity

Our aim was to compare the capacity of two M2e-specific monoclonal antibodies, that bind a similar epitope with similar affinity but are of different isotype, to engage FcγRs *in vitro* and *in vivo*. Ala-scan revealed that for both mAb 37 and mAb 65 the recognition of M2 involves residues Thr5, Glu6, Pro10, Ile11 and Trp15 and to a lesser extent Val7, Glu8, Thr9 and Glu14 (Figure 1) [17].
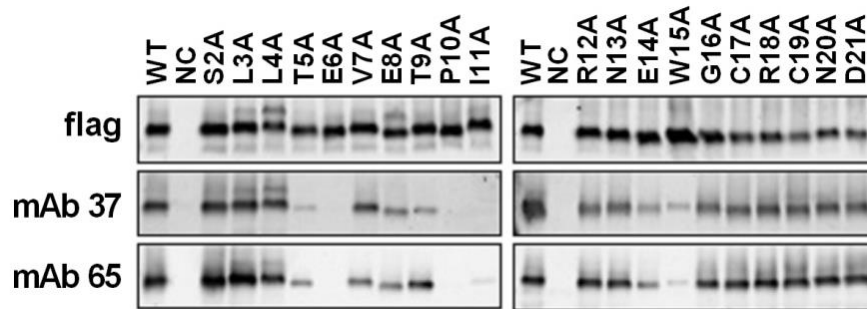

Figure 1: mAb 37 and mAb 65 have a similar epitope specificity. Lysates of HEK293T cells transfected with Flag-tagged M2 wild type (WT) and M2e Ala-scan mutants were analyzed in Western Blot for reactivity with mAb 37 or mAb 65. As a loading control, reactivity with an anti-Flag antibody was used (flag). NC: lysate from non-transfected cells; S2A: serine at position 2 in M2e changed to alanine, L3A: leucine at position 3 in M2e changed to alanine, etc. The result shown for anti-Flag and mAb 65 is the same as published before [17]. Copyright © American Society for Microbiology [17]

Next, the affinity of the two mAbs for M2e was determined. Based on Surface Plasmon Resonance measurements, mAb 37 and mAb 65 bound M2e peptide with similar equilibrium dissociation constant (Table 1). We also calculated the affinity for immobilized M2e peptide by ELISA. This method confirmed that mAb 37 and 65 bound M2e peptide with similar affinity with an estimated K_D of 0.526 nM for mAb 37 and 0.375 nM for mAb 65 (Figure 2A).

Table 1. Affinities of anti-M2e antibodies for the M2e consensus sequence measured by Surface Plasmon Resonance.

Sample	K_{on} (M^{-1}s^{-1})	K_{off} (s^{-1})	K_D (nM)	χ^2 (RU^2)
mAb 37	$3.49*10^5$ ($\pm6.60*10^2$)	$1.48*10^{-4}$ ($\pm1.70*10^{-6}$)	0.423	0.123
mAb 65	$2.24*10^5$ ($\pm4.70*10^2$)	$9.16*10^{-5}$ ($\pm3.50*10^{-6}$)	0.409	0.0593

K_{on} = Association rate constant, K_{off} = dissociation rate constant, K_D = equilibrium dissociation constant (K_{off}/K_{on}). Numbers between brackets are the standard error calculated based on measurements at seven different peptide concentrations. χ^2 = goodness of fit. RU = resonance units.
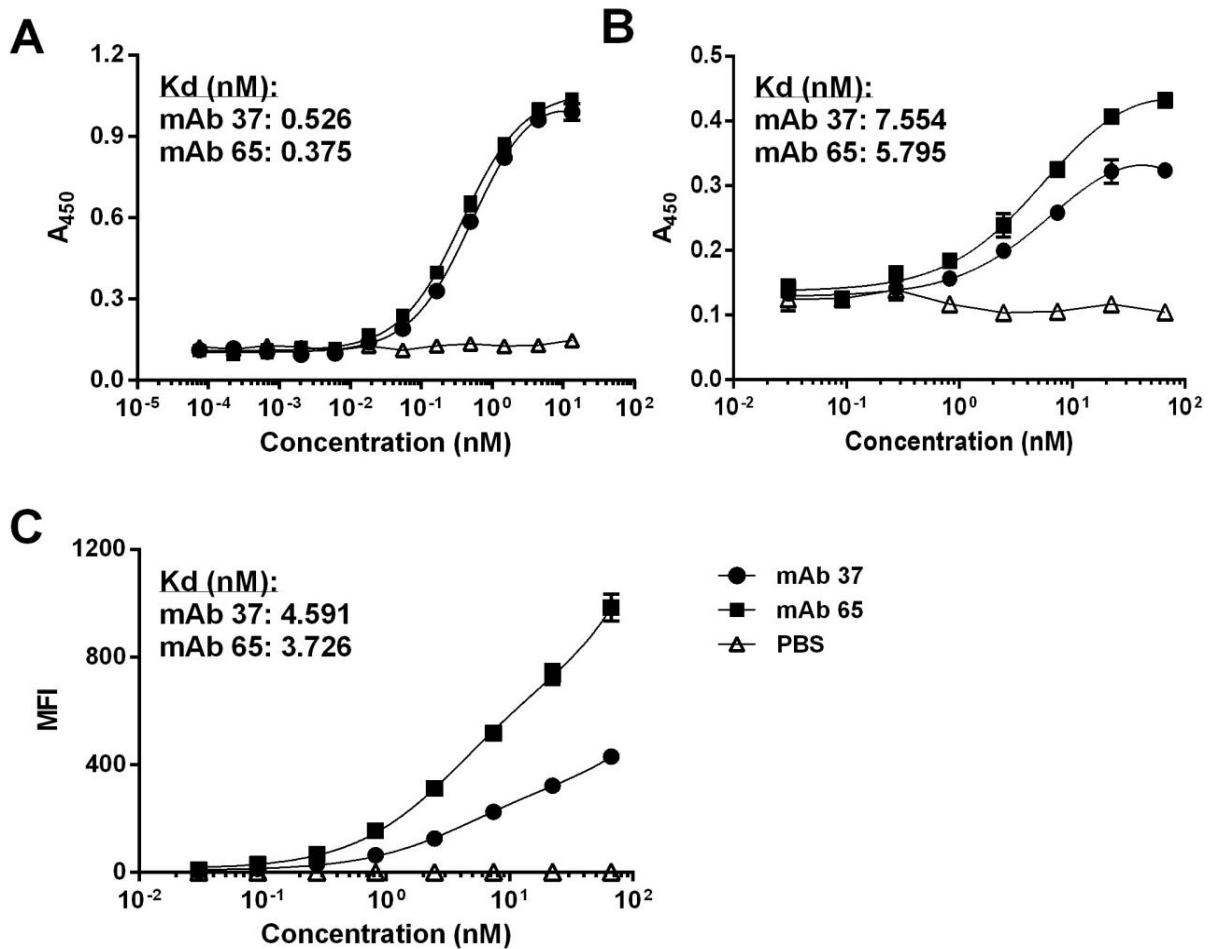
Figure 2: mAb37 and mAb65 bind with similar affinity to M2e and M2. (A) M2e peptide ELISA using high-performance liquid chromatography (HPLC)-purified M2e peptide (SLLTEVETPIRNEWGCRCNDSSD) for coating. Binding of mAb 37 or mAb 65 to M2e-coated wells was determined using a secondary horseradish peroxidase (HRP)-conjugated anti-mouse antibody. (B) MDCK cells were infected with A/Puerto Rico/8/34 virus. Twenty-four hours later, the cells were incubated with a dilution series of mAb 37 or mAb 65, followed by fixation with 4% paraformaldehyde and detection by cellular ELISA. (C) HEK293T cells were infected with A/Puerto Rico/8/34 virus. Sixteen hours later, the cells were incubated with a dilution series of mAb 37 or mAb 65, followed by fixation with 4% paraformaldehyde, permeabilisation and staining with goat anti-vRNP polyclonal serum. Binding of mAb 37 and mAb 65 was detected with donkey anti-mouse IgG coupled to AlexaFluor 488 and binding of anti-vRNP was detected with donkey anti-goat IgG coupled to AlexaFluor 647. Y-axis depicts Median fluorescence intensity (MFI) which corresponds to the median fluorescence of binding of mAb 37 or mAb 65 to infected cells subtracted with the median fluorescence of uninfected cells bound by mAb 37 or mAb 65. The dissociation constant (Kd) of mAb 37 or mAb 65 in each condition is shown. Data points represent averages (M2e peptide ELISA and Flow cytometry analysis: n = 2, cell ELISA: n =3) and error bars represent standard deviations. The graphs in A-C are representative of one out of 3 repeat experiments.

The affinity of mAb 37 and 65 for M2 on virus-infected cells, where M2 assembles as a tetrameric membrane protein, was also determined. mAb 37 and 65 bound to M2 expressed on the surface of PR8 virus-infected cells with an estimated K_D of 7.554 nM and 5.795 nM, respectively, based on cell ELISA (Figure 2B). In line with this, the estimated K_D values deduced from flow cytometry analysis of PR8 virus-infected cells were 4.591 nM and 3.726 nM for mAb 37 and 65, respectively (Figure 2C). From these *in vitro* binding studies, we conclude that the two M2e-specific mAbs bind to M2e peptide and PR8-infected cells with comparable affinity.

Murine and human IgGs have a conserved N-glycosylation site in their Fc region [18], where the associated N-glycan plays an important role in Fc-dependent effector functions [19]. Especially core-fucosylation strongly influences the binding of antibodies to FcγRs, and subsequent antibody-mediated cellular cytotoxicity or antibody-dependent cellular phagocytosis [20-23]. We therefore profiled the N-glycans of mAb 37 and 65, and show that they contained comparable levels of total terminal galactose residues and that they were completely core-fucosylated (Figure 3).

mAb 37 and 65 differentially activate FcγRs in vitro

To compare the potency of mAb 37 versus mAb 65 to activate individual FcγRs *in vitro,* a recently developed cell-based activation assay was applied [24, 25]. We compared the engagement of individual FcγRs by graded concentrations of mAb 37 and mAb 65 bound to PR8-infected MDCK cells. Control IgG1 and IgG2a monoclonal antibodies did not activate any of the FcγR-ζs in this assay (Figure 4). IgG2a mAb 65 activated all FcγR-ζs with very similar dose-response curves ranging from 0.0065 nM mAb to plateau values at 6.7 nM antibody (Figure 4). In contrast, mAb 37 (IgG1) only triggered the inhibitory FcγRII-ζ and the activating FcγRIII-ζ, and completely failed to activate FcγRI and FcγRIV even at very high concentrations of opsonizing mAb. The latter result is in line with the report that monomeric mouse IgG1 binds very poorly to FcγRI and -IV [26]. mAbs 37 and 65 activated FcγRIII-ζ equally well, but mAb 37 was approximately 10 fold more potent in activating FcγRII-ζ than mAb 65 (Figure 4), which accords well with the reported higher affinity of monomeric IgG1 for FcγRII [27]. Taken together, mAb 65 and mAb 37 exhibited similar F(ab)2-mediated M2e-binding characteristics and had clearly distinct Fc-mediated functions *in vitro*.
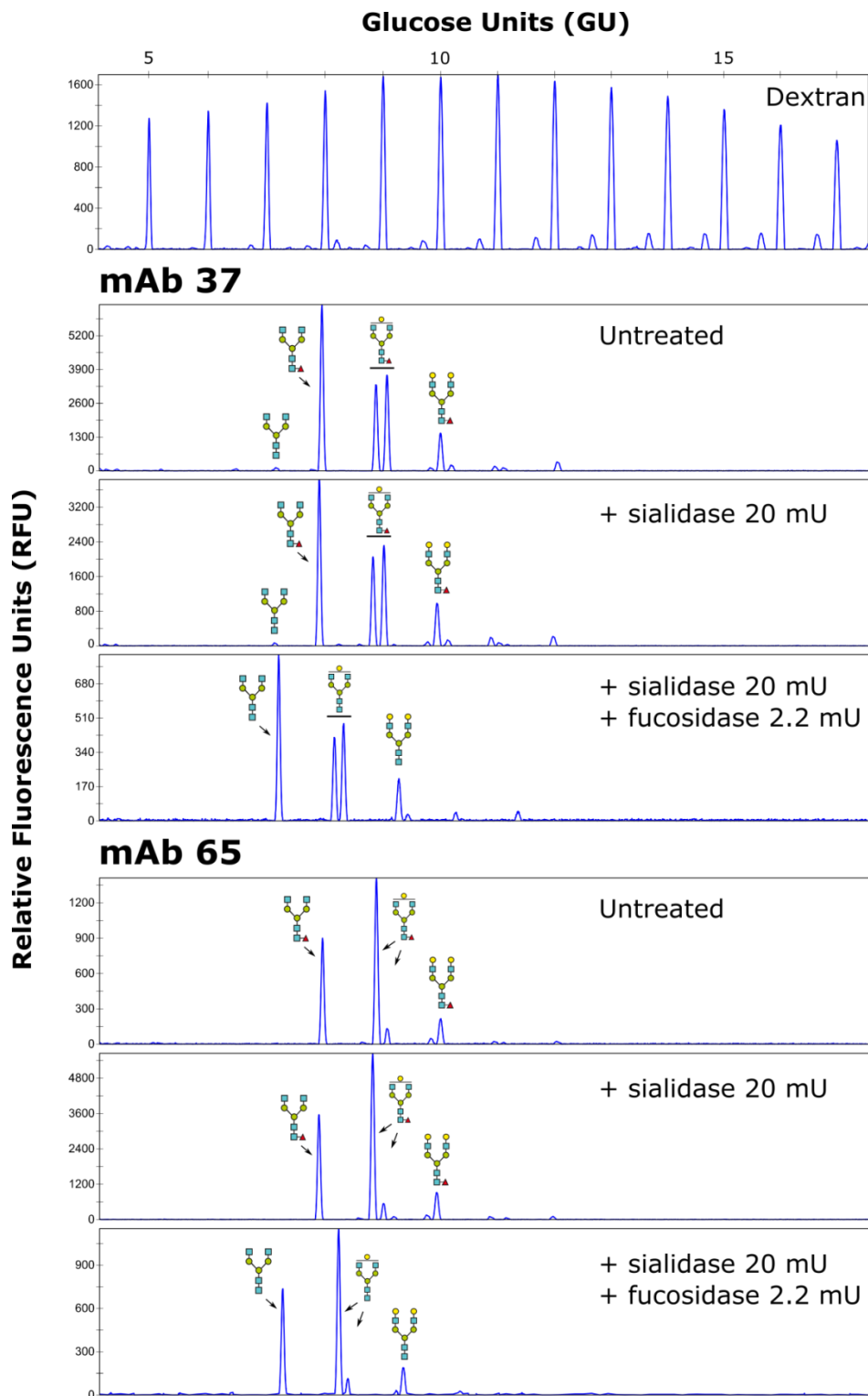
Figure 3: N-Glycan profiling of mAb 37 and mAb 65. DSA-FACE chromatograms of the fluorescently labeled dextran ladder (top) and N-glycans on mAb 37 and mAb 65. For each mAb, the upper panel corresponds to the untreated sample, which was subjected to exoglycosidase digests using an α-2,3/6/8-sialidase alone or combined with an α-1,2/3/4/6-fucosidase, as indicated. The glucose units from the ladder were annotated using N-glycans from bovine RNaseB (not shown), and the annotated structures were confirmed using additional exoglycosidase digests. Blue squares correspond to N-acetylglucosamine residues, green circles to mannose, yellow circles to galactose and red triangles to fucose residues.
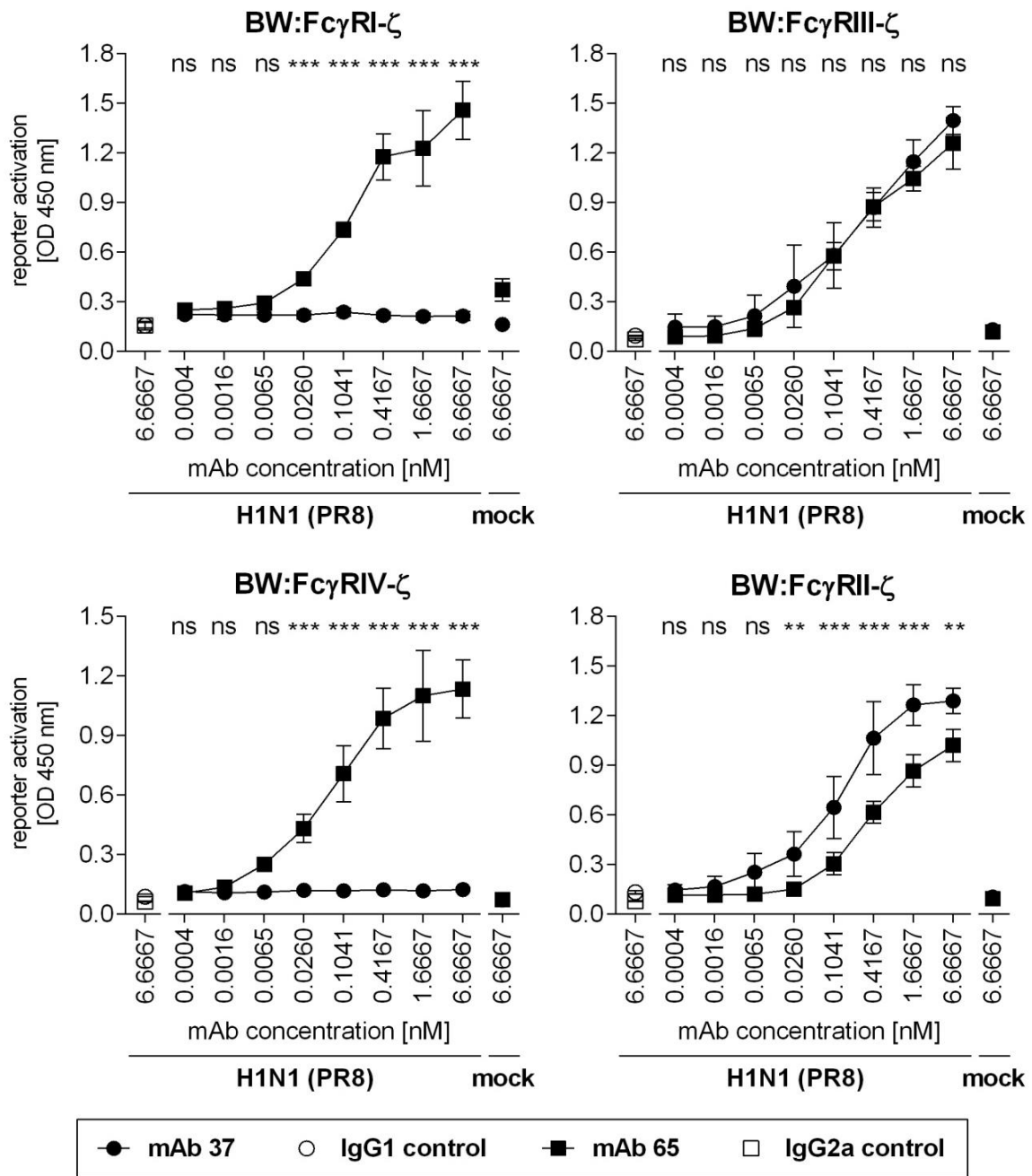
Figure 4: MAb 37 and mAb 65 differentially activate FcγRs *in vitro*. MDCK cells were infected with A/Puerto Rico/8/34 virus (MOI 5) for one hour, washed and co-cultured overnight with the FcγR-ζ expression reporter cells and graded concentrations of the mAbs or the corresponding isotype controls. Produced IL-2 was quantified by sandwich ELISA and represents a measure for the magnitude of FcγR activation. Data points are averages from triplicates and error bars represent 95% confidence interval. Two way ANOVA with Sidak correction for multiple comparisons (* p < 0.05; ** p < 0.01; *** p<0.001).
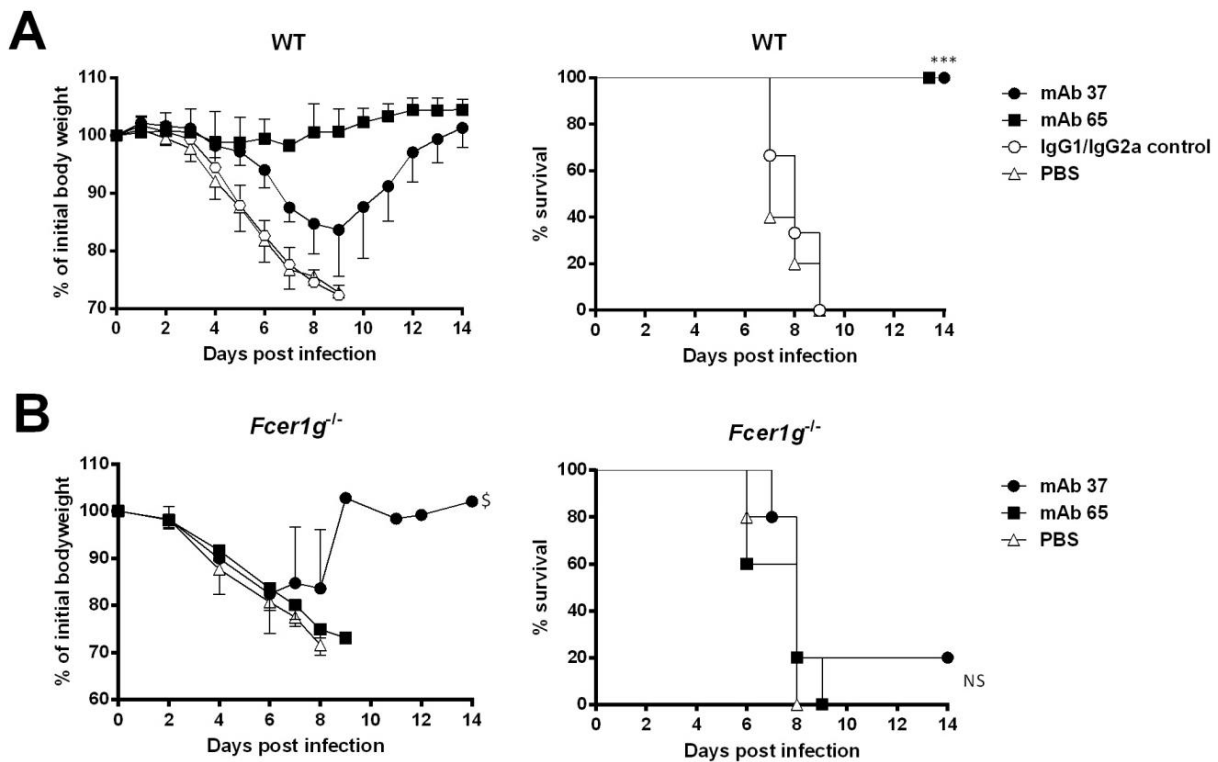
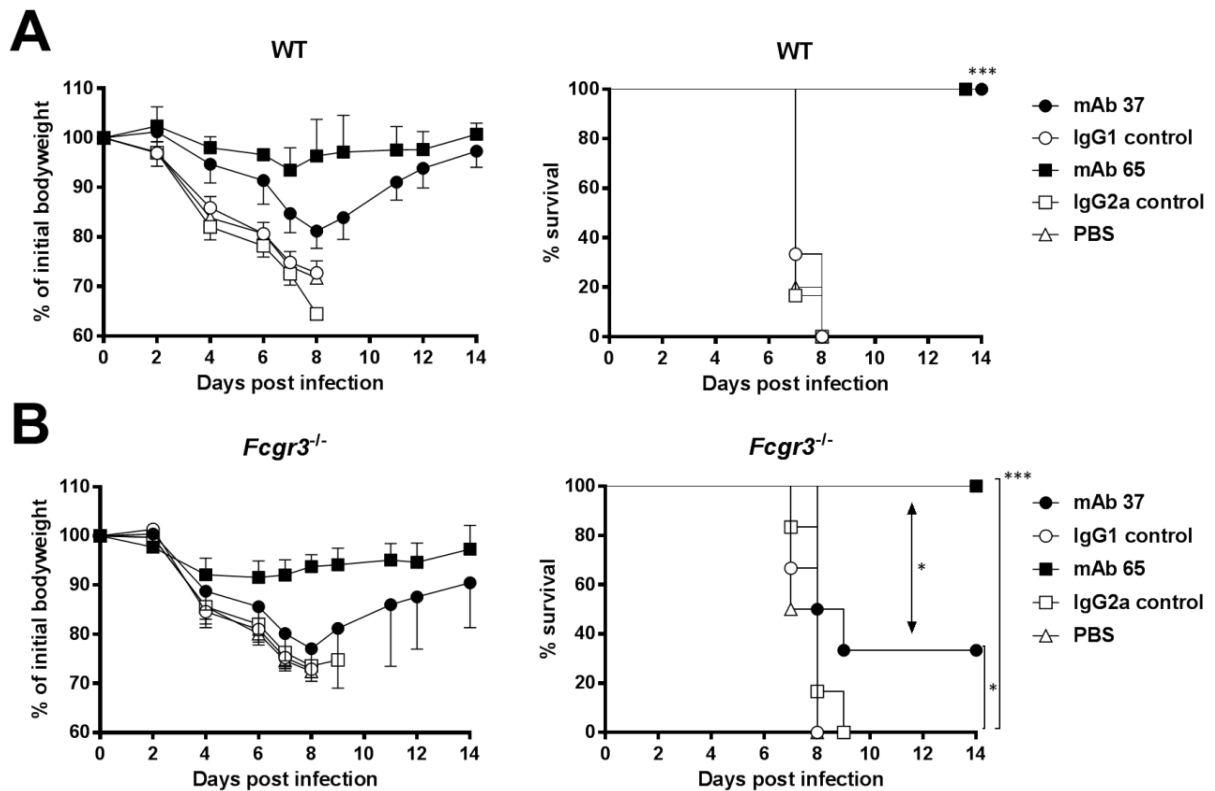**Figure 5: M2e-specific IgG2a protects better than M2e-specific IgG1 and protection depends on FcRγ.** (A) M2e-specific IgG2a protects better than M2e-specific IgG1 against influenza A virus challenge. BALB/c mice (WT) received 100 µg of either mAb 37 (IgG1) or mAb 65 (IgG2a) by i.p. injection. The IgG1/IgG2a group was treated with 50 µg of control IgG1 and 50 µg of control IgG2a per mouse. Twenty-four h later, mice were challenged with 4 LD_{50} of mouse-adapted X47 virus. Body weight (left) and survival (right) were monitored for two weeks after challenge. In the left hand graph data points represent averages and error bars represent standard deviations. Differences in body weight between the negative control groups on the one hand and the mAb 37 and 65 groups on the other hand are statistically significant (p < 0.001, REML variance components analysis, n = 6 per group on day 0 except for the PBS group which had 5 mice). The differences in body weight curves between groups that received mAb 37 or mAb 65 are statistically significant (p < 0.001; REML variance components analysis). The survival rate of the mice that received mAb 65 or mAb 37 is significantly different from the control groups (***: p<0.001, Log-rank test). The graphs represent one out of 2 repeat experiments which had similar results. (B) M2e-specific mAbs fail to protect mice lacking the Fc common ? chain. *Fcer1g*^-/-^ BALB/c mice were treated with 100 µg of mAb 37, mAb 65, or PBS by i.p. injection. Twenty-four h later, mice were challenged with 4 LD_{50} of mouse-adapted X47 virus. Body weight (left) and survival (right) were monitored for two weeks after challenge. In the left hand graph data points represent averages and error bars represent standard deviations. $: starting from day 9 after infection, data based on only one surviving mouse. The survival rates of the groups that received mAb 37 or mAb 65 are not significantly different from the control groups (p>0.05, Log-rank test, n = 5 per group on day 0). "NS": not significant.

Protection by mAb 37 and mAb 65 requires FcγRs

Next, we compared the protective efficacy of passively transferred mAb 37 and 65 against a potentially lethal influenza A virus challenge (4 LD_{50}) of BALB/c mice with X47, a mouse-adapted H3N2 virus [1]. Compared to isotype control mAbs, both mAb 37 and 65 protected the animals from lethal infection (Figure 5A). However, mice that were treated with the M2e-specific IgG2a mAb 65 displayed significantly less body weight loss compared to M2e-specific IgG1 mAb 37 recipients (p < 0.001, Figure 5A). Next, we evaluated the *in vivo* requirement of FcRγs for protection by the two mAbs. To define the requirement for one or more activating FcγRs for *in vivo* protection mediated by the two anti-M2e mAbs, we performed passive transfer experiments in *Fcer1g*$^{-/-}$ mice that lack the common γ chain and cannot express any functional FcγR. Twenty-four hours after antibody administration, the mice were challenged with 4 LD_{50} of X47. *Fcer1g*$^{-/-}$ mice that received mAb 65, mAb 37 or the isotype controls showed no significant difference in body weight loss (p = 0.541, Figure 5B). Except for one mouse in the mAb 37 recipient group, all *Fcer1g*$^{-/-}$ mice succumbed to the challenge infection (Figure 5B), demonstrating that FcγRs are essential for protection against influenza A virus challenge by both M2e-specific mAbs.

MAb 65 can protect in the absence of FcγRI and FcγRIII and involves FcγRIV

MAb 37 largely failed to protect *Fcgr3*$^{-/-}$ mice, although the difference in survival rates of challenged mice that had been treated with the irrelevant control antibody was significantly different (p = 0.0289, Figure 6). In contrast to mAb 37, mAb 65 performed significantly better (p < 0.001) and completely protected *Fcgr3*$^{-/-}$ mice from lethal infection and severe morbidity (Figure 6). To further narrow down the differential requirement for the activating FcγRs, we performed challenge studies in mice deficient in *Fcgr1* and *Fcgr3*. In these double-deficient mice, mAb 65 provided full protection against virus challenge, whereas all control-treated and mAb 37-treated mice died (Figure 7).

Figure 6: Protection by M2e-specific IgG1 mAb depends on FcγRIII. (A) Wild type BALB/c mice (n = 6 per group, except PBS: n = 5) received 100 µg of mAb 37, mAb 65, negative control IgG1 or IgG2a mAb or PBS by i.p. injection. Twenty-four h later, mice were challenged with 4 LD50 of mouse-adapted X47 virus and body weight (left) and survival (right) were monitored. In the left hand graph, data points represent averages and error bars standard deviations. (B) M2e-specific IgG2a protects better than M2e-specific IgG1 against influenza A virus challenge. *Fcgr3*[-/-] BALB/c mice (n = 6 per group) received the same treatment as in A and morbidity and mortality were monitored for 2 weeks after challenge. Both mice groups were treated with antibody and challenged in parallel in the same experiment. Body weight changes of the *Fcgr3*[-/-] mice that received mAb 65 are significantly different from those of mice that received mAb 37 or control mAbs (p<0.001; REML variance components analysis). The survival rate of the mice that received mAb 65 is significantly different from the mAb 37 recipient mice (*: p = 0.0178, Log-rank test) and from the control groups (***: p<0.001, Log-rank test). The survival rate of the mAb 37 recipient mice is significantly different from the control groups (*: p<0.05, Log-rank test).
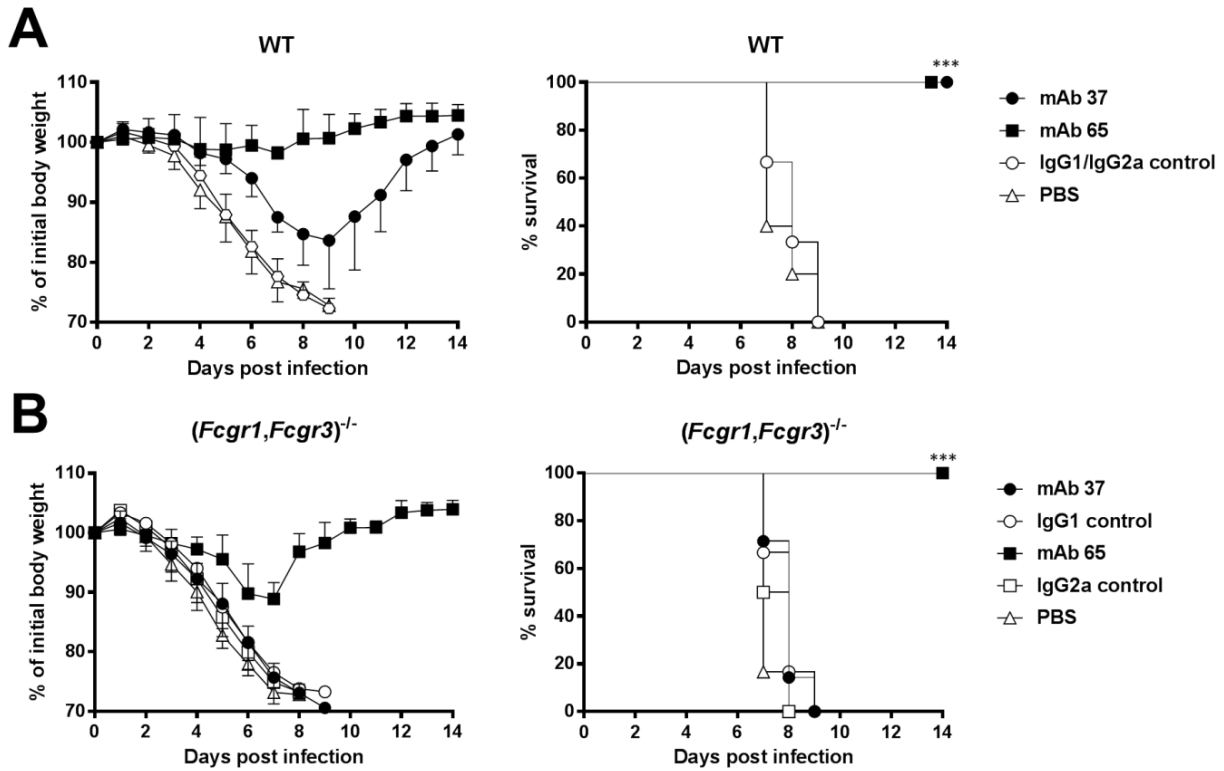
Figure 7: Protection by IgG2a subclass antibodies in the absence of FcγRI and FcγRIII. (A) Wild type BALB/c mice (n = 6 per group) received 100 μg of mAb 37, mAb 65, a 1:1 mix of negative control IgG1 and IgG2a mAb or PBS by i.p. injection. Twenty-four h later, mice were challenged with 4 LD50 of mouse-adapted X47 virus and body weight (left) and survival (right) were monitored. In the left hand graph, data points represent averages and error bars standard deviations. (B) IgG2a mAb 65 protects against influenza A virus challenge in the absence of FcγRI and –III. (*Fcgr1, Fcgr3*)$^{-/-}$ BALB/c mice (n = 6 per group, except mAb37: n = 7) received 100 μg of mAb 37, mAb 65, negative control IgG1 or IgG2a mAb or PBS by i.p. injection, followed by viral challenge as in A, with both mice groups treated with antibody and challenged in parallel in the same experiment. Differences in body weight between the (*Fcgr1, Fcgr3*)$^{-/-}$ mice that received mAb 65 are significant from mAb 37 (p <0.001; REML variance components analysis). The survival rate of the groups that received mAb 65 is significantly different from mAb 37 and the control groups (***: p<0.001, Log-rank test). The graphs are representative of one out of 2 repeat experiments with similar results. The data in (A) are the same as in Figure 5.A but are also included here to make direct comparison between the protective effect of the M2e-specific mAbs in mice differing in their FcγR compartment possible.
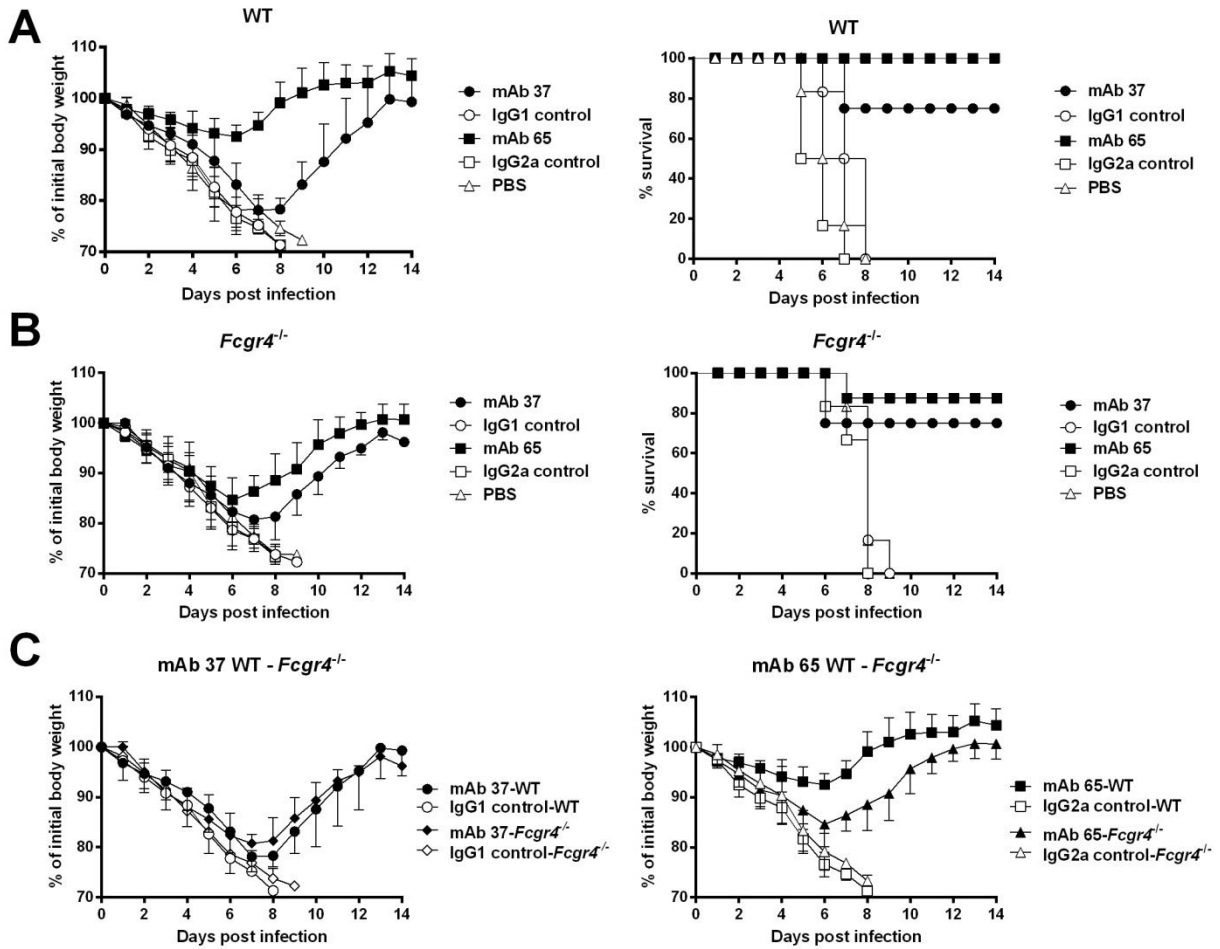
Figure 8: FcγRIV contributes to protection by M2e-specific IgG2a antibodies against influenza A virus challenge. (A) Wild type mice were treated with 100 µg of mouse IgG1 mAb 37, IgG2a mAb 65, isotype control mAb or PBS 24 h before intranasal infection with 4 LD$_{50}$ of influenza A virus strain X31. (B) *Fcgr4*$^{-/-}$ mice were treated with 100 µg of mouse IgG1 mAb 37, IgG2a mAb 65, isotype control mAb or PBS 24 h prior to challenge with 4 LD$_{50}$ of X31. Wild type and *Fcgr4*$^{-/-}$ mice were treated with antibody and challenged in parallel in the same experiment. In (A) and (B) body weight (left) and survival (right) of mice are shown. In the left hand graphs, data points represent averages and error bars standard deviations. (C) Reduced protection by IgG2a mAb 65 in *Fcgr4*$^{-/-}$ mice. The data are from panels (A) and (B) but represented as a direct comparison of wild type and *Fcgr4*$^{-/-}$ mice treated in the same way. The data are pooled from two independent experiments. In total 6 mice were used in the PBS groups and 8 or 9 in mAb treated groups. Differences in body weight change between wild type and *Fcgr4*$^{-/-}$ mice that had received mAb 65 prior to challenge were statistically significant (p < 0.001, REML variance components analysis).

We hypothesized that FcγRIV could be responsible for protection by mAb 65 in the absence of FcγRI and FcγRIII. To test this wild type and C57BL/6 *Fcgr4*[-/-] mice were treated with mAb 65 or mAb 37 and infected with a lethal dose of X31 (H3N2). Both mAb 37 and 65 protected, whereas control-treated wild type C57BL/6 and C57BL/6 *Fcgr4*[-/-] mice did not survive the virus challenge (Figure 8). Similar to wild type mice, *Fcgr4*[-/-] mice treated with mAb 65 displayed significantly less body weight loss than mAb 37 recipients ($p < 0.001$, Figure 8). Body weight loss after challenge was the same in *Fcgr4*[-/-] and wild type C57BL/6 mice treated with mAb 37 ($p = 0.114$), in line with a lack of FcγRIV engagement by mouse IgG1 isotype antibodies (Figure 4 and 8). Protection by mAb 65 was partially dependent on FcγRIV since *Fcgr4*[-/-] mice showed significantly different body weight loss after infection when compared to wild type mice ($p < 0.001$, Figure 8). Taken together, these results suggest that FcγRIV contributes to protection by M2e-specific IgG2a.

DISCUSSION

We isolated and characterized a pair of M2e-specific murine mAbs with similar affinity for M2e and a comparable N-glycan profile of their Fc parts. We compared the potency of this antibody pair to activate individual FcγRs *in vitro* in the context of a viral infection, and their protective potential in wild type and FcγR-deficient mice. mAb 65 activated all FcγR-ζs in the presence of influenza A virus-infected cells. In contrast, mAb 37 only triggered the activating FcγRIII-ζ and the inhibitory FcγRII-ζ, with the latter activation being remarkably higher compared to mAb 65. The two M2e-specific mAbs thus differentially activate FcγRs *in vitro*. These *in vitro* results also highlighted that FcγRIV can contribute to anti-M2e immune complex recognition on influenza A virus-infected target cells.

The *in vitro* results correlated surprisingly well with the *in vivo* experiments in which we compared protection by mAb 37 and 65 against influenza A virus challenge in mice with different deficiencies in their FcγR compartment. M2e-specific IgG2a antibody protected better than IgG1 against influenza A virus challenge, presumably because as detected *in vitro*, mAb 65 could engage all three activating receptors, which are expressed on natural killer cells, neutrophils, monocytes and macrophages [28, 29]. This is in agreement with the finding that active vaccination strategies with M2e fusion constructs that predominantly induce M2e-specific IgG2a/c antibodies result in better protection against challenge [30, 31].

In the absence of FcγRIII, the IgG1 mAb largely failed to protect against influenza A virus challenge while mAb 65 was still protective. Therefore, FcγRIII is not strictly required for mAb 65-mediated protection, but significantly contributes to mAb 37-mediated protection This suggests also that natural killer cells, which in mice only express FcγRIII [28], do not play an indispensable role in M2e-based immune protection provided by IgG2a isotype antibodies. In the absence of both FcγRI and FcγRIII, mice were still fully protected against death caused by influenza virus challenge. However, compared to wild type, *Fcgr1*[-/-], *Fcgr3*[-/-] double deficient mice exhibited significantly more body

174

weight loss (Figure 6). The possible contribution of FcγRIV in protection by M2e-specific antibodies has not yet been reported. We observed that *Fcgr4*[-/-] mice are protected by mAb 65, but displayed significantly more body weight loss after influenza A virus challenge than wild type mice. In contrast, mAb 37 protected wild type and *Fcgr4*[-/-] mice equally well, although worse than mAb 65 did in terms of weight loss (Figure 7). These results also accord with the finding that FcγRIV, which is expressed on monocytes, macrophages and neutrophils, plays an important role for IgG2a-dependent effector activities *in vivo*, including IgG2a-mediated killing of tumor cells [32]. FcγRI and FcγRIV are both expressed on alveolar macrophages, which play an essential role in M2e antibody-mediated immune protection [2, 27, 33, 34]. In our model, we find a contribution but not a determining role for FcγRIV in protection by M2e-specific IgG2a against influenza A virus challenge. Future studies comparing FcγRI mice and triple-deficient mice lacking functional FcγRI, FcγRIII and FcγRIV will be required to determine the relative contribution of the latter two high affinity activating Fc receptors.

What are the implications of our findings for the clinical development of M2e-based vaccines? M2e immunity appears to operate in the absence of demonstrable virus-neutralizing activity but rather engages Fcγ receptor-expressing myeloid cells. Human IgG1 and -3 isotype antibodies can be considered the counterparts of mouse IgG2a antibodies. Therefore, vaccine formulations should be used that promote the induction of antigen-specific IgG1 and -3 in humans. MF59, AS03 and AS04 adjuvants promote such a Th1-specific response [35]. Human IgG1 and -3 have the highest affinity for FcγRI, which, as in mouse, has a broad expression pattern (dendritic cells, monocytes and macrophages) [27]. The sequence of mouse FcγRIV suggests that it is related to human FcγRIIIA (expressed on natural killer cells, monocytes and macrophages) [26, 34]. Therefore, protection by M2e-specific IgG antibodies could be possible through multiple effector cells that are resident or recruited to the site of infection.

We still know surprisingly little about how effective antimicrobial vaccines work. It was reported that Fcγ receptor-dependent phagocytosis of influenza A virions opsonized with HA-specific antibodies is a strong contributor of protection by a conventional influenza A vaccine [36]. In addition, protection against influenza A virus infection by broadly neutralizing antibodies directed against the HA stalk largely depends on FcγRs [16]. Recently, another report even suggested that broadly neutralizing anti-influenza antibodies, regardless of their epitope, require interaction with FcγRs to mediate *in vivo* protection [37]. Therefore, future developments towards antibody-based universal influenza vaccines should consider the role of the Fcγ receptor repertoire in vaccine efficacy.

MATERIAL AND METHODS

Ethics statement.

All animal experiments described in this study were conducted according to the national (Belgian Law 14/08/1986 and 22/12/2003, Belgian Royal Decree 06/04/2010) and European legislation (EU Directives 2010/63/EU, 86/609/EEC). All experiments on mice and animal protocols were approved by the ethics committee of Ghent University (permit numbers LA1400091 and EC2012-034).

Monoclonal antibodies and their epitope specificity.

Hybridomas that produce M2e-specific mAbs 37 and 65 were isolated as described [17]. After subcloning, these two hybridoma cultures were scaled up and mAb 37 and 65 were purified from the culture supernatant by protein A sepharose (GE Healthcare). M2e-Ala scan analysis was performed as described [17] and visualized by Western blot using mAb 37, mAb 65 or anti-Flag (Sigma-Aldrich) antibody followed by HRP-conjugated sheep anti-mouse IgG (GE Healthcare). Isotype control mAbs directed against Hepatitis B core (IgG1) or the Small Hydrophobic protein of human respiratory syncytial virus (IgG2a) were produced and purified as above.

Affinity measurement by ELISA and flow cytometry.

The affinity of mAb 37 and mAb 65 for M2e was determined by ELISA with M2e peptide (SLLTEVETPIRNEWGCRCNDSSDSG, used at 2 μg/ml in 50 μl/well), as described in [17], or MDCK cells infected with A/Puerto Rico/8/34 (H1N1) (PR8). A dilution series of mAb 65 or mAb 37 was applied to infected cells on ice. Subsequently, cells were fixed with 4% paraformaldehyde and antibody binding was detected using HRP-conjugated sheep anti-mouse IgG (GE Healthcare).

In flow cytometry analysis, PR8-infected HEK 293T cells were incubated on ice with a dilution series of mAb 65 or mAb 37 in 0.5% BSA in PBS, and subsequently fixed with 4% PFA. After permeabilization (10x Permeabilization buffer diluted in bidi, eBioscience), cells were stained with 1/2000 diluted polyclonal goat anti-influenza ribonucleoprotein (RNP) (Biodefense and Emerging Infections Resources Repository, NIAID, NIH, NR-4282). Binding of primary antibodies was revealed with donkey-anti-mouse IgG coupled to Alexa Fluor 488 (Invitrogen, 1/600) and donkey-anti-goat IgG coupled to Alexa Fluor 647 (Invitrogen, 1/600). The median fluorescence intensity (MFI) of the cells was determined with a FACS Calibur (BD) flow cytometer. The influenza virus-specific MFI was calculated by subtracting the MFI of mAb 37 or mAb 65 positive cells in the RNP negative population (uninfected HEK cells stained with 10 μg/ml mAb 37 or mAb 65 and anti-RNP) from the MFI of mAb 37 or mAb 65 positive cells in the RNP positive population (infected HEK cells stained with dilution series of mAb 37 or mAb 65 and RNP).

Affinity measurement by Surface Plasmon Resonance.

The affinities of mAbs 37 and 65 for M2e peptide were determined using a Biacore T200 instrument (GE Healthcare). Anti-M2e mAbs were immobilized on flow cells of a CM5 sensor chip (GE Healthcare) by amine coupling according to the manufacturer's instructions. A flow cell blocked with

ethanolamine served as a reference. M2e peptide in HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) was added at concentrations of 0.39, 0.78, 3.13, 6.25, 12.5, 25 and 50 nM. The samples were injected at 50 µl/min for 180s, after which dissociation was monitored for 1000s. The sensor chip surface was regenerated by injecting 10 mM HCl for 60s, and 20 mM HCl for another 60s. The Biacore T200 Evaluation Software v1.0 was used to calculate the association rate constants (k_{on}), dissociation rate constants (k_{off}) and the equilibrium dissociation constants (KD = k_{off}/k_{on}) by fitting a 1:1 binding model with drifting baseline.

N-glycan analysis.

N-linked oligosaccharides were isolated, derivatized with APTS and analyzed by capillary electrophoresis on an ABI3130 capillary DNA sequencer using in-house produced PNGaseF (15.4 IUB mU/µl) as described [38]. As electrophoretic mobility references, labeled dextran ladder and N-glycans from bovine RnaseB were included. The data were analyzed using the Genemapper software (Applied Biosystems) and N-glycan profiles were exported as Scalable Vector Graphics (svg) for manual alignment and annotation in Inkscape 0.91. Exoglycosidase treatments to determine the degree of fucosylation were carried out overnight at 37°C in 20 mM sodium acetate (pH 5.0) on labeled glycans using 20 mU *Arthrobacter ureafaciens* α-2,3/6/8-sialidase (produced in-house), 2.2 mU α-1,2/3/4/6-fucosidase from bovine kidney (Prozyme), or both.

In vitro FcγR activation assay.

FcγR activation by mAb 37 and 65 was compared using a recently developed *in vitro* FcγR activation assay [24, 25, 39]. Cloning of FcγR-ζ constructs and generation of FcγR-ζ BW5147 reporter cells was performed as reported [24, 25]. Activation of stably transduced FcγR-ζ BW5147 reporter cells by immune complexes results in the production of IL-2 which is quantified by ELISA [24, 25]. MDCK cells were seeded in 96 well flat bottom plates and infected with PR8 virus (MOI 5). After one h incubation at 37°C, unbound virus particles were removed by washing and serial dilutions of the mAbs were added followed by 1.5×10^5 BW-FcγR-ζ BW5147 reporter cells in a total volume of 200 µl RPMI with 10% FCS per well. Co-cultures were incubated overnight at 37°C in a 5% CO_2 atmosphere. To increase the release of produced IL-2 from reporter cells, 100 µl PBS with 0,1% Tween was added to each well and 150 µl was used in an anti-IL2 sandwich ELISA as described [24, 25].

Challenge experiments in mice.

BALB/c mice were purchased from Harlan (The Netherlands), BALB/c *Fcer1g*[-/-] mice from Taconic (Denmark) and C57BL/6 mice from Charles River (France). BALB/c *Fcgr3*[-/-] and BALB/c *Fcgr1*[-/-] *Fcgr3*[-/-] double-deficient mice and C57BL/6 *Fcgr4*[-/-] mice were bred in-house under SPF conditions. Mice were used at the age of 7-8 weeks and were housed in individually ventilated cages, in a temperature-controlled environment with 12 h light/dark cycles with food and water ad libitum. To evaluate protection, mice were injected intraperitoneally with 100 µg of mAb 37, mAb 65, or negative control mAbs. Twenty-four h later, mice were anesthetized with a mixture of ketamine (10 mg/kg) and xylazine (60 mg/kg) and challenged by intranasal administration of 50 µl PBS containing 4

LD_{50} of mouse-adapted X47 (A/Victoria/3/75 (H3N2) x PR8) influenza A virus for the wt BALB/c, BALB/c $Fcer1g^{-/-}$, BALB/c $Fcgr3^{-/-}$ and BALB/c $Fcgr1^{-/-}$ $Fcgr3^{-/-}$ double knock-out mice. Four LD_{50} of mouse-adapted X31 (A/Aichi/2/68 (H3N2) x PR8) influenza A virus was used to challenge wt C57BL/6 and C57BL/6 $Fcgr4^{-/-}$ mice. X31 and X47 thus have an identical gene segment 7. Body weight and survival of mice was monitored for two weeks after challenge and animals that had lost more than 25% body weight compared to the time of challenge, were euthanized.

Statistics.

Statistical significance between the different mAbs in the FcγRs activation assay was analysed using two-way ANOVA with Sidak correction for multiple comparisons. Statistical analysis of the differences in survival rates was performed by comparing Kaplan-Meier curves using the Log-rank test. Both tests were performed in GraphPad Prism version 6.07 for Windows (GraphPad Software, San Diego California; www.graphpad.com). Statistical comparison of differences in bodyweight loss was performed using Restricted Maximum Likelihood (REML) variance components analysis in Genstat 64.bit version 18.1. The following linear mixed model (random terms underlined) was fitted to the data: $Y_{ijkt} = \mu + \text{genotype}_j + \text{treatment}_k + \text{time}_t + (\text{genotype.treatment})_{jk} + (\text{genotype.time})_{jt} + (\text{treatment.time})_{kt} + (\text{genotype.treatment.time})_{jkt} + \underline{(\text{mouse.time})_{it}} + \underline{\text{residual}_{ijkt}}$, where Y_{ijkt} is the relative body weight of *i*-th mouse having genotype *j*, *k*-treated and measured at time point *t* (*t* = 1 – 14 days; equally spaced), and μ is the overall mean calculated for all mice considered across all time points. A first order autoregressive covariance structure was used to model the within-subject correlation and allowed for heterogeneity across time. Significance of the fixed main and interaction effects was assessed by an *F*-test. A value of $p \leq 0.05$ was considered statistically significant.

Acknowledgements

References

1. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.
2. El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.
3. Treanor JJ, Tierney EL, Zebedee SL, Lamb RA, Murphy BR: Passively transferred monoclonal antibody to the M2 protein inhibits influenza A virus replication in mice. *Journal of virology* 1990, 64(3):1375-1377.
4. Lamb RA, Zebedee SL, Richardson CD: Influenza Virus-M2 Protein Is an Integral Membrane-Protein Expressed on the Infected-Cell Surface. *Cell* 1985, 40(3):627-633.
5. Zebedee SL, Lamb RA: Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of virology* 1988, 62(8):2762-2772.
6. Feng J, Zhang M, Mozdzanowska K, Zharikova D, Hoff H, Wunner W, Couch RB, Gerhard W: Influenza A virus infection engenders a poor antibody response against the ectodomain of matrix protein 2. *Virology journal* 2006, 3:102.
7. Zhong W, Reed C, Blair PJ, Katz JM, Hancock K: Serum antibody response to matrix protein 2 following natural infection with 2009 pandemic influenza A(H1N1) virus in humans. *The Journal of infectious diseases* 2014, 209(7):986-994.
8. De Filette M, Martens W, Roose K, Deroo T, Vervalle F, Bentahir M, Vandekerckhove J, Fiers W, Saelens X: An influenza A vaccine based on tetrameric ectodomain of matrix protein 2. *The Journal of biological chemistry* 2008, 283(17):11382-11387.
9. Fan J, Liang X, Horton MS, Perry HC, Citron MP, Heidecker GJ, Fu TM, Joyce J, Przysiecki CT, Keller PM *et al*: Preclinical study of influenza virus A M2 peptide conjugate vaccines in mice, ferrets, and rhesus monkeys. *Vaccine* 2004, 22(23-24):2993-3003.
10. Jegerlehner A, Schmitz N, Storni T, Bachmann MF: Influenza A vaccine based on the extracellular domain of M2: weak protection mediated via antibody-dependent NK cell activity. *J Immunol* 2004, 172(9):5598-5605.
11. Schotsaert M, De Filette M, Fiers W, Saelens X: Universal M2 ectodomain-based influenza A vaccines: preclinical and clinical developments. *Expert review of vaccines* 2009, 8(4):499-508.
12. Wang BZ, Gill HS, Kang SM, Wang L, Wang YC, Vassilieva EV, Compans RW: Enhanced influenza virus-like particle vaccines containing the extracellular domain of matrix protein 2 and a Toll-like receptor ligand. *Clinical and vaccine immunology : CVI* 2012, 19(8):1119-1125.
13. Turley CB, Rupp RE, Johnson C, Taylor DN, Wolfson J, Tussey L, Kavita U, Stanberry L, Shaw A: Safety and immunogenicity of a recombinant M2e-flagellin influenza vaccine (STF2.4xM2e) in healthy adults. *Vaccine* 2011, 29(32):5145-5152.
14. Wang R, Song A, Levin J, Dennis D, Zhang NJ, Yoshida H, Koriazova L, Madura L, Shapiro L, Matsumoto A *et al*: Therapeutic potential of a fully human monoclonal antibody against influenza A virus M2 protein. *Antiviral research* 2008, 80(2):168-177.
15. Schmitz N, Beerli RR, Bauer M, Jegerlehner A, Dietmeier K, Maudrich M, Pumpens P, Saudan P, Bachmann MF: Universal vaccine against influenza virus: linking TLR signaling to anti-viral protection. *European journal of immunology* 2012, 42(4):863-869.
16. DiLillo DJ, Tan GS, Palese P, Ravetch JV: Broadly neutralizing hemagglutinin stalk-specific antibodies require FcgammaR interactions for protection against influenza virus in vivo. *Nature medicine* 2014, 20(2):143-151.
17. Cho KJ, Schepens B, Seok JH, Kim S, Roose K, Lee JH, Gallardo R, Van Hamme E, Schymkowitz J, Rousseau F *et al*: Structure of the extracellular domain of matrix protein 2 of influenza A virus in complex with a protective monoclonal antibody. *Journal of virology* 2015, 89(7):3700-3711.
18. Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P: Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *Journal of molecular biology* 2003, 325(5):979-989.
19. Jefferis R, Lund J, Pound JD: IgG-Fc-mediated effector functions: molecular definition of interaction sites for effector ligands and the role of glycosylation. *Immunological reviews* 1998, 163:59-76.

20.   Forthal DN, Gach JS, Landucci G, Jez J, Strasser R, Kunert R, Steinkellner H: Fc-glycosylation influences Fcgamma receptor binding and cell-mediated anti-HIV activity of monoclonal antibody 2G12. *J Immunol* 2010, 185(11):6876-6882.

21.   Herter S, Birk MC, Klein C, Gerdes C, Umana P, Bacac M: Glycoengineering of therapeutic antibodies enhances monocyte/macrophage-mediated phagocytosis and cytotoxicity. *J Immunol* 2014, 192(5):2252-2260.

22.   Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SH, Presta LG: Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *The Journal of biological chemistry* 2002, 277(30):26733-26740.

23.   Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M *et al*: The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *The Journal of biological chemistry* 2003, 278(5):3466-3473.

24.   Corrales-Aguilar E, Trilling M, Hunold K, Fiedler M, Le VT, Reinhard H, Ehrhardt K, Merce-Maldonado E, Aliyev E, Zimmermann A *et al*: Human cytomegalovirus Fcgamma binding proteins gp34 and gp68 antagonize Fcgamma receptors I, II and III. *PLoS pathogens* 2014, 10(5):e1004131.

25.   Corrales-Aguilar E, Trilling M, Reinhard H, Merce-Maldonado E, Widera M, Schaal H, Zimmermann A, Mandelboim O, Hengel H: A novel assay for detecting virus-specific antibodies triggering activation of Fcgamma receptors. *Journal of immunological methods* 2013, 387(1-2):21-35.

26.   Nimmerjahn F, Bruhns P, Horiuchi K, Ravetch JV: FcgammaRIV: a novel FcR with distinct IgG subclass specificity. *Immunity* 2005, 23(1):41-51.

27.   Guilliams M, Bruhns P, Saeys Y, Hammad H, Lambrecht BN: The function of Fcgamma receptors in dendritic cells and macrophages. *Nature reviews Immunology* 2014, 14(2):94-108.

28.   Bruhns P: Properties of mouse and human IgG receptors and their contribution to disease models. *Blood* 2012, 119(24):5640-5649.

29.   Nimmerjahn F, Ravetch JV: FcgammaRs in health and disease. *Current topics in microbiology and immunology* 2011, 350:105-125.

30.   Ibanez LI, Roose K, De Filette M, Schotsaert M, De Sloovere J, Roels S, Pollard C, Schepens B, Grooten J, Fiers W *et al*: M2e-displaying virus-like particles with associated RNA promote T helper 1 type adaptive immunity against influenza A. *PLoS one* 2013, 8(3):e59081.

31.   Song JM, Van Rooijen N, Bozja J, Compans RW, Kang SM: Vaccination inducing broad and improved cross protection against multiple subtypes of influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108(2):757-761.

32.   Nimmerjahn F, Lux A, Albert H, Woigk M, Lehmann C, Dudziak D, Smith P, Ravetch JV: FcgammaRIV deletion reveals its central role for IgG2a and IgG2b activity in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(45):19396-19401.

33.   Lee SM, Gardy JL, Cheung CY, Cheung TK, Hui KP, Ip NY, Guan Y, Hancock RE, Peiris JS: Systems-level comparison of host-responses elicited by avian H5N1 and seasonal H1N1 influenza viruses in primary human macrophages. *PloS one* 2009, 4(12):e8072.

34.   Mancardi DA, Iannascoli B, Hoos S, England P, Daeron M, Bruhns P: FcgammaRIV is a mouse IgE receptor that resembles macrophage FcepsilonRI in humans and promotes IgE-induced lung inflammation. *The Journal of clinical investigation* 2008, 118(11):3738-3750.

35.   Coffman RL, Sher A, Seder RA: Vaccine adjuvants: putting innate immunity to work. *Immunity* 2010, 33(4):492-503.

36.   Huber VC, Lynch JM, Bucher DJ, Le J, Metzger DW: Fc receptor-mediated phagocytosis makes a significant contribution to clearance of influenza virus infections. *J Immunol* 2001, 166(12):7381-7388.

37.   DiLillo DJ, Palese P, Wilson PC, Ravetch JV: Broadly neutralizing anti-influenza antibodies require Fc receptor engagement for in vivo protection. *The Journal of clinical investigation* 2016, 126(2):605-610.

38.   Laroy W, Contreras R, Callewaert N: Glycome mapping on DNA sequencing equipment. *Nature protocols* 2006, 1(1):397-405.

39.   Storcksdieck genannt Bonsmann M, Niezold T, Temchura V, Pissani F, Ehrhardt K, Brown EP, Osei-Owusu NY, Hannaman D, Hengel H, Ackerman ME *et al*: Enhancing the Quality of Antibodies to HIV-1 Envelope by GagPol-Specific Th Cells. *J Immunol* 2015, 195(10):4861-4872.

Chapter 8

Influenza A virus escape routes from immune

selection by M2e-specific monoclonal antibodies

Influenza A virus escape routes from immune selection by M2e-specific monoclonal antibodies.

Silvie Van den Hoecke[1,2], Bram Vrancken[3], Lei Deng[1,2,4], Emma Job[1,2], Kenny Roose[1,2], Bert Schepens[1,2], Philippe Lemey[3] and Xavier Saelens[1,2].

[1] Department of Biomedical Molecular Biology, Ghent University, Ghent, B-9052, Belgium

[2] Medical Biotechnology Center, VIB, Ghent, B-9052, Belgium

[3] KU Leuven - University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Evolutionary and Computational Virology, B-3000 Leuven, Belgium

[4] Current affiliation: Department of Microbiology and Immunology, School of Medicine, Emory University, GA 30322 Atlanta, USA

Relative contributions of the authors:

SVDH performed the experiments and data analysis. BV performed the Shannon Entropy, distance-based clustering and logistic regressions analysis. LD assisted with the mice experiments. SVDH, BV, EJ, KR, BS, PL and XS designed the experiments. SVDH, BV and XS co-wrote the manuscript. XS and PL supervised the research.

Manuscript in preparation

182

Abstract

The ectodomain of matrix protein 2 (M2e) of influenza A viruses is a universal influenza A vaccine candidate. In this study, we explored the potential evasion strategies of influenza A viruses under *in vivo* M2e-based immune selection pressure induced by passively transferred M2e-specific IgG antibodies. PR8-infected SCID mice were chronically treated with one of three anti-M2e mouse IgG monoclonal antibodies (mAb) that differ in epitope and/or antibody isotype. mAb65 (IgG2a) and mAb37 (IgG1) recognize an M2e epitope that involves amino acids 5 to 15 and bind to M2e with a similar affinity. mAb148 (IgG1) binds to the invariant N-terminus of M2e. Treatment of challenged SCID mice with these mAbs significantly prolonged survival compared to control IgG treatment. Furthermore, M2e-specific IgG2a protected significantly better than IgG1, and was associated with virus-clearance in some of the SCID mice. Whole virus genome next-generation sequence analysis of the virus population that persisted in mice treated with mAb37 or mAb65, revealed that viruses emerged with a proline to histidine or leucine mutation at position 10 and/or an isoleucine to threonine mutation at position 11 in M2. These mutations abolished recognition by mAb37 and -65 and occurred at diverse frequencies, either alone or combined in the viral population. Polyclonal anti-M2e serum, induced by vaccination with an M2e-virus-like particle vaccine, still recognized these M2e variants. Remarkably, in half of the samples isolated from moribund mAb37-treated mice and in all mAb148-treated mice, virus was isolated with a wild type M2 sequence but with non-synonymous mutations in the polymerases and/or the hemagglutinin. Some of these mutations, when combined, were associated with a delay in M2 expression. We conclude that immune protection by M2e-specific mAbs selects for viruses with limited variation in M2e, which can still be recognized by polyclonal anti-M2e serum. Our results also suggest that influenza A viruses may undergo an alternative escape route from M2e-specific mAbs by acquiring mutations elsewhere in their genome that results in delayed M2 expression.

Introduction

Human influenza is a highly contagious respiratory disease and remains a major cause of morbidity and mortality each year. The licensed influenza vaccines are mainly based on inducing virus neutralizing antibodies against hemagglutinin (HA), the major membrane protein of influenza [1]. However, these vaccines are strain-specific and their protective effect is limited by antigenic drift, which necessitates annual vaccination with an updated vaccine composition. Expanding the antigenic breath of influenza vaccines by targeting more conserved antigenic portions of the virus, could lead to a new generation of influenza vaccines that protect against disease caused by antigenically drifted viruses or even provide protection against multiple influenza A virus subtypes. The protective potential of the conserved ectodomain of the matrix protein 2 (M2, ectodomain of M2: M2e) has been explored extensively, and supports the inclusion of an M2e-based vaccine in a yet to be licensed 'Universal' influenza A vaccine [2-6].

M2 is a 97-amino acid homotetrameric transmembrane protein of influenza A viruses with selective ion channel activity [7, 8]. In contrast to HA and neuraminidase (NA), only a small number of M2 molecules (20 to 60/virion) are present in the influenza virion, but M2 is highly expressed on the infected cell surface [9-12]. M2e is 23 amino acids long and its sequence is evolutionary conserved between human, avian as well as swine influenza viruses [13]. This sequence conservation of M2e can be explained by the very limited immune pressure imposed by the host immune system against M2e and the overlap of its coding information with the open reading frame of M1, one of the most conserved influenza proteins [14, 15]. M2 is translated from a spliced M1 mRNA, with the first 9 aa shared between both proteins [16-18]. These residues are almost completely conserved in all influenza A virus strains, even in the recently isolated H17N10 and H18N11 viruses from bats [13, 16, 19]. Amino acids 10 to 23 of M2e are located downstream of the 3' splice acceptor site and are translated out-of-frame with the M1 protein, with somewhat more sequence variation observed at these positions [13].

Influenza viruses can escape from the current vaccination pressure as a consequence of the relatively high error rate (2.3×10^{-5}) of their RNA-dependent-RNA-polymerase, together with the fact that their HA and NA can accommodate almost any amino acid substitution in their dominant antigenic sites, without the loss of functions essential for virus replication [20]. Such substitutions can result in escape from immune pressure, which provides a selective advantage over the parental virus population and can result in outgrowth of the mutant virus. The mechanisms of influenza virus escape from monoclonal and polyclonal antibodies that target HA can be recapitulated *in vitro* and has been extensively studied [1, 21, 22]. In contrast, M2e-specific antibodies are not virus neutralizing but are protective by binding to Fcγ Receptor expressing immune cells such as alveolar macrophages and natural killer cells [23-25].

Only two studies have reported on the characteristics of mutant influenza A viruses that emerged under M2e-specific immune selection pressure, both using M2e-specific mAbs. In the first study, anti-M2e escape mutants were generated by *in vitro* selection of A/Udorn/72 virus, which, unlike most other influenza A viruses, is sensitive *in vitro* to the direct action of anti-M2e IgG [26]. The resistant phenotype was in most escape viruses linked to single amino acid substitutions in the cytoplasmic domain of M2 or the N terminus of M1 [26]. In only one escape virus, resistance was linked to a glutamic acid to glycine mutation at position 8 in M2 (and M1) and resulted in loss of anti-M2e IgG binding [26]. The second study, performed by Zharikova and colleagues, made use of SCID mice that were infected with PR8 virus and that were treated with anti-M2e mAbs [27]. Escape viruses presented with a change in M2e at position 10, where a proline was mutated to either a histidine or leucine [27]. Although only the M2e sequence was thoroughly investigated, sequencing of 10,641 M1-encoding nucleotides of M2e-escape mutants also revealed six synonymous and two nonsynonymous mutations (A127T and V228L) in M1 [27]. These mutations were based on sequence analysis of a virus sample that had been amplified on MDCK cells using virus present in the lung homogenates of treated mice as the starting inoculum. As such, cell culture adaptation may have disturbed the original variant frequencies of the virus that was present in the lung and therefore some of the *in vivo* viral diversity may have been missed. Sequence analysis of the MDCK grown virus was performed by direct Sanger sequencing of purified RT-PCR product [27]. Although state of the art in 2005, this method has a relatively limited sensitivity of approximately 20 % to detect variant sequences. In addition, the anti-M2e mAbs used in the study of Zharikova *et al.* were all IgG2a and recognized the same epitope. It is now clear that the B cell response against M2e-based vaccines is oligoclonal, and M2e-specific mAbs that bind to different parts of M2e have been described [28-30]. Moreover, FcγRs play a very important role in anti-M2e based immunity and these receptors differentially engage different antibody isotypes [23, 31, 32]. Therefore, it would be interesting to explore the escape routes of influenza A viruses from selection imposed by different anti-M2e IgG isotypes.

Here, we used a next-generation sequencing (NGS) approach to explore the possible escape routes of influenza viruses in SCID mice under anti-M2e mAb pressure. We previously described a workflow to study the genetic diversity of influenza A viruses using NGS [33]. This approach allows whole genome sequence analysis with higher sensitivity compared to Sanger sequencing of PCR products, and thus may allow the identification of minor variants in the viral population. Importantly, bias introduced by *in vitro* amplification of the virus on cells and cloning of viral isolates is bypassed, since sequence analysis is performed directly on RT-PCR products prepared from influenza viral genomic RNA present in BAL fluid.

Anti-M2e mAbs of the IgG2a (mAb65) and IgG1 (mAb37) isotype that recognize a similar epitope in M2e with the same affinity were used for *in vivo* escape selection [28, 34] (see also Chapter 7). In addition, an anti-M2e mAb (mAb148, IgG1) was included that binds to the very conserved first 8 aa of the N-terminal part of M2e [29]. We found that treatment of infected mice with anti-M2e mAbs

resulted in significant longer survival when compared to control mAb treated mice. In addition, M2e escape mutants were selected *in vivo* in mAb37 or mAb65 treated mice, but not in mice treated with mAb148. However, escape in M2e was limited to a proline to histidine or leucine mutation at position 10 and/or a isoleucine to threonine mutation at position 11. These M2e-mutants retained sufficient antigenicity for recognition by a polyclonal anti-M2e immune serum. Remarkably, in half of the mAb37 treated and in all mAb148 treated mice that eventually succumbed to infection, only virus with wild type M2e was detected. Instead, several mutations with a frequency above 50% were observed in PB2, PA and HA in these samples. We found that a combination of some of these mutations was associated with delayed M2 expression.

Results

Experimental conditions that result in sufficient rounds of viral replication to select for M2e escape mutants.

A well-characterized, plasmid-derived PR8 viral stock was used to infect SCID mice [33]. To select for M2e escape variants in this virus upon treatment with M2e mAbs, a suitable infection dose had to be determined which allows multiple replication cycles before the ethical experimental endpoint of 25% body weight loss is reached. Therefore, SCID mice were infected with 10 or 50 PFU of PR8 virus. BALB/c mice infected with 50 PFU of PR8 virus were included as control. The ethical experimental endpoint was reached in 75% of the SCID mice on day 11 after infection and in 50% of the SCID mice on day 7 after infection with 10 and 50 PFU of PR8 virus, respectively (Figure 1). One mouse still retained 92.59% of its initial bodyweight on day 11 after infection with 10 PFU of PR8 virus, the time at which the remaining three mice in that group had reached the ethical endpoint. This difference in morbidity within that group may have been due to the random variation in the amount of infectious virus particles (10 PFU) present in the inoculum (50 µl). The possible variation in the inoculum stock was therefore determined by performing a plaque assay in triplicate immediately after the mice were infected. This revealed that the number of plaques in 50 µl of the 10 PFU/50 µl inoculum ranged from 4 to 8. Therefore, it is possible that the SCID mouse that displayed limited weight loss after challenge may in practice have had received too few PFUs.
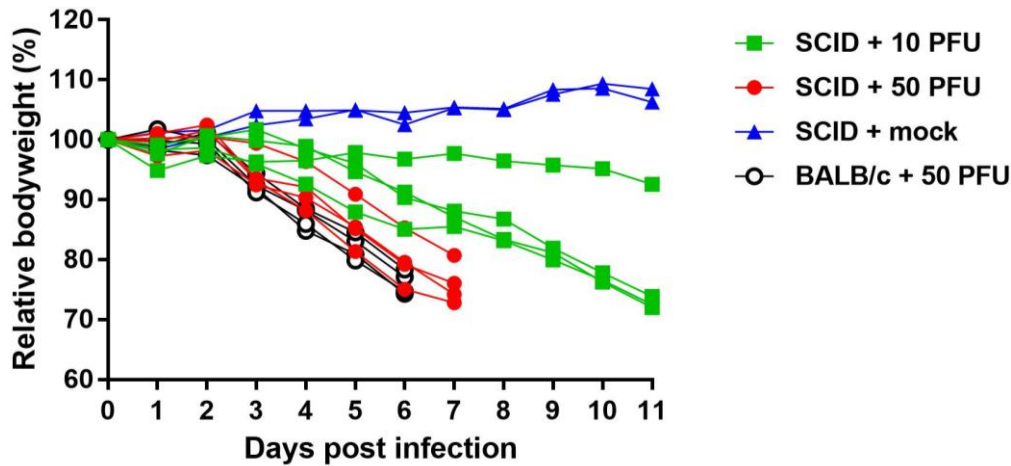
Figure 1: 10 PFU of PR8 virus is a suitable infection dose to allow multiple rounds of viral replication in SCID mice. SCID (n = 4 per group) or BALB/c (n = 2) mice were infected with 10 or 50 PFU of PR8, or mock-infected (n = 2). Shown is the relative bodyweight per mouse after infection.

A viral dose of 10 PFU allows at least 11 days of viral replication before the mice reach their ethical experimental endpoint. We reasoned that this was a convenient time span to allow the emergence of anti-M2e IgG escape mutants. Next, the influenza-specific RT-PCR protocol was optimized for samples with a low viral titer, such as bronchoalveolar lavage (BAL) fluid isolated from infected mice, to ensure successful amplification of all eight genomic RNA segments (Figure 2) [33].
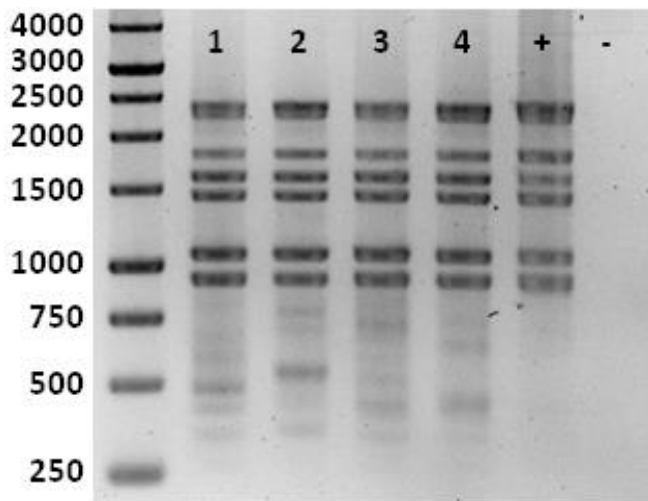
Figure 2: Whole genome amplification of influenza A viruses isolated from BAL fluid of infected mice. Bronchoalveolar lavage (BAL) fluid was isolated 11 dpi from SCID mice infected with 10 PFU of PR8. An influenza-specific RT-PCR was performed on 200 µl of BAL fluid using primers targeting the conserved 5' and 3' termini of each genomic segment. After DNA purification, 500 ng of each sample was separated on length by gel electrophoresis on a 1% agarose gel. DNA fragments were stained using ethidium bromide. 1-4: influenza specific RT-PCR products on vRNA extracted from BAL fluid isolated from four different SCID mice. + = 500 ng of RT-PCR product of RNA isolated from PR8 stock virus. - : negative control for RT-PCR.
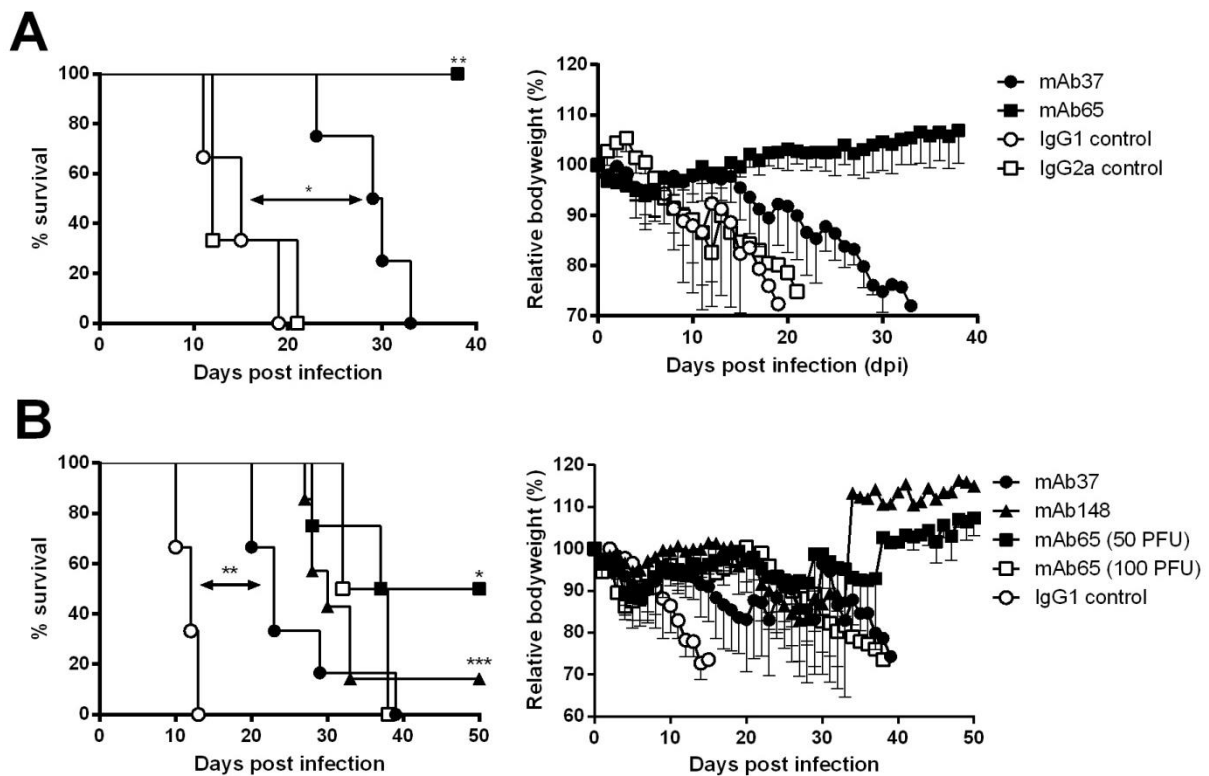
Figure 3: Anti-M2e mAbs significantly prolong survival in SCID mice. One day before infection, SCID mice were i.p. injected with 100 μg and at weekly intervals thereafter with 50 μg of mAb37, mAb65, mAb148 or negative control mAbs. Twenty-four h after the first mAb injection, mice were challenged with 10 PFU (two groups of mice treated with mAb65 were infected with either 50 (n = 4) or 100 PFU (n = 2), as mentioned in panel B) of PR8 virus. Survival (left) and body weight (right) were monitored daily. In the right hand graphs, data points represent averages and error bars standard deviations. (A) Treatment with mAb65 significantly prolonged survival compared to all other groups (n = 4, **: p = 0,0067). Treatment with mAb37 significantly prolonged survival compared to control treated groups (mAb37: n = 4, *: p = 0,0101; IgG1 and IgG2a controls: n = 3). (B) Treatment with mAb37, mAb148 and mAb65 (when infected with 50 PFU) significantly prolong survival compared to control treated groups (mAb37: n = 6, **: p = 0,0018; mAb148: n = 7, ***: p = 0,0008; mAb65: n = 4, *: p = 0,0101, IgG1 control: n = 3). The survival between groups was compared using the Logrank test.

Treatment with anti-M2e mAbs significantly prolongs survival of challenged SCID mice

Next, the protective effect of and genetic diversity induced by treatment of PR8-infected SCID mice with three different well-characterized anti-M2e mAbs was analysed. mAb37 (IgG1) and mAb65 (IgG2a) recognize a part of M2e that involves residues 5 to 15 in M2e and bind human consensus M2e with similar affinity, while mAb148 (IgG1) binds the 8 N-terminal aa residues in M2e [28, 29, 34]. Mice were treated with 100 μg of one of the three anti-M2e mAbs or isotype control mAbs one day before infection by intraperitoneal injection. This resulted in a serum antibody concentration of 40 to 80 μg/ml, based on M2e peptide ELISA, on the day of challenge (unpublished results). To compensate for the IgG half-life and to ensure constant M2e antibody pressure over time, subsequent antibody injections of 50 μg per mice were given in weekly intervals [35].

All anti-M2e mAbs significantly prolonged survival of PR8 challenged SCID mice compared to isotype control treated animals in two independent experiments (Figure 3). All mice that had been challenged with 10 PFU of virus and treated with mAb65 survived and displayed limited transient body weight loss early after infection (Figure 3.A). Using a five times higher infection dose, half of the mice treated with mAb65 were still alive at 50 days post infection whereas two of the SCID mice had to be euthanized on days 28 and 37 after challenge (Figure 3.B). mAb65-treated mice that had been challenged with 100 PFU of PR8 virus succumbed to infection by day 38 after infection. Treatment with mAb37 and mAb148, both IgG1 but recognizing a different epitope in M2e, resulted in comparable, significant prolongation of survival compared to control treatment and a statistically insignificant trend in favour of mAb148 (Figure 3.B).

mAb65 treatment of PR8 infected SCID mice reduces lung viral loads

Vaccination strategies based on M2e enhance the clearance of virus from the lungs and result in reduced lung viral titers in immune competent mice [2]. To evaluate the influence of M2e-specific mAbs on viral titer in the SCID model, we determined the viral titer in BAL fluid on day 7 and day 14 after infection and when mice had lost 25% of their initial body weight in the first experiment. Time points for the assessment of lung virus loads in the second mice experiment were 13 dpi (the day that the last control mouse had lost 25% of her initial bodyweight) and when the mice had lost 25% of their initial body weight. The viral titer in mice infected with 10 PFU and treated with mAb65 was significantly different from the IgG2a control group 7 dpi (p = 0.0378), 14 dpi (p = 0.0380) and at the end of the experiment (38 dpi, p = 0.0148) (Figure 4.A). Administration of mAb 65 even resulted in virus-clearance in most of the SCID mice in the first experiment. Challenge with 50 PFU of PR8 virus was still associated with viral clearance in half of the samples (Figure 4.B). In all anti-M2e treated moribund mice, viral titers were similar to the control antibody treated mice. No significant differences in viral titer between mice treated with M2-specifc mAbs mAb37 or mAb148 or control antibodies could be detected.
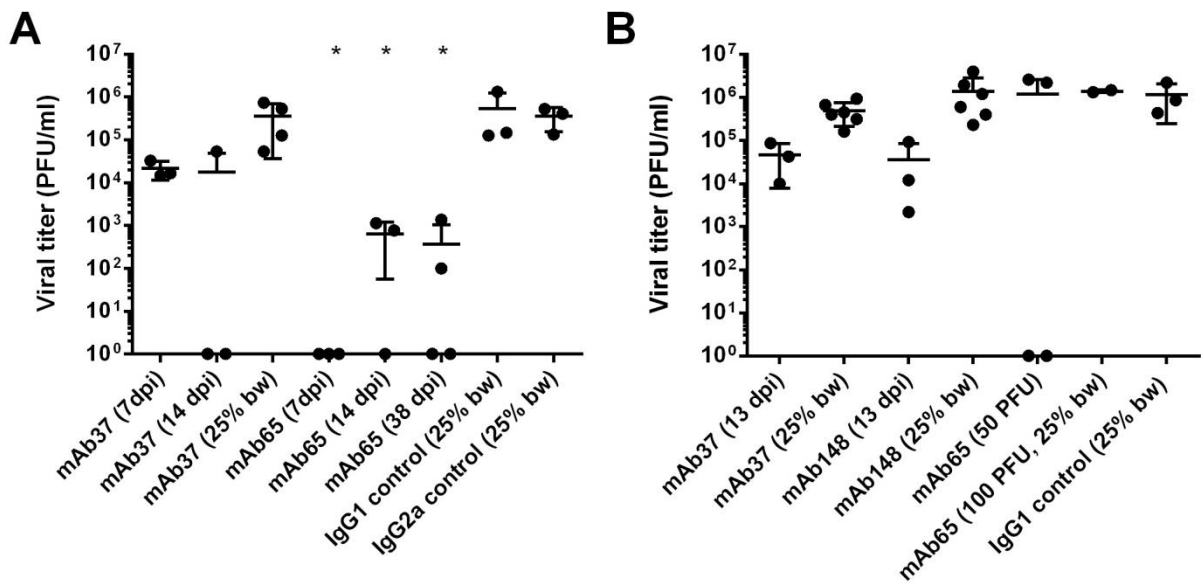
Figure 4: mAb65 treatment of PR8 infected SCID mice reduces lung viral titer. Mice were infected with 10 PFU of PR8 (unless otherwise mentioned) and BAL fluid was isolated on the indicated time point. The viral titer in BAL fluid was determined by viral plaque titration on MDCK cells. Panel A and B represent data obtained from independently performed mice experiments. (A) The viral titer in mAb65 treated mice was significantly different (Unpaired T-test, * = p < 0.05) from the IgG2a control group after infection with 10 PFU on 7 dpi (p = 0.0378), 14 dpi (p = 0.0380) and at the end of the experiment (38 dpi, p = 0.0148). (B) No significant differences in viral titer between mice treated with M2e-specific or control mAbs could be detected (unpaired T-test, p > 0.05). BAL fluid from mAb65 treated mice infected with 50 PFU was isolated at 50 dpi (n = 2, no virus detected) or when mice had lost at least 25% of their initial bodyweight (n = 2). 25% bw: viral titer in BAL fluid samples isolated from mice on the day when the ethical experimental endpoint of at least 25% loss of the initial body weight loss was reached.

Whole viral genome NGS sequence analysis

Whole genome analysis was performed using the workflow described before [33], with some adaptations in the influenza-specific RT-PCR protocol to allow amplification of viral RNA in samples with a lower viral titer. The purified RT-PCR samples of both mouse experiments (first experiment: n = 18, second experiment: n = 25) were sequenced in two separate Illumina MiSeq runs and the minimum, maximum and average coverage depth are listed in supplementary table S1. All variants detected at a frequency above 10% in these samples are listed in supplementary tables S2-S14.

A high sequence diversity within a given virus sample is often associated with increased viral fitness of the viral population. This has been shown *e.g.* for poliovirus, where a mutant virus carrying a polymerase with a higher than natural replication fidelity, which resulted in lower genetic diversity in the viral population, turned out to be less fit in competition assays [36, 37]. We therefore first determined the Shannon entropy, which is a direct measure for the variation in a given sample, for each position in the PR8 sequence to investigate if any of the viral populations present in the BAL fluid had an altered nucleotide sequence diversity, *e.g.* as a result of anti-M2e IgG selection pressure.

190

Within the same experiment, the Shannon entropy was similar in all samples for all conditions and all eight segments (Supplementary Figure S1). However, the mean Shannon entropy was significantly higher in the second compared with the first experiment (Supplementary Figure S1; p<0.001, unpaired two-sided Student t-test). In both experiments a different version of the 'MiSeq Reagent Kit' was used (v1 in the first and v2 in the second experiment), in which a different DNA polymerase is present and can perhaps explain the observed higher Shannon entropy in the second experiment

We next mapped all sequence variants that were detected at frequencies above 50% in M2e mAb treated samples to the primary structural maps of the viral proteins (Figure 5). The highest number of variants were present in the polymerases PB2 and PA and in HA. Distance-based clustering showed that the variants in the M2e antibody and control treated groups clustered in the same group. The variants detected in the M2e treated mice are thus not more closely related to each other than to the variants that were present in the control treated groups. In addition, no predictive polymorphism for either anti-M2e mAb-treated or control mAb treated viral samples could be determined using logistic regression.

Figure 5 (next page): Mutations selected under M2e-specifc mAbs mapped to the primary structures of the viral proteins. Mutations detected above 50% are mapped to the primary structural maps of the viral proteins. Mutations are indicated with arrowheads and mutations that were detected in more than one sample are marked with an asterisk. The interaction domains or functional sites are labeled and colour coded as described in [38], except for HA which is labeled as in [39]. Nuclear localisation signals (NLS) are indicated in black. The cap-binding domain of PB2 is marked in orange. Impα5: importin subunit α5, hCLE: human transcription factor, RBS: Receptor binding site, F: fusion peptide, TM: transmembrane domain, D: dsRNA/PABP1/RIG-1/E1B-AP5 interacting domain, E: eIF4G1/CPSF interacting domain, PABPI and II: poly-A binding protein 1, respectively 2; RIG-I: retinoic acid inducible gene I, E1B-AP5: E1B associated protein 5; CPSF: cleavage and polyadenyaltion specificity factor; eIF4GI: eukaryotic translation initiation factor 4G1 and PKR: Protein kinase R and NES: nuclear export signal.

PB2

309 315 373* 434 493,494 504* 726 740*

PB1 NP

124 242 269 282 363 404 449 495 538 577 580 683 687 736 753

PB1 NP Impα5

CAP binding NLS NLS

PB1

40 336 536

PA PB2

48 187 211 297 312 396 483 506 659

NLS Pol activity

PA

3 127 374 405 550 556

PB1

124 139 186 154 247 493 512 557 574 601 692

NLS NLS hCLE

HA

169* 203 209 231, 232 238* 257* 2 18 46 55 110 112, 113

HA1 HA2

1 23 185 214

RBS F TM

NP

303 377 395*

PB2 PB2 NP-NP NP-NP

13 161 189 198 216 338 358 371 405

NLS NLS Actin

M2

10* 11

24 42

Ion channel

NS1

NLS1 NES 204 NLS2

34 38 73 81 113 123 138 188 216

D E PKR CPSF PABPII

NS2

47

11 23

NES

M2e escape mutants are selected in vivo with mAb37 or mAb65 but not with mAb148

The sequence variation of M2e in the viral stock used to infect the mice and in all mice that received control treatment in the first mice experiment, was below the detection limit [33]. Three types of variation were detected in M2e in virus samples that were derived from mAb37 and mAb65 treated mice: M2-P10H, -P10L and -I11T. These M2e mutations were either present alone, or in combination in the samples (Figure 6). The M2e mutations were detected in 6 out of 10 mice that were treated with mAb37 that reached the ethical endpoint at a frequency that ranged from 0.62% to 90.26% (Figure 6). In the second experiment, where mice were infected with a viral dose of 50 or 100 PFU, a change in predicted M2e sequence was detected in all four mice that had lost 25% of their initial body weight (and thus were euthanized) after mAb65 treatment with a frequency that ranged from 4.11% to 99.96% (Figure 6). Interestingly, virus in the BAL fluid of all mAb148 treated mice still had a wild type M2e sequence.
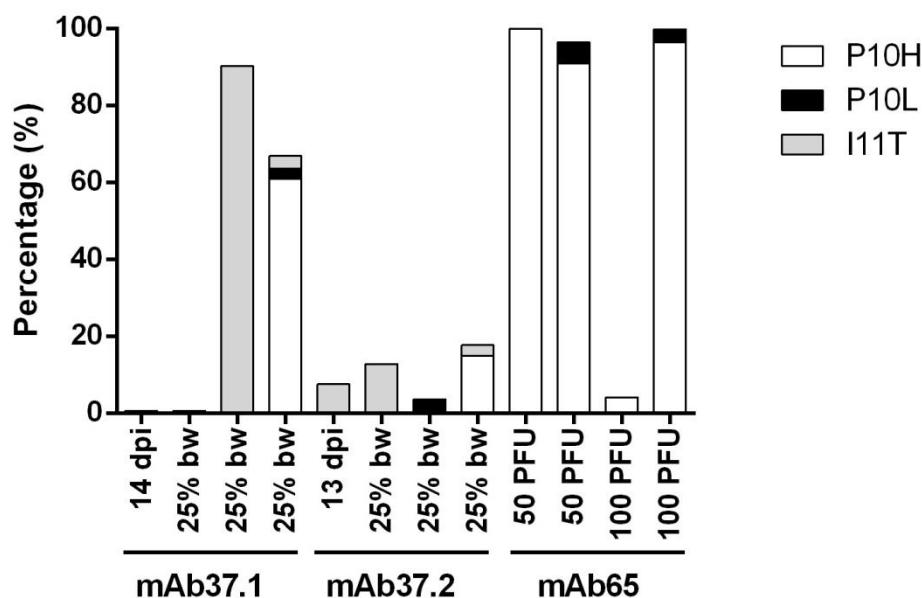
Figure 6: Changes in M2e after treatment with M2e-specific mAbs is limited to P10H, P10L and I11T. Infected SCID mice were treated with M2e-specific mAbs (mAb37 or mAb65). BAL fluid was isolated at the indicated time points, or when mice had lost 25% of their original body weight, followed by influenza-specific RT-PCR and the variation in M2e was determined by Illumina MiSeq sequencing. The results for mAb37.1 and mAb37.2 are from two independent experiments and mice were infected with 10 PFU of PR8 virus. BAL fluid from mAb65 treated mice was isolated when mice had lost at least 25% of their initial body weight after infection with either 50 or 100 PFU of PR8 virus.

The variation in the M2e sequence was limited to position C762A (P10H) or C762T (P10L) for mAb65. In mAb37 treated mice, also an additional mutation at position T765C (I11T) was detected. These three mutations are all synonymous in M1, hinting for a strong genetic selection pressure on M1. We previously reported that mAb37 and -65 recognize an overlapping epitope in M2e and bind to M2e as present in PR8 virus with similar affinity [34]. However, the I11T mutation in M2e is only detected

in mAb37 treated mice, suggesting that position 11 of M2e is less critical for binding of mAb65 to M2e. A difference in epitope specificity between mAb37 and mAb65 was confirmed by ELISA using different M2e variant peptides (Supplementary Figure S2).

To confirm that the observed mutations in M2e that were present in virus samples from mAb37 and -65 treated SCID mice are genetically stable and abolish binding by these anti-M2e mAbs, we generated recombinant PR8 virus with these mutations by reverse genetics [40]. The rescued viruses were plaque purified and the presence of the introduced mutations after large-scale virus production of the viruses was confirmed by Sanger sequencing of the RT-PCR-amplified M segment. All three mutations were homogenously present in the respective virus stocks, which suggests that they are not detrimental for *in vitro* viral replication (Figure 7).
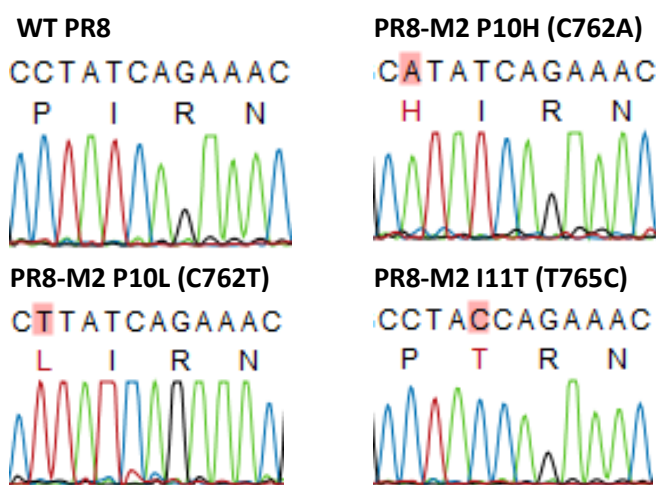
Figure 7: The P10H, P10L and I11T mutations in M2e of PR8 virus are stable during *in vitro* virus growth. Wild type PR8, PR8-M2 P10H, PR8-M2 P10L and PR8-M2 I11T viruses were amplified on MDCK cells to produce the respective stock viruses. The Sanger sequencing profile of RT-PCR amplified M segment, isolated from the different stock viruses with the introduced mutation highlighted in red, is shown. The introduced single nucleotide substitution (M segment numbering) which results in the mutated M2e sequence is added in parenthesis.

The ability of the M2e-specific mAbs to bind to these M2 mutant viruses was investigated using cell-ELISA. mAb148 bound to cells that had been infected with any of the four M2e variant PR8 viruses (Figure 8). In contrast to wild type virus, foci of PR8 viruses with M2e-P10L or -P10H were not recognized by mAb37 or mAb65 (Figure 8). In addition, mAb37 bound only weakly to PR8-M2 I11T while mAb65 retained a similar binding capacity as for the wild type PR8 foci. mAb148 could bind to all tested M2e variants (Figure 8) This differential recognition of virus with M2e-I11T is in line with the NGS results of the *in vivo* escape selection.

194
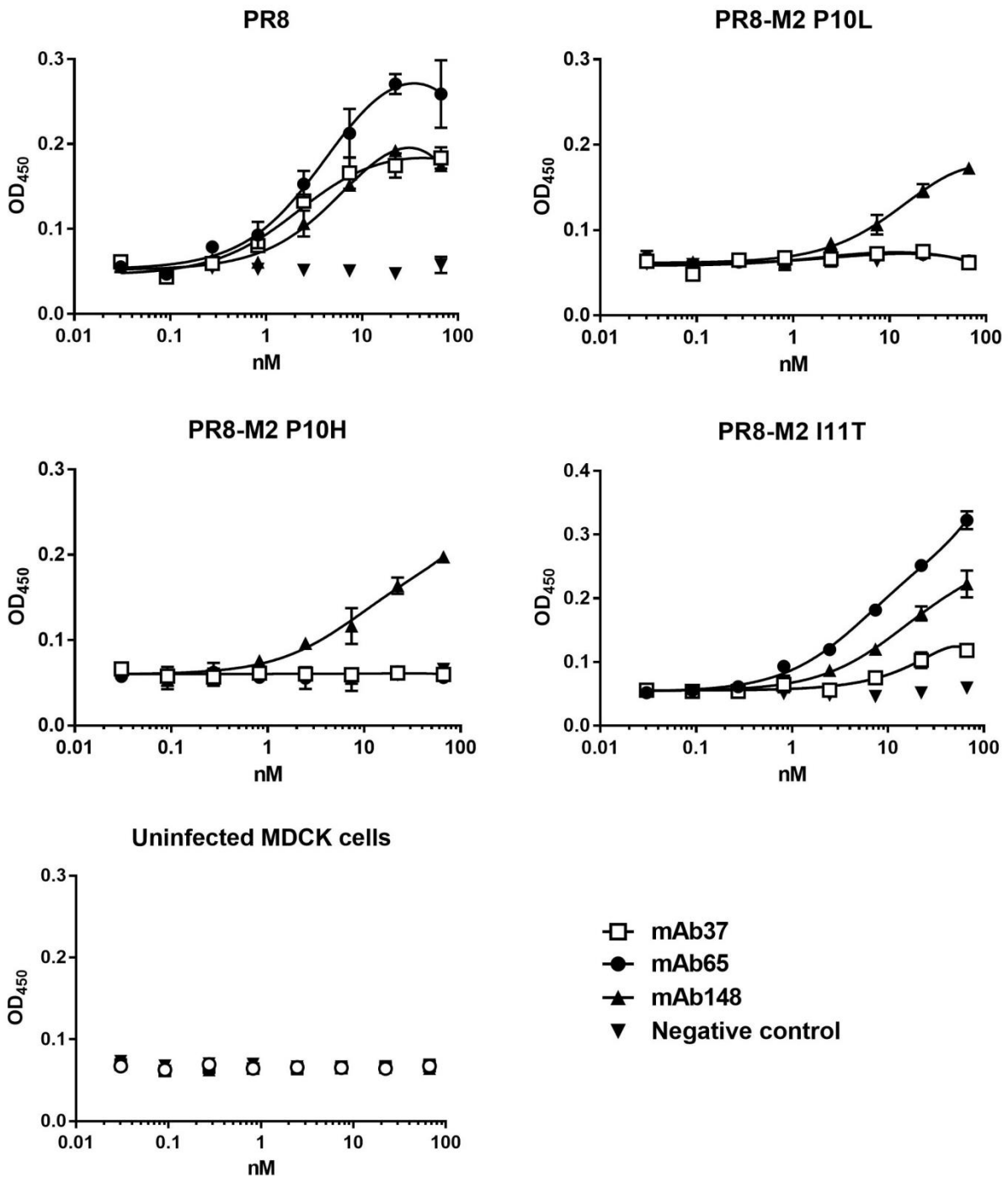
Figure 8: mAb65 and mAb37 fail to bind to M2e when a proline or histidine is present at position 10. MDCK cells were infected with A/Puerto Rico/8/34 (PR8) virus or a PR8 virus that carries a P10L, P10H or I11T M2e variant. Twenty-four hours later, the cells were incubated with a dilution series of mAb37, mAb65, mAb148 or negative control (0.5% BSA in PBS) and fixed with 4% paraformaldehyde. Binding of the M2e-specific mAbs was determined using horseradish peroxidase-conjugated sheep anti-mouse IgG as secondary Ab.

The M2e escape mutants were selected using monoclonal antibodies whereas the humoral immune response elicited by M2e-based vaccines is polyclonal. Consequently, the potential of polyclonal M2e immune serum to bind these M2e variant viruses was analyzed using cell-ELISA. Foci of cells that had

been infected with PR8 viruses carrying wild type M2e, M2e-P10L, -P10H or -I11T mutations were still recognized by serum IgG derived from mice that had been vaccinated with M2e-HBc virus-like particles that contained three tandem copies of M2e: human consensus M2e, P10L M2e and M2e containing both P10L and I11T (Figure 9). In addition, antibodies of both the IgG1 and IgG2a isotype in this M2e vaccine induced mouse serum could bind to the different M2e escape viruses (Figure 9). These results suggest that monoclonal antibody escape in the M2e epitope still retains sufficient antigenicity for recognition by a polyclonal anti-M2e immune serum.
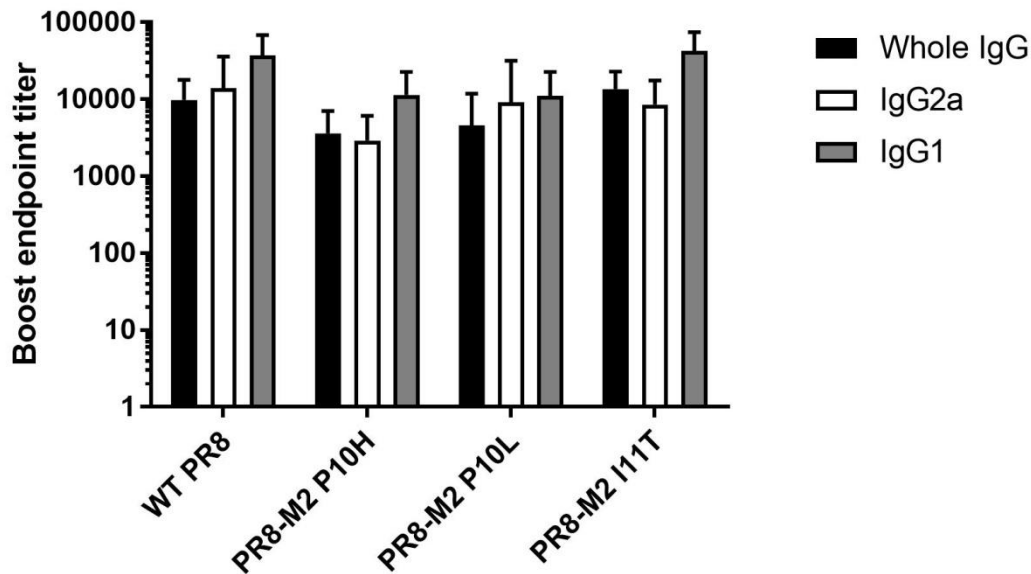
Figure 9: M2e variant viruses are recognized by polyclonal M2e immune serum. MDCK cells were infected with A/Puerto Rico/8/34 (PR8) virus or a PR8 virus carrying an M2 variant (PR8 M2-P10H, -P10L or -I11T). Twenty-four hours later, the cells were fixed with 4% paraformaldehyde and incubated with an individual dilution series of immune serum per M2e-HBc vaccinated mouse (n = 10). Binding of the M2e-specific mAbs was determined using horseradish peroxidase-conjugated sheep anti-mouse IgG, goat anti-mouse IgG2a or goat anti-mouse IgG1 as secondary Ab. Error bars represent the standard deviation.

Delayed M2 expression represents a possible alternative route of escape from anti-M2e IgG immune pressure.

While determining the virus titers present in the BAL samples of the SCID mice that had been treated with mAb37 and which had a wild type M2 sequence, we noticed a remarkable plaque morphology. The plaque immune-staining with mAb37 resulted in a smaller halo compared to plaques from control antibody-treated mice (Figure 10.A). NGS analysis of virus in this sample, had revealed five mutations that were present in more than half of the viruses in the sample: 62.7 % K443R in PB2, 98.09% I550T in PA and 99.80% silent mutation (G679A), 96.32% A231S and 96.15% I361M (I18M in HA2) in HA. Recombinant mutant viruses were created containing one or a combination of these sequence variants. The recombinant virus containing all five mutations was named PR8-HPP virus. Compared to wild type PR8, plaques of PR8-HPP infected MDCK cells had a smaller halo after anti-

M2e mAb staining (Figure 10.B). This smaller halo compared to the one visualized with anti-vRNP serum suggests that the timing or levels of M2 expression could possibly be altered in the PR8-HPP virus (Figure 10.B and Supplementary Figure S3).
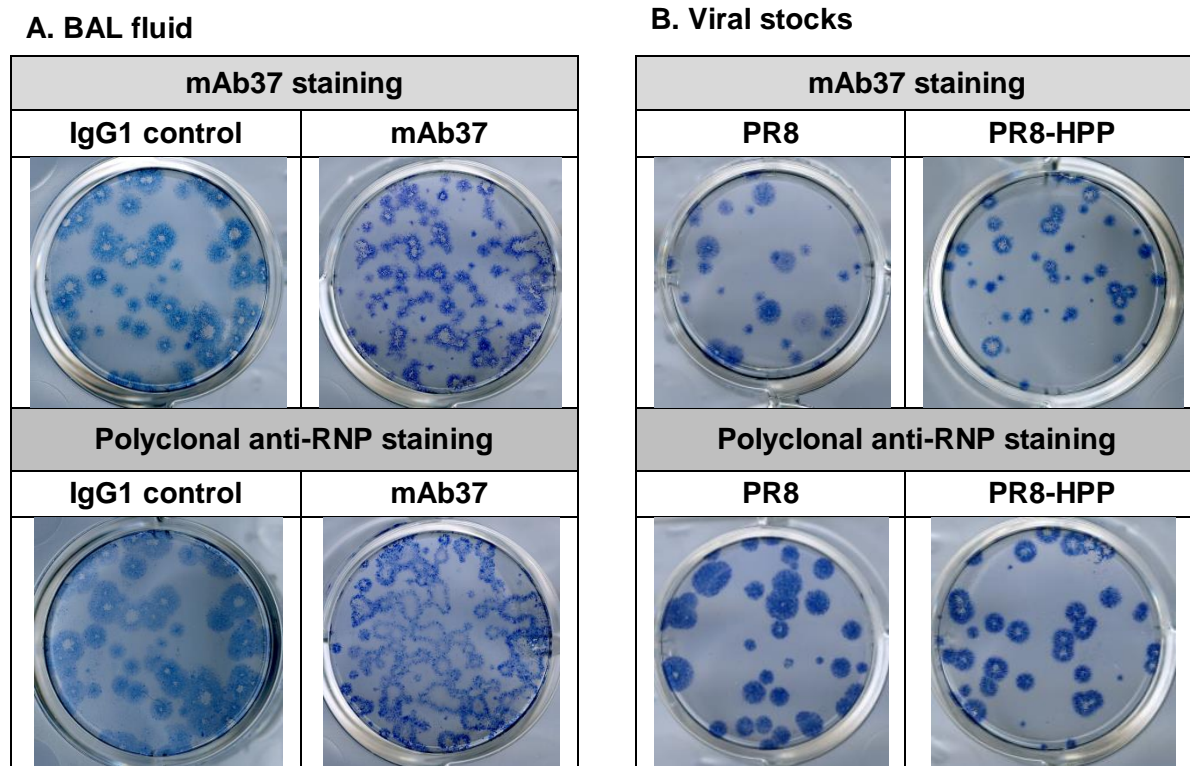
Figure 10: Plaques of virus isolated from a PR8-infected, mAb37-treated mouse and PR8-HPP show reduced M2e staining. MDCK cells were infected and fixed with 4% paraformaldehyde two days later. The plaques were revealed by immunostaining using an anti-M2e specific mAb (mAb37) or polyclonal anti-RNP Abs, followed by horseradish peroxidase-conjugated sheep anti-mouse IgG, respectively donkey anti-goat IgG, and TMB staining. (A) MDCK cells were infected with virus present in BAL fluid isolated from mice that had to be euthanized after infection with PR8 and IgG1 control mAb (left panel) or mAb37 (right panel) treatment. Plaques were first stained with mAb37 (upper panel) and subsequently with polyclonal anti-RNP (lower panel). Sequence analysis revealed that viruses in these BAL samples had a wild type M2e sequence. However, five mutations were present above 50% in the viral population of the mAb37-treated sample: K443R in PB2, I550T in PA and a silent mutation (G679A), A231S and I361M (I18M in HA2) in HA. (B) MDCK cells were infected in parallel with PR8 (left panel) or PR8-HPP virus (right panel). The latter virus was generated by reverse genetics and carries mutations K443R in PB2, I550T in PA and a silent mutation (G679A), A231S and I361M (I18M in HA2) in HA. Plaques from different wells were stained with mAb37 (upper panel) or polyclonal anti-RNP (lower panel).

The hypothesis that influenza viruses can escape from anti-M2e mAbs selection pressure by delaying M2 display on the infected cell surface was investigated by performing *in vitro* virus growth kinetics experiments, in which surface expression of M2 and NA was determined at different time points after infection. Flow cytometric analysis of these samples did not show delayed surface expression of M2 for the PR8-HPP virus, compared to surface expression of NA (Figure 11). In addition, the kinetics

of surface expression of M2 and NA are similar to the ones observed for wild type PR8 virus, suggesting similar replication kinetics for both viruses (Supplementary Figure S4).
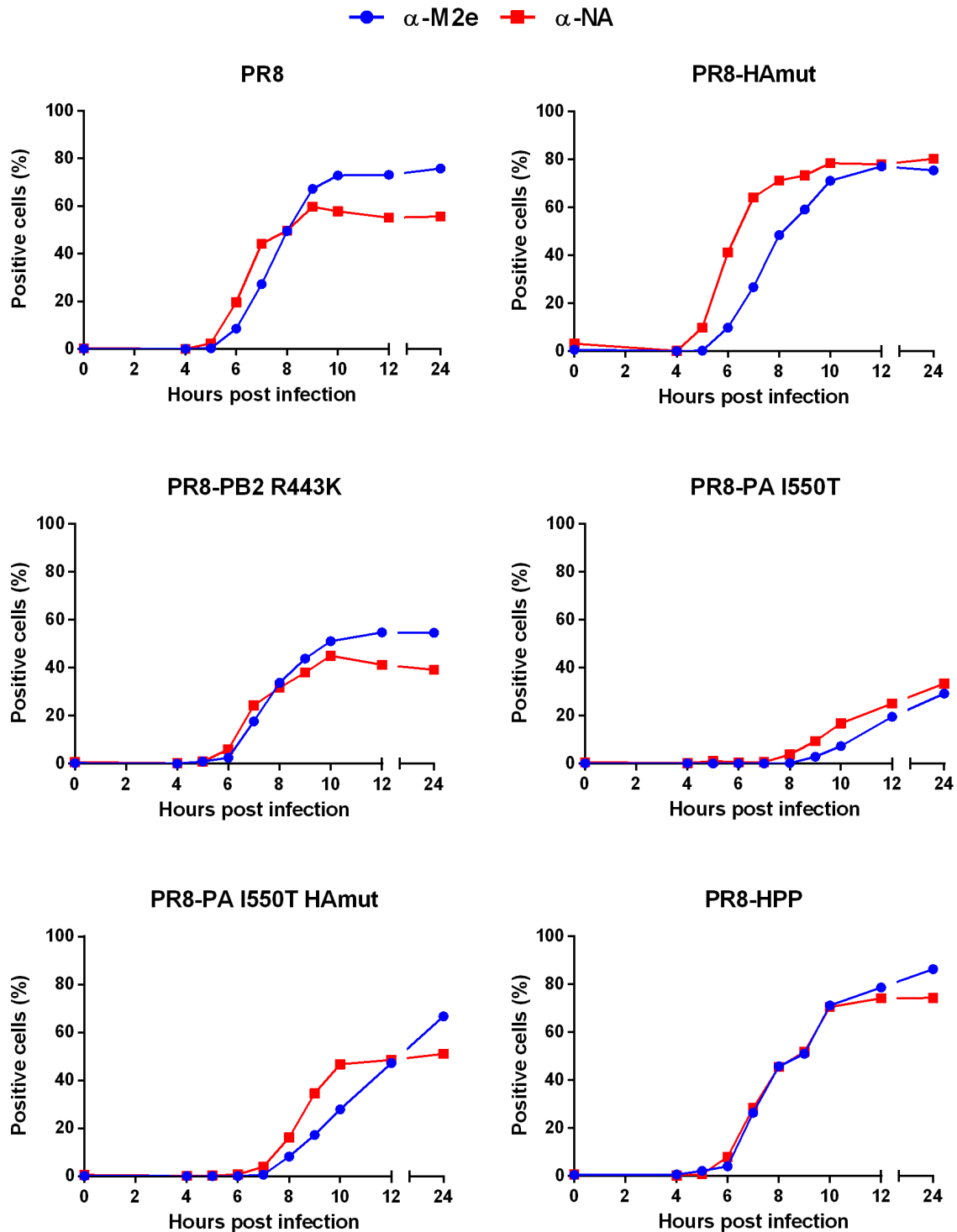
Figure 11: The kinetics for surface expression of M2e and NA are similar for PR8-HPP. HEK cells were infected with a moi of 0.5 and fixed with 2% paraformaldehyde at the indicated time points. The cells were either stained with mAb37 (M2e-specific mAb, blue) or 7D3 (NA-specific mAb, red), followed by a secondary anti-mouse antibody coupled to Alexa Fluor 488 and analysis on an LSRII HTS (BD) flow cytometer. The graphs are the representative of one out of two experiments.
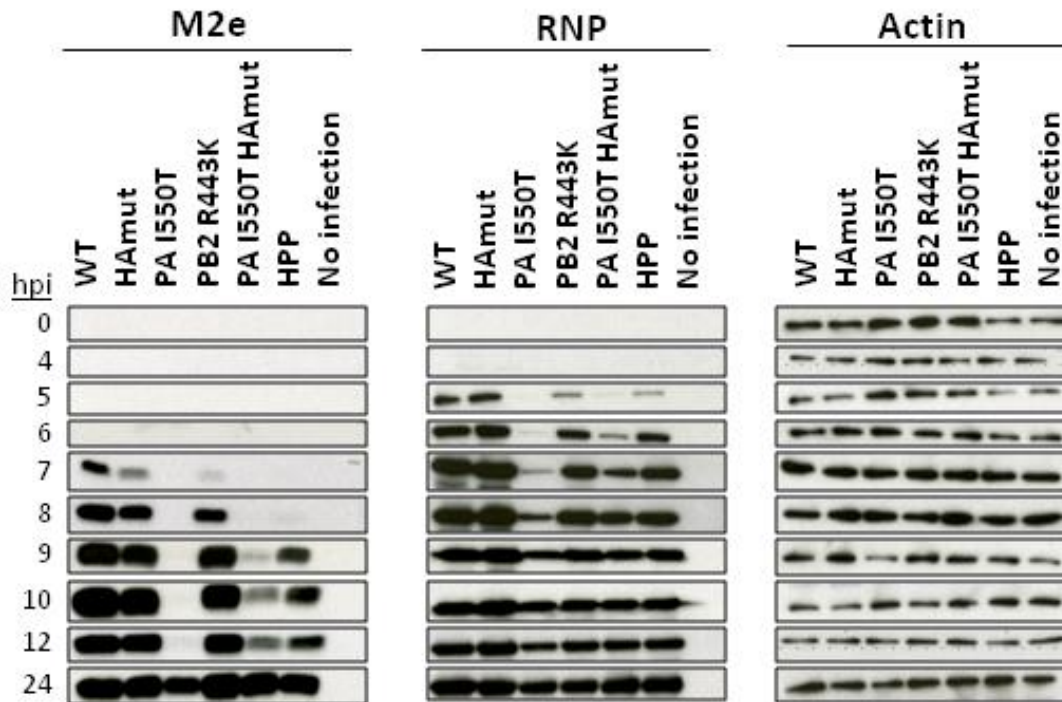
Figure 12: Expression of M2 is delayed in the PR8-HPP virus. HEK cells were infected at a moi of 0.5 with the indicated wild type or mutant PR8 viruses. Cell lysates were made on the indicated time points, followed by separation of the proteins on size on a 15% SDS-PAGE gel, western blotting and detection of M2e, RNP and actin. The graphs are the representative of one out of two experiments.

A delay in M2 expression of PR8-HPP was, however, evident after analyses of the cell lysates from the kinetics experiment by Western blotting (Figure 12). RNP-staining shows detectable expression of the vRNP proteins for both PR8 and PR8-HPP starting from five hours post infection. In contrast, M2 is detected starting from 7 hpi for PR8 and only from 9 hpi onwards for PR8-HPP (Figure 12).

In summary, these data shows that M2 expression is delayed in the PR8-HPP virus, when compared to wild type virus by Western blotting analysis. However, this observation is not reflected in the M2e surface expression based on flow cytometric analysis.

Discussion

Influenza is a vaccine preventable disease, but the current vaccination strategies are limited by their strain specificity and the ease by which the virus can mutate and evade the evoked humoral immune response. Currently, several influenza vaccines are being developed against conserved epitopes, aiming for broad-spectrum, long-lasting immunity. An important question to answer is: "Can influenza viruses also escape from immune pressure directed against more conserved epitopes?". The sequence conservation of an influenza peptide or protein sequence can be due to functional constraints or the consequence of a poor immune pressure against such epitopes. It is thus

important to investigate if and how influenza viruses can escape from the immunity elicited by broadly protective influenza vaccines.

In this study, we analyzed the genetic variability of influenza A viruses under immune pressure evoked by three different M2e-specific monoclonal antibodies in SCID mice, at levels similar to the M2e-specific serum antibody titers obtained after prime-boost vaccination of mice with M2e-MAP [41]. The absence of T- and B-cells in SCID mice creates a very controlled environment to study the viral diversity under chronic M2e treatment. All three M2e-specific mAbs significantly prolonged the survival of infected mice when compared to the control group. In addition, IgG2a mAb65 significantly prolonged the survival of challenged mice when compared to mAb37 (IgG1) treatment. This is in line with our previous results, showing that mAb65 can bind to all activating FcγRs, while mAb37 is limited in binding to activating receptor FcγRIII [34].

Whole genome analysis of the isolated viruses suggests that the virus can follow different pathways to circumvent the M2e-specific antibody pressure. The most obvious one is by favouring the outgrowth of a virus that has a mutation in the epitope recognized by the mAbs. Mutations in M2 were detected in mice that had been treated with mAb37 and mAb65, but not in virus derived from mice that were treated with mAb148. In the mAb37 and mAb65 treated mice, diversity was only observed at positions 10 and 11 and limited to P10L/H and I11T, which appeared either alone or combined. These mutations are all silent for M1, suggesting that variation in M2e is limited by genetic restriction to M1. A threonine at position 11 in M2 is the consensus in avian influenza viruses, demonstrating the genetic flexibility at that position. The observed variants are also in line with the crystal structure of mAb65 that showed a critical role for Glu6, Pro10, Ile11 and Trp15 for mAb binding, although it was surprising that none of the mAb65-treated mice had virus with a Ile11 alteration in M2 as was found in mAb37-treated mice [28]. The absence of variation at Glu6 and Trp15 is due to genetic constraints in M1: Glu6 is translated in frame with M1 and every nucleotide substitution in the codon encoding for Trp15 results in alteration of Met244 or Gly245 in M1.

Active vaccination strategies induce an oligoclonal anti-M2e response [42]. We showed that the M2e escape mutants are still recognized by M2e immune serum that was raised with an M2e-virus like particle construct. In addition, Wolf *et al.* reported that mice vaccinated with M2e-MAP, which carries four identical human consensus M2e side chains, were fully protected against challenge with PR8-M2 P10L or PR8-M2 P10H virus [43]. Moreover, in a review article by Gerhard *et al.*, it was stated that when PR8 infected SCID mice were treated with a combination of mAbs which can bind to wild type M2e and the P10H and P10L M2e variants, also no escape mutants emerged [44]. The same review also states that eleven consecutive passages of PR8 virus in M2e-vaccinated BALB/c mice did not result in selection of a single M2e escape virus [44].

Virus isolated from mAb148 treated mice carried wild type M2. This suggests that there is no or very little room for alterations in the sequence encoding the first nine amino acids of M2. In addition,

mAb148 could still bind to cells that had been infected with the P10H/L and I11T M2e escape variant viruses. This is in line with the crystal structure of mAb148 bound to M2e, where the deep and narrow binding pocket formed by the complementarity-determining regions (CDRs) of mAb148 accommodates the N-terminal part of M2, with Pro10 and Ile11 emerging out of the binding pocket [29]. Furthermore, treating experimentally infected individuals with a human M2e-specific monoclonal antibody (TCN-032), did not result in low frequency variations within the SLLTE (M2e amino acid residues 2-6) epitope or elsewhere in the M2 sequence in the virus that was present at the latest time points in positive nasal swab samples, as determined by 454 deep sequencing [3]. All these findings suggest that only limited variability in M2e is tolerated, presumably as a consequence of the genetic constraint imposed by M1.

In half of the BAL fluid samples of moribund mAb37 and mAb65 treated mice, and in all moribund mAb148-treated mice, virus was isolated with a wild type M2 sequence, but with non-synonymous mutations in other gene segments, principally in the polymerases and HA. In addition, mutations at high frequency were also detected in other viral proteins in samples containing variation in the M2 sequence. The high diversity in these selected variants demonstrates that the virus can follow multiple evolutionary pathways to escape to M2e immune pressure. The functions of some of these mutations have been described in literature. *E.g.* the I504V mutation in PB2 of PR8 displays a higher replication efficiency than the parental virus [45, 46]. Several of the mutations that were found in the current study (Supplementary tables S1-S13) have also been acquired during viral mouse-adaptation experiments, *e.g.* D740N in PB2, M205I in PB1 and P199S in HA [38, 47, 48].

Our results reveal that escape by delayed M2 production compared to wild type viruses represents a possible alternative escape route. It should be mentioned that the kinetic experiments were performed in HEK cells, which are of human kidney origin. To better mimic the *in vivo* situation, the kinetic should be repeated using a more homologous system, such as the mouse MLE-15 cell line. Determination of the viral titer in the supernatants should then also be included since this will learn if the window between the start of M2 expression and viral release in the mutant virus is shorter compared to that of wild type virus. A shorter window may enable virus release before the immune cells can attack the antibody-bound M2 expressing cells. Determination of the viral dose that results in 50% lethality (LD_{50}) will allow us to compare the pathogenicity of PR8-HPP with that of wild type virus. Subsequently, the delayed M2 expression should be verified *in vivo* by immunochemistry. In addition, the potential of PR8-HPP to evade M2e immune pressure should be studied in vaccinated, immunocompetent mice.

In all the treatment groups, both control mice and mice treated with M2e-specific mAbs, virus with mutations in the polymerases emerged above 10%. The I504V mutation in PB2 was present above 10% in virus isolated from half of the anti-M2e mAb treated mice, and virus isolated from 20 % of the mice carry a mutated position 550 in PA. It has been reported that the I504V mutation in PB2 and I550L in PA have enhanced *in vitro* polymerase activity and higher pathogenicity in mice [45, 46]. In

addition, several other mutations that have been linked to increased virulence are present in the polymerases at frequencies above 10%: *e.g.* D309N [49], R318K [47, 50], R355K [50-52] in PB2 and V127I [47, 53] in PA. Another strategy to escape to M2e immune pressure could thus be by gaining faster replication and associated enhanced pathogenicity in the host. HvPR8, a PR8 variant which is virulent in Mx1-congenic mice although it exhibits normal sensitivity to growth restriction by this protein, can escape to the induced antiviral response by enhanced replication [48]. Several variants in PB2 and PA - some with unknown function together with some for which the increased polymerase activity has been described - were tested using a minireplicon system in HEK cells. However, in all tested combinations, the mutant polymerase complex showed no significant increased activity compared to the wild type polymerases (unpublished results). It remains possible, however, that the selected mutants are optimized for replication in murine cells. This should be verified using the minireplicon in murine cells, *e.g.* L929 cells.

The fitness of influenza viruses can also increase by a mutation in the receptor binding domain (RBD) of HA that increases the binding to its receptor. The crystal structure of HA from the PR8 virus revealed that it can bind similarly to α2,3- and α2,6 linked sialic acids (SA) [54]. However, the mouse lung predominantly contains α-2,3 linked SA's [55]. A D238G (D225G in H3 numbering) mutation was detected in several samples at a high concentration. This mutation is described to increase binding to α2,3-linked SA [56, 57]. In addition, the detected E169K (E156K in H3 numbering) mutation in HA increases its avidity for receptor binding [58].

Escape from M2e-specific immune pressure in a chronically infected immune compromised host is thus a complex process for which influenza viruses may use several strategies. Escape in the M2e epitope can likely be prevented by including different M2e variants in the vaccine. The chance that viruses will emerge with increased virulence in the human population when applying M2e vaccination, is unclear. Since M2e-based vaccines are not virus neutralizing but infection-permissive, both the humoral and cellular immune system will be stimulated upon infection, which will likely result in viral clearance before virus adaptation can take place. Consequently, an important next step is to mimic a more natural situation by evaluating the viral diversity in immunocompetent mice after M2e-based vaccination. This would require serial passaging of the virus, using samples taken prior to virus control by the adaptive immune response that is elicited in such a host.

Material and Methods

Ethics statement.

All animal experiments described in this study were conducted according to the national (Belgian Law 14/08/1986 and 22/12/2003, Belgian Royal Decree 06/04/2010) and European legislation on the protection of animals used for scientific purposes (EU Directives 2010/63/EU, 86/609/EEC). All experiments on mice and animal protocols were approved by the ethics committee of Ghent University (permit numbers EC2012-034). All efforts were made to avoid and ameliorate suffering of animals.

Cell lines.

MDCK cells were cultured in Dulbecco's Modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, non-essential amino acids, 2 mM L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in 5% CO_2. HEK293T cells were cultured in the same medium, with the addition of 0.4 mM sodium pyruvate.

Generation of monoclonal antibodies.

Hybridomas were derived from splenocytes isolated from BALB/c mice i.p. immunized three times (three weeks interval) with 10 μg M2e-tGCN4, a soluble tetrameric M2e-antigen described in [59], adjuvanted with MPL+TDM adjuvant system (Sigma). Hybridomas were screened in M2e peptide and M2e-tGCN4 ELISA for the presence of M2e-specific IgG. After subcloning, scaled up cultures of hybridomas 37, 65 and 148 were used for purification of mAb37, -65 and -148 using protein A sepharose columns (GE Healthcare, Uppsala, Sweden). Isotype control mAb directed against the ectodomain of NB of influenza B virus (IgG1) or the Small Hydrophobic protein of human respiratory syncytial virus (IgG2a) were produced and purified as above and were used as irrelevant antigen-specific controls in passive transfer experiments.

Infection, treatment, and analysis of mice.

C.B-17/IcrHan®Hsd-Prkdc^{scid} (SCID) mice were purchased from Harlan and BALB/c mice from Charles River (France). Mice were used at the age of 7-8 weeks and were housed in individually ventilated cages, in a temperature-controlled environment with 12 h light/dark cycles and food and water *ad libitum.* Influenza virus challenge experiments were performed in biosafety level 2$^+$ facilities. To determine the viral inoculum which allows sufficient rounds of viral replication to select for M2e escape mutants, mice were anesthetized with a mixture of ketamine (10 mg/kg) and xylazine (60 mg/kg) and challenged by intranasal administration of 50 μl PBS containing 10 or 50 PFU of PR8 or mock-infected with PBS only. To select for M2e escape mutants, mice were first injected i.p. with 100 μg (in a 200 μl volume) of mAb37, -65, -148, or isotype control mAbs. Twenty four h later, mice were anesthetized with a mixture of ketamine (10 mg/kg) and xylazine (60 mg/kg) and challenged by intranasal administration of 50 μl PBS containing 10 PFU of PR8 (50 or 100 PFU for the mAb65 treatment in the second experiment). Body weight and survival of mice was monitored daily after

challenge. When mice lost more than 25% of their body weight, or on different dates as mentioned otherwise, the mice were euthanized and bronchoalveolar lavage (BAL) fluid prepared by flushing the lungs two times with 600 µl of PBS. Cells in this BAL fluid were pelleted by centrifugation and the supernatants stored at -80°C.

Peptide ELISA and cell-based ELISA.
For peptide ELISA, 50 µl of 2 µg/ml wild-type M2e (SLLTEVETPIRNEWGCRCNDSSDSG) or M2e variants were used to coat the wells of a 96-well Maxisorp plate overnight at 37°C. Wells were blocked with 3% skim milk in PBS buffer, followed by incubation with a dilution series of mouse serum as primary antibody. After vigorous washing with 0.1% Tween-20 in PBS buffer, binding was detected using horseradish peroxidase (HRP)-conjugated sheep anti-mouse IgG Abs. For cell-based ELISA, MDCK cells were seeded in a 96-well plate at 25.000 cells per well. Sixteen hours later, the cells were infected with a multiplicity of infection (MOI) of 1 of PR8 virus or the PR8-M2e mutant viruses (PR8-M2 P10H, PR8-M2 P10L or PR8-M2 I11T). For cell-ELISA using mAb37, mAb65 and mAb148: Twenty-four hours later, the cells were blocked with 0.5% BSA in PBS and incubated with mAb37, mAb65 or mAb148, followed by fixation with 4% PFA. For cell-ELISA using M2e-HBc immune serum: Twenty-four hours later, the cells were fixed with 4% paraformaldehyde (PFA), blocked with 3% skim milk in PBS buffer, followed by incubation with a dilution series of polyclonal M2e-HBc serum. This serum was obtained two weeks after prime-boost vaccination (intramuscular immunization using Alum adjuvant, with 3 weeks interval) with 10 µg M2e-HBc, containing three tandem copies of M2e: wild type M2e, P10L M2e and M2e containing both P10L and I11T. The detection was performed in the same way as described above in the peptide ELISA. For the cell-ELISAs using M2e-HBc immune serum, also detection with HRP-conjugated sheep anti-mouse IgG1 or IgG2a was included. The antibody titer in the individual immune serum of mice vaccinated with M2e-HBc was determined as the last antibody dilution having at least a two-fold higher $OD_{450-655}$ compared to the same dilution of negative control HBc-only serum.

Plaque assay.
MDCK cells were seeded in complete DMEM in 12-well plates at 3×10^5 cells per well. After 18 h, the cells were washed once with serum-free medium and incubated (in triplicate) with a ten-fold dilution series of the virus (made in serum-free cell culture medium containing 0.1% BSA) in 500 µl medium. After 1 h incubation at 37°C, an overlay of 500 µl of 1.2% Avicel RC-591 (FMC Biopolymer) in serum-free medium with 4 µg/ml TPCK-treated trypsin (Sigma) was added. After incubation at 37°C for 40 h, the overlay was removed and the cells were fixed with 4% paraformaldehyde and permeabilized with 20 mM glycine and 0.5% (v/v) Triton X-100. Plaques were stained with anti-M2e mAb37 (final concentration 0.5 µg/ml) or polyclonal goat anti-influenza ribonucleoprotein (RNP) (Biodefense and Emerging Infections Resources Repository, NIAID, NIH, NR-4282, dilution: 1/3000), followed by a secondary anti-mouse, respectively anti-goat, IgG HRP-linked antibody (GE Healthcare). After washing, TrueBlue peroxidase substrate (KPL) was used to visualize the plaques.

RT-PCR on BAL fluid samples

The previously described RT-PCR protocol for full genome influenza virus amplification was used to amplify the viral RNA present in BAL fluid samples [33]. However, the lower viral titer in BAL fluid required further optimization of the RT-PCR conditions and the following adaptations were made to the protocol: the addition of 20 µg polyA RNA carrier to the lysis buffer used for RNA extraction, including the DNase I digestion step during RNA isolation, preheating of the RNA elution buffer at 70°C and the first 5 PCR cycli should be performed at a lower annealing temperature (45°C instead of 72°C). In addition, two separate cDNA synthesis reactions were performed, using primers specific for the influenza A vRNAs: CommonUni12G (GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG) and CommonUni12A (GCCGGAGCTCTGCAGATATCAGCAAAAGCAGG). Subsequently, all eight genomic segments were amplified in one PCR reaction of 100 µl using a mix containing 5 µl of CommonUni12G and CommonUni12A cDNA, 200 nM primer CommonUni13 (GCCGGAGCTCTGCAGATATCAGTAGAAACAAGG) and the Phusion High Fidelity polymerase (Thermo Scientific) [33, 60, 61]. PCR products were purified using the High Pure PCR Product Purification kit (Roche) according to the manufacturer's instructions, and the product was eluted in 50 µl sterile ultrapure water (preheated to 65°C).

Illumina MiSeq library preparation and sequencing

300 ng (first experiment) or 150 ng (second experiment) of each RT-PCR sample was sheared with an M220 focused-ultrasonicator (Covaris) set to obtain peak fragment lengths of 300-400 bp. Next, the NEBNext Ultra DNA Library Preparation kit (New England Biolabs, second experiment: dual-indexing using NEBNext Multiplex oligos for Illumina) was used to repair the ends and to add the Illumina MiSeq-compatible barcode adapters to 100 ng of fragmented DNA. The resulting fragments were size-selected using Agencourt AMPure XP bead sizing (Beckman Coulter). Afterwards, indexes were added in a limited-cycle PCR (10 cycles), followed by purification on Agencourt AMpure XP beads. Fragments were analysed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies). The multiplex sample was heat denatured for 2 min at 96°C before loading on the Illumina MiSeq chip. After the 2×250 bp Illumina MiSeq paired-end sequencing run, the data were base called and reads with the same barcode were collected and assigned to a sample on the instrument, which generated Illumina FASTQ files (Phred +64 encoding).

Data analysis

The downstream data analyses were performed on the resulting Illumina FASTQ files (Phred +64 encoding) using CLC Genomics Workbench (Version 7.0.3) following the analysis pipeline as described in Van den Hoecke, *et al.*, with quality trimming of the sequencing reads to a Phred score of 30 [33]. For variant calling, the A-to-G variant introduced by the primer at position 24 in the HA, NP, NA, M and NS segments was not taken into account during the influenza quasispecies variant analysis. In addition, the Val458Met variant can be neglected since this mutation was already present for 95% in the PR8 virus stock and is possibly the result of plaque purification [33].

Generation and production of plasmids with escape mutations.

Reverse genetics plasmids for PR8 virus were kindly provided by Dr. Robert G. Webster (St. Jude Children's Research Hospital, Memphis, USA) [40]. The mutations in pHW197-M, pHW193-PA and pHW191-PB2 were introduced using quickchange site-directed mutagenesis (Stratagene). The HA segment carrying the G679A (silent), G743T (Ala231Ser), A1135G (Ile361Met) and G1424A (Val458Met) mutations was introduced by Gibson cloning into the pHW2000 vector. The HA gene was first amplified by RT-PCR on vRNA using HAfw (GAAGTTGGGGGGGAGCAAAAGCAGGgga) for cDNA synthesis (Transcriptor First Strand cDNA synthesis kit), and HAfw and HArv (CCGCCGGGTTATTAGTAGAAACAAGGgtg) for PCR. In parallel, overlapping ends were added to the pHW2000 vector by PCR following the manufacturer's protocol and flu-pHW-R (CCTGCTTTTGCTCCCCCCCAACTTC) and flu-pHW-F (CCTTGTTTCTACTAATAACCCGGCGG) as primers. DNA products were either purified from gel (for pHW2000) or from solution (for HA) using the High Pure PCR Product Purification kit (Roche). The Gibson reaction was performed according to the manufacturer's instructions. All plasmids encoding either the wild type or mutant PR8 genome segments were transformed and amplified in *E. coli* DH5α. Plasmid DNA was isolated with the Plasmid Midi Kit (Qiagen) according to the manufacturer's instructions. The resulting air-dried pellet was dissolved in 50 µl of sterile ultrapure water. The presence of the introduced mutations was confirmed by Sanger sequencing on a capillary sequencer (Applied Biosystems 3730XL DNA Analyzer).

Generation of wild type or mutant recombinant PR8 virus.

To generate recombinant wild type PR8 virus, 1 µg of pHW191-PB2, pHW192-PB1, pHW193-PA, pHW194-HA, pHW195-NP, pHW196-NA, pHW197-M and pHW198-NS was transfected using calcium phosphate co-precipitation into a HEK293T-MDCK cell co-culture in Opti-MEM (3 x 10^5 HEK293T and 2 x 10^5 MDCK cells in a 6-well plate). To generate mutant PR8 virus, the same set-up was followed, but one or more of the wild type pHW plasmids were replaced by the corresponding mutant plasmid. After 30 h, L-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin (Sigma) was added to a final concentration of 2 µg/ml. After 72 h, the culture medium was collected and the presence of virus was confirmed by hemagglutination of chicken red blood cells. Reverse genetics-generated PR8 and PR8 mutant viruses were plaque-purified on MDCK cells as follows. Confluent MDCK cells in a six-well plate were infected with a serial dilution series of virus. After 1 h, an overlay of low melting agarose (Type VII agarose, Sigma; final concentration 1%) in serum-free cell culture medium containing 4 µg/ml TPCK-treated trypsin (Sigma) was added. After 56 h, cytopathic effect was checked, agar overlaying viral plaques were selected with a pipette tip, and virus was allowed to diffuse from the agar for 24 h at 4°C in serum-free medium. Afterwards, virus derived from one plaque was amplified on MDCK cells in serum-free cell culture medium in the presence of 2 µg/ml TPCK-treated trypsin (Sigma). After 96 h, the culture medium was collected, and cell debris was removed by centrifugation for 10 min at 2,500 g at 4°C, and the virus was pelleted from the supernatants by overnight centrifugation at 16,000 g at 4°C. The pellet was dissolved in sterile 20%

glycerol in PBS, aliquoted and stored at -80°C. The infectious titer of the obtained virus stocks was determined by plaque assay on MDCK cells, on three different aliquots each performed in triplicate.

Determination of M2e sequence in viral stocks of PR8, PR8-M2 P10H, PR8-M2 P10L and PR8-M2 I11T

RNA was isolated using the High Pure RNA Isolation Kit (Roche) according to the manufacturer's instructions. cDNA synthesis was performed using the influenza Uni12 (AGCAAAAGCAGG) primer and the Transcriptor First Strand cDNA synthesis kit (Roche) according to the manufacturer's instructions for cDNA synthesis using gene-specific primers. The M segment was amplified using pHW-MPf (GAAGTTGGGGGGGGAGCAAAAGCAGGTAG), pHW-MPr (CCGCCGGGTTATTAGTAGAAACAAGGTAG), Phusion polymerase (Thermo Scientific) and the following conditions: an initial denaturation step of 98°C for 30'', followed by 5 cycli of 98°C for 10'', 45°C for 30'' and 72°C for 2' and 30 cycli of 98°C for 10'' and 72°C for 2'30'', and a final elongation step of 72°C for 7' [62]. Subsequently, the M segment was purified from 1% agarose gel (High Pure PCR Product Purification Kit, Roche) and the presence of the introduced mutations confirmed by Sanger sequencing on a capillary sequencer (Applied Biosystems 3730 XL DNA Analyzer).

Flow cytometric analysis

HEK293T cells were infected with moi 0.5 of either PR8, PR8-HAmut, PR8-PB2 R443K, PR8-PA I550T, PR8-PA I550T HAmut, PR8-HPP or uninfected. One hour later, the inoculum was removed, the cells washed once with PBS and replaced with Optimem (Gibco). Cells were detached with EDTA and successively washed with Optimem and PBS. One-tenth of these cells was used to make lysates for Western blot analysis. The remaining cells were fixed in 2% paraformaldehyde, followed by blocking the aspecific binding sites on the cells with 1% BSA in PBS. Half of the cells were stained with M2e-specific mAb37 (10 µg/ml) and the other half with NA-specific mAb 7D2 (10 µg/ml) in 0.5% BSA in PBS. Donkey anti-mouse Alexa-Fluor 488 (1/600, Invitrogen) was used as secondary antibody, followed by analysis on a LSRII HTS (BD) flow cytometer.

Western blot analysis

HEK cells were lysed in low salt lysis buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1% Igepal (NP-40) and 5 mM EDTA. The samples were separated by SDS-PAGE (15%) and visualized by Western Blotting with antibodies directed against M2e (1,3 µg/ml, mAb37), RNP (1/3000, polyclonal goat anti-RNP, Biodefense and Emerging Infections Resources Repository, NIAID, NIH, NR-4282) or actin (1/3000, mouse anti-actin monoclonal (clone: C4), Bio-connect).

Shannon entropy

To detect instances of convergent evolution we compared the mother stock's segment sequences with the newly assembled sample-specific majority rule consensus sequences [63]. The *de novo* assembly was successful for 340/344 (98.8%) segments. Failure due to a wet-lab error was ruled out because a check with the Mosaik aligner revealed that numerous reads (>20,000) mapped to the corresponding segment of the mother stock for each of the four erroneous assemblies [64]. All

assemblies were visually inspected and, if required, manually edited in AliView [65]. For optimal alignment accuracy, the newly generated segment assemblies were concatenated into sample-specific full genomes that served for the read mappings upon which a site-specific measure of nucleotide diversity (the standard Shannon entropy) was calculated [66]. The corresponding segments of the closely related mother stock served in place of the four failed segment assemblies, and in the full genomes the segments were separated from each other by strings of 50 Ns. Estimates of the synonymous and non-synonymous nucleotide diversity were obtained with SNPGenie [67]. The entropy levels of the virus populations from mice that received the same treatment (*i.e.* IgG1 and mAb37) in both sequencing experiments were contrasted with unpaired two-sided t-tests, and for these populations the proportion of sites with non-zero Shannon entropy levels was compared with Pearson's Chi-squared test with Yates' continuity correction. R was used for the statistical analyses, and $p \leq 0.05$ was taken as the cut-off for statistical significance [68].

Distance-based clustering

To *in silico* evaluate whether lower frequent variants can have a synergetic anti-M2e vaccine effect we clustered the virus populations using several distance measures that are based on the presence and prevalence of SNPs measured against the mother stock reference sequence. The used rules are: *1)* The distance is defined by the sum of unique polymorphic sites. In this setup, a site that is polymorphic in one population but homogenous in the other increases the distance between both populations with 1. *2)* As in 1, but the type of polymorphism is also taken into account. In an example in which a site in population 1 has 7% A and 9% G as polymorphic content, and the corresponding site in population 2 has 9%A and 0% G as non-reference nucleotides, the distance increases with 1. *3)* As in 2, but now the frequency of polymorphisms is also considered, and the distance increases with increasing prevalence differences. In the previous example the distance increases with 2+9=11%. *4)* Same as in 3, but the distance is averaged over the number of involved characters. In the example, the distance increases with 5,5%. *5)* Instead of the count, the distance is defined by the proportion of unique polymorphic sites. For example, population 1 has 10/40 popymorphic sites shared with population 2, and population 2 has 10/50 of its polymorphic sites in common. The distance between both is calculated as 30/40+40/50=1.55. *6)* Same as in 5, but the average proportion is used as a proxy for the distance. In the previous example the distance becomes 0.775. Distance calculations and hierarchical clustering visualisations were done in R for each of three SNP prevalence cutoffs (> 5%, > 25% and > 50%).

Logistic regression

We determined the association between the frequencies of SNPs and anti-M2e mAb treatment using a logistic regression. To increase the observation count for each outcome the results of both sequencing experiments were combined; the no compound, IgG1 and IgG2a treated populations were labeled as 'controls', and the other populations as 'treatment'. Because using the same lower thresholds for polymorphic sites as above (*e.g.* > 5%, > 25% and > 50%) created a quasi-complete segregation, parameter estimates were obtained with penalised likelihood methods [69].

ACKNOWLEDGEMENTS

References

1. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W: The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* 1982, 31(2 Pt 1):417-427.
2. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.
3. Ramos EL, Mitcham JL, Koller TD, Bonavia A, Usner DW, Balaratnam G, Fredlund P, Swiderek KM: Efficacy and safety of treatment with an anti-m2e monoclonal antibody in experimental human influenza. *The Journal of infectious diseases* 2015, 211(7):1038-1044.
4. De Filette M, Min Jou W, Birkett A, Lyons K, Schultz B, Tonkyro A, Resch S, Fiers W: Universal influenza A vaccine: optimization of M2-based constructs. *Virology* 2005, 337(1):149-161.
5. Schotsaert M, De Filette M, Fiers W, Saelens X: Universal M2 ectodomain-based influenza A vaccines: preclinical and clinical developments. *Expert review of vaccines* 2009, 8(4):499-508.
6. Acambis: Positive Phase I and Pre-Clinical Data Suggest Acambis' M2e-Based Universal Influenza Vaccine, ACAM-FLU-A™, Could Tackle Influenza Pandemics. 2008.
7. Holsinger LJ, Nichani D, Pinto LH, Lamb RA: Influenza A virus M2 ion channel protein: a structure-function analysis. *Journal of virology* 1994, 68(3):1551-1563.
8. Pielak RM, Chou JJ: Influenza M2 proton channels. *Biochimica et biophysica acta* 2011, 1808(2):522-529.
9. Lamb RA, Zebedee SL, Richardson CD: Influenza Virus-M2 Protein Is an Integral Membrane-Protein Expressed on the Infected-Cell Surface. *Cell* 1985, 40(3):627-633.
10. Zebedee SL, Lamb RA: Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of virology* 1988, 62(8):2762-2772.
11. Bouvier NM, Palese P: The biology of influenza viruses. *Vaccine* 2008, 26 Suppl 4:D49-53.
12. Hutchinson EC, Charles PD, Hester SS, Thomas B, Trudgian D, Martinez-Alonso M, Fodor E: Conserved and host-specific features of influenza virion architecture. *Nature communications* 2014, 5:4816.
13. Deng L, Cho KJ, Fiers W, Saelens X: M2e-Based Universal Influenza A Vaccines. *Vaccines* 2015, 3(1):105-136.
14. Feng J, Zhang M, Mozdzanowska K, Zharikova D, Hoff H, Wunner W, Couch RB, Gerhard W: Influenza A virus infection engenders a poor antibody response against the ectodomain of matrix protein 2. *Virology journal* 2006, 3:102.
15. Ito T, Gorman OT, Kawaoka Y, Bean WJ, Webster RG: Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins. *Journal of virology* 1991, 65(10):5491-5498.
16. Lamb RA, Choppin PW: Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. *Virology* 1981, 112(2):729-737.
17. Inglis SC, Brown CM: Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. *Nucleic acids research* 1981, 9(12):2727-2740.
18. Lamb RA, Lai CJ, Choppin PW: Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proceedings of the National Academy of Sciences of the United States of America* 1981, 78(7):4170-4174.
19. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J *et al*: New world bats harbor diverse influenza A viruses. *PLoS pathogens* 2013, 9(10):e1003657.
20. Nobusawa E, Sato K: Comparison of the mutation rates of human influenza A and B viruses. *Journal of virology* 2006, 80(7):3675-3678.
21. Das SR, Hensley SE, Ince WL, Brooke CB, Subba A, Delboy MG, Russ G, Gibbs JS, Bennink JR, Yewdell JW: Defining influenza A virus hemagglutinin antigenic drift by sequential monoclonal antibody selection. *Cell host & microbe* 2013, 13(3):314-323.
22. Li C, Hatta M, Burke DF, Ping J, Zhang Y, Ozawa M, Taft AS, Das SC, Hanson AP, Song J *et al*: Selection of antigenically advanced variants of seasonal influenza viruses. *Nature Microbiology* 2016, 1:16058.
23. El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.
24. Song JM, Wang BZ, Park KM, Van Rooijen N, Quan FS, Kim MC, Jin HT, Pekosz A, Compans RW, Kang SM: Influenza virus-like particles containing M2 induce broadly cross protective immunity. *PloS one* 2011, 6(1):e14538.

25.   Jegerlehner A, Schmitz N, Storni T, Bachmann MF: Influenza A vaccine based on the extracellular domain of M2: weak protection mediated via antibody-dependent NK cell activity. *J Immunol* 2004, 172(9):5598-5605.

26.   Zebedee SL, Lamb RA: Growth restriction of influenza A virus by M2 protein antibody is genetically linked to the M1 protein. *Proceedings of the National Academy of Sciences of the United States of America* 1989, 86(3):1061-1065.

27.   Zharikova D, Mozdzanowska K, Feng J, Zhang M, Gerhard W: Influenza type A virus escape mutants emerge in vivo in the presence of antibodies to the ectodomain of matrix protein 2. *Journal of virology* 2005, 79(11):6644-6654.

28.   Cho KJ, Schepens B, Seok JH, Kim S, Roose K, Lee JH, Gallardo R, Van Hamme E, Schymkowitz J, Rousseau F *et al*: Structure of the extracellular domain of matrix protein 2 of influenza A virus in complex with a protective monoclonal antibody. *Journal of virology* 2015, 89(7):3700-3711.

29.   Cho KJ, Schepens B, Moonens K, Deng L, Fiers W, Remaut H, Saelens X: Crystal Structure of the Conserved Amino Terminus of the Extracellular Domain of Matrix Protein 2 of Influenza A Virus Gripped by an Antibody. *Journal of virology* 2016, 90(1):611-615.

30.   Grandea AG, 3rd, Olsen OA, Cox TC, Renshaw M, Hammond PW, Chan-Hui PY, Mitcham JL, Cieplak W, Stewart SM, Grantham ML *et al*: Human antibodies reveal a protective epitope that is highly conserved among human and nonhuman influenza A viruses. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(28):12658-12663.

31.   Lee YN, Lee YT, Kim MC, Hwang HS, Lee JS, Kim KH, Kang SM: Fc receptor is not required for inducing antibodies but plays a critical role in conferring protection after influenza M2 vaccination. *Immunology* 2014, 143(2):300-309.

32.   Guilliams M, Bruhns P, Saeys Y, Hammad H, Lambrecht BN: The function of Fcgamma receptors in dendritic cells and macrophages. *Nature reviews Immunology* 2014, 14(2):94-108.

33.   Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X: Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC genomics* 2015, 16:79.

34.   Van den Hoecke S, Ehrhardt K, Kolpe A, El Bakkouri K, Deng L, Grootaert H, Schoonooghe S, Smet A, Bentahir M, Roose K *et al*: Hierarchical and redundant roles of activating FcγRs in protection against influenza disease by M2e-specific IgG1 and IgG2a antibodies *Under revision for Journal of Virology* 2016.

35.   Vieira P, Rajewsky K: The half-lives of serum immunoglobulins in adult mice. *European journal of immunology* 1988, 18(2):313-316.

36.   Pfeiffer JK, Kirkegaard K: Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS pathogens* 2005, 1(2):e11.

37.   Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R: Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006, 439(7074):344-348.

38.   Ping J, Keleta L, Forbes NE, Dankar S, Stecho W, Tyler S, Zhou Y, Babiuk L, Weingartl H, Halpin RA *et al*: Genomic and protein structural maps of adaptive evolution of human influenza A virus to increased virulence in the mouse. *PloS one* 2011, 6(6):e21740.

39.   Sriwilaijaroen N, Suzuki Y: Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proceedings of the Japan Academy Series B, Physical and biological sciences* 2012, 88(6):226-249.

40.   Hoffmann E, Krauss S, Perez D, Webby R, Webster RG: Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine* 2002, 20(25-26):3165-3170.

41.   Mozdzanowska K, Feng J, Eid M, Kragol G, Cudic M, Otvos L, Jr., Gerhard W: Induction of influenza type A virus-specific resistance by immunization of mice with a synthetic multiple antigenic peptide vaccine that contains ectodomains of matrix protein 2. *Vaccine* 2003, 21(19-20):2616-2626.

42.   Zhang M, Zharikova D, Mozdzanowska K, Otvos L, Gerhard W: Fine specificity and sequence of antibodies directed against the ectodomain of matrix protein 2 of influenza A virus. *Molecular immunology* 2006, 43(14):2195-2206.

43.   Wolf AI, Mozdzanowska K, Williams KL, Singer D, Richter M, Hoffmann R, Caton AJ, Otvos L, Erikson J: Vaccination with M2e-based multiple antigenic peptides: characterization of the B cell response and protection efficacy in inbred and outbred mice. *PloS one* 2011, 6(12):e28445.

44.   Gerhard W, Mozdzanowska K, Zharikova D: Prospects for universal influenza virus vaccine. *Emerging infectious diseases* 2006, 12(4):569-574.

45. Llompart CM, Nieto A, Rodriguez-Frandsen A: Specific residues of PB2 and PA influenza virus polymerase subunits confer the ability for RNA polymerase II degradation and virus pathogenicity in mice. *Journal of virology* 2014, 88(6):3455-3463.

46. Rolling T, Koerner I, Zimmermann P, Holz K, Haller O, Staeheli P, Kochs G: Adaptive mutations resulting in enhanced polymerase activity contribute to high virulence of influenza A virus in mice. *Journal of virology* 2009, 83(13):6673-6680.

47. Lycett SJ, Ward MJ, Lewis FI, Poon AF, Kosakovsky Pond SL, Brown AJ: Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data. *Journal of virology* 2009, 83(19):9901-9910.

48. Grimm D, Staeheli P, Hufbauer M, Koerner I, Martinez-Sobrido L, Solorzano A, Garcia-Sastre A, Haller O, Kochs G: Replication fitness determines high virulence of influenza A virus in mice carrying functional Mx1 resistance gene. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104(16):6806-6811.

49. Chin AW, Li OT, Mok CK, Ng MK, Peiris M, Poon LL: Influenza A viruses with different amino acid residues at PB2-627 display distinct replication properties in vitro and in vivo: revealing the sequence plasticity of PB2-627 position. *Virology* 2014, 468-470:545-555.

50. Chen H, Bright RA, Subbarao K, Smith C, Cox NJ, Katz JM, Matsuoka Y: Polygenic virulence factors involved in pathogenesis of 1997 Hong Kong H5N1 influenza viruses in mice. *Virus research* 2007, 128(1-2):159-163.

51. Katz JM, Lu X, Tumpey TM, Smith CB, Shaw MW, Subbarao K: Molecular correlates of influenza A H5N1 virus pathogenesis in mice. *Journal of virology* 2000, 74(22):10807-10810.

52. Lee MS, Deng MC, Lin YJ, Chang CY, Shieh HK, Shiau JZ, Huang CC: Characterization of an H5N1 avian influenza virus from Taiwan. *Veterinary microbiology* 2007, 124(3-4):193-201.

53. Subbarao K, Shaw MW: Molecular aspects of avian influenza (H5N1) viruses isolated from humans. *Reviews in medical virology* 2000, 10(5):337-348.

54. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A *et al*: The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 2004, 303(5665):1838-1842.

55. Ibricevic A, Pekosz A, Walter MJ, Newby C, Battaile JT, Brown EG, Holtzman MJ, Brody SL: Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells. *Journal of virology* 2006, 80(15):7469-7480.

56. Glaser L, Stevens J, Zamarin D, Wilson IA, Garcia-Sastre A, Tumpey TM, Basler CF, Taubenberger JK, Palese P: A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *Journal of virology* 2005, 79(17):11533-11536.

57. Zheng B, Chan KH, Zhang AJ, Zhou J, Chan CC, Poon VK, Zhang K, Leung VH, Jin DY, Woo PC *et al*: D225G mutation in hemagglutinin of pandemic influenza H1N1 (2009) virus enhances virulence in mice. *Exp Biol Med (Maywood)* 2010, 235(8):981-988.

58. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR *et al*: Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* 2009, 326(5953):734-736.

59. De Filette M, Martens W, Roose K, Deroo T, Vervalle F, Bentahir M, Vandekerckhove J, Fiers W, Saelens X: An influenza A vaccine based on tetrameric ectodomain of matrix protein 2. *The Journal of biological chemistry* 2008, 283(17):11382-11387.

60. Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P: Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci* 2013, 368(1614):20120205.

61. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE: Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *Journal of virology* 2009, 83(19):10309-10313.

62. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR: Universal primer set for the full-length amplification of all influenza A viruses. *Arch Virol* 2001, 146(12):2275-2289.

63. Yang X, Charlebois P, Gnerre S, Coole MG, Lennon NJ, Levin JZ, Qu J, Ryan EM, Zody MC, Henn MR: De novo assembly of highly diverse viral populations. *BMC genomics* 2012, 13:475.

64. Lee WP, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT: MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PloS one* 2014, 9(3):e90581.

65.   Larsson A: AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014, 30(22):3276-3278.

66.   Archer J, Rambaut A, Robertson D: Segminator: the analysis of ultra-deep next generation sequence data. *Submitted* 2010.

67.   Nelson CW, Hughes AL: Within-host nucleotide diversity of virus populations: insights from next-generation sequencing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2015, 30:1-7.

68.   Team RC: R: A Language and Environment for Statistical Computing. In. Vienna, Austria; 2012.

69.   Heinze G, Ploner M, Dunkler D, Southworth H: logistf: Firth's bias reduced logistic regression. In., R package version 1.21 edn; 2013.

Supplementary figures
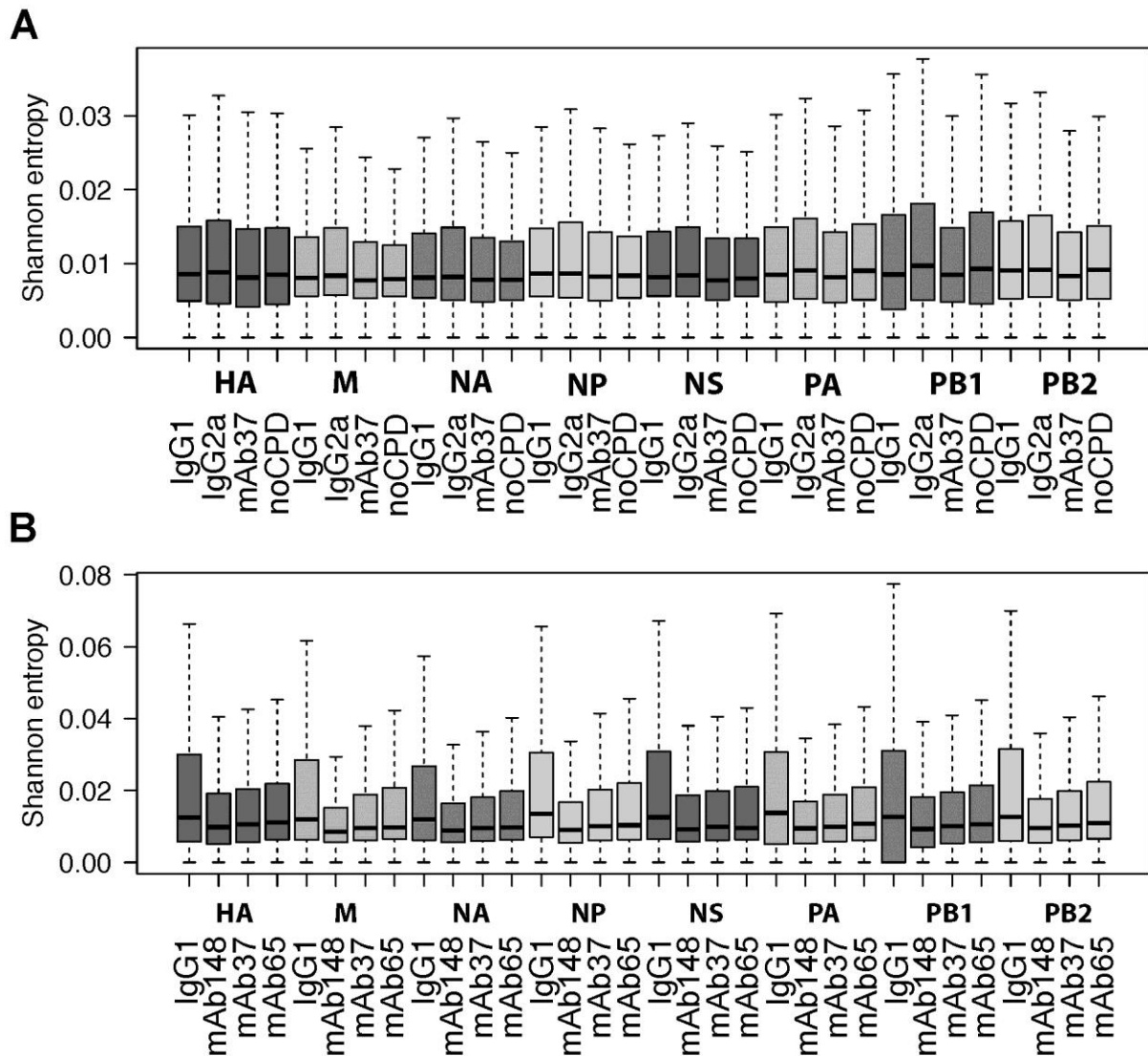
Supplementary Figure S1: Influenza A virus populations isolated from infected SCID mice treated with anti-M2e mAbs or isotype controls have a similar sequence diversity. The standard Shannon entropy was calculated per segment and treatment (IgG1 or IgG2a isotype controls, anti-M2e mAbs (mAb37, mAb65 or mAb148) or no treatment (noCPD: no compound)) after mapping the reads to the sample-specific majority rule consensus sequence [67]. Panel A and B represent data obtained from independently performed mice experiments.
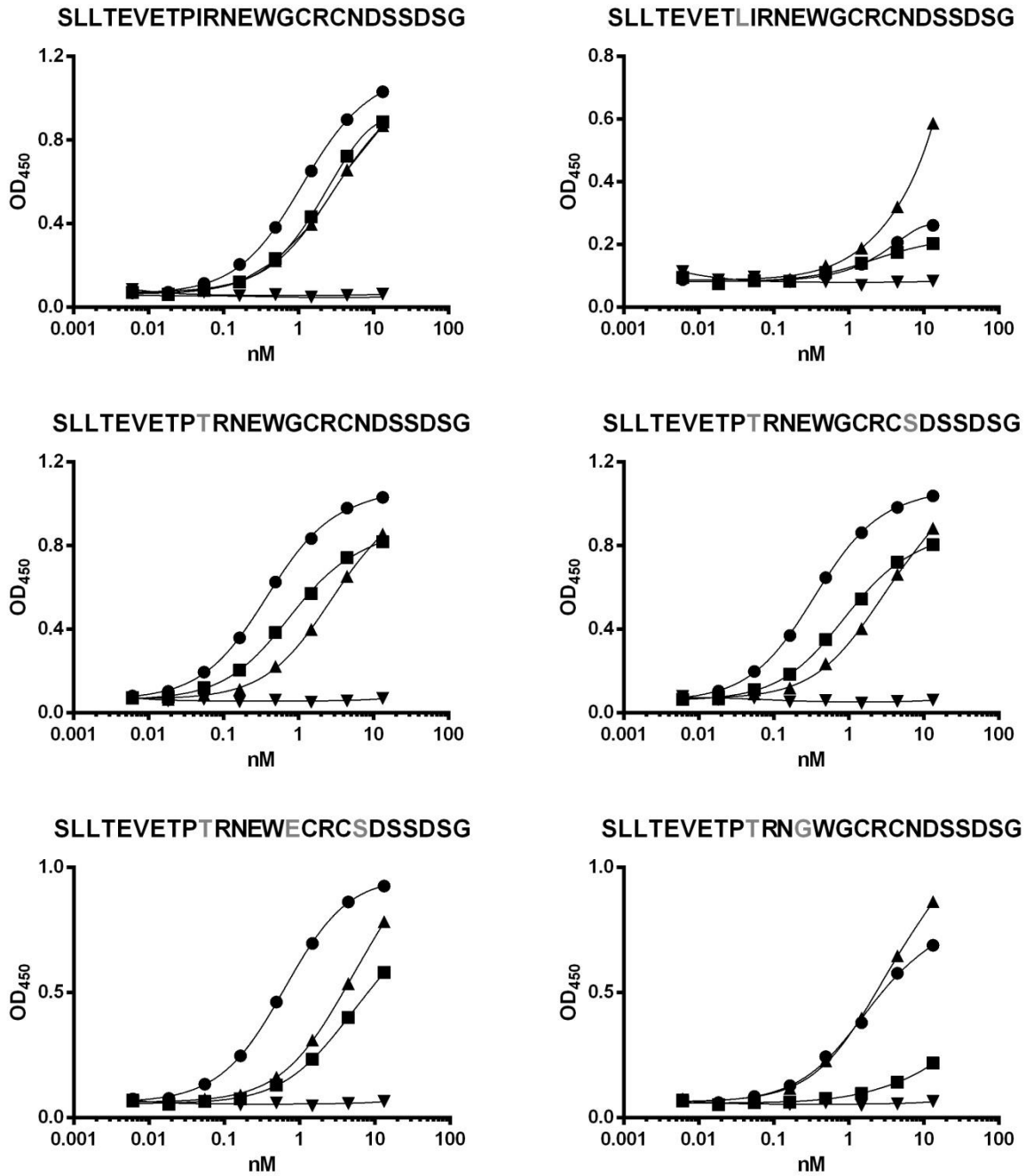
Supplementary Figure S2: mAb65 and mAb37 bind a different epitope in M2e. Human consensus M2e (SLLTEVETPIRNEWGCRCNDSSD) and variants hereof (variation marked in grey) were used for coating in M2e peptide ELISA. Binding of mAb37, mAb65 and mAb148 was determined using horseradish peroxidase-conjugated sheep anti-mouse IgG as secondary Ab.

Supplementary Figure S2 (continued): mAb65 and mAb37 bind a different epitope in M2e. Human consensus M2e (SLLTEVETPIRNEWGCRCNDSSD) and variants hereof (variation marked in grey) were used for coating in M2e peptide ELISA. Binding of mAb37, mAb65 and mAb148 was determined using horseradish peroxidase-conjugated sheep anti-mouse IgG as secondary Ab.

Supplementary Figure S2 (continued): mAb65 and mAb37 bind a different epitope in M2e. Human consensus M2e (SLLTEVETPIRNEWGCRCNDSSD) and variants hereof (variation marked in grey) were used for coating in M2e peptide ELISA. Binding of mAb37, mAb65 and mAb148 was determined using horseradish peroxidase-conjugated sheep anti-mouse IgG as secondary Ab.

mAb37 staining			
PR8-HAmut	**PR8-PB2 R443K**	**PR8-PA I550T**	**PR8-PA I550T HAmut**

polyclonal anti-RNP staining			
PR8-HAmut	**PR8-PB2 R443K**	**PR8-PA I550T**	**PR-PA I550T HAmut**

Supplementary Figure S3: Plaque staining of mutant PR8 viruses. Confluent layers of MDCK cells were infected in parallel with mutant PR8 viruses. Two days later, the cells were fixed with 4% paraformaldehyde and stained with M2e-specific mAb (mAb37, upper panel) or polyclonal anti-vRNP (lower panel). Plaques were revealed using horseradish peroxidase-conjugated sheep anti-mouse IgGs, respectively donkey anti-goat IgGs, followed by visualization with TrueBlue Peroxidase substrate (KPL).
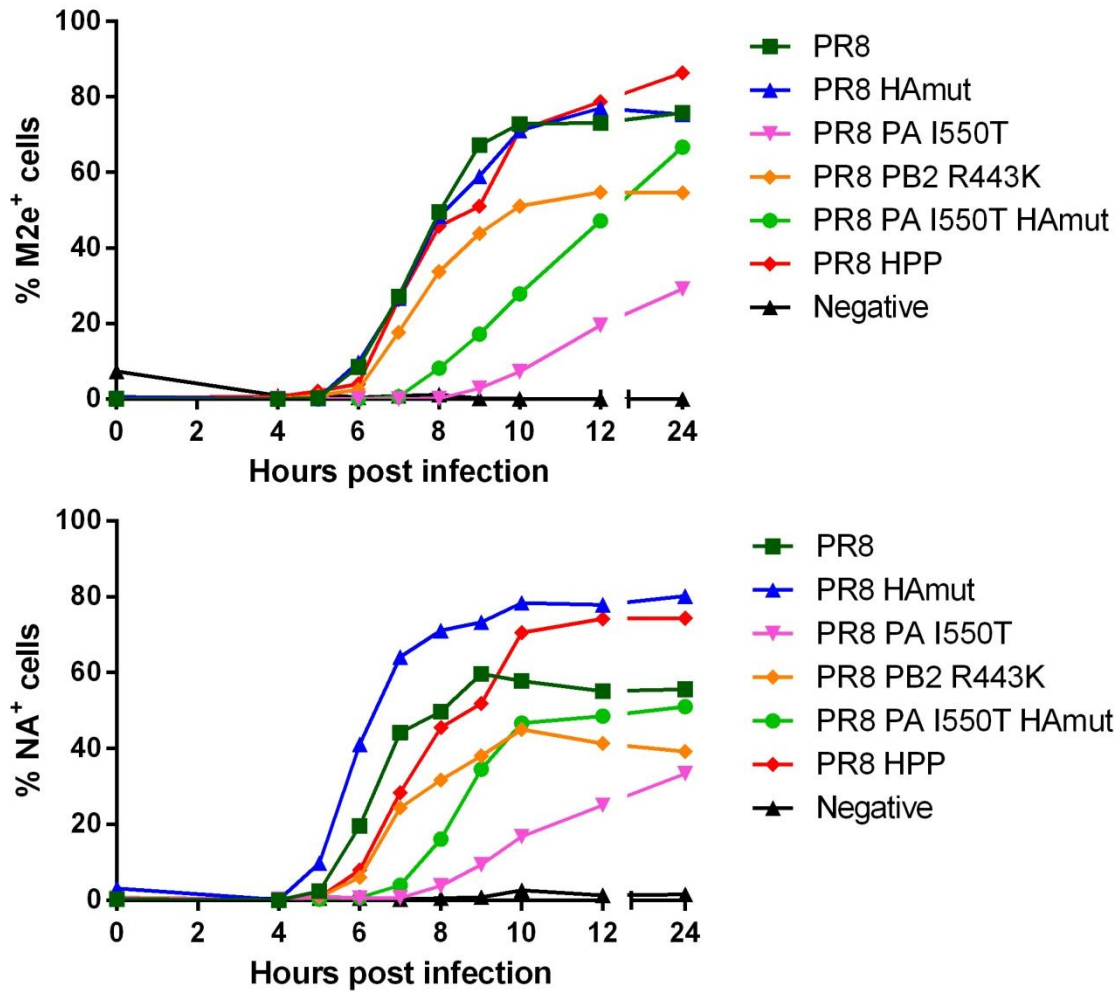
Supplementary Figure S4: Kinetics of M2e and NA expression on the surface of infected cells. HEK cells were infected with a moi of 0.5 and fixed with 2% paraformaldehyde at the indicated time points. The cells were either stained with mAb37 (M2e-specific mAb, upper panel) or 7D3 (NA-specific mAb, lower panel), followed by a secondary anti-mouse antibody coupled to Alexa Fluor 488 and analysis on an LSRII HTS (BD) flow cytometer.

Supplementary Table S1: minimum, maximum and average sequencing coverage depth of PR8 virus genome after influenza-specific RT-PCR on RNA isolated from BAL fluid of mice infected with 10 PFU of PR8

1ˢᵗ mice experiment

Treatment	Mouse	Dpi	Coverage			SD (average)
			Minimum	Maximum	Average	
Untreated	1	11	1619	57657	13005.40	12709.66
Untreated	2	11	2710	55245	13884.97	11506.39
Untreated	3	11	2222	64189	15135.92	14878.89
Untreated	4	11	2836	54232	13955.97	11021.56
IgG1 control	1	11	1359	88877	16647.49	16018.01
IgG1 control	2	15	1873	64429	15462.90	12603.65
IgG1 control	3	19	2056	59312	14943.63	11954.40
IgG2a control	1	12	2663	54214	14101.85	10369.01
IgG2a control	2	21	4781	51052	15286.96	9641.71
IgG2a control	3	12	4707	54629	15250.58	10161.71
mAb37	1	7	3423	63688	14835.94	11152.37
mAb37	2	7	3288	63826	14894.45	11863.82
mAb37	3	7	1795	81620	14582.97	13612.95
mAb37	1	14	5042	50911	14898.04	8684.95
mAb37	1	33	3801	51537	13466.80	11267.94
mAb37	2	23	5011	46412	15005.92	8403.85
mAb37	3	30	6450	45178	14918.95	8500.65
mAb37	4	29	5183	38011	14585.29	6299.66

2ⁿᵈ mice experiment

Treatment	Mouse	Dpi	Coverage			SD (average)
			Minimum	Maximum	Average	
IgG1 control	1	12	2782	38222	11108.53	8591.69
IgG1 control	2	10	287	4125	1369.99	838.77
IgG1 control	3	13	2255	35760	11113.39	8062.81
mAb37	1	13	4065	27855	10199.13	5189.11
mAb37	2	13	2749	33030	10877.47	5856.87
mAb37	3	13	2722	35154	10890.29	6177.49
mAb148	1	13	3453	38264	11670.28	6955.30
mAb148	2	13	3644	34769	10252.30	5950.33
mAb148	3	13	2229	32028	10294.68	5757.00
mAb37	1	23	3656	30083	12114.02	5725.89
mAb37	2	39	2880	28117	11167.71	5770.04
mAb37	3	20	3365	23340	8569.41	4455.49
mAb37	4	29	2804	28465	10403.31	5662.53
mAb37	5	23	4404	30516	11989.59	5585.34
mAb37	6	20	3672	27747	11047.84	5350.47
mAb148	1	27	3248	27188	10132.95	5965.50
mAb148	2	33	4499	31041	11259.96	6287.84
mAb148	3	28	3838	34936	13185.26	6993.80
mAb148	4	30	4723	36606	13174.48	7726.44
mAb148	5	33	2877	27126	10216.37	5347.08
mAb148	6	28	3000	26067	9699.04	5555.26
mAb65 (50 PFU)	1	37	2754	61111	16909.90	13575.29
mAb65 (50 PFU)	2	28	3162	27188	9459.08	5384.61
mAb65 (100 PFU)	1	38	4504	34914	11287.38	6016.42
mAb65 (100 PFU)	2	32	5149	49035	16497.54	9415.52

SD = standard deviation

220

Supplementary Table S2: Variants detected above 10% in BAL fluid from untreated, PR8-infected SCID mice, isolated at 11 dpi (First experiment)

Mouse 1

Segment	Position	Frequency	Amino acid change
PB2	181	12.85	PB2:p.Leu45Pro
PB2	184	15.61	PB2:p.Arg46fs
PB2	340-354	10.22	PB2:p.Trp98_Gly103delins*
PB1	74-87	26.05	PB1:p.Leu10fs
HA	648	11.66	HA:p.Pro199His
HA	749	13.79	HA:p.Arg233Gly
HA	765	24.44	HA:p.Asp238Gly
HA	1424	97.55	HA:p.Val458Met
NP	1249	32.99	NP:p.Asn395Ser

Mouse 2

Segment	Position	Frequency	Amino acid change
PB2	50	14.24	PB2:p.Met1?
HA	217	12.43	
HA	495	13.9	HA:p.Val148Ala
HA	765	12.16	HA:p.Asp238Gly
HA	770	18.02	HA:p.Ala240Thr
HA	847	12.83	HA:p.Ile265Met
HA	1424	86.94	HA:p.Val458Met
M	149	10.86	M1:p.[Lys35Arg]

Mouse 3

Segment	Position	Frequency	Amino acid change
PB1	95-96	12.09	PB1:p.Ala17fs
HA	647	23.23	HA:p.Pro199Phe
HA	765	14.29	HA:p.Asp238Gly
HA	1424	99.94	HA:p.Val458Met
M	453	27.32	

Mouse 4

Segment	Position	Frequency	Amino acid change
PB2	2036	11.95	
PA	1265	10.09	
HA	497	11.06	HA:p.Thr149Ala
HA	749	14.38	HA:p.Arg233Gly
HA	765	21.37	HA:p.Asp238Gly
HA	1424	99.88	HA:p.Val458Met
NA	1240	12.89	

Supplementary table S3: Variants detected above 10% in BAL fluid from IgG1 control-treated mice infected with PR8, when mice lost 25% of their initial body weight (First experiment).

Mouse 1	11 dpi		
Segment	Position	Frequency	Amino acid change
PB1	675	11.99	PB1:p.[Arg211Gly]; PB1-N40:p.[Arg172Gly]
HA	493	18.87	
HA	765	26.48	HA:p.Asp238Gly
HA	770	21.87	HA:p.Ala240Thr
HA	847	10.54	HA:p.Ile265Met
HA	1424	75.32	HA:p.Val458Met
NP	1161	12.62	NP:p.Ala366Ser
NP	1249	16.42	NP:p.Asn395Ser
M	657	15.95	

Mouse 2	15 dpi		
Segment	Position	Frequency	Amino acid change
PB2	2240	10.44	
PB1	675	13.51	PB1:p.[Arg211Gly]; PB1-N40:p.[Arg172Gly]
HA	765	45.79	HA:p.Asp238Gly
HA	823	11.68	HA:p.Ile257Met
HA	1424	90.79	HA:p.Val458Met
NP	1249	15.08	NP:p.Asn395Ser

Mouse 3	19 dpi		
Segment	Position	Frequency	Amino acid change
PB2	716	18.4	
PB2	972	10.07	PB2:p.Asp309Asn
PB1	675	13.58	PB1:p.[Arg211Gly]; PB1-N40:p.[Arg172Gly]
PA	188	49.4	
PA	2035	49.46	PA:p.[Lys664Arg]; PA-N155:p.[Lys510Arg]; PA-N182:p.[Lys483Arg]
HA	497	17.12	HA:p.Thr149Ala
HA	689	17.85	HA:p.Ala213Thr
HA	765	32.95	HA:p.Asp238Gly
HA	1424	99.94	HA:p.Val458Met
NP	434	45.62	

Supplementary table S4: Variants detected above 10% in BAL fluid isolated from IgG2a control-treated mice infected with PR8, when mice lost 25% of their initial body weight (First experiment).

Mouse 1	12 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1355	16.14	
PB1	1187	25.89	
PB1	1331	12.2	
HA	442	34.4	
HA	749	10.64	HA:p.Arg233Gly
HA	765	12.15	HA:p.Asp238Gly
HA	847	36.14	HA:p.Ile265Met
HA	1424	98.44	HA:p.Val458Met
HA	1474	33.64	
NA	1251	11.29	NA:p.Arg404Lys
M	603	28.4	

Supplementary table S4 (continued): Variants detected above 10% in BAL fluid isolated from IgG2a control-treated mice infected with PR8, when mice lost 25% of their initial body weight (First experiment).

Mouse 2	21 dpi		
Segment	Position	Frequency	Amino acid change
PB2	416	18.01	PB2:p.Glu123Asp
PB2	1557	17.21	PB2:p.Ile504Val
PA	102	14.61	PA-X:p.[Thr20Ala]; PA:p.[Thr20Ala]
PA	104	11.41	
HA	765	29.54	HA:p.Asp238Gly
HA	1424	99.88	HA:p.Val458Met
NP	1208	12.46	

Mouse 3	12 dpi		
Segment	Position	Frequency	Amino acid change
HA	765	22.62	HA:p.Asp238Gly
HA	1424	94.28	HA:p.Val458Met

Supplementary table S5: Variants detected above 10% in BAL fluid from mAb37-treated mice infected with PR8, isolated at 7 dpi (First experiment).

Mouse 1			
Segment	Position	Frequency	Amino acid change
HA	1424	100	HA:p.Val458Met
NP	1249	37.76	NP:p.Asn395Ser

Mouse 2			
Segment	Position	Frequency	Amino acid change
PB1	159-167	12.63	PB1:p.[Thr39_Asp41del]; PB1-F2:p.[Thr7_Ile10delinsThr]; PB1-N40:p.[Met1?]
PB1	164	59.42	PB1:p.[Met40fs]; PB1-F2:p.[Trp9fs]; PB1-N40:p.[Met1?]
PA	136-139	12.19	PA-X:p.[Glu31fs]; PA:p.[Glu31fs]
PA	314-320	24.28	PA-X:p.[Val90fs]; PA:p.[Val90fs]
HA	1424	100	HA:p.Val458Met
NP	1249	73.26	NP:p.Asn395Ser

Mouse 3			
Segment	Position	Frequency	Amino acid change
HA	1424	100	HA:p.Val458Met
HA	1426	99.72	
NP	1249	99.93	NP:p.Asn395Ser

Supplementary table S6: Variants detected above 10% in BAL fluid isolated from mAb37-treated mice infected with PR8, isolated at 14 dpi (First experiment).

Mouse 1			
Segment	Position	Frequency	Amino acid change
PB2	2218-2220	13.98	PB2:p.Val724_Leu725delinsVal
PA	166-175	14.55	PA-X:p.[His41fs]; PA:p.[His41fs]
PA	398	25.34	
HA	856	20.82	HA:p.Met268Ile
HA	1424	99.93	HA:p.Val458Met

Supplementary table S7: Variants detected above 10% in BAL fluid isolated from mAb37-treated mice infected with PR8, when mice lost 25% of their initial body weight (First experiment).

Mouse 1 **33 dpi**

Segment	Position	Frequency	Amino acid change
PB2	1375	62.7	PB2:p.Lys443Arg
PB1	737	13.03	
PA	1693	98.09	PA:p.[Ile550Thr]; PA-N155:p.[Ile396Thr]; PA-N182:p.[Ile369Thr]
HA	557	12.75	HA:p.Glu169Lys
HA	679	99.8	
HA	743	96.32	HA:p.Ala231Ser
HA	748	34.89	HA:p.Glu232Asp
HA	1135	96.15	HA:p.Ile361Met
HA	1424	99.6	HA:p.Val458Met
NP	267	22.87	NP:p.Leu68Ile

Mouse 2 **23 dpi**

Segment	Position	Frequency	Amino acid change
PB2	1557	94.76	PB2:p.Ile504Val
PB1	1050	78.6	PB1:p.[Val336Ile]; PB1-N40:p.[Val297Ile]
PA	2006	10.64	
HA	191	18.29	HA:p.Val47Ile
HA	823	44.45	HA:p.Ile257Met
HA	1424	99.93	HA:p.Val458Met
M	765	90.26	M2:p.[Ile11Thr]

Mouse 3 **30 dpi**

Segment	Position	Frequency	Amino acid change
PB2	796	12.21	PB2:p.Val250Ala
PB2	1557	67.64	PB2:p.Ile504Val
PB2	2102	13.29	
PB1	686	18.26	
PB1	1457	17.52	
HA	1244	11.82	HA:p.Val398Ile
HA	1424	99.81	HA:p.Val458Met
HA	1486	38.83	
M	762	60.95	M2:p.[Pro10His]

Mouse 4 **29 dpi**

Segment	Position	Frequency	Amino acid change
PB2	1525	88.17	PB2:p.Arg493Lys
PB2	1557	12.88	PB2:p.Ile504Val
PB1	925	12.03	PB1:p.[Gln294Arg]; PB1-N40:p.[Gln255Arg]
HA	738	10.49	HA:p.Glu229Val
HA	749	21.77	HA:p.Arg233Gly
HA	765	36.33	HA:p.Asp238Gly
HA	1039	25.66	
HA	1424	99.72	HA:p.Val458Met
HA	1426	11.09	

224

Supplementary table S8: Variants detected above 10% in BAL fluid isolated from IgG1 control-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 1   12 dpi

Segment	Position	Frequency	Amino acid change
PB2	2009	11.8	
HA	1424	88.5	HA:p.Val458Met
NP	1187	19.98	NP:p.Met374Ile

Mouse 2   10 dpi

Segment	Position	Frequency	Amino acid change
PB2	1000	10.6	PB2:p.Arg318Lys
PA	104-106	13.31	PA-X:p.[Thr20_Met21delinsThr]; PA:p.[Thr20_Met21delinsThr]
HA	500	10.43	HA:p.Ala150Thr
HA	765	14.49	HA:p.Asp238Gly
HA	770	10.61	HA:p.Ala240Thr
HA	1424	92.43	HA:p.Val458Met
HA	1528	25.22	HA:p.Met492Ile

Mouse 3   13 dpi

Segment	Position	Frequency	Amino acid change
PB2	561	39.56	PB2:p.Val172Met
PB2	1169	23.33	
PB2	2158-2167	24.37	PB2:p.Tyr704fs
PB1	79-91	10.99	PB1:p.Val12fs
HA	672	27.73	HA:p.Leu207His
HA	765	49.35	HA:p.Asp238Gly
HA	1424	99.98	HA:p.Val458Met
NP	1249	13.44	NP:p.Asn395Ser

Supplementary Table S9: Variants detected above 10% in BAL fluid from mAb37-treated mice infected with PR8, isolated at 13 dpi (Second experiment).

Mouse 1

Segment	Position	Frequency	Amino acid change
PB2	165-182	13.72	PB2:p.Glu40_Leu45del
PB2	238	12.97	PB2:p.Thr64fs
PB2	1557	72.44	PB2:p.Ile504Val
PB1	1364	18.6	
HA	933	10.75	HA:p.Lys294Arg
HA	1424	99.93	HA:p.Val458Met
HA	1426	11.31	
HA	1619	99.88	HA:p.Ser523Pro

Mouse 2

Segment	Position	Frequency	Amino acid change
PB1	794	35.47	
HA	611	20.64	HA:p.Lys187Glu
HA	1424	99.73	HA:p.Val458Met
NP	1324	24.49	NP:p.Phe420Cys
M	1024	10.87	

Supplementary Table S9 (continued): Variants detected above 10% in BAL fluid from mAb37-treated mice infected with PR8, isolated at 13 dpi (Second experiment).

Mouse 3			
Segment	Position	Frequency	Amino acid change
PB2	1111	22.57	PB2:p.Arg355Lys
PB1	1650	99.87	PB1:p.[Asn536Asp]; PB1-N40:p.[Asn497Asp]
HA	740	20.69	HA:p.Ile230Val
HA	1424	99.91	HA:p.Val458Met

Supplementary table S10: Variants detected above 10% in BAL fluid from mAb148-treated mice infected with PR8, isolated at 13 dpi (Second experiment).

Mouse 1			
Segment	Position	Frequency	Amino acid change
PB2	1525	23.08	PB2:p.Arg493Lys
HA	689	66.48	HA:p.Ala213Thr
HA	1424	99.91	HA:p.Val458Met
NP	972	95.2	NP:p.Pro303Ser
NA	1180	21.04	

Mouse 2			
Segment	Position	Frequency	Amino acid change
HA	1424	100	HA:p.Val458Met
NP	284	15.79	
NS	278	10.75	NS1:p.Lys78Glu

Mouse 3			
Segment	Position	Frequency	Amino acid change
PB2	972	14.51	PB2:p.Asp309Asn
PB2	1000	46.72	PB2:p.Arg318Lys
PB2	1110	10.69	PB2:p.Arg355Gly
HA	765	25.59	HA:p.Asp238Gly
HA	790	10.85	
HA	1424	100	HA:p.Val458Met
NA	916	11.23	

Supplementary table S11: Variants detected above 10% in BAL fluid isolated from mAb37-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 1	23 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1557	15.7	PB2:p.Ile504Val
PB2	2265	72.61	PB2:p.Asp740Asn
HA	765	99.96	HA:p.Asp238Gly
HA	1424	99.91	HA:p.Val458Met
NP	212	96.29	

Supplementary table S11 (continued): Variants detected above 10% in BAL fluid isolated from mAb37-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 2	39 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1557	89.1	PB2:p.Ile504Val
PA	413	70.91	
PA	423	65.84	PA-X:p.[Val127Ile]; PA:p.[Val127Ile]
PA	1258	64.84	PA:p.[Ser405Asn]; PA-N155:p.[Ser251Asn]; PA-N182:p.[Ser224Asn]
HA	427	85.24	
HA	660	87.75	HA:p.Glu203Gly
HA	747	10.4	HA:p.Glu232Gly
HA	1424	98.25	HA:p.Val458Met
NA	415	90.07	
NA	492	27.7	NA:p.Ala151Val
M	765	12.86	M2:p.[Ile11Thr]

Mouse 3	20 dpi		
Segment	Position	Frequency	Amino acid change
PB2	992	51.5	PB2:p.Met315Ile
PB1	490	11.15	PB1:p.[Val149Ala]; PB1-N40:p.[Val110Ala]
PA	51	33.07	PA-X:p.[Asp3Asn]; PA:p.[Asp3Asn]
PA	1421	42.22	PA:p.[Ile459Met]; PA-N155:p.[Ile305Met]; PA-N182:p.[Ile278Met]
HA	765	70.77	HA:p.Asp238Gly
HA	1424	100	HA:p.Val458Met
NP	925	12.71	NP:p.Ser287Asn

Mouse 4	29 dpi		
Segment	Position	Frequency	Amino acid change
PB2	864	23.52	PB2:p.Ser273Thr
PB2	1132	22.55	PB2:p.Glu362Gly
PB2	2120	62.96	
PB2	2216-2217	31.18	PB2:p.Asn723fs
PB1	1434	11.42	PB1:p.[Asp464Asn]; PB1-N40:p.[Asp425Asn]
PB1	2214	23.74	PB1:p.[Ile724Val]; PB1-N40:p.[Ile685Val]
PA	753	47.09	PA-X:p.[Glu237Lys]; PA:p.[Glu237Lys]; PA-N155:p.[Glu83Lys]; PA-N182:p.[Glu56Lys]
HA	359	13.66	HA:p.Glu103Lys
HA	1217	24.86	HA:p.Asn389Asp
HA	1409	26.78	HA:p.Phe453Leu
HA	1424	72.17	HA:p.Val458Met
NA	820	26.12	
M	155	42.29	M1:p.[Thr37Ile]
M	762	15.25	M2:p.[Pro10His]
NS	592	15.31	NS2:p.[Ser25Leu]

Supplementary table S11 (continued): Variants detected above 10% in BAL fluid isolated from mAb37-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 5	23 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1091	13.68	
PB2	1557	33.28	PB2:p.Ile504Val
PB2	1570	12.85	PB2:p.Arg508Gln
PB1	67	11.77	PB1:p.Leu8Pro
PB1	1244	10.55	
PA	1887	17.54	PA:p.[Lys615Glu]; PA-N155:p.[Lys461Glu]; PA-N182:p.[Lys434Glu]
HA	748	99.96	HA:p.Glu232Asp
HA	1424	99.95	HA:p.Val458Met
NP	110	13.76	

Mouse 6	20 dpi		
Segment	Position	Frequency	Amino acid change
PB1	2214	15.49	PB1:p.[Ile724Val]; PB1-N40:p.[Ile685Val]
PA	1292	10.41	PA:p.[Glu416Asp]; PA-N155:p.[Glu262Asp]; PA-N182:p.[Glu235Asp]
PA	1661	11.37	
PA	1692	19.66	PA:p.[Ile550Val]; PA-N155:p.[Ile396Val]; PA-N182:p.[Ile369Val]
PA	1694	11.45	PA:p.[Ile550Met]; PA-N155:p.[Ile396Met]; PA-N182:p.[Ile369Met]
HA	678	89.88	HA:p.Gln209Arg
HA	1424	99.79	HA:p.Val458Met

Supplementary table S12: Variants detected above 10% in BAL fluid isolated from mAb148-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 1	27 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1557	98.42	PB2:p.Ile504Val
PA	1534	19.07	PA:p.[Lys497Arg]; PA-N155:p.[Lys343Arg]; PA-N182:p.[Lys316Arg]
HA	765	84.22	HA:p.Asp238Gly
HA	1135	24.05	HA:p.Ile361Met
HA	1217	23.22	HA:p.Asn389Asp
HA	1424	99.91	HA:p.Val458Met
NP	1063	11.24	NP:p.Cys333Phe
NA	206	15.48	NA:p.Tyr56His
NS	657	92.63	NS1:p.[Arg204Lys]; NS2:p.[Glu47Lys]

Mouse 2	33 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1557	25.88	PB2:p.Ile504Val
PA	981	40.03	PA:p.[Thr313Ala]; PA-N155:p.[Thr159Ala]; PA-N182:p.[Thr132Ala]
PA	1784	25.55	PA:p.[Glu580Asp]; PA-N155:p.[Glu426Asp]; PA-N182:p.[Glu399Asp]
HA	647	18.29	HA:p.Pro199Ser
HA	719	18.85	HA:p.Asn223Tyr
HA	823	15.2	HA:p.Ile257Met
HA	1424	99.79	HA:p.Val458Met
NP	542	12.71	NP:p.Met159Ile
NA	201	15.62	NA:p.Ile54Asn

Supplementary table S12 (continued): Variants detected above 10% in BAL fluid isolated from mAb148-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 3	28 dpi		
Segment	Position	Frequency	Amino acid change
PB2	741	99.82	
PB2	2223	53.35	PB2:p.Ile726Val
PB1	1586	11.89	
PA	1052	10.46	
PA	1071	43.77	PA:p.[Ala343Thr]; PA-N155:p.[Ala189Thr]; PA-N182:p.[Ala162Thr]
PA	1293	12.33	PA:p.[Leu417Met]; PA-N155:p.[Leu263Met]; PA-N182:p.[Leu236Met]
PA	1711	10.34	PA:p.[Gln556Arg]; PA-N155:p.[Gln402Arg]; PA-N182:p.[Gln375Arg]
HA	775	10.43	
HA	1085	61.7	HA:p.Leu345Ile
HA	1419	38.43	HA:p.Ser456Leu
NP	212	99.97	
NS	151	12.18	

Mouse 4	30 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1111	44.08	PB2:p.Arg355Lys
PB1	2164	18.48	PB1:p.[Arg707Lys]; PB1-N40:p.[Arg668Lys]
PA	1692	28.84	PA:p.[Ile550Val]; PA-N155:p.[Ile396Val]; PA-N182:p.[Ile369Val]
PA	1887	24.51	PA:p.[Lys615Glu]; PA-N155:p.[Lys461Glu]; PA-N182:p.[Lys434Glu]
HA	611	44.59	HA:p.Lys187Glu
HA	765	85.81	HA:p.Asp238Gly
HA	1424	98.35	HA:p.Val458Met
HA	1598	37.88	HA:p.Val516Ile
NP	921	13.13	NP:p.Ala286Ser
NP	972	69.78	NP:p.Pro303Ser
NP	1420	18.4	NP:p.Arg452Ile

Mouse 5	33 dpi		
Segment	Position	Frequency	Amino acid change
PB2	782	10.39	
PB2	1184	21.59	
PB2	1557	75.24	PB2:p.Ile504Val
PA	51	49.46	PA-X:p.[Asp3Asn]; PA:p.[Asp3Asn]
PA	971	12.62	PA:p.[Lys309Asn]; PA-N155:p.[Lys155Asn]; PA-N182:p.[Lys128Asn]
HA	823	57.75	HA:p.Ile257Met
HA	856	15.1	HA:p.Met268Ile
HA	1135	10.93	HA:p.Ile361Met
HA	1415	15.23	HA:p.Asp455Tyr
HA	1417	50.34	HA:p.Asp455Glu
HA	1424	20.36	HA:p.Val458Met
HA	1537	49.52	

Supplementary table S12 (continued): Variants detected above 10% in BAL fluid isolated from mAb148-treated mice infected with PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 6	28 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1165	59.34	PB2:p.Ile373Thr
PB1	659	13.03	PB1:p.[Met205Ile]; PB1-N40:p.[Met166Ile]
PA	1887	34.76	PA:p.[Lys615Glu]; PA-N155:p.[Lys461Glu]; PA-N182:p.[Lys434Glu]
PA	1933	26.24	PA:p.[Glu630Gly]; PA-N155:p.[Glu476Gly]; PA-N182:p.[Glu449Gly]
HA	557	77.46	HA:p.Glu169Lys
HA	765	16.91	HA:p.Asp238Gly
HA	1424	99.11	HA:p.Val458Met
NA	190	21.83	NA:p.Asn50Lys
NA	670	14.03	
NS	151	28.01	

Supplementary table S13: Variants detected above 10% in BAL fluid isolated from mAb65-treated mice infected with 50 PFU of PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 1	37 dpi		
Segment	Position	Frequency	Amino acid change
PB2	972	93.76	PB2:p.Asp309Asn
PB2	1112	45.16	PB2:p.Arg355Ser
PB1	157-158	11.89	PB1:p.[Tyr38fs]; PB1-F2:p.[Thr7fs]
PA	524	32.55	
PA	1165	95.69	PA:p.[Met374Lys]; PA-N155:p.[Met220Lys]; PA-N182:p.[Met193Lys]
PA	1712	95.55	PA:p.[Gln556His]; PA-N155:p.[Gln402His]; PA-N182:p.[Gln375His]
HA	497	22.57	HA:p.Thr149Ala
HA	501	10.28	HA:p.Ala150Glu
HA	765	91.15	HA:p.Asp238Gly
HA	771	31.66	HA:p.Ala240Asp
HA	834	31.44	HA:p.Asn261Ser
HA	1424	98.03	HA:p.Val458Met
NP	64	28.89	
NA	973	11.9	
M	762	99.96	M2:p.[Pro10His]
M	1014	11.22	M2:p.[Ile94Thr]

Mouse 2	28 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1165	54.83	PB2:p.Ile373Thr
PB2	1495	18.28	PB2:p.Met483Thr
PB2	1527	53.78	PB2:p.Val494Ile
HA	823	96.38	HA:p.Ile257Met
HA	1424	98.85	HA:p.Val458Met
NP	1390	12.75	NP:p.Thr442Ile
M	762	90.99	M2:p.[Pro10His]

Supplementary table S14: Variants detected above 10% in BAL fluid isolated from mAb65-treated mice infected with 100 PFU of PR8, when mice lost 25% of their initial body weight (Second experiment).

Mouse 1	38 dpi		
Segment	Position	Frequency	Amino acid change
PB2	1243	36.9	PB2:p.Ile399Thr
PB2	1557	59.6	PB2:p.Ile504Val
PB2	1570	37.57	PB2:p.Arg508Gln
PB1	1025	15.36	
PB1	1397	30.03	
PB1	1583	15.01	
PA	1739	96.89	
HA	765	88.98	HA:p.Asp238Gly
HA	1217	85.99	HA:p.Asn389Asp
HA	1424	78.96	HA:p.Val458Met
NP	1037	29.96	

Mouse 2	32 dpi		
Segment	Position	Frequency	Amino acid change
PB2	972	13.75	PB2:p.Asp309Asn
PB2	2265	89.13	PB2:p.Asp740Asn
HA	557	72.51	HA:p.Glu169Lys
HA	563	17.87	HA:p.Glu171Lys
HA	719	17.94	HA:p.Asn223Asp
HA	1003	21.76	
HA	1424	99.96	HA:p.Val458Met
NP	1195	67.07	NP:p.Ser377Asn
M	762	96.47	M2:p.[Pro10His]
NS	681	12.46	NS1:p.[Pro212Leu]; NS2:p.[Leu55Phe]
NS	703	26.74	NS1:p.[Lys219Asn]; NS2:p.[Asn62Thr]

Part IV:
Conclusions, discussion and future perspectives

M2e as candidate for a 'universal influenza A vaccine'

The high economical and social cost associated with the annual disease burden of influenza epidemics, together with the fear for a next influenza pandemic, demonstrate the need for a so called 'universal influenza vaccine'. Such a vaccine should ideally result in effective, broad-spectrum and long-lasting protection against disease caused by influenza virus infection. Our lab, followed by several others, has demonstrated that a vaccine based on the conserved ectodomain of the influenza virus membrane protein M2 is a valuable 'universal influenza A vaccine' candidate [1, 2]. Some M2e-vaccine constructs as well as M2e-specific monoclonal IgG antibodies have been evaluated in early phase clinical trials [3-6]. To further optimize M2e-based immune protection strategies against influenza, it is important to have a detailed understanding of their mechanism of action. This is a challenging task because M2e-specific antibodies lack virus neutralizing activity, which is the conventional correlate of protection for most licensed anti-microbial vaccines. In addition, the high genetic diversity of influenza viruses - a consequence of their error-prone replication - necessitates to investigate how these viruses may evolve under M2e-based immune pressure. These two key issues concerning M2e-based vaccines were addressed in this PhD project.

NGS as tool to study the genetic diversity of influenza A viruses

The study of the evolution of influenza and other viruses has been facilitated by the recent progress in the field of next-generation sequencing (NGS). At the beginning of this PhD project, controversial results on the accuracy of the different sequencing platforms were reported [7, 8]. Consequently, we first compared the suitability of the two most potent NGS platforms to study the genetic diversity in a viral population with high sensitivity: the Illumina MiSeq and the Ion Torrent Personal Genome Machine (PGM). In addition, reports on how the sequencing data is processed or analyzed are often either lacking or ill-explained in the literature, which impedes the comparison of results obtained by different NGS studies. Consequently, we designed an NGS data analysis pipeline for variant calling of genetically diverse RNA virus populations. We opted for the user-friendly bioinformatics program CLC Genomics Workbench (Qiagen Bioinformatics), which makes it possible for virologists to generate and analyse their data themselves. We think that such a user-friendly program will lower the threshold for virologist to perform NGS experiments, since knowledge on command lines, which is mostly linked to large data analysis, is not required. Multiple open-source software tools are also available to perform NGS data analysis [9]. However, their use is often restricted by the limited number of possible applications, the scarce maintenance of the software and/or the limited availability of computation and data storage resources on their servers.

We first applied this NGS data analysis pipeline on sequencing reads obtained for plasmid DNA samples and found that the accuracy of the reads sequenced on the Illumina MiSeq was one and a half times higher than on the Ion Torrent PGM. A higher accuracy for Illumina MiSeq was also observed by Loman *et al.* and Quail *et al.,* who compared the characteristics of both sequencers by

sequencing either the O104:H4 isolate of *Escherichia coli*, isolated during an event of food poisoning in Germany in 2011, or four different microbial genomes differing in mean GC-content from 19.3 to 67.7% [7, 8]. However, we observed a lower substitution error rate on the Ion Torrent PGM compared to the Illumina MiSeq, which is in contrast with Loman *et al.* who compared the characteristics of both sequencers on the raw sequencing data [7]. The obtained sequence read length was similar on both sequencers (+/- 215 nucleotides) after processing the data, although the 400 base pair kit was used for sequencing on the Ion Torrent PGM. This relative short average sequence length obtained on the Ion Torrent PGM is not the result of the small size of the sequenced DNA fragments, since the peak fragment length of the DNA fragments before emPCR was situated around 450 base pairs. However, the default settings in the Ion Torrent PGM sequencing software ('Ion Torrent suite') were used to obtain the sequencing output. These default settings also include the removal of low quality 3' ends (mean Phred score of at least 15 in a base window of 30). This suggests that the advantage of long reads obtained on the Ion Torrent PGM is cancelled by their relatively low quality. This finding is in agreement with a report by Jünemann *et al.*, where read lengths of 194 (+/- 89.08) nucleotides were obtained with the 400-bp PGM chemistry [10].

To sequence the viral diversity in an influenza virus population on the NGS platforms, the segmented RNA genome had to be converted to DNA. We designed an RT-PCR protocol in which we took advantage of the conserved sequences at the influenza genome ends, which resulted in efficient amplification of all eight genome segments of the influenza A/Puerto Rico/8/34 (PR8) virus. Critical steps in this protocol appeared to be the primer concentration and annealing and elongation times. The primers used for full-length amplification of all eight influenza A vRNA segments have a relatively limited sequence homology with their target RNA sequences. Therefore, it is very important to run the RT-PCR at precise primer concentrations and annealing temperature conditions. Indeed, we noticed that tuning these conditions was important to avoid the "aspecific" amplification of the first 847 nucleotides of the HA segment, due to partial overlap of one of the primers with the coding region of HA. This sequence of nine nucleotides is also present in the HA consensus sequence of H1N1 influenza A viruses (unpublished results). An advantage of this influenza-specific RT-PCR protocol compared to others, is the amplification of the full influenza genome in a single reaction, using a single set of primers. Consequently, performing multiple PCRs is omitted, which decreases the change of contamination [11-13]. The sequence of the segment ends is conserved between all subtypes of influenza A viruses and can thus be used to study all influenza A virus strains. The broad applicability of this RT-PCR protocol was also verified for other influenza A viruses: mouse-adapted PR8 (H1N1), H9N2/CA09 (A/quail/Hong Kong/G1/1997 (H9N2) x A/California/04/2009 (H1N1)), A/quail/Hong Kong/G1/1997 (H9N2), A/Belgium/145-ma/2009(H1N1), X47 (A/Victoria/3/75 (H3N2) x PR8), NIBRG-14 (A/Vietnam/1194/04 (H5N1) x PR8), PR8-NS1(1–73)GFP (H1N1) and A/Brisbane/59/2007 (H1N1) (unpublished results, and Mancera Gracia *et al.,* submitted) [14]. This demonstrated that the RT-PCR protocol, along with the NGS data analysis pipeline, is broadly applicable and can be used *e.g.* for viral resistance testing and the study of virus evolution.

The influenza-specific RT-PCR protocol was performed on the PR8 virus, created by reverse genetics, to compare the suitability of both sequencers to determine the variants present in a viral population. The influenza reference sequence could be obtained for both sequencers by *de novo* assembly using de Bruijn graphs. Mapping the reads to the influenza reference genome, based on the eight plasmids to generate the virus, resulted in ample coverage across all segments. However, the segments ends were underrepresented after transposase-based Nextera XT DNA fragmentation followed by Illumina MiSeq sequencing. This is the result of the intrinsic nature of the transposition reaction since the transposon sequence cannot be inserted at the genome ends. In addition, a significant coverage dip near the middle of the NP segments as well as a dip around position 600 of the PA segment was observed. When the same sample was sequenced after Covaris shearing, this bias in sequencing coverage was absent, indicating that Nextera XT fragmentation is partially sequence dependent as a consequence of the sequence bias of its transposase [15]. A bias in genomic coverage was also observed after Nextera fragmentation and Illumina sequencing of the small genome of the PA1 bacteriophage [16]. Random fragmentation by mechanical shearing is thus preferred for smaller genomes to ensure comparable coverage across the genome.

It is important to distinguish variants in the viral population from technical errors introduced during RT-PCR amplification and the NGS chemistry. Consequently, we verified our NGS data analysis pipeline on wild type and mutant samples in duplicate, and mixed at different ratios. We concluded that variants in the influenza virus genome that appear with a frequency below 0.5% are very difficult to distinguish from the background noise and that a threshold of 0.5% should be used to interpret the genetic diversity of RNA viruses. Applying this threshold to our samples, resulted in 19 detected mutations in PR8 virus and 29 mutations (including the two tracer mutations) in PR8mut virus. The highest variation was detected in HA, being 13 out of 19 mutations for PR8 and 16 out of 29 for PR8mut, which is also the most variable viral protein in nature [17]. The empirical determined threshold of 0.5% makes it thus possible to sensitively study the variants present in a viral population. However, the errors introduced during RT-PCR, together with the error-rate of the NGS platforms, limits their use for the identification of each variant present in the viral quasispecies. Another drawback of the NGS platforms is their relatively short read length which limits their use for detecting mutations that are linked in a genome segment. The long read-length reported for the single-molecule sequencers will probably fill this gap in the future, although the use of these sequencers for the detection of viral diversity is currently limited by their low sequencing accuracy.

Recently, different techniques have been applied to further reduce the error threshold of NGS, each with their own benefits and drawbacks. One of these techniques is CirSeq, developed by Acevedo and colleagues [18]. In this technique, RNA fragments of 80 to 90 bases are circularized, followed by 'rolling circle' reverse transcription, resulting in DNA fragments containing three repeats of physically linked cDNA fragments [18]. Consequently, real mutations that are present in the viral RNA will be shared by all repeats in the DNA fragment. In contrast, variation introduced during RT-PCR or sequencing, will not be present in all repeats and can be filtered out. This reduces the estimated

error probability to 10^{-6} per base, which is below the estimated mutation rate of influenza viruses and makes it thus possible to study each variant in a viral population [19]. This technique has been successfully applied to define the mutation rate of poliovirus and to resolve the mutational landscape of its viral population [19]. However, this technique requires a large amount of pure viral RNA (1 to 5 μg) and is not applicable for *de novo* assembly, since a reference genome is required to map the tandem repeat reads [18]. In addition, due to the limited read lengths of current NGS platforms, only tandem fragments with units of 80 to 90 nucleotides can be sequenced. Furthermore, since each fragment is sequenced three times, the sequencing throughput drops with the same factor.

Another method to reduce the error-threshold is 'primer ID sequencing' [20]. In this technique, a random barcode, 'the primer ID', is added to the primer used for cDNA synthesis. This makes it possible to cluster the sequenced DNA fragments carrying the same primer barcode afterwards. By doing so, a real mutant can be discriminated from technical errors as it will be present in all sequencing reads of a barcoded cluster. A major limitation of this technique is that errors introduced during reverse transcription, which is an error-prone step, cannot be corrected. Ideally, to eliminate all errors and biases introduced during RT-PCR and sequencing, direct RNA sequencing of genomic viral RNA should be performed [21]. Oxford Nanopore Technologies recently released the first results of direct RNA sequencing on their MinION platform [22]. The long read length of this platform makes it also possible to study linkage of viral variants. However, the use of the MinION system to study variants in a viral population is currently limited because of its high reported error rate of about 10%, although its accuracy is steadily improving [23-25].

From the NGS platform comparison, we concluded that the Illumina MiSeq is most suitable to detect variant sequences in a viral population since this platform has the lowest total error rate. In addition, the output of the Illumina MiSeq, which is at least four times larger than the Ion Torrent, makes it possible to analyse many more samples in parallel. However, the time to data of Ion Torrent PGM is significantly shorter, which can be an important advantage in a clinical setting.

The field of NGS is fast evolving with the first NGS platform - the 454 Genome Sequencer 20 (Roche) - only released eleven years ago. This platform was characterized by a read length of approximately 100 bases and a total throughput of 20 Mb per run. In the meanwhile, the 454 pyrosequencing platforms have been discontinued in 2016, the sequencing read length is increased up to 400 nucleotides (Ion Torrent PGM and S5) and the maximal sequencing throughput on the market is obtained on the Illumina HiSeq X (up to 1800 Gb). In addition, single-molecule sequencers were brought to the market, which eliminate the sequencing bias introduced during library preparation and make it possible to sequence DNA fragments of several kilobases [26]. However, their use in influenza research is currently limited by their high error rate.

Illumina MiSeq sequencing disfavours the 'CCNGCC' sequence motif

Influenza reporter viruses are of use in several applications, *e.g.* to study the kinetics of viral replication, virus tropism or to screen for influenza antivirals. However, the reporter gene has no selective advantage for the virus and can thus be deleted upon viral replication. Gradual loss of reporter gene expression has been reported for GFP-expressing influenza viruses [27]. However, we note that recently, the group of Dr. Kawaoka, described a PR8-based backbone vector for reporter genes that seems to be exceedingly more stable *in vitro* and *in vivo* than the earlier developed influenza reporter viruses [28].

The high sequencing throughput of NGS makes it a useful tool to study the genetic stability of influenza reporter viruses. We verified the suitability of NGS to study the genetic stability of a GFP-expressing influenza reporter virus, recently developed in our lab, using the aforementioned influenza-specific RT-PCR protocol and NGS data analysis pipeline [14]. The detected total variation and the variation per viral genome segment between the virus stock of this GFP-expressing influenza reporter virus and the wild type PR8 virus was similar, which suggests that both viruses are equally genetic stable. Interestingly, a remarkable drop in sequencing coverage was detected at the GFP coding sequence, although this virus appeared to be genetically stable based on phenotypic analysis of this virus upon multiple cycles of viral replication [14]. This drop in sequencing coverage could be linked to a sequencing bias of the Illumina MiSeq for the 'CCCGCC' motif in the GFP coding sequence, since mutating this motif resolved the bias in sequencing coverage. A similar drop in sequencing coverage upstream of the 'CCNGCC' motif was previously reported by Ekblom and colleagues in mitochondrial DNA, after deep sequencing the whole genome of wolverine (*Gulo gulo*) and collared flycatcher (*Ficedula albicollis*) [29]. This study demonstrated an inverse relation between sequence coverage and error rate, based on the presence of a position with high error rate and low sequencing coverage directly upstream of the 'CCNGCC' sequence motif [29]. A high error rate for the 'GGCGGG' motif was also reported by Quail *et al*. after Illumina MiSeq sequencing of microbial genomes which differ in mean GC-content from 19.3 to 67.7% [8].

We have tried to explain why the drop in sequencing coverage was observed only at two out of the twelve 'CCNGCC' motifs in the pHW-NS1(1-73)Dmd-GFP-NEP plasmid, and why this motif is required but not sufficient for the adjacent drop in sequencing coverage. We hypothesised that the sequences upstream and downstream of the 'CCNGCC' motif may also play a role. Therefore, we tried to identify a recurrent primary sequence pattern in the flanking regions of such motifs. We generated a sequence logo based on the 150 nucleotides upstream and downstream, since the sequenced DNA fragments have a length of approximately 300 bases, of the 'CCNGCC' and the shorter 'CNGCC' motifs present in the pHW-NS1(1-73)Dmd-GFP-NEP plasmid encoding the NS-GFP segment and in the plasmids containing the other GFP variants. Unfortunately, due to the limited number of motifs, no additional sequence pattern that was under or overrepresented in the regions flanking a 'CCNGCC' or 'CNGCC' motif could be revealed (unpublished results). More extensive studies on coverage analysis

are thus required to investigate if it is possible to come up with a sequence pattern that can predict the sequence bias at a given 'CCNGCC' motif. Such a discriminatory sequence motif would be widely applicable to correct for sequencing bias *e.g.* during RNA sequencing data analysis, where sequencing coverage is linked to gene expression levels.

We can conclude from this study that it is important to take a potential sequencing bias into account when performing NGS coverage analysis. In our study, we could have come to the false conclusion that our GFP-expressing influenza reporter virus was genetically unstable. The real genetic stability of the GFP sequence in this recombinant virus can be determined by creating a variant virus through reverse-genetics which carries a C504T and/or C507T (GFP numbering) mutation in the sequence encoding for GFP. Alternatively, one can use an alternative reporter gene to create a reporter virus after having ensured that the corresponding coding sequence is not subject to potential coverage bias for the anticipated NGS analysis. Next to this, the genetic stability of this GFP-expressing influenza reporter virus can be analyzed on a sequencing platform that uses a different sequencing chemistry, such as the Ion Torrent PGM.

The protective potential of M2e-based vaccines can be enhanced by directing the elicited immune response towards the induction of IgG2a antibodies

The protective mechanism underlying M2e-based vaccines was further unravelled in this PhD project by investigating the functional engagement of FcγR family members by two mouse monoclonal antibodies (mAbs) that bind M2e with similar affinity, but are either of the IgG1 or IgG2a antibody isotype. In collaboration with the group of Dr. Harmut Hengel, we could directly correlate findings from an *in vitro* assay in which all mouse FcγRs were evaluated one by one for the binding of these two mAbs, which were bound in turn to M2 on influenza A virus infected cells; to protection against viral challenge in mice that differ in their deficiencies in the FcγR compartment. We showed that M2e-specific antibodies of the mouse IgG1 isotype require low affinity receptor FcγRIII to accomplish protection upon infection, whereas the IgG2a antibodies can protect against influenza A virus challenge via any of the three activating FcγRs. This study thus confirms the essential role of Fcγ receptors in protection by M2e-specific IgG. In addition, the superior protection of M2e-specific IgG2a over IgG1 against influenza A virus challenge can be linked to its potential to bind all three activating receptors on effector cells, such as NK cells, macrophages and neutrophils [30].

An important role has also been assigned to FcγRs in protection against influenza A virus infection by broad-spectrum HA-stem specific IgG mAbs, with a higher protective potential for IgG2a mAbs compared to their IgG1 counterpart [31, 32]. DeLillo *et al.* even suggested that all broadly neutralizing anti-influenza antibodies require FcγRs to protect against influenza infection, based on an *in vivo* comparison of broadly neutralizing as well as strain-specific anti-HA and anti-NA IgG2a mAbs, with their mutant counterparts that lack FcγR binding [33].

The anti-M2e antibody pair used in this study originates from splenocytes isolated after vaccination of mice with soluble tetrameric M2e and are hence not identical in their antigen binding site. However, we have proven that both antibodies bind peptide M2e and M2e expressed on infected cells with similar affinity. In addition, Ala-scan analysis showed that both antibodies require similar amino acid residues in M2e for their binding. The impact of antibody isotype on binding to the specific FcγRs should ideally be studied using recombinant produced mAbs which share the same variable region and only differ in their constant region. However, it has been shown that the constant heavy chain domains may affect the structure of the antigen binding site and the herewith associated antibody binding specificity and affinity [34, 35]. In addition, we demonstrated that both mAbs carry a similar glycosylation pattern at the conserved N-glycosylation site in their antibody Fc tail, which has an important role in Fc-dependent effector functions [36-39]. Consequently, we considered that this antibody pair was well suited to investigate the impact of antibody isotype and their binding to specific FcγRs on effector cells, on the protective effect of M2e-specific antibodies.

We demonstrated an essential role for FcγRs in M2e-based protection, which is in agreement with former studies [40, 41]. Furthermore, we showed that IgG2a mAbs were still protective in mice that lack FcγRIII, while M2e-specific IgG1 mAbs largely failed to protect these mice [30]. These results are in line with prefacing studies, which indicate that NK cells may contribute to M2e-based protection, although their role is probably not essential [30, 41-45]. We also reported for the first time that FcγRIV, which plays an important role in IgG2a-dependent effector activities *in vivo*, contributes to M2e-based protection [46]. A possible role for FcγRIV in protection by M2e-specific IgG2a mAbs was first suggested in the FcγRI and FcγRIII double knock-out mice, since these mice were fully protected against viral infection when treated with M2e-specific IgG2a mAbs. This was later confirmed in *fcgr4$^{-/-}$* mice, which were protected by IgG2a but displayed significantly more body weight loss after influenza A virus challenge than wild type mice. In contrast, wild type and *fcgr4$^{-/-}$* mice were equally well protected by M2e-specific IgG1 mAbs, although worse than IgG2a in terms of weight loss. FcγRIV thus contributes to protection, but has no determining role. An important role has also been assigned to FcγRIV in protection against influenza viruses by the broadly neutralizing HA stalk-specific antibodies [31]. The higher morbidity observed in challenged FcγRI and FcγRIII double knock-out mice after treatment with M2e-specific IgG2a mAbs, compared to wild type mice and mice that lack FcγRIII, also suggests that FcγRI significantly contributes to their protection. However, future studies comparing FcγRI knock-out mice, FcγRI and FcγRIV double knock-out mice and triple-deficient mice lacking functional FcγRI, FcγRIII and FcγRIV, will be required to determine the relative contribution of the two high affinity activating FcγRs in M2e-based protection.

We demonstrated that FcγRI and FcγRIV are both important for the protective effect of M2e-specific IgG2a antibodies. These receptors are highly expressed on alveolar macrophages [30]. An essential role in M2e-based protection has previously been assigned to these cells, since anti-M2e immune serum failed to protect mice in which the alveolar macrophages were eliminated by clodronate-loaded liposomes, although NK cells were still present [41, 47]. Moreover, protection by M2e-

immune serum was restored in FcγRI and FcγRIII knock-out mice, depleted of alveolar macrophages, after transfer of wild type alveolar macrophages [41]. The essential role of alveolar macrophages in M2e-based immune protection, should be further exploited. The *csf2rb*[-/-] mice, which lack alveolar macrophages, are an interesting mouse model for this purpose [48, 49]. However, the influence of pulmonary alveolar proteinosis, as a result of defective clearance of surfactant, should be excluded by transferring alveolar macrophages isolated from *fcer1g*[-/-] or wild type mice, as positive control [50-53]. To further dissect the cell-specific contribution of FcγRs in M2e-based protection, it will also be interesting to create conditional FcγR knock-out mice for alveolar macrophages, and possibly NK cells. In addition, to extrapolate these finding to humans, which show a distinct FcγR expression pattern, the functional contribution of human FcγRs in the protective effect of M2e-specific antibodies should be investigated. Since human IgG1 and -3 isotype antibodies can be considered to be the functional counterparts of mouse IgG2a antibodies, it will be interesting to create humanized IgG1 and IgG3 anti-M2e mAbs which carry the variable domain of our murine anti-M2e mAbs. The FcγR humanized mice created by Smith and colleagues, in which all murine FcγRs have been deleted and all human FcγRs have been inserted under the control of their human regulatory elements, are a good mouse model to evaluate the protective effect of these humanized anti-M2e mAbs [54].

The results from our study can be implemented to optimize M2e-based vaccine formulations. We showed that M2e-specific IgG2a antibodies have a higher protective effect compared to IgG1, with minimal morbidity and absence of mortality in wild type mice. Moreover, increased protection upon viral challenge is linked to the induction of IgG2a antibodies following M2e-based vaccination [55-58]. Translating these findings to human vaccine design, will result in more potent M2e-based vaccines. Induction of human IgG1 an -3 isotype antibodies should thus be promoted by using an M2e-based vaccine formulation which stimulates a Th1-specific immune response, such as the licensed MF59, AS03 and AS04 adjuvants [59]. Currently, the majority of the licensed influenza vaccines are administered intramuscularly, which results in systemic protective immunity. However, the highest protective potential for M2e-based vaccines in mice is obtained when they are administered intranasally, which is also the site where initial influenza infection takes place [60]. In addition, intranasal vaccination induces mucosal as well as systemic immunity and can be applied needle-free, which allows self-vaccination. Further research on intranasal vaccination of humans is thus required.

Next to optimizing M2e-based vaccines, there is also an important need for a read-out system to test their vaccine effectiveness. The efficacy of conventional influenza vaccines is currently evaluated by measuring the hemagglutination inhibition titers. However, this assay measures only the elicited number of antibodies that target epitopes in or close to the receptor binding site of influenza HA and can thus not be applied to evaluate the protective effect of M2e-based antibodies. The same holds true for broadly neutralizing antibodies elicited by vaccines that target the conserved stalk domain of HA [31]. Since these vaccines mainly exert their protective effect by binding to FcγRs, a novel read-out to determine the vaccine efficacy is required. An *in vitro* assay that can correlate the presence of

antibodies in immune serum with their potential to induce ADCC and ADPC would be suitable for this. Consequently, a test based on the *in vitro* FcγR activation assay, as used in this study, could be implemented for this. However, further clinical studies, in the first place demonstrating a clinical benefit of an M2e-based vaccine, are required to reveal a correlation between the *in vivo* protective potential of M2e immune serum and the *in vitro* activation of FcγRs.

M2e-based immune pressure selects for limited variation in M2e and alternative escape routes in immuno-deficient mice

Influenza viruses can escape from the selection pressure imposed by the neutralizing antibodies targeting the highly variable antigenic sites of HA, which are elicited upon natural infection or vaccination with the licensed influenza vaccines. The effectiveness of the current vaccination strategies is limited by this antigenic drift and requires annual vaccination with the updated vaccine. Since M2e-based vaccines differ from the current vaccination strategies which mimic the natural immune response, it is important to investigate how influenza viruses evolve when immune pressure is imposed on a conserved epitope such as M2e.

In this PhD project we started to investigate how influenza viruses evolve under M2e selection pressure. The different mode of action of the M2e-based vaccines, compared to the licensed influenza vaccines where antigenic drift can be mimicked *in vitro*, necessitates for an *in vivo* M2e escape selection model [61-63]. We opted here for immunodeficient SCID mice since they are proficient in antibody-dependent effector functions, on which anti-M2e immunity relies, but lack B and T cells and thus cannot mount an adaptive immune response that would normally lead to the clearance of the influenza viruses. Influenza viruses can thus replicate for a prolonged time in these SCID mice and are thus nicely suited as mouse model to select for escape under M2e-immune pressure, in the form of passively transferred M2e-specific mAbs. In the future, it will be important to investigate how influenza viruses evolve in the presence of a polyclonal M2e immune response in an immune competent host by serially passaging virus present in BAL fluid, *e.g.* isolated on day five post infection, of infected M2e-vaccinated mice. In this set-up it would also be interesting to include mice vaccinated with the inactivated whole virus vaccine and a vaccine based on the conserved HA-stalk domain, to compare the kinetics of viral escape between the different vaccine strategies. Up to now, there is no experimental evidence that influenza viruses can escape from a polyclonal anti-M2e immune reaction [64].

The survival of infected SCID mice was significantly prolonged in the presence of M2e-specific mAbs. In agreement with the aforementioned study, a significantly higher impact on survival was observed for M2e-specific IgG2a antibodies than IgG1. The viral diversity present in the BAL fluid of infected mice was analysed by NGS, and escape in the M2e sequence was only detected in mice that were treated with mAbs that bind an epitope spanning amino acids 5 to 15 in M2e. However, the variation in M2e in these samples was limited to a proline to histidine or leucine mutation at position 10 or a

leucine to threonine mutation at position 11, mutations that are all silent in M1. These results are in line with the crystal structure of M2e peptide in complex with the Fab fragment of one of these mAbs, which showed a critical role for Glu6, Pro10, Ile11 and Trp15 for mAb binding [65]. The impact of these M2e escape mutants in immune competent hosts will probably be limited, since we subsequently showed that these M2e escape mutants retained enough antigenicity to be bound by polyclonal serum of mice vaccinated with M2e-HBc. Moreover, it has been shown that the polyclonal immune response elicited by vaccinating mice with M2e-MAP, comprising four copies of the human consensus M2e sequence, resulted in cross-protection against infection with a mutant PR8 virus carrying M2e-P10L or P10H [66]. In nature, some genetic flexibility is also allowed at position 10 and 11. A histidine or leucine at position 10 is usually found in the M2e sequence of the highly pathogenic avian influenza H5 and H7 viruses [2]. In addition, a threonine is commonly found at position 11 in the M2e sequence of avian and swine influenza A viruses (Chapter 2, figure 1). An M2e-vaccine should thus likely comprise a mixture of M2e sequences, for example arranged in tandem repeats on virus-like particles as reported by Kim and colleagues, to ensure maximal coverage of the M2e sequence diversity in circulating influenza A viruses and to reduce the risk of M2e escape virus emergence [67].

There was no variation detected in M2e when SCID mice were treated with an M2e-specific antibody that binds the eight N-terminal amino acids in M2e. This finding is in line with a phase 2 clinical trial in experimentally infected, healthy volunteers who received a human M2e-specific mAb (TCN-032) that binds to the SLLTE sequence at the N-terminus of M2e [68]. Deep sequencing of the last positive nasal swab sample did not result in the detection of low frequency variation in the M2 sequence [68]. The M2e sequence is thus remarkably stable, even in the presence of an immune pressure.

The selected variation in M2e is limited to substitutions which are silent in M1, which suggests a strong natural selection pressure for the virus to maintain its M1 sequence conserved [69]. The coding sequence for the N-terminus of the M1 protein and the eight N-terminal amino acids of the M2 protein overlaps with the conserved 5' packaging signals of the M segment, which will contribute to the genetic constraint imposed on this sequence [70-72]. The three different anti-M2e mAbs used in this study were derived from splenocytes of mice vaccinated with tetrameric M2e (M2e-tGCN4) and thus reflect some of the epitope diversity elicited upon vaccination. The absence of variation in M2e when mice were treated with mAbs binding to its N-terminus and the presence of limited variation when mice were treated with the mAbs binding to amino acids 5 to 15 in M2e, suggests that escape in the M2e sequence will be hindered by a polyclonal immune response in an immune-competent host.

The wild type M2e sequence was retained in all mice that succumbed to infection after treatment with M2e-mAbs which recognize the N-terminus and in half of the mice treated with M2e-specific mAbs binding to amino acids 5 to 15. However, mutations at high frequency were observed in these samples, mainly in the polymerases PB2 and PA and in HA. Therefore, we considered the possibility

that an alternative escape route from M2e immune pressure was possible for the virus in our model. The detection of various combinations of different mutations suggests that the virus can follow multiple genetic pathways to escape from M2e-based immune pressure. A small number of these mutations were independently selected in multiple mice, which suggest that these mutations are advantageous for escape. Several of these mutations have been linked to enhanced viral replication or receptor binding. For example, an enhanced *in vitro* polymerase activity and higher pathogenicity in mice have been reported for the I504V mutation in PB2 [73, 74]. In addition, we noticed a remarkable difference in plaque morphology between the parental virus and a PR8 virus carrying the mutations observed in PB2, PA and HA when plaques were immunostained with anti-M2e antibodies and subsequently with polyclonal immune serum directed against the vRNPs. This difference in plaque morphology (essentially the ratio of the plaque size as revealed by anti-M2e immune staining over the size of the plaques visualized by anti-vRNP immune serum) was smaller in the HPP mutant virus compared with wild type virus. These ratios still require more objective quantification by appropriate imaging software, such as Volocity (Perkin Elmer). Western blot analysis of samples prepared at different time points after infection, appear in line with a delayed M2 expression compared to new vRNP synthesis. However, also here, quantification of Western blot signals (*e.g.* direct quantitative analysis of protein levels using the Amersham Imager 600) are required to substantiate the finding. On the other hand, flow cytometric analysis of membrane expressed M2 on the surface of infected HEK293T cells determined at different time points after infection, did not reveal such a delay in M2 display. There are several explanations possible for this seeming discrepancy between the total protein analysis and the flow cytometric approach that provides information on M2 expression at the cell surface. First, the kinetics experiments were performed using a human cell line and should be repeated using a mouse cell line, *e.g.* MLE-15 cells, to better recapitulate what may be going on in the SCID model. In addition, the amount and timing of virus released from the infected cells should be included in this kinetics experiment. This could lend support to the hypothesis that a shorter time between the start of M2 expression and virus budding could allow mutant virus to be shed before FcγR-expressing immune cells can attack infected cells that are opsonized with M2e-specific antibodies. The mutations in PA and PB2 that arise under anti-M2e IgG selection pressure in SCID mice may raise the concern that these could increase the replication rate of the virus and possibly alter their pathogenicity. This can be tested by defining the LD_{50} of these viruses in immune competent mice. The delayed expression of M2 should also be verified *in vivo* by immunochemistry and the potential of this virus to evade M2e immune pressure should be confirmed either in SCID mice treated with anti-M2e mAbs and in M2e-vaccinated, immunocompetent mice. The polymerase complex has also been suggested to play a role, albeit controversial, in splicing of the M1 mRNA, hence it will also be interesting to study the impact of the selected mutations in PB2 and PA on M1 mRNA splicing [75, 76].

In vivo escape selection from M2e-specific mAbs was first described by Zharikova and colleagues [77]. In this study, infected SCID mice were chronically treated with M2e-specific IgG2a antibodies

which all recognize a similar epitope within amino acids 4 to 16 of M2e, as determined by a competition assay where N- and C-terminal truncated M2e-peptides in solution interfered with coated full-length M2e for binding to these mAbs [78]. Our results confirm some of the findings in this pioneering study but also revealed new findings. First, in both studies survival was prolonged when infected mice were treated with M2e-specific mAbs, demonstrating their protective effect. In addition, we could demonstrate that M2e-specific IgG2a mAbs have a higher protective effect compared to their IgG1 equivalent. In both studies, escape viruses with a mutated M2e epitope, which abolishes recognition by the mAbs, were selected. The variation in the study of Zharikova *et al.* was restricted to a proline to histidine or leucine mutation at position 10 in M2e in 65% of all treated mice. We got a better understanding of the tolerated variation in M2e by using M2e-specific mAbs that bind different epitopes and/or are from a different antibody isotype, which resulted in the detection of an additional isoleucine to threonine mutation at position 11. In addition, as a consequence of the higher sensitivity of our deep sequencing approach compared to direct Sanger sequencing of purified RT-PCR products, we were able to detect a mixture of these M2e variants in a single sample. Furthermore, we also demonstrated that variation in the very conserved eight N-terminal amino acids of M2e is not tolerated. Importantly, The study of Zharikova and colleagues was limited to Sanger sequencing of purified M2e RT-PCR products, derived from viral RNA after amplifying the virus present in lung homogenates on cells. Using whole genome deep sequencing, we were able to analyze the variation across the whole influenza virus genome, which also resulted in the detection of compensatory mutations outside the M2e epitope. In addition, by directly analyzing the shed virus in the lung, we omitted cell culture adaptation. Hence, our study demonstrates the benefits of whole genome sequencing of influenza viral genomes in M2e escape selection.

Interestingly, the level on which escape from M2e-based immune pressure takes place is different from the current influenza vaccine strategies. Escape from the neutralizing antibodies elicited by the licensed influenza vaccines is limited to the level of viral entry and as such, only viruses that escape the immune response before the initial infection will result in viral progeny. These antigenically drifted escape mutants will carry substitutions in HA.This conventional type of escape is well known and can be taken care of by including the variant in the vaccine. However, M2e-based vaccines are infection permissive, and as such, selection takes place on the level of the infected cell, full of progeny influenza viruses. The pool of viruses on which selection takes place is thus higher when compared to the licensed influenza vaccines. However, the infective-permissive character of the M2e-based vaccines also allows the induction of B cells and a cross-reactive cellular immune response upon infection, which will impose an additional layer of immune selection on the challenge virus and little opportunity for the viruses to escape anti-M2e immunity.

Influenza viruses which can escape to M2e-based immunity require sustained host-to-host transmission to emerge in the host population. Hence, it is important to investigate the impact of anti-M2e immunity on influenza virus transmission and to study the effect of selected M2e escape variants on this process. Ferrets are the favored animal model to study influenza virus transmission

since they are highly susceptible to these viruses and show similar disease symptoms as in humans, moreover, successful transmission can be obtained via aerosols and respiratory droplets ('airborne transmission') or direct contact [79-83]. Nachbagauer *et al.* recently demonstrated that ferrets with HA-stalk immunity, induced by a universal influenza vaccine approach using vaccination with chimeric HA vaccine constructs, were either protected against aerosol-transmitted H1N1 influenza virus infection or displayed significantly reduced virus shedding in a shorter time span in the nasal washes [84]. Mice can also be used to study influenza virus transmission. However, the success of influenza virus transmission in mice is variable and depends strongly on the mouse model and influenza virus strain used [85-87]. Recently, effective direct-contact transmission of mouse-adapted A/FM/1/47(H1N1) and A/Udorn/307/72 (H3N2) in CFW mice was reported by the group of Epstein [86]. Challenging CFW mice with A/FM/1/47(H1N1) resulted in significantly reduced virus transmission after vaccination with rAd vectors expressing NP or M2, when compared to mice receiving the control vaccine [86].

Concluding remark

The work described in this thesis provides a workflow to study the genetic diversity of influenza viruses by NGS and shows that the Illumina MiSeq, due to its high accuracy, is currently the preferred method for this. In addition, we also showed that potential sequencing biases should be taken into account when performing NGS coverage analysis. This PhD thesis also contributes to the understanding of the protective mechanism of the M2e-based vaccines. We demonstrated that M2e-specific IgG2a antibodies have a higher protective potential compared to IgG1, which correlates with their potential of binding to all activating FcγRs on effector cells. This information will be of use for vaccine design and vaccines should thus be designed to elicit a robust Th1-biased immune response. In this PhD project we also demonstrated that only limited variation is tolerated in M2e under immune pressure and that these M2e escape mutants can still be recognized by polyclonal anti-M2e immune serum. The chance that viruses will emerge with a mutated M2e sequence in an immune competent host upon M2e-vaccination will thus probably be low. However, the possibility that viruses that follow alternative escape routes will emerge in the human population should be further investigated.

A vaccine based on M2e is a good candidate for a 'universal influenza A vaccine' because of their broad-spectrum protective effect and the genetic constraints towards variation in M2e. Several of these M2e-based vaccines can also be produced in bacteria at large scale and low cost, omitting the need for eggs or cell culture as is the case for the licensed influenza vaccines. Moreover, the sequence conservation of M2e between influenza A viruses, makes stockpiling of these vaccines possible, which is an important advantage in pandemic preparedness. Ideally, the licensed influenza vaccines should be replaced by an influenza vaccine targeting conserved epitopes on both the

influenza A and B viruses. However, the protective potential of conserved epitopes in influenza B viruses is less exploited.

References

1. Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W: A universal influenza A vaccine based on the extracellular domain of the M2 protein. *Nature medicine* 1999, 5(10):1157-1163.
2. Deng L, Cho KJ, Fiers W, Saelens X: M2e-Based Universal Influenza A Vaccines. *Vaccines* 2015, 3(1):105-136.
3. Safety Study of Recombinant M2e Influenza-A Vaccine in Healthy Adults (FLU-A) [https://clinicaltrials.gov/ct2/show/results/NCT00819013?sect=Xa0156#outcome4]
4. Turley CB, Rupp RE, Johnson C, Taylor DN, Wolfson J, Tussey L, Kavita U, Stanberry L, Shaw A: Safety and immunogenicity of a recombinant M2e-flagellin influenza vaccine (STF2.4xM2e) in healthy adults. *Vaccine* 2011, 29(32):5145-5152.
5. Talbot HK, Rock MT, Johnson C, Tussey L, Kavita U, Shanker A, Shaw AR, Taylor DN: Immunopotentiation of trivalent influenza vaccine when given with VAX102, a recombinant influenza M2e vaccine fused to the TLR5 ligand flagellin. *PloS one* 2010, 5(12):e14442.
6. Dynavax Reports New Phase 1a and Phase 1b Data for Universal Flu Vaccine Candidate [http://investors.dynavax.com/releasedetail.cfm?ReleaseID=551606]
7. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ: Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology* 2012, 30(5):434-439.
8. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* 2012, 13:341.
9. Software [http://seqanswers.com/wiki/Software]
10. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J *et al*: Updating benchtop sequencing performance comparison. *Nature biotechnology* 2013, 31(4):294-296.
11. Croville G, Soubies SM, Barbieri J, Klopp C, Mariette J, Bouchez O, Camus-Bouclainville C, Guerin JL: Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *Journal of clinical microbiology* 2012, 50(9):2881-2887.
12. Bidzhieva B, Zagorodnyaya T, Karagiannis K, Simonyan V, Laassri M, Chumakov K: Deep sequencing approach for genetic stability evaluation of influenza A viruses. *Journal of virological methods* 2014, 199:68-75.
13. Hoper D, Hoffmann B, Beer M: A comprehensive deep sequencing strategy for full-length genomes of influenza A. *PloS one* 2011, 6(4):e19075.
14. De Baets S, Verhelst J, Van den Hoecke S, Smet A, Schotsaert M, Job ER, Roose K, Schepens B, Fiers W, Saelens X: A GFP expressing influenza A virus to report in vivo tropism and protection by a matrix protein 2 ectodomain-specific monoclonal antibody. *PloS one* 2015, 10(3):e0121491.
15. Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS: Tn5/IS50 target recognition. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(18):10716-10721.
16. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X *et al*: Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology* 2010, 11(12):R119.
17. Ellebedy AH, Webby RJ: Influenza vaccines. *Vaccine* 2009, 27 Suppl 4:D65-68.
18. Acevedo A, Andino R: Library preparation for highly accurate population sequencing of RNA viruses. *Nature protocols* 2014, 9(7):1760-1769.

19.   Acevedo A, Brodsky L, Andino R: Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 2014, 505(7485):686-690.

20.   Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R: Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108(50):20166-20171.

21.   Barnard R, Futo V, Pecheniuk N, Slattery M, Walsh T: PCR bias toward the wild-type k-ras and p53 sequences: implications for PCR detection of mutations and cancer diagnosis. *BioTechniques* 1998, 25(4):684-691.

22.   Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M, Lloyd J, Warland A, Pantic N, Admassu T, Ciccone J *et al*: Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* 2016.

23.   Magi A, Giusti B, Tattini L: Characterization of MinION nanopore data for resequencing analyses. *Briefings in bioinformatics* 2016.

24.   Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M: Improved data analysis for the MinION nanopore sequencer. *Nature methods* 2015, 12(4):351-356.

25.   Hargreaves AD, Mulley JF: Assessing the utility of the Oxford Nanopore MinION for snake venom gland cDNA sequencing. *PeerJ* 2015, 3:e1441.

26.   Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. *Genome biology* 2013, 14(5):R51.

27.   Manicassamy B, Manicassamy S, Belicha-Villanueva A, Pisanelli G, Pulendran B, Garcia-Sastre A: Analysis of in vivo dynamics of influenza virus infection in mice using a GFP reporter virus. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(25):11531-11536.

28.   Fukuyama S, Katsura H, Zhao D, Ozawa M, Ando T, Shoemaker JE, Ishikawa I, Yamada S, Neumann G, Watanabe S *et al*: Multi-spectral fluorescent reporter influenza viruses (Color-flu) as powerful tools for in vivo studies. *Nature communications* 2015, 6:6600.

29.   Ekblom R, Smeds L, Ellegren H: Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC genomics* 2014, 15:467.

30.   Guilliams M, Bruhns P, Saeys Y, Hammad H, Lambrecht BN: The function of Fcgamma receptors in dendritic cells and macrophages. *Nature reviews Immunology* 2014, 14(2):94-108.

31.   DiLillo DJ, Tan GS, Palese P, Ravetch JV: Broadly neutralizing hemagglutinin stalk-specific antibodies require FcgammaR interactions for protection against influenza virus in vivo. *Nature medicine* 2014, 20(2):143-151.

32.   Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, Vachieri SG, Pinna D, Minola A, Vanzetta F *et al*: A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 2011, 333(6044):850-856.

33.   DiLillo DJ, Palese P, Wilson PC, Ravetch JV: Broadly neutralizing anti-influenza antibodies require Fc receptor engagement for in vivo protection. *The Journal of clinical investigation* 2016, 126(2):605-610.

34.   Torres M, Casadevall A: The immunoglobulin constant region contributes to affinity and specificity. *Trends in immunology* 2008, 29(2):91-97.

35.   Tudor D, Yu H, Maupetit J, Drillet AS, Bouceba T, Schwartz-Cornil I, Lopalco L, Tuffery P, Bomsel M: Isotype modulates epitope specificity, affinity, and antiviral activities of anti-HIV-1 human broadly neutralizing 2F5 antibody. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(31):12680-12685.

36.   Chung AW, Crispin M, Pritchard L, Robinson H, Gorny MK, Yu X, Bailey-Kellogg C, Ackerman ME, Scanlan C, Zolla-Pazner S *et al*: Identification of antibody glycosylation structures that predict monoclonal antibody Fc-effector function. *AIDS* 2014, 28(17):2523-2530.

37.   Forthal DN, Gach JS, Landucci G, Jez J, Strasser R, Kunert R, Steinkellner H: Fc-glycosylation influences Fcgamma receptor binding and cell-mediated anti-HIV activity of monoclonal antibody 2G12. *J Immunol* 2010, 185(11):6876-6882.

38.   Herter S, Birk MC, Klein C, Gerdes C, Umana P, Bacac M: Glycoengineering of therapeutic antibodies enhances monocyte/macrophage-mediated phagocytosis and cytotoxicity. *J Immunol* 2014, 192(5):2252-2260.

39.   Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SH, Presta LG: Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *The Journal of biological chemistry* 2002, 277(30):26733-26740.

40.   Lee YN, Lee YT, Kim MC, Hwang HS, Lee JS, Kim KH, Kang SM: Fc receptor is not required for inducing antibodies but plays a critical role in conferring protection after influenza M2 vaccination. *Immunology* 2014, 143(2):300-309.

41.   El Bakkouri K, Descamps F, De Filette M, Smet A, Festjens E, Birkett A, Van Rooijen N, Verbeek S, Fiers W, Saelens X: Universal vaccine based on ectodomain of matrix protein 2 of influenza A: Fc receptors and alveolar macrophages mediate protection. *J Immunol* 2011, 186(2):1022-1031.

42.   Jegerlehner A, Schmitz N, Storni T, Bachmann MF: Influenza A vaccine based on the extracellular domain of M2: weak protection mediated via antibody-dependent NK cell activity. *J Immunol* 2004, 172(9):5598-5605.

43.   Tompkins SM, Zhao ZS, Lo CY, Misplon JA, Liu T, Ye Z, Hogan RJ, Wu Z, Benton KA, Tumpey TM *et al*: Matrix protein 2 vaccination and protection against influenza viruses, including subtype H5N1. *Emerging infectious diseases* 2007, 13(3):426-435.

44.   Wang R, Song A, Levin J, Dennis D, Zhang NJ, Yoshida H, Koriazova L, Madura L, Shapiro L, Matsumoto A *et al*: Therapeutic potential of a fully human monoclonal antibody against influenza A virus M2 protein. *Antiviral research* 2008, 80(2):168-177.

45.   Fu TM, Freed DC, Horton MS, Fan J, Citron MP, Joyce JG, Garsky VM, Casimiro DR, Zhao Q, Shiver JW *et al*: Characterizations of four monoclonal antibodies against M2 protein ectodomain of influenza A virus. *Virology* 2009, 385(1):218-226.

46.   Nimmerjahn F, Lux A, Albert H, Woigk M, Lehmann C, Dudziak D, Smith P, Ravetch JV: FcgammaRIV deletion reveals its central role for IgG2a and IgG2b activity in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(45):19396-19401.

47.   Song JM, Wang BZ, Park KM, Van Rooijen N, Quan FS, Kim MC, Jin HT, Pekosz A, Compans RW, Kang SM: Influenza virus-like particles containing M2 induce broadly cross protective immunity. *PloS one* 2011, 6(1):e14538.

48.   Robb L, Drinkwater CC, Metcalf D, Li R, Kontgen F, Nicola NA, Begley CG: Hematopoietic and lung abnormalities in mice with a null mutation of the common beta subunit of the receptors for granulocyte-macrophage colony-stimulating factor and interleukins 3 and 5. *Proceedings of the National Academy of Sciences of the United States of America* 1995, 92(21):9565-9569.

49.   Schneider C, Nobs SP, Heer AK, Kurrer M, Klinke G, van Rooijen N, Vogel J, Kopf M: Alveolar macrophages are essential for protection from respiratory failure and associated morbidity following influenza virus infection. *PLoS pathogens* 2014, 10(4):e1004053.

50.   Dranoff G, Crawford AD, Sadelain M, Ream B, Rashid A, Bronson RT, Dickersin GR, Bachurski CJ, Mark EL, Whitsett JA *et al*: Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis. *Science* 1994, 264(5159):713-716.

51.   Stanley E, Lieschke GJ, Grail D, Metcalf D, Hodgson G, Gall JA, Maher DW, Cebon J, Sinickas V, Dunn AR: Granulocyte/macrophage colony-stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology.

Proceedings of the National Academy of Sciences of the United States of America 1994, 91(12):5592-5596.

52. Happle C, Lachmann N, Skuljec J, Wetzke M, Ackermann M, Brennig S, Mucci A, Jirmo AC, Groos S, Mirenska A *et al*: Pulmonary transplantation of macrophage progenitors as effective and long-lasting therapy for hereditary pulmonary alveolar proteinosis. *Science translational medicine* 2014, 6(250):250ra113.

53. Suzuki T, Arumugam P, Sakagami T, Lachmann N, Chalk C, Sallese A, Abe S, Trapnell C, Carey B, Moritz T *et al*: Pulmonary macrophage transplantation therapy. *Nature* 2014, 514(7523):450-454.

54. Smith P, DiLillo DJ, Bournazos S, Li F, Ravetch JV: Mouse model recapitulating human Fcgamma receptor structural and functional diversity. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(16):6181-6186.

55. Fiers W, De Filette M, Birkett A, Neirynck S, Min Jou W: A "universal" human influenza A vaccine. *Virus research* 2004, 103(1-2):173-176.

56. Schmitz N, Beerli RR, Bauer M, Jegerlehner A, Dietmeier K, Maudrich M, Pumpens P, Saudan P, Bachmann MF: Universal vaccine against influenza virus: linking TLR signaling to anti-viral protection. *European journal of immunology* 2012, 42(4):863-869.

57. Ibanez LI, Roose K, De Filette M, Schotsaert M, De Sloovere J, Roels S, Pollard C, Schepens B, Grooten J, Fiers W *et al*: M2e-displaying virus-like particles with associated RNA promote T helper 1 type adaptive immunity against influenza A. *PloS one* 2013, 8(3):e59081.

58. Wang BZ, Gill HS, He C, Ou C, Wang L, Wang YC, Feng H, Zhang H, Prausnitz MR, Compans RW: Microneedle delivery of an M2e-TLR5 ligand fusion protein to skin confers broadly cross-protective influenza immunity. *Journal of controlled release : official journal of the Controlled Release Society* 2014, 178:1-7.

59. Coffman RL, Sher A, Seder RA: Vaccine adjuvants: putting innate immunity to work. *Immunity* 2010, 33(4):492-503.

60. Mozdzanowska K, Zharikova D, Cudic M, Otvos L, Gerhard W: Roles of adjuvant and route of vaccination in antibody response and protection engendered by a synthetic matrix protein 2-based influenza A virus vaccine in the mouse. *Virology journal* 2007, 4:118.

61. Li C, Hatta M, Burke DF, Ping J, Zhang Y, Ozawa M, Taft AS, Das SC, Hanson AP, Song J *et al*: Selection of antigenically advanced variants of seasonal influenza viruses. *Nature microbiology* 2016, 1(6):16058.

62. Das SR, Hensley SE, Ince WL, Brooke CB, Subba A, Delboy MG, Russ G, Gibbs JS, Bennink JR, Yewdell JW: Defining influenza A virus hemagglutinin antigenic drift by sequential monoclonal antibody selection. *Cell host & microbe* 2013, 13(3):314-323.

63. Gerhard W, Yewdell J, Frankel ME, Webster R: Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* 1981, 290(5808):713-717.

64. Gerhard W, Mozdzanowska K, Zharikova D: Prospects for universal influenza virus vaccine. *Emerging infectious diseases* 2006, 12(4):569-574.

65. Cho KJ, Schepens B, Seok JH, Kim S, Roose K, Lee JH, Gallardo R, Van Hamme E, Schymkowitz J, Rousseau F *et al*: Structure of the extracellular domain of matrix protein 2 of influenza A virus in complex with a protective monoclonal antibody. *Journal of virology* 2015, 89(7):3700-3711.

66. Wolf AI, Mozdzanowska K, Williams KL, Singer D, Richter M, Hoffmann R, Caton AJ, Otvos L, Erikson J: Vaccination with M2e-based multiple antigenic peptides: characterization of the B cell response and protection efficacy in inbred and outbred mice. *PloS one* 2011, 6(12):e28445.

67. Kim MC, Song JM, O E, Kwon YM, Lee YJ, Compans RW, Kang SM: Virus-like particles containing multiple M2 extracellular domains confer improved cross-protection against

various subtypes of influenza virus. *Molecular therapy : the journal of the American Society of Gene Therapy* 2013, 21(2):485-492.

68. Ramos EL, Mitcham JL, Koller TD, Bonavia A, Usner DW, Balaratnam G, Fredlund P, Swiderek KM: Efficacy and safety of treatment with an anti-m2e monoclonal antibody in experimental human influenza. *The Journal of infectious diseases* 2015, 211(7):1038-1044.

69. Ito T, Gorman OT, Kawaoka Y, Bean WJ, Webster RG: Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins. *Journal of virology* 1991, 65(10):5491-5498.

70. Hutchinson EC, Curran MD, Read EK, Gog JR, Digard P: Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. *Journal of virology* 2008, 82(23):11869-11879.

71. Gog JR, Afonso Edos S, Dalton RM, Leclercq I, Tiley L, Elton D, von Kirchbach JC, Naffakh N, Escriou N, Digard P: Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic acids research* 2007, 35(6):1897-1907.

72. Ozawa M, Maeda J, Iwatsuki-Horimoto K, Watanabe S, Goto H, Horimoto T, Kawaoka Y: Nucleotide sequence requirements at the 5' end of the influenza A virus M RNA segment for efficient virus replication. *Journal of virology* 2009, 83(7):3384-3388.

73. Llompart CM, Nieto A, Rodriguez-Frandsen A: Specific residues of PB2 and PA influenza virus polymerase subunits confer the ability for RNA polymerase II degradation and virus pathogenicity in mice. *Journal of virology* 2014, 88(6):3455-3463.

74. Rolling T, Koerner I, Zimmermann P, Holz K, Haller O, Staeheli P, Kochs G: Adaptive mutations resulting in enhanced polymerase activity contribute to high virulence of influenza A virus in mice. *Journal of virology* 2009, 83(13):6673-6680.

75. Shih SR, Nemeroff ME, Krug RM: The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. *Proceedings of the National Academy of Sciences of the United States of America* 1995, 92(14):6324-6328.

76. Bier K, York A, Fodor E: Cellular cap-binding proteins associate with influenza virus mRNAs. *The Journal of general virology* 2011, 92(Pt 7):1627-1634.

77. Zharikova D, Mozdzanowska K, Feng J, Zhang M, Gerhard W: Influenza type A virus escape mutants emerge in vivo in the presence of antibodies to the ectodomain of matrix protein 2. *Journal of virology* 2005, 79(11):6644-6654.

78. Zhang M, Zharikova D, Mozdzanowska K, Otvos L, Gerhard W: Fine specificity and sequence of antibodies directed against the ectodomain of matrix protein 2 of influenza A virus. *Molecular immunology* 2006, 43(14):2195-2206.

79. Belser JA, Katz JM, Tumpey TM: The ferret as a model organism to study influenza A virus infection. *Disease Models & Mechanisms* 2011, 4(5):575-579.

80. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, Munster VJ, Sorrell EM, Bestebroer TM, Burke DF, Smith DJ *et al*: Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* 2012, 336(6088):1534-1541.

81. Herlocher ML, Elias S, Truscon R, Harrison S, Mindell D, Simon C, Monto AS: Ferrets as a transmission model for influenza: Sequence changes in HA1 of type A (H3N2) virus. *Journal of Infectious Diseases* 2001, 184(5):542-546.

82. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, Zhong G, Hanson A, Katsura H, Watanabe S *et al*: Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* 2012, 486(7403):420-428.

83. Smith W, Andrewes CH, Laidlaw PP: A virus obtained from influenza patients. *The Lancet* 1933, 222:66-68.

84. Nachbagauer R, Miller MS, Hai R, Ryder AB, Rose JK, Palese P, Garcia-Sastre A, Krammer F, Albrecht RA: Hemagglutinin Stalk Immunity Reduces Influenza Virus Replication and Transmission in Ferrets. *Journal of virology* 2016, 90(6):3268-3273.

85.   Edenborough KM, Gilbertson BP, Brown LE: A mouse model for the study of contact-dependent transmission of influenza A virus and the factors that govern transmissibility. *Journal of virology* 2012, 86(23):12544-12551.

86.   Price GE, Lo CY, Misplon JA, Epstein SL: Mucosal immunization with a candidate universal influenza vaccine reduces virus transmission in a mouse model. *Journal of virology* 2014, 88(11):6019-6030.

87.   Schulman JL: Experimental transmission of influenza virus infection in mice. IV. Relationship of transmissibility of different strains of virus and recovery of airborne virus in the environment of infector mice. *The Journal of experimental medicine* 1967, 125(3):479-488.

Part V: Addendum

Single-Domain Antibodies Targeting Neuraminidase Protect Against An H5N1 Influenza Virus Challenge

Francisco Miguel Cardoso[a,b], Lorena Itatí Ibañez[a,b*], <u>Silvie Van den Hoecke</u>[a,b], Sarah De Baets[a,b], Anouk Smet[a,b], Kenny Roose[a,b], Bert Schepens[a,b], Francis J. Descamps[a,b*], Walter Fiers[a,b], Serge Muyldermans[c,f], Ann Depicker[d,e] and Xavier Saelens[a,b].

[a] VIB Inflammation Research Center, Ghent, Belgium

[b] Department for Biomedical Molecular Biology, Ghent University, Ghent, Belgium

[c] Structural Biology Research Center, VIB, Brussels, Belgium

[d] Department of Plant Systems Biology, VIB, Ghent, Belgium

[e] Department of Biotechnology and Bioinformatics, Ghent, Belgium

[f] Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium

Contribution of Silvie Van den Hoecke:

Selection and characterization of H5N1 N1-VHHm escape mutant viruses (Figure 8).

Single-Domain Antibodies Targeting Neuraminidase Protect against an H5N1 Influenza Virus Challenge

Francisco Miguel Cardoso,[a,b] Lorena Itatí Ibañez,[a,b*] Silvie Van den Hoecke,[a,b] Sarah De Baets,[a,b] Anouk Smet,[a,b] Kenny Roose,[a,b] Bert Schepens,[a,b] Francis J. Descamps,[a,b*] Walter Fiers,[a,b] Serge Muyldermans,[c,f] Ann Depicker,[d,e] Xavier   Saelens[a,b]

VIB Inflammation Research Center, Ghent, Belgium[a]; Department for Biomedical Molecular Biology, Ghent University, Ghent, Belgium[b]; Structural Biology Research Center, VIB, Brussels, Belgium[c]; Department of Plant Systems Biology, VIB, Ghent, Belgium[d]; Department of Biotechnology and Bioinformatics, Ghent, Belgium[e]; Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium[f]

ABSTRACT

Influenza virus neuraminidase (NA) is an interesting target of small-molecule antiviral drugs. We isolated a set of H5N1 NA-specific single-domain antibodies (N1-VHHm) and evaluated their *in vitro* and *in vivo* antiviral potential. Two of them inhibited the NA activity and *in vitro* replication of clade 1 and 2 H5N1 viruses. We then generated bivalent derivatives of N1-VHHm by two methods. First, we made N1-VHHb by genetically joining two N1-VHHm moieties with a flexible linker. Second, bivalent N1-VHH-Fc proteins were obtained by genetic fusion of the N1-VHHm moiety with the crystallizable region of mouse IgG2a (Fc). The *in vitro* antiviral potency against H5N1 of both bivalent N1-VHHb formats was 30- to 240-fold higher than that of their monovalent counterparts, with 50% inhibitory concentrations in the low nanomolar range. Moreover, single-dose prophylactic treatment with bivalent N1-VHHb or N1-VHH-Fc protected BALB/c mice against a lethal challenge with H5N1 virus, including an oseltamivir-resistant H5N1 variant. Surprisingly, an N1-VHH-Fc fusion without *in vitro* NA-inhibitory or antiviral activity also protected mice against an H5N1 challenge. Virus escape selection experiments indicated that one amino acid residue close to the catalytic site is required for N1-VHHm binding. We conclude that single-domain antibodies directed against influenza virus NA protect against H5N1 virus infection, and when engineered with a conventional Fc domain, they can do so in the absence of detectable NA-inhibitory activity.

IMPORTANCE

Highly pathogenic H5N1 viruses are a zoonotic threat. Outbreaks of avian influenza caused by these viruses occur in many parts of the world and are associated with tremendous economic loss, and these viruses can cause very severe disease in humans. In such cases, small-molecule inhibitors of the viral NA are among the few treatment options for patients. However, treatment with such drugs often results in the emergence of resistant viruses. Here we show that single-domain antibody fragments that are specific for NA can bind and inhibit H5N1 viruses *in vitro* and can protect laboratory mice against a challenge with an H5N1 virus, including an oseltamivir-resistant virus. In addition, plant-produced VHH fused to a conventional Fc domain can protect *in vivo* even in the absence of NA-inhibitory activity. Thus, NA of influenza virus can be effectively targeted by single-domain antibody fragments, which are amenable to further engineering.

Zoonotic influenza A virus infections are a persistent threat because of their pandemic potential. In particular, highly pathogenic avian influenza viruses (HPAIV) of the H5N1, H7N1, and H7N7 subtypes occasionally cross the species barrier between domesticated birds and humans. These viruses could become transmissible between humans through reassortment with circulating swine or human influenza viruses or by gradually accumulating mutations (1, 2). In the last decade, zoonotic outbreaks have had a major effect on public health. HPAIV H5N1 (3), the swine influenza (H1N1) outbreak in 2009 (4), and more recently, human infections with H7N9 in southern Asia (5) illustrate our poor preparedness for pandemic influenza (6). HPAIV H5N1 infection in humans has a confirmed case fatality rate of approximately 60%. The high pathogenicity of HPAIV H5N1 in humans can be attributed to a high replication rate and a broad cellular tropism that can lead to systemic virus spread. In addition, deregulated induction of proinflammatory cytokines and chemokines ("cytokine storm") is associated with severe HPAIV H5N1 infections and can result in a disproportionate immunological response (7).

Influenza virus neuraminidase (NA) is a homotetrameric type II membrane glycoprotein with sialidase activity. The NA catalytic site is located at the top of each monomer, opposite the tetramer interface. NA plays an essential role in the spread of influenza viruses by cleaving sialic acids from the host cell receptors and from virions. NA activity also contributes to virus entry by cleaving decoy receptors present in mucins that line the layer of respiratory epithelial cells (8). Immunologically, NA is the second major humoral antigenic determinant (after hemagglutinin [HA]) and is subject to antigenic drift and occasional shift. In addition, experimental influenza vaccines supplemented with NA have im-

258

A GFP Expressing Influenza A Virus to Report In Vivo Tropism and Protection by a Matrix Protein 2 Ectodomain-Specific Monoclonal Antibody

Sarah De Baets[1,2,3☺], Judith Verhelst[1,2☺], <u>Silvie Van den Hoecke</u>[1,2], Anouk Smet[1,2], Michael Schotsaert[1,2,3], Emma R. Job[1,2], Kenny Roose[1,2], Bert Schepens[1,2], Walter Fiers[1,2], Xavier Saelens[1,2*]

[1] Department of Medical Protein Research, VIB, Ghent, Belgium

[2] Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

[3] Current address: Labo Medische Analyse, Centrum voor Radio-Immunologie, Ghent, Belgium

[4] Current address: Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, United States of America

☺ These authors contributed equally to this work.

* corresponding author: xavier.saelens@irc.vib-ugent.be

Contribution of Silvie Van den Hoecke:

Deep sequencing analysis of PR8 and PR8-NS1(1–73)GFP virus (Figure 4 and Table 1).

PLOS ONE

RESEARCH ARTICLE

A GFP Expressing Influenza A Virus to Report *In Vivo* Tropism and Protection by a Matrix Protein 2 Ectodomain-Specific Monoclonal Antibody

Sarah De Baets[1,2☯¤a], Judith Verhelst[1,2☯], Silvie Van den Hoecke[1,2], Anouk Smet[1,2], Michael Schotsaert[1,2¤b], Emma R. Job[1,2], Kenny Roose[1,2], Bert Schepens[1,2], Walter Fiers[1,2], Xavier Saelens[1,2]*

1 Department of Medical Protein Research, VIB, Ghent, Belgium, 2 Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

¤a Current address: Labo Medische Analyse, Centrum voor Radio-Immunologie, Ghent, Belgium
¤b Current address: Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, United States of America
☯ These authors contributed equally to this work.
* xavier.saelens@irc.vib-ugent.be

Abstract

The severity of influenza-related illness is mediated by many factors, including *in vivo* cell tropism, timing and magnitude of the immune response, and presence of pre-existing immunity. A direct way to study cell tropism and virus spread *in vivo* is with an influenza virus expressing a reporter gene. However, reporter gene-expressing influenza viruses are often attenuated *in vivo* and may be genetically unstable. Here, we describe the generation of an influenza A virus expressing GFP from a tri-cistronic NS segment. To reduce the size of this engineered gene segment, we used a truncated NS1 protein of 73 amino acids combined with a heterologous dimerization domain to increase protein stability. GFP and nuclear export protein coding information were fused in frame with the truncated NS1 open reading frame and separated from each other by 2A self-processing sites. The resulting PR8-NS1 (1–73)GFP virus was successfully rescued and replicated as efficiently as the parental PR8 virus *in vitro* and was slightly attenuated *in vivo*. Flow cytometry-based monitoring of cells isolated from PR8-NS1(1–73)GFP virus infected BALB/c mice revealed that GFP expression peaked on day two in all cell types tested. In particular respiratory epithelial cells and myeloid cells known to be involved in antigen presentation, including dendritic cells (CD11c[+]) and inflammatory monocytes (CD11b[+] GR1[+]), became GFP positive following infection. Prophylactic treatment with anti-M2e monoclonal antibody or oseltamivir reduced GFP expression in all cell types studied, demonstrating the usefulness of this reporter virus to analyze the efficacy of antiviral treatments *in vivo*. Finally, deep sequencing analysis, serial *in vitro* passages and *ex vivo* analysis of PR8-NS1(1–73)GFP virus, indicate that this virus is genetically and phenotypically stable.

260

CURRICULUM VITAE

- ## PERSONAL INFORMATION

Name	SILVIE VAN DEN HOECKE
Personal address	KERKSTRAAT 93, 9060 ZELZATE, BELGIUM
Phone	+32 472 438 736
E-mail	silvievandenhoecke@gmail.com
Date of Birth	October 18, 1988
Place of Birth	Ghent, Belgium
Nationality	Belgian

- ## EDUCATION

2011 - 2016   **PhD in Biochemistry**

Doctoral thesis: Deep sequencing of influenza A viruses tempered by antibodies against the M2 ectodomain

Medical Biotechnology Center, VIB, Ghent, Belgium.

Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

2011   **Certificate laboratory animal science FELASA category B and C**

Ghent University, Ghent

2009 - 2011   **Master in Biochemistry and Biotechnology**

Major option: Biomedical Biotechnology, Minor option: Microbial biotechnology

Masters thesis: 'Nanobodies against influenza NA(H5N1): Antiviral, molecular and functional characterization'

Ghent University, Ghent

2006 - 2009   **Bachelor in Biochemistry and Biotechnology**

Ghent University, Ghent

2000 - 2006   **Science and Mathematics**

Sint-Laurensinstituut ASO, Zelzate

- **PUBLICATIONS IN SCIENTIFIC JOURNALS**

  <u>Van den Hoecke S</u>, Verhelst J, Saelens X (2016) Illumina MiSeq sequencing disfavours a sequence motif in the GFP reporter gene. Scientific reports, 6:26314.

  <u>Van den Hoecke S</u>, Verhelst J, Vuylsteke M and Saelens X (2015) Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. BMC genomics 16: 79. [Highly accessed]

  De Baets S, Verhelst J, <u>Van den Hoecke S</u>, Smet A, Schotsaert M, *et al*. (2015) A GFP expressing influenza A virus to report *in vivo* tropism and protection by a matrix protein 2 ectodomain-specific monoclonal antibody. PloS one 10: e0121491.

  Cardoso FM, Ibanez LI, <u>Van den Hoecke S</u>, De Baets S, Smet A, *et al.* (2014) Single-domain antibodies targeting neuraminidase protect against an H5N1 influenza virus challenge. Journal of virology 88: 8278- 8296.

  Manuscripts in preparation:

  <u>Van den Hoecke S</u>, Ehrhardt K, Kolpe A, El Bakkouri K, Deng L, Grootaert H, Schoonooghe S, Smet A, Bentahir M, Roose K, Schotsaert M, Schepens B, Callewaert N, Nimmerjahn F, Staeheli P, Hengel H, and Saelens X. Hierarchical and redundant roles of activating FcγRs in protection against influenza disease by M2e-specific IgG1 and IgG2a antibodies. (under revisions for Journal of Virology)

  Carlos Mancera Gracia JC, <u>Van den Hoecke S</u>, Saelens X, Van Reeth K. Effect of serial pig passages on the adaptation of an avian H9N2 influenza virus to swine. (Submitted to PLoS ONE)

- **TEACHING ACTIVITIES**

2015	Guidance Master project (Jullie Van Coillie) 'Analysis of influenza A viruses escaping anti-M2e immune pressure'
2014	Guidance Master project (Lana Hellebaut) 'Analyse van de stabiliteit van het NS-GFP segment in het influenza GFP quasispecies'
2013	Guidance bachelor projects
2011 – 2012	Guidance practical course for Moleculaire Biologie 1 (Prokaryoten), UGent.
Others	Dag Van De Biotechnology (2012), Wetenschap In De Kijker (2013), Wetenschap Op Stap (2013)

- **PARTICIPATION (INTER)NATIONAL CONFERENCES**

  - **ORAL PRESENTATIONS**

    Successful Planning of Large Data Generating Experiments
    16 March 2016, Leuven, Belgium

    Annual Belvir Meeting
    18 December 2015, Brussels, Belgium

    3th European Seminar in Virology (EuSeV): (Re)-Emerging Viruses
    20th June 2015, Bertinoro, Italy

    MBC Lunch Seminar
    17 June 2015, Ghent, Belgium

    Illumina Technical User Experience Day
    4th June 2015, Brussels

    DMBR Pecha Kucha Evening
    25 April 2013, Ghent, Belgium

  - **POSTER PRESENTATIONS**

    Departmental Evaluation Board of Medical Biotechnology Center
    3 November 2015, Ghent, Belgium

    VIB Seminar 2015
    30 March 2015- 1 April 2015, Blankenberge, Belgium

    VIB: Revolutionizing Next-Generation Sequencing: Tools and Technologies
    15-16 January 2015, Leuven

    BSM: Cell Signaling in Host-Microbe Interactions
    18th November 2014, Brussels

    4th International Influenza Meeting
    21-23 September 2014, Münster, Germany

    VIB seminar 2014
    28-30 April 2014, Blankenberge, Belgium

    UGent doctoraatssymposium 2014
    20 March 2014, Ghent, Belgium

    BSM: Microbial Diversity for Science and Industry
    26-27 November 2013, Brussels, Belgium

NSV Meeting 2013
16-21 June 2013, Granada, Spain

- **OTHERS**

    Annual Belvir Meeting
    8 December 2014, Brussels, Belgium

    IRC Predoc Symposium
    20 June 2014, Ghent, Belgium

    Annual Belvir Meeting
    8 November 2013, Brussels, Belgium

    VIB Seminar
    6-8 February 2013, Blankenberge, Belgium

    Inflammation and vaccination seminar
    19 September 2012, Ugent Faculty of Veterinary Medicine, Merelbeke, Belgium

    3rd Internation Influenza Meeting
    2-4 September, Münster, Germany

    VIB Seminar
    18-20 April 2012, Blankenberge, Belgium

    VIB-BGI Genomics Meeting
    15 February 2012, Leuven, Belgium

- **COURSES AND TRAININGS FOLLOWED**

2016   Successful Planning of Large Data Generating Experiments
            VIB - Leuven, Belgium

2015   Molecular Epidemiology of Infectious diseases
            VIB - Antwerp, Belgium

            Career Guidance for PhDs and Postdocs
            VIB - Ghent, Belgium

2013   Introductory statistics in Graphpad Prism
            VIB - Ghent, Belgium

            Introduction to the analysis of NGS data
            VIB - Leuven, Belgium

Hands-on introduction to NGS variant analysis

VIB - Leuven, Belgium

2012   Introduction to CLC Main Workbench

VIB - Ghent, Belgium

Next Generation Sequencing Workshop

VIB - Leuven, Belgium

Dankwoord

Met een glimlach op mijn gezicht rond ik vijf jaar doctoreren met deze doctoraatsthesis af. Dit was echter niet mogelijk geweest zonder de hulp en steun van mijn vele collega's, vrienden en familie, daarom aan iedereen die op de één of andere manier bijgedragen heeft aan dit werk: Dankjewel! Dat ik met trots het werk in dit boekje kan voorstellen, heb ik ook aan jullie te danken.

Eerst en vooral, Xavier, bedankt om na mijn masterthesis nog steeds in mij te geloven en me de kans te geven om te doctoreren in jouw onderzoeksgroep. Gedurende die bijna zes jaar heb ik zeer veel van jou en jouw kritisch denken geleerd. Bedankt om me in deze periode de vrijheid te geven om me te laten groeien als wetenschapster en me hierbij te sturen waar nodig. Het was ook een fijn gevoel om te weten dat jouw deur altijd openstond voor zowel positieve als negatieve resultaten. Ook bedankt om me steeds te steunen en te ondersteunen in de experimentele set-up van mijn dure sequencing experimenten, want zoals je het zo mooi verwoorde: 'Je kan geen lekkere omelet maken zonder eieren te kopen'. Daarnaast wil jou ook bedanken voor het steeds kritisch lezen en verbeteren van mijn manuscripten en thesis, wat resulteerde in dit mooi eindresultaat. Prof. Fiers, bedankt voor uw kritische blik op mijn werk tijdens de labvergaderingen. Ik bewonder jouw blijvende interesse in de wetenschap.

I also would like to thank the members of my examination board, Prof. Paul Digard, Prof. Johan Grooten, Prof. Martin Guilliams, Prof. Philippe Lemey and Prof. Kristien Van Reeth, for taking the time to critically read my PhD thesis and to challenge me with interesting questions during my internal PhD defense. Prof. Paul Digard and Prof. Kristien Van Reeth, it was a pity that both of you could not make it for my internal defense, but hopefully you can be present on my public PhD defense.

Ik zou ook graag VIB Tech Watch bedanken voor de financiële steun waardoor het mogelijk werd om de geschiktheid van de Illumina MiSeq met die van de Ion Torrent PGM voor het bepalen van varianten in een influenza viruspopulatie te vergelijken. Daarnaast wil ik graag ook de vele collega's buiten het labo bedanken voor hun inbreng in dit project. Deze samenwerkingen speelden een belangrijke rol bij het behalen van de resultaten in deze thesis. Hiernaast wil ik ook Marnik Vuylsteke bedanken voor de vele, verschillende statistische analyses op mijn data. Tenslotte wil ik Liesbeth Martens en Paco Hulpiau bedanken voor het uitvoeren van hun analyses naar het uitbreiden van het 'CCNGCC' motief.

Ik had ook het geluk om mijn doctoraat te mogen uitvoeren omringd door fijne collega's, daarom wil ik graag ook iedereen van het labo bedanken voor de vele hulp, gezellige middagpauzes en babbeltjes tussendoor. De fijne tijd die jullie me bezorgden, zal me steeds bijblijven. Koen, ik herinner me nog goed hoe we allebei gefascineerd waren door microbiologie en daardoor dolgraag in het labo van Xavier een masterthesis wilden lopen. Op kousevoetjes gingen we daarom Xavier zijn

bureau binnen en zonder probleem werd er een tweede thesisonderwerp gecreëerd. Tijdens deze thesis nam onze interesse in virussen alleen maar toe en hadden we beiden het geluk dat we ook een doctoraat mochten starten. Ik wens jou ook alle succes bij het afronden van je thesis, nog even doorbijten en ook Dr. Sedeyn is een feit! Iebe, kersverse Dr. Rossey, het was fijn dat ik samen met jou mijn doctoraat kon afronden. Ik bewonder jouw rust en kalmte in elke situatie, net als jouw geografisch talent. Zonder jou liep ik nu misschien nog steeds verloren in Bertinoro. Onze zoektocht naar de shuttlebus was alvast een sportieve afsluiter van onze fijne tijd tijdens het congres in Noord-Italïe. Dorien, de gedreven sportvrouw van het labo met een voorliefde voor lama 'Winter' en efficiënte mollenvanger. Dankjewel dat ik steeds met mijn M2e of FcγR gerelateerde vragen bij jou terecht kon. Iebe en Dorien, het was ook zeer fijn om samen met jullie de Warmathon te lopen. Oorspronkelijk gingen we enkel een paar rondjes lopen, maar door elkaar te motiveren, verbraken we alledrie onze maximale loopafstand. Bert, jouw wetenschappelijke interesse lijkt onuitputtelijk en werkt aanstekelijk. Naast jouw rol als mentor voor het hele RSV-team, denk je ook mee met alle andere projecten in het labo. Bedankt voor jouw vele ideeën en jouw inbreng in mijn thesis. Soraya, de nieuwste aanwinst van het RSV-team, het is fijn dat je na je stage nu ook deel mag uitmaken van ons labo. Ik wens je veel succes met het verdedigen van je masterthesis! Anouk, Kenny and Emma, the 'Holy' Trinity for late lunch and cava parties. Anouk, bedankt voor de vele miljoenen cellen die je me steeds zonder problemen bezorgde en om het labo te voorzien van koffie en koekjes, waardoor ochtendhumeur geen kans krijgt in onze groep. Emma and Kenny, thank you so much for reading the introduction of my thesis and providing me with your very useful comments, which were highly appreciated and increased the quality of the thesis. I also want to thank you for always taking the time to answer all my (influenza-related) questions. Ioanna, I wish you all the best with the finalizing of your project and the thesis writing next year! Goedlachse, lieve Tine, door de zwangerschap en geboorte van jouw twee mannen heb ik jou helaas eventjes in het labo moeten missen, maar het is mooi om te zien hoe jouw mannen jou doen stralen en ik kijk ernaar uit om jou terug in het labo te zien, zodat we onze kleine babbeltjes tussendoor terug kunnen hervatten. Annasaheb, thank you for your help with the mice experiments and the several chats we had in the P2. I wish you and your wife all the best! Jan, dankjewel voor jouw hulp met mijn flow cytometry experimenten, waarbij je niet enkel helpt maar ook steeds de tijd neemt om je kennis hieromtrent te delen door rustig en duidelijk uit te leggen welke settings je voor welke reden toepast. Daarnaast zorgt jouw vrolijk gefluit of tafelgetrommel steeds voor een muzikale noot in het labo. Lien, plots maakte jij deel uit van ons labo en wat zijn we blij met jou in ons team! Jouw talent voor multitasken heb je met het succesvol afronden van verschillende projecten alvast bewezen. Misschien is er na je vertolking van Iebe en de Jove-reportage nog een carrière in de filmindustrie voor jou weggelegd? Eline, de nieuwste aanwinst van het Mx-team, jouw doorzettingsvermogen siert je, ik wens je veel succes met je doctoraat! Sella, thank you for your help with the figures of the crystal structures of M2e for my thesis. I hope you will recover very soon from your back pain.

Naast de vele collega's, wil ik ook mijn ex-collegas - Judith, Sarah, Karim, Michael, Miguel, Farzaneh, Florencia, Ki Joon en Liesbeth - bedanken. Judith en Sara, jullie zijn nu beiden elk al een tijdje weg uit het lab, maar het is echt fijn dat jullie nu en dan nog eens langskomen op het VIB om samen te lunchen. Ik wens jullie veel succes met jullie gezinnetje en jullie carrières. Miguel, thank you for guiding me during my master thesis and introducing me in the virology field. It was never boring with you in the lab, thank you for all these fun moments. Michael, op-en-top man met enkele vrouwelijke interesses, jouw wetenschappelijke en niet-wetenschappelijke input in het labo wordt nog steeds gemist. Ik wens je nog veel succes in New York! Lei, my Chinese friend, thank you for helping me with my experiments and it was also very nice getting to know you and your family. Thank you for inviting us for a typical Chinese dinner, we really enjoyed it. I wish you and your family all the best in Atlanta! Liesbeth, het is prettig om te horen dat je het goed stelt op je nieuwe job en geloof gerust in jezelf want de functie van projectleider kan je zeker aan. Ik wens je ook veel succes met de bouw van jullie droomhuis.

Ik zou ook graag alle collega's en vrienden binnen het VIB bedanken om dit zo een aangename en inspirerende werkplek te maken. Binnen de vier muren van onze bureaumodule brengt dit ons naar de 'Callewaert' unit en verder in de wandelgangen wil ik ook alle onderzoeksgroepen en ondersteunend personeel bedanken voor hun bijdrage aan mijn project. In het bijzonder wil ik hierbij de 'biochemietjes' bedanken. Tien jaar geleden startten we allen aan het biochemie-avontuur, werden we vrienden en vervolgens ook collega's. Onze vaste afspraak op de eerste dinsdag van de maand is steeds iets om naar uit te kijken. Ik wens jullie ook veel succes bij het afronden van jullie doctoraat! Next to this I also want to thank the UPVA-URBE party squad for the many after work chats and drinks.

Laura, Alice, Behrouz and Ans although the thesis writing was sometimes hard, it was nice to share this experience with you. I wish you all the best with your PhD defenses in the coming weeks! We did it! Morgane, good luck with defending your PhD thesis, you will do great! I wish you and Yves all the best with your international careers!

Net als de hulp binnen het labo, speelde ook de steun en begrip van mijn vrienden buiten het labo een belangrijke rol bij het behalen van dit doctoraat. Daarom wil ik ook graag al mijn vrienden bedanken, voor de fijne tijd die jullie me steeds bezorgen. Mieke DB, Pieter, Eleni, Ineke, Mieke V, Lisa, Adriaan, Lien, Jolien, Lana, Jasper, Leen, Maarten, Nathalie, Niels, Stein, Cedrick, Sandy en Bjorn, jullie ben ik extra knuffels verschuldigd!

Mama en papa, zonder de kansen die jullie mij gegeven hebben had ik hier vandaag niet gestaan. Daarom ben ik jullie énorm dankbaar om steeds in mij te blijven geloven en me steunen in alles wat ik doe. Ik kon me geen betere ouders wensen, jullie zijn top! Ook bedankt aan mijn zusjes (en hun partners), om deel uit te maken van dit warme nest, waar we één voor één aan het uitvliegen zijn. Ik

wens jullie beiden het allerbeste in jullie nieuw (t)huis. Hiernaast heb ik ook het geluk deel uit te maken van een fantastische plus-familie, Marc, Rina, Tim en Emmely: Bedankt!

Giel, dankjewel om van mij het gelukkigste meisje ter wereld te maken en steeds weer het beste in mij naar boven te halen, zowel binnen als buiten het labo. Wij zijn een top team! Ik zie je zo graag.

Bedankt!
Silvie