Transcriptional analysis of *chIFITM* knock-out cell lines

Martina Hadrovic Sinthalapadi

St Catherine College

Thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2022

Abstract

Interferon-inducible transmembrane (IFITM) proteins are restriction factors known to confer antiviral resistance against a wide range of viral pathogens. The chicken *IFITM* locus is found on *Gallus gallus* chromosome 5. It contains 4 genes of *chIFITM* family: *chIFITM1, 2, 3, and 5*. They have been shown to restrict avian influenza virus *in vitro* and significantly ameliorate clinical signs in animals *in vivo.* Here I hypothesise that chIFITM knock-out models could support higher viral titres due to the ablation of *chIFITM* genes. I develop two models: *chIFITM* KO DF-1 and *chIFITM* KO PGC. *chIFITM* KO DF-1 fibroblasts representing ubiquitous avian fibroblast line used in veterinary research and *chIFITM* KO PGC modelling an approximation to an animal model. chPGCs are chicken primordial germ cells: precursors of sperm and ova which are very amenable to gene editing and for that reason frequently used for production of transgenic animals in research. Subsequent to gene editing they were differentiated into fibroblasts to facilitate comparison with the DF-1 fibroblasts. The study focussed on the investigation of the effects of the *chIFITM* KO on the permissiveness of the cells to the influenza virus and their ability to support viral replication. Differential expression analysis and the subsequent pathway analysis was conducted on the differentially expressed genes in native conditions as well as during the influenza infection for both models to compare and contrast differentially regulated pathways for each model.

Declaration

I declare that the work presented in this D.Phil thesis entitled, “Transcriptional analysis of chIFITM knockout cell lines” is entirely my own except where the contributions of collaborators have been clearly acknowledged.. No part of this thesis has been submitted for any degree of other qualification at the University of Oxford or elsewhere. My research has been conducted at the Pirbright Institute under the supervision of Prof John Hammond.

Martina Hadrovic Sinthalapadi

Michaelmas Term 2022

Acknowledgements

Firstly, I want to thank Dr Mark Fife who took a bet on me and made this PhD happen for me despite the hurdles. I also want to thank Dr Jonathan Moore for believing in me and truly being a mentor in every sense of the word. Thank you. I am very grateful to Prof Adrian Smith for giving me this opportunity, for always being available with advice and never thinking twice about how he can be of help. Thank you. Lastly, to Prof John Hammond for accepting the challenge of taking the group on, for all the technically sharp scientific advice and critical assessment of the work, for his patience and flexibility, and most of all for his optimistic attitude when times were dire. It was truly a gift to be a part of your group.

I would also like to acknowledge BBSRC, The Pirbright Institute and the University of Oxford for the provision of financial support for this project, and Horizon Discovery and Dr Michael McGrew at the Roslin Institute for the cells.

I want to express my deepest gratitude to (future Dr) Abigail Hay – there isn't an area where you did not extend a helping hand: the lab and the science, the paperwork and logistics of our work, the social life of the group. Your capabilities will never cease to astound me, from scientific thinking to emotional acuity. It is an honour to have met you. Thank you to Dr Jean-Rémy Sadeyen and Dr Carina Conceição for always helping without ever wanting anything in return. I was touched. To Dr Andreas Alber for all his advice and resources he selflessly shared. To Dr Richard Borne for his code. It was the skeleton upon which I built half of my bioinformatic work. To Dr Chandana Tennakoon for always helping me to make the cluster obey and for being light-hearted on occasions when I "broke" it. To Dr Daniel Dorey-Robinson and Dr John Schwartz for debugging the code and sharing tips and tricks. To Dr Graham Freimanis for teaching me everything I know about sequencing. Thank you to all the guys for always making me feel included in the Bioinformatics group. Thank you to Dr Andrew Broadbent for all the scientific discussions, encouragement, and support; it meant more than you know. I will be eternally grateful to Dr Archana Jadhav and Dr Veronica Martini for being truly amazing friends, through thick and thin, for life. Thank you to Kevan Hanson for sincerely caring as a human, for all the cake and for all the technical advice. You are truly a treasure chest of old school deep knowledge of molecular biology techniques.

I will be forever indebted to Dr Michael Love, Dr Joshua Starmer, Dr Kevin Blighe, Dr Devon Ryan, Dr Simon Andrews, Dr Gordon Smyth, Dr Ian Sudbery, Dr Ram, Dr GenoMax and Dr ATPoint. You embody what science and spreading knowledge is about. Without your expertise, time, and the willingness to help practically with the code, so many bioinformatic projects would never see the light of day. Your

contribution is beyond quantifiable. Åsa Nordgren – you ignited my love for R. Thank you for your time, willingness to selflessly share your knowledge and my first code chunk in R that grew into what I produced in this thesis.

Thank you to my first friends in the UK. How lucky I was to have met you! Thank you to Dr Isabel Okoye and Dr Sheila Govind for bringing me the first PhD adverts and planting in my head the first idea that I could do this, and to Harsha Siani for making me put money where my mouth is. You really exemplify how to lead a life with bravery and competence. To Dr Omodele Ashiru and Dr Déla Nacer – you inspired me then and you continue to inspire me today to bite my teeth and take on challenges and never stop learning. You are true friends. To Dr Joanna Waldock, Dr Jacqueline McDonald, Ryaka Poonawala for believing in me and keeping me sane throughout the duration of this PhD, for proof reading and showing me not to sweat the small stuff. Most heart-felt thanks go to Dr Othmar Engelhardt and Carolyn Nicolson, for all the advice about influenza but also for being mentors who propelled me to accomplish things I thought were out of my reach. It was a privilege to work for you.

A mis amigos hispanohablantes, gracias: a Dra Ledicia Botas Pérez – por modelar qué es una mujer de palabra, con principios y “work ethic” sin comparación y por envalentondandome a cada paso, a Profa Luna Santos Roldán y Ana Amaral– por nunca haber dudado de mis capacidades, al Dr Alfonso Gallo Bueno – mi amigo de siempre que estuvo aqui desde el principio, eres una incarnación del “Hombre del renacimiento” e inspiras a nunca dejar de aprender – tu apoyo significa más de lo que te imaginas, a Dr Danilo Villanueva Navarrete – una definición de amigo cuyos actos hablan por si mismos – nunca sabré cómo me merecí tu amistad y a Dra Kristel Ramirez Valdez – una guerrera que nunca se rinde, gracias por creer en mi y mostrarme que incluso los heroes tienen miedo a veces.

Thank you to the Panacea Innovation team – it was one of the most amazing experiences during my PhD. Thank you to Dr Mina Bekheet – for your leadership and the bravery to entrust us with your “baby”. I hope the team made you proud. Thanks to Dr Joe Healy for being so approachable and sharing what it takes to run a biotech start-up, to Dr Kyriakos Tzafestas – for your Balkan humour I miss so much from home and for all the career advice, to (future Dr) Isabel Zhang – my twin soul sister I did not know I had, for being a friend in the trenches, to Amaia Cadinanos Garai for showing me how to quietly amaze by doing your own thing, to Nathan Wiles and Dr Tom Gorgy for sharing your knowledge about commercialisation. It was an honour to be thought of as your equal and even more so a part of this Dream Team.

Hvala mojim prijateljicama: Kristini Balenovic, Mariji Serajlic, Renati Grenkovic, Mirni Petretic, Kasii Sturlan, Jasni Filipovic, Dr Ivi Krizaic, Dr Srdjani Grgurevic, I jednom prijatelju, ;) Dr Janku Gospocicu.

Puno mi je znacilo da ste uvijek I oduvijek tu za mene, da nase prijateljstvo, istinsko I vjecno, nadilazi udaljenostii I vrijeme.

Hvala teti Maji, teti Nadi, Kumi Meri I teti Tajki sta su uvijek bile ponosne na mene.

Hvala mojoj obitelji koja je voljna biti moja zastitna mreza za svaki skok koji se usudim. Bez vas to ne bi bilo moguce.

Contents

List of figures

Abbreviations

APC	Antigen Presenting Cells
ATHL1	acid trehalase-like 1
B4GALNT4	β – 1-4-N-acetyl-galactosaminyl transferase 4
chIFITM	Chicken IFITM
chPGC	Chicken Primordial Germ Cells
DEGs	Differentially Expressed Genes
DF-1	David Forster - cell
FCS	Functional Class Scoring
FDR	False Discovery Rate
GLM	General Liner Model
GO	gene ontology
GSEA	Gene Set Enrichement Analysis
HA	Haeamglutinin
HPC	High Performance Computing
hpi	hours post infection
IFITM	Interferon Inducible Transmembrane protein
IFNa	Interferon alpha
IFNb	Interferon beta
IL	interleaukin
IRF	interferon regulatory factor
ISG	interferon Stimulated Gene
KO	Knock-out
LGP2	laboratory of genetics and physiology 2
MAP	mitogen activated protein
MAVS	mitochondrial anti-viral sginalling
MDA5	melanoma differentiation associated gene
MDCK	Madin Darby Canine Kidney cell
MOI	multiplicity of infection
NFkB	nuclear factor kappa-light-chain-enhancer of activated B-cells
NLR	NOD-like receptor
PADOG	Pathways Analysis with Downweighing of Overlapping Genes
PAMP	Pathogen Associated Molecular Patterns
PCA	Principal Component Analysis
PFU	plaque forming units
PGC	progenitor germ cells
PGGHG	protein -glucosylgalactosylhydroxylysine glucosidase
PRR	Pathogen Recognition Receptor
RLR	retinoic acid inducible gene I
ROS	reactive oxigen species
SNV	single nucleotide variant
TF	transcription factor
TLR	Toll Like receptor
TPM	Transcript Per Million
GSEA	Gene Set Enrichment Analysis

1

Introduction

The immune system of complex organisms is layered to maximise the defence against wide variety of pathogens. It deploys various strategies, spatially and temporally separated, as well as mechanistically different. It can be broadly divided into intrinsic, innate and adaptive.

Intrinsic immune system consists of constitutively expressed restriction factors whose function is to antagonise viral infections. Some examples are shown in **Figure 1**. They do not require specific activation. They can interact with the virus particle even before the replication of the viral genome takes place and therefore halt the transmission in an immunologically naïve host such as laboratory cell lines [1,2]

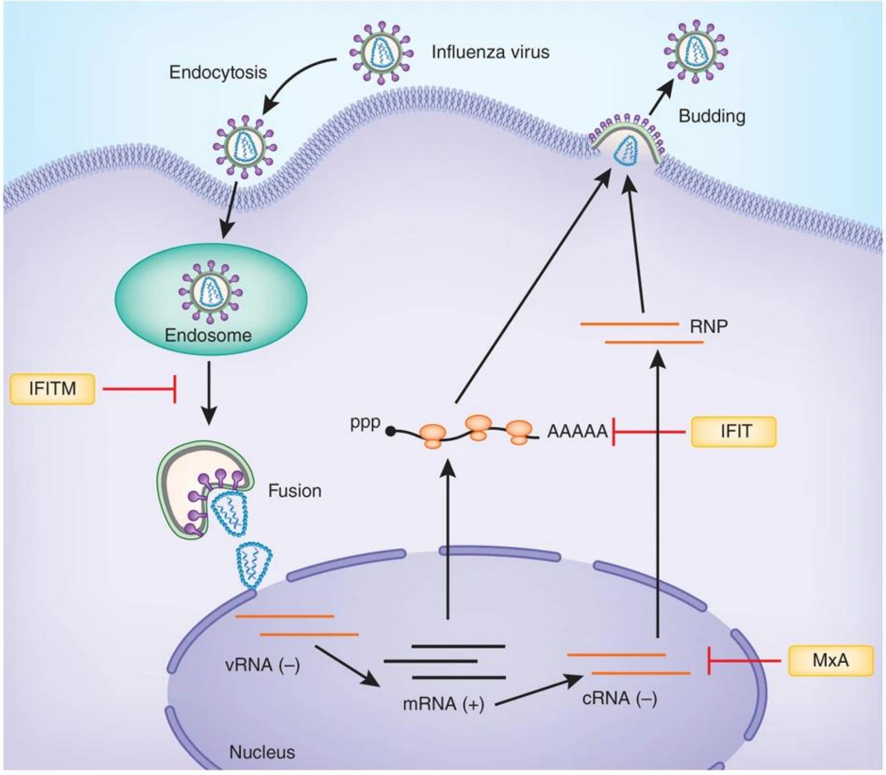


Figure 1 Intrinsic immunity against influenza Schema of the influenza life cycle depicting the stages in which IFITM, Mx and IFIT exert their restrictive action against infection. Borrowed from [2].

In contrast, the innate and adaptive immune system require activation to mount a response. The innate immune system recognises motifs that are common in invading pathogens and has no adaptive capability. It is triggered within hours from the initial exposure to a pathogen. The adaptive immune response develops in 7-14 days during an initial infection and is specific for the pathogen against which it is raised. Innate immune cells such as macrophages activate B and T lymphocytes through a process called antigen presentation. Upon activation B-lymphocytes produce antigen specific antibodies and T-lymphocytes perform distinct functions: helper T-cells and suppressor T-cells regulate the activities of other immune cells and killer T-cells eliminate infected cells. This cellular response provides specific and long-lasting protection and reacts rapidly upon subsequent exposure to the same pathogen [3].

1.1 The innate immune system

The innate immune system plays a critical role in immunity against all microorganisms. In this work the focus will be placed on characterisation in context of viral infections. Activation of the innate immune system is triggered by detection of viral nucleic acid by PRRs (pattern recognition receptors) after the replication of the virus genome starts. PRRs recognise PAMPs (pathogen associated molecular patterns). PRRs are expressed by APCs (antigen presenting cells) but can also be found in non-immune cells such as epithelial cells. Activated cell signalling causes apoptosis of the infected cell and simultaneously alerts the neighbouring cells of incoming infection through upregulation of paracrinally secreted cytokines [4]. There is an overlap between the intrinsic and the innate branch of the immune system: restriction factors which are constitutively expressed as part of the intrinsic immune system can be upregulated by interferon (IFN), a key signalling molecule of innate immune system [1], to reinforce their effect.

1.1.1 Pattern Recognition Receptors (PRRs)

PRRs can be divided into several groups according to their localisation and structure: (i) TLR (Toll-like receptors): localised in the plasma membrane and endosomes, e.g.TLR3, TLR7, TLR9. They recognise wide array of PAMPS: proteins, nucleic acids, glycans: and activate MAP (mitogen activated protein) kinase, NFκB (nuclear factor kappa-light-chain-enhancer of activated B cells) and IRF (interferon regulatory factor) pathways [5]. (ii) RLR (RIG-I like receptors), e.g. MDA5 (melanoma differentiation associated gene 5), LGP2 (laboratory of genetics and physiology 2) and RIG-1 (retinoic acid inducible gene I) are cytoplasmatic sensors of viral RNA (vRNA) that activate IFN I production through the mitochondrial anti-viral signalling (MAVS) pathway [6]. (iii) NLR (NOD-like receptors), e.g. NOD2 (nucleotide-binding oligomerisation domain 2), are also located in the cytoplasm. They were long thought to be bacterial sensors because they detect dsDNA and activate NFκB pathway to stimulate cytokine production, but recent evidence shows they also successfully target viruses [7].

Viral pathogen recognition is mostly mediated by RLRs, which share many similarities with TLRs in terms of signalling pathways they activate (NFκB, IRF3/7) and the genes whose expression they induce (cytokines and type I IFNs). RLRs contain an RNA helicase domain and a C-terminal regulatory domain (CTD), with RIG-I and MDA5 also containing a caspase activation and recruitment domain CARD). LGP2 lacks a CARD and acts as a regulator of RIG-I and MDA5. The CTD of RIG-I and MDA5 binds to vRNA and this interaction activates the CARD [8]. RIG-I activation is modulated by polyubiquitination by either TRIM25 (tripartite motif containing protein 25) [9], Riplet [10] or MEX3C [11]. Following activation, downstream signal comprises of IFNβ promoter stimulator (IPS)-1 (also known as MAVS), IKK (I kappa B kinase) and TBK1 (TANK binding kinase 1) who phosphorylate NFκB and IRF3/7 (interferon regulatory factor 3 and 7). These transcription factors in turn dimerise, translocate to the nucleus and activate the transcription of cytokines and type I IFNs [6] as illustrated on **Figure 2**.

The latest research also points to the importance of involvement of NLRs in viral infections [12]. NOD2 associates with MAVS to activate IRF3 thereby upregulating the production of IFNβ [7] and promotes the formation of the "inflammasome" [5]. NLR viral sensing is mostly by proxy rather than direct activation by viral components. NLRs are activated by ROS (reactive oxygen species), membrane disintegration and other stress related cellular changes. They react upon unsuccessful IFN response by activating the catalytic activity of caspase 1 that cleaves pro-IL1β (pro-interleukin 1β) into its active form and thus promote inflammation.

The most conspicuous difference between chicken and mammalian PRR systems is the lack of RIG-I in chickens [13]. However, MDA5 can compensate for the absence of RIG-I and recognise short and long double stranded RNA (dsRNA)[14]. This signalling can additionally be enhanced by positive regulation of LGP2. Chickens also lack TLR8, another single stranded RNA (ssRNA) sensor as well as a transcription factor (TF) that is activated through RIG-I pathways: IRF3. IRF7 and IRF10 are the only members of this TF family that have been identified so far [15]. A few more deficiencies are reviewed in [16]. Nevertheless, chickens are immunologically capable to respond to infection, confirming functional redundancy of the immune system.

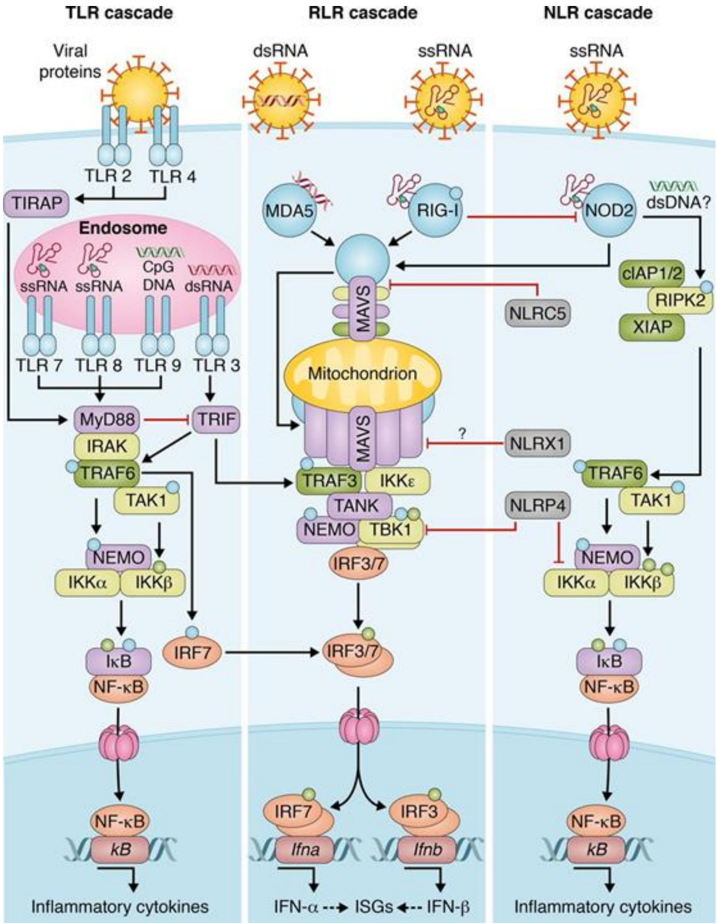


Figure 2 Three classes of pattern recognition receptors (PRRs) involved in viral RNA recognition: TLR (Toll-like receptors), RLRs (RIG-I-like receptors) and NLRs (NOD-like receptors). Upon detection of viral RNA TLR 7 and TLR3 initiate a signalling cascade through MyD88 (Myeloid Differentiation factor 88) and TRIF (Toll∕IL-1 receptor (IL-1R) homology domain-containing adapter inducing IFN-β), respectively. RLRs; RIG-I (Retinoic acid-inducible gene I) and MDA5 (melanoma differentiation-associated gene 5) associate with an adaptor protein IPS-1 (IFNβ promoter stimulator) (not denoted). NLRs; NOD2 (nucleotide-binding oligomerisation domain) binds to MAVS(mitochondrial anti-viral signalling) to activate IRF3 (IFN regulatory factor 3). Kinases used for phosphorylation of respective transcription factors: NFκB (nuclear factor kappa-light-chain-enhancer of activated B cells) and IRF3/7 are shared between TLR and RLR pathways: IKK (I kappa B kinase) and TBK1 (TANK binding kinase 1) act as intersection of these 2 pathways which enables them to activate either transcription factor. Adapted from [4].

1.1.2 Interferon signalling

Interferons are a group of cytokines divided into types according to their structure and type of receptor they bind: Type I (α, β, ε, κ, and ω), Type II (γ), and Type III (λ). There are 13 distinct proteins within the IFNα family in humans, two in the IFNβ family, one member in each of the IFNε, κ, ω, and γ families, and three in IFNλ family.

In chickens, IFNα is multigene family albeit with less members than in mamals. IFNβ is s single gene family, as are IFNγ and IFNλ [17].

Type I, type II and type III IFNs bind to their respective receptors: IFNAR, IFNGR and IFNLR. See **Figure 3**. All receptors are associated either with Jak1 (Janus activated kinase 1) or Tyk2 (tyrosine kinase 2) both of whom downstream converge at Stat 1 (signal transducer and activator of transcription 1 [18]. For Type I and III signalling, phosphorylated Stat1 and 2 bind to IRF9 and form the IFN stimulated gene factor 3 (ISGF3), a transcription factor that recognises ISRE (interferon stimulated responsive element) found in the promoter sequences of interferon stimulated genes (ISGs). This process works as a positive feedback loop, exacerbating IFN I response [19]. For Type II signalling Stat1 homodimers bind to interferon gamma activated sequence (GAS). Type II IFN does not influence the expression of genes under control of an ISRE promoter (see **Figure 3**).

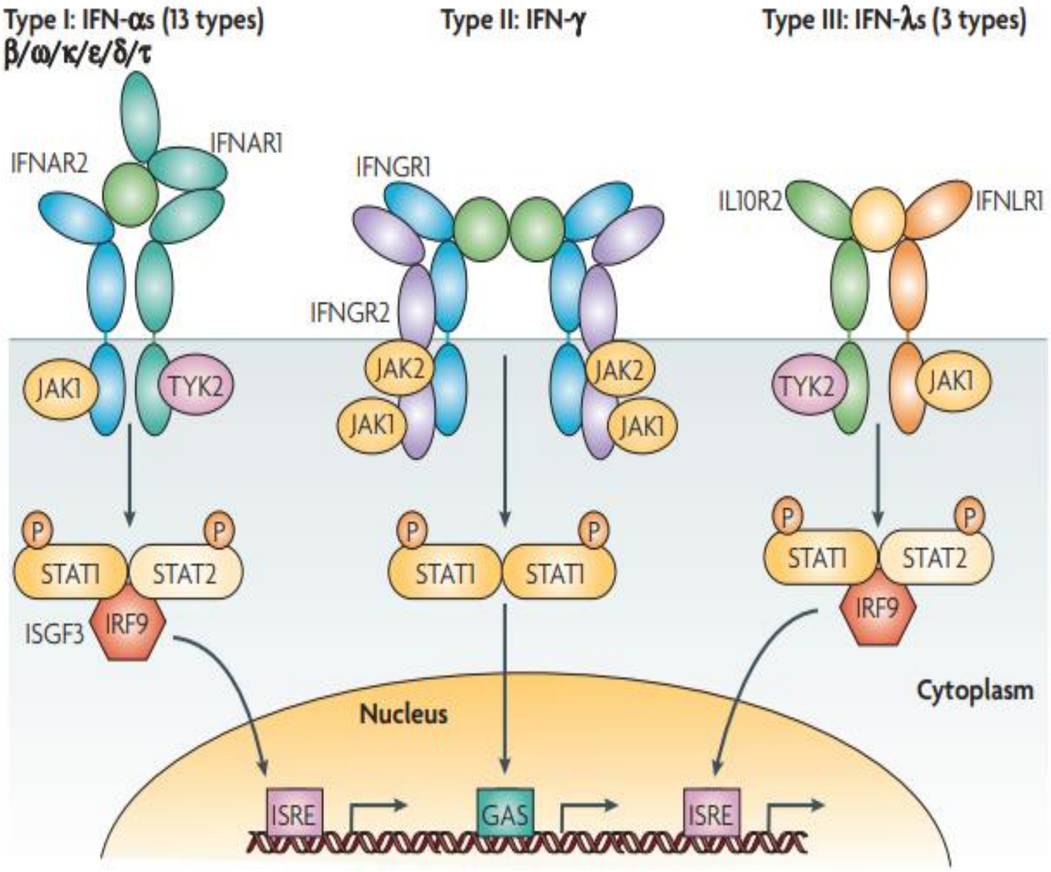


Figure 3 ***Classical JAK-STAT signalling pathway.*** The type I IFN receptor is composed of two subunits, IFNAR1 and IFNAR2, associated with Janus activated kinases (JAKs), tyrosine kinase 2 (TYK2) and JAK1, respectively. IFN-γ, binds type II IFN receptor, composed of IFNGR1 and IFNGR2, associated with JAK1 and JAK2, respectively. Type III IFN, iFN λ, binds its receptor composed of IFNLR1 and IL10R2 (interleukin 10 receptor 2). Activation of the JAKs and TYKs that are associated with the type

I and III IFN receptor results in tyrosine phosphorylation of STAT2 (signal transducer and activator of transcription 2) and STAT1. This leads toformation of ISGF3 (IFN-stimulated gene (ISG) factor 3) complex. ISGF3 translocates to the nucleus and binds IFN-stimulated response elements (ISREs) to initiate transcription. Type II IFNs induce the formation of STAT1 homodimers that translocate to the nucleus to bind GAS (IFN-γ-activated site) elements that are present in the promoter of certain ISGs (interferon stimulated genes) and promote their transcription. Adapted from (19).

ISGs, whose expression is modulated via IFN, perform a myriad of antiviral and anti-proliferative functions [20], many of which remain unknown.

Viruses attempt to evade the immune response by coding for antagonists. Antagonists employ various mechanisms to achieve immune escape: the inhibition of ubiquitination of RIG-I [21] causing the arrest of downstream signalling, the antagonism of the AP-1 [22] – a transcription factor that induces expression of several cytokine genes, a complete transcriptional shutoff [23] which dampens antiviral responses, the association with MAVS, an interaction that decreases the mitochondrial membrane potential [24] and disables polymerisation necessary for downstream signal, etc.

Interferon signalling is a crucial process for host defence from pathogens and as such is under significant selective pressure. Immune genes evolve quicker than the other host genes [25] because pathogens they are co-evolving with have much higher mutation rates. In addition to regulation of ISG expression, IFN also modulates expression of regulatory RNAs: miRNAs and lncRNAs [26] as well as influences the cellular immune response (neutrophils, macrophages, NK cells and dendritic cells).

1.1.3 Interferon-stimulated genes in chicken

ISGs are genes that have ISRE or GAS elements present in their promoters and as such are subject to ISGF3 and GAF targeting, respectively. Their main function is to inhibit viral infection by interfering with various stages of viral life cycle as illustrated on **Figure 4**.

Below is a brief overview of restriction factors that are commonly encountered in avian immunology research.

ZAP (zinc finger antiviral protein) recognises GC rich genomes [27] or a specific ZAP responsive element present in viral genomes [28]. Both interactions lead to degradation of vRNA via RNA processing exosome. ZAP exists in 2 isoforms: long and short, however, only the long isoform has been identified in chickens [29].

PKR (Protein kinase R) targets translation by phosphorylating EIF2a, a translation initiation factor[30]. Ko et al [31] confirmed the polymorphic genetic structure of chicken PKR and its ability to restrict VSV in chicken cell culture. However those results did not translate to *in vivo* system [32].

2'-5'-OAS (2'-5'-oligoadenylate synthetase) is associated with RNAse L and acts via directing vRNA towards degradation. The resulting free RNA fragments stimulate RIG-I and MDA25 and consequently initiate the IFN pathway. The role of related gene OASL (oligoadenylate synthetase like) has also been described [33]. Duck and ostrich OASL significantly reduced the replication of AIV, NDV and IBDV in DF-1$^{dRIG\text{-}I+/+}$ cells [34]. Chicken OASL has also been shown to have antiviral activity against West Nile virus, when ectopically expressed in mammalian system that has a functional RIG-I pathway [35].

Viperin affects viral replication [36] and/or egress [37]. Goossens et al characterised chicken viperin and found it contains same domains as the mammalian counterpart [38] They observed significant upregulation upon stimulation with IAV and IBDV both *in vivo* and *in vitro*.

IFITMs (Interferon Inducible Transmembrane Proteins) were discovered in chicken cells in 1984 by Friedman [39]. IFITMs have not attracted much interest until late 2009 when Brass [40] and Shapira (41) published their studies on IFITM restriction of wide range of viruses. This range was further expanded to many enveloped viruses; influenza [41,42], respiratory syncytial virus (RSV) [43], West Nile virus, dengue virus [40], SARS coronavirus [44], Marburg virus and Ebola virus and non -enveloped ones; such as reovirus [45]. This ignited the field of research studying the mechanism of restriction as IFITMs seemed a promising candidate to target to enhance the host response during infection.

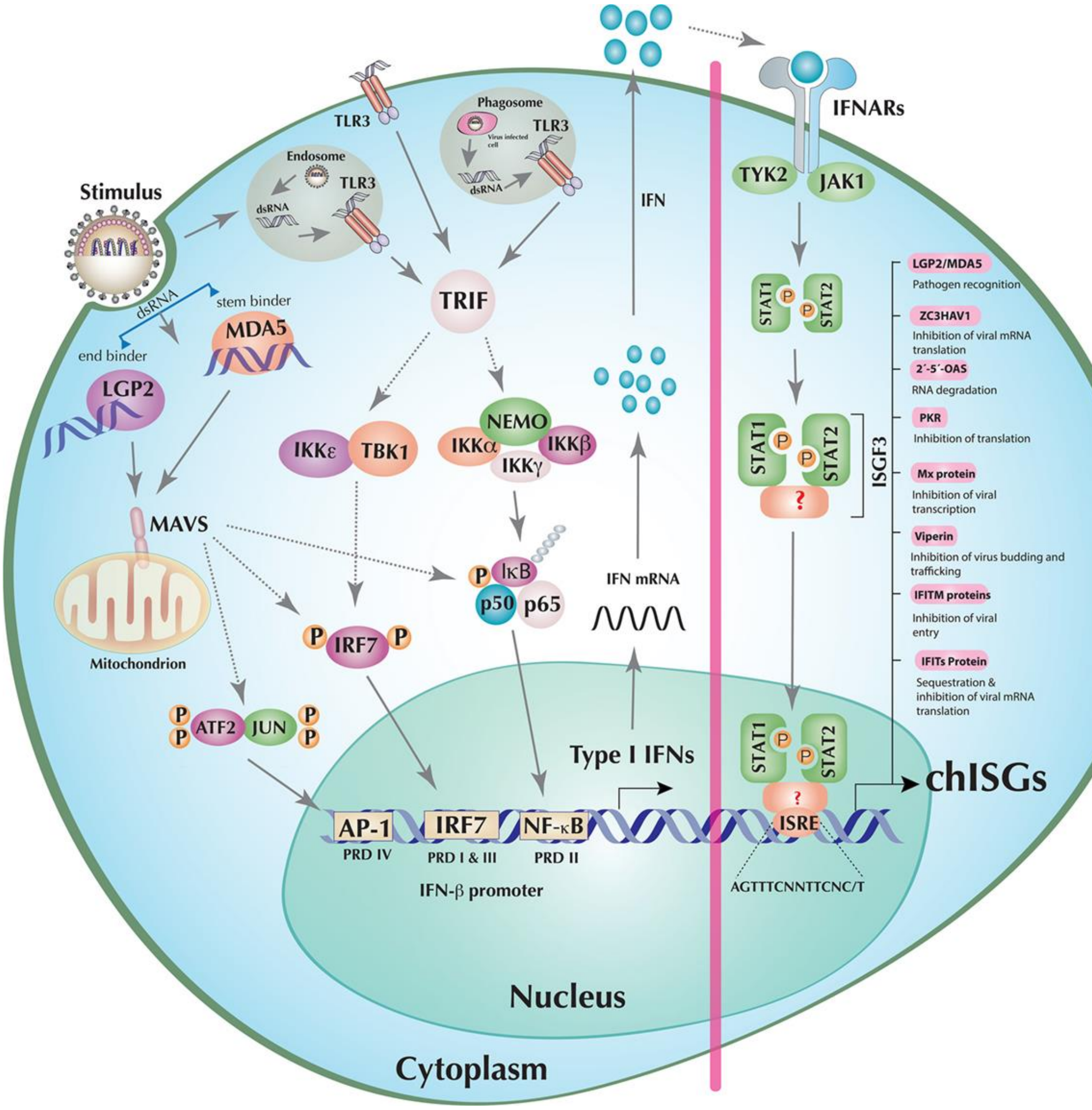


Figure 4 Induction of interferons (IFNs) and upregulations of chicken interferon stimulated genes (chISGs). Taken from [14]. Note the absence of RIG-I, IRF3 and TLR8 does not affect the IFN production due to redundancy of pathways in the network. Vertical pink line separate Interferon induction pathways from interferon signalling that result in upregulation of chISGs. chISGs whose transcription is activated by ISGF3 are depicted at the most right with proposed mode of action.

1.2 IFITMs

1.2.1 Evolution and genetic structure of IFITMs

The human and avian IFITM family is composed of 5 proteins: namely, IFITM1, IFITM2, IFITM3, IFITM5 and IFITM10 belonging to an evolutionary conserved protein family called dispanins [46]. Level of dispanin sequence conservation between bacteria and eukarya is high. Several domain motifs are conserved across, one of which is double cysteine motif that is frequently palmitoylated [47]. Palmitoylation confers antiviral activity and enhances stability, crucial properties for its function [48].

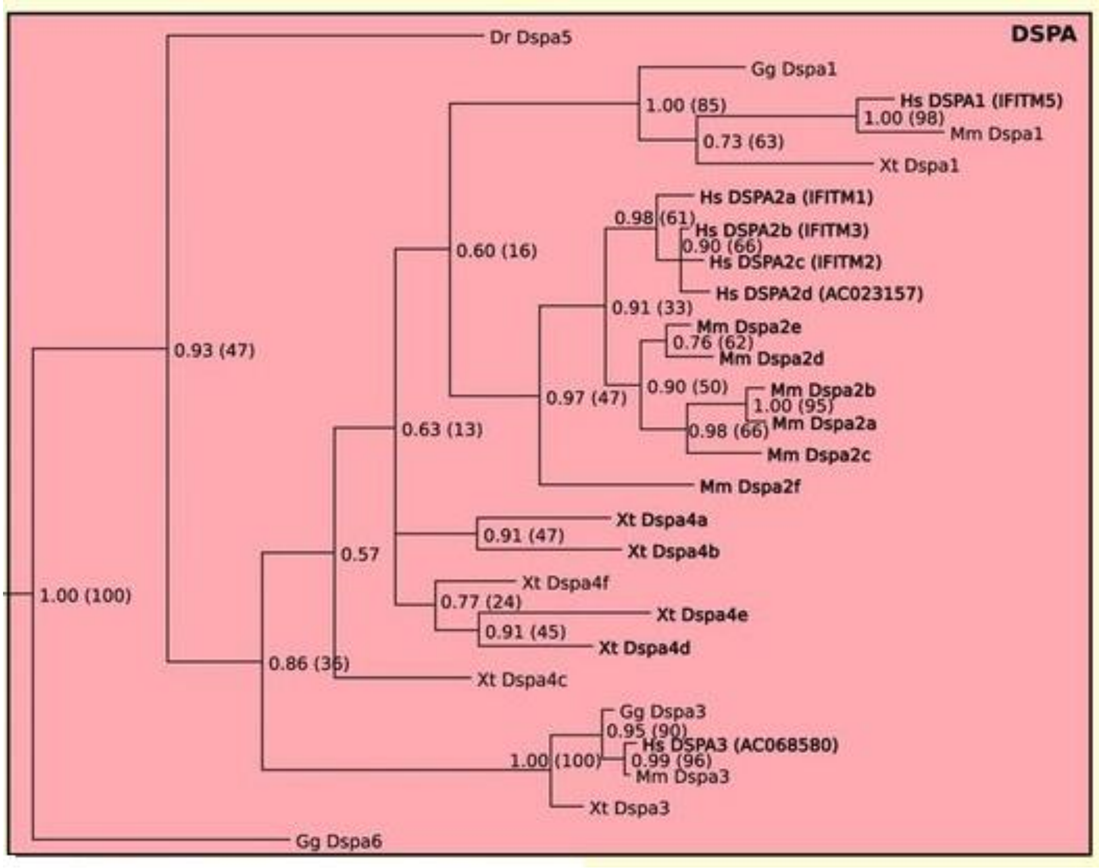


Figure 5 Phyologenetic tree of DSPA subgroup of dispanins. Hs = Homo sapiens, Mm = Mus musculus, Xt = Xenopus tropicalis, Gg = Gallus gallus. Adapted from [46].

IFITM locus is flanked by *B4GALNT4* (β-1,4-N-acetyl-galactosaminyl transferase 4) and *ATHL1* (acid trehalase-like 1) in all studied vertebrates, an evidence of common ancestor duplication event.

DISPA2 subgroup is what we refer to as IR-IFITMs (immunity related) subset of IFITMs because they can be upregulated by IFN I and II. *M. musculus* and *H. sapiens* individual *DISPA2/IFITM* genes show greater similarity in sequence identity within a species than to orthologs in other species hence current naming consensus does not accurately indicate evolutionary history. Chicken (Gg) *Dspa1* branches off late and shows greatest sequence similarity to *huIFITM5*, so it was the first annotated chicken *IFITM (chIFITM*). Gg *Dispa6* separates first and clusters with no other members of the *DSPA* family. *Gg Dspa3* cluster together with Hs *Dispa3* and Mm *Dispa3*, outside of the *IFITM/Dspa2* family. No Gg *Dispa2* was identified. *IFITM* gene family is a rich paralogous family with many copies and pseudogenes [49]. This makes phylogenetic studies problematic as pseudogenes can be mistaken for functional genes.

First attempts at annotating these genes in chicken by Smith et al [42] used syntenic approach and *chIFITMs* were named after their human counterparts. However, concerns were raised that *chIFITM2* is actually analogous to *huIFITM1*, judging by their cellular localisation data. Closer inspection by Bassano et al [49] in 2017 confirmed these suspicions using PacBio and MiSeq sequencing platforms and additional immunofluorescence assays. Genetic analysis has determined that most polymorphisms are found in N-terminal cytoplasmatic domain suggesting functional significance. *chIFITM2* is most variable while *chIFITM3* the least.

1.2.2 Protein structure and posttranslational modifications of IFITMs

Structural motifs determined by sequencing and protein modelling defined a number of conserved domains: NTD (N-terminal domain), CTD (C-terminal domain), CIL (conserved intracellular loop), IM (intramembrane domain) and TMD (transmembrane domain). For a number of years, it was accepted that the topology must include 2 transmembrane domains given *IFITMs'* background as dispanins as well as research done by Yount et al [48], see **Figure 6**, model (a). Further studies on huIFITM3 by the same group opened possibilities for intramembrane topology model [47], as illustrated in **Figure 6** (b). Additional work on murine IFITM3 proposed a model that combines the two [50]. Research by Ling et al [51] from 2016 using EPR (electron paramagnetic resonance) and NMR (nuclear magnetic resonance) agrees with type II membrane topology comprising one TM domain and one IM domains where CIL and NTD are located cytoplasmatically and CTD is extracellular (model (c) in **Figure 6**.

All three models possess conserved cysteine motif (C71, 72 and 105) on the junction of the CIL domain and the putative TM domain. This is a well-known palmitoylation site needed for full viral restriction, as mentioned previously [48]. IFITM3 can also be ubiquitinated on lysines found on the domains exposed to the cytosol. Ubiquitination will render IFTM3 more instable and reduce its antiviral properties by promoting its relocation from endolysosomes to autophagosomes.

Studies looking into subcellular localisation of chIFITMs have been hindered by the lack of suitable antibodies, so tagged constructs are used. Tags, however, can affect the localisation so interpretation of data must be undertaken with care. Smith et al [42] showed that chIFITM 1 localises on the plasma membrane and chIFITM3 to the endosomes. This is in keeping with the entry site of viruses they restrict.

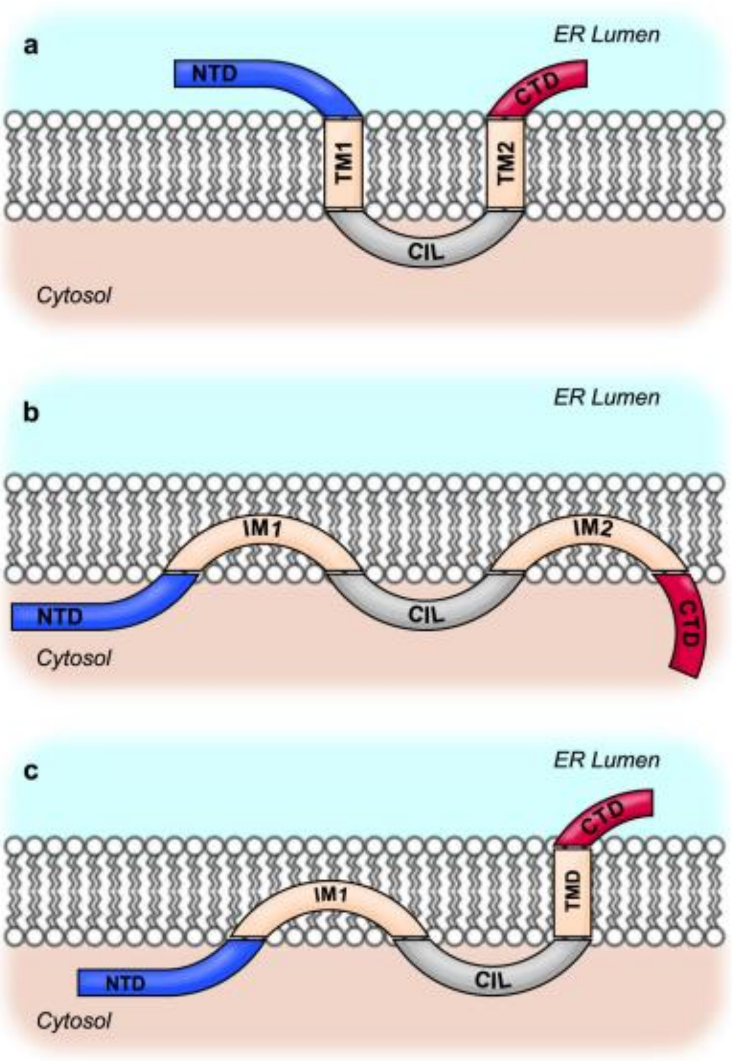


Figure 6IFITM protein membrane topology *(a) First model with 2 transmembrane domains (TM) and C-terminal domain (CTD) and N-terminal domain (NTD) in the endoplasmic reticulum lumen. (b) A model with 2 intramembrane domains with CTD and NTD on the cytoplasmatic side of the membrane. (c) A topology model with one TM domain, and one IM domain and CTD in endoplasmic reticulum and NTD in the cytoplasm. Adapted from (53).*

1.2.3 Mode of action of IFITMs

One of the proposed modes of antiviral restriction is the inhibition of the fusion of cellular and viral membranes. IFITMs localised in the endosomal membranes stabilise the hemifusion intermediate and thus entrap the virus within the endosome [52].

An alternative theory, so called “tough membrane model” [53,54] states that the IFITMs localised in rafts of plasma membrane can associate with one another [55] and with tetraspanins. Tetraspanins are a family of proteins that anchor other proteins to the membrane and stabilise them in order to enable further interactions. This net of IFITMs and tetraspanins creates curvature in the membrane towards the cytoplasm, away from the virus particle, and therefore generates an additional energetic barrier to successful membrane fusion. Tetraspanins and IFITMs, as well as other plasma-membrane components, get incorporated into extracellular vesicles (EVs) during budding from the plasma membrane and in this way extend their effector roles to neighbouring cells which take up these vesicles [56]. In some conditions, like cancer, IFITMs expressed on the surface can affect the uptake of EVs [57], a phenomenon that could be utilised in therapeutic applications [58].

Another mechanism to modulate physical properties of lipid membranes employed by IFITMs is to affect cholesterol trafficking. Cholesterol facilitates membrane fluidity. IFITM protein overexpression leads to accumulation of cholesterol in the endosomal membranes and increases its rigidity [53]. Possibly this influences various processes that take place on the membranes that may or may not be a part of the immune response. This model expands the scope for potential roles of IFITMs. However, involvement of that property of IFITMs in their restriction activity has been called into question [59].

One of the alternative functions is promoting the infection [60]. Zhao et al identified 3 mutations as well as posttranslational modifications in huIFITM1 and 3 that facilitate conversion of restriction function into enhancing function and enable SARS-CoV and MERS-Cov entry [61].

The most potent antiviral members of the family are IFITM3 and 1, depending on the site of entry for a particular virus. Various roles have been proposed for these proteins beyond the one they play in immune response [62]. Human IFITM5 has a role in bone mineralisation and is not induced by interferon, while chIFITM5 is an immune related IFITM. Function of chIFITM10 is not precisely determined but there is some indication that it may be involved in antiviral activity. Okuzaki et al [63] have found chicken IFITM10 exhibits viral restriction properties in chicken embryonic fibroblasts (CEFs) not induced by interferon, suggesting that an interferon independent pathway may be involved. This finding was substantiated by Jang et al [64] who got a 10-fold overexpression in tracheal epithelial cells (TECs) from 20-day old chick embryos infected with H1N1 influenza A virus (AIV).

1.2.4 Influenza virus and IFITMs

Influenza virus is a single stranded negative RNA virus, belonging to *Orthomyxoviridae* family. Its genome is segmented and encoding 8 proteins [65]: two surface antigens: HA (Haemagglutinin) and NA (Neuraminidase), three RNA-dependent RNA polymerase (RdRp) subunits: PB1, PB2 and PA, the non-structural proteins NS1 (immune system modulator) and NEP (nuclear export protein) which share a segment, as do two M proteins M1 (matrix) and M2 (an ion channel). Last segment is encoding the NP (nucleoprotein). Accessory proteins, PB1-F2 and PA-X, are expressed through alternative open reading frames from the segments 2 and 3, respectively [66]. Accessory proteins have a role in the pathogenicity of the virus and the suppression of the host immune response [23,67]. Each segment is bound to RNP (ribonucleoprotein) complex along with all components of the RdRp. During budding from the plasma membrane the eight segments are packaged together into an enveloped virion, structurally supported by the M1 protein. During the infection the antigen proteins are readily expressed on the cell surface and so the envelope contains proteins normally found on the plasma membrane, out of which the HA and NA are the most abundantly represented. An infection is initiated when the HA binds to sialic acid receptor on the host cell surface and the virus is internalised into an endocytic compartment. M2 proteins form ion channels that go through the viral envelope and mediate virion acidification required for the release of the viral genome into the cytoplasm. Once the vRNPs are not bound to the M1 protein they can be transported into the nucleus where the replication and transcription take place. These processes are mediated through RdRp. Replication proceeds via self-complementarity of 5’ and 3’ end that form a panhandle. Synthesised positive sense intermediary copy RNA cRNA is then used for mRNA synthesis for subsequent transcription and for vRNA that gets packaged into the progeny virions [68].

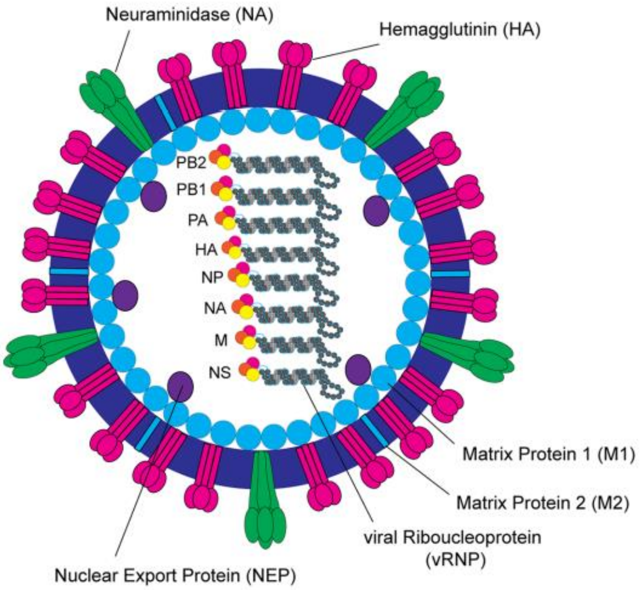


Figure 7 Schematic representation of a prototypic influenza virus Taken from [65].

The evidence for AIV restriction by IFITMs is abundant. Brass et al [40] confirmed that in primary chicken fibroblasts (CEFs) and in Madin-Darby canine kidney cells (MDCKs), influenza infection is successfully restricted by just basal levels of IFITM3. Conversely, Smith et al [42] showed that transient reduction of IFITM gene expression using RNAi results in higher viral titre in DF-1 cells (chicken embryonic fibroblasts) infected with pseudotyped avian influenza A virus. A seminal *in vivo* study on

Ifitm3 $^{-/-}$ mice by Everitt et al [41] directly correlated poor clinical outcomes with absence of IFITM3. Low pathogenicity influenza A virus (LPAIV) does not normally cause severe infection in wild type (WT) mice. However, in mice lacking IFITM3 (KO) infection with LAIV resulted in fulminant pneumonia. To investigate association of IFITM3 and severity of influenza infection they sequenced samples from human individuals who required hospitalisation in time of 2009 pandemic as well as the following season. They identified a SNP rs12252 in the *ifitm3* where the major allele thymine (T) is replaced by cytosine (C) that resulted in translation of a truncated product. Statistically significant proportion of hospitalised patients possessed the minor allele. Functional testing in cell culture confirmed reduced restriction of IAV by IFITM3 with minor CC genotype. However, this finding remains controversial as many subsequent studies and meta-analyses showed conflicting results [69–72]. Blyth et al [73] functionally demonstrated that duck IFITM3 (dIFITM3) reduced the number of IAV infected DF-1 cells, and not dIFITM1 or 2. Interestingly, they also showed that YEML motif, indispensable for correct localisation and function in mammals, is not essential in ducks.

Smith et al [74] examined chIFITM and dIFITM expression in lung samples of chickens and ducks respectively, infected with high pathogenicity influenza virus, using RNA-seq. Level of chIFITM transcription was only slightly upregulated in comparison to dIFITMs. The authors hypothesise that it is the robust upregulation of dIFITMs that enables ducks to resist influenza, but they do not go deeper to investigate what causes this difference in upregulation. It is conceivable that different strains of AIV differentially regulate the expression of IFITMs. In humans, a genetic difference in IFITM3 promoter was discovered to be the reason for differential regulation of expression. A SNP in IFTM3 identified as rs34481144 in the 5'UTR decreased the binding of IRF3 and caused a drop in expression levels of IFITM3 that was correlated with more severe clinical outcomes in three distinct cohorts of patients infected with influenza [75]. Authors subsequently found that CTCF (a protein involved in organisation of chromosomal structure) binding was increased in risk allele carriers. This disrupted long range (700kb) correlation of expression levels of neighbouring genes. This illustrates how regulation of IFITM locus may impact multiple pathways that potentially cooperate during an infection.

A general lack of suitable reagents such as antibodies against chIFITMs or characterised cell lines has been hindering research in avian immunology. Suitable antibodies to individual IFITMs are difficult to generate due to high degree of structural similarity among them in the regions that are sufficiently immunogenic. Transcriptomic studies are attractive because they depend less on chicken-specific reagents. To characterise transcriptional profile of the host in relation to infection with AIV, numerous comparative studies were conducted. The differentially regulated antiviral genes seem to differ depending on the viral strain used, time points selected for observation, model system (*in vivo or in vitro)* as well as which host was subjected to infection i.e. duck [76,77] or chickens [78,79].

1.3 Overall aims

IFITMs have been shown to severly restrict avian influenza *in vivo* [41,44,52] and significantly ameliorate clinical symptoms of influenza *in vivo* [41,69,80]. We wanted to investigate the potential of application of that observation in vaccine manufacture. Due to yearly release of influenza vaccines, manufactures are under extreme pressures to increase vaccine yields because of time and profitability constraints. Cell-culture based vaccines are being heavily developed with some already on the market. However, great majority of vaccines are still produced in embryonated chicken eggs. This piece of research addresses both of those aspects of vaccine production. We hypothesise that ablation of chicken IFITM locus in DF-1s and PGCs would lead to increased viral titres.

DF-1 are proposed for cell culture-based vaccines because they are a natural host for avian influenza. Currently influenza vaccines are produced in MDCK cells, meaning the avian HAs evolve to adapt to "human-type" receptors i.e. α2-6 sialic acid receptor. It quickly becomes clear that growing avian vaccines in an avian host would be a welcome advantage.

PGCs (Primordial Germ Cells) are germ cells that are very easily retrievable from chicken. They are also very amenable to genetic engineering and thus very frequently used for production of transgenic chickens. We propose to test chicken IFITM KO PGC-derived fibroblasts for influenza restriction as an approximation to testing in an animal model. In the future a transgenic animal could be generated, and preliminary feasibility studies conducted for egg-based vaccine production. Differentiating PGCs into fibroblasts facilitates comparison with DF-1 fibroblasts, while maintaining more natural signalling pathways, thus PGC-derived fibroblasts are used for this arm of the study.

Overall aims:

- Generate suitable *chIFITM* KO models
- Establish infection system suitable for testing the effects of the *chIFITM* knock-out in terms of cell culture conditions
- Characterise the KO cell lines relative to their WT counterparts for:
    - permissiveness to influenza virus
    - potential to grow virus in multicycle infection experiment
    - their response to infection on the transcriptomic level using RNA-Seq
    - IFITM expression

2

Materials and Methods

2.1 Materials

2.1.1 Cell lines

Table 1 Cell lines used in this study

Cell line	Cell strain	Genotype	Cell type	Source
DF-1	DF-1, parental	WT	Immortal chicken adherent fibroblasts	ATCC
DF-1	DF-1_608	WT	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
DF-1	DF-1_609	WT	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
DF-1	DF-1_610	WT	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
DF-1	DF-1_225	ChIFITM KO	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
DF-1	DF-1_250	ChIFITM KO	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
DF-1	DF-1_600	ChIFITM KO	Immortal chicken adherent fibroblasts	The Pirbright Institute/Horizon Discovery
PGC	PGC-WT	WT	Finite primordial chicken germ cell -derived adherent fibroblasts	The Roslin Institute, kind gift by Dr Mike McGrew
PGC	PGC-KO	ChIFITM KO	Finite primordial chicken germ cell -derived adherent fibroblasts	The Roslin Institute, kind gift by Dr Mike McGrew
MDCK	MDCK	WT	Immortal adherent canine kidney epithelial cells	The Pirbright Institute, central Research Services

2.1.2 Influenza virus strains

All viruses were sourced from the Influenza Group at the Pirbright Institute and are wild-type (non-reassortant) strains.

- A/Chicken/Pakistan/UDL-01/08 (H9N2)
- A/duck/Ukraine/63 (H3N8)
- A/duck/Sing-Q/119/97 (H5N3)
- A/Puerto Rico/8/1934 (H1N1)

2.1.3 Primers

2.1.3.1 RT-qPCR primers and probes for reference genes

Table 2 Primers and probes for reference genes

Gene	Primer	Sequence	Chemistry
ACTB	Forward	5’-CAGGTCATCACCATTGGCAAT-3’	Taqman
	Reverse	5’-GCATACAGATCCTTACGGATATCCA-3’	
	Probe	5’-(**JUN**)-CACAGGACTCCATACCCAAGAAAGATGGC-(QSY)-3’	
B2M	Forward	5’-AAGGAGCCGCAGGTCTAC-3’	Taqman
	Reverse	5’-CTTGCTCTTTGCCGTCATAC-3’	
	Probe	5’(**FAM**)-CCGGGATGAGCACGGTCTGAAGAATT-(QSY)-3’	
HMBS	Forward	5’-GGTTGAGATGCTCCGTGAGTTT-3’	Taqman
	Reverse	5’-GGCTCTTCTCCCCAATCTTAGAA-3’	
	Probe	5’-(**ABY**)-CCTGACCTCTGCTTTGAGATTGTTGCCA-(QSY)-3’	
HPRT1	Forward	5’-TGGTCAAAAGAACTCCTCGAAGT-3’	Taqman
	Reverse	5’-TGTAATCGAGGGCGTATCCAA-3’	
	Probe	5’-(**FAM**)-TCCAACAAAGTCTGGCCGATATCCCA-(QSY)-3’	
PGK1	Forward	5’-GTTTATGTCAATGATGCTTTTGGAA-3’	Taqman
	Reverse	5’-GCCTTTGCAAAATAATCCAGTTCT-3’	
	Probe	5’-(**FAM**)-CATCGTGCTCACAGCTCCATGGTAGGT-(QSY)-3’	
PLA2	Forward	5’-GCACAAGACATTTGGCAGTTGT-3’	Taqman
	Reverse	5’-TGTGACATTTGTGGCTTTCCTTA-3’	
	Probe	5’-(**JUN**)-CAACACATTGTGGTGGAACACCAGTACTCA-(QSY)-3’	
PPIA	Forward	5’-CCCGTCGTGTTCTTCGACAT-3’	Taqman
	Reverse	5’-CCCTTGTAGCCAAATCCCTTCT-3’	
	Probe	5’-(**VIC**) -CACCTTCGAGCTCTTCGCTGACAAGG-(QSY)-3’	
RPL13	Forward	5’-TCGTGCTGGCAGAGGATTC-3’	Taqman
	Reverse	5’-TCGTCCGAGCAAACCTTTTG-3’	
	Probe	5’-(**VIC**)-TAATGCCCGCCAGTTTAAGCTCTTCTAGGC-(QSY)-3’	
RPLP0	Forward	5’-TTGGGCATCACCACAAAGATT-3’	Taqman
	Reverse	5’-CCCACTTTGTCTCCGGTCTTAA-3’	
	Probe	5’-(**JUN**)-CATCACTCAGAATTTCAATGGTCCCTCG GG-(QSY)-3’	
TBP	Forward	5’-CTTCGTGCCCGAAATGCT-3’	Taqman
	Reverse	5’-GCGCAGTAGTACGTGGTTCTCTT-3’	
	Probe	5’-(**FAM**)-CTCATAATAACAGCAGCAAAACGCTTGGGA-(QSY)-3’	

2.1.3.2 RT-qPCR primers and probes for genes of interest

Table 3 Primers and probes for genes of interest

gene	Forward Primer	Reverse Primer	Probe	reporter	quencher	reaction chemistry
IFITM1	TGTGCTCTCTCATACTCCATCA	TCTTCGCTGTCCTCCCATAG	CCTACGAAGTCCTTGGCGATGATCCTATCC	ABY	QSY	Taqman
IFITM2	GCTTCGCTCTCGCCTACC	CAGTCGCGGGCCTTGAT	TGCTTCTGCTTCCCCTCGCTCATCTT	JUN	QSY	Taqman
IFITM3	GAAGTCCAGGTCTTCTCCGT	GGCGATGATGAGGATGATGAG	ATGGCTCCACTGCGAAGTACCTGAACA	VIC	QSY	Taqman
IFITM5	CACAAGCGGGACTCATCTCC	CCGGTCCCGTGCCTTGA	CACCATCTACATGAACTTCTGCTGCCTCGG	FAM	QSY	Taqman
IFNa	GACAGCCAACGCCAAAGC	GTCGCTGCTGTCCAAGCATT	CTCAACCGGATCCACCGCTACACC	ABY	QSY	Taqman
IFNb	CCTCCAACACCTCTTCAACATG	TGGCGTGTGCGGTCAAT	CCACACTGCAGCTGGAGGAAGCC	JUN	QSY	Taqman
Mx	CACTGCAACAAGCAAAGAAGGA	TGATCAACCCCACAAGGAAAA	ACAAAGCACACACCCAACTGTCAGCG	FAM	TAM	Taqman
IL-6	GCTCGCCGGCTTCGA	GGTAGGTCTGAAAGGCGAACAG	AGGAGAAATGCCTGACGAAGCTCTCCA	FAM	TAM	Taqman

2.1.3.3 end-point PCR primers

Gene	Primer name	Primer sequence	5'-3' Orientation
ATHL	ATHL_F_11870	CTGCTCTGTACAATTCCAGGAG	Forward
IFITM5	IFITM5_F_500	CTGAGGCTGGGCTGGAGA	Forward
IFITM5	IFITM5_R_1100	ACCCCATGGGCACCCCTA	Reverse
IFITM1	IFITM1_F_1140	GAGAGGGCCTCAAAACACAAAG	Forward
IFITM1	IFITM1_R_600	GGATGAGATAGAATGGGACGGAG	Reverse
IFITM2	IFITM2_R_960	GGGTTGGAGGATCACGTTGT	Reverse
IFITM2	IFITM2_F_ 2	ACTGTCCCACGCAGATAGGT	Forward
IFITM2	IFITM2_F_ 290	TGAGCTCTGGGAGCCCG	Forward
IFITM3	IFITM3_F_15	ACCAGTTCCCTCCGTCCTCCC	Forward
IFITM3	IFITM3_R_620	TCCACAGGATTCTGTGGGGTCCACG	Reverse
B4GALNT	B4GAL_R_31	CAGCTCTTATAGCTGTCCCACCG	Reverse

2.1.4 Antibodies

2.1.4.1 Flow cytometry

- Anti-influenza A Virus Nucleoprotein Mouse Monoclonal Antibody [C43] (#REF ab128193, Abcam)
- Goat anti-mouse IgG2a Cross-Adsorbed Secondary Antibody, Alexa-Flour 647 (#REF A-21241, Invitrogen)

2.1.5 Media and Buffers

2.1.5.1 Media

DF-1 growth medium

- DMEM (Gibco #REF 41965039)
- 10% FBS (Gibco #REF 10099-141)
- 1% L-Glutamine (Sigma #REF G7513)

PGC growth medium

- Knock-Out DMEM (Gibco #REF 10829018)
- 10% Embryonic stem-cell FBS (Gibco #REF 16141079)
- 1% chicken serum (Gibco #REF 16110082)
- 1x Non-Essential Amino Acids (NEAA) (Gibco #REF 11140050)

- 1x Glutamax (Gibco #REF 35050061)
- 1x Conalbumin (10mg/ml stock) (Sigma #REF C7786)

MDCK growth medium

- DMEM (Sigma #REF D6429)
- 10% FBS (Gibco #REF 10099-141)
- 1% Penicillin/Streptomycin (Gibco #REF 15140122)
- 1% L-Glutamine (Sigma, #REF G7513)

Plaque overlay medium

- 10x EMEM 100ml (Sigma #REF M0275-100ML)
- 7.5% BSA (Sigma #REF A8412-100ML)
- 1% L-Glutamine (Sigma, #REF G7513)
- 4 % of 7.5% $NaHCO_2$ (Sigma, #REF S8761)
- 2% 1M HEPES(Sigma, #REF H0887)
- 1% DEAE-Dextran hydrochloride (Sigma #REF D9885)
- 1% Pen/Strep Penicillin/Streptomycin (Gibco #REF 15140122)
- Distilled water to final volume
- 1-2ug/ml (final) TPCK-trypsin (Sigma #REF T8802)

Infection medium (DMEM, TPCK, BSA)

- DMEM (Gibco #REF 41965039)
- 0.3% BSA (Sigma #REF A9576)
- 0.1ug/ml TPCK-trypsin (Sigma #REF T8802)

Freezing medium (DMEM, FBS, DMSO)

- 10% DMEM (Gibco #REF 41965039)
- 80% FBS (Gibco #REF 10099-141)
- 10% DMSO (Sigma #REF D4540)

2.1.5.2 Buffers and solutions

PBS"A" (Ca/Mg free) (Central Research Services at the Pirbright Institute)

TAE buffer (50x TAE, ThermoFisher REF #B49)

FACS buffer

- 1% BSA in PBS"A" (Sigma #REF A7906)

SFACS buffer

- 1% BSA (Sigma #REF A7906)
- 0.5% saponin (Sigma #REF 47036)

Crystal Violet solution

- 0.1% Crystal Violet (Sigma #REF HT901-8FOZ)
- 20% Methanol (VWR Intenational #REF K977-1L)

0.25% Trypsin EDTA (Sigma #REF T4049)

2.2 Methods

2.2.1 Cell culture

2.2.1.1 Culturing, passaging, freezing

All cell culture techniques were performed in a class II biosafety cabinet, unless otherwise stated.

Resucitation

Cell lines listed in the Table 1 were revived from liquid nitrogen storage by flash thawing in a water bath at 37°C. They were subsequently resuspended in a generous amount of growth medium to remove DMSO and centrifuged for 5min at 300xg. Supernatant was discarded and cells resuspended in 5ml of the growth medium. They were seeded into 25cm^2 tissue culture flasks and incubated at 37°C and 5% CO_2.

Passaging

When the confluency was reached, the growth medium was removed. The cells were rinsed with PBS"A" and detached from the flask using 0.25% trypsin-EDTA for 5-10min at 37°C. Bright-light microscope was used to confirm the release of the cells from the flask surface upon which the trypsinisation was stopped by addition of serum containing growth medium. The cell suspension was subsequently centrifuged for 5min at 1200 rpm, supernatant removed and the cells resuspended in a fresh aliquot of the growth medium. 10ul of the cell suspension was used to count the cells and the appropriate number of cells was calculated to be plated into new flasks according to the desired confluency.

Cryopreservation

Cryopreserved stocks were maintained for all cell lines used in this work. Cells were harvested from confluent flasks as previously and counted. Aliquots of 10^6 cells per 0.5ml of freezing medium were distributed into cryovials suitable for liquid nitrogen storage (Nalgene Cryoware, Fisher scientific, UK). The vials were frozen at -80°C for 24h in a Nalgene Mr Frosty freezing container containing 100% isopropanol (VWR international, Lutterworth, UK) followed by long-term storage in the vapour phase of liquid nitrogen.

2.2.1.2 *Cell counting*

In preparation for plating, DF-1 cells were dissociated from the flask with a 0.25% trypsin-EDTA solution (Life technologies) and PGC cells in TrypLE Express. Cells were resuspended in appropriate medium and 10µl of the cell suspension was used for counting on BioRad TC20 automated cell counter according to the machine instructions. The size of total population of cells was calculated accounting for the dilution factor used.

2.2.1.3 *Cell growth kinetics*

Cells were seeded in 24-well plates at doubling densities starting from 5000cells/well in replicate wells and imaged real-time using the IncuCyte Live-cell Analysis System (Sartorius, UK) in 1h intervals to monitor the growth. The software calculates the increasing ratio of well surface occupied by the growing cells over the appropriate interval of time and computes the speed of cell doubling.

2.2.2 Virology techniques

2.2.2.1 *Propagation of virus in embryonated chicken eggs*

Embryonated chicken eggs (VALO GmbH, Germany) were incubated at 37°C, 40-50% humidity, for 10 days prior to the inoculation of the influenza virus. Viability was checked by candling at the Animal services at the Pirbright Institute. The procedure was conducted following internal SOP of the Pirbright Institute SHAC-MA-01, but here I provide the summarised steps. Virus to be inoculated was diluted to 10_4 pfu/ml in PBS. Eggs were sprayed with 70% (v/v) ethanol and punctured just above the line of the air sac. 100µl of diluted virus stock was inoculated into the allantoic cavity at a 45° angle using 1ml syringe with a 22G needle. Autoclave tape was used to seal the puncture. The inoculated eggs were incubated at 37°C, 40-50% humidity, on the egg turner for 72h after which the embryos were culled by chilling overnight at 4°C. The allantoic fluid was harvested, clarified by centrifugation at 3000 rpm, 30 min at 4°C, aliquoted into 1ml cryovials, 24abelled and stored at -80°C.

2.2.2.2 *Propagation of influenza virus in Madin-Darby Canine Kidney cells (MDCK)*

MDCK cells were grown to 80% confluency in cell culture flasks. Upon removal of the growth medium on the day of the experiment, the cell layer was washed with sterile PBS "A". Virus dilution was prepared in the infection medium to yield the MOI of 0.01 and dispensed over the washed cells. Flasks were incubated at 37°C and 5% CO_2 for 1 hour, and shaken every 20 minutes. Subsequently virus was removed, cells washed in PBS"A" and incubated in infection medium containing 2ug/ml TPCK for the duration of the experiment. Cells were monitored under the light microscope daily for CPE. After two to three days, depending on the healthy appearance of the cells, the supernatant was collected and centrifuged for 10min at 1200 rpm to remove the debris. Aliquots of the virus suspension were titred using the HA assay on the same day. Upon a satisfactory result they were frozen at -80°C for long term storage. Virus stock was always titred anew before experimentation using a plaque assay to assess the extent of viral particle decay caused by long-term exposure to freezing temperatures.

2.2.2.3 Haemagglutination assay (HA)

Haemagglutination assay (HA) was used to estimate the titre of a virus stock. Using U-bottom 96-well plates, a volume of 50uL of a two-fold serial dilution of the allantoic fluid in PBS”A” was further serially diluted two fold in 50ul of 1% chicken red blood cells (RBCs) suspension in PBS”A”, obtained from the Pirbright Institute Animal Facility. Each sample was assayed in triplicates. Plates were gently shaken to ensure mixing and left covered for 30min at room temperature. Titre readout was expressed as the dilution of the last well where hemagglutination was observed in form mobile lattice of red blood cells on the suspension surface. Settled red blood cells at the bottom of the well in form of a “button” indicated no haemagglutination occurred.

2.2.2.4 Infection assays

24h prior to infection the cells were seeded at 80-90% confluency. On the day, the plates were washed once with sterile PBS to remove any residual serum. Virus inoculum was diluted in the appropriate volume of the infection medium according to desired MOI and added on top the cell layer. Cells were incubated at 37°C and 5% CO_2 for 1 hour, and shaken every 20 minutes. Upon the end of the incubation the virus was removed and cell monolayer washed twice with sterile PBS”A”. In case of single cycle infections appropriate volume of maintenance media was added to each well and the plates were incubated at 37°C and 5% CO_2 without shaking for the time indicated. For the multicycle infections the protocol was the same with the exception of the medium, which in those cases was substituted by the infection medium.

2.2.2.5 Influenza virus plaque assay

Titration of Influenza virus was carried out in Madin-Darby Canine Kidney cells (MDCK). Cells were grown to ~90% confluency in 12-well plates. Virus material to be assayed was prepared by 10-fold serial dilution of the allantoic fluid in cold serum-free DMEM medium. At the time of infection the cells were washed twice with PBS”A” (Ca/Mg free) and overlaid with virus in the volume of 100ul/well in duplicates. Plates were incubated for 1h at 37°C and rocked every 20min. Following the incubation the inoculum was removed, the cells were washed with PBS”A” twice and coated with the overlay gel containing 2µg/ml of TPCK trypsin (Sigma) made from 2% agarose and infection medium in ratio 30%-

70% v/v. 72h post infection the solid overlay was removed by flicking and plaques were visualised using 0.1% crystal violet stain. Plaque forming units (PFU/ml) were counted and averaged and the titration calculated as the mean of pfu divided by the product of the inoculum volume and the dilution factor of the virus solution used in the infection.

2.2.2.6 Multiplicity of Infection calculation

Plaque forming unit (PFU) is a measure of the number of functional infectious viral particles used in the calculation of the multiplicity of infection (MOI). MOI is the critical parameter in the virus infection studies used to estimate the likely number of virus particles targeting the individual cell expressed as the ratio of the PFU and the number of cells per well. In this work the volume of the virus dilution to be used to achieve the desired MOI was calculated using the below formula $V(virus) = \frac{MOI * cell\ number}{PFU\ (virus)}$

2.2.3 Molecular biology techniques

2.2.3.1 DNA extraction

To dissociate the cells from the flask, they were incubated in 0.25% trypsin-EDTA solution (Life technologies) at 37°C for 5-10min. The reaction was stopped using complete medium. The cell suspension was collected into a tube and centrifuged at 1200rpm for 5 minutes. The pellet was resuspended in 200µl of PBS containing proteinase K. Total DNA was extracted using the Dneasy® Blood & Tissue kit (Qiagen) according to the manufacturer’s instructions. The DNA was eluted in 100µl of sterile Dnase-free water and was stored at -20 °C. DNA quantification was performed on a Nanodrop Lite (ThermoFisher).

2.2.3.2 Polymerase chain reaction (PCR)

End-point PCR was performed using Q5® High-Fidelity DNA Polymerase (NEB, ref #M0491) according to manufacturer’s instructions. Briefly, a reaction mix was prepared from 5X Q5 reaction Buffer, 5X GC Enhancer, 10mM dNTP mix, alongside of 10uM forward and reverse primers and the Q5 High-Fidelity DNA Polymerase. Template DNA of mass ranging from 50 to 100ng in appropriate volume was finally added and the reaction was topped with Nuclease-free water to final volume of 25uL. PCR was run according to the following cycling conditions: initial denaturation for 98°C for 30sec, followed by 35 cycles of denaturation for 98°C for 10sec, annealing [95°C for 1 min, 50°C for 1 min, 72°C for 1 min/kb] then 75°C for 15 mins, where annealing temperature and the extension time were adjusted according to the primers used and product size expected.

2.2.3.3 *Agarose gel electrophoresis*

1% weight/volume agarose gels in TAE were used to separate DNA products of the PCR. Electrophoresis was run at 100V for approximately an hour. DNA bands ware visualised using GelDoc^{TM} EZ Imager (BioRad) and their sizes estimated relative to QuickLoad 1kb+ DNA Ladder (NEB).

2.2.3.4 *PCR product purification and gel extraction*

PCR products used for Sanger sequencing were purified from the agarose gels using QIAquick Gel Extraction Kit (Qiagen, #REF 28706X4). DNA was eluted in 30uL of elution buffer that comes with the kit and its concentration quantified with NanoDrop^{TM} spectrophotometer (Thermo Scientific).

2.2.3.5 *RNA extraction*

RNA was extracted using RNeasy kit (Qiagen, #REF 74106) according to manufacturer's instructions. Cells were released from the surface using RLT buffer that comes with the kit. Vigorous pipetting up and down was used to aid the dissociation with mechanical force. Dissolved cellular material was stored at -80 °C until extraction, and never longer than three months. All extractions included DNase treatment. RNA concentration was estimated using NanoDrop Lite Spectrophotometer (Thermo Scientific) before further manipulation.

2.2.3.6 *RNA quantification with Qubit*

In instances of scarce material or need for highly precise quantification of RNA the Qubit™ RNA HS Assay Kit (Invitrogen™, #REF Q32852) was used. Briefly, RNA samples and RNA standards (used for calibration of the fluorometer) were diluted in Qubit® assay tubes (Invitrogen™, #REF Q32856) using Qubit® working solution prepared according to the protocol supplied by the manufacturer. All samples and standards were incubated for exactly 2 minutes at room temperature. RNA concentrations were calculated using the below formula:

$$c(RNA) = QF\ value \times \left(\frac{200}{A}\right) x\ DF$$

Where, QF value = the value generated by the Qubit® 2.0 Fluorometer
A = sample volume
DF = Dilution factor for RNA

2.2.3.7 *RNA quantification with TapeStation*

RNA ScreenTape Assay (Agilent, #REF 5067-5576) was used for the assessment of RNA quality. Samples were prepared by mixing 5 µL RNA Sample Buffer (Agilent, #REF 5067-5577) with 1 µL of RNA, as per manufacturer's protocol, subsequently spinned and vortexted for 1min on IKA vortexer. Spinned samples were incubated at 72 °C for 3min followed by 2min on ice. After the additional spin they were loaded onto the 2200 TapeStation instrument for the electrophoresis.

2.2.3.8 cDNA synthesis

cDNA for RT-qPCR was synthesised from total RNA with the SuperScript™ IV First-Strand Synthesis System (Invitrogen™, #REF 18080051) according to the manufacturer's instructions, using random hexamers. Concentration of the resulting cDNA was quantified on NanoDrop Lite spectrophotometer as before and stored at -20°C.

2.2.3.9 Quantitative RT-PCR (TaqMan)

Quantitative RT-PCR was performed with TaqMan Multiplex Master Mix (Applied Biosystems, #REF 4461882) in duplex reactions (optimised in the Avian Genomics group): *IFITM1* and *IFITM2*, *IFITM3* and *IFITM5*, *IFNa* and *IFNb*, and singleplex *IL6* and *Mx* (for primers and probes see **Table 3**) as well as all reference genes from **Table 2**. Briefly, a gene-specific 20x assay was prepared from primers (50uM, final concentration 0.5uM) and a probe (10uM, final concentration 0.25uM) in appropriate volume and mixed with multiplex-specific master mix and template cDNA topped with ddH_20 to final volume of 20uL. For duplex reactions, water volume was reduced by the volume of suitable additional gene-specific 20x assay mix. All samples were prepared in technical triplicates. Plates were sealed and ran on a 7500/7500 Fast Real-Time PCR Systems (Life Technologies). Thermocycling conditions included enzyme activation step of 20sec at 95°C and 40 cycles of denaturation of 3sec at 95°C and annealing of 30sec at 60°C. Ct values were calculated using QuantStudio 5 Software v1.4.3., normalised against those of the reference genes and exported to MS Excel where fold change in expression was calculated using ΔΔCt method.

2.2.3.10 Reference gene stability analysis using geNorm algorithm

A panel of 10 reference genes suitable for avian transcriptomic research [81] was tested in quantitative RT-PCR to determine reference genes to be used in this project (see the list in **Table 2**). Ct values were calculated using QuantStudio 5 Software v1.4.3, same as before. Excel plugin for geNorm algorithm [82] was employed to calculate the M value representing the gene stability and V value indicating the recommended number of reference genes to be used per assay.

2.2.3.11 Flow cytometry for quantification of viral infection

Cells from the infections studies in cell culture were used to assess the cell line permissiveness to influenza virus by flow cytometry. To ensure no loss of cells; the supernatant, the wash and the adhered cells, detached after trypsin treatment, were pooled together for flow cytometry analysis. The supernatant was collected and cells washed once with PBS”A”. Appropriate volume of TrypLE Express (GIBCO REF 12605-028) was dispensed into each well and the plate incubated at 37°C until cells detached from the plate surface. Cell suspension was aspirated and pooled with collected supernatant and the PBS”A” wash. Samples were centrifuged at 300g for 5min. Following centrifugation, the supernatant was discarded and cell pellet resuspended in 180uL PBS”A” and transferred to a well of a U-bottom 96-well plate. LIVE/DEAD™ Fixable Violet Dead Cell Stain (Invitrogen, REF L34963) was prepared according to manufacturer’s instructions and 20uL was added per well and thoroughly mixed. The plate was left covered with aluminium foil, at room temperature. After 30-min incubation, the samples were centrifuged 300g for 5min at room temperature. Infectious supernatant was disposed of inside the cabinet by flicking, followed by another wash with 200uL PBS”A”. The plate was centrifuged as previously to remove the remnants of the stain and the PBS”A”. Cells were then resuspended in 100ul/well of BD Cytofix/Cytoperm™ solution (BD, REF 554722) containing 4.2% formaldehyde and incubated in the dark at room temperature for 20min. Following the fixation the plates were centrifuged and supernatant removed as before. Cell pellets were resuspended in 100-200ul/well of PBS”A” and subjected to another round of centrifugation and supernatant removal. In some instances at this point the cells were kept at 4°C overnight for convenience and stained the following day after an additional washing step. Otherwise, the cells were stained immediately. Cells used as controls were resuspended in FACS buffer containing 1% bovine serum albumin (BSA) and 0.5% saponin in PBS”A” (SFACS buffer) and infected samples were stained with a primary antibody; Anti-influenza A Virus Nucleoprotein mouse monoclonal antibody [C43] (ab128193, Abcam), in dilution 1:1600 of FACS buffer at 50ul/well and left in the dark at room temperature for 45min. Upon incubation the plates were centrifuged as before, wells checked for pellets and supernatant removed by flicking. After two additional washing steps, cell pellets were resuspended in the solution of the secondary antibody (Goat anti-mouse IgG2a Cross-Adsorbed Secondary Antibody, Alexa-Flour 647, from Invitrogen, #REF A-21241) diluted in ratio 1:1000 in SFACS. Plates were incubated in the dark, at room temperature for 30min after which they were centrifuged. All the samples were taken through two subsequent washing steps described previously followed by the final resuspension in 100ul of PBS”A” and transfer into 1.2ml FACS tubes containing 300ul PBS”A” for flow cytometry measurement on LSR Fortessa (BD Biosciences). DIVA v9.0.1 software was used for sample processing and FC Express 7 (De Novo Software, USA) for the analysis. Results were further statistically analysed and visualised in GraphPad Prism v8 [83].

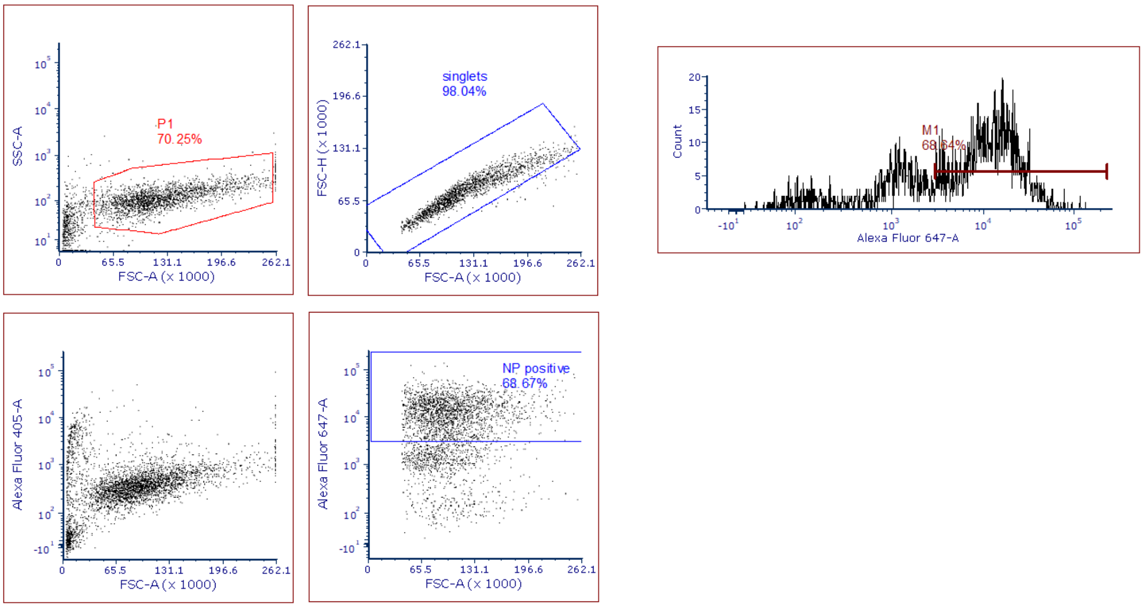


Figure 8 ***Outline of the gating strategy used for the detection of NP protein in influenza infected cells*** Minimum 10 000 events ware collected for a representative sample of parental WT DF-1 cells on a Fortessa machine. SSC-A vs FSC-A gating was used to exclude debris. Gated cells were then gated for FSC-H vs FSC-A for single cells and Alexa-405 for live cells. Final plot depicts the gating for NP protein stained with Alexa-Flour 647, facilitating the enumeration of infected cells.

2.2.4 Statistical analysis

Statistical analysis was performed by me using GraphPad Prism 8 software [83] or R, as convenient. The standard error was reported when only technical replicates were available, otherwise bars represent standard deviation estimated from data obtained from biological replicates. p-value <0.05 was considered statistically significant. Degrees of significance are indicated with stars: *p<0.05, **p<0.01, ***p<0.001 and ****p<0.0001.

2.2.5 RNA-Seq

Two RNA-Seq experiments were conducted during this study. First contained WT-DF1 samples discussed in 3.3.2. and was conducted by Dr Graham Freimanis and myself at the Sequencing facility at the Pirbright Institute. The second sequencing experiment was conducted at the sequencing facility at the Earlham Institute; samples included in this run are discussed in Chapters 4 and 5, specifically Sections 4.3.1.4 and 5.3.3.

2.2.5.1 RNA-Seq at the Pirbright Institute

RNA was extracted from samples using the RNeasy kit (Qiagen) as per kit instructions and the quality and concentration evaluated using RNA ScreenTape on TapeStation 2000 (Agilent) following manufacturer's manual. 1ug of RNA per sample was taken further for QC prior to sequencing.

RNA-Seq Library preparation was carried out with the help of Dr Graham Freimanis at The Pirbright Institute Sequencing Unit. Total RNA was checked for quality using a Bioanalyser 2100 RNA Pico kit (Agilent, #REF 5067-1513). Concentrations were estimated using a Qubit (Life Technologies) and normalised for a 200 ng input per sample for enrichment. Total RNA was enriched for mRNA using the NEBNext® Poly(A) RNA Magnetic Isolation Module (NEB #REF E7490L). Libraries were prepared from the mRNA enriched samples using a NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (NEB, #REF E7760L). Quality and concentrations of generated libraries were assessed with Bioanalyser 2100 DNA 1000 kit and Qubit, prior to pooling. Library pools were quantified using a NEBNext® Library Quant Kit for Illumina® (NEB, #REF E7630S). Equimolar quantities were sequenced on a High output (150 cycles) NextSeq 500/550 v2.5 flow cell (Illumina, #REF 20024907). Approximately 30 million 75 bp PE reads were generated per sample.

2.2.5.2 RNA-Seq at the Earlham Institute

RNA samples were extracted from cell culture by myself using the RNeasy kit (Qiagen) following manufacturer's instructions and eluted in water. I assessed the RNA quality using RNA ScreenTape and corresponding reagents on TapeStation 2000 (Agilent) as per protocol recommended by the manufacturer. Aliquots containing approximately 1µg of RNA were shipped to Earlham Institute for further processing.

Library preparation was performed on the Perkin Elmer (formerly Caliper LS) Sciclone G3 (PerkinElmer PN: CLS145321). 1µg of RNA was purified to extract mRNA with a Poly(A) mRNA Magnetic Isolation Module. The technology is based on the coupling of Oligo d(T)25 to paramagnetic beads which is then used as the solid support for the direct binding of poly(A)+ RNA. Isolated mRNA was then fragmented for 12 minutes at 94°C, and first strand cDNA was synthesised. This process reverse transcribes the RNA fragments primed with random hexamers into first strand cDNA using reverse transcriptase and random primers. The second strand synthesis process removes the RNA template and synthesizes a replacement strand to generate ds cDNA. Directionality is retained by adding dUTP during the second strand synthesis step and subsequent cleavage of the uridine containing strand using USER Enzyme (a combination of UDG and Endo VIII). NEBNext Adaptors were ligated to end-repaired, dA-tailed DNA. The NEBNext Adaptors with novel hairpin loop structure are designed to ligate with high efficiency and minimize adaptor-dimer formation. The loop contains a U, which is removed by treatment with USER Enzyme to open the loop and make it available as a substrate for PCR. The ligated products were subjected to a bead-based purification using Beckman Coulter AMPure XP beads (A63882) to remove most of the un-ligated adaptors. Adaptor Ligated DNA was then enriched by receiving 10 cycles of PCR (30 secs at 98°C, 10 cycles of: 10 secs at 98°C _75 secs at 65°C _5 mins at 65°C, final hold at 4°C). Barcodes are incorporated during PCR using NEBNext Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs) thereby allowing multiplexing. The quality of the resulting libraries was determined using a Perkin Elmer DNA High Sensitivity Reagent Kit (CLS760672) with DNA 1K / 12K / HiSensitivity Assay LabChip (760517) and the concentration measured with a Quant-iT™ dsDNA Assay Kit, high sensitivity (Plate Reader) assay from ThermoFisher (Q-33120). The final libraries were equimolar pooled, a q-PCR was performed and the pool was sequenced on a one Illumina NovaSeq 6000 S4 flow cell lane with 150bp read pairs.

2.2.5.3 RNA-Seq data analysis

I analysed all the RNA-Seq data discussed in this study, using the HPC facility at the Pirbright Institute. Briefly, demultiplexed reads were downloaded from Illumina's cloud BaseSpace and assessed for quality using FastQC from Babraham Institute [84]. Reads were then aligned using STAR [85] to a reference genome *Gallus gallus* GRCg6a downloaded from Ensembl [86,87] by itself or concatenated with H5N3 influenza sequences. Read quantification was accomplished using RSEM [88] or *featureCounts* function of the *Rsubread* package [89], as appropriate. Quality of the alignment was evaluated with Qualimap [90,91] and all reports were summarised with MultiQC [92].

Differential expression analysis was carried out in R [93] using DESeq2 [94]. Unless otherwise stated, for accurate fold change estimation *ashr* (adaptive shrinkage) method was used [95] because it accommodates contrasts used in the design formula for GLMs. Additionally, effect size shrinkage is more robustly estimated because the *ashr* method takes as input two parameters: calculated effect size and associated standard error, in contrast to the other methods that take only one (e.g. either p-value or a Z-score). Non-informative multiple testing was reduced by omitting all genes that had less than 10 reads across all samples. In case of DESeq2, an independent filtering step is additionally included to further reduce the burden of multiple testing.

2.2.5.4 Transcript Quantification

Transcript abundance quantification was done with RSEM [88] to make use of its flexibility in terms of input, accuracy achieved by its algorithm and the convenient output in TPMs. The newest software releases include modelling of the paired-end and variable-length reads as well as considering the quality scores if the reads are provided in fastq file format, a feature that was used in this analysis. Accurate estimation of transcript abundance involves calculating the steady-state quantities of individual transcripts which is proportionally more difficult for low expression transcripts due to the uncertainty in their quantification which stems from high signal-to-noise ratio typical for low counts. To deal with this, RSEM employs an expectation maximisation algorithm (EM). The EM algorithm considers alternative alignments and associated probabilities, not only the uniquely mapped reads as decided by the aligner. The statistical model of RSEM is better suited to use this additional information in its probability estimations to "decide" which is the most likely accurate alignment, rather than the aligners such as STAR. The output, counts in TPMs (transcripts per million), is valuable for comparison of expression levels of genes within a sample. Note that the TPMs are not suitable for comparisons among samples. In such circumstances an additional normalisation step is required that takes into consideration the variability of expression level across all samples, e.g. a Z-score.

Software used for differential expression calculation used as input raw counts quantification estimates, because it employs internal normalisation procedure. For those purposes, the *featureCounts* function of the *Rsubread* [96] R package was used, as mentioned previously in Section 2.2.5.3. Raw counts were input into DESeq2 [94] and normalised by accounting for RNA composition [97] as well as sequencing depth by calculating so called size factors and gene-wise dispersion [94]. Note that henceforth expressions sequencing depth and library size are used interchangeably.

2.2.5.5 Data transformations

Normalised read counts for each gene come in an extremely wide range of values spanning orders of magnitude. For an efficient visualisation and detection of trends in RNA-Seq data, data are transformed. Unsupervised clustering methods such as PCA work best on homoscedastic data, i.e. data that have same variance across their whole range of mean values. The DESeq2 package offers three approaches to transform the data to render them homoscedastic: normal $\log_2$ (*log*), regularised log transformation (*rlog*) and variance stabilising transformation (*vst*). Regularised log transformation is recommended when the range of library sizes among samples exceeds an order of magnitude as well as when the sample size per group is small (n<30), the latter of which is the case in this study. However, *vst* is more dynamic, less influenced by outliers and its algorithm is much more time efficient. Thus, both *vst* and *rlog* transformations were tested for every RNA-seq dataset in this work to establish which method renders the data more homoscedastic.

2.2.5.6 Exploratory Analysis

Various diagnostic plots are generated for the purposes of RNA-Seq QC. They normally show distribution of counts along the whole range of values among all samples. They can also include clustering and/or dimensionality reduction methods used to investigate variability among samples.

Dimensionality reduction

Typically, MDS (multi-dimensional scaling) and PCA (principal component analysis) are employed to detect sample outliers. MDS is a mapping class of methods that fits the data to a predefined number of dimensions (normally two) while preserving the distances. PCA is a narrower concept within the mapping analyses that also incorporates a factor analysis and preserves the covariance of data. It does so through retention of all dimensions and calculation of the loadings. Loadings are values that represent the contribution of a particular dimension to overall variability in the data. Thus, the singular value decomposition (SVD), a procedure used for dimensionality reduction, is conducted in MDS and PCA on Euclidian distances and loadings, respectively. Both methods are used to detect sample outliers which could negatively affect differential expression calculation and generate spurious results. To preserve as much information as possible (within the covariance of data), PCA was performed for the high level exploratory analysis of the RNA-seq data in this thesis.

Clustering

A selection of clustering methods along with various distance metrics are available for clustering. Clustering is used to extract maximum amount of information from the available data thus there are no strict rules for the selection of either the method or the metric. However, the majority of the software tools habitually used in transcriptomic research employ the complete linkage method for hierarchical clustering. More flexibility is afforded to the user in the selection of the distance metric. In this work, Euclidean distances (the default) were used in most analyses. Deviations from this approach are stated where appropriate.

2.2.5.7 Functional Analysis

Functional annotation can be performed with a variety of method broadly divided into three categories : ORA (Over Representation Analysis), Functional Class Enrichement (e.g GSEA – Gene Set Enrichment Analysis) and topology based methods (e.g. SPIA – Signalling Pathway Impact Analysis).

Selection of method to use depends on the question the experiment is designed to answer, available expertise to conduct the analysis, accessibility of tools and functionality of their dependencies, such as the databases that store information on feature classes one wishes to test. Many comparative studies were conducted to determine which method yields most reliable results [98]. However, such analyses are usually benchmarked on synthetic data, or the list of methods is not extensive and ultimately with the rapid development of new methods their utility becomes debatable.

ORA is the most widely available method, contained in many programs with some variations among them. It requires only a list of genes of interest and a list of background genes. It is useful for comparing results of published research where the raw data is not available. However, it is not of quantitative nature and the threshold used to generate the gene list (p-value, logFC) can severely affect the analysis. Lastly, perhaps most importantly, for statistical testing ORA assumes independence of each differentially expressed gene; an assumption which is biologically highly unlikely, thereby decreasing the accuracy of the result.

GSEA [99] and its derivatives address the quantitative aspect as well as qualitative. It does not require a background list of genes and it uses the list of all differentially expressed genes without the inclusion of any threshold. Rank file with all DEGs is used as input; it contains associated effect sizes for every differentially expressed gene.

Lastly, topology based methods (e.g. GGEA [100] or SPIA [101] are the third generation of functional analysis tests. They incorporate information of regulation of enriched pathways (activation or suppression) and the knowledge about interactions in the wider signalling network. These methods appear to be most promising and representative of reality, however they are not widely used due to the paucity of data, especially outside of human research, about the regulatory mechanisms or available networks.

2.2.5.7.1 GSEA

Given its superiority over ORA and the impracticality of application of topology based methods in avian research, GSEA was selected as the approach to analyse enriched pathways in this doctoral thesis. To strengthen the analysis and attempt to limit false positives, two approaches were used: classical GSEA [99] and PADOG [102].

Enrichment analysis for the GO biological processes was conducted in R employing GSEA [99,103,104] and utilising Gene Ontology [105–107] and Reactome [108–110] databases, projecting chicken gene identifiers onto human pathways.

In Chapter 3 clones are being compared. Detected differences are expected to be small, thus a strict cut-off for statistical significance was applied to ensure accuracy.: false discovery rate (FDR) was kept at 0.05. Benjamini-Hochberg multiple testing correction was applied in all analyses throughout the pipeline. Gene set range was set from 3 to 800. This range may seem large; however gene sets are based on human data and it is not clear how much information gets lost in the ontological processes with which chicken gene functions are inferred. For the sake of capturing as many relevant pathways possible, precision was sacrificed where necessary. This was balanced with strict FDR of 0.05. All visualisations were created in R with either *clusterprofiler* or *enrichr* package. Deviations from this approach are stated where appropriate.

2.2.5.7.2 PADOG

PADOG [102] – Pathway Analysis with Downweighing of Overlapping Genes retains all the advantages of GSEA with the additional benefit of proportionally down-weighting the computed score for genes that appear in more than one gene set. Unique genes represent an unambiguous signal and are scored higher. PADOG was implemented using the Reactome web application [111,112] with default parameters and the FDR of 0.05. At the time of writing, PADOG analysis does not support multifactorial design or samples sizes smaller than three. Instances where these constraints made some analyses in this thesis impossible are indicated in the text.

3

Clonal variability of WT DF-1

3.1 Introduction

Challenges in avian immunology are in great measure due to paucity of available reagents. Looking specifically at research into *chIFITMs*, the absence of suitable antibodies is a very conspicuous obstacle. A common approach used to bypass that issue is the use of tagged constructs, but this can yield misleading results at times[113]. Protein tags affect localisation of proteins [114], their function [115] and physical properties such as solubility or thir proclivity to aggregation [116]. Nevertheless, this remains an acceptable risk as antibodies against huIFITMs have been reported to cross-react [59,60,117] and anti-chIFITM antibodies being otherwise poor. During this project attempts to generate antibodies have been unsuccessful. It proved difficult to generate antibodies of the desirable type as well as binding specificity and avidity. This makes a transcriptomic approach attractive. It opens possibilities of a discovery of related mechanisms that could be followed up by functional studies that could offer learnings about IFITMs by proxy. It is an approach frequently taken in avian immunology research [74,118–120]. Analyses like these offer insight into the diversity of mechanisms hosts use in the response to infection and with the continuous improvements in technologies and decrease in costs, they represent an accessible avenue in the research into genetic functions and mechanisms.

Laboratory cell lines are frequently composed of mixed populations of cells giving rise to heterogeneity in desired phenotype (expression level, resistance to infection, growth kinetics, etc) which in turn generates mixed signal in detection, lower yield in production or irreproducibility of results in general. Cellular cloning is used to mitigate this issue and maximally reduce variability. Intrinsic tendency of cell lines towards heterogeneity is unavoidable and it usually develops in clones as well during their life span because of epigenetic changes, stochastic changes in their response to external stimuli and time in culture, laboratory or storage conditions. Attempts to control for this include limiting the shelf-life of cell lines, tightly controlling the laboratory conditions and stringent monitoring for change in phenotype among the clonal cell lines.

With the goal of analysing the effect of *chIFITM* knock-out in later chapters, which can have off target consequences on the transcriptome potentially different in every targeted cell, we employ cellular cloning of WT DF-1 cells to gauge their clonability and assess underlying variability in their transcriptomic profile during an influenza infection. This approach provides information about the "background" level of disruption to transcriptomic landscape while allowing modular design of bioinformatic analysis where every cell line can be analysed separately or heterogeneity can be modelled by pooling different clonal cell lines together.

The focus of this section is the characterisation of the innate immune response of WT DF-1 clonal cell lines infected with the influenza virus on the transcriptomic level. An RNA-Seq pilot study was designed to investigate following questions:

- Does the single cell cloning significantly affect the expression profile of WT DF-1 clonal cell lines? Can statistically significant and biologically meaningful differences be detected among the transcriptomic profiles of daughter cell lines? Can we identify any signatures that are in agreement with phenomena identified in the laboratory?
- Do the individual clones mount a significantly altered immune response to the influenza infection to that of the parental cell line? Are the main hallmarks of the DF-1 cell line preserved in all clones?
- What is the sequencing depth required for identification of the more nuanced differences in genetic expression among chicken origin cell lines? Does the viral transcription significantly disturb the host machinery in a way that would obstruct the goal of identifying differentially expressed genes?

Continuous cell lines are historically generated by induced immortalisation i.e. transformation using viral approaches (e.g. MDV or RSV), physical agents (e.g. UV) or oncogenic chemicals (e.g. methylcholanthrene). However, DF-1 (also known as UMNSAH/DF-1) is a cell line of spontaneously immortalised fibroblasts originating from endogenous-virus-negative 10-day old East Lansing line 0 eggs [121]. The absence of known and endogenous and exogenous viral sequences (e.g. avian sarcoma and leukosis virus) in their genome makes them suitable for many applications in avian research.

DF-1 are fast growing and exhibit a typical spindle-shaped fibroblast morphology (**Figure 9**). They are characterised by dysfunctional regulation of various pathways such as apoptosis [122], cell proliferation [118] and oxidative stress [119,123] and they possess high transfection efficiency [124]. DF-1 are also susceptible to infection by many viruses [120,124–130] as well as to the virus-induced death [131]. These properties make the cell line amenable to virus propagation, protein expression, retroviral research, vaccine production and many other purposes.

DF-1 cells have been passaged more than 300 times [119] and currently various lineages of different derivative karyotype are used in research. O'Hare and Delany found at least three ploidy levels: haploid (14%), diploid (78%), and tetraploid (8%) in the lineage they were using [132]. They also observed DF-1 had much longer telomeres in comparison to wild-type or transformed chicken cells. Genomic rearrangements can affect gene expression directly by altering gene-dosage or indirectly by changing structural organisation of genetic elements involved in the regulation of expression. In fact, many examples of dysregulated signalling pathways have been reported as characteristic of DF-1:

upregulation of SOCS1 18)[120], MnSOD [133], low basal expression of TLR3 [129,134], downregulation of catalase [122] to name a few.

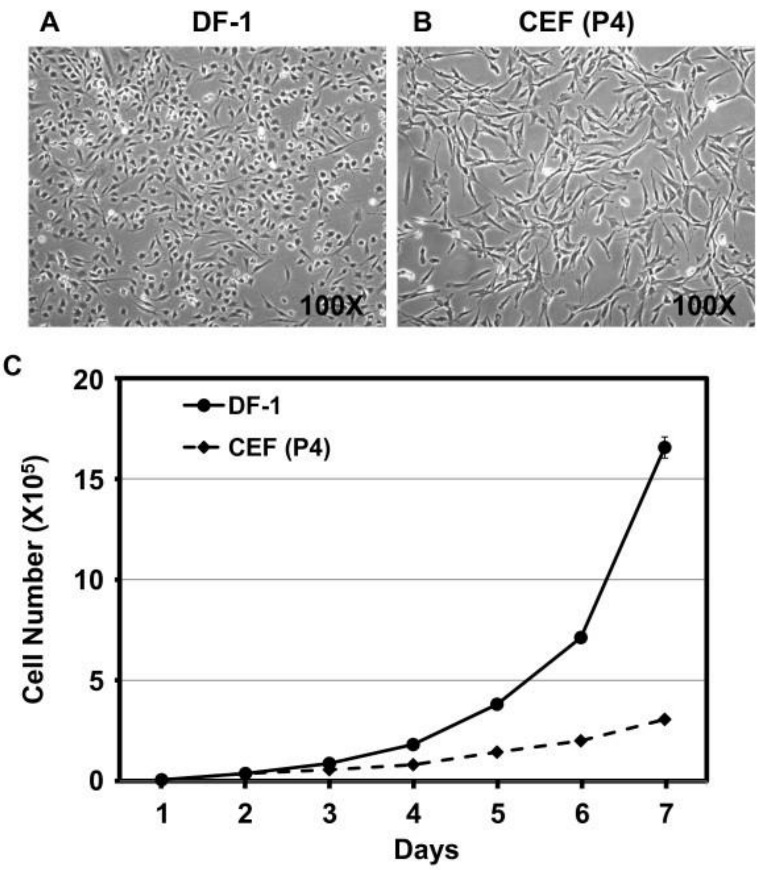


Figure 9 ***Comparison of the morphology and growth kinetics of DF-1 and CEF cell lines*** Both cell lines exhibit spindle-shaped morphology with DF-1 being smaller in size than CEF (A and B). DF-1 have comparatively much higher growth potential than CEF (C). Figure and legend taken from [119]

Due to the population heterogeneity, intrinsic and/or transient, derived lineages differ, sometimes greatly, from the parental line in their biological characteristics, chromosomal composition, and transcriptional profile. This phenomenon is even more frequent in cell lines derived through oncogenic methods (substances or viruses) [135,136]. It is not clear how much of this heterogeneity can be eliminated by cellular cloning. Nevertheless, single-cell cloning remains the most widely used approach due to its relative accessibility and the lack of alternatives. For this reason, it is important to characterise the effects single cell cloning has on the properties that have a significant impact on the immune response to influenza infection.

Subcloning is used as a method of decomposition of that heterogeneity to elucidate whether the subpopulations contain meaningful information or the averaged measurements of individual responses across the mixed population represent sufficient characterisation of the phenotype. It is not clear from the outset how many clonal cell lines would be sufficient to model DF-1 heterogeneity so for feasibility reasons and technical limitations we start with three. Subpopulations exist within the parent population on virtually all measured dimensions such as gene expression, cell cycle stage or other biological states. Heterogeneity becomes important when the deviation from the average measurement represents a change in the biological state being studied. Otherwise, it may just constitute cell-to-cell fluctuations irrelevant for the study of the parameter at hand [137]. This study focusses heavily on the effect of viral infection on the transcriptomic response and viral yield in different cells. Viruses heavily manipulate and influence the cell cycle progression and the activated pathways in each stage of the cell cycle, thereby affecting the parameters being studied. To ensure that the measured variability in the transcriptomic response and the virus yield is truly a reflection of heterogeneity in the transcriptomic response and not a stage of cell cycle, it would be important to synchronise the cell cycle.

Following sections detail the differences and similarities among the WT DF-1 cell lines and the parental DF-1 cell line in their susceptibility to infection by various strains of the influenza virus, ability to support viral replication and generate high viral titres, their homeostatic levels of *chIFITM* mRNA and their karyotypic profile.

Subsequently, a bulk RNA-seq experiment is described with detailed characterisation of transcriptomic profile at 6hpi with H5N3 subtype influenza virus aiming to relate the changes in phenotype to transcriptomic differences.

3.2 Objectives

- Identify and characterise phenotypic differences among clonal WT DF-1 cell lines
- Establish the experimental conditions for the investigation of transcriptomic profile during an infection with influenza virus:
    - Select the model influenza strain, required MOI and time point for RNA collection
    - Confirm suitable sequencing depth and the number of replicates required for robust detection of differentially expressed genes using bulk RNA-Seq
    - Devise a pipeline to be used for differential expression analysis
- Identify and describe the variability in the response to the influenza infection among clonal cell lines

3.3 Results

The parental cell line of DF-1, originally sourced from ATCC by the Genetics and Genomics group at the Pirbright Institute, was sent to Horizon Discovery, Cambridge, UK for single cell cloning. Three cell lines of the WT genotype (referred to as: WT DF-1) were banked, namely: 608, 609 and 610. Each cell line was subsequently expanded at the Pirbright Institute and aliquots frozen to be used in research.

3.3.1 Effect of single cell cloning on the DF-1 phenotype

Morphologically all cell lines resembled the parent in their fibroblast appearance, with WT DF-1 608 being slightly smaller in size and WT DF-1 610 having shorter doubling time (personal observation). All cell lines were equally sensitive to TPCK-trypsin (data not shown) and exhibiting consistent phenotype throughout the culturing which showed promise of genetic stability.

3.3.1.1 Cytopathic effect narrows down the selection to H5N3 as the model virus

The infection system to be used for generation of transcriptomic data was set up previously in the research group. Influenza was chosen as the model virus for its ecological importance in human and veterinary health. Dr Ahmed Ibrahim infected the cells at MOI 0.01 and 0.01 ug/ml TPCK-trypsin for multi-cycle infection and monitored the cells for CPE to determine the level of pathogenicity. Three different subtypes of influenza virus were tested: A/Chicken/Pakistan/UDL-01/08 (H9N2), A/duck/Ukraine/63 (H3N8) and A/duck/Sing-Q/119/97 (H5N3).

H5N3 subtype caused moderately paced CPE in all WT DF-1 clonal cell lines; a phenotype which was possible to correlate with the progression of infection via visual inspection (**Figure 10**). For this reason, H5N3 strain was used for subsequent studies except when the research question required use of more than one virus strain.

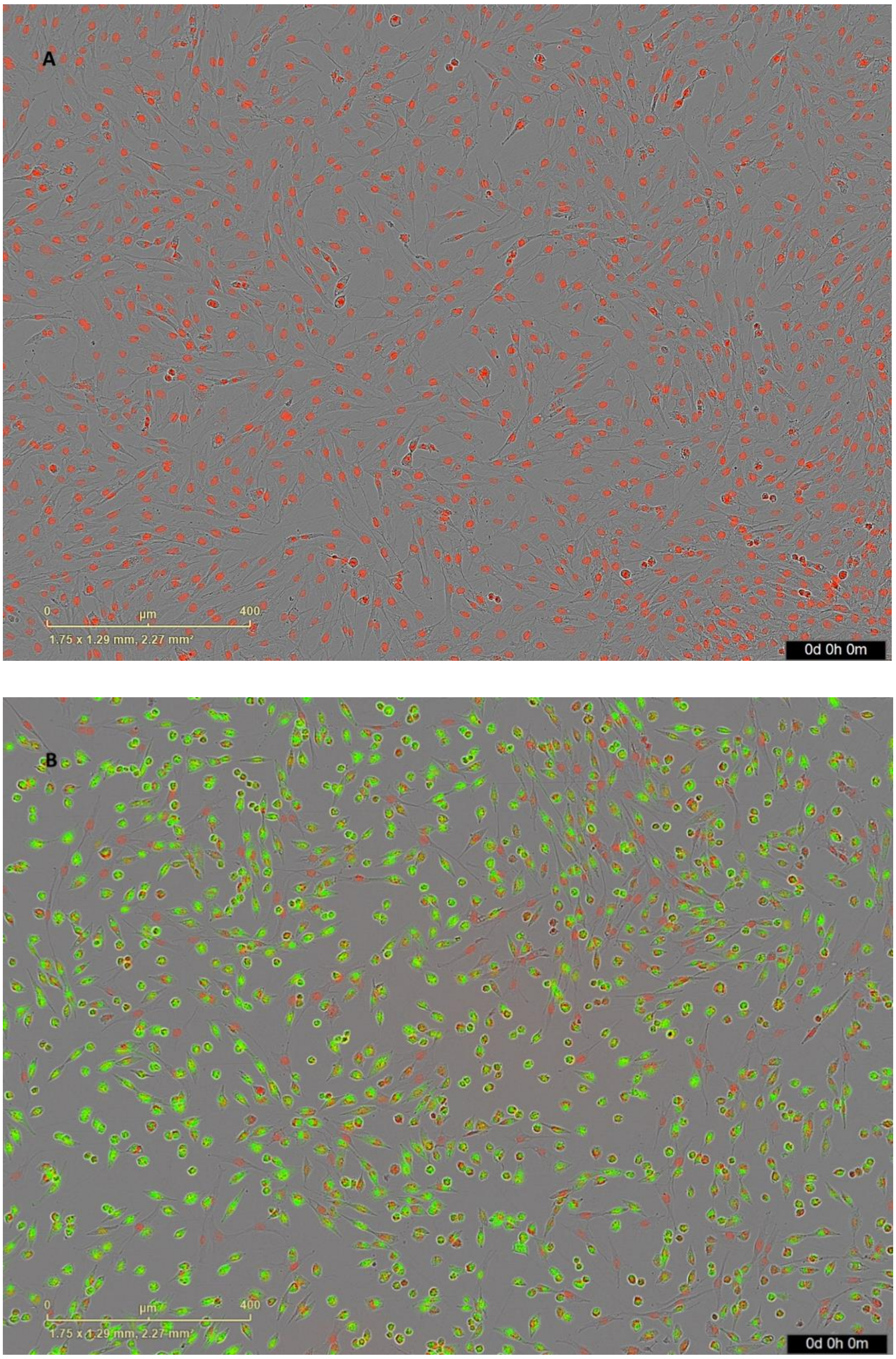


Figure 10 H5N3 shows moderate CPE in WT-DF-1 clonal cell line WT DF-1 clonal cell line 609 was infected with H5N3 influenza subtype and stained for surface expressed HA with Alexa fluor 488 (green). Nuclei are stained with TO-PRO-3 (red). A. Mock infected cells. B. Cells infected with A/duck/Sing-Q/119/97 (H5N3)

3.3.1.2 All single cell clonal cell lines are equally permissive to influenza virus

Viral yield in a specific cell substrate is dependent on many factors, among which is the ability of the virus to enter the host cell. Quantifying the permissiveness eliminates the uncertainty whether the differential ability of clonal cell lines to support efficient viral entry is the cause of lower virus yield in some cell lines. In addition to that, influenza viruses display different sensitivities to the host innate immune response, depending on the strain and its NS1 protein, which also has downstream consequences on the virus yield.

Desirable properties of a cell line includes the ability to support the life cycle of a pan-subtype range of influenza viruses, so I investigated the susceptibility of DF-1 WT clonal cell lines to the following influenza strains: A/Chicken/Pakistan/UDL-01/08 (H9N2), A/duck/Ukraine/63 (H3N8) and A/duck/Sing-Q/119/97 (H5N3). Cells were infected with three subtypes of influenza that were used in the experiment evaluating the cytopathic effect (section 3.3.1). Following the protocol delineated in 2.2.2.4., cells were infected at MOI of 1 and collected for analysis at 6 hours post infection. To confirm infection, cells were stained with an anti-NP antibody and analysed by flow cytometry following the protocol in 2.2.3.11.

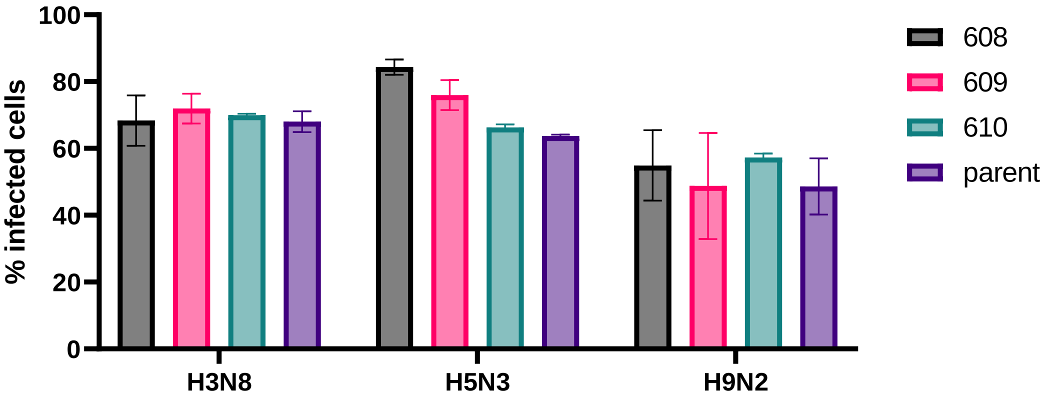


Figure 11 Permissiveness of WT DF-1 clonal cell lines Parental WT DF-1 and clonal cell lines were infected with A/Chicken/Pakistan/UDL-01/08 (H9N2), A/duck/Ukraine/63 (H3N8) and A/duck/Sing-Q/119/97 (H5N3 at MOI 1, 3 and 5 for 6h. Cells were collected, fixed, and analysed by flow cytometry for influenza virus NP staining. Bars represent ±1SEM. Experiment was conducted once in duplicate wells. Data were analysed using unpaired T-test with statistical significance denoted by * $p<0.05$, ** $p<0.01$ and ***$p<0.001$.

There was no statistically significant difference in permissiveness to the influenza among the WT DF-1 clonal cell lines and the parental cell line for any of the viruses tested. All subtypes appear to exhibit similar capacity to infect the cells (**Figure 11**), allowing us to remove this as a further consideration for the selection of the model virus for subsequent RNA-seq experiments. These data reiterated the results of the previous experiments conducted within the laboratory by Dr Alber and Dr Ibrahim (data not shown) so further replication was not required. In this way, we ensured that all the experiments concerning WT DF-1 clonal cell lines were conducted with the same stock of virus, thereby reducing non-informative biological variability.

3.3.1.3 Single-cell cloning affects cell line capabilities to support viral growth

It is not possible to infer viral yield capabilities of the clonal lineages with any certainty from the historical data available for the parental cell line. It is crucial that any estimations are generated from an appropriate data set. To address this in the context of the viral yield, a multicycle infectivity study was conducted in addition to plaque assay to titre the virus, following the same protocols referenced in the Materials and Methods chapter (2.2.2.4 and 2.2.2.5). As previously mentioned in the section 3.3.1.1, the cytopathic effect study showed H5N3 to be the optimal influenza subtype to visually monitor the progression of the infection under a light microscope. Additionally, the permissiveness studies confirmed that H5N3 subtype possesses a satisfactory ability to enter the cell (**Figure 11**). Thus, H5N3 subtype was selected for the multicycle infection experiment as well as experiments used to generate the RNA-seq data.

All the WT clonal cell lines: 608, 609 and 610 as well as the parental DF-1, were infected with A/duck/Singapore/3/1997 H5N3 at MOI of 0.01 and TPCK-trypsin concentration of 0.01ug/ml, as described (section 2.2.2.4). Supernatant fluid was collected at the 24h and 48h time point and titred with a plaque assay. Virus titration showed capacity of the WT DF-1 clonal cell line 608 to support higher levels of viral growth that the other clonal cell lines. Titres at 24h post infection (hpi) reached values in the range of 2 $\log_{10}$ and were statistically significant when compared with the parental cell line, with titres around 1.5 $\log_{10}$ PFU/ml (unpaired T-test, p value of <0.05 was considered significant). The same effect was observed for cell line WT DF-1 609, albeit to a lesser degree and without statistical significance (**Figure 12**). WT DF-1 610 clonal cell line, showed greater similarity to the parental line, retaining viral titres in the same range around 1.5 $\log_{10}$ PFU/ml at 24hpi and below 2 $\log_{10}$ PFU/ml for 48hpi; the difference was not statistically significant at any time point tested, most likely due to the variability between both cell lines, resulting in high standard deviation. For the clonal cell lines 608 and 609, we observed that the significance and the extent of the effect progressed over time: the viral titres increased by a $\log_{10}$ for both, reaching mean values of 2.5 $\log_{10}$ PFU/ml while 610 and the parent remained below 2 $\log_{10}$ PFU/ml at 48h post infection.

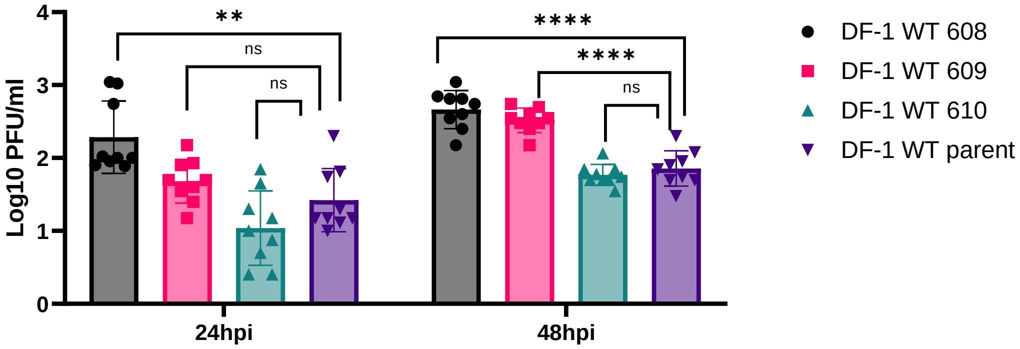


Figure 12 DF-1 cell lines differentially support the growth of avian influenza virus. Parent WT DF-1 and WT clones (608, 609 and 610) infected with A/duck/Singapore/3/1997 H5N3 at MOI 0.01. Supernatant fluid was collected at 24 and 48 hours post infection and assayed by plaque assay in MDCK cells. Four replicate experiments were conducted in duplicates. Data were analysed using unpaired T-test with statistical significance denoted by * $p<0.05$, ** $p<0.01$ and ***$p<0.001$. Data were kindly supplied by Dr Ibrahim.

3.3.1.4 Karyotyping of the Wild Type clonal DF-1 cell lines

As elaborated in the Chapter Introduction 3.1, the heterogeneity is a good predictor of genetic instability. To characterise chromosomal structure, deviations in their composition and organisation, KaryoLogic, Inc., NC, USA was contracted to assess ploidy during passaging as well as noticeable chromosomal irregularities.

The diploid set of chicken chromosomes contains 78 chromosomes: 38 autosomes and one pair of sex chromosomes, termed ZW and ZZ in case of female and male birds, respectively. Autosomes in birds have an unusual feature: a very wide range of sizes, spanning from less than 10Mb to more than 200Mb. International Chicken Genome Sequencing Consortium divides them into three groups: microchromosomes (<10Mb), intermediate choromosomes (10-50Mb) and macrochromosomes (50-200Mb) [138]. Microchromosomes make the complete karyotyping virtually impossible due to their very small size and poor banding, a technique used for identification of individual chromosomes. Thus, the analysis presented here focuses on the eight macrochromosomes and the sex chromosomes.

Approximately 20 spreads were analysed per passage per cell line. Two passages were analysed, one early (passage 14) and one late (passage 30) to monitor the karyotype dynamics as the cells age. The DF-1 parental cell line overall showed an atypical tendency towards diploidy as the number of passages increased, even though the number of macrochromosomes was not strictly 2n (diploid) across all spreads analysed. At passage 14 the spreads showed a mixture of equally represented diploid and tetraploid cells. At passage 30 diploid cells overtook the population almost entirely thereby signalling at some, at this time unidentified, selective advantage over tetraploidy for fitness in culture

(**Figure 13**). This behaviour was similar for clone 610, which consistently displayed characteristics most alike the parental line across many features (e.g. cell morphology, doubling time, virus yield). 608 had all the cells harbouring a tetraploid set of chromosomes at the later stage, as did 609 a diploid set. All cell lines maintained the same aberrations of chromosomes 1, 2 and 4 which did not change during passaging.

Overall, the parental DF-1 cell line exhibited very variable results in various parameters measured: permissiveness as determined by flow cytometry and viral load titred by plaque assay. This characteristic is retained in karyotyping as well. Karyological data confirmed that the widely used parental DF-1 cell line is a mixed population of aneuploidic cells. This genomic instability could be mitigated by single cell cloning as confirmed by the constant set of chromosomes that clones 608 and 609 harboured throughout their passaging. It is possible that the tetraploidy of the DF-1 clone 608 confers some of its observed features, such as the ability to support higher levels of viral replication, via elevated levels of expression of relevant genes. Further studies would be necessary to investigate this because karyotyping did not provide any information about the structure and the maintenance of the integrity of the microchromosomes throughout the life-span of the cell lines assayed. Nevertheless, these data are informative in the greater context of characterisation of the WT DF-1 clonal cell lines and serve as preliminary signpost of the origin of the variability observed in functional assays.

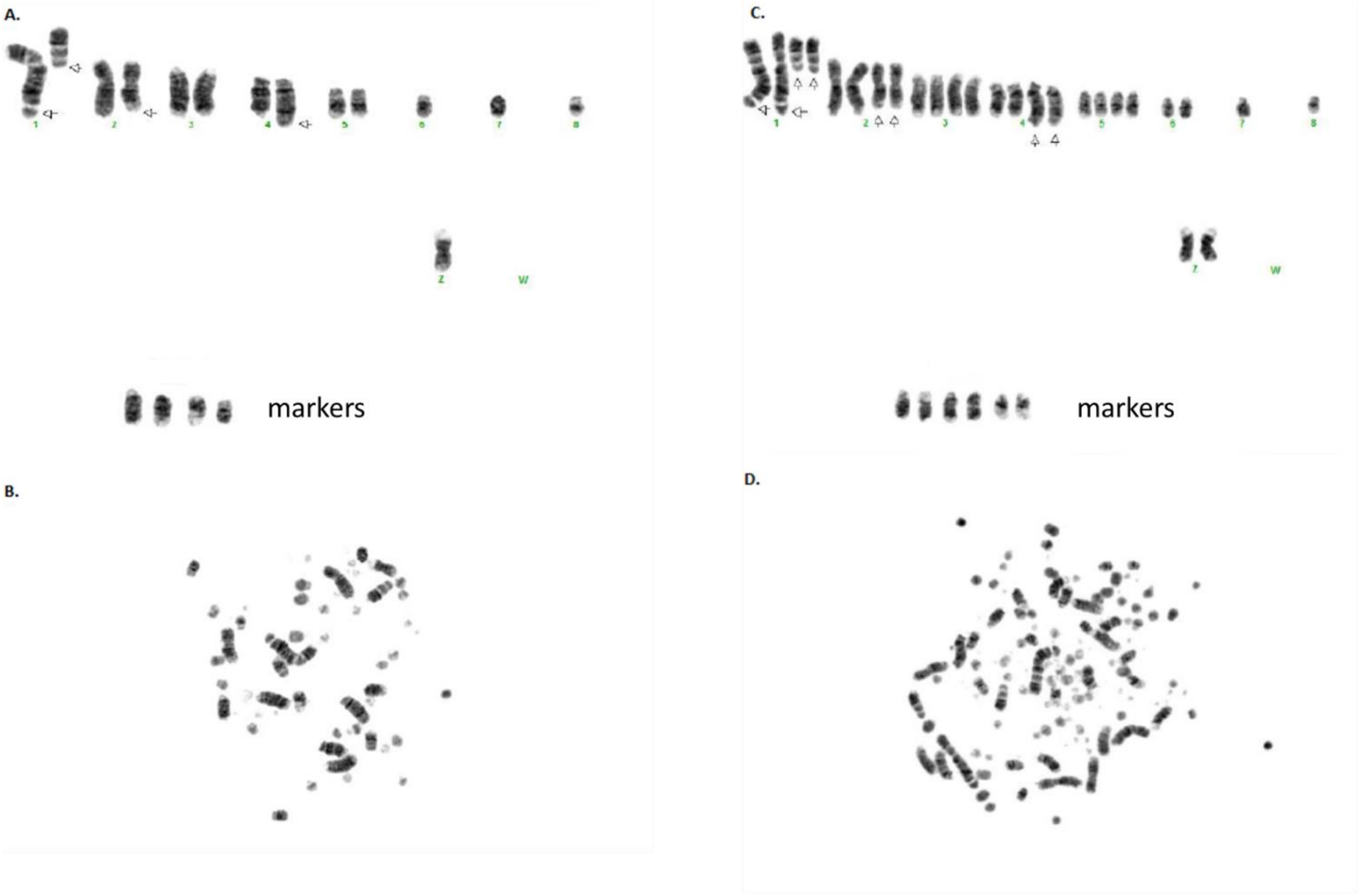


Figure 13 Karyotype of WT DF-1 parental cell line Cytogenetic analysis was performed on twenty G-banded metaphase spreads of the parental WT DF-1 at passage 14. Panels A and C show the karyotype of a full metaphase spread imaged in the B. and D. panels, respectively. Arrows indicate a chromosomal aberration (partial deletion of the q-arm) on one member of the chromosomal pairs 1, 2 and 4. A marker is defined as "a structurally abnormal chromosome that cannot be unambiguously identified or characterized by conventional banding cytogenetics."[139]

3.3.1.5 DF-1 clonal cell lines have equal homeostatic levels of expression of chIFITMs

Significant restriction of the influenza virus by IFITMs has been extensively demonstrated in many models [41,44,52]. Smith *et al* specifically looked at differences between ducks and chickens in their response to high (H5N1) and low (H5N2) pathogenic influenza infection *in vivo* [74]. They analysed lung and ileum samples by bulk RNA-seq and found that chickens only moderately upregulated their chIFITM expression [74]. Given that higher levels of basal expression of IFITMs render the host more refractory to infection, investigating if clonal cell lines exhibit the same levels of expression of IFITMs will offer insight into potential reasons for differential ability of clonal cell lines to support viral growth.

To investigate the basal levels of *chIFITM* gene expression, the counts retrieved from the RNA-Seq were compared to the quantitative RT-PCR results. The presence of reads mapping to all genes is difficult to visualise from plotting the RPKM values as the expression levels for *chIFITM3* are orders of magnitude higher than all the other *chIFITM* genes, masking the signal (**Figure 15**). To facilitate visualisation, SeqMonk (v1.48.0.); a software for analysis of high-throughput sequencing experiments from Babraham Institute, UK [140] was used. The selected quantitation pipeline optionally considers only uniquely mapped reads (mapping quality 255, as designated by STAR) over exons and normalises them solely for the library depth. For the visualisation purposes, the data are subsequently transformed using a logarithm function, base 2, resulting in $\log_2$ RPKM as a measure of transcript abundance. For this visualisation, the normalisation for gene length was not used so that the levels of observation would be loosely comparable among samples. Plotting the values on the $\log_2$ scale dynamically scales the data in a way that allows for detection of change in expression between the two tested conditions for all samples: infected vs uninfected (designated “Ctrl”). There is a significantly greater abundance of *chIFITM3* transcripts relative to other genes of the *chIFITM* locus (**Figure *14***). This observation is in agreement with the literature and confirms that the regulation of expression of *chIFITMs* is affected by the influenza infection. Note that no conclusions regarding the extent and significance can be drawn at this stage given no statistical analysis for differential expression has been conducted. Nevertheless, RPKMs are informative insomuch we can confirm that all clones have a functional *chIFITM* locus whose expression does not significantly deviate among the clonal cell lines.

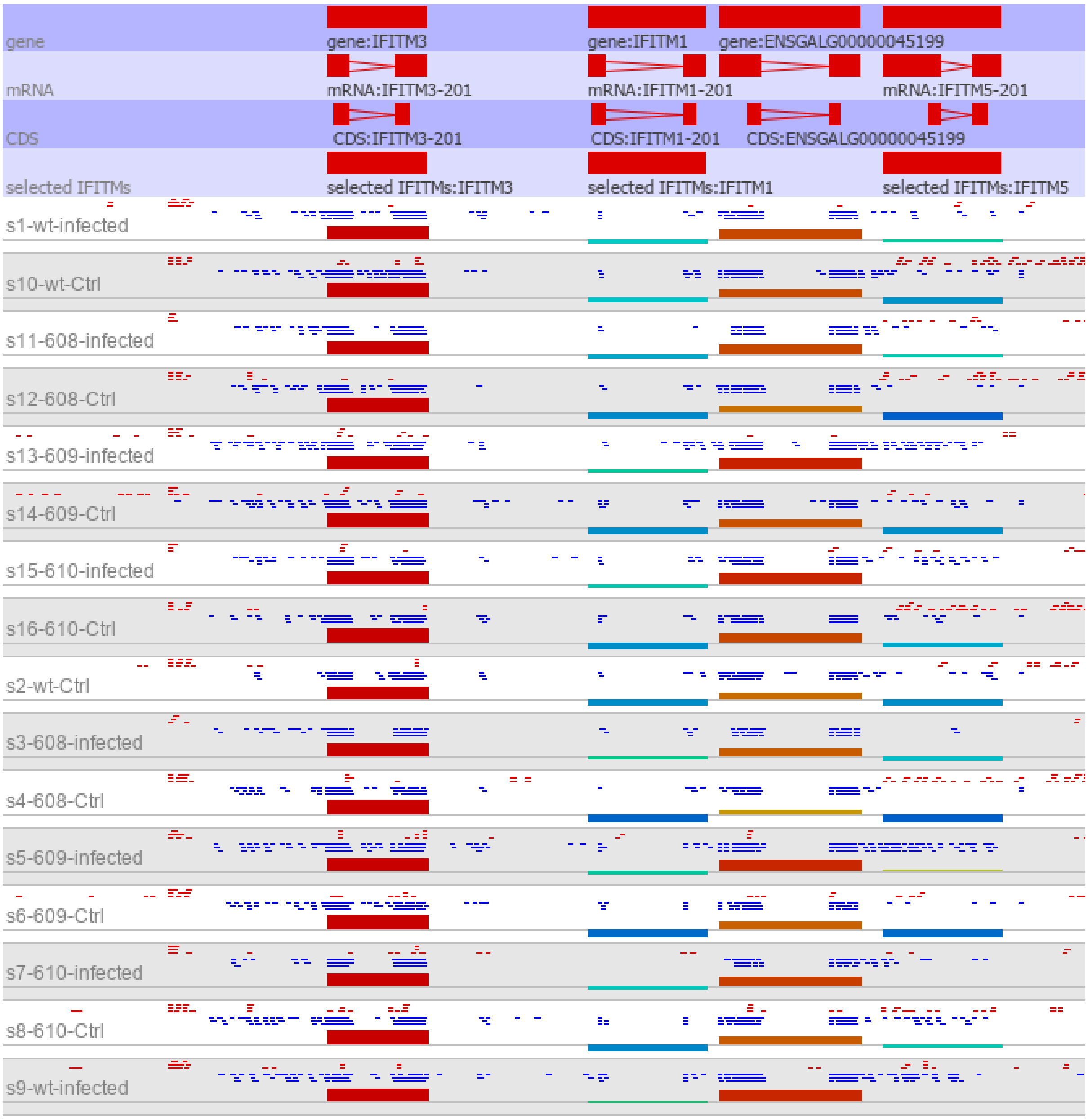


Figure* 14 *Visualisation of RNA-Seq reads mapping to chIFITM locus Bulk RNA-Seq was conducted on Illumina NovaSeq and reads were aligned to Gallus gallus GRCg6a_v95 using STAR vSTAR 2.7.2a. Visualisation of reads mapping to the region of interest was produced with SeqMonk v1.48.0, using only the reads with mapping quality 255, indicating uniquely mapped reads. S designates sample number in the RNA-seq experiment. Raw reads are coloured red or blue indicating that they map to forward or reverse strand, respectively. Tracks show the *chIFITM* locus with individual genes from left to right: IFITM3, IFITM1, IFITM2 and IFITM5 and corresponding features from top to bottom: gene, mRNA, CDS and custom track (chIFITM locus). IFITM2 did not have the annotation information available, thus Ensembl identifier is provided. The height and colour of the bars is indicative of the levels of observation: from low (blue) to high (red), indicating an increase. Quantitation was conducted over exons, for a directional reverse library. Counts were normalised for library depth and subsequently log transformed (log2 RPKM). They have not been normalised for gene length to facilitate a high-level comparison among samples. For practical reasons, only representative samples for every condition are visualised.

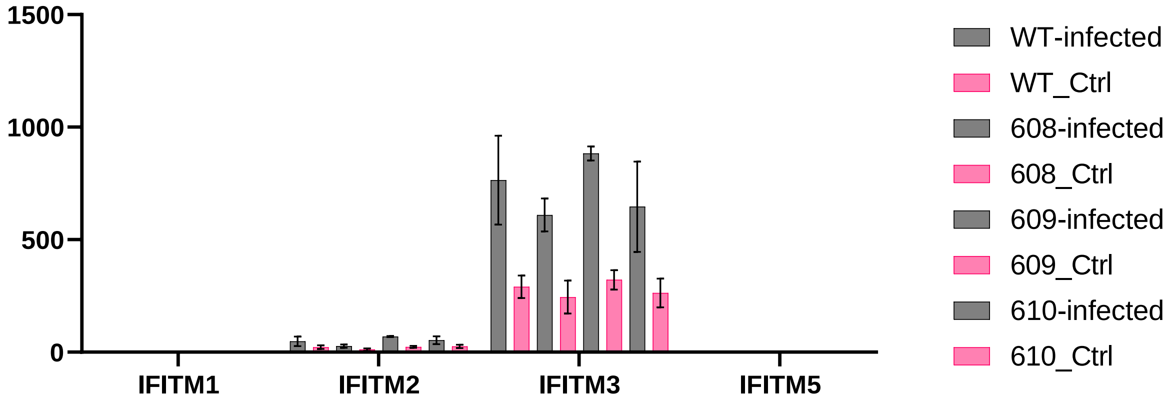


Figure 15 Levels of observation of chIFITM reads in WT DF-1 cell lines RNA-Seq reads were normalised for library depth to assess relative level of observation. Note that reads have not been normalised for gene length to enable high-level comparison among samples/cell lines and are not informative of differences among genes within a sample. Summary values represent the mean from the duplicate sample and error bars indicate SD.

To compare RNA-Seq and the quantitative RT-PCR results, all the read counts were subsequently normalised for effective gene length, producing transcript abundance values expressed in TPMs (transcript per million). These data allow for comparison of expression levels among genes within the sample, thus confirming extremely low levels of expression for *chIFITM1* and *chIFITM5*, relative to *chIFITM2* and particularly *chIFITM3 (* **Figure 16**).

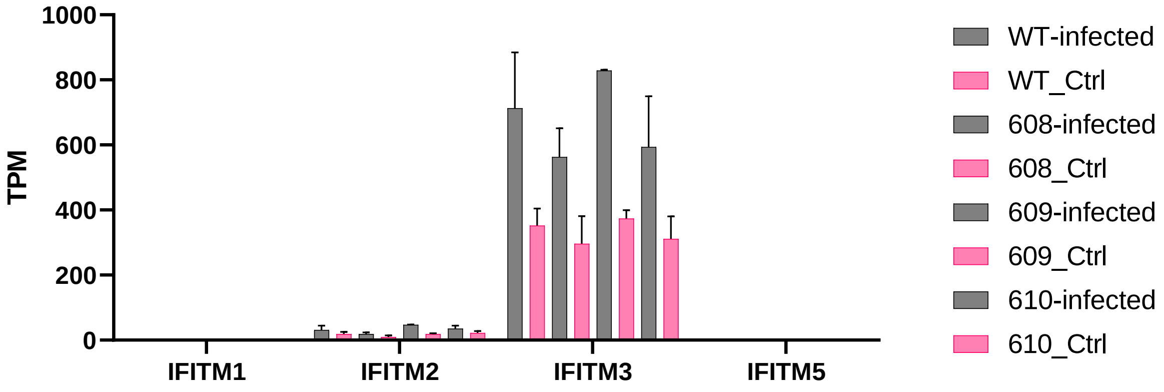


Figure 16 Quantification of chIFITM reads in WT DF-1 cell lines RNA-Seq reads were mapped with STAR and quantified using RSEM with the output in TPMs (transcript per million). TPM is a unit of transcript abundance normalised for gene length, library depth and type. Note that TPMs allow comparison of relative expression levels of genes within a sample, not among samples. Bars represent SE from duplicate samples.

I next compared the RNA-Seq read counts to quantitative RT-PCR data, previously generated by the research group, to investigate if there is an identifiable trend in expression levels of *chIFITMs* among

WT DF-1 clonal cell lines. Briefly, RNA was extracted from unstimulated WT DF-1 parental and clonal cell lines in biological triplicates and subjected to quantitative RT-PCR using Syber-green in technical duplicates. Ct values have been normalised against reference genes β-actin and *RPLP0* and the averaged Ct values subtracted from the total number of cycles (40 cycles).

chIFITM3 was the most highly expressed *chIFITM* gene tested, followed by *chIFITM2* and lastly *chIFITM1* (**Figure 17**). In addition, it is clear that the single-cell cloning has not altered the expression of *chIFITM* genes as there is no significant difference in expression levels when the parental line is compared to any of the clonal cell lines, in any of the transcriptomic methods employed or in any of the genes tested.

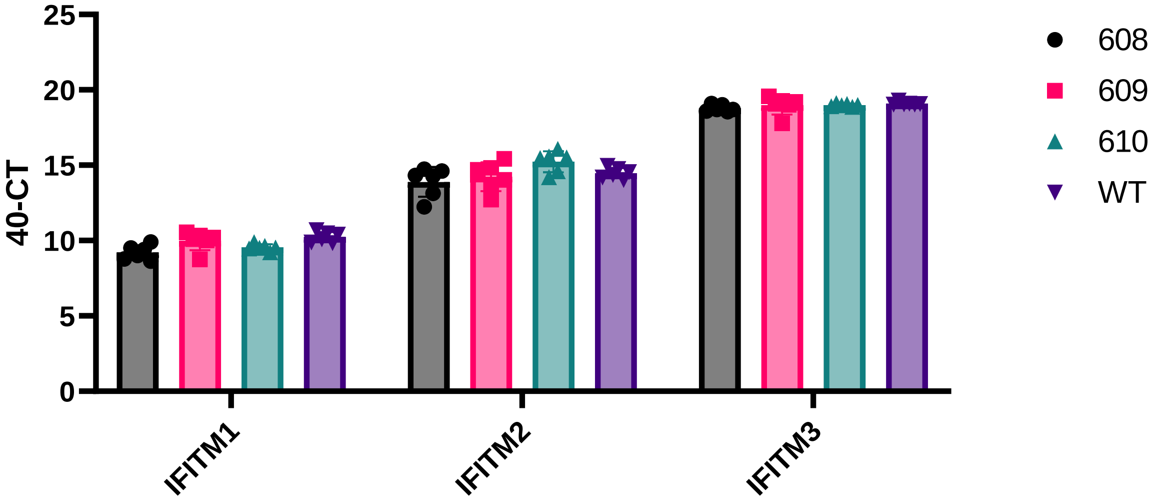


Figure 17 Basal level of IFITM expression in WT DF-1 clonal cell lines RNA was extracted from unstimulated cells using column extraction and used in a SyberGreen quantitative RT-PCR assay for 40 cycles. Values were normalised against those of reference genes: β-actin and RPLP0. Obtained Ct values were subtracted from the number of total cycles run for visualisation purposes. Bars represent standard deviations of the means of three biological replicates, assayed in duplicates. Data were kindly provided by Dr Ahmed Ibrahim.

3.3.2 Whole transcriptome sequencing of Wild-Type DF-1 cell lines

This section focuses on profiling the transcriptome of the model chicken fibroblasts at 6h post infection with the influenza virus. For this purpose, clonal cell lines of WT DF-1 named 608, 609 and 610 and the parental line from which they originated were being used. Because DF-1 is a mixed population, extraction of the transcriptomic signature that confers the phenotype being investigated becomes difficult. Clonal cell lines are used to eliminate some of the intrinsic variability with the aim of detection of expression patterns that generate the behaviour observed in the laboratory.

Concomitantly, the depth of sequencing required for subsequent analysis of differential expression is being assessed.

3.3.2.1 Mapping quality assessment and inter-sample variability of WT DF-1 clonal cell lines

Experiment set-up and RNA quality assessment

As described in section 3.3.1, an optimised infection system was set up to assess the variability of the transcriptomic profile in clonal (608, 609, 610) and parental WT DF-1 cell lines grown in 6-well tissue culture plates in duplicates to 80% confluency and infected with H5N3 A/duck/Singapore/3/1997 at the MOI of 1. The cells and the supernatant fluid were harvested at 6h post-infection and frozen at -80°C until processing. Three infection experiments were conducted, separated by a minimum of one week. The supernatant fluid from all samples was assayed by plaque assay following the protocol detailed in section 2.2.2.5. All infected samples were successfully confirmed positive for influenza infection by plaque assay (data not shown). Cells, frozen in RLT buffer (RNeasy kit, Qiagen), were used to extract the RNA following the protocol briefly explained in 2.2.3.5. RNA quality was assessed on TapeStation 2100 (Agilent) as per manufacturer's instructions summarised in 2.2.3.7. All samples were of very high quality (see Appendix **Figure 98**) and 2 replicate samples per condition were taken forward for RNA-seq. Additional validation of the quality and concentration was conducted by Dr Graham Freimanis at the Sequencing facility at the Pirbright Institute using Bioanalyzer 2100 (Agilent). Results affirmed high RNA quality for all 16 samples with RIN values around 9 (for a representative analysis results see Appendix **Figure 99**).

Library preparation and sequencing quality control (QC)

All the samples were taken forward for library preparation performed by myself and Dr Freimanis at the Pirbright Institute. 75 bp PE libraries were constructed using NEBNextUltra II RNA-Seq kit (NEB), with steps described in 2.2.5.1 and sequenced on a NextSeq machine (Illumina) to generate libraries of approximately 30 million (M) reads per sample. Generated reads were assessed for quality using FastQC [84] (a representative result of the analysis is available in the Appendix **Figure 100**). All results were summarised by MultiQC [92] and visualised in the **Figure 101** in the Appendix. The FastQC analysis showed the sequencing was successful and no trimming was necessary prior to alignment.

Alignment and mapping QC

Sequencing reads were aligned to *Gallus gallus* reference genome GRCg6a downloaded from Ensembl [86,87] database using STAR [85] . Quality control (QC) for the mapping step was performed with Qualimap [90,91]. For an example of the QC output for a typical sample for this dataset see Appendix

Figure 107. Results of the QC showed that the reads mapping to the chicken reference genome comprised on average one third of the library of infected samples (**Table 4**). In the uninfected samples, more than 90% of reads mapped to the reference, confirming the sequencing was successful, specifically in terms of the quality of the reads allowing for straightforward alignment. In terms of the number of reads produced, 50% of infected samples on average did not reach the targeted library size of 30M; most extreme case being sample 7, infected WT DF-1 610, with library of 13.5 M reads. 10M reads are still sufficient for high level differential expression analysis, especially given paired end libraries were used thereby facilitating alignment for rare transcripts. However, care must be taken when interpreting results for lowly expressed genes if those are of interest.

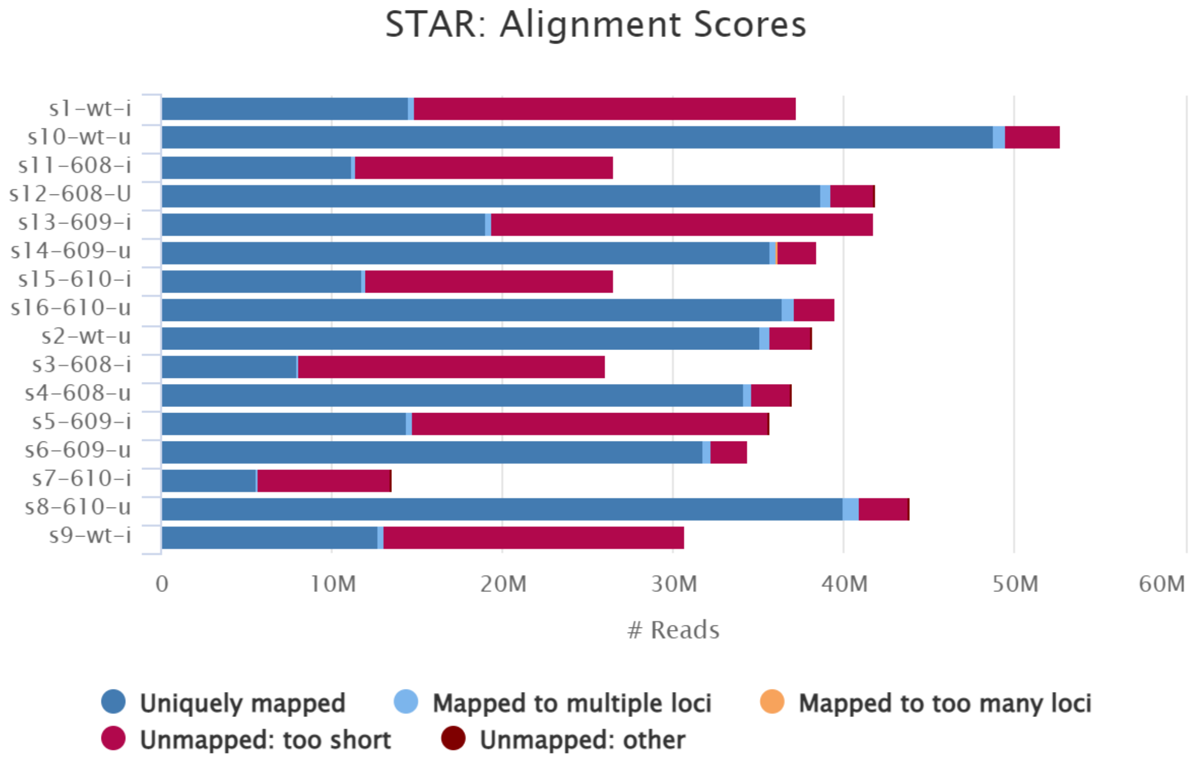


__Figure 18 Alignment results from STAR with chicken genome reference__ WT DF-1 RNA-Seq reads were aligned to chicken reference genome GRCg6a using STAR. Infected samples are denoted with an "i" and uninfected with a "u". Library sizes in millions of reads (M) are plotted on the x-axis. Reads successfully mapped to the chicken genome are designated as uniquely mapped and coloured in blue, unmapped reads are coloured in wine red. Summarisation of alignment reports and the figure was produced with MultiQC.

Table 4 Aligned reads for WT DF-1 clonal cell lines data set Sample names have a designation "i" indicating infection, and "u" for uninfected if the sample belongs to the control group. Library size is expressed as number of reads in millions (M).

Sample Name	% Aligned	M Aligned	total in M
s1-wt-i	38.9	14.5	37.30874
s2-wt-u	92.0	35.2	38.211931
s3-608-i	30.3	7.9	26.112102
s4-608-u	92.1	34.1	37.075247
s5-609-i	40.3	14.4	35.710369
s6-609-u	92.3	31.8	34.413011
s7-610-i	41.0	5.6	13.545515
s8-610-u	91.0	40.0	43.952617
s9-wt-i	41.6	12.8	30.741358
s10-wt-u	92.5	48.8	52.770814
s11-608-i	42.1	11.2	26.581565
s12-608-U	92.5	38.7	41.860239
s13-609-i	45.6	19.1	41.795844
s14-609-u	92.6	35.7	38.497589
s15-610-i	44.5	11.8	26.564327
s16-610-u	92.2	36.4	39.534879

Low mapping percentages were restricted to infected samples, suggesting a great proportion of the libraries is occupied by viral sequences. To investigate the origin of the unmapped reads fasta sequences of each of the eight segments of the H5N3 A/duck/Singapore/3/1997 virus, as retrieved from the NCBI (**Table 5**) were concatenated to the chicken reference genome GRCg6a. Using this genome as a reference, mapping quality reached the same levels for all samples, confirming the unaccounted reads had viral origin (**Figure 19**).

Table 5 *GenBank references of A/duck/Singapore/3/1997 segments used for bespoke genome reference generation*

Influenza Segment	Gene Product	GenBank Reference
1	PB2	GU052808.1
2	PB1	GU052807.1
3	PA	EU271944.1
4	HA	EF619972.1
5	NP	EF566039.1
6	NA	GU052804.1
7	M1, M2	GU052803.1
8	NS1	EU271945.1

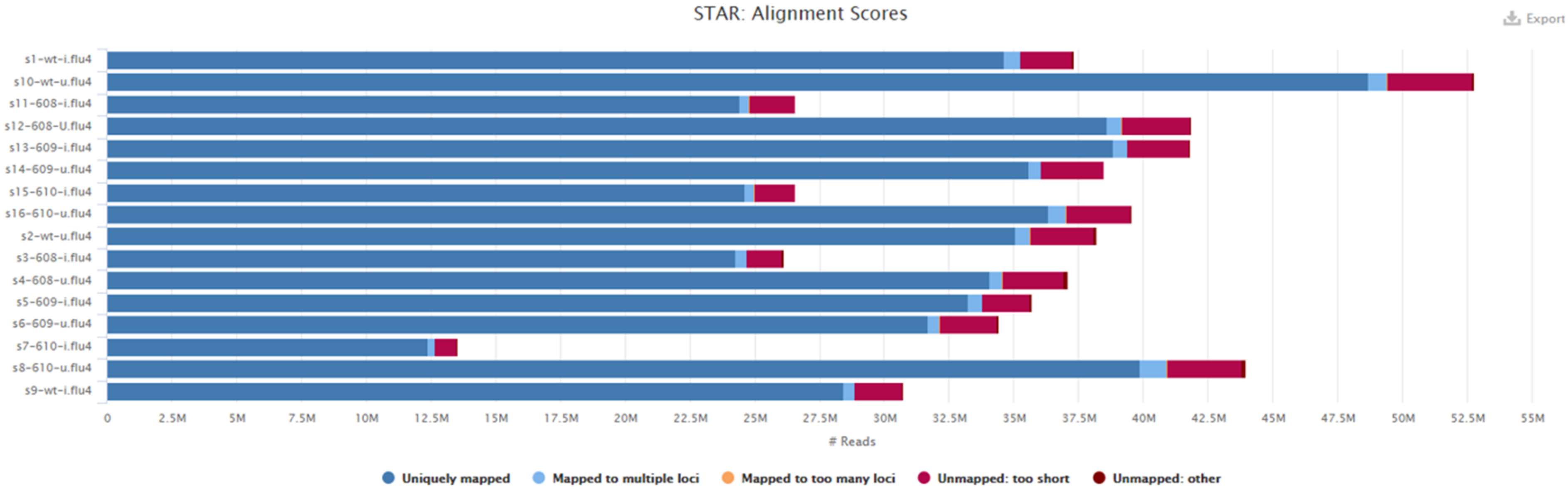


Figure 19 Mapping reaches 90% for all samples when bespoke reference genome is used WT DF-1 RNA-Seq reads were aligned to chicken reference genome GRCg6a concatenated to influenza sequences using STAR. “i” and “u” in the sample name on the y-axis indicate whether the sample is infected or uninfected, respectively. Infected samples are denoted with an “i” and uninfected with a “u”. Library sizes are expressed in millions (M) and plotted on the x-axis. Summarisation of alignment reports and the figure was produced with MultiQC.

Transcript Quantification and Data normalisation

Transcript abundance was estimated with RSEM as described in section 2.2.5.4 and reported in TPMs. Raw counts of the WT DF-1 data set were normalised using internal DESEq2 method that estimates *size factors* (a parameter normalising for library size and RNA composition).

Data Transformations and Exploratory Analysis

Both *rlog* and *vst* transformations of the raw counts were tested as they both have features useful for this dataset. The mean counts for each gene were plotted against its standard deviation (**Figure 20**). Note that the mean values are habitually represented by ranks. Ranks facilitate more even distribution of the data which in turn allows for easier detection of trends of the dependence of standard deviation on the mean. For the WT DF-1 data set *vst* ever so slightly outperformed *rlog* in flattening the variance-mean trend, illustrated by the red line depicting the running median estimator. The flexibility of the *vst* is notable in the low mean count range, an area where it is difficult to extract the signal form the noise. Henceforth, the values from the *vst* transformation were used for all clustering methods. To identify potential sample outliers and visually assess library sizes, the WT DF-1 RNA-Seq data set was additionally analysed with SeqMonk [140]. Analyses showed high level of quality of data and no sample outliers were identified (Appendix **Figure 102**).

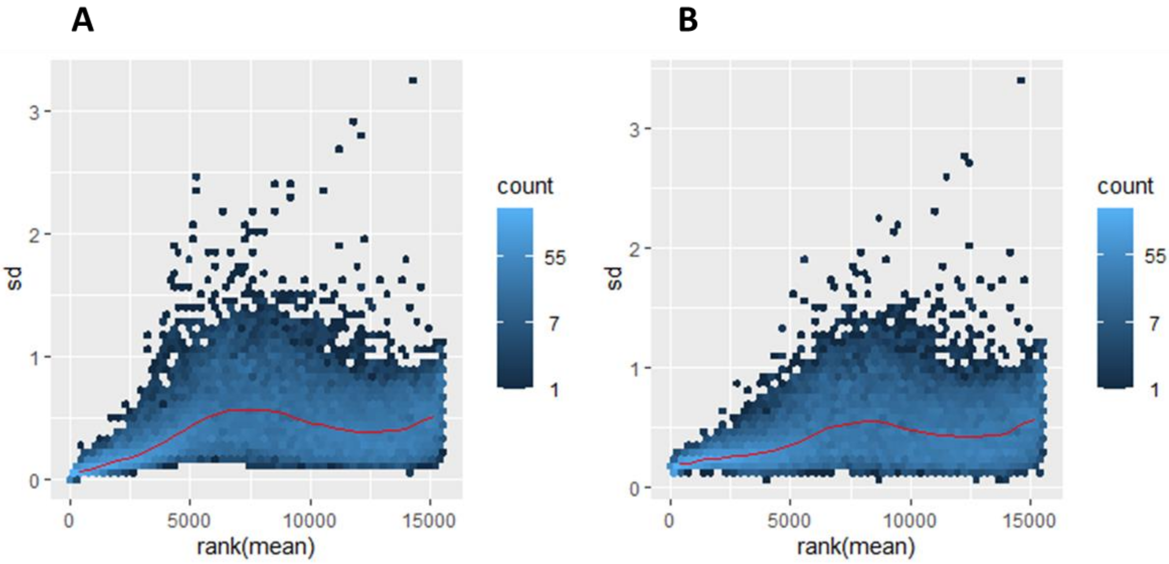


Figure 20 vst stabilises the mean-variance trend slightly better than rlog Normalised data were transformed using regularised log function (A) and variance stabilisation transformation (B). Gene counts are modelled as populations across samples whose mean values are plotted on x-axis and standard deviation on the y-axis. Red lines represents the trend of variance-mean dependency (measure of heteroskedasticity).

Count estimates, as normalised by DESeq2 using median of ratios method, and transformed using *vst*, were used to assess relatedness among all the samples by principal component analysis (PCA). The top 500 variable genes were included to estimate the contribution of each sample to the overall variation in the data. Results can vary slightly among iterations because the data are randomly sampled for generation of the covariance matrix. Nevertheless, overall distance among samples implies relative similarity in the measured parameter. First principal component contributed 71% of total variation showing that the condition, in this case the influenza infection, is the strongest driver of change in the gene expression patterns. The second most significant dimension contributed 18% to the total variation and it is almost exclusively dominated by the WT DF-1 608 cell line. WT DF-1 608 samples segregate visibly from all the others in both conditions (**Figure 21**) suggesting that some inherent difference in the expression patterns is present. All other cell lines showed little variation in the transcriptomic profile in the control condition, as shown by their close clustering, an observation aligning with their clonal origin. WT DF-1 609 appears to be more alike to the parental line than WT DF-1 610. This feature is preserved in both infected and uninfected condition and easily detected by visual inspection of the data (for an additional confirmation see diagnostic plot of sample-to-sample distances, Appendix **Figure 103** where the level of similarities is derived using Euclidian distances calculated from vst transformed data). Note that the replicates cluster together, indicating no considerable batch effect is present.

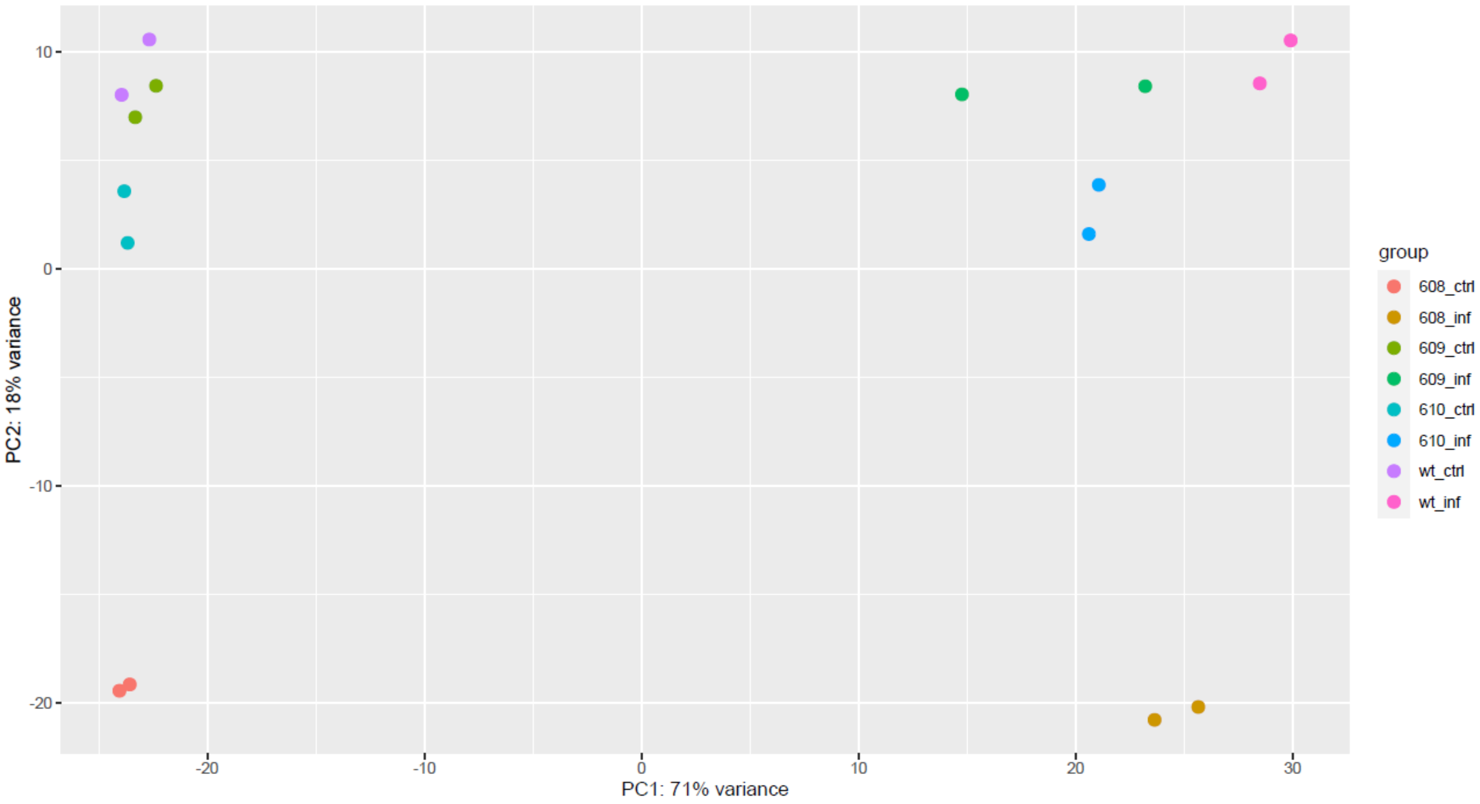


Figure 21 PCA plot for WT DF-1 RNA-Seq data Axes in PCA plot represent similarity parameters across which samples cluster. Note this graphic is a bi-plot of PC1 and PC2, two most significant principal components. Legend indicates the extent of representation of the variable (sample) in the principal component; higher values indicate greater contribution of the variation of the variable to the total variation of the whole data set. Note that “U” and “I” at the end of a sample name denote condition: uninfected and infected, respectively.

Standard diagnostic plots for outlier detection did not detect any outliers (data not shown). Gene-wise dispersion estimates were calculated; WT DF-1 data set exhibited typical dispersion trend (i.e. high variance/standard deviation in the low count region that dropped and stabilised in the high count region) visible in **Figure 104** in the Appendix. Ultimately, Cook's distances were calculated, and all samples showed extremely similar distribution profile (**Figure 105**) thus satisfying requirements for progression to differential expression analysis.

To analyse more closely the PCA results, principal components one and two were extracted to analyse the genes that contribute the most to the transcriptomic variability in response to the condition. For the PC1 the genes that account for the top 12% of variation were extracted and plotted in **Figure 22**; namely: *#LOC112532844, CMPK2, ERBB3, CCN2, SOCS1, CITED4, CALHM2, RHOB, PDP2, SH3BP4, FDFT1, PDP1, CENPJ, ADM, ZNF518B, GPR146, LRRC32, NET1, IRS4, FBXO32, TOB2, DUSP6, VPS18, MAVS, RNF24, DUSP7, SPRY2, LIG4, PLK2, TBC1D20, RWDD2A, CEBPB, RIN3, ENSGALG00000040546, LIMK2, ID4, NFKBIA, SOCS6, LRRC8A, SPSB1, PPP1R3C, LOC422150, SMURF2, BTG1, PDZRN3, TIPARP, URB2, ARL4C, TOPORS, ARV1, TUBGCP6, FEM1A, SLC38A2, TICAM1, CAPRIN2, FAM46A, BTG2, TFAP2C, RGS3, ACTR5, LZTS1, FCHSD2, TAF4, JUN, FASTKD3, ENSGALG00000017071, HAUS3, ZNF839, FBXO5, SNAPC3, ENSGALG00000003206, ALG2, CCNF, EPC2, TTI1, TMEM170A*, for PC1 and for PC2: *HINTW, LOC107049046, ATP5F1AW, JAC1, ENSGALG00000041221, CDH5, PAPPA2, LSP1P1, PPL, IL2RA, CHMP4C, IL31RA, VSIG10L, KHDRBS3, FSTL5.*

To visualise the patterns of gene expression across samples, a heatmap was constructed plotting Z-scores of top 50 most variable genes. Variance was calculated from vst normalised values to create a ranking of all genes. For those genes, TPM values were normalised with the Z-score method and clustered. The Euclidian distances were used to cluster the data to further elucidate which genes drive the similarities in the transcriptome among samples (**Figure 23**). Five clusters emerged. Cluster 1 drove the separation of WT DF-1 608 cell line from the rest of the samples in both conditions. It contained following genes: *WISP2, JAC1, ADH1C, CRYL1, PPEF1, ABCA2, UBAP2L2, ATP5F1AW, UBE2R2L, LOC420716, HINTW, LOC107049046, GOLM1, COLEC12.* These genes had lower expression levels in WT DF-1 608 than all the other cell lines, independent of the infection. The second cluster was characteristic of the infection condition and the genes contained therein (*LOC112532844, IFIT5, CMPK2*) were upregulated relative to the control. Cluster 3 comprised of *TAGLN, LOXL3, SLC39A8, ARL8BL* – genes whose upregulated expression in WT and 610 DF-1 drove the clustering of those two cell lines. Fourth cluster is also dominated by WT DF-1 608, as was cluster one, but here the relevant genes were more highly expressed in 608 than in all other cell lines. Cluster four contained EFEMP*1, HSPB9, IL8L2, VSIG10L, LSP1P1, PAPPA2, KHDRBS3, C1QTNF3, CDH5, PPL.* Cluster five was the biggest cluster containing *ADM, FDFT1, CALHM2, SH3BP4, RIN3, IRS4, VPS18, RHOB, SOCS1, LRRC32, CCN2,*

CITED4, PDP1, DUSP6, TOB2, ADRA1B, ADAMTSL2. These genes had increased levels of expression in infection which were approximate across all cell lines.

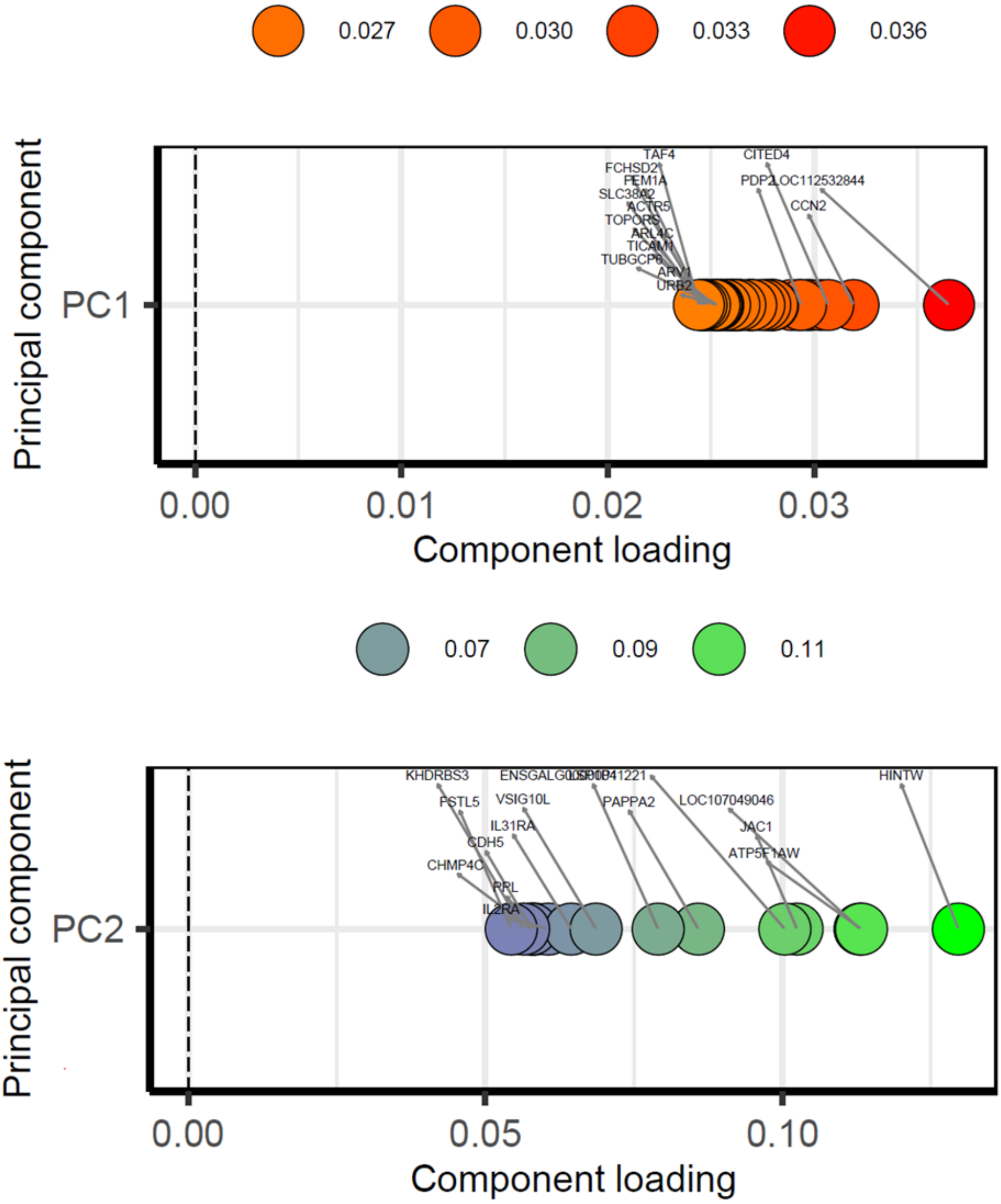


Figure 22 Loadings plot for PC1 and PC2 Topmost 12% individual genes that contribute to the variation of each component were extracted. For easier visualisation not all are represented. Full list for **PC1** includes *#LOC112532844, CMPK2, ERBB3, CCN2, SOCS1, CITED4, CALHM2, RHOB, PDP2, SH3BP4, FDFT1, PDP1, CENPJ, ADM, ZNF518B, GPR146, LRRC32, NET1, IRS4, FBXO32, TOB2, DUSP6, VPS18, MAVS, RNF24, DUSP7, SPRY2, LIG4, PLK2, TBC1D20, RWDD2A, CEBPB, RIN3, ENSGALG00000040546, LIMK2, ID4, NFKBIA, SOCS6, LRRC8A, SPSB1, PPP1R3C, LOC422150, SMURF2, BTG1, PDZRN3, TIPARP, URB2, ARL4C, TOPORS, ARV1, TUBGCP6, FEM1A, SLC38A2, TICAM1, CAPRIN2, FAM46A, BTG2, TFAP2C, RGS3, ACTR5, LZTS1, FCHSD2, TAF4, JUN, FASTKD3, ENSGALG00000017071, HAUS3, ZNF839, FBXO5, SNAPC3, ENSGALG00000003206, ALG2, CCNF, EPC2, TTI1, TMEM170A ;* and for **PC2** *HINTW, LOC107049046, ATP5F1AW, JAC1, ENSGALG00000041221, CDH5, PAPPA2, LSP1P1, PPL, IL2RA, CHMP4C, IL31RA, VSIG10L, KHDRBS3, FSTL5.*

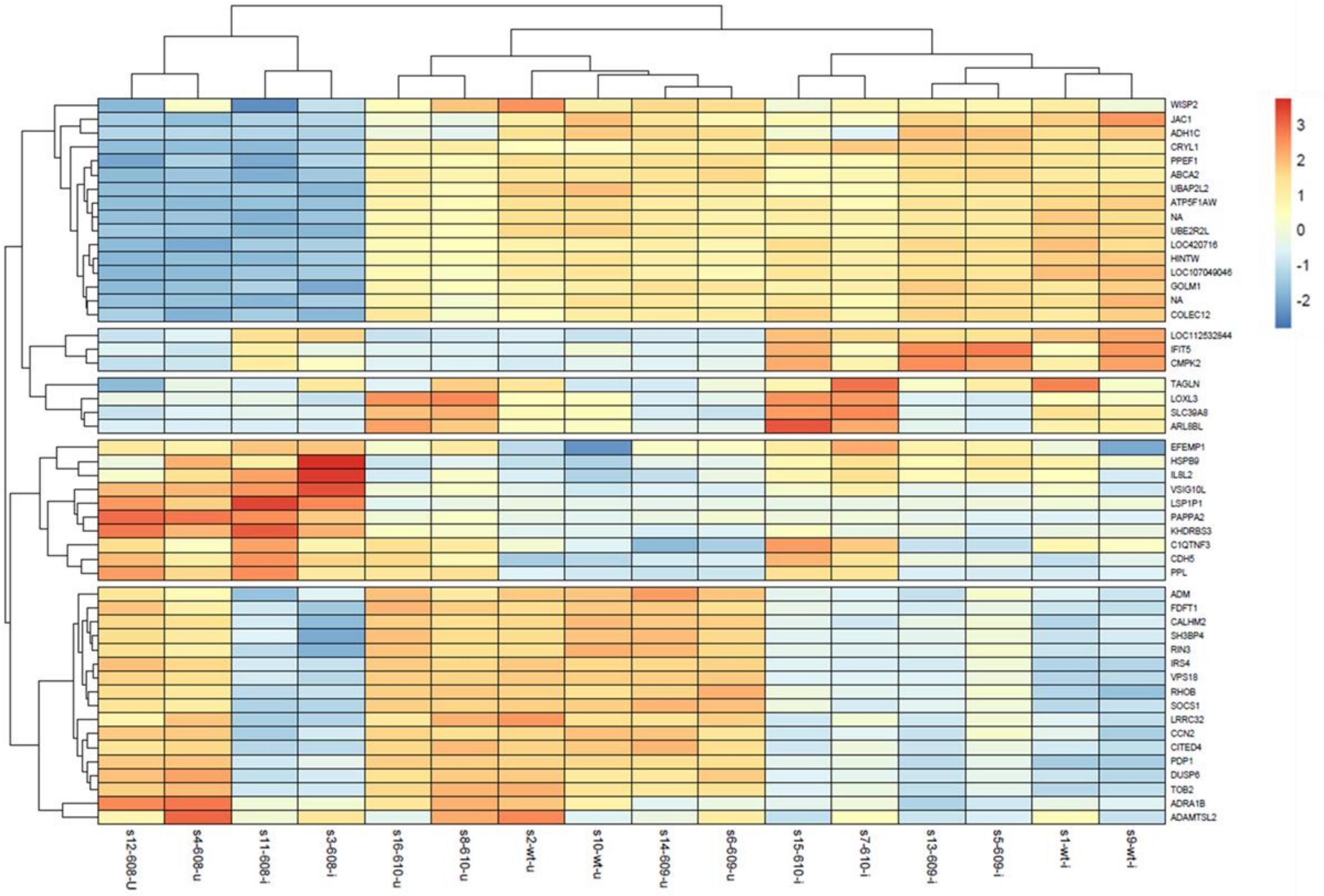


Figure 23 Most "differently" expressed genes in WT DF-1 cell lines at 6h post infection with the influenza virus *Z-scores were calculated for 50 genes that showed the greatest variance in expression values expressed in TPMs. Clustering was done using Euclidian distances. Legend indicates a deviation from the mean across all samples. Overall increase in level of transcription from the mean across samples is coloured in warm tones towards red while a decrease is coloured in cooler tones towards blue. Visualisation was produced in R.*

3.3.3 The impact of influenza infection on the global transcriptome of DF-1 cells

Statistical modelling

To characterise overall transcriptomic response to infection, the data for all WT DF-1 clonal cell lines were pooled for the analysis, resulting in eight replicates per condition. This approach increases the power of the calculation of differential expression and consequently the resolution at which a gene can be declared as differentially expressed. Because WT DF-1 608 cell line clustered away from other cell lines in both conditions in the PCA, therefore implicating a clone-specific behaviour, clonal origin was deemed to be a meaningful variable and genotype was included in the statistical modelling. A means model without intercept, represented by design formula ~ 0 + condition + genotype, was employed in DESeq2 to calculate differential expression.

Differential expression

For hypothesis testing the Wald test and an FDR= 0.05 (false discovery rate) were employed together with Benjamini-Hochberg correction for multiple testing. For shrinkage of log fold changes *ashr* method [141] (adaptive shrinkage) was used. *ashr* supports contrasts as means of comparing treatments (as opposed to coefficients). Out of 15 561 genes with non-zero total read count, 5839 (38%) were upregulated and 4784 (31%) were downregulated.

To further investigate the clustering of samples heatmaps were constructed by plotting 100 most upregulated and downregulated differentially expressed genes. TPM values were normalised across all samples using Z-score method and clustered using Euclidean distances unless stated otherwise.

Analysis of upregulated genes in **Figure 24** uncovered two main sample clusters: the infected and the uninfected group. Overall change in gene expression between the two groups seemed weak with some, potentially spurious, extreme data points. Replicates in the control group clustered together due to less variation in the base level expression. Among all WT DF-1 cell lines in the control condition two loose groupings were formed: WT DF-1 608 with 609, and 610 with the parental WT DF-1 cell line. Replicates in the infection condition had less consistent transcriptomic response and did not exhibit any more granular grouping within the cluster. Statistically significant downregulated genes displayed very stark clustering: both between the two conditions and among the cell lines (**Figure 25**). Samples are segregated distinctly into two clusters according to the infection status. Within each cluster/condition the replicates clustered together.

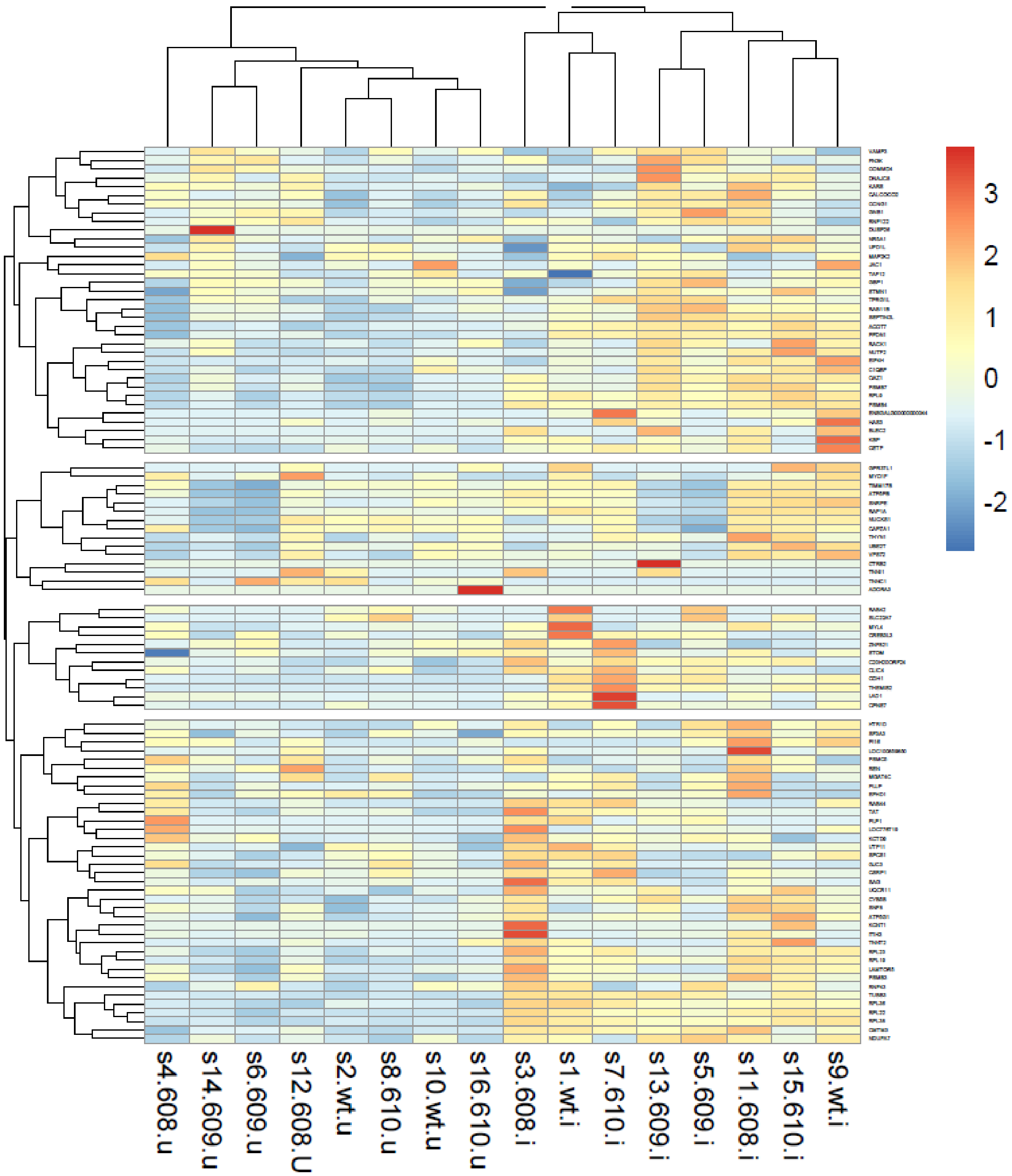


Figure 24 Heatmap of 100 most upregulated differentially expressed genes in WT DF-1 cell lines in influenza infection
Differential expression was calculated at FDR of 0.05 as the cut-off for statistical significance. TPM values for 100 most upregulated genes were normalised using Z-scores method across all samples. Clustering was carried out with *pheatmap* package in R using Euclidian distances.

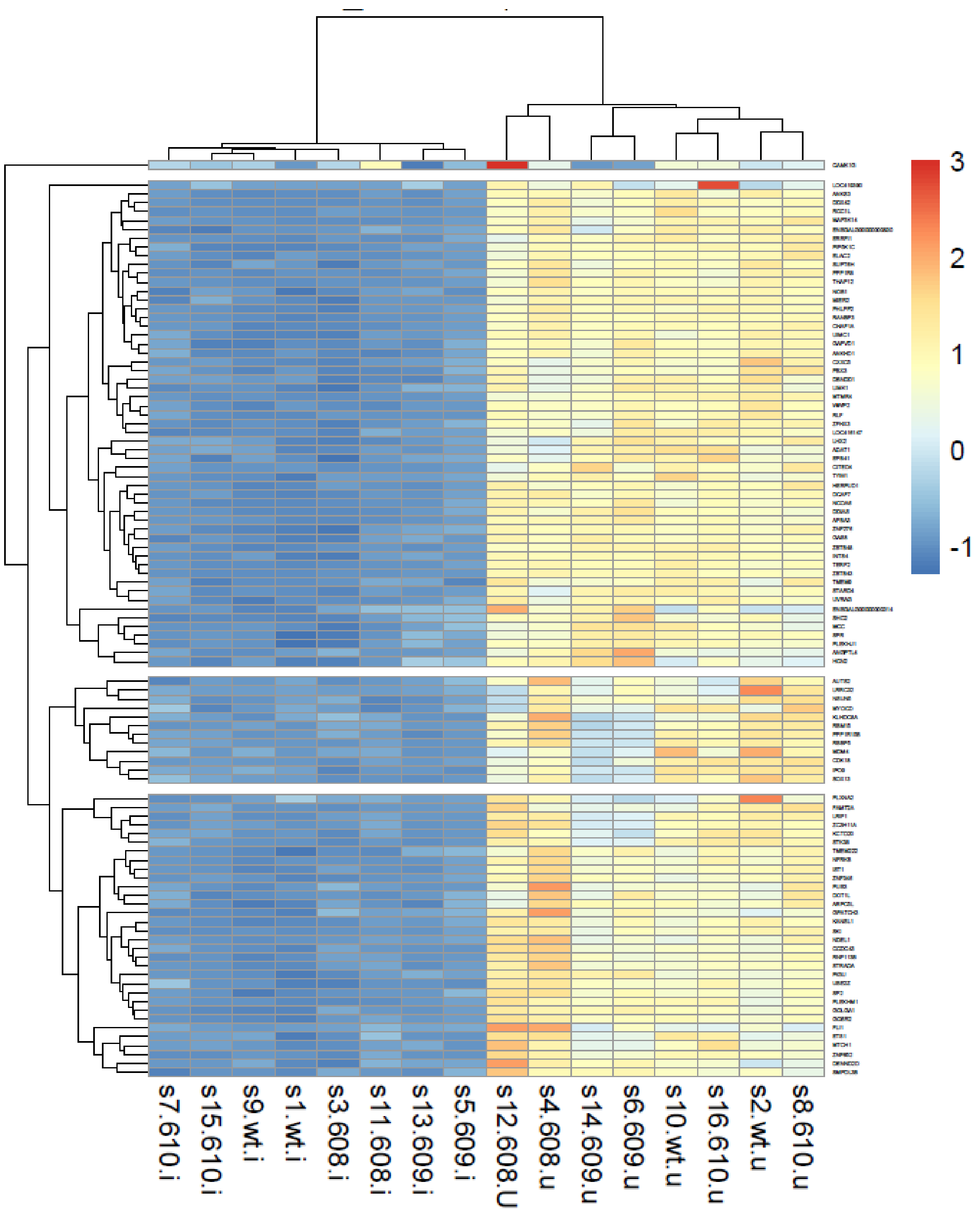


Figure 25 100 most downregulated genes in WT DF1 cell lines at 6h post infection with influenza Z-scores were calculated from TPM values for 100 most downregulated genes that were statistically significant at FDR = 0.05. Values were clustered using Euclidean distances and the heatmap produced with *pheatmap* package in R. Legend indicates deviation from the mean across all samples, in positive and negative direction, coloured in warm and cold gradient, respectively

.

3.3.3.1 Pathway analysis

GSEA in clusterProfiler

Differential expression analysis returned a list with more than thousand genes; it was necessary to distil the information further. Log fold changes of differentially expressed genes were shrunken using the method *ashr* [141] to obtain significantly more accurate estimation of the effect sizes. Subsequently they were filtered by FDR values and genes with FDR less than 0.05 were put forward for subsequent analysis. They were tested for pathway enrichment using Gene Set Enrichment Analysis (GSEA) [99] in *clusterProfiler* [142,143] against gene ontology (GO) [105,107] database using chicken annotation. Significant biological processes were sorted by q-value (FDR-adjusted p-value) and plotted against their gene ratio (number of genes in the dataset as a fraction of total number of genes in the term). Many terms associated with nucleic acid metabolism were highly represented in the experimental dataset at 30% gene ratio and they were found to be suppressed. Activated biological processes were highly statistically significant and were mainly concerned with lipid and protein organisation. Immune response was activated and highly statistically significant with the q-value of 0.03 (**Figure 26**).

PADOG in ReactomeGSA

To further mine the experimental dataset for information, a Pathway Analysis with Down-weighting of Overlapping Genes (PADOG) was conducted with ReactomeGSA [112]. The principal advantage of this analysis is the reduction of the ranking of genes that are represented in multiple gene sets, thus controlling the number of false positives. To increase the statistical power, the differential expression analysis is conducted on the pathway level as opposed to on the individual gene level. Chicken gene identifiers were translated into human UniProt identifiers and projected onto communal pathways. Translating chicken identifiers into human is necessary because databases of pathways are based on human pathways. Reactome [109] database was queried, and 2394 pathways were found significantly regulated with the FDR cut-off of 0.05.

Infectious disease (R-HSA-5663205), Cellular response to stimuli (R-HSA-9711123) and RNA metabolism (R-HSA-8953854) were the largest terms identified as significantly upregulated. Within those Cellular responses to stress (R-HSA-2262752), Cellular responses to stimuli (R-HSA-8953897), Downstream signalling events of B Cell Receptor (BCR) (R-HSA-1168372), Major pathway of rRNA processing in the nucleolus and cytosol (R-HSA-6791226), rRNA processing in the nucleus and cytosol (R-HSA-8868773), FCERI mediated NF-kB activation (R-HSA-2871837), Nucleotide Excision Repair (R-HSA-5696398), Activation of NF-kappaB in B cells (R-HSA-1169091), G1/S DNA Damage Checkpoints

(R-HSA-69615), p53-Dependent G1 DNA Damage Response (R-HSA-69563), p53-Dependent G1/S DNA damage checkpoint (R-HSA-69580), Interleukin-1 signalling (R-HSA-9020702), C-type lectin receptors (CLRs) (R-HSA-5621481), CLEC7A (Dectin-1) signalling (R-HSA-5607764), UCH proteinases (R-HSA-5689603), SARS-CoV Infections (R-HSA-9679506), Degradation of GLI2 by the proteasome (R-HSA-5610783).

More biological processes were downregulated – mostly concerning protein processing and transcription, cell cycle progression and DNA-repair as well as the vesicular transport. Pathways with lowest FDR (<0.001) were : Processing of Capped Intron-Containing Pre-mRNA (R-HSA-72203), Mitochondrial biogenesis (R-HSA-1592230), TICAM1 deficiency - HSE (R-HSA-5602566), Metabolism of proteins (R-HSA-392499), Cell Cycle Checkpoints (R-HSA-69620), DNA Repair (R-HSA-73894), Antigen processing: Ubiquitination & Proteasome degradation (R-HSA-983168), SUMOylation of transcription cofactors (R-HSA-3899300), Defective DNA double strand break response due to BRCA1 loss of function (R-HSA-9663199), Cell Cycle (R-HSA-1640170), Circadian Clock (R-HSA-400253), Cellular Senescence (R-HSA-2559583), Class I MHC mediated antigen processing & presentation (R-HSA-983169), Deubiquitination (R-HSA-5688426), Regulation of TP53 Activity (R-HSA-5633007), Transcriptional Regulation by TP53 (R-HSA-3700989), Transport of Mature Transcript to Cytoplasm (R-HSA-72202).

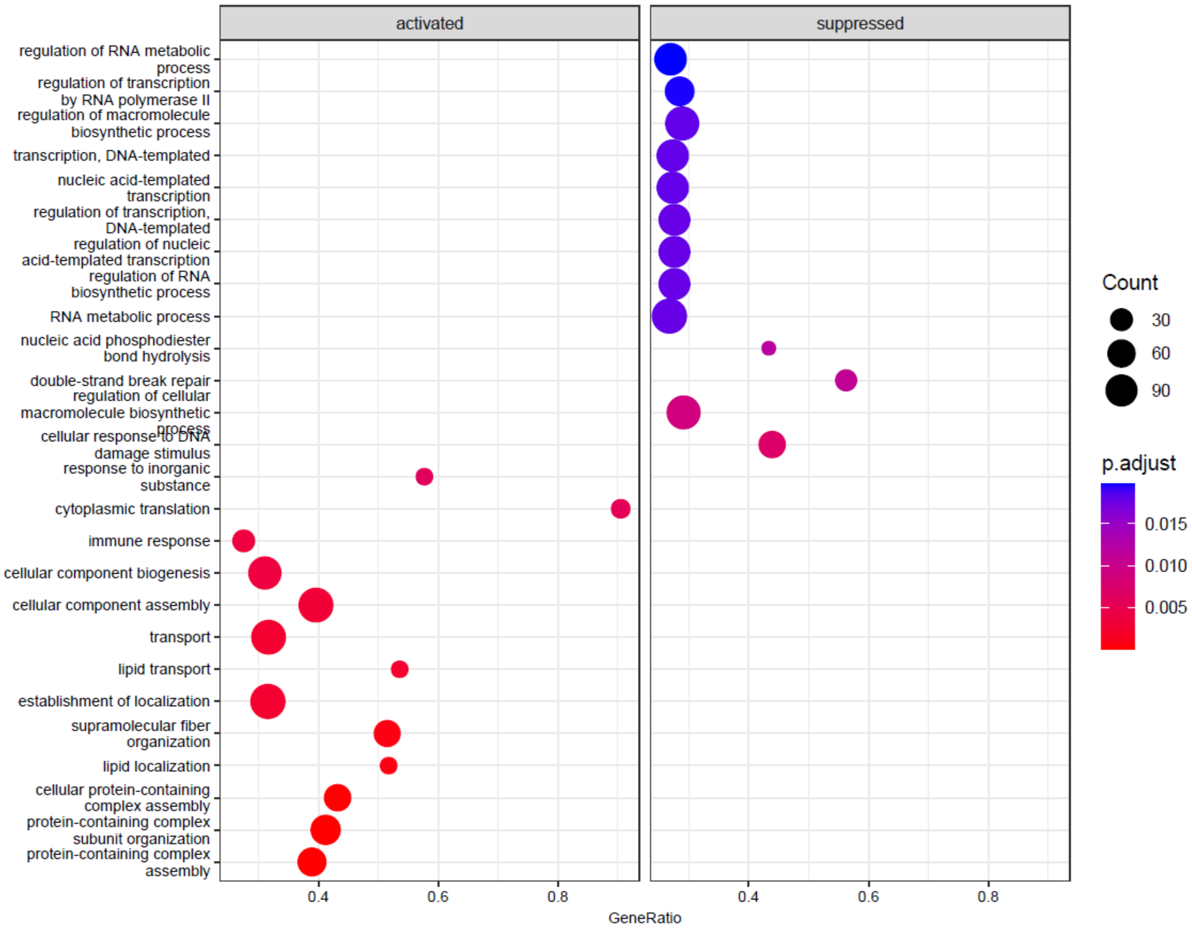


Figure 26 Gene Set Enrichment Analysis (GSEA) of differentially expressed genes in WT DF-1 cell lines infected with influenza GSEA for Gene Ontology terms associated with biological processes was conducted in clusterProfiler in R against all differentially expressed genes in pairwise comparison of infected vs uninfected WT DF-1 cell lines. Only genes with FDR <0.05 were included in the analysis. Gene ratio plotted on x-axis indicates the proportion of genes from the experimental dataset represented in the gene-set from the database. Size of the circles illustrates the absolute number of genes matched and the colour of the circle the p-adjusted value (FDR) for the individual test. Graph was produced with *geneProfiler* package in R.

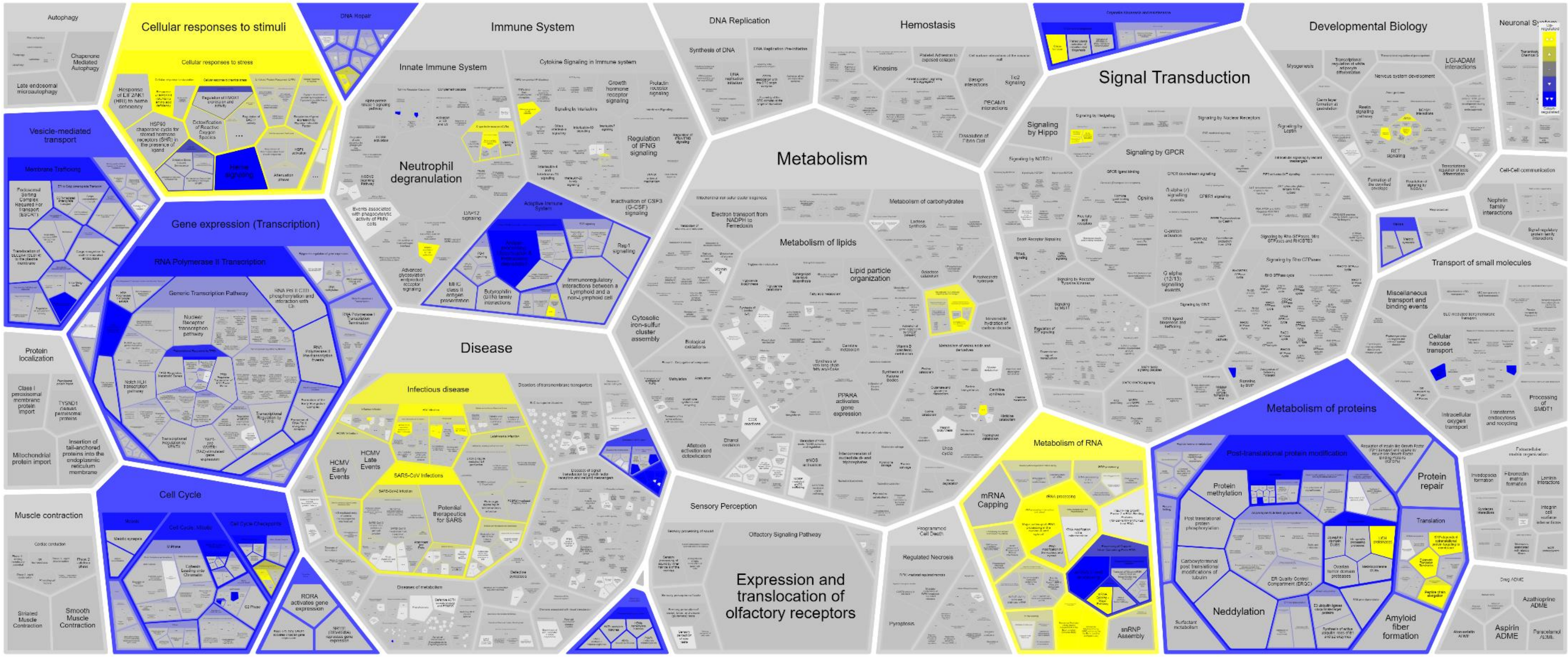


Figure 27 ReactomeGSA Pathway analysis of differentially expressed genes in WT DF-1 cell lines infected with influenza Raw RNA-seq reads were imported into Reactome GSA and PADOG analysis was conducted for condition: infected vs uninfected. FDR was set to 0.05 and differential expression analysis was done on the pathway level. Upregulated pathways are designated in yellow and the downregulated in blue.

3.3.4 WT DF-1 608-specific response

Differential expression

WT DF-1 608 clonal cell line exhibited properties that deviated on the whole from the other cell lines tested. WT DF-1 608 was able to support higher levels of viral replication as estimated by the plaque assay (**Figure 12**) and its transcriptional profile was significantly different from other cell lines, demonstrated by the PCA (**Figure 21**). These observations warranted a further characterisation of the gene expression in the clone 608. A separate differential expression analysis was conducted for this clonal cell line only, following the protocol described in 2.2.5.3. Infected samples were compared to mock using the contrast function in DESEq2. The resulting differentially expressed genes were deemed statistically significant at FDR of 0.05. To adjust the estimates of fold changes *ashr* function was used.

To visualise potential patterns, Z-scores of TPM values of most upregulated genes were clustered using Euclidean distances (**Figure 28**). Genes which are specifically upregulated in WT DF-1 608 do not follow any pattern of regulation in other cell lines. Clustering resulted only in separation of infected and uninfected samples, confirming that the expression of the WT DF-1 608 specific differentially expressed genes is influenced by the infection significantly more than by the genotype. Relatively similar expression profile of those genes among all cell lines indicates a communal part of the immune response to the influenza infection characteristic for all DF-1 cell lines. Using this experimental design no strong 608-specific signal was detected.

Downregulated genes were analysed in the same manner described previously. Resulting clusters are visualised in **Figure 29**. Downregulation seems to be more tightly controlled as is evidenced by great majority of the most downregulated genes having same extent of downregulation across all cell lines. Some less orderly clusters of genes are dominated by an outlier whose extreme counts influence the average too much to make the Z-score an informative way to represent the information (see cluster containing *HNRNPM, ADIPOR1, MAPT, KCTD9, MAP3K3, MYO1D, PSMC5, DMA* and *ISOC1*). Other clusters house the true variability in expression levels that are significant between infected and non-infected clones of WT DF-1 608, but not among all the other cell lines, e.g. cluster containing genes *ENSGALG00000000044, HAS3, UBE2T, RACK1, TUBB3, RPL36, CYB5B, RPL9, RPL22, RAB11B, ACOT7, CMTM3* and *NDUFA7*. Genes *TMEM183A, NUCKS1, CSRP1, LAMTOR5, TIMM17B* and *SNRPE* seem to be consistently less expressed in WT DF-1 609 clonal cell line than in all other cell lines, regardless of the condition.

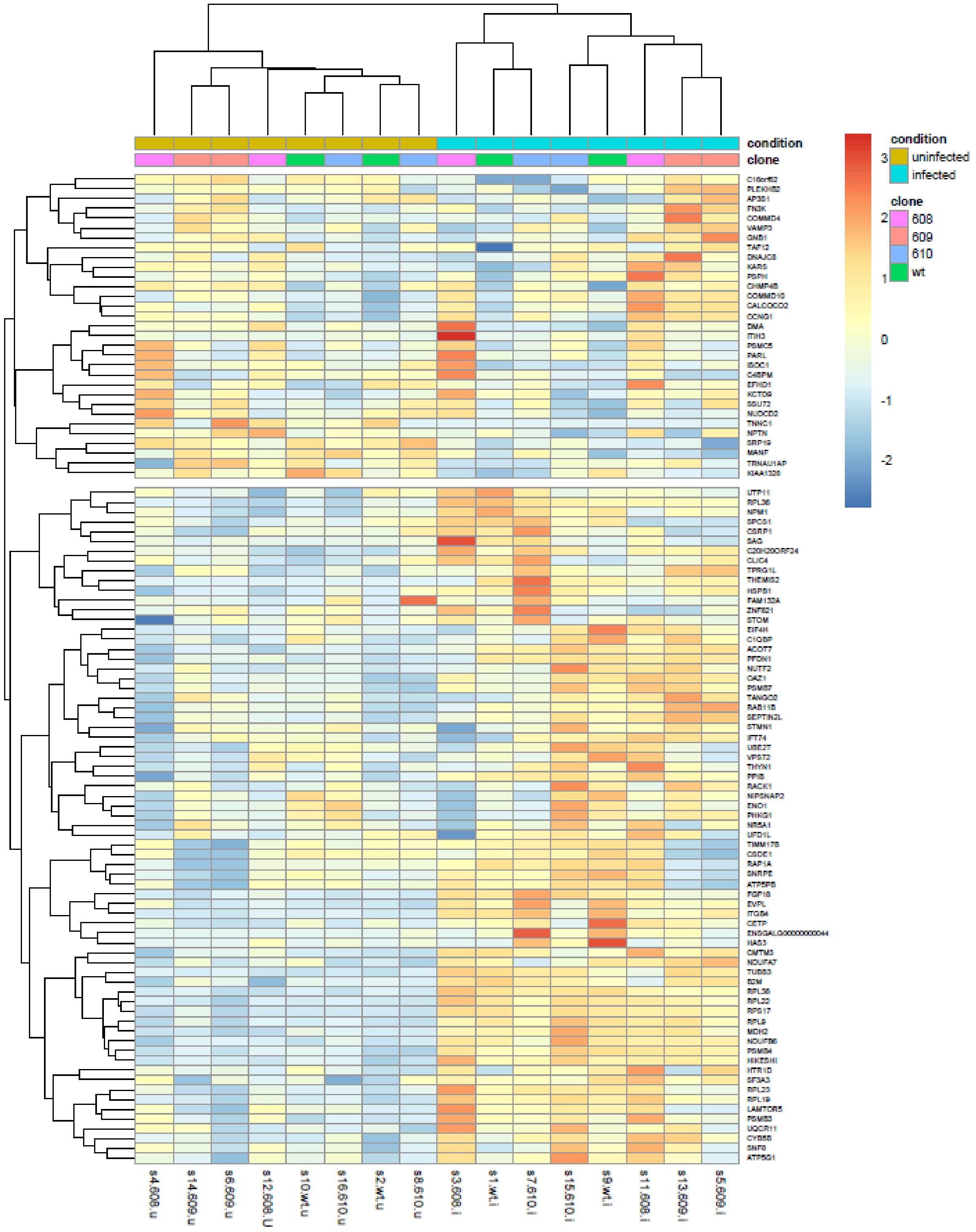


Figure 28 Most upregulated differentially expressed genes in WT DF1 608 clonal cell line at 6h post infection with H5N3 influenza virus *TPM values of 100 most differentially expressed genes at FDR 0.05 were Z-scored across all samples.*Z-score values were then clustered by Euclidean distances with *pheatmap* in R. Deviation from the mean is indicated in warm and cold gradient depending on the direction of the change.

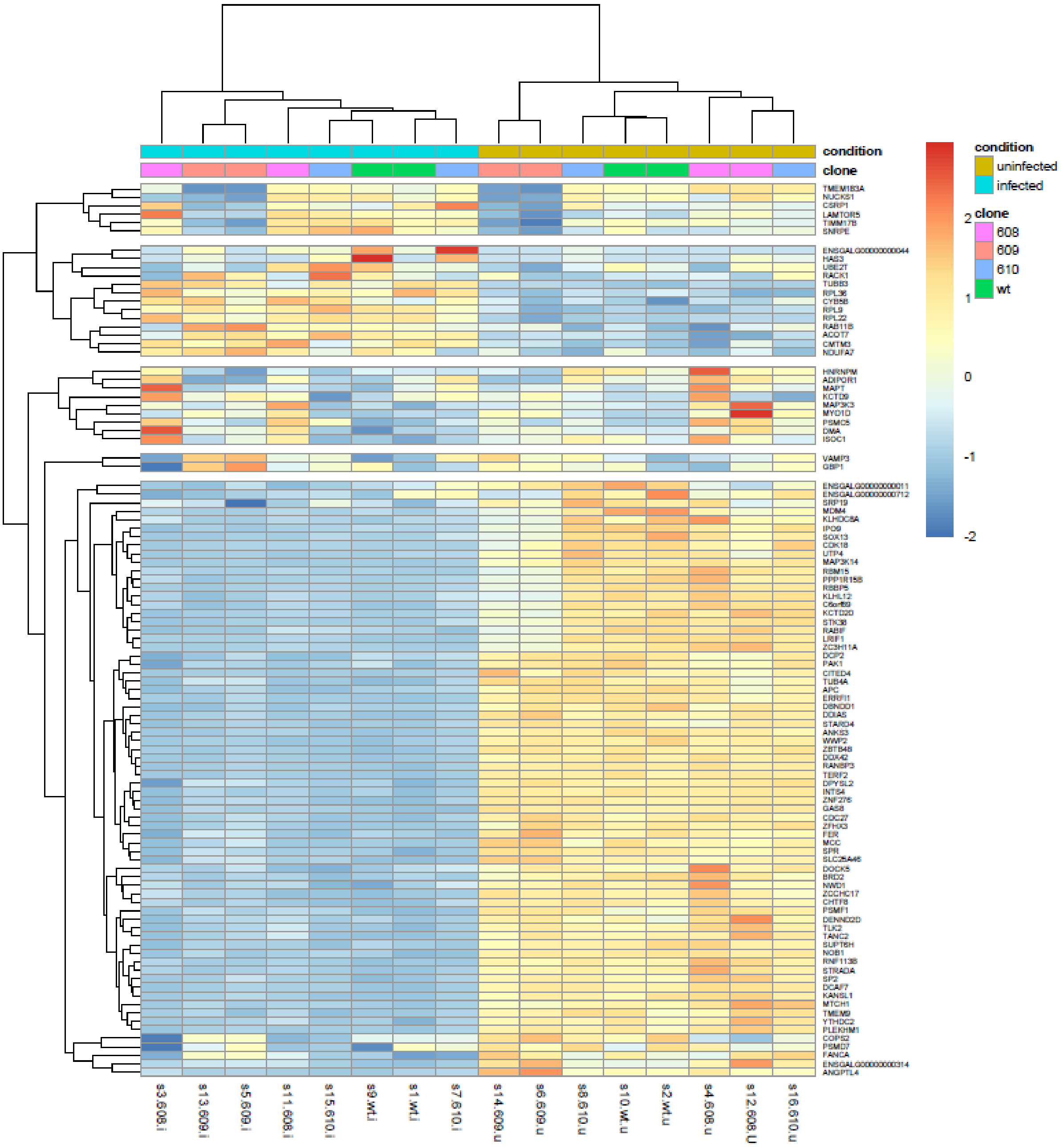
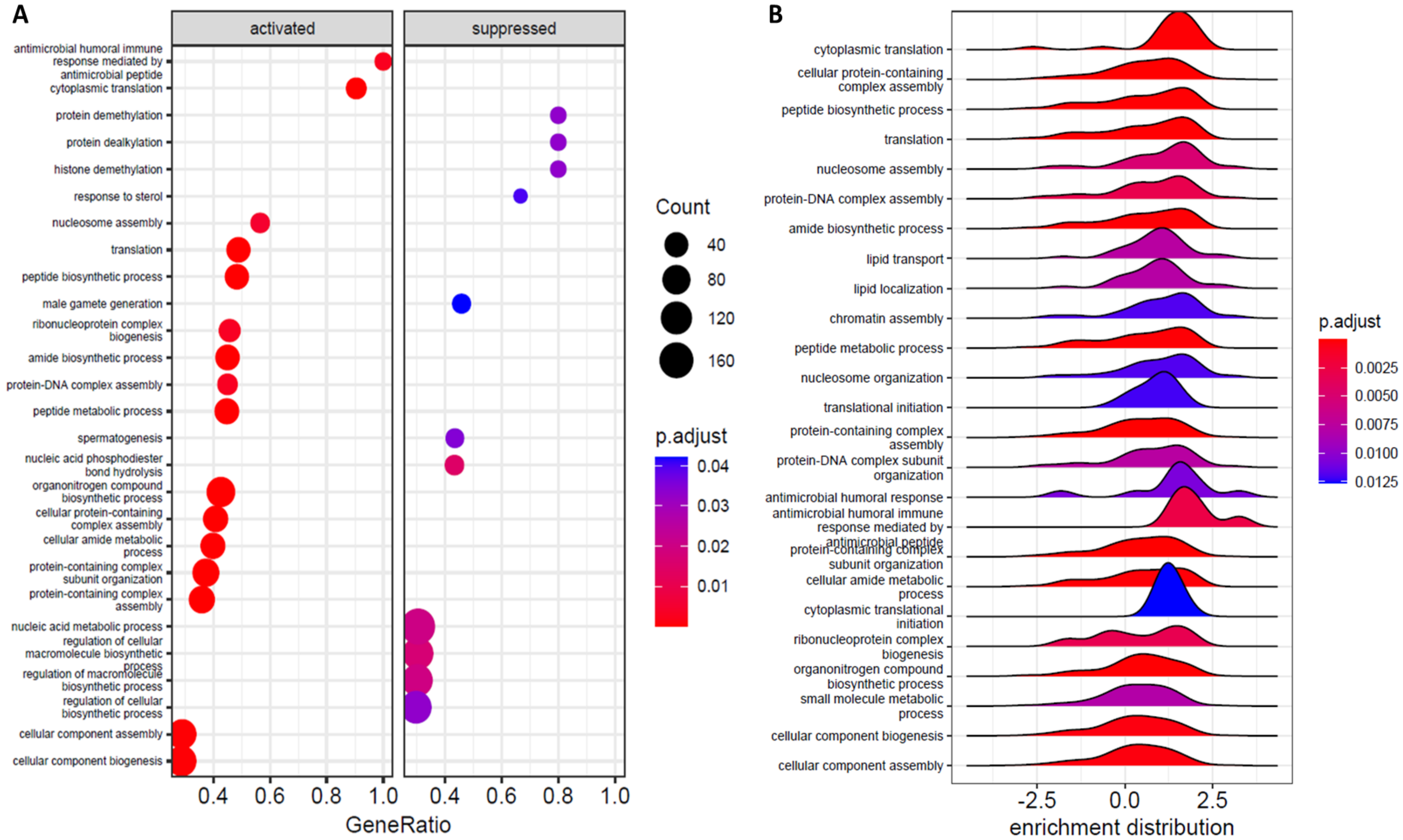


Figure 29 100 most downregulated genes in WT DF-1 608 clonal cell line at 6h time point during an infection with H5N3 subtype of influenza 100 most downregulated genes at FDR cut-off of 0.05 as determined by DESeq2 -

Functional analysis

GSEA

clusterProfiler was used to conduct GSEA to identify differentially regulated pathways in WT DF-1 608 clonal cell line. List of differentially expressed genes with *ashr* adjusted logFCs was tested against gene sets in Gene Ontology database, Biological Process category with a p-value cut off 0.05. As indicated in the Methods section 2.2.5.7.1, maximum gene set size was 800 and minimum 3.


A
activated
suppressed
antimicrobial humoral immune response mediated by antimicrobial peptide
cytoplasmic translation
protein demethylation
protein dealkylation
histone demethylation
response to sterol
nucleosome assembly
translation
peptide biosynthetic process
male gamete generation
ribonucleoprotein complex biogenesis
amide biosynthetic process
protein-DNA complex assembly
peptide metabolic process
spermatogenesis
nucleic acid phosphodiester bond hydrolysis
organonitrogen compound biosynthetic process
cellular protein-containing complex assembly
cellular amide metabolic process
protein-containing complex subunit organization
protein-containing complex assembly
nucleic acid metabolic process
regulation of cellular macromolecule biosynthetic process
regulation of macromolecule biosynthetic process
regulation of cellular biosynthetic process
cellular component assembly
cellular component biogenesis
0.4
0.6
0.8
1.0
0.4
0.6
0.8
1.0
GeneRatio
Count
40
80
120
160
p.adjust
0.04
0.03
0.02
0.01
B
cytoplasmic translation
cellular protein-containing complex assembly
peptide biosynthetic process
translation
nucleosome assembly
protein-DNA complex assembly
amide biosynthetic process
lipid transport
lipid localization
chromatin assembly
peptide metabolic process
nucleosome organization
translational initiation
protein-containing complex assembly
protein-DNA complex subunit organization
antimicrobial humoral response
antimicrobial humoral immune response mediated by antimicrobial peptide
protein-containing complex subunit organization
cellular amide metabolic process
cytoplasmic translational initiation
ribonucleoprotein complex biogenesis
organonitrogen compound biosynthetic process
small molecule metabolic process
cellular component biogenesis
cellular component assembly
-2.5
0.0
2.5
enrichment distribution
p.adjust
0.0025
0.0050
0.0075
0.0100
0.0125

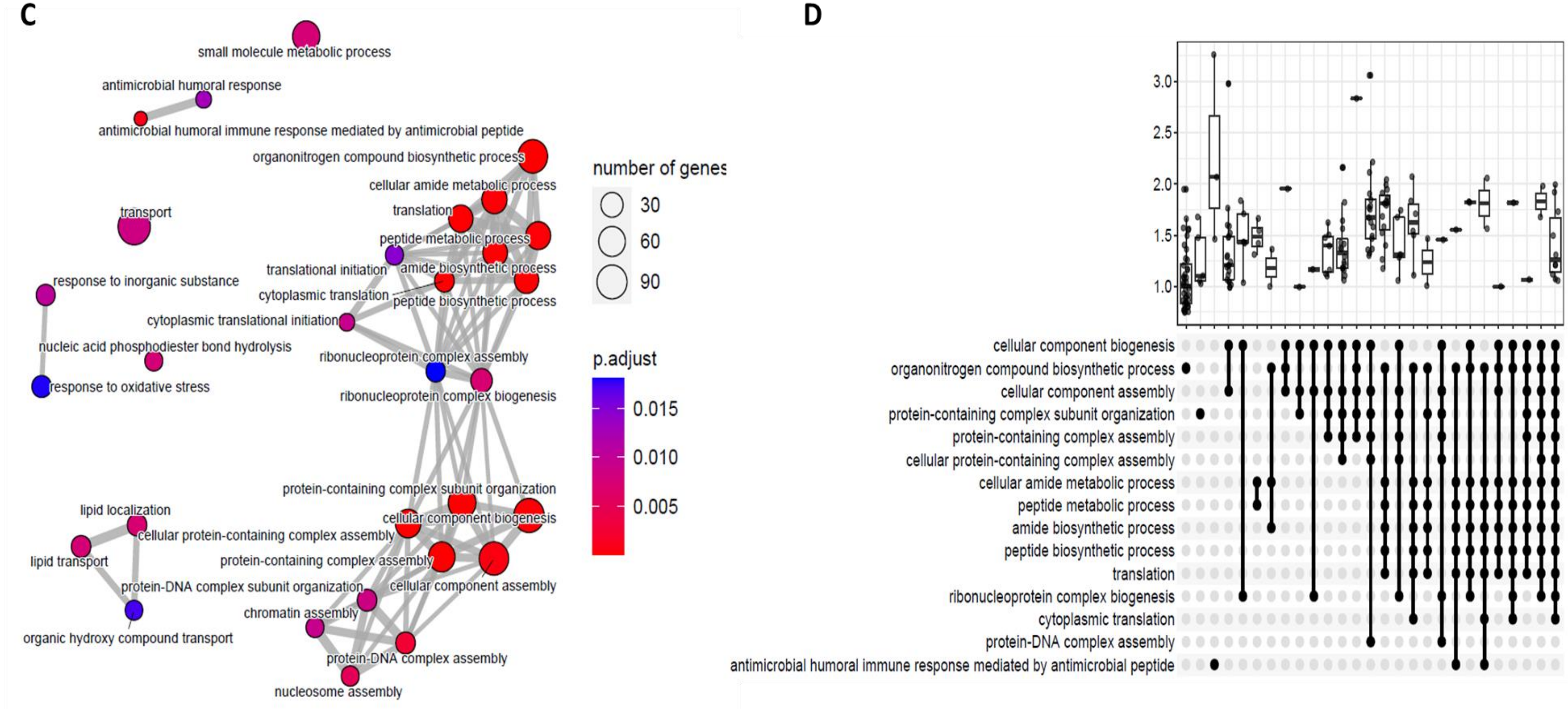


Figure 30 Pathway analysis for comparison infected vs uninfected WT DF1 608 cell line List of differentially expressed genes in comparison infected vs uninfected WT DF-1 608 cell line that passed the threshold of 0.05 FDR were analysed in clusterProfiler in R using GSEA algorithm interrogating GO database: Biological Process category. -P-value cut off was 0.05 and Benjamini-Hochberg method was used for multiple testing correction. **A**) Dotplot of significant enriched pathways. Size of the circle indicates the number of genes in the experimental dataset that was found in the Gene set associated with the pathway. The ratio of those two parameters is plotted on the x-axis. P-adjusted values for each enriched process are indicated by colour of the circle. **B**) Ridgeplot shows the distribution of enrichment scores belonging to a particular pathway that exhibited a corresponding enrichment score plotted on the x-axis. **C**) Emapplot identifies biological themes via creation of a network of enriched pathways. **D**) Upset plot depicts the distribution of fold changes for genes identified in the dataset belonging to a particular gene set. If a gene is a member of various gene sets, the corresponding pathway is marked with a black dot.

Pathway enrichment analysis identified significant upregulation (up to three-fold for some genes) of the antimicrobial humoral immune response (**Figure 30 D).** Antimicrobial immune response and cytoplasmic translation had a very high gene ratio and statistical significance (**Figure 30 A),** despite the low replicate number used in this analysis. Gene sets describing these pathways have a decent size (>30 genes), providing additional level of confidence. Most statistically significant suppressed biological processes in infection of WT DF-1 608 clonal cell line were associated with protein processing (

Figure 30 A). Interestingly, downregulated processes were always less statistically significant in terms of adjusted p-value than the upregulated ones. Genes with the highest fold change belonged uniquely to the microbial immune response and to the complex assembly (

Figure 30 D).

PADOG

It was not possible to conduct the PADOG analysis as the minimal sample size required for the analysis is a triplicate. Triplicate is generally considered an absolute minimum for any statistical analysis. This requirement is even more important in this instance given PADOG conducts a higher-level analysis by calculating differential expression on the pathway level, thereby losing the a degree of freedom that would otherwise be retained on the single gene level. In this pilot experiment duplicates were used due to practical and technical limitations. Nevertheless, the results are still informative for the design of subsequent experiments.

Variance

Identified dysregulated pathways were, on the whole, communal to all clonal cell lines. To drill down further into the source causing the observed variability in phenotype, I investigated the individual genes that drive it.

Infection, specifically with a high MOI, poses an extremely strong assault on the cellular homeostasis. The change in the expression patterns was strongly dominated by the responses to the infection shared by all clonal cell lines, making the extraction of the WT DF-1 608-specific signal difficult. It is plausible to believe that the 608-specific differences are subtle in comparison to overall changes induced by the infection, thus the differential expression of those genes would be more difficult to detect, especially in the absence of more replicates. For the detection of changes at greater resolution a better powered experiment would be required. All the genes driving the variability contained in PC2, which was representative of WT DF-1 608, were also differentially expressed and statistically significant. To investigate to what extent the expression of these genes differs in other cell lines, their

TPM values were Z-scored and clustered using Manhattan distances (**Figure 31**). Manhattan distances are less sensitive to outliers, a feature important in smaller datasets. Given the significantly reduced size of the data set (PC2 vs the whole transcriptome) and the presence of the extreme values in some samples which could not objectively be eliminated, Manhattan distances were used as proximity measure for hierarchical clustering, thus deviating from the procedure explained in Materials and Methods, section 2.2.5.6. *CDH5, PPL, CHMP4C, VSIG10L, LSP1P, KHDRBS3 IL2RA, IL31RA, PAPPA2* and *FSTL5* were more highly expressed in WT DF-1 608 than in all other cell lines, independently of the condition. *HINTW, LOC107049046, ENSGALG00000041221* and *ATP5F1AW* seem to be less expressed in WT DF-1 608, but it is not clear if that is influenced by higher expression of the same genes in the parental line.

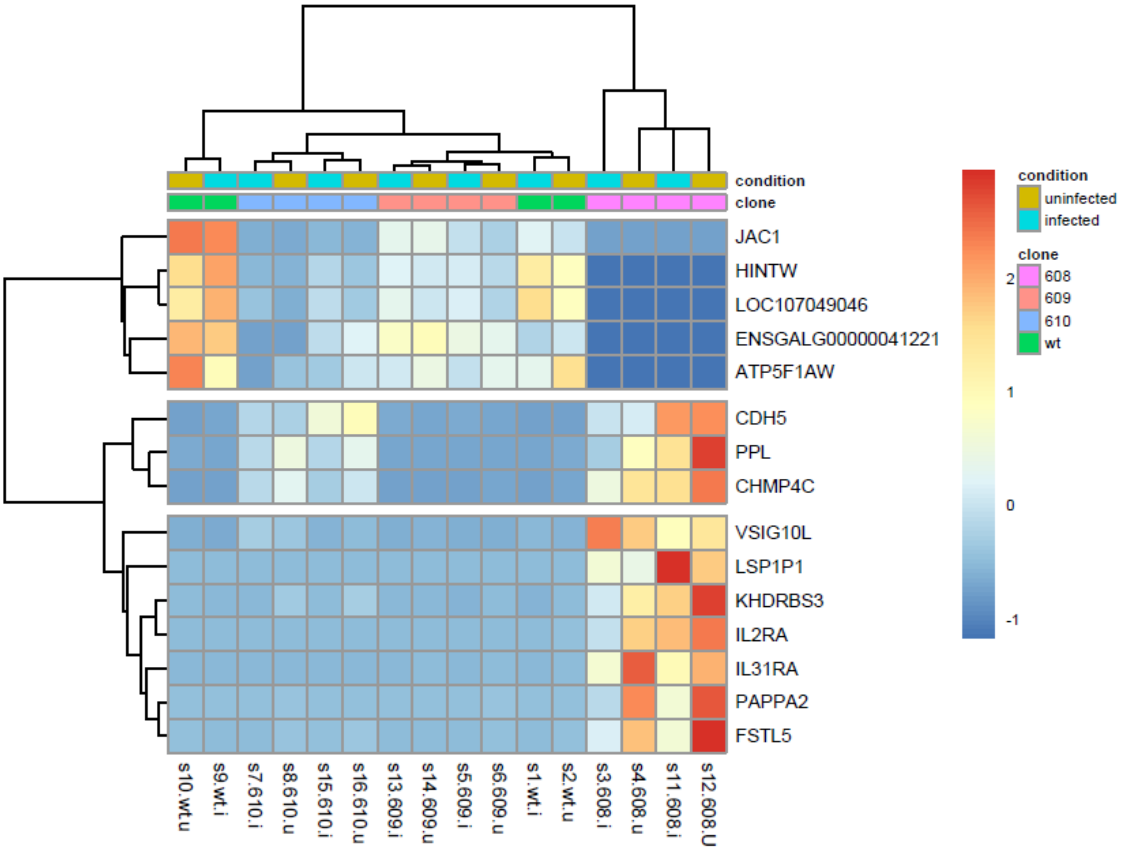


Figure 31 Z-scores for differentially expressed genes that drive the PC2 TPM values for genes that drive PC2 in the WT DF-1 dataset were Z-scored and clustered with *pheatmap* in R using Manhattan distances.

Genes driving the variability contained within PC1 were also processed in the same way and plotted (**Figure 32**). PC1 accounts for the infection and the genes that are differentially regulated as a response to it. This is evidenced by the separation of all sample groups into two clusters: the infected and the uninfected. Note that almost all PC1 genes are downregulated in the infected samples, irrespective of the cell line. Any differences among the cell lines potentially only exist in the homeostatic levels of expression. *CMPK2* and *ERBB3* are the only two upregulated genes among the plotted top 12% of the range of the vector. *CMPK2* is involved in immunomodulatory response [144,145] and *ERBB*3 is a kinase involved in cell proliferation [146,147].

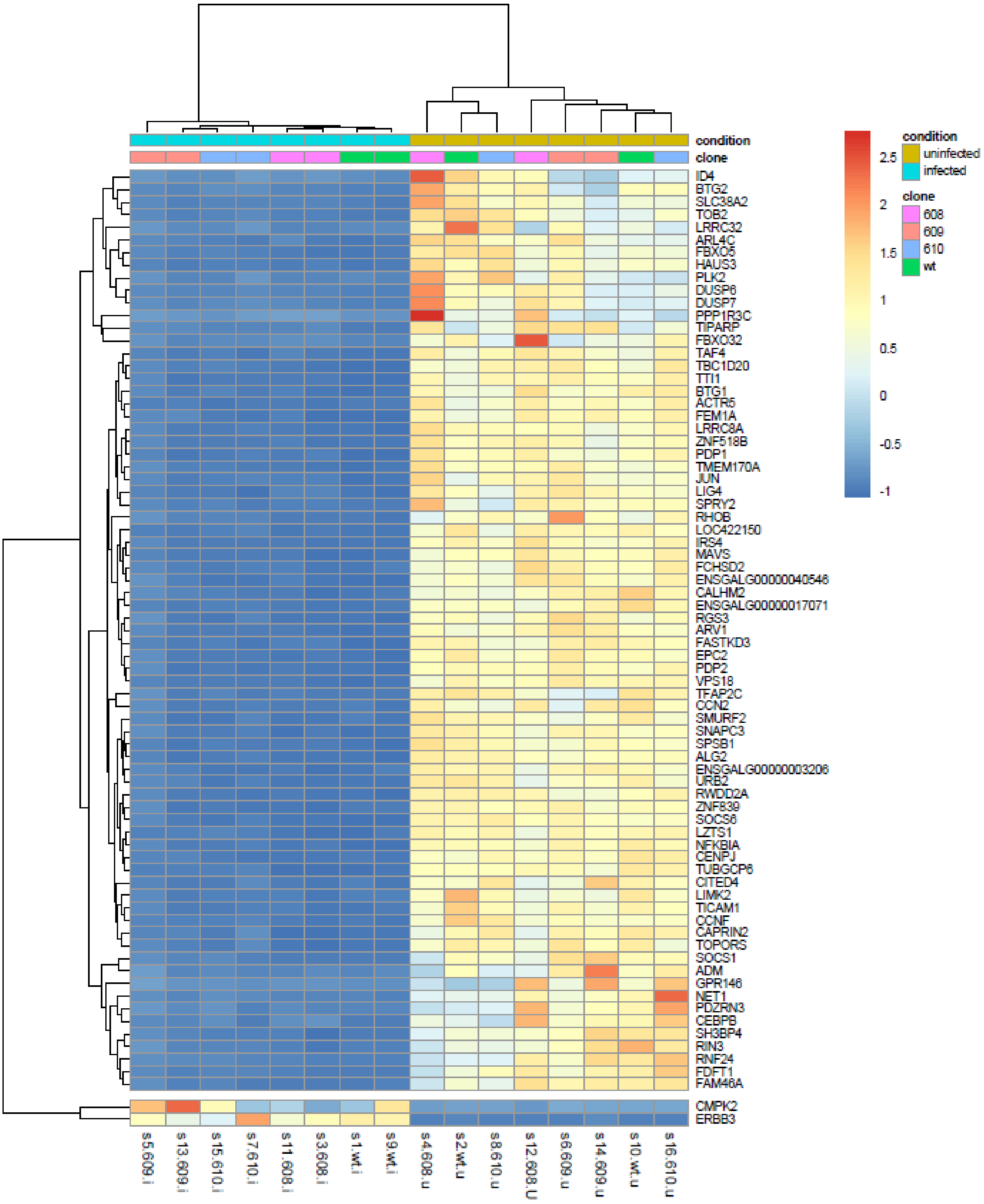


Figure 32 Clustering of Z-scores for genes driving the PC1 TPM values for genes that drive PC1 in the WT DF-1 dataset were Z-scored and clustered using Manhattan distances as metric. Calculation and visualisation were produced in *pheatmap* package in R.

3.4 Discussion

As the ongoing paucity of continuous cell lines free of endogenous viral sequences persists, the DF-1 cell line remains the main workhorse in avian immunology research with a lot of resources invested into their characterisation. The heterogeneity of DF-1 is reflected in non-uniform behaviour exhibited across various properties – morphology, ploidy, growth kinetics etc. Heterogeneity of cell lines used in research, and fibroblasts among them, is not a new phenomenon. Heterogeneity can transpire from different origin cells, their age and epigenetic markup, conditions of culture etc [148,149]. For sophisticated experiments it is crucial to have the cell line characterised with great detail to ensure a sensible design of experiments and the reproducibility of results. A seemingly simple solution for transcriptomic charcterisation, single-cell RNA-seq, is often not accessible due to cost or available expertise for analysis. Here we approach this issue by using single cell clonal cell lines to untangle some of the underlying variation and simultaneously create a stable reagent for avian virology research.

Three clonal cell lines were generated from the parental DF-1 and individually named 608, 609 and 610. True to their clonal origin, few of the examined properties were different among the cell lines. They retained the morphology of a typical fibroblast with a slight variation in size, WT DF-1 608 being the smallest (data not shown). Assessing the permissiveness to the influenza virus had two-pronged rationale for the research presented in this thesis; it allowed the estimation of the required MOI for approximately 100% infection rate as well as providing additional information that could correlate with other investigated parameters, e.g. viral yield. The three clonal cell lines and the parental DF-1 were shown to be almost equally permissive to the influenza infection by flow cytometry. A slight difference in permissiveness was present in individual comparisons of each clone vs the parental cell line, but it very much depended on the virus and it was not statistically significant. We observed a deviation in the transcriptomic profile of WT DF-1 608 and attempt to associate it with phenotype of faster replication, smaller size and higher viral yield. All these properties have multifactorial causal agents and without functional experiments no correlations can be deduced. Pilot experiments serve as exploratory expeditions for hypothesis generation. For confident conclusions better powered experiments should be conducted.

Levels of expression of *chIFITM* genes were also individually comparable among all cell lines, albeit different between the technologies used to assess them. Quantitative RT-PCR detected upregulation of *chIFITM1, chIFITM2* and *chIFITM3* in all WT DF-1 cell lines, while the analysis of the RNA-Seq data showed upregulation of *chIFITM3* and to a much lesser extent *chIFITM2* when the cells were infected with the influenza virus. RNA-seq results reiterated the findings of various studies summarised by Bassano *et al.* [49]. Their analysis showed the expression of *chIFITM1* was virtually absent in many tested *in vivo* and *in vitro* systems, prompting the authors to agree with Smith et al [42] on the hypothesis that an inversion happened in the *chIFITM* locus and that *chIFITM2* is syntenic to *huIFITM1*. A consensus has still not been reached and *chIFITM2* remains un-annotated officially.

It would be interesting to investigate the correlation of expression levels of chIFITMs and the virus yield, especially for WT DF-1 608. However, I speculate that upregulation of IFITMs would happen very early in the infection and either drop or not increase in the long term because they form part of the immediate response. Even the cells which are not infected would upregulate the expression of their innate immune genes as a response to paracrine signalling from neighbouring cells inducing the antiviral state. The measurements of expression levels either by qPCR or RNAseq were done at 6h post infection whereas virus titres were measured later during a multi-cycle infection experiment. The replication kinetics of the clonal cell lines could have a more direct effect on the virus yield and would be worth exploring, especially in the context of the differing phenotype of WT DF-1 608. Synchronising the cell cycles would be necessary to examine the question whether the higher viral yields follow faster doubling times or if they are a consequence of another mechanism.

Statistically significant differences were observed in virus yield: for WT DF-1 608 supported higher level of virus growth than the parental DF-1 throughout the time-course of the experiment. This phenomenon was also observed for WT DF-1 609 later in the infection. These characteristics reflect the underlying heterogeneity of the parental line and the ability to extract a particular phenotype by cellular cloning. This has been reflected to an extent in the transcriptomic profile as judged by the PCA, however no statistically sound conclusions could be drawn from individual comparisons of cells due to low number of replicates.

Karyotyping results showed the same split as the observation of unequal ability of the clonal cell lines to support viral propagation; WT DF-1 608 and 609 trended towards tetraploidy over time while 610 and the parental line trended towards diploidy. O'Hare and Delaney, while studying telomere organisation in chicken models, observed the same trend with the DF-1 cells used in their laboratory: 78% diploid, 14% haploid and 8% tetraploid [132]. This finding is rare; polyploidy is a hallmark of continuous cell lines. DF-1 cells also have unusually high percentage of telomeric sequences – 17% [132], where other chicken lines harbour 3-4% of telomeric content per diploid genome [150]. Ostensibly this confers advantageous characteristics for fitness in culture given that longer telomeres are associated with greater proliferation potential. It is conceivable that multiple chromosome sets are transcriptionally active in 608 and 609 and so relative content of mRNA stays the same. However, elevated levels of genomic content affect almost every aspect of cell biology, cell cycle, cell size, cell surface and distances within the cell, all of which could have secondary effects on viral proliferation.

The aim of the research described in this chapter was to uncover transcriptional changes induced by the process of cellular cloning and the infection as well as to establish the optimal experimental design for subsequent RNA-seq experiments.

Quantifying the infectivity level via flow cytometry ensured that a sufficient proportion of cells were infected for the RNA-seq experiment. In this way the mixed information originating from both infected and non-infected cells in the pool was maximally reduced. Similarity in the response, given that clonal cell lines are being investigated, provides a rationale for pooling the samples to determine the most consistent, thus ostensibly the most important, changes in the transcript levels. This approach concomitantly simulates the heterogeneity of the mixed population of the parental cell line. Finally, the inbuilt flexibility of the design may allow for a separate exploratory analysis of each clonal cell line. However, albeit informative, the underpowered experimental design is a limitation in such an instance thus care must be taken when interpreting the results beyond the hypothesis generating level.

Mapping the sequencing reads to the chicken reference genome resulted in a great proportion of reads remaining unmapped in the infected samples. They were subsequently successfully mapped to the influenza genome sequences, confirming that the influenza virus sequesters host's transcription machinery extremely efficiently. In terms of library size, all samples had more than 10M reads mapping to chicken genome. Only sample 7 (infected replicate of WT DF-1 610 clone) had a low sequencing depth of 5.6 M reads. Wang *et al* [151] found that as low as 10M 75 bp reads are sufficient for detection of 80% chicken annotated genes, equivalent to what is possible using the microarray approach. However, higher sequencing depth of 30M is required to detect lowly expressed genes annotated in the GRCg6a [151]. To ensure a complete coverage of chicken transcriptome, given the influenza sequences occupy a significant proportion of the sequencing real-estate, RNA-seq experiments in following chapters of this thesis contain libraries of 70M reads.

Analyses of RNA-seq experiments, which are very information-rich, are almost always subjective procedures that depend on the scientific question at hand. Well powered transcriptomic experiments are able to detect subtle changes in differential expression which can be statistically significant, but which may not have much biological relevance for the studied condition. Alternatively, genes that have a role in coordinated processes are frequently regulated by a mechanism that does not produce a singular output in terms of either upregulation or downregulation and/or the extent may not meet the cut-off threshold. Additionally, authors generate curated lists of genes of interest from the literature which are heavily influenced by the experimental design, tools used and arbitrary thresholds selected in the original research.

Here, the RNA-seq approach has been taken as means of indirect quantitation of variance seen in cytology and viral yield among clonal cell lines. Rationale for the functional annotation approach taken in this work is elaborated in the section 2.2.5.7. Applying the differential expression and the subsequent gene set enrichment on the gene level and on the pathway level offer the opportunity to confirm whether the results point in the same direction. Reactome extracts its data from Orphanet, Protein Ontology (PRO), IntAct, RHEA, DOCKBlaster, FlyBase, The Human Metabolome Database (HMDB), Zinc, KEGG, UniProt, ENSEMBL, BRENDA and IntEnz to obtain the cross-references for entities annotated in its database. This results in Reactome's ability to define more detailed and precise terms than e.g. KEGG (which contains more broad terms). The ontological entities are additionally cross-referenced against protein interactions, providing another layer of reliability. It is possible to reanalyse the data using some more established tools such as Quigen's IPA and/or against other frequently used databases such as KEGG, both of which require a paid subsription. Within technical and logistical constraints of this work, I believe the selected approach offers the greatest level of confidence given the strictest statistical cut-offs applied and widest selection of databases (via Reactome) covered. With the increase of the amount of information available in the literature, databases become richer and the inferences about functions and interactions more reliable. Especially in the case of chicken, once more functional data becomes available, a repeated analysis of this dataset with better tools may provide even more insights than what I was able to uncover at the time of writing.

Differential expression analysis of WT DF-1 cell lines revealed that thousands of genes are affected by the influenza infection. PCA showed that 78% of the variability among samples can be explained by the infection. Selecting relevant targets to pursue further in the validation experiments or additional analyses can prove to be challenging and biassed. In such cases a more objective approach may be to focus on the pathway analyses and examine the data on the level of biological processes.

Downregulated genes had a relatively stable range of effect-sizes distributed across the range of p-values with no conspicuous data points or trends (**Figure 109**). Upregulated genes were much "noisier" with a selection of genes at the extremes of both the effect size and the statistical significance. However, they were all similarly expressed in all cell lines and while helpful to describe the metabolic response to the viral assault, they were not indicative of differences between the cell lines.

A similar effect was detected at the pathway level where GSEA showed that the dispersion of fold-changes for upregulated processes is wider than for the downregulated (data not shown). This could be a consequence of two-pronged mechanism of control; the reduction of transcription and silencing of already produced transcripts, allowing for stricter control of downregulation.

Both approaches to pathway analysis identified some of the well-known processes which take place during the influenza infection: increased translation and the nucleic acid metabolism, sequestered by the virus for its own multiplication. DNA repair was supressed [152], as was the healthy progression of the cell cycle [153] as the virus is disrupting healthy host transcription while still trying to preserve its replication niche. Cytokine signalling was upregulated, but antigen presentation downregulated [154]. These processes are all manipulated by NS1, the influenza protein whose properties are strain specific and modes of action temporal and multi-levelled [155] allowing for seemingly opposing strategies to take place. This affords fantastic opportunities for viral adaptation depending on the specifics of the host. PADOG additionally identified vesicle mediated transport as downregulated [156,157].

Looking at genes that showed the greatest variance (**Figure 23**) preserves a potentially informative but weak signal that could be otherwise lost in the statistical analysis due to noise. *WISP2, JAC1, ADH1C, CRYL1, PPEF1, ABCA2, UBAP2L2, ATP5F1AW, UBE2R2L, LOC420716, HINTW, LOC107049046, GOLM1 and COLEC12* were less expressed in DF-1 608 than in all other cell lines. The case was inverse for *EFEMP1, HSPB9, IL8L2, VSIG10L, LSP1P1, PAPPA2, KHDRBS3, C1QTNF3, CDH5, PPL.*

Here we highlight *KHDRBS3*, for its role in inhibiting cell proliferation and posttranslational regulation of viral gene-expression via binding of the polyA tail [158] in alignment with what we observed for WT DF-1 608: slower growth kinetics and higher viral yields. *JAC1* is also a positive regulator of growth [159] as is *PAPPA2. LSP1P* and *CDH5* participate in fibrinogen adhesion and organisation of intercellular junctions, respectively. Their upregulation also contributes to the proliferation patterns observed. Upregulation of *CHMP4C (***Figure 31**) may also be significant for phenotype of WT DF-1 608. *CHMP4C* is part of the ESCORT-III – endosomal sorting complex, involved with endocytic vesical formation and sometimes sequestered by viruses for egress. Influenza uses an independent mechanism for budding via its own protein M2 [160], nevertheless the upregulation of *CHMP4C* is indicative of increased membrane activity symptomatic of autophagy, lysosome degradation and transport of cellular components among others. It is conceivable that the observed phenotype is a result of fine-tuned co-expression of many synergic signalling pathways, but further experiments are necessary to confirm that.

In conclusion: despite the WT DF-1 608 showing some clear difference from other cell lines in its ability to support infection, genes that showed most different (higher) expression in WT DF-1 608 did not show condition-specific upregulation, despite having potential roles in infection. And reciprocally, the most significantly differentially expressed genes, especially those with maximum effect size, were not WT DF-1 608 specific.

With greater number of replicates and potentially increased sequencing depth, it will be possible to confidently identify transcriptional signature that translates into the ability of WT DF-1 608 to support virus replication more successfully than the parental cell line. This in turn contributes to the knowledge about heterogeneity of DF-1 cell line.

Analysis in this section showed that the downregulated or the upregulated genes do not show any grouped behaviour that could be detected by clustering, but rather overall similar trend. No specifically immune genes stood out. Identified genes perform functions in signalling pathways in ubiquitous biological processes such as cell proliferation or regulation of cell cycle and apoptosis. Upregulated genes, particularly, were frequently affected by presence of an outlier sample for an individual gene. Given that only duplicate samples per group were used in this pilot experiment, it is likely that a better powered experiment would reduce the noise in the data, facilitating identification of genes of interest.

4

Characterisation of downstream effects of full locus *chIFITM* KO in DF-1

4.1 Introduction

IFITMs are one of many restriction factors that together orchestrate the defence against viruses in many organisms as part of the innate immune system. They are transmembrane proteins localised on plasma membrane or endosomes. They can affect the fluidity of membranes, increasing their rigidity and in so doing disable the virus to release its genetic material into the cytoplasm. Their contribution to the overall immune response is not minimal as proven by the loss-of function experiments across different cell lines and mice models [40,41,161]. Supporting data for the same claim in avian research is less abundant but present nonetheless [42,43,73]. Blyth *et al* expressed duck IFITM in DF-1 cells and infected the cells with H6N2 and H11N9. They saw relative infection drop by 40% in cells overexpressing duck IFITM3 specifically [73]. IFITM knock-down or knock-out studies were mostly concerned with consequences on the progression of infection. Nevertheless, there was one study in 2015 that looked into disruption of homeostatic levels of expressed transcripts caused by knock-down of endogenous IFITM3. Shein *et al* used shRNAs to silence the expression of IFITM3 in HeLa cells and detected differentially expressed genes with microarray [161]. They found the complement pathway to be downregulated as well as some heat shock proteins and KIR receptors. Antigen processing and presentation were upregulated as were families of transcription factors regulating expression of genes involved cellular proliferation, apoptosis, differentiation and oncogenic transformation.

With this work we aim to provide more information about the effect of IFITM knock-out using CRISPR/Cas9 in a relatively less represented model organism: *Gallus gallus*. *G. gallus* holds a very important place in the agricultural landscape as a farming animal used for meat and egg production. Poultry farms benefit from breeding programs and vaccination campaigns that make animals more resistant to infection, thereby reducing losses. In addition to furthering research about the roles of chicken IFITMs in infection, this work aims to provide information about potential roles of IFITM proteins outside of that in innate immune response.

chIFITM overexpression experiments previously conducted in the Genetics and Genomic group showed that chIFITMs play a substantial role in the restriction of the influenza infection. To validate those results a collection of knock-out clones were produced by Horizon Discovery, Cambridge, UK in various cell lines using CRISPR technology. Clones were produced from the mixed population for logistical reasons and the unavailability of functional characterisation data for the individual WT DF-1 clones. Admittedly, producing the KO clones from a single WT clone would constitute a better experimental design and would provide a more direct investigation of the KO success rate and the off-target effects of a particular set of gRNAs However, here we were interested in the downstream effects of the *chIFITM* KO, at a higher resolution, making the cost-benefit analysis of the production of the KO from a single WT clone not favourable in terms of feasibility.

In this chapter we characterise the transcriptomic profile of the resulting KO clones in relation to the response to the infection in the absence of chIFITM genes/products as well as in terms of downstream effects of the chIFITM KO by comparing basal expression alterations between the WT and KO cell lines in the uninfected state.

4.2 Objectives

- Generate stable *chIFITM* KO clonal cell lines
- Investigate the phenotypical changes of the *chIFITM* KO DF-1 cell lines
- Characterise changes in the transcriptomic profile in the KO DF-1 clones relative to the WT DF-1 parental cell line
- Establish the effect of the *chIFITM* locus knock-out in DF-1 cell lines on transcriptomic level during an infection with an influenza virus

4.3 Results

4.3.1 Downstream effects of the full locus chIFITM KO in DF-1 cell lines

4.3.1.1 Confirmation of the chIFITM locus KO

To confirm that the *chIFITM* knock-out was successful, end-point PCR and RNA-Seq techniques were used to detect the presence and expression of *chIFITM* genes, respectively.

Successful knock-out of the chIFITM locus in the three DF-1 cell lines was confirmed by RT-qPCR by Dr Ahmed Ibrahim (data not shown). Here described is an additional PCR experiment designed to confirm the knock-out in the Primordial Germ Cell (PGC) derived chicken fibroblasts (see 5.3.1.2). Primers were designed to target the genes that flank the locus resulting in a short product in the KO genotype (see 2.1.3.3., and **Table 6**). No product was expected for the WT control because because the elongation time in the PCR protocol is too short for a product larger than 2kb (see Materials and Methods 2.2.3.2). When combined with primers targeting respective *chIFITM* genes (**Table 7),** the design predicts two products in the WT genotype: one from each end of the locus, and none in the KO given the premise in this case is that their target binding sequence is absent.

Table 6 Primers including flanking genes for endpoint PCR to confirm chIFITM KO in PGC-derived fibroblasts

Primer name	Primer sequence	5'-3' Orientation
ATHL1_F_11870	CTGCTCTGTACAATTCCAGGAG	Forward
IFITM5_F_500	CTGAGGCTGGGCTGGAGA	Forward
IFITM3_R_620	TCCACAGGATTCTGTGGGGTCCACG	Reverse
B4GALNT4_R_31	CAGCTCTTATAGCTGTCCCACCG	Reverse

Table 7 Predicted sizes of the PCR products

Primer combinations:	Approximate product size
1. ATHL1_IFITM5	1878 bp
2. IFITM3_B4GALNT4	1458 bp
3. ATHL1_B4GALNT4	KO strategy dependent

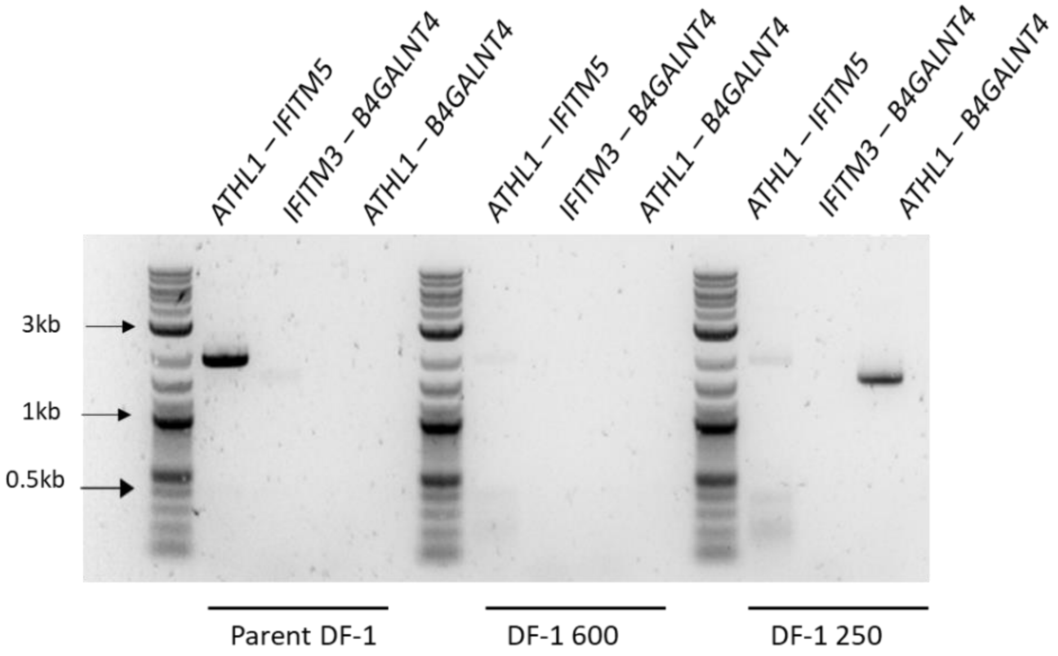


***Figure 33 Figure 44 Confirmation of the chIFITM locus KO in DF-1 fibroblasts** Primers targeting ATHL1, chIFITM5, chIFITM3 and B4GALNT4 used in different combinations. Resulting bands are indicative of the KO-strategy applied in each clone. Combination 2 failed* to produce amplicons due to the suboptimal melting temperature in the PCR protocol.

This PCR was applied to two DF-1 KO clones; DF-1 250 and DF-600 only. The DNA from the DF-1 225 clonal cell line was not available at the time of the experiement due to a contamination incident. It was reasoned that given the KO strategy being the same as for the clone 250, the PCR design if successful would be suitable for 225 as well. Successful KO in 225 was previously confirmed at Horizon Discovery via PCR (data not shown) and subsequently by me with the RNAseq data (***Figure 35***). PCR generated a band of approximately 1.7kb in clone 250 only.

As expected, clone DF-1 600 showed no bands in PCR 1 and 2 because IFITM5_F_500 and IFITM3_R_620 primers bind in the region of exon 1 of *chIFITM5* and exon2 of *chIFITM3* that have been excised in the CRISPR strategy used to produce the DF-1 600 (**Figure 34 A.**).

For PCR primer combination three no product was expected due to technical limitations of this particular protocol (2.2.3.2). Primer combination three ecompasses the entire locus (>10kb) and the product could not be amplified reliably. CRISPR strategy deployed for generation of clone DF-1 250 abolishes the *IFITM5*, thus shortening the final sequence (**Figure 34 B.**) which is hence successfully amplified.

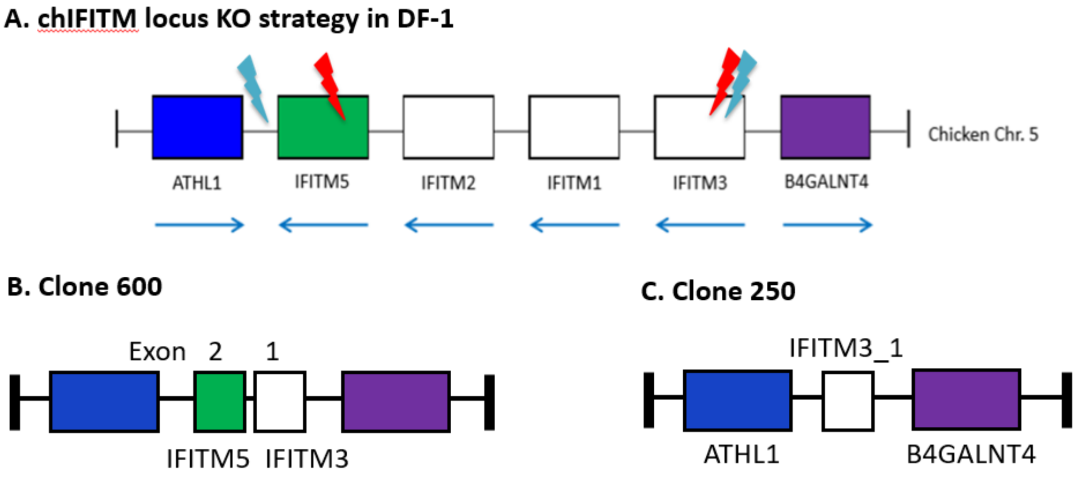


Figure 34 Schematic of CRISPR chIFITM whole locus KO strategy for DF-1 cells Two alternative KO strategies for full chIFITM locus KO **A.** Depiction of the chIFITM locus (borrowed from [42]. Light blue and red bolts mark the approximate target locations of gRNAs used to produce clone 600 and 250, respectively. **B.** KO strategy for Clone 600 consisted of gRNAs targeting exon 2 of chIFITM5 and exon 1 of chIFITM3 thus generating a product capable of producing a fusion protein potentially retaining some residual function **C.** KO strategy for clone 250 targeting sequences outside the locus resulting in a truncated sequence of exon2 of chIFITM3 thus avoiding the possibility of generation of a novel fusion protein

To confirm there was no functional mRNA originating from chIFITM genes, we utilised the RNA-Seq data whose generation is described in the following section. Briefly – RNA from WT and KO DF-1 clonal cell lines was sequenced on an Illuumina machine and the reads aligned to chicken genome reference using STAR. Resulting BAM files were imported into SeqMonk [140] and put through RNA-quantitation pipeline. **Figure 35** shows the relevant annotation tracks for chIFITM locus. Sample names on the left indicate which tracks belong to which cell line. Breakdown of the naming convention is as follows:

e.g. S13DF1WTU1_KO_BothAligned

> S13 is the sample number
> DF1WT is the cell line
> U means uninfected (while infected samples are marked with "I")
> 1 is the replicate number
> KO_ BothAligned is the mapping strategy used with STAR

Raw reads are depicted by red and blue dashed lines above the tracks. Red signifies the reads originate from the forward strand and blue from the reverse. *chIFITM* genes are found in the locus in the following order from left to right: IFITM3, IFITM1, IFITM2 and IFITM5. Annotation track contains the following features: gene, mRNA, CDS and custom track (*chIFITM* locus). chIFITM2 was not officially annotated at the time and thus is missing from the track. Below the reads we can identify the bars representing the relative levels of expression. Height and the warmth of the colour indicate an increase in the detected levels of mRNA. Quantitation included only the reads mapping to exons, taking into account the directionality of the library. Resulting values were log transformed for visualisation purposes.

Using this approach, I was unable to detect the presence of *chIFITM* reads in all KO DF-1 clonal cell lines 225 and 250, indicated by the blue, "upside-down" bars representing a negative log. Weak signal was detected in KO DF-1 600 cell line for reads mapping to *chIFITM3* and *chIFITM5*. This was expected because the CRISPR approach employed in this cell line results in a fusion product between exon 1 of *chIFITM3* and exon 2 of *chIFITM5* (**Figure 34**). A relatively higher level of expression of *chIFITM3* and *chIFITM2* was observed in comparison to *chIFITM1* and *chIFITM5* in the same sample (third replicate was omitted for all groups for visualisation purposes).

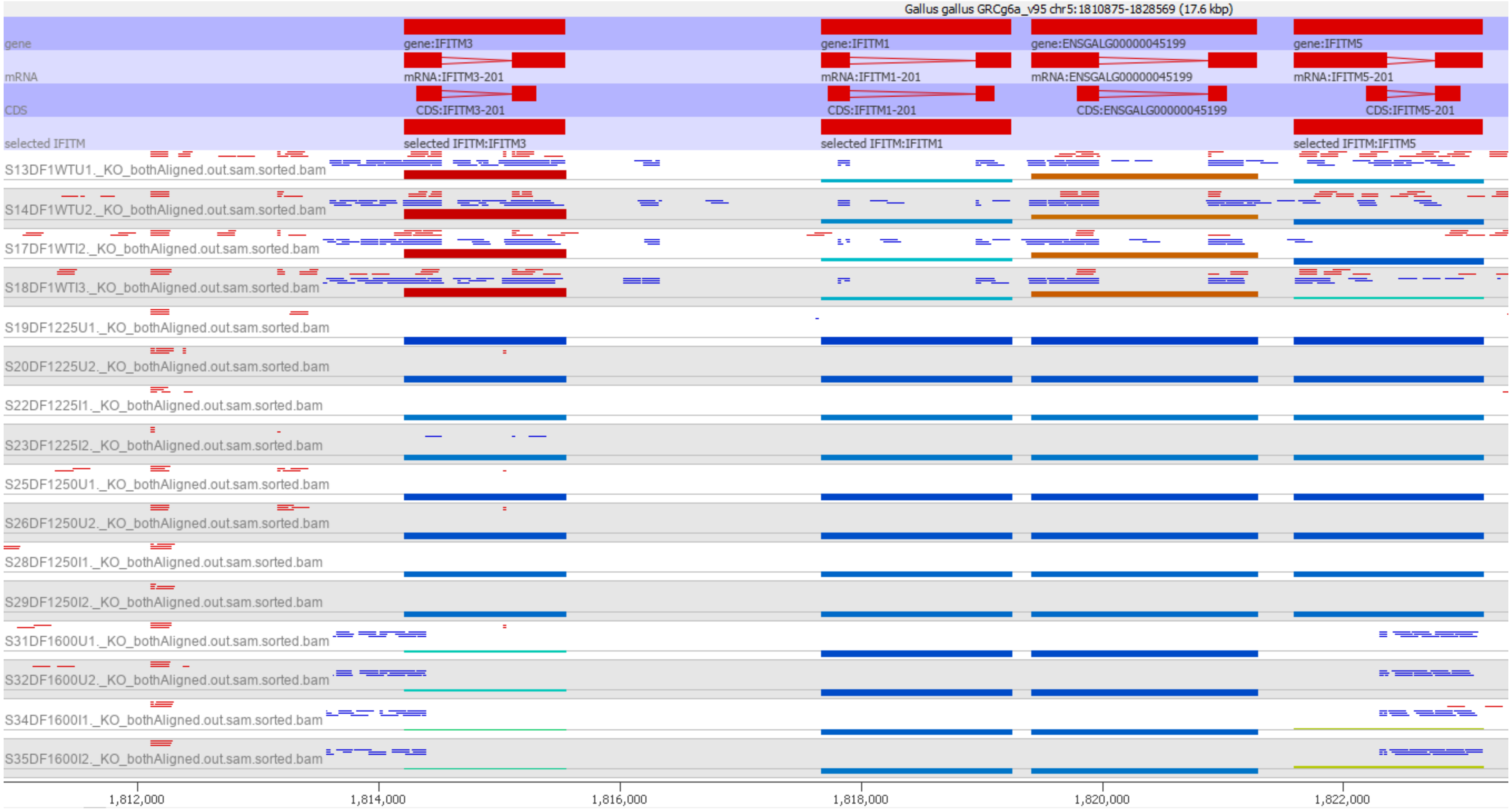


Figure 35 **Visualisation of reads mapping to *chIFITM* locus** Bulk RNA-Seq was conducted on Illumina NovaSeq and reads were aligned to Gallus gallus GRCg6a_v95 using STAR vSTAR 2.7.2a. Visualisation was produced with SeqMonk v1.48.0, using only the uniquely mapped reads. Raw reads are coloured red or blue indicating they map to forward or reverse strand, respectively. Tracks show the *chIFITM* locus with individual genes from left to right: *IFITM3, IFITM1, IFITM2* and *IFITM5* and corresponding features from top to bottom: gene, mRNA, CDS and custom track (*chIFITM* locus). *IFITM2* is missing from the "selected IFITMs" track because the annotation information that was available in the reference did not include the gene name. The height of the bars underneath the reads reflects relative expression levels for the genes, as does their colour: ranging from red (hot) to blue (cold). Quantitation was conducted over exons, assuming a directional reverse library and represents read count normalised for library depth and subsequently log transformed.

4.3.1.2 chIFITM KO confers the ability to support higher levels of virus replication in DF-1 cells

As elaborated already, one of the main functions of IFITMs is that of a restriction factor. Previous work in the laboratory confirmed this in DF-1 cells with over-expression experiments and transient knock-down (data not shown). To validate those results chIFITM KO cell lines were generated and infected with a range of viruses to test if the loss of those genes confers greater permissiveness and/or the ability to support viral infection.

chIFITM KO DF-1 clonal cell lines 225, 250 and 600 as well as the WT parental cell line all showed similar properties in terms of growth kinetics and overall morphology. Splitting ratios and frequency of subculturing were always the same for all cell lines. Morphologically, in terms of size and shape, they did not differ from the parental line (personal communication with dr Ahmed Ibrahim).

were infected with A/duck/Singapore/3/1997 H5N3 and A/Hong Kong/4801/2014/PR8 6:2 H3N2 influenza viruses at MOI 0.01 following the protocol detailed in section 2.2.2.4. Supernatant fluid was collected at 15, 24 and 48h post infection and titred using the plaque assay (section 2.2.2.5).

All clonal cell lines are able to support higher viral replication than the parent (**Figure 39**). Clone 250 had higher viral titres which were statistically significant at all tested time points and for both strains of influenza, as did clonal cell line 250. This was similarly the case for clone 600 as well, except for the earliest time point, 15h post infection with H3N2 subtype of influenza, which did not reach statistical significance. Viral titres appear stable across all time point with no increase over time for all cell lines tested. Differences in viral titres between the clonal cell line and the parent DF-1 range from 0.5 to 1.5 log10 and do not show any meaningful reproducible trend of increase or decrease for any of the viruses or cell lines tested.

In addition to similar profile in terms of being able to support higher viral titres for all KO clonal cell lines, they also exhibited very similar growth kinetics and morphology (personal communication with Dr Ahmed Ibrahim).

4.3.1.3 Karyotyping of the chIFITM KO clonal DF-1 cell lines

CRISPR/Cas9 is known to cause off target effects which can range from SNPs to large-scare rearrangements and chromosome loss in extreme cases [162]. To assess genomic stability of the chIFITM KO DF-1 clonal cell lines, karyotypic assessment was conducted by KaryoLogic, Inc., NC, USA, using the same protocol as for WT DF-1 clonal cell lines (see section 3.3.1.4). Briefly, eight macrochromosomes and the sex chromosomes were analysed in 20 spreads of two passages each: one early and one late for each *chIFITM* KO DF-1 cell line: 225, 250 and 600.

The most unstable clone was 225; moreover, the variation increased with the passage number. At late passage (p53) a partial deletion of the Z chromosome arm appeared in 15% of all spreads analysed. There were other aberrations, such as additional chromatin of unknown origin attached to the q arm of chromosome 2 and 4, that were not consistently present in all spreads (**Figure 36**). In terms of ploidy, there was no observable trend over time in either direction i.e. increase or decrease. The number of chromosomes varied from 18 to 39 among spreads with no particular grouping.

chIFITM KO DF-1 250 was the most stable clone. No significant changes in the chromosomal integrity or number were observed between the two stages. Spreads in both stages displayed 60-70% a diploid number of chromosomes and 40-30% tetraploid. All identified aberrations were persistent through passaging and were displayed in duplicate in tetraploid cells (**Figure 37**).

chIFITM KO DF-1 600 trended from tetra-ploidy towards diploidy as the cells aged. At passage p39 one half of the spreads were near-diploid, and one half were near-tetraploid or polyploid (**Figure 38**). This ratio shifted to 95% diploid at the passage 59. Chromosome count varied from 16 to 46. In addition, some unique aberrations appeared in 10% of the spreads by passage 59.

Consistent features that persisted across all clones and time span were the markers and a single Z chromosome, as well as additional chromatin at the q arm of chromosome 1 and 4.

Chicken *IFITM* genes are localised on chromosome 5. The deletion of the *chIFITM* locus did not cause any structural changes that could be detected by karyotyping. Chromosome 5 was readily identified in all analysed spreads except one: passage 30 of KO DF-1 250 cells. Its ploidy relative to the other chromosomes remained stable across passaging for all cell lines. Stable cell lines frequently display a non-singular ploidy profile through time. Tendencies towards increased ploidy are a hallmark of tumorigenic origins of many laboratory adapted cell lines. Here we show that some variability among cell lines examined exists, but it is not possible to confidently relate it to the observed phenotype.

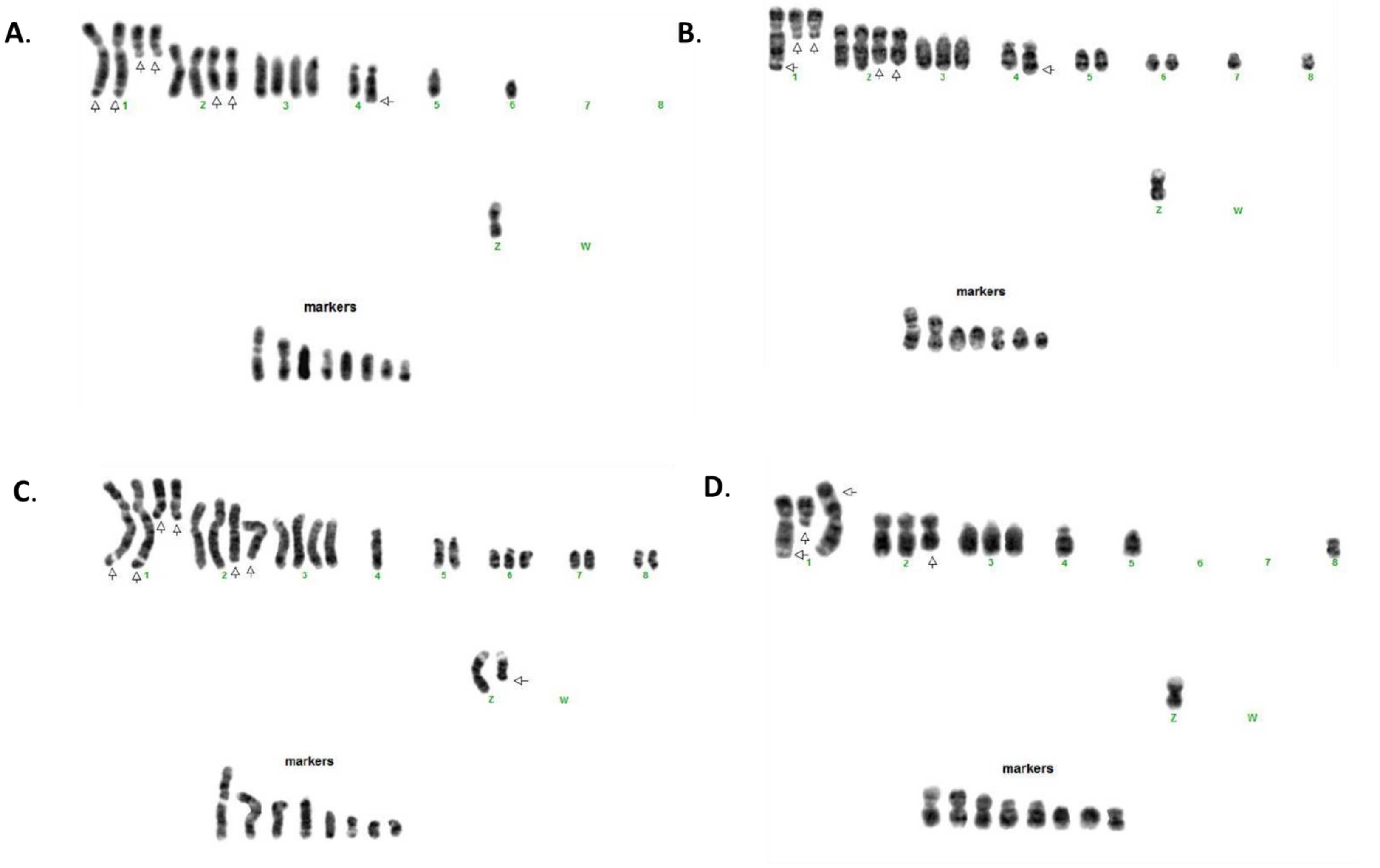


Figure 36 Karyotype for chIFITM KO DF-1 225 clonal cell line A representative spread of metaphase stage chromatin stained with Giemsa stain. Panels A and B show two characteristic karyotypes at passage 33. C. and D. panels show examples at assage 53. Arrows indicate deviations from the normal chromosomal structures (e.g. deletions and insertions of chromatin). A marker is defined as “a structurally abnormal chromosome that cannot be unambiguously identified or characterized by conventional banding cytogenetics” ([139].

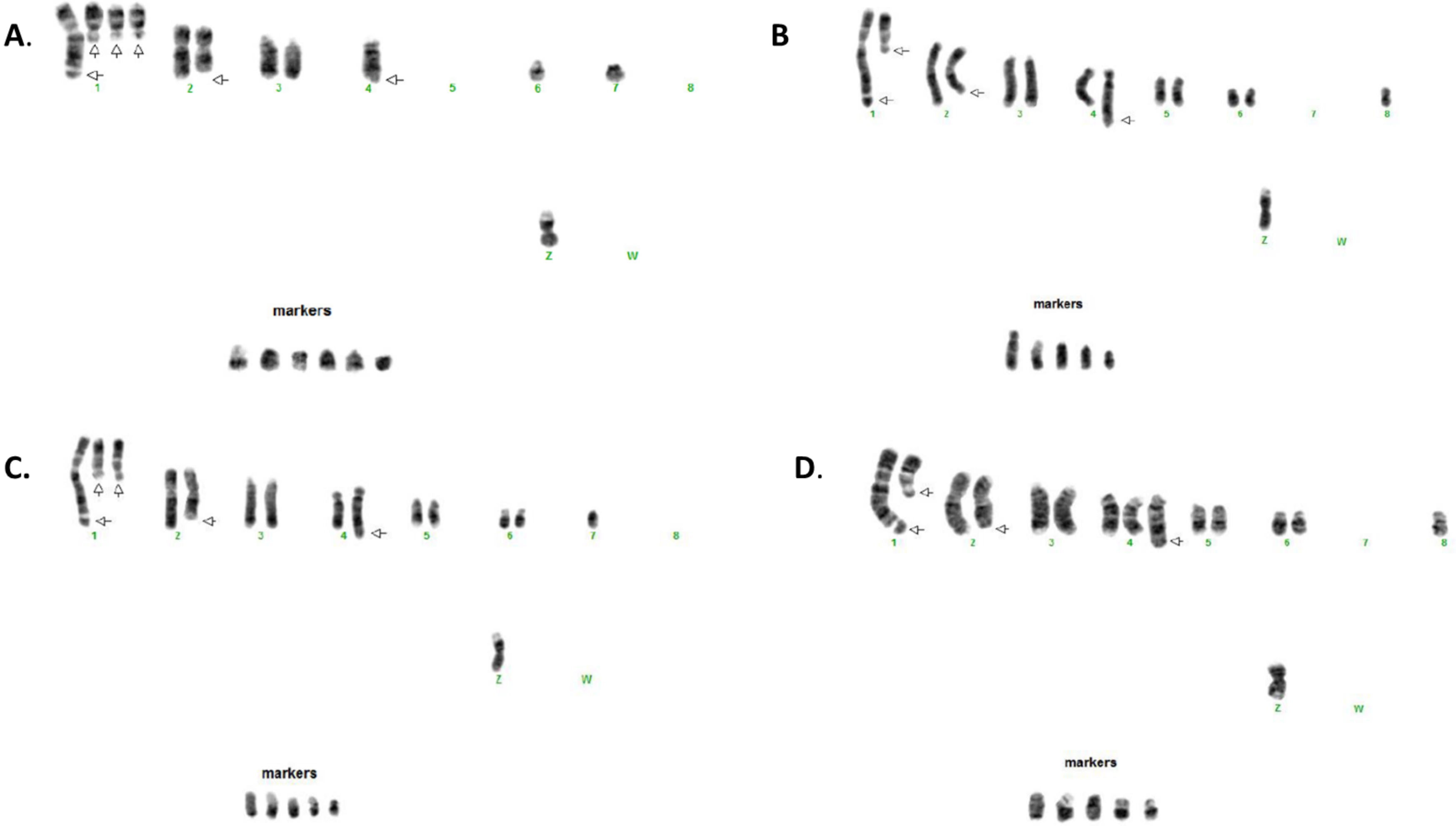


Figure 37 Karyotype for chIFITM KO DF-1 250 clonal cell line Representative karyotypes of metaphase stage G-stained chromatin. Panels A and B show two characteristic karyotypes at passage 30. C. and D. panels show examples at passage 50. Arrows point to structural aberrations (e.g. deletions and insertions of chromatin). A marker is defined as “a structurally abnormal chromosome that cannot be unambiguously identified or characterized by conventional banding cytogenetics” ([139].

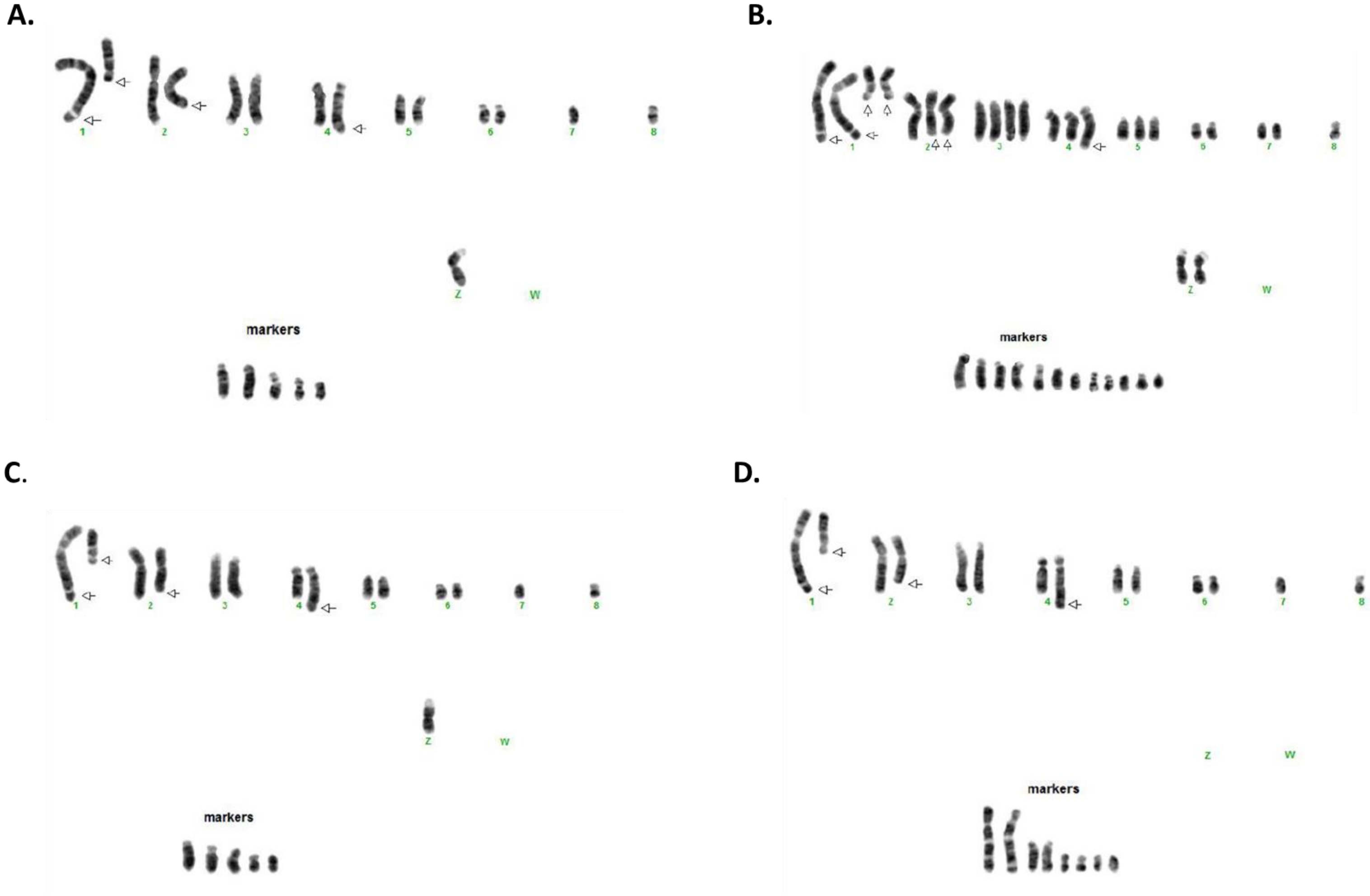


Figure 38 Karyotype for chIFITM KO DF-1 600 clonal cell line Representative karyotypes of metaphase stage G-stained chromatin. Panels A and B show two characteristic karyotypes at passage 39. C. and D. panels show examples at passage 59. Arrows point to structural aberrations (e.g. deletions and insertions of chromatin). A marker is defined as "a structurally abnormal chromosome that cannot be unambiguously identified or characterized by conventional banding cytogenetics" ([139].

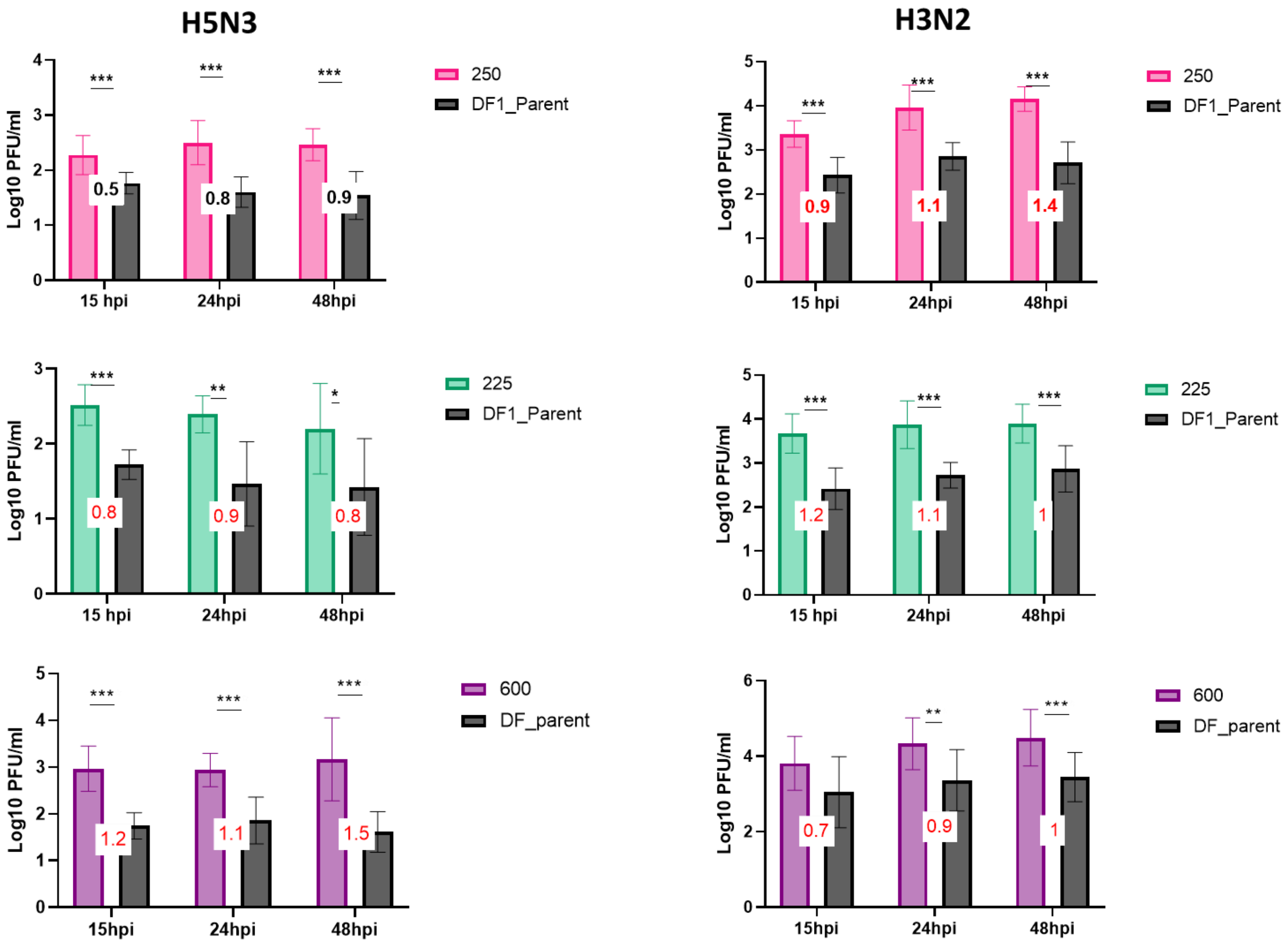


Figure 39. ***DF-1 parental and KO cell lines infected with 2 subtypes of AIV*** *Parent WT DF-1 and chiFITM KO clonal cell lines: 225, 250 and 600 were infected with A/duck/Singapore/3/1997 H5N3 and A/Hong Kong/4801/2014/PR8 6:2 H3N2 at MOI 0.01. Supernatant fluid was collected at 15, 24 and 48 hours post infection and assayed by plaque assay in MDCK cells. Numbers in red within the white box indicate the difference in titre (log10) between the KO clonal and the parental cell line. T-tests were performed for comparison to DF-1 parental line, with statistical significance denoted by * p<0.05, ** p<0.01 and ***p<0.001. Data were kindly supplied by Dr Ahmed Ibrahim.*

4.3.1.4 *Whole transcriptome sequencing*

Experimental set-up

Experimental set-up for the assessment of the transcriptomic response of the IFITM knock-out cell lines was identical to the previous chapter. Briefly – clonal cell lines KO DF1: 225, 250 and 600 together with the parental WT DF1 cell line were grown in 6-well plates to 80% confluency. They were infected with A/duck/Sing-Q/119/97 (H5N3) influenza following the procedure detailed in 2.2.2.4. At 6h post infection, the cells and the supernatant were harvested. Half of cells ware frozen at -80°C in RLT buffer (RNeasy kit, Qiagen) for RNA extraction at a later date and half ware processed immediately for flow cytometry to confirm the infection status (**Figure 40**). The procedure for staining the influenza NP-protein is described in 2.2.3.11. Infection experiments were conducted three times in technical duplicates, separated by at least one week to emulate biological variation. Flow cytometry confirmed successful infection for infected samples. The level of infection (>80% of infected cells) was deemed suitable for RNA-Seq thus all samples were taken forward for processing. RNA was extracted (following protocol in 2.2.3.5) and its quality assessed with Tapestation (section 2.2.3.7) on the same day for all 36 samples to eliminate any potential batch effect. Infections and the RNA extractions were done simultaneously for both KO systems: the DF-1 and the PGC (see section 5.3.2).

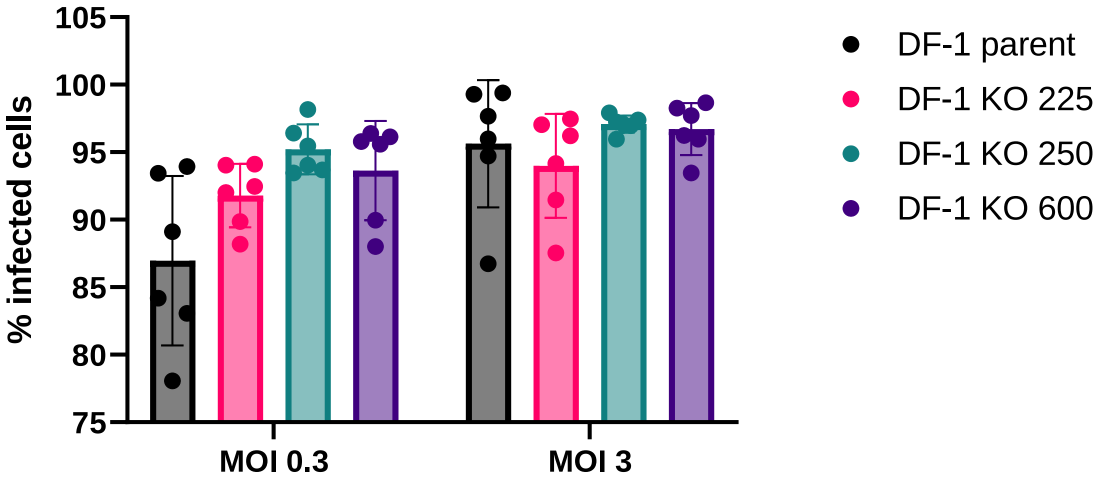


Figure 40 A/duck/Sing-Q/119/97 (H5N3) influenza successfully infects all DF-1 cell lines equally Parental WT DF-1 cell line and KO DF-1 clonal cell lines 225, 250 and 600 were infected with A/duck/Sing-Q/119/97 (H5N3) influenza virus at MOIs 0.3 and 3. Cells were collected at 6h post infection and processed for flow cytometry for detection of influenza NP protein. Bars represent ±1SD. Experiment was conducted three times in technical duplicates.

Library preparation

Library preparation was done at the Earlham Institute using their robotised pipeline described in 2.2.5.2. All RNA samples passed the quality assessment and were taken forward to RNA-sequencing with the aim of generating a library of 70M 150bp PE reads per sample.

Sequencing and alignment quality control (QC)

RNA-Seq data was downloaded from the cluster at the Earlham Institute and assessed for quality using FastQC [84]. All samples exhibited very high quality therefore no trimming was applied. *Gallus gallus* GRCg6a downloaded from Ensembl [86,87] was used as the genome reference and the alignment was performed with STAR [85] to maintain consistency with previous work. All QC was conducted with Qualimap ([90,91] and the reports summarised with MultiQC [92].

With this data set as well, alignment to only the chicken reference resulted in 90% mapping in the uninfected samples and on average 40% for infected samples. To address this, a reference genome comprising of GRCg6a reference with concatenated influenza sequences belonging to the A/duck/Sing-Q/119/97 (H5N3) strain was used. This improved the mapping rate to 90% for all samples (**Figure 41**). BAM files resulting from the alignment were put into SeqMonk ([140] to visualise library size distributions and mapping quality (**Figure 42**). Results show on average 90% of reads mapped to exons, indicating a successful mapping. No contamination with mtRNA or rRNA was detected. Library sizes spanned one order of magnitude; the smallest libraries were approximately 30% of the largest library number of aligned reads. This range is acceptable for any transformations subsequently used as well as the differential expression calculations.

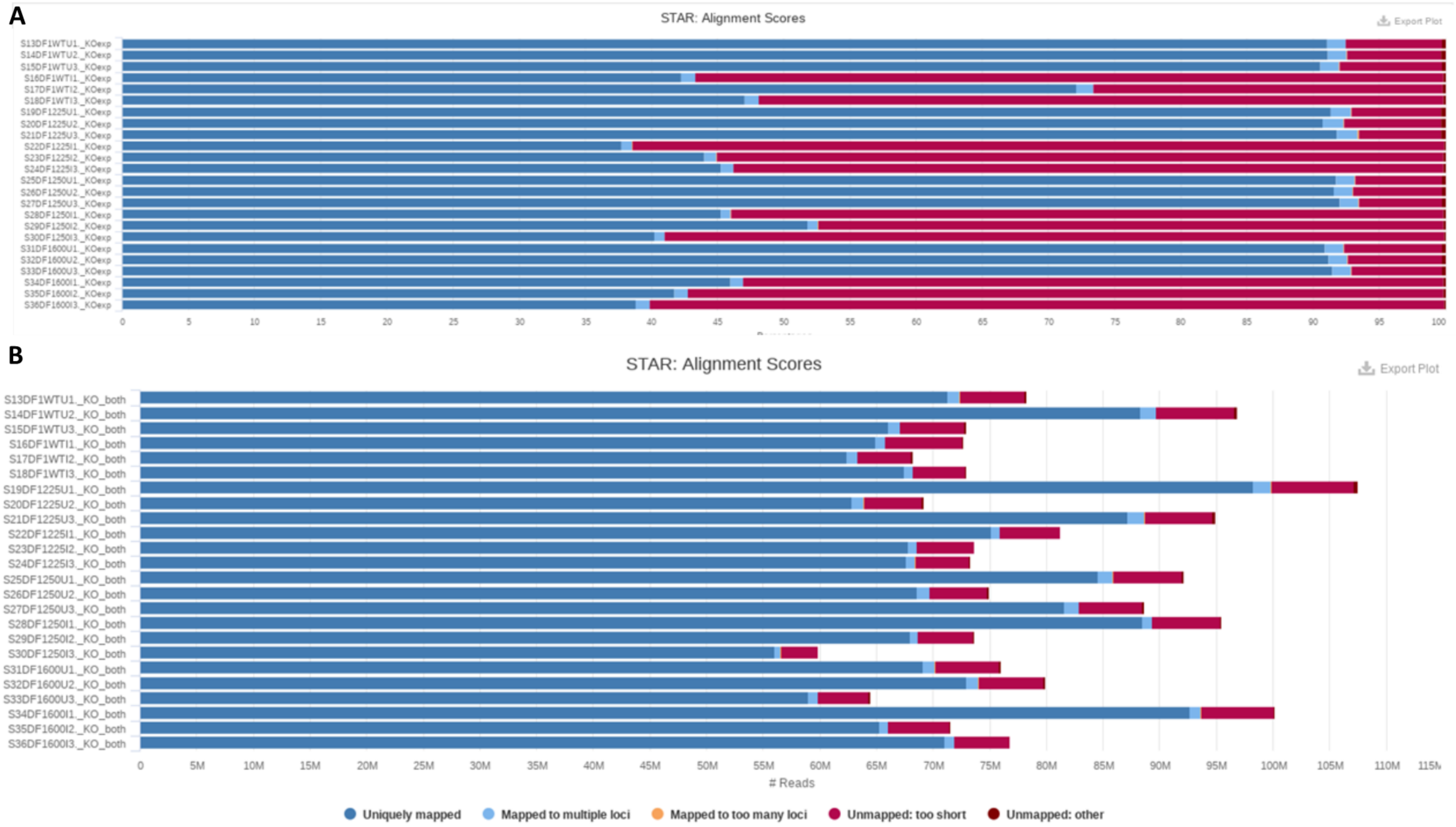


Figure 41 MultiQC summarisation of mapping quality for KO DF-1 RNA-seq dataset WT DF-1 STAR was used for alignment of RNA-seq reads to the reference genome GRCg6a. Panel A shows the percentages of reads successfully aligned to chicken genome. Panel B shows mapping outcome when a reference genome is comprised of chicken genome and the concatenated influenza sequences. Successfully matched reads are represented by the blue segment of the bar and unmapped by red. Library sizes are plotted on the x-axis and expressed in millions (M). Summarisation of individual alignment reports and figure were produced with MultiQC.

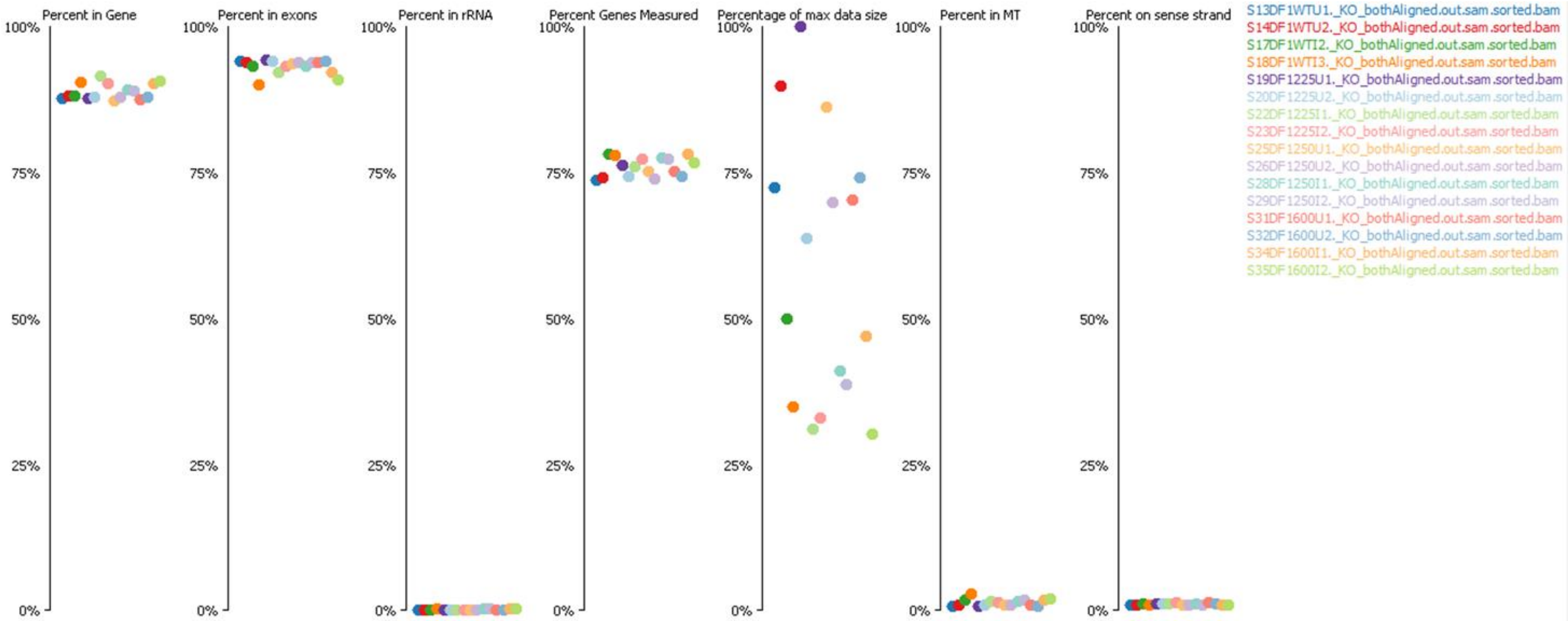


Figure 42 Mapping QC of the KO DF-1 RNA-Seq dataset WT parental and KO DF-1 clonal cell lines 225, 250 and 600 were infected with the influenza virus. 6h post infection RNA was extracted and subjected to RNA-Seq. Reads were mapped to GRCg6a reference genome using STAR. The resulting BAM files were assessed using SeqMonk RNA-Seq QC pipeline.

Exploratory RNA-seq data analysis

The data set sequenced at the Earlham Institute consisted of 36 samples: 12 PGC samples and 28 DF-1 samples. The RNA extraction and the RNA sequencing was conducted for all samples together. This had the potential of being beneficial for data normalisation and differential expression analysis because of the increased sample size. Bigger sample sizes facilitate more accurate estimation of the population parameters resulting in higher statistical power. Because all samples went through the same treatment during processing and the cell lines went through the same experimental set-up as well as both being chicken fibroblasts, it was reasonable to believe the gene expression would behave in a similar fashion.

To test this assumption, a PCA was conducted to estimate the variation in the gene expression levels in the whole dataset. Additionally, PCA allows for identification of variables that meaningfully contribute to the phenotype and thus should be included in statistical modelling. Dataset was normalised using *vst* and rlog (data not shown) using DESEq2 [94] package in R. *Vst* was found to better stabilise the variance and was subsequently used to transform the data and proceed to the PCA. PCA analysis showed extreme separation of PGC and DF-1 cell lines across the PC1 which accounted for 93% of total variation in the data (**Figure 43**). PC2 explained 3% of the variation.

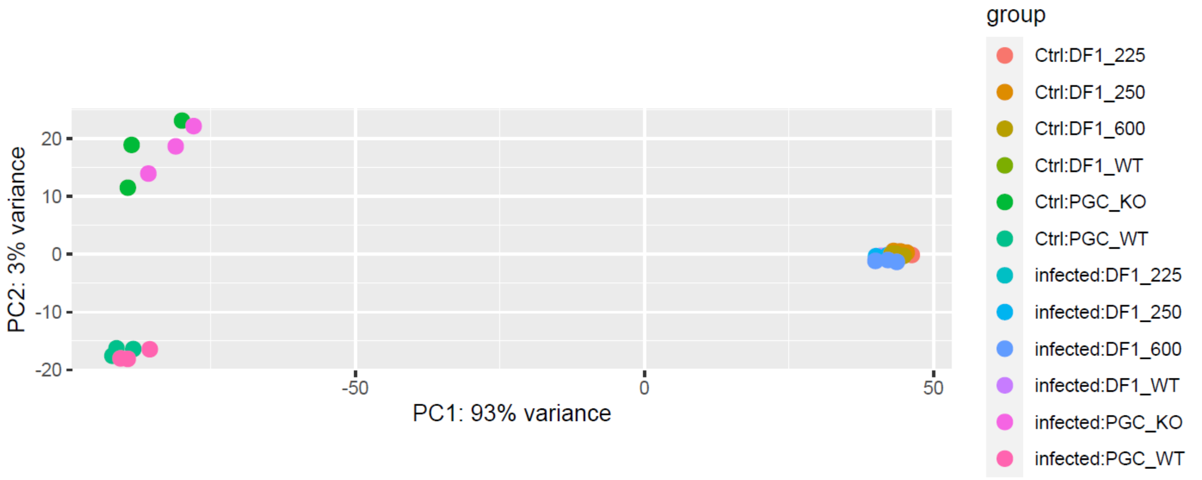


Figure 43 PCA shows extreme inter-group variability among DF-1 and PGC cell lines Vst transformed raw RNA-Seq counts from merged PGC and DF-1 datasets were used in the PCA performed with the DESeq2 package in R. PC1 and PC2 are plotted on the x and y axis of the bi-plot, respectively. Data points represent sample groups (each of which in biological triplicates).

Great intergroup variability influences the dispersion estimation negatively, resulting in suboptimal estimation thus decreasing statistical power. Retaining all the samples in the same analysis would drown out any other signal we may wish to detect. Hence, the data set was split accordingly between the two cell types which were subsequently processed separately.

4.3.1.5 *KO DF-1 data set has a potential outlier*

Raw RNA-Seq reads from the KO DF-1 data subset were transformed using *vst* and *rlog* transformations to investigate which renders the data more heteroskedastic (**Figure 106** in the Appendix). *Vst* proved to stabilise the variance better so those values were used to conduct the PCA (**Figure 44**).

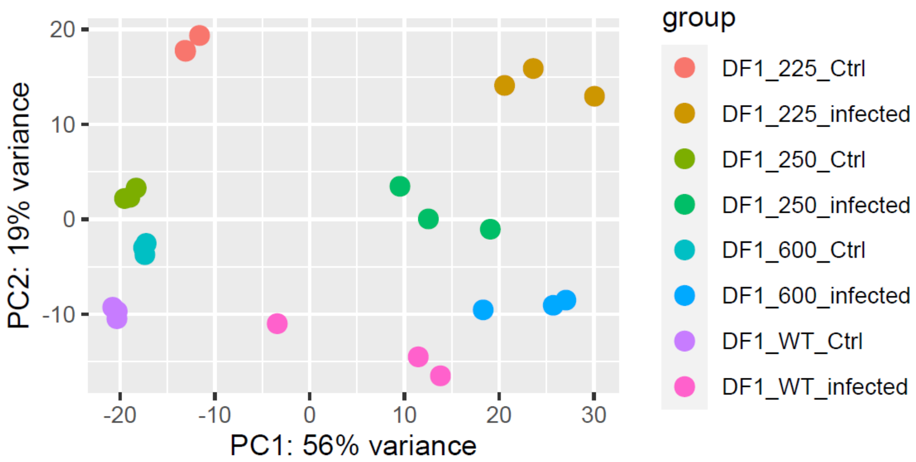


Figure 44 PCA analysis uncovers an outlier in the KO DF-1 data set Vst transformed raw counts were used to conduct PCA in DESeq2 package in R. PC1 and PC2 are plotted on the X and Y axes, respectively. Datapoints are representing sample groups consisting of three biological replicates.

PC1 explained 56% of the variance; it was consistent with the condition, mainly separating the infected from the uninfected samples. One replicate of the infected DF-1 parental cell line showed greater similarity in its transcriptional profile with the uninfected population. To investigate the extent of the dissimilarity a sample-distance matrix was constructed (**Figure 45**). The infected replicate of WT DF-1 cell line segregated from the infected samples very significantly. To determine if it should be considered an outlier and discarded from the analysis, Cooks distance was calculated for all samples. Results (**Figure 46**) showed all samples have equal leverage. The source of variation was not technical, but rather biological, thus all samples were taken forward for differential expression analysis.

PC2 shows a separation of KO DF-1 clonal cell lines 225 and 250 the parental cell line and the clone 600, whose dissimilarity accounted for 19% of variation in the data set. To see how the remainder of the variation is distributed among the principal components, an elbow plot was constructed in PCAtools package in R (results not shown). Results indicated 4 PCs explain almost all the variation. Hence, we looked into PC3 and PC4.

To characterise the contribution of individual genes across samples to the overall variation of the data 15% of the top of the loadings range for relevant PCs: PC1, PC2, PC3 and PC4 were plotted in **Figure 47**. Highest portion of the loadings range was occupied by a greater number of genes in PC1, namely: *ERBB3, DHH, PTX3, LIPG, SOCS1, SOCS3, PIM1, LOC422150, PPP1R3B, ID1, RND3, SIK1, ADM, LOC425795, FDFT1, CCNE2, DUSP7, CYR61, CITED4, KLHL25, ZNF469, RIN3, LRRC32, FBXO5, ENSGALG00000017071, AMOTL2, TFAP2C, HES1, ENSGALG00000017376, MAT2L2.* Top 15% of the loadings range for PC2 was occupied by *WIF1, EFEMP1, TNFAIP8L3, CYP26B1, LOC770612* and *JAC1. Genes contained in PC3 were: LOC770612, C1QTNF3, SUCNR1, UNC5C, IL8L2, KIT, IL1R1 and PC4 retained LOC770612, GPRIN1, COMTD1, EMILIN2, TMEM132E, ENSGALG00000005180, EMID1, LAPTM5, CNTNAP1, PRKCB, CD1C, TNFAIP6.*

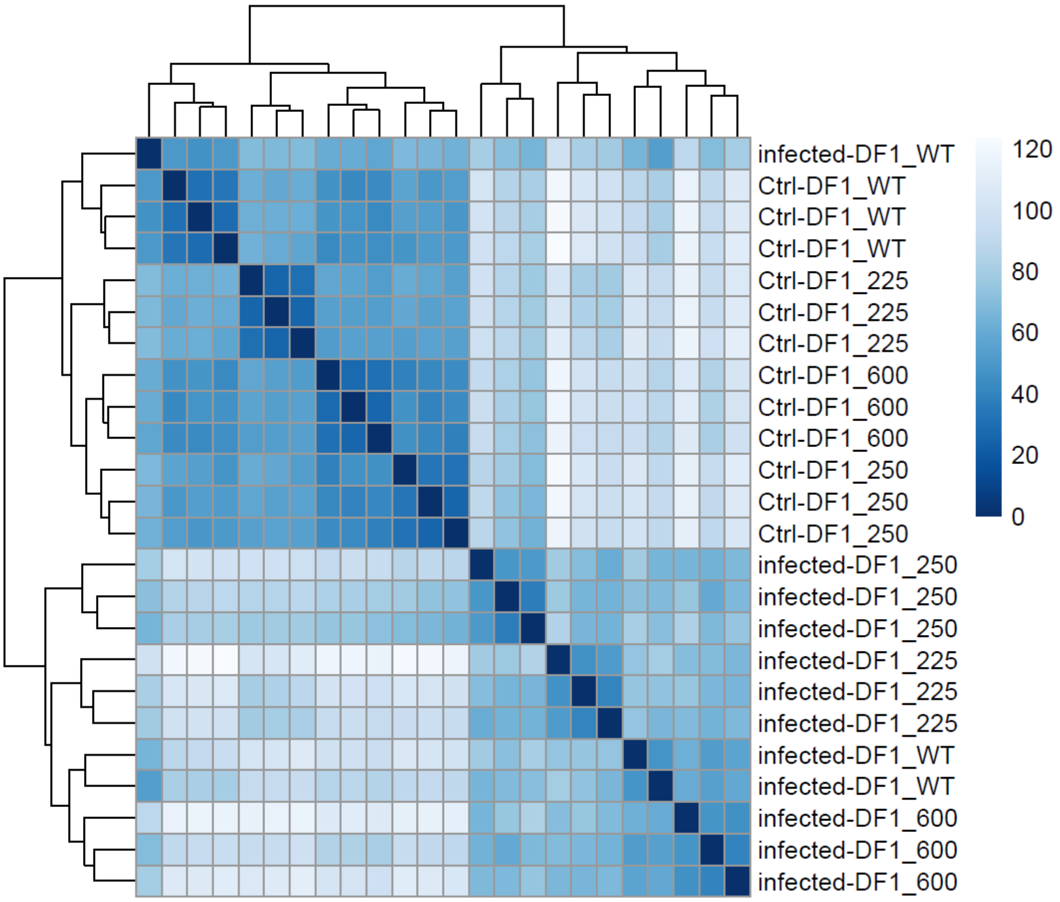


Figure 45 Infected DF-1 parental line contains an outlier. Sample distance matrix was constructed in DESeq2 in R from *vst* values of the KO DF-1 RNA-seq counts, using Euclidean distances as the distance metric. Legend gradient indicates similarity among samples from dark to light tone, indicating increasing dissimilarity.

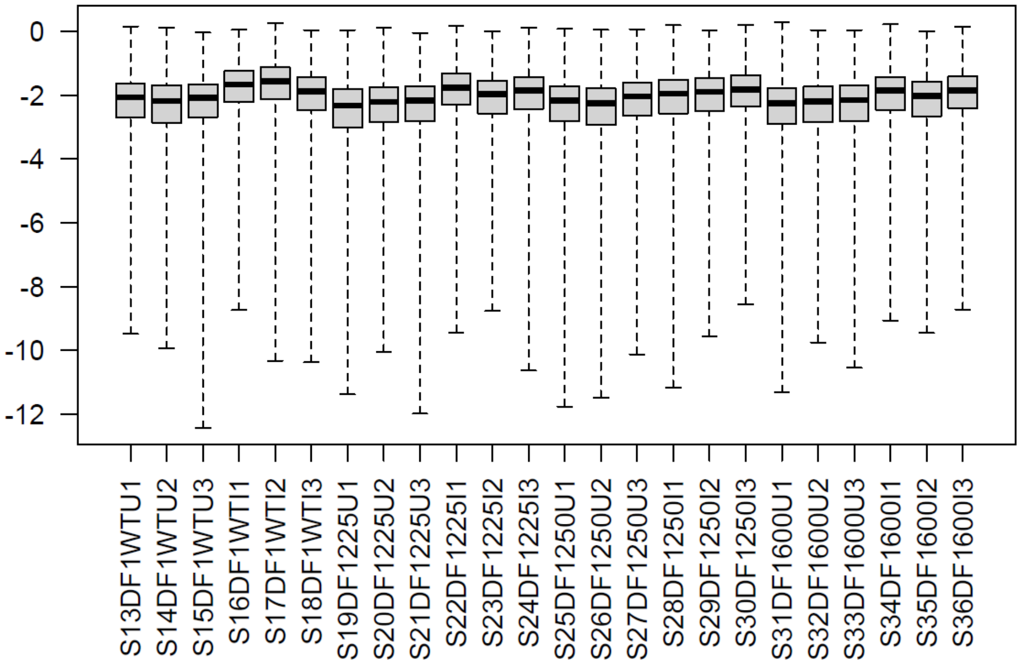


Figure 46 Cook's distances for KO DF-1 dataset

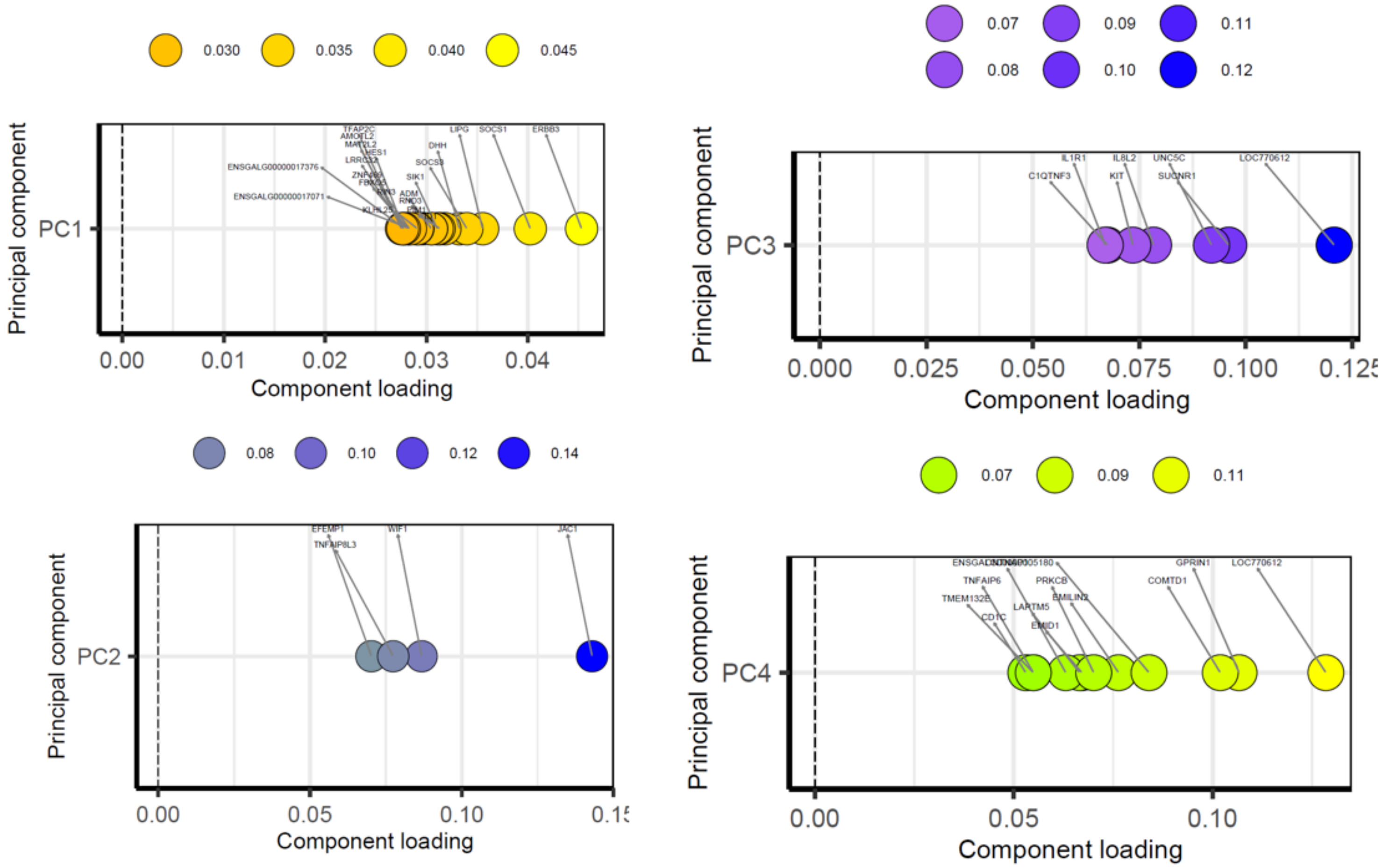


Figure 47 Loadings for main 4 PCs for KO DF-1 data set Component loadings from the PCA are plotted for the top 15% of the loadings range for each PC1 and PC2. Retained variables for PC1 are *ERBB3, DHH, PTX3, LIPG, SOCS1, SOCS3, PIM1, LOC422150, PPP1R3B, ID1, RND3, SIK1, ADM, LOC425795, FDFT1, CCNE2, DUSP7, CYR61, CITED4, KLHL25, ZNF469, RIN3, LRRC32, FBXO5, ENSGALG00000017071, AMOTL2, TFAP2C, HES1, ENSGALG00000017376, MAT2L2*. Retained variables for PC2 are *WIF1, EFEMP1, TNFAIP8L3, CYP26B1, LOC770612, and JAC1*. PC3 retained following variables: *LOC770612, C1QTNF3, SUCNR1, UNC5C, IL8L2, KIT, IL1R1* and PC4 retained *LOC770612, GPRIN1, COMTD1, EMILIN2, TMEM132E, ENSGALG00000005180, EMID1, LAPTM5, CNTNAP1, PRKCB, CD1C, TNFAIP6. Plotting was done in R using PCAtools package*

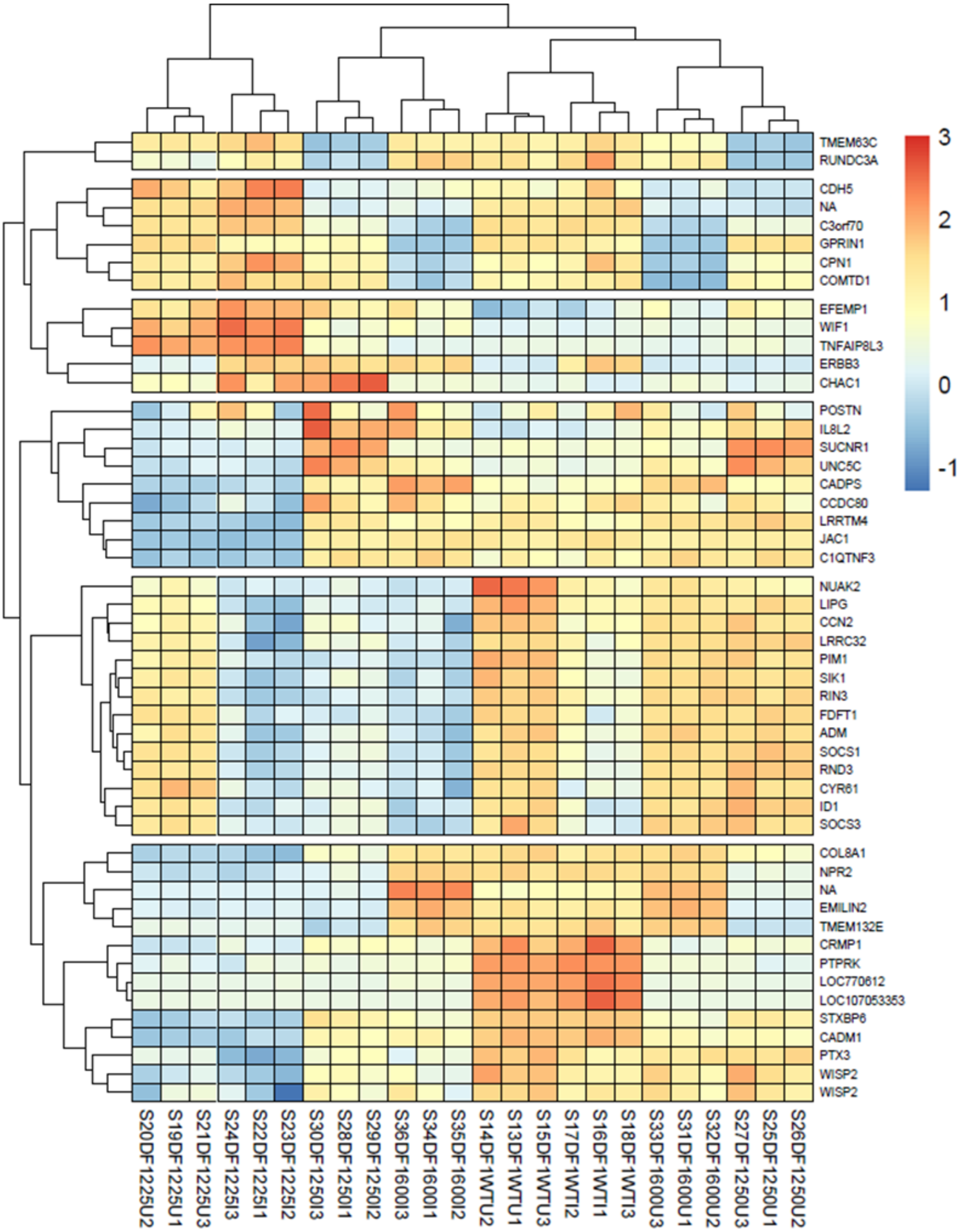


Figure 48 Top 100 most variable genes in KO DF-1 RNA-Seq dataset *Vst* normalised read counts from KO DF-1 data set were sorted by variance. TPM values of the 100 most variable genes across all samples were Z-score normalised and plotted against all samples after clustering using Euclidean distances. Heatmap was produced in the pheatmap package in R.

For a more granular view of dissimilarities of transcriptomic profiles among all sample groups, a heatmap was constructed to visualise the clustering (**Figure 48**). 100 most variable genes were selected by sorting all the *vst* normalised read counts by variance. TPM values of those genes were Z-score normalised to standardise the values across samples and clustered using Euclidean distances. Six gene clusters emerged as well as 4 sample type clusters which did not always match the cell lines. KO DF-1 225 clustered away from all the others in both conditions. Top-most cluster contains *TMEM63C* and *RUNDC3A* genes which are expressed at a lower level in KO DF-1 225 than in all other cell lines. Second cluster contained *CDH5, C3orf70, GPRIN1, CPN1, COMTD1*; a set of genes that were less expressed in KO DF-1 600 than the average and slightly more highly expressed in KO DF-1 225. *EFEMP1, WIF1, TNFAIP8L3, ERBB3* and *CHAC1* also clustered together and were characteristic of KO DF-1 225. *ERBB3* had higher levels of expression in infected condition in all cell lines. *POSTN, IL8L2, SUCNR1, UNC5C, CADPS, CCDC80, LRRTM4, JAC1* and *C1QTNF3* clustered together due to their lower expression in both conditions in KO DF-1 225. Genes in Cluster 5 (*NUAK2, LIPG, CCN2, LRRC32, PIM1, SIK1, RIN3, FDFT1, ADM, SOCS1, RND3, CYR61, ID1* and *SOCS3)* were exhibiting lower levels expression in infected conditions in all cell lines except the WT parental DF-1 cell line. Lower expression levels of *COL8A1, NPR2, EMILIN2, TMEM132E, STXBP6, CADM1, PTX3, WISP2 and WISP2* in KO DF-1 225 were another feature that separated that cell line from all others. On the other hand*, LOC107053353, CRMP1, PTPRK* and *LOC770612* showed higher expression only in WT DF-1 parental line.

4.3.2 WT and *chIFITM* KO DF1 cells respond differently to influenza infection

4.3.2.1 Early immune response in WT DF-1 to influenza infection

Differential expression WT DF1 infected vs non-infected

To further characterise the WT specific response to infection, genes that were significantly differentially expressed in WT DF-1 at 6h post infection were investigated via clustering. Differential expression was conducted in DESeq2 in R with the FDR cut-off <0.05. Data was modelled with a means model without intercept and all the factors were rearranged into unique groups. Desired contrast, DF-1_WT_infected vs DF-1_WT_Ctrl, was supplied to the *ashr* function for a more accurate estimation of effect size. Out of 18143 genes with nonzero total read count and the adjusted p-value < 0.05, 3904 (22%) genes were upregulated, and 3371 (19%) genes were downregulated. 2111 (12%) genes were not included in the analysis due to their low read count. No outliers were detected. Resulting list was for filtered by shrunken fold changes for 100 most upregulated and 100 most downregulated genes. Their TPM values were normalised using Z-score method and clustered using Euclidean distances.

The heatmap of hundred most upregulated genes (**Figure 49**) shows the Z-scores are heavily influenced by a single value in a row. No strong clustering emerged using this method beyond separation of infected and uninfected samples. Most upregulated genes show that the differential expression is dominated by low count genes, which when normalised do not show great differences between groups due to low absolute numbers (**Figure 108** in the Appendix). When looking at the low count genes, the precision of measurement becomes paramount. DESeq2 method of normalisation accounts for that by taking the RNA composition into the equation. To ensure the normalisation employed in DESeq2 addressed the outlier values successfully for appropriate differential expression, normalised values were clustered using “manhattan” distances (**Figure 50**) because they are more resistant to the influence of outliers [163]. Some outlier values remain present after normalisation, but overall contrast in expression levels between the infected and the uninfected group is clearer. However, no further clustering within groups was observed.

On the contrary downregulated genes reveal individual groupings (**Figure 51**) within two major clusters of infected and uninfected samples. First cluster contains *BLEC2, AFMID, NDUFA7, TIMM17B, SNRPE, RACK1, ENSGALG00000000044, RPL9, RPL36, RAB44, RPL22, TUBB3, CDH1, LAMTOR5 and RAB11B.*

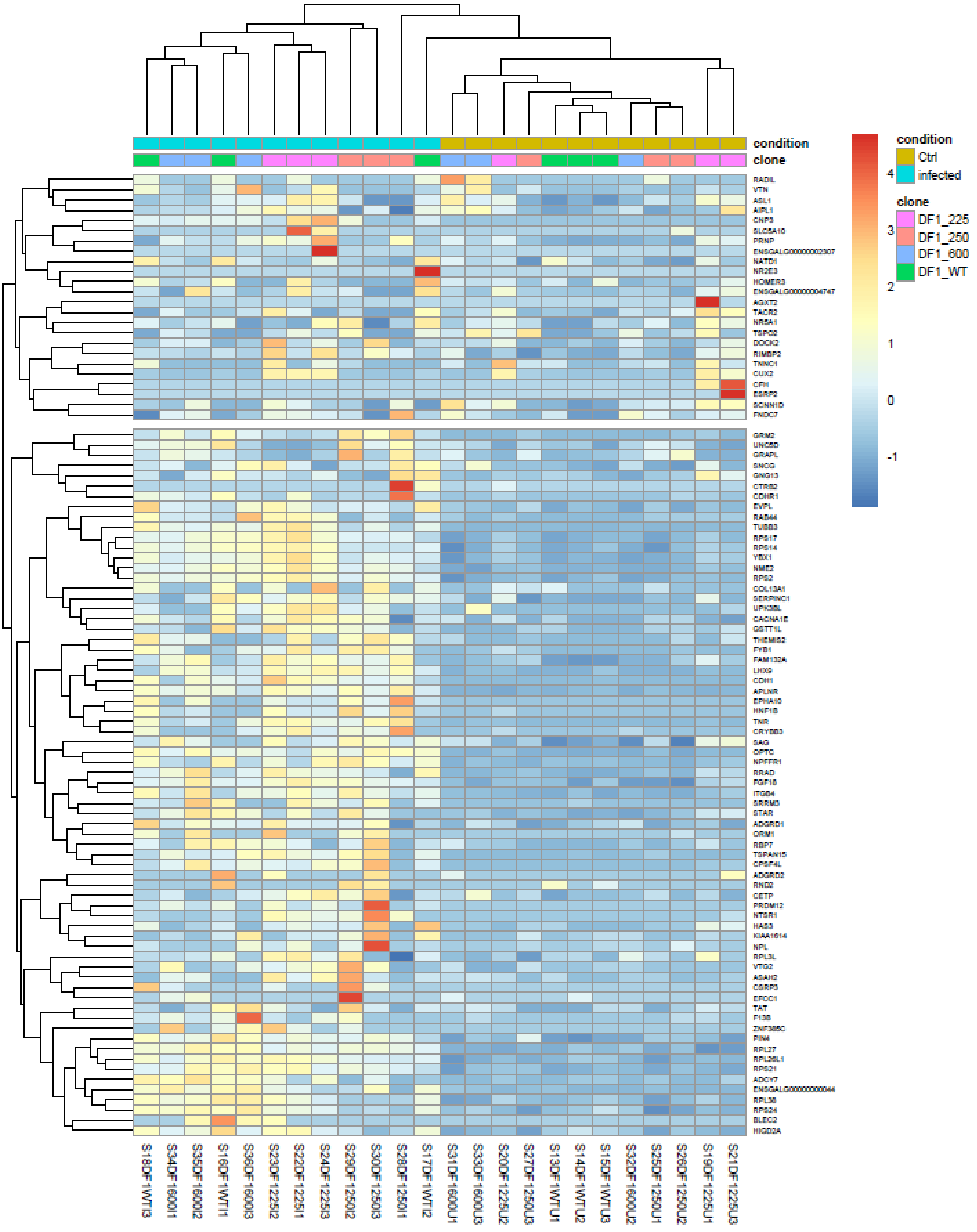


Figure 49 Heatmap of 100 most upregulated differentially expressed genes in WT DF-1 cell lines in influenza infection
Differential expression was calculated for pairwise comparison infected vs uninfected. FDR <0.05 was considered significant. TPM values were normalised using the Z-score method and clustered by Euclidean distances in pheatmap package in R.

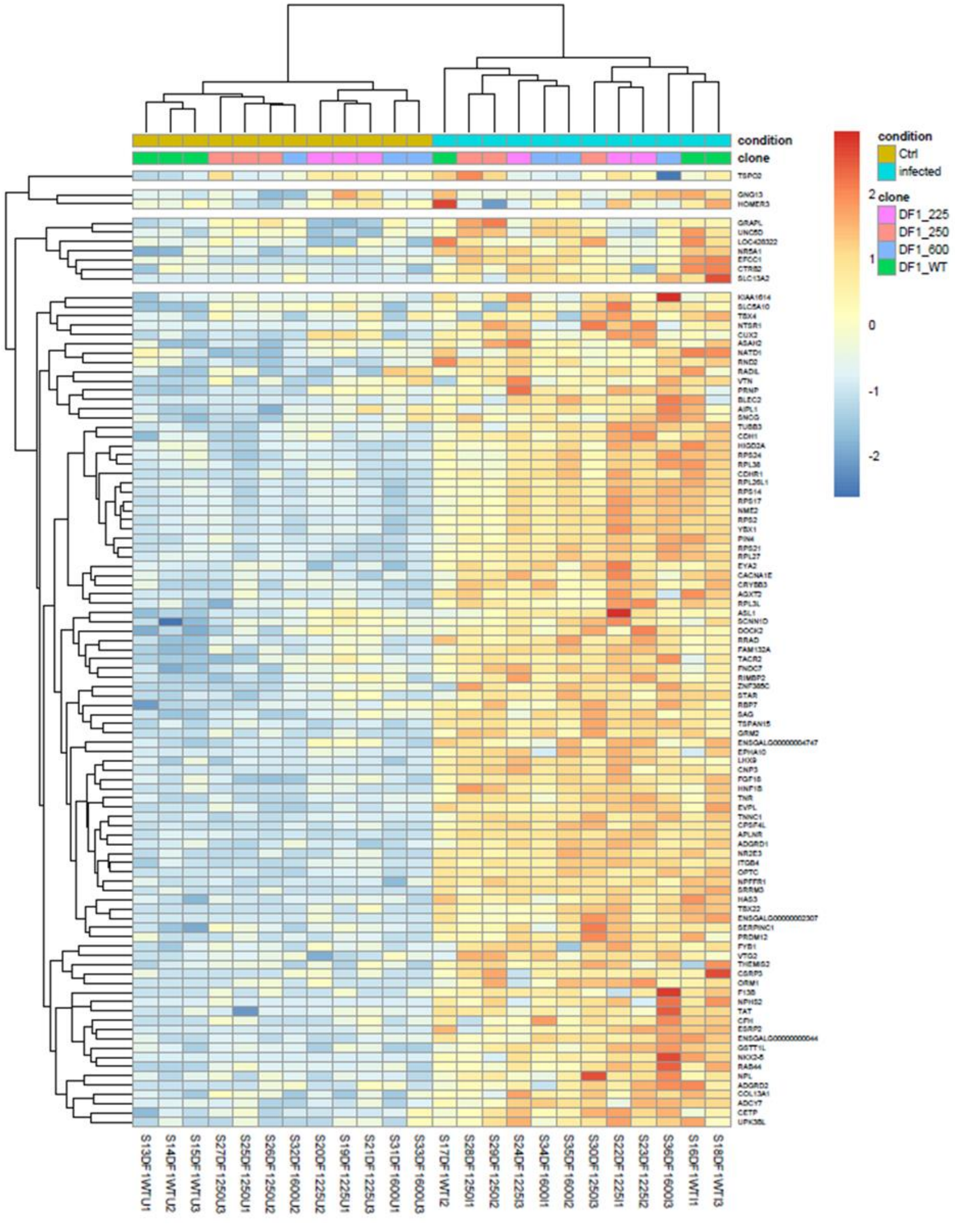


Figure 50 Heatmap of DESeq2 normalised raw RNA-Seq counts for WT DF-1 pairwise comparison of infected vs uninfected Differential expression was calculated for contrast infected vs uninfected WT DF-1. For 100 top differentially expressed genes, the raw RNA-Seq counts were normalised using median of ratios, Z-score centred and plotted using "manhattan" distances in pheatmap package in R. Cells are coloured on a scale of colours from cold (blue) to hot (red) indicating increasing values.

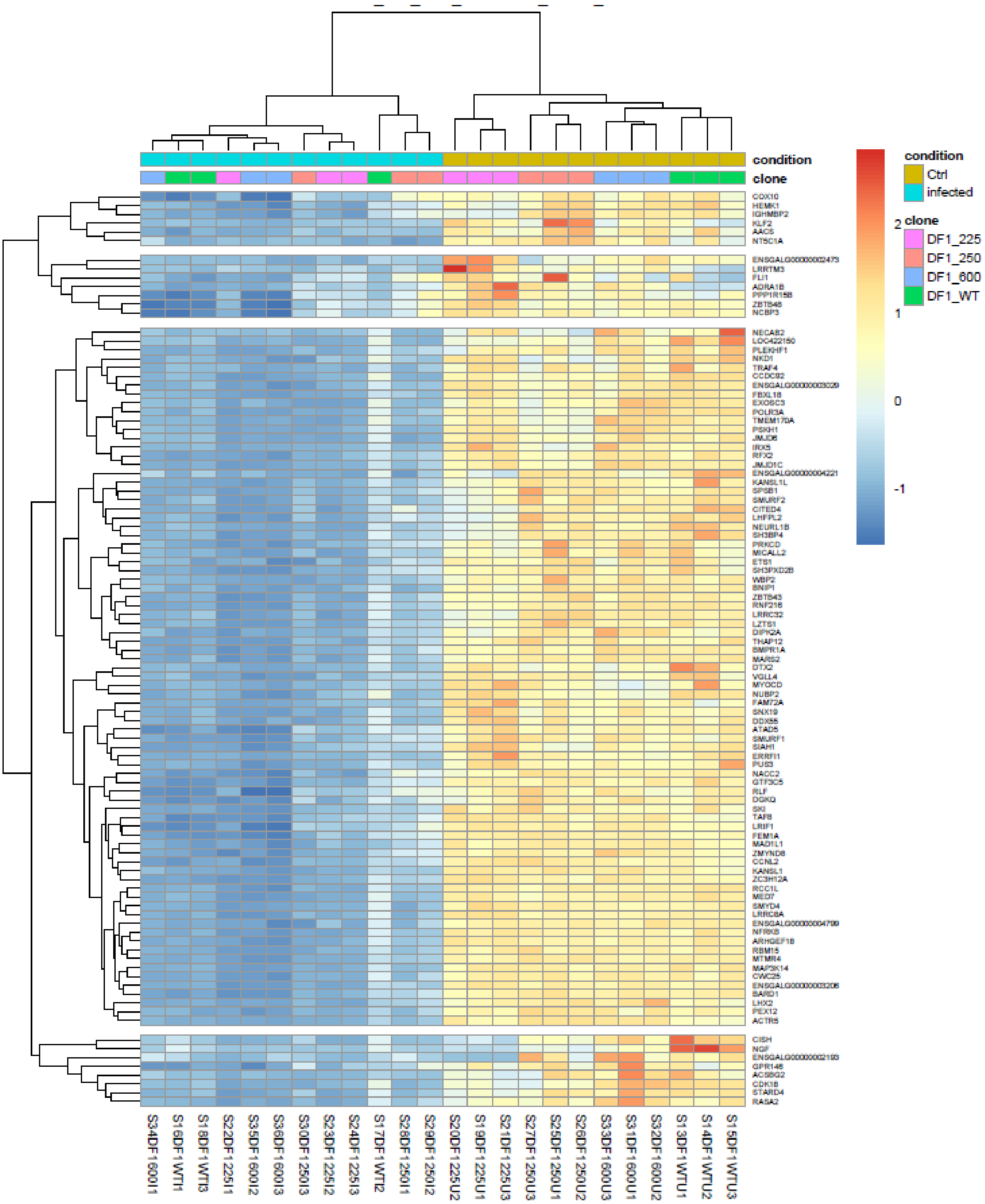


Figure 51 Most down regulated genes in pairwise comparison between infected and uninfected parental DF-1 cell line
DESeq2 was used to calculate differential expression for contrast infected vs uninfected WT DF-1. 100 most downregulated genes that were statistically significant at FDR <0.05 were selected. TPM values for those genes were Z-scored and clustered using Euclidean distances in pheatmap package in R.

Pathway analysis

GSEA

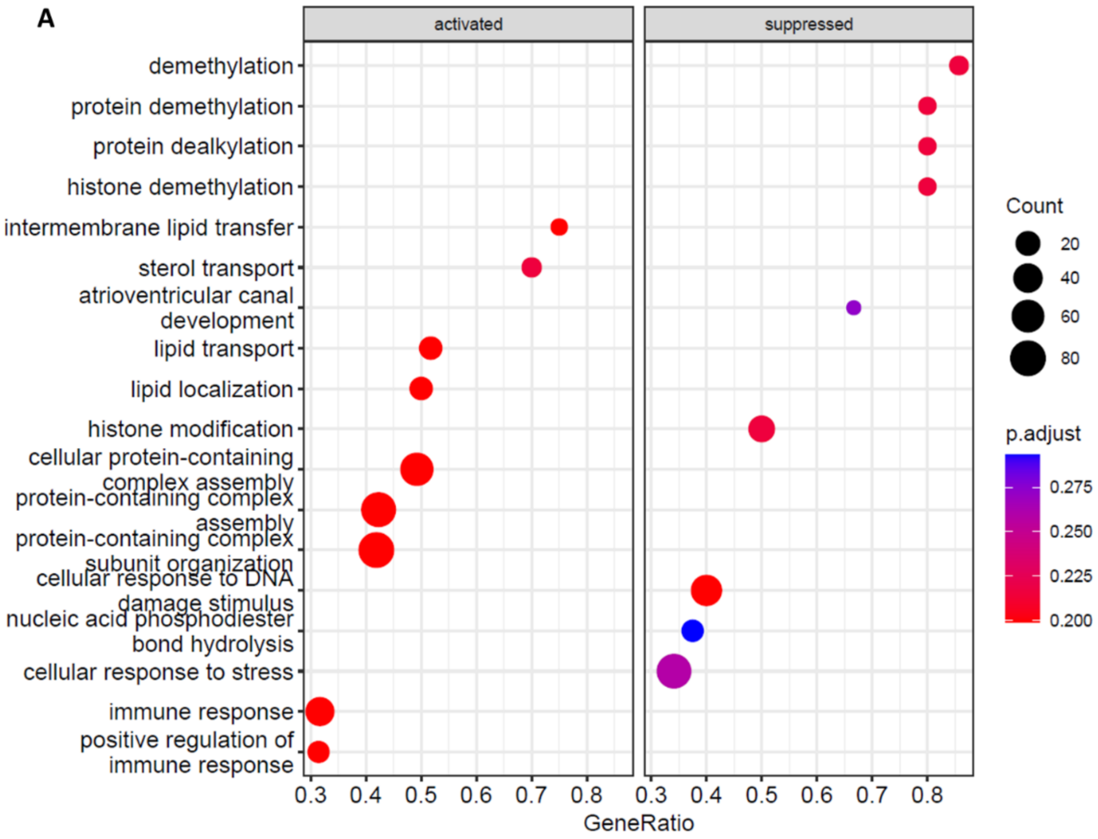
The high MOI used in this infection experiment represents a great source of stress to the cell. Activated defence systems affect many processes in the cell which are subsequently dysregulated. As in the previous chapter, differential expression resulted in a long list of differentially expressed genes. They were filtered for FDR <0.05 and the final list was used for Gene Set Enrichment Analysis (GSEA) in *clusterProfiler* against category of Biological Processes (BP) in Gene Ontology (GO) database. The analysis was run with following parameters: ont ="BP", keyType = "ENSEMBL", minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.5, verbose = TRUE, OrgDb = org.Gg.eg.db, pAdjustMethod = "BH", eps = 0.

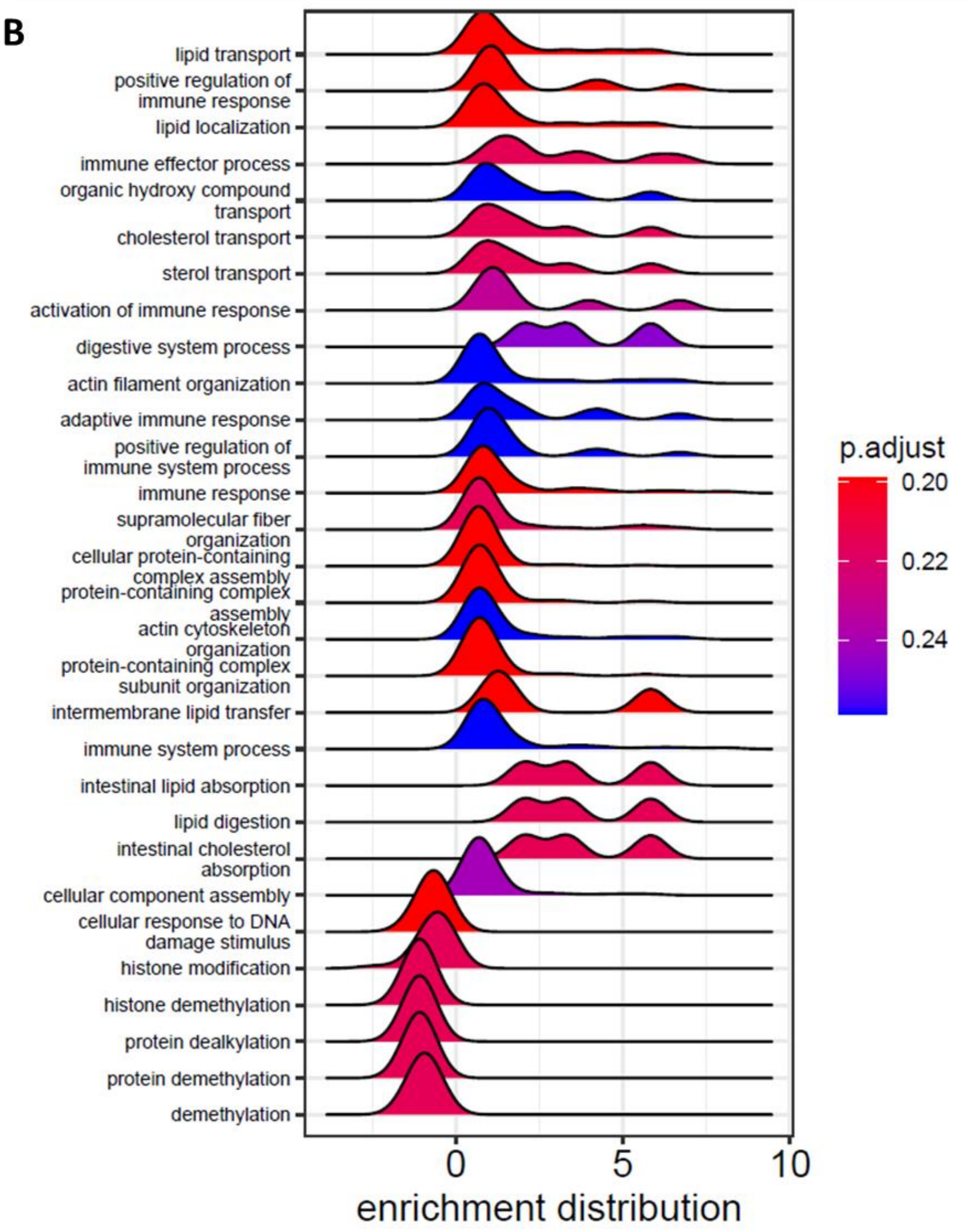
Results show significant (FDR <0.5) downregulation of processes concerned with protein processing, as indicated by the negative NES score in the ridgeplot (**Figure 52 B**). Processes involved in immune response and lipid metabolism were highly upregulated. Emapplot shows network results of how these processes are coordinated through term “immune effector process”. Activation of other separate processes was detected as well: many protein processing pathways and cellular component organisation. To investigate if significant representation of the same genes in various gene sets results in detection of many pathways an UpSet plot was generated to plot the fold changes of most statistically significant differentially regulated pathways. Note that all genes in gene sets plotted are regulated only in one direction. Genes detected as belonging to cellular response to DNA damage and histone modification are all uniquely detected only in those categories as well as singularly downregulated. Genes with the highest fold changes were identified in gene sets belonging to immune response. Immune response is also a biological theme with the highest number of detected associated gene sets. These gene sets were also identified as participating in other processes albeit through an extremely small number of genes/proteins, potentially a signalling molecule.

PADOG

To investigate the same question with a different statistical approach, PADOG analysis was conducted. WT DF-1 parental cell lines was compared in infected and mock condition. RNA-Seq raw reads were uploaded to ReactomeGSA and differential expression was conducted on the pathway level using *limma* algorithm in the background. FDR <0.05 was considered statistically significant. Main upregulated biological themes that were statistically significant were Cytokine signalling within the term Immune System and Signal transduction (**Figure 53**). Interestingly no major themes were downregulated, just more narrow terms such as: RAF-independent MAPK1/3 activation (R-HSA-112409), Golgi Cisternae Pericentriolar Stack Reorganization (R-HSA-162658), Frs2-mediated activation (R-HSA-170968), p75NTR negatively regulates cell cycle via SC1 (R-HSA-193670), Signaling by Hippo (R-HSA-2028269), NFG and proNGF binds to p75NTR (R-HSA-205017), Axonal growth stimulation (R-HSA-209563), AKT-mediated inactivation of FOXO1A (R-HSA-211163), Signaling by Leptin (R-HSA-2586552), Scavenging by Class F Receptors (R-HSA-3000484), SUMOylation of transcription factors (R-HSA-3232118), TET1,2,3 and TDG demethylate DNA (R-HSA-5221030), TRAF3 deficiency - HSE (R-HSA-5602571), Defective SLC2A1 causes GLUT1 deficiency syndrome 1 (GLUT1DS1) (R-HSA-5619043), Defective SLC24A4 causes hypomineralized amelogenesis imperfecta (AI) (R-HSA-5619055), Negative regulation of MAPK pathway (R-HSA-5675221), Phosphorylation of proteins involved in G1/S transition by active Cyclin E:Cdk2 complexes (R-HSA-69200), Death Receptor Signalling (R-HSA-73887), PTK6 Activates STAT3 (R-HSA-8849474)

Upregulated terms were: Cytokine Signaling in Immune system (R-HSA-1280215), Extrinsic Pathway of Fibrin Clot Formation (R-HSA-140834), Translocation of SLC2A4 (GLUT4) to the plasma membrane (R-HSA-1445148), Signal Transduction (R-HSA-162582), Triglyceride catabolism (R-HSA-163560), NRAGE signals death through JNK (R-HSA-193648), p75 NTR receptor-mediated signalling (R-HSA-193704), Signalling to STAT3 (R-HSA-198745), G-protein activation (R-HSA-202040), Role of phospholipids in phagocytosis (R-HSA-2029485), Cell death signalling via NRAGE, NRIF and NADE (R-HSA-204998), POU5F1 (OCT4), SOX2, NANOG activate genes related to proliferation (R-HSA-2892247), Scavenging by Class B Receptors (R-HSA-3000471), Glycogen storage diseases (R-HSA-3229121), Detoxification of Reactive Oxygen Species (R-HSA-3299685), Glycogen synthesis (R-HSA-3322077), Tachykinin receptors bind tachykinins (R-HSA-380095), Glycogen storage disease type 0 (liver GYS2) (R-HSA-3858516), Glycogen storage disease type IV (GBE1) (R-HSA-3878781) and G alpha (12/13) signalling events (R-HSA-416482).


A
activated
suppressed
demethylation
protein demethylation
protein dealkylation
histone demethylation
intermembrane lipid transfer
sterol transport
atrioventricular canal development
lipid transport
lipid localization
histone modification
cellular protein-containing complex assembly
protein-containing complex assembly
protein-containing complex subunit organization
cellular response to DNA damage stimulus
nucleic acid phosphodiester bond hydrolysis
cellular response to stress
immune response
positive regulation of immune response
0.3
0.4
0.5
0.6
0.7
0.8
GeneRatio
Count
20
40
60
80
p.adjust
0.275
0.250
0.225
0.200

B
lipid transport
positive regulation of immune response
lipid localization
immune effector process
organic hydroxy compound transport
cholesterol transport
sterol transport
activation of immune response
digestive system process
actin filament organization
adaptive immune response
positive regulation of immune system process
immune response
supramolecular fiber organization
cellular protein-containing complex assembly
protein-containing complex assembly
actin cytoskeleton organization
protein-containing complex subunit organization
intermembrane lipid transfer
immune system process
intestinal lipid absorption
lipid digestion
intestinal cholesterol absorption
cellular component assembly
cellular response to DNA damage stimulus
histone modification
histone demethylation
protein dealkylation
protein demethylation
demethylation
0
5
10
enrichment distribution
p.adjust
0.20
0.22
0.24

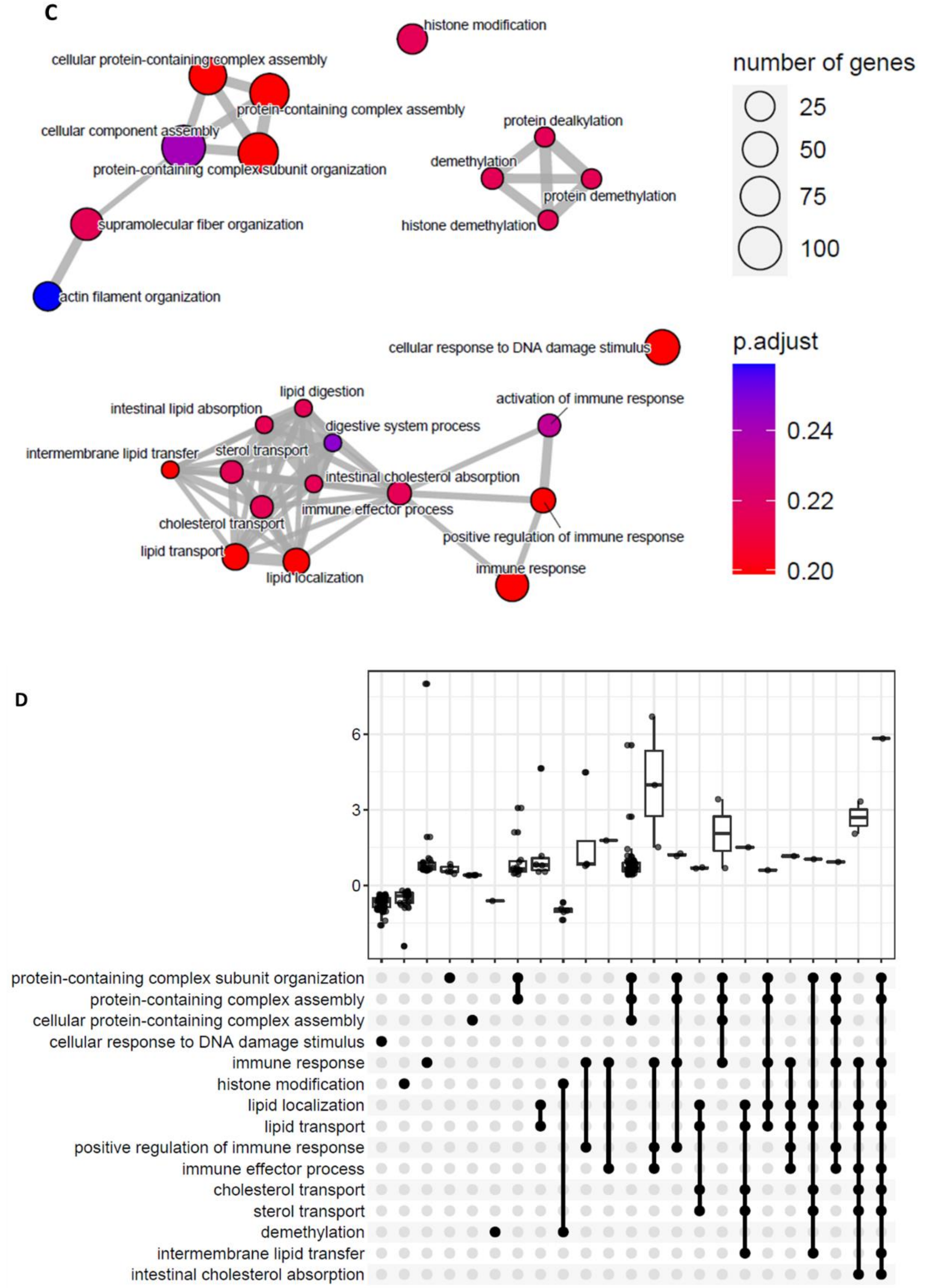


Figure 52 GSEA for infected vs uninfected WT DF-1 cell line Differential expression analysis was conduced for comparison infected vs non-infected WT DF-1 cell line. Significant (FDR >0.05) differentially expressed genes were supplied to the ashr function and their fold changes re-estimated. Resulting list was analysed in clucterProfiler using fgsea algorithm against biological process category in the GO database with chicken annotation. A) Dotplot – most significant categories are split in activated and supressed and plotted against their gene ratios. Size of the circle represents the number of genes identified within a category and the colour the statistical significance. B) Ridgeplot – most significant detected categories are plotted against their NES score (negative values indicate downregulation and positive values upregulation). Curves represent distribution of foldchanges of genes detected in the gene set. C) Emapplot – circles represent the size of the gene sets detected and the colour the FDR. D) UpsetPlot – biological categories are plotted against the fold changes of genes found in gene set categorising the biological process.

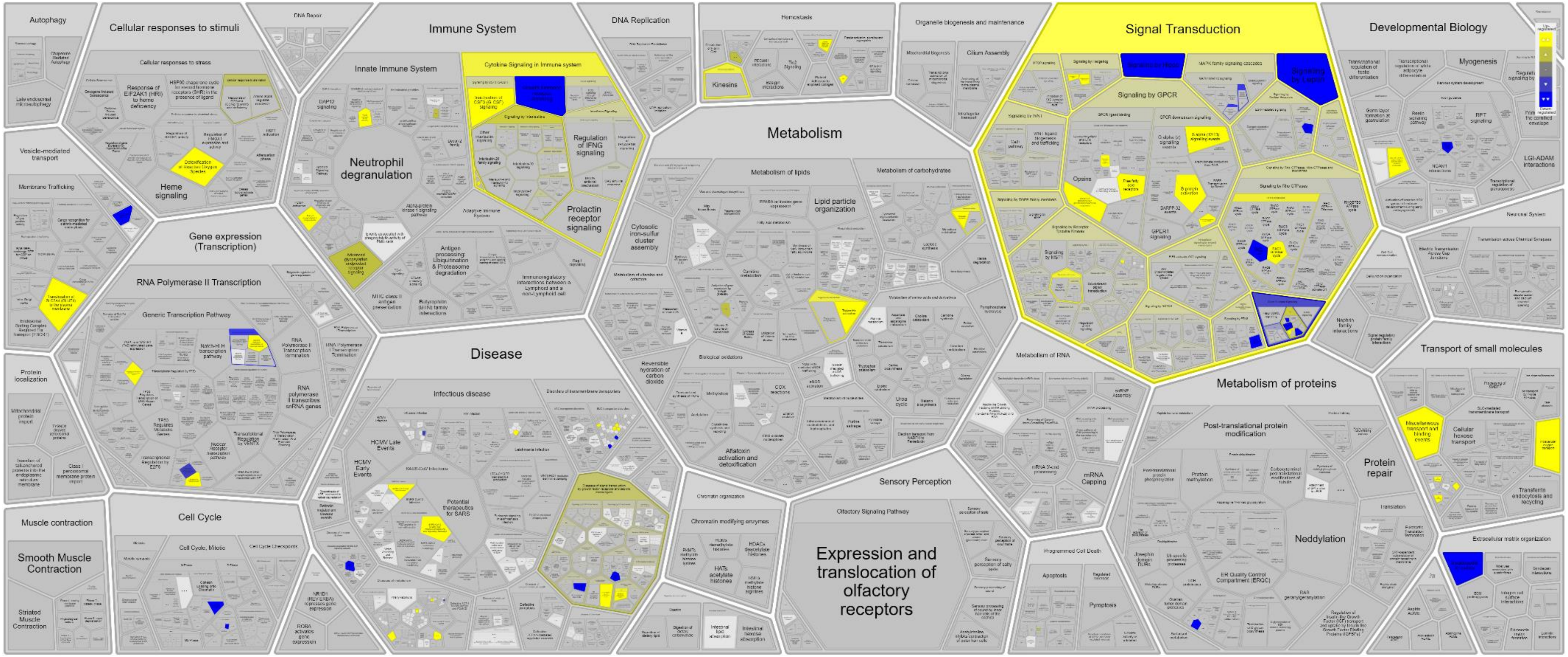


Figure 53 PADOG analysis for differentially expressed pathways in WT DF-1 parental cell line at 6h post infection PADOG analysis using Reactome data base was conducted with raw RNA-Seq data. FDR<0.05 was considered statistically significant. Chicken EnsemblIDs were translated into human and all the results were obtained based on human pathway annotations.

4.3.2.2 Early innate immune response in chIFITM KO DF-1 cell lines

Experimental set-up

To characterise the immune response in the absence of chIFITM locus, all three chIFITM KO DF-1 clones, designated with unique numbers 225, 250 and 600, were infected with H5N3 subtype of influenza virus following the protocol outlined in 2.2.2.4. The samples were processed in the same way to maintain consistency and ensure comparability among experiments and analyses.

Differential expression DF1 KO infected vs non-infected

Comparison of gene expression levels was conducted in DESeq2 using a Wald test at a significance level of 0.05. The experiment was modelled with formula *~0+group* where the multi factorial design (condition and genotype) was reformulated into unique sample groups in triplicates. To obtain the average response of the three clonal cell lines, care was taken to split the contribution of every coefficient to one third. “Contrast” approach was used to obtain the results hence the effect sizes were appropriately shrunken with *ashr* function. Differentially expressed genes were sorted by logFoldChange and most strongly upregulated and downregulated genes were clustered. To standardise the values across samples and experiments TPM values were normalised using the Z-score method and clustered to be visualised with the heatmap package in R. This allows inspection of expression levels in all samples, including ones for which comparison was not conducted.

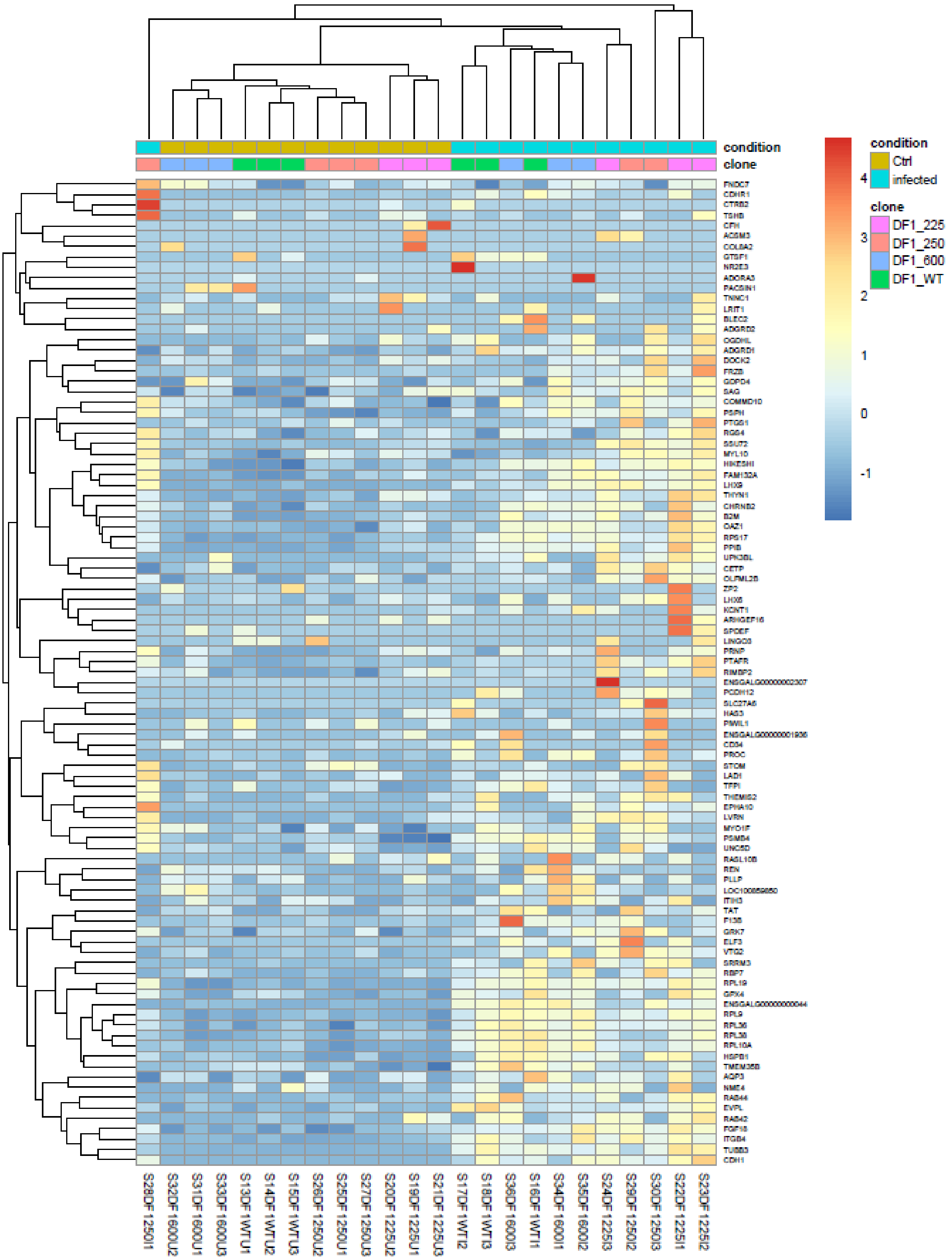


Figure 54 Most upregulated genes in chIFITM KO DF-1 clonal cell lines at 6h post infection with H5N3 subtype of influenza
TPM values for most upregulated genes that were statistically significant were normalised using Z-score and clustered using Euclidean distances with pheatmap package in R.

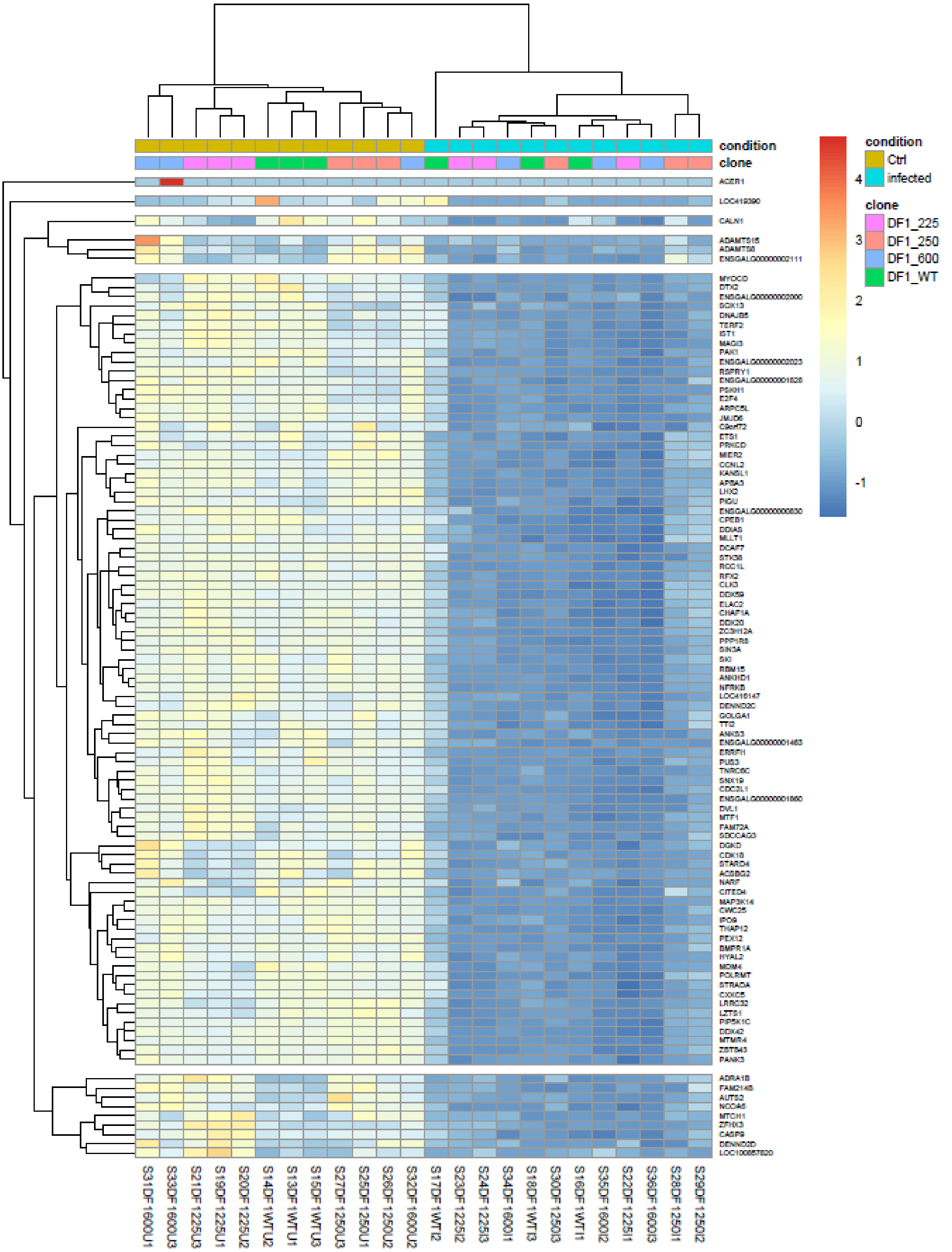


Figure 55 Downregulated genes in comparison infected vs uninfected KO DF-1 clonal cell lines TPM values were Z-scored for most downregulated genes that had FDR < 0.05 in comparison infected vs uninfected chIFITM KO DF-1 cell lines. Clustering was conducted and values plotted with pheatmap package in R.

Upregulated genes show a high degree of variability among samples (**Figure 54**). Even though they roughly cluster according to the condition, with the exception of sample 28 – KO DF-1 250 infected replicate 1, the levels of observation are varied across both conditions for all samples. This fact also impedes tighter clustering of the replicates. On occasion, normalisation and accurate observation of differential expression levels are further hindered by an outlier measurement that negatively influences the sample average and the standard deviation (see genes: *FNDC7, CDHR1, CTRB2, TSHB, CFH, ACSM3, COL8A2, GTSF1, NR2E3* and *ADORA3*). No patterns in expression that could indicate co-regulatory mechanisms were detected.

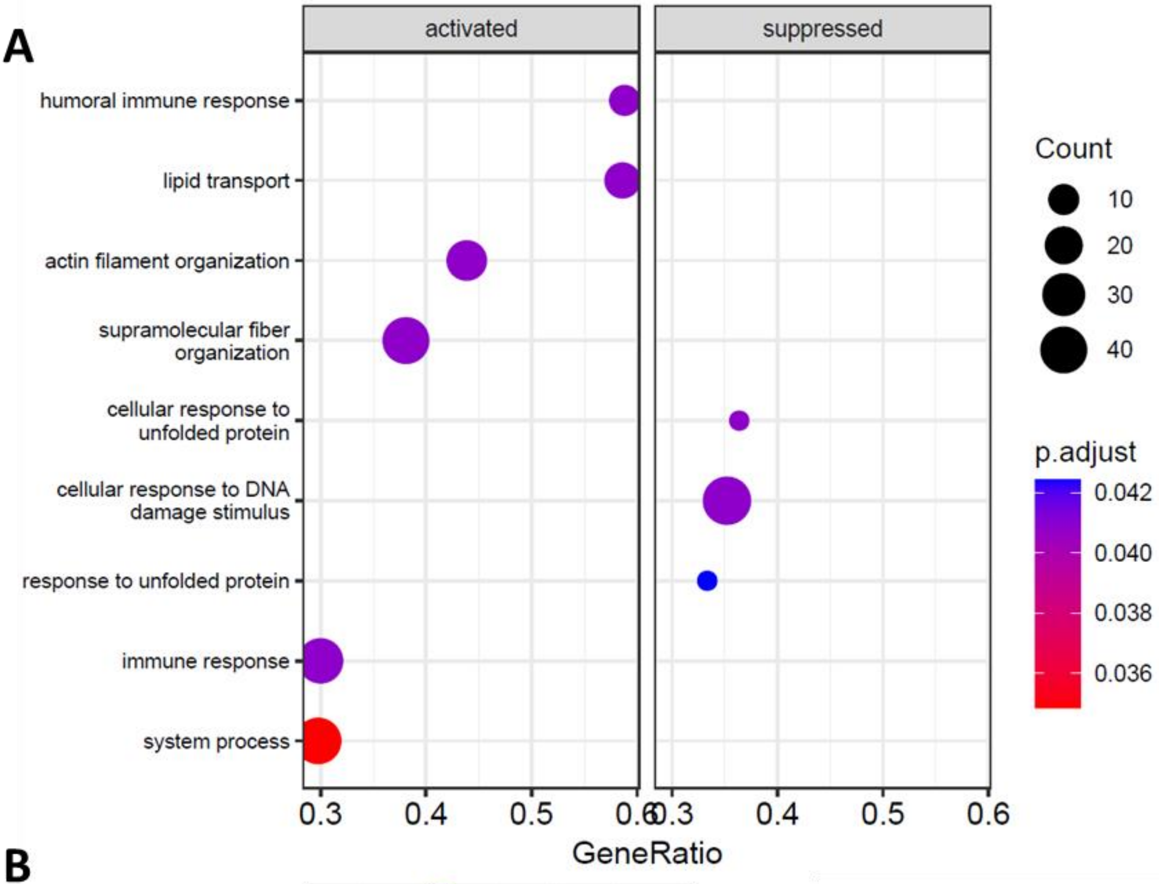
Downregulated genes are also primarily clustered into infected and uninfected groups (**Figure 55**). Similarly to the upregulated genes, clustering of the triplicates was more consistent in the control condition compared to the treatment. Overall, all the cell lines exhibit similar transcriptional profile. Some more finer clusters can be observed, e.g. *ADAMTS15, ADAMTS8* and *ENSGALG0000000211* or *ADRA1B, FAM214B, AUTS2, NCOA6 and MTCH1* – which separate the parental DF-1 cell line from the clones. However, this clustering is observed in the control condition, and it is unclear if it has any influence on the response to the infection.

Functional analysis

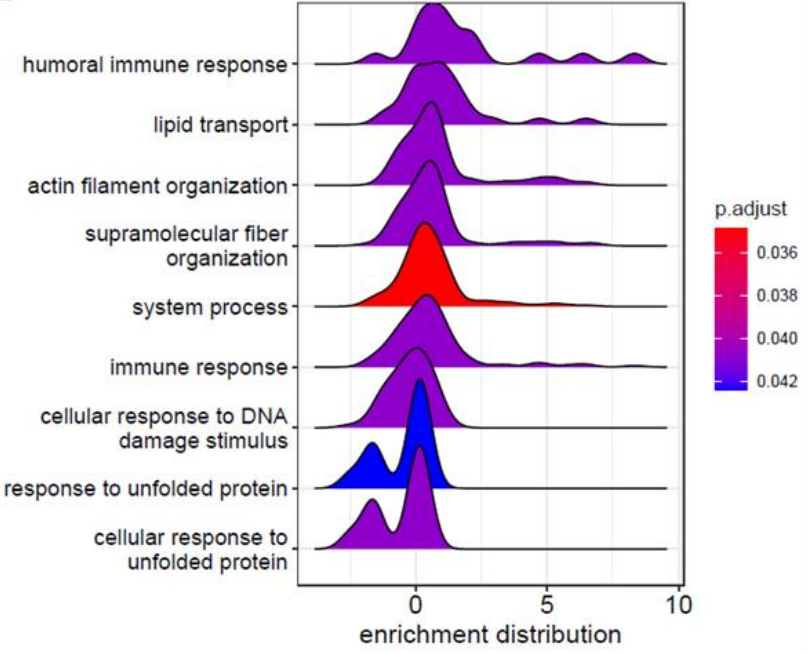
Finally, the Pathway Analysis was performed to elucidate processes that take place during an infection which are differentially regulated in the absence of the *chIFITM* locus in the DF-1 avian fibroblasts. Gene Set Enrichment Analysis (GSEA) was conducted using the ranked list of differentially expressed genes among all clones pooled together. As previously, clusterProfiler was used with the following parameters: minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.05, pAdjustMethod = "BH", eps = 0. Category Biological Process in the Gene Ontology database was interrogated to test for enrichment.

A relatively low number of pathways was detected as enriched, possibly because of a strict p-value cutoff and a conservative approach to multiple testing correction (Benjamini-Hochberg method). Enriched pathways at p-value <0.05 were immune response, lipid transport, actin filament organisation, cellular response to DNA damage and unfolded protein (**Figure 56**). In noisier systems such as clinical samples, it is justifiable to relax the thresholds for the purpose of generating result that can be narrowed down later in the pipeline. Given that systematic analyses such as these are generally used for the hypothesis generation, weaker statistics are usually acceptable with the goal of capturing as much information as possible. In later chapter concerning PGC-derived fibroblasts we encounter such a situation and take a more relaxed approach and interpret the results with appropriate caveats accordingly..

A


activated
suppressed
humoral immune response
lipid transport
actin filament organization
supramolecular fiber organization
cellular response to unfolded protein
cellular response to DNA damage stimulus
response to unfolded protein
immune response
system process
0.3
0.4
0.5
0.6
0.3
0.4
0.5
0.6
GeneRatio
Count
10
20
30
40
p.adjust
0.042
0.040
0.038
0.036

B


humoral immune response
lipid transport
actin filament organization
supramolecular fiber organization
system process
immune response
cellular response to DNA damage stimulus
response to unfolded protein
cellular response to unfolded protein
0
5
10
enrichment distribution
p.adjust
0.036
0.038
0.040
0.042

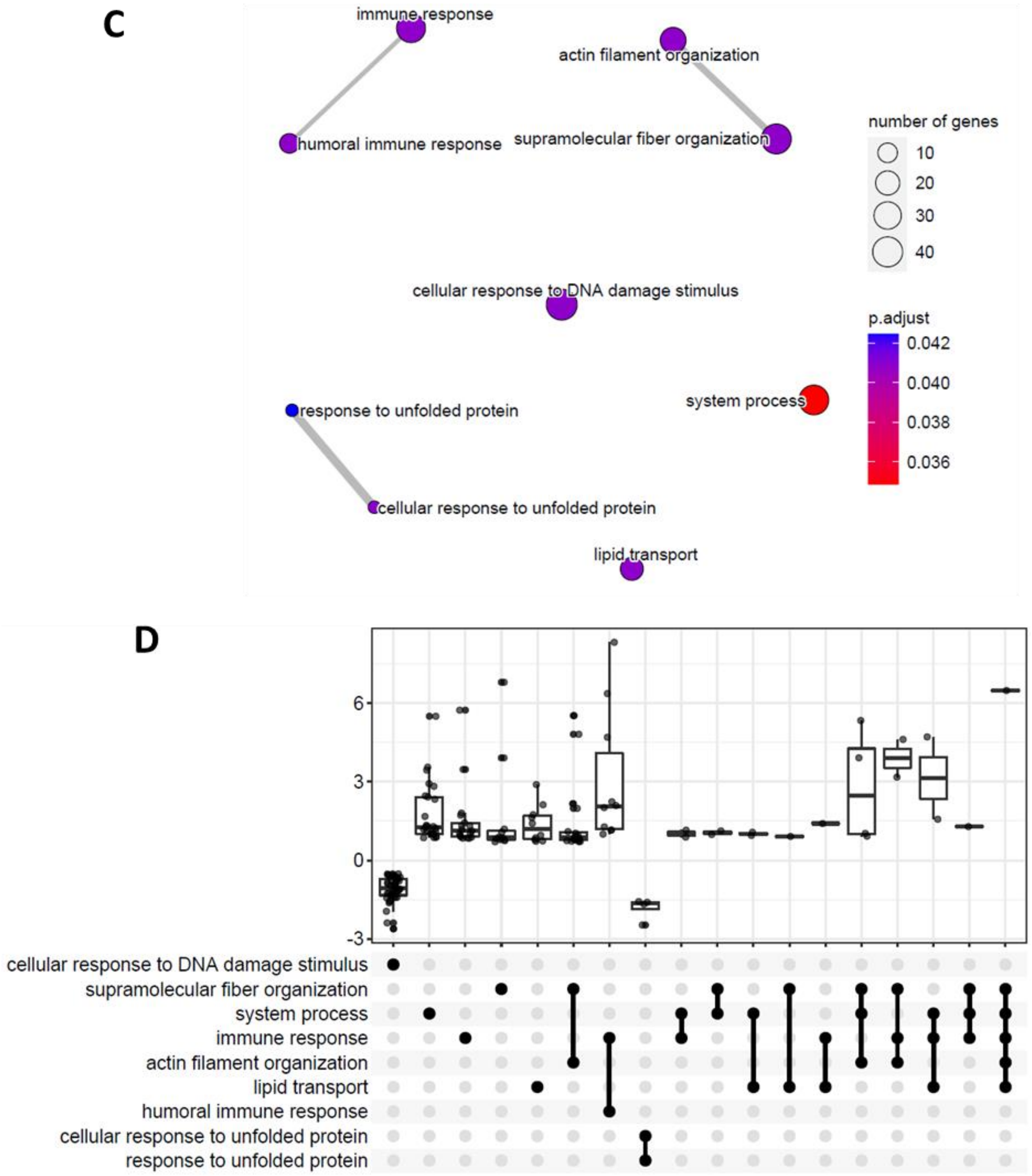


Figure 56 GSEA analysis for differentially expressed genes in infected chIFITM KO DF-1 GSEA analysis was conducted in clusterProfiler on the ranked list of differentially expressed genes generated with DESeq2 for the comparison infected vs uninfected KO DF-1 clonal cell lines. A) Dotplot – GeneRatios plotted on the x-axis indicate the fraction of genes from the GeneSet that are recognised in the experimental dataset. Size of the circle is meant to reflect their absolute number. B) Ridgeplot illustrates distribution of enrichment scores for genes identified in a particular pathway. C) Emapplot. Enriched pathways are networked via nodes where same genes are recognised in both. Their size reflect the number of genes represented in the pathway and colour. Colour indicates statistical significance in all plots. D) UpSet plot. Top panel displays the distribution of fold changes of enriched genes. Black dots indicate communal elements. If same genes are detected as enriched in more than one pathway, dots are connected with a black line.

PADOG analysis

To investigate differentially regulated pathways further, a pathway analysis using PADOG was performed in ReactomeGSA [111,112]. Differential expression analysis on the pathway level found that 2419 pathways are differentially regulated. Fold changes were small (1>) for the most statistically significant pathways (FDR <0.001)

Most upregulated pathways were concerned with cellular response to stimuli and infectious disease, specifically SARS-Cov2 and Influenza. Some examples are: Defective SLC26A4 causes Pendred syndrome (PDS) (R-HSA-5619046), TNF receptor superfamily (TNFSF) members mediating non-canonical NF-kB pathway (RHSA-5676594), Cellular responses to stress (R-HSA-2262752), Cellular responses to stimuli (R-HSA-8953897), Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex (R-HSA-75035), PTK6 Regulates RTKs and Their Effectors AKT1 and DOK1 (R-HSA-8849469), Signaling by NOTCH (R-HSA-157118), Defective SLC2A2 causes Fanconi-Bickel syndrome (FBS) (R-HSA-5619098), MECP2 regulates transcription of neuronal ligands (R-HSA-9022702), RHO GTPases activate PKNs (R-HSA-5625740), Synthesis of DNA (R-HSA-69239), Switching of origins to a post-replicative state (R-HSA-69052), Assembly of the pre-replicative complex (R-HSA-68867), Interleukin-1 signaling (R-HSA-9020702), Metabolism of proteins (R-HSA-392499), FBXL7 down-regulates AURKA during mitotic entry and in early mitosis (R-HSA-8854050), SARS-CoV-2-host interactions (R-HSA-9705683), Cyclin E associated events during G1/S transition (R-HSA-69202), Cyclin A:Cdk2-associated events at S phase entry (R-HSA-69656) and Orc1 removal from chromatin (R-HSA-68949).

Downregulated pathways were part of the following processes DNA replication, translation, cell cycle regulation, membrane trafficking and post-translational protein modification and antigen presentation and interferon gamma singalling. Most strongly regulated pathways included: HDMs demethylate histones (R-HSA-3214842), Defective MPI causes MPI-CDG (R-HSA-4043916), Post-transcriptional silencing by small RNAs (R-HSA-426496), TET1,2,3 and TDG demethylate DNA (R-HSA-5221030), TICAM1 deficiency - HSE (R-HSA-5602566), PTK6 Down-Regulation (R-HSA-8849472), Regulation of RUNX1 Expression and Activity (R-HSA-8934593), Competing endogenous RNAs (ceRNAs) regulate PTEN translation (R-HSA-8948700), Rab regulation of trafficking (R-HSA-9007101), WNT mediated activation of DVL (R-HSA-201688), NR1H2 & NR1H3 regulate gene expression linked to lipogenesis (R-HSA-9029558), Inhibition of TSC complex formation by PKB (R-HSA-165181), SUMOylation of chromatin organization proteins (R-HSA-4551638), Regulation of PTEN mRNA translation (R-HSA-8943723), Post-translational protein modification (R-HSA-597592), Transcriptional Regulation by MECP2 (R-HSA-8986944).

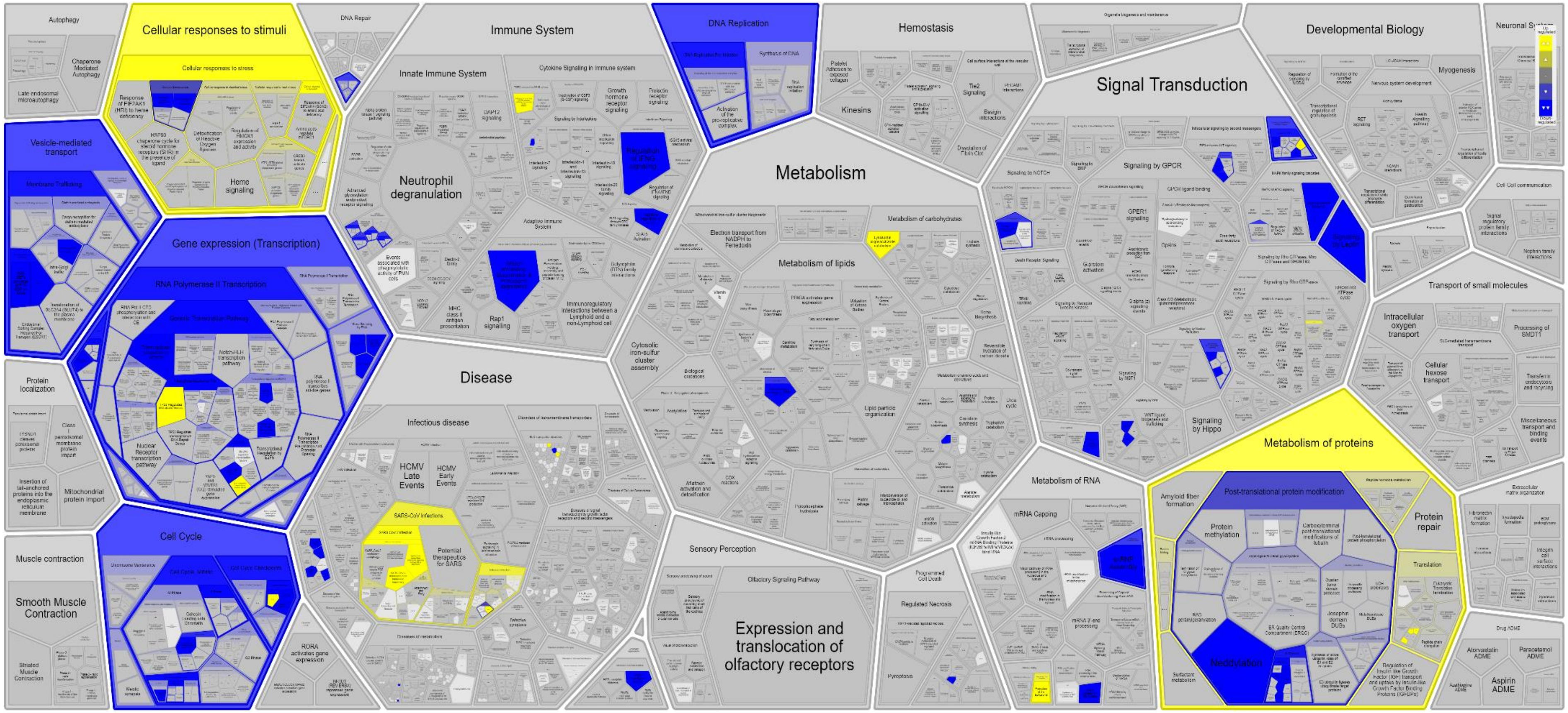


Figure 57 PADOG analysis of the KO DF-1 fibrobalsts at 6h post infection with H5N3 influenza PADOG analysis was conducted with ReactomeGSA on raw RNA-seq counts. Cpunts from all clones were pooled and compared infected vs non-infected. Statistical significance was capped at 0.05. Downregulated terms are coloured in blue and upregulated in yellow.

4.3.2.3 Effect of the chIFITM knock-out on the transcriptome of the DF-1 cell line

Differential expression

To establish the differences in the transcriptome between the WT and the KO DF-1 cell lines in basal levels of gene expression, differential expression analysis was performed in DESeq2 using the Wald test. Means model without intercept, represented by the design formula: ~0 + group, was used. FDR was kept at 0.05. Results were extracted using contrasts, thus the *ashr* shrinkage method was employed to adjust the effect sizes. WT DF-1 cell line in the control condition was compared to the three clonal KO cell lines pooled; the coefficients were appropriately adjusted for their contribution.

Differentially expressed genes that were not statistically significant were removed and the remainder was sorted by decreasing logFoldChange. Extreme one hundred genes at either end of the list were extracted and their TPM values standardised using the Z-score method in baseR. Values were clustered using Euclidean distances and the heatmap was produced in R with the *pheatmap* package.

1126 (6.2%) genes were upregulated, 1389 (7.7%) genes were downregulated and 2463 (14%) were not considered in the analysis due to low counts.

Genes that are more highly expressed in the KO cell lines on average show varied distributions of expression among individual clonal cell lines. They are expressed to a lower level in the WT irrespective of the condition. KO clonal cell lines 250 and 600 show most similar transcriptional profile between themselves and they separate by the condition – indicating they have similar response to infection. Because the infected samples of KO DF-1 250 and 600 cluster together with the WT infected samples, we can conclude they share similar pattern of upregulation during infection. Note that we did not test for condition in this comparison; we are solely observing the levels of expression of those genes in infected samples. Interestingly, a big cluster of genes that differentiate the WT DF-1 from the KO DF-1 cell lines are most highly expressed in the clone 225 (*VDAC1, GGT7, CACNG5, PLA2G4A, CHRNA7, HVCN1, CLDN11, LOC416197, NEFH, CYP19A1, PATL2, EPS8, VAT1, SNX10, HPD, CD3D, MYOM1, LRRC9, ETV7, EPHA1, HHATL, AOX1, VILL, HEBP1, ADAP1L, PROX1, LRRTM3, CYTH4, BIRC7, ADRA1B, ENSGALG00000000314, TOR4A, SCNN1D, LOC417937, RARB, LRIG3, PDE7B, AHR1A, AHNAK2, PDZD2, MID2, GLIS1, SLC5A1, TRPV3, CTNNAL1, ENSGALG00000011792, RELN, NEK10*). This pattern is more prominent and consistent in the control condition.

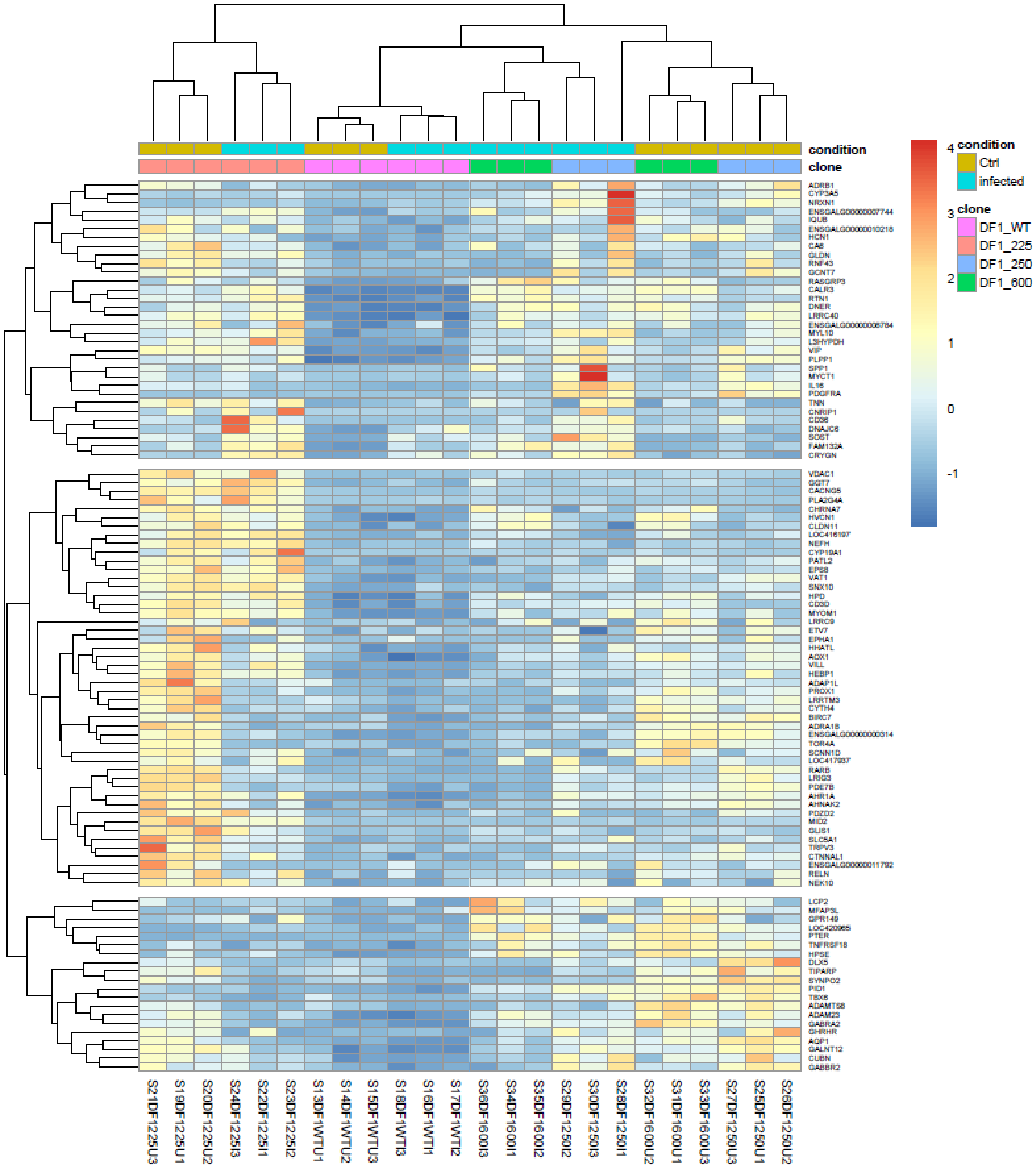


Figure 58 Most upregulated genes in comparison KO vs WT DF-1 cell line Expression values in the control condition between the WT and the pooled three KO clonal DF-1 lines were compared. Differentially expressed genes with adjusted p-value >0.05 were deemed statistically significant. Their TPM values were extracted and normalised across all samples using Z-score method. Heatmap was produced by clustering their Euclidean distances in pheatmap package in R.

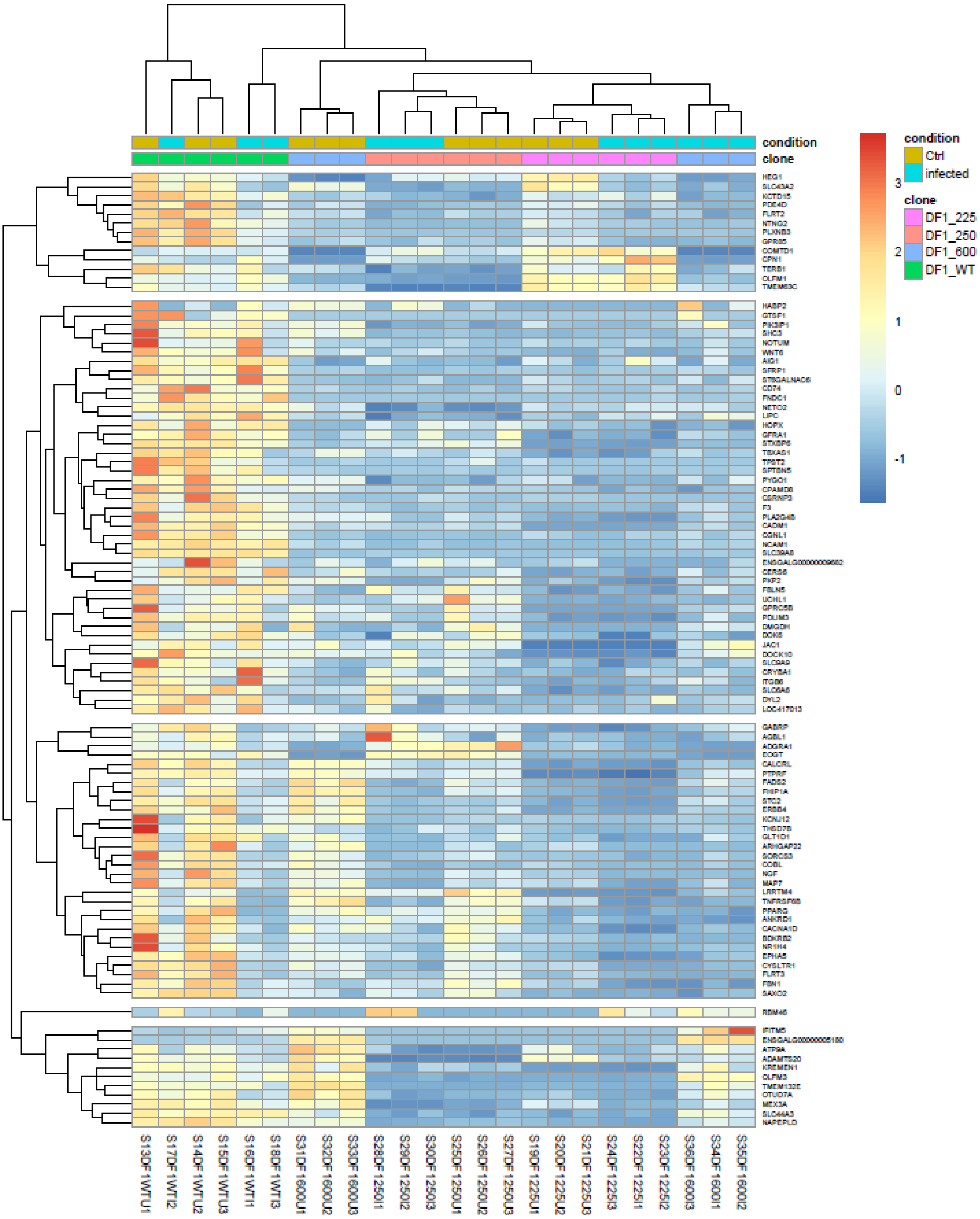


Figure 59 Most downregulated genes in comparison WT vs KO DF-1 cell lines Differential expression was performed between WT and pooled KO DF-1 cell lines in cmock condition. P-adjusted value 0.05 was used as cutoff for significance and ashr method was used for shrinking. Statistically significant genes were sorted by logFoldChange. TPM values of hundred most downregulated genes were Z-scored and clustered by Euclidean distances in *pheatmap* package in R..

Genes that are expressed to a higher level in WT as opposed to the K0 cell lines on average, show less consistent levels of expression among replicates in comparison to the genes that get upregulated in the K0 clones. The expression profile of the most downregulated genes in comparison WT vs KO shows a trend of down regulation in the infection as well, in doing so even further lowering the level of expression in the KO clones. The clustering is not straightforward by condition or genotype, indicating very mixed gene regulatory dysregulation is taking place in the clonal cell lines.

Overall, no uniquely WT-specific pattern emerged. Nevertheless, genes downregulated in the clones, exhibit an expression behaviour similar among all KO clonal cell lines. In contrast, for the upregulated genes: KO_DF1 225 seems to be expressing higher levels of upregulated genes than the other c*hIFITM* KO DF-1 cell lines.

Functional analysis

GSEA

To elucidate if the knock-out of the *chIFTIM* locus causes changes on the gene network level in DF-1 chicken fibroblasts GSEA was conducted in clusterProfiler using the differentially expressed gene list with shrunken effect sizes with following parameters: minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.05, pAdjustMethod = "BH", eps = 0.

Most statistically significant pathways were activated and concerned with developmental processes and related signalling pathways (**Figure 60**). Animal organ development, even though not statistically significant at FDR 0.05, was represented by a meaningful gene set size that was not overlapping with other processes indicating a strong signal as detected by GSEA. Supressed processes were always represented by single genes or very small gene sets.

PADOG

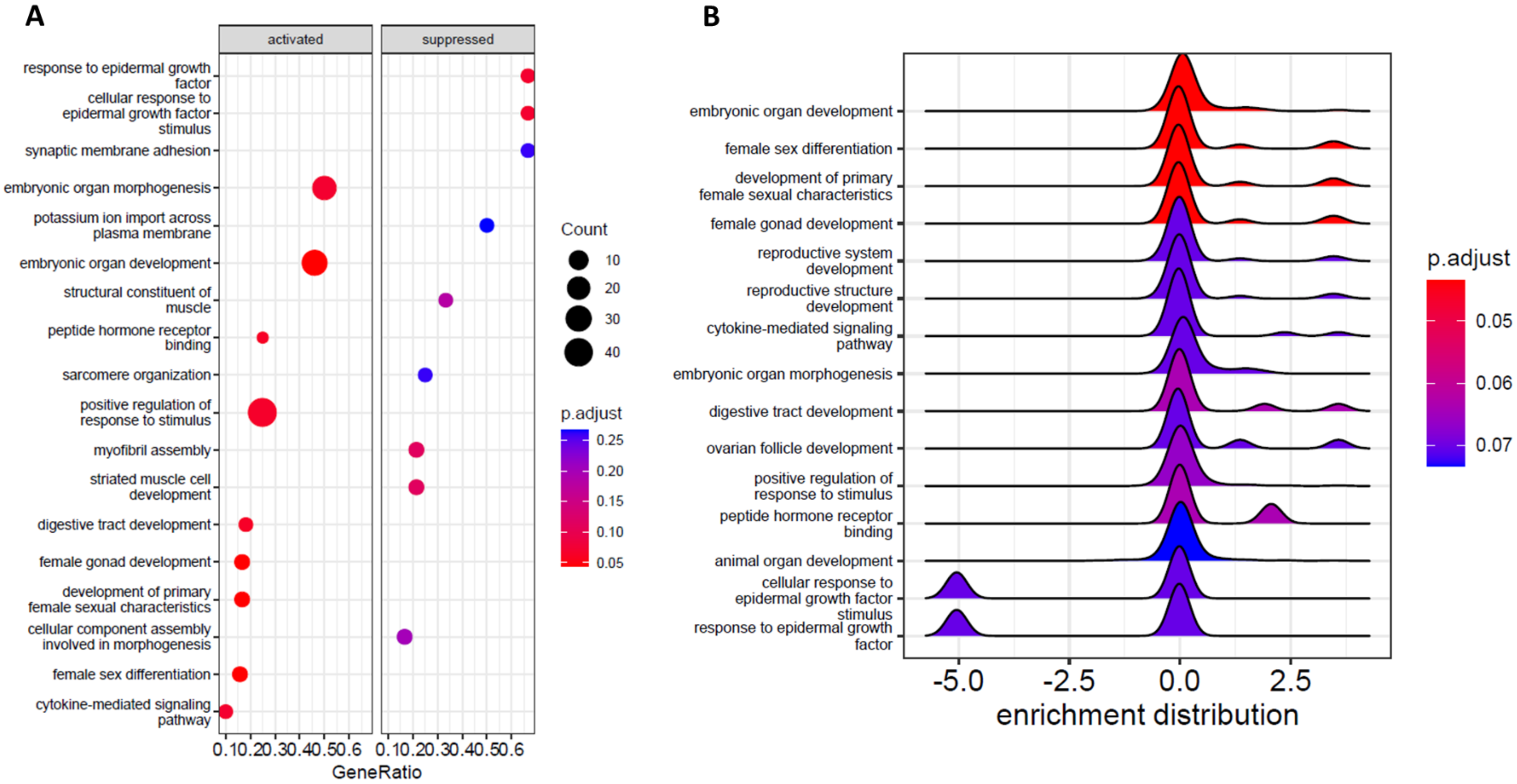
Pathway Analysis with Downweighing of Overlapping Genes was conducted with ReactomeGSA using raw RNA-Seq counts. The method conducts differential expression analysis on the pathway level using FDR of 0.05 as the cut-off for statistical significance. 2403 pathways were found to be differentially regulated passing the FDR filter. As is frequent in pathway analyses, very significant pathways were signalling pathways with multiple functions and more specific processes were represented by small gene-set sizes.

Top upregulated pathways included Activation of RAS in B cells (R-HSA-1169092), Calcineurin activates NFAT (R-HSA-2025928), Chaperonin-mediated protein folding (R-HSA-390466), Protein folding (R-HSA-391251), Ras activation upon Ca2+ influx through NMDA receptor (R-HSA-442982), Defective GSS causes GSS deficiency (R-HSA-5579006), CLEC7A (Dectin-1) induces NFAT activation (R-HSA-5607763), RAF/MAP kinase cascade (R-HSA-5673001), Oncogenic MAPK signaling (R-HSA-6802957), NR1H2 &

NR1H3 regulate gene expression to control bile acid homeostasis (R-HSA-9623433), BMAL1:CLOCK,NPAS2 activates circadian gene expression (R-HSA-1368108), Adrenoceptors (R-HSA-390696), Defective PGM1 causes PGM1-CDG (R-HSA-5609974), ARL13B-mediated ciliary trafficking of INPP5E (R-HSA-5624958), Blockage of phagosome acidification (R-HSA-9636467), Translation of Replicase and Assembly of the Replication Transcription Complex (R-HSA-9679504), Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell (R-HSA-198933), Downstream signalling events of B Cell Receptor (BCR) (R-HSA-1168372), MAPK family signalling cascades (R-HSA-5683057).

Down-regulated processes were concerned with signalling, transport and proliferation: PLCG1 events in ERBB2 signaling (R-HSA-1251932), Synthesis of bile acids and bile salts via 27-hydroxycholesterol (R-HSA-193807), GRB2 events in ERBB2 signaling (R-HSA-1963640), EGFR Transactivation by Gastrin (R-HSA-2179392), Metal ion SLC transporters (R-HSA-425410), Regulation of commissural axon pathfinding by SLIT and ROBO (R-HSA-428542), Inactivation of CDC42 and RAC1 (R-HSA-428543), Role of ABL in ROBO-SLIT signaling (R-HSA-428890), Zinc transporters (R-HSA-435354), Zinc influx into cells by the SLC39 gene family (R-HSA-442380), Cargo trafficking to the periciliary membrane (R-HSA-5620920), Trafficking of myristoylated proteins to the cilium (R-HSA-5624138), Constitutive Signalling by EGFRvIII (R-HSA-5637810), Signalling by EGFRvIII in Cancer (R-HSA-5637812), Downstream signalling of activated FGFR1 (R-HSA-5654687), Signalling by FGFR1 (R-HSA-5654736), Ethanol oxidation (R-HSA-71384), Gastrin-CREB signalling pathway via PKC and MAPK (R-HSA-881907), SLIT2:ROBO1 increases RHOA activity (R-HSA-8985586), Signalling by ERBB2 ECD mutants (R-HSA-9665348).

Wider biological themes that were statistically significant were Autophagy, MPAK Signalling and Organelles Biogenesis and Maintenance, all three upregulated (**Figure 61**).


A
activated
suppressed
response to epidermal growth factor
cellular response to epidermal growth factor stimulus
synaptic membrane adhesion
embryonic organ morphogenesis
potassium ion import across plasma membrane
embryonic organ development
structural constituent of muscle
peptide hormone receptor binding
sarcomere organization
positive regulation of response to stimulus
myofibril assembly
striated muscle cell development
digestive tract development
female gonad development
development of primary female sexual characteristics
cellular component assembly involved in morphogenesis
female sex differentiation
cytokine-mediated signaling pathway
0.1 0.2 0.3 0.4 0.5 0.6
0.1 0.2 0.3 0.4 0.5 0.6
GeneRatio
Count
10
20
30
40
p.adjust
0.25
0.20
0.15
0.10
0.05
B
embryonic organ development
female sex differentiation
development of primary female sexual characteristics
female gonad development
reproductive system development
reproductive structure development
cytokine-mediated signaling pathway
embryonic organ morphogenesis
digestive tract development
ovarian follicle development
positive regulation of response to stimulus
peptide hormone receptor binding
animal organ development
cellular response to epidermal growth factor stimulus
response to epidermal growth factor
-5.0
-2.5
0.0
2.5
enrichment distribution
p.adjust
0.05
0.06
0.07

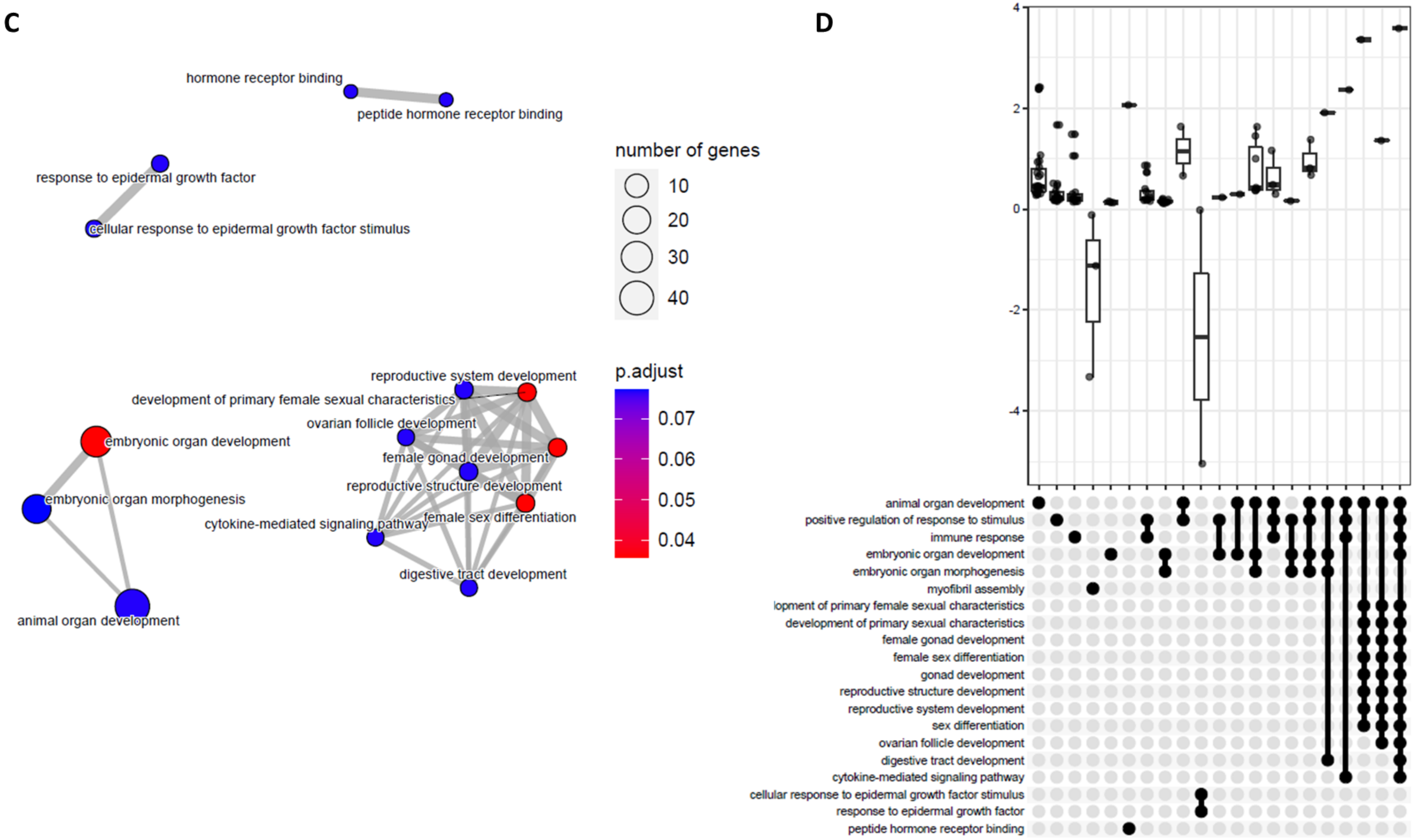


Figure 60 Pathway analysis for comparison WT vs KO DF-1 fibroblasts GSEA was conducted in clusterProfiler with p-value cutoff 0.3 and BH-correction. A) Dotplot B) Ridgeplot C) Emapplot D)UpSetplot

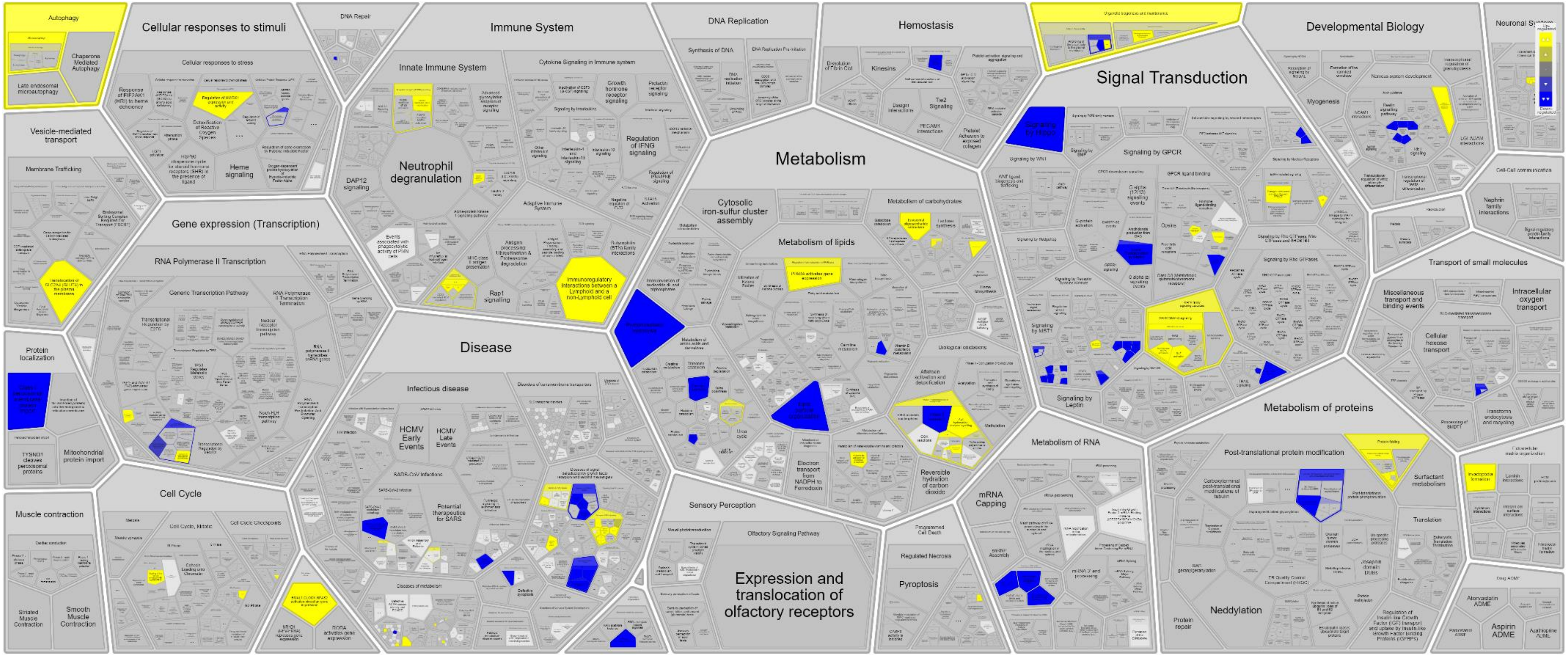


Figure 61 PADOG analysis for comparison WT DF-1 vs KO DF-1 in uninfected condition Raw RNA-Seq counts were uploaded to Reactome and the differential expression analysis on the pathway level was conducted using the web interface with default parameters. Blue highlighted pathways are downregulated and yellow upregulated. Strong tone indicates statistical significance, muted tones mark differentially regulated pathways which are not statistically significant.

4.4 Discussion

It has been more than a decade since the 2009 studies by Brass *et al* [40] and Shapira *et al* [164] that ignited the research of IFITMs and their role in the innate immune system against viral infection. Research interest took off significantly and numerous studies were conducted subsequently to characterise the host range, mechanism of action and regulation of gene expression, susceptible viral species and strains as well as the spectrum of roles the IFITMs perform beyond the immune system.

Significantly less resources have been invested towards research into avian IFITMs relative to mammalian IFITMs. Given the 300 million years of evolutionary history between birds and mammals, and the high selection pressure constantly being exerted on the immune genes, it is not clear how many parallels can be drawn between characteristics of mammalian and avian IFITMs. Gaps are still present in our knowledge of their localisation, functional residues that mediate the mechanism of restriction, the regulation of their expression and even the host and pathogen range within which they operate. Concerning the localisation of avian IFITM proteins, two pieces of work stand out. Smith *et al* [42] identified chIFITM1 and 2 at the plasma membrane and chIFITM3 at the endosome. This echoed the information known about huIFITM1 and 3 but was novel for chIFITM2. HuIFITM2 has been shown to localise on the endosomal membrane. Blyth *et al* [73] looked into duck IFITMs (dIFITM) and found a similar distribution to what was previously observed in the chicken, i.e. that dIFITM1 localises to the plasma membrane and dIFITM3 to LAMP1-containing compartments. Duck IFITM2 and 5 were partially localised to the endosomal membrane, but authors did not draw definitive conclusions but recommend further research.

Uncovering the localisation and the associated mechanism of action of chIFITMs is severely hindered by the lack of suitable antibodies. Generation of anti-chIFITM antibodies is extremely challenging because of high sequence similarities among homologues. Those obstacles notwithstanding, Kim *et al* [165] have had some success recently using the phage display. Should this antibody prove to be functional in wider experimental conditions, it will be an immense step forward in the research of chIFITMs.

Until now, all studies identifying chIFITM proteins for purposes of localisation and quantification of protein products relied on usage of tags in overexpression experiments. Tags can obstruct native protein folding, hinder partner interactions, affect transportation or suffer from cleavage. Nevertheless, when used with these considerations in mind, tags are an irreplaceable tool in molecular biology.

Aforementioned constraints directed significant research efforts towards approaches such as transcriptomics or indirect measurements of chIFITM activity, e.g. quantification of the viral yield. Here we generated stable *chIFITM* KO clonal cell lines and used bulk RNA-seq to investigate the effects of loss of *chIFITM* function on the transcriptomic level.

As previously shown by Bassano *et al* [49], sequencing the *chIFITM* locus is not a straightforward task rendering the genetic engineering time-consuming. Nevertheless, three *chIFITM* KO DF-1 cell lines were generated by Horizon Discovery; namely 225, 250 and 600 each with different combination of gRNAs. Ablation of the locus was confirmed by quantitative RT-PCR and RNA-Seq; and no structural damage to the chromosome 5 was detected by karyotyping. All clones consistently supported higher titres of H5N3 and H3N2 compared to the parental WT DF-1 cell line. These results confirm that chIFITMs effectively restrict the influenza infection and that the magnitude of the effect depends on the strain [42,166]. Meaningful difference among the clonal cell lines in virus yield or in permissiveness to influenza virus were not detected; they were equally permissive to the influenza virus (**Figure 39**, **Figure 40**).

DF-1 *chIFITM* KO clones were generated from parental DF-1 cell line which is a mixed population. This increases the variation in the phenotype by another level of complexity beyond the random effects that cause variation in the phenotype even in isogenic cell lines such as growing conditions, age of the cells, off target CRISPR effects and stochastic genetic changes that take place as the cells live in culture. Awareness of these factors that contribute to the intrinsic variability in phenotype reduces the likelihood of false attribution of an observed change to the chIFITM KO.

PCA showed that the infection accounts for 56% of variability as opposed to 70% for WT DF-1 clones suggesting the presence of intrinsic variation unrelated to the condition. The distance among clonal cell lines within the same condition was 18%. This is similar to WT DF-1 clones with the caveat that this variability is occupied by a single clone: 608, whereas KO DF-1 clones seem to house more uniformly distributed variability, with no clone singularly standing out. The overlap of the replicate samples in the PCA indicates that the level of experimental noise was reduced to a minimum. Simultaneously, the separation of replicate samples in the infected condition suggests individual adaptation strategy which each cell line utilised to respond to infection in the absence of IFITMs.

Infection is an immense assault on the homeostasis of the cell. Consequently, the expression levels of many genes get affected. On the level of biological processes differential analysis revealed a small number of informative pathways to be significantly differentially regulated. Greater statistical power can be attained by increasing the number of replicates in future experiments. This is particularly the case for the study of transcriptomic response of the WT DF-1 to the infection.

Many pathways were similarly regulated as a response to infection regardless of presence or absence of IFITMs – immune response, cytokine signalling, downregulation of cell cycle regulation and proliferation. Interestingly, influenza immune response was specifically detected only in KO DF-1 cell lines using the PADOG method, potentially indicating that in the absence of IFITMs more immune-specific functions get upregulated as a compensation mechanism thereby facilitating easier detection during differential expression analysis. DNA repair was downregulated in both WT DF-1 and KO DF-1 cell lines. IAV induces host cell DNA damage through production of reactive oxidative species [167]. Subsequent DNA repair inhibition is one of the relatively typical strategies of the influenza NS1 pre-transcriptional modulation of host's immune response [168,169]. By inhibiting DNA repair the expression of all genes is reduced, including the immune genes. Golgi cisternae restacking was downregulated in the WT DF-1. This is in agreement with negative regulation of the cell cycle progression during infection, also detected in both WT and KO DF-1 fibroblasts. Golgi restacking is a process taking place during mitosis that ensures both daughter-cells inherit the organelle. Interestingly this was not detected in the KO DF-1 cells. Recently, Zhong *et al* [170] uncovered a role of IFITM3 in secretory pathway of glycoproteins from the Golgi. They found that in the conditions of higher levels of expression of IFITMs (e.g. ectopic over-expression or IFN stimulation) there is significant accumulation of glycoproteins in the peri-nuclear space. IFITMs, whose egress from the Golgi was impaired due to a mutation, affected the trafficking of glycoproteins in general. This hypothesis was validated when much higher distribution of intrinsically expressed VSV-G glycoproteins on the outer membrane was detected by confocal microscopy in *IFITMs* KO HeLa cells. This hindrance of correct processing of viral glycoproteins by the IFITMs is one of the strategies to contain the infection, one that would yet need to be proven in chicken.

Membrane trafficking is affected negatively in the *IFITM* KO DF-1 (as detected by PADOG). The most widely accepted mechanism of action of IFITMs posits that they recruit cholesterol to regulate membrane fluidity [53] making it obvious that IFITMs participate in other processes that involve membrane manipulation beyond those identified during an infection. Currently no evidence has been collected for chicken IFITMs, but mammalian IFITMs have been implicated in many processes in addition to viral infection: in trophoblast fusion during formation of the placenta [171], PI3K-signalling [172], cell migration, proliferation and adhesion predominantly in the context of cancer [173–176].

When the parental WT DF-1 cell line was compared with *chIFITM* KO DF-1 cell lines in mock infected condition major upregulated biological themes detected were Development, Autophagy, MPAK Signalling and Organelles biogenesis. Most statistically significant pathways were activated and concerned with developmental processes and related signalling pathways (**Figure 60**). Animal organ development, even though not statistically significant at FDR 0.05, was represented by a meaningful gene set size that was not overlapping with other processes indicating a strong signal detected by GSEA. The connection of *chIFITM* in DF-1 to animal organ development in this context is not clear. Leading edge analysis would provide information about which genes contribute the most to the enrichment signal and they could be investigated further thereby shedding some light on processes or proteins that work at the intersection of the animal organ development and the immune response. To date, no such work has been conducted in an avian model. Pathway analysis pipeline assumes projection of differentially expressed genes onto human pathways thus care must be taken when interpreting the results of the analysis by taking into consideration alternative functions of the same genes in birds or their absence altogether. IFITMs have been implied in the embryo development in mammals [177,178], specifically PGC migration which may explain the detection of developmental pathways as differentially regulated in this system. However, a few studies have disproven this hypothesis by using ifitm -/- mice who subsequently developed normally [179] or *Ifitm1*$^{tm1IEG/tm1IEG}$ mutant mice, with ifitm1 loss of function, whose fertility was not affected [180].

Supressed processes in comparison WT vs KO DF-1 were always represented by single genes or very small gene sets, diminishing the statistical confidence. Given that IFITM mechanism of action lends itself to many processes in the cell, it is conceivable many may be subsequently affected [181] thereby diluting the signal needed for precise detection. It is reasonable to expect transcriptomic differences based on different phenotypes among *chIFITM* KO DF-1 clonal cell lines observed in the laboratory. They were indeed detected by the PCA. However, strong communal response to infection among clones did not allow for extraction of confident clone-specific information that would explain differential ability of *chIFITM* KO DF-1 clonal cell lines to support influenza infection.

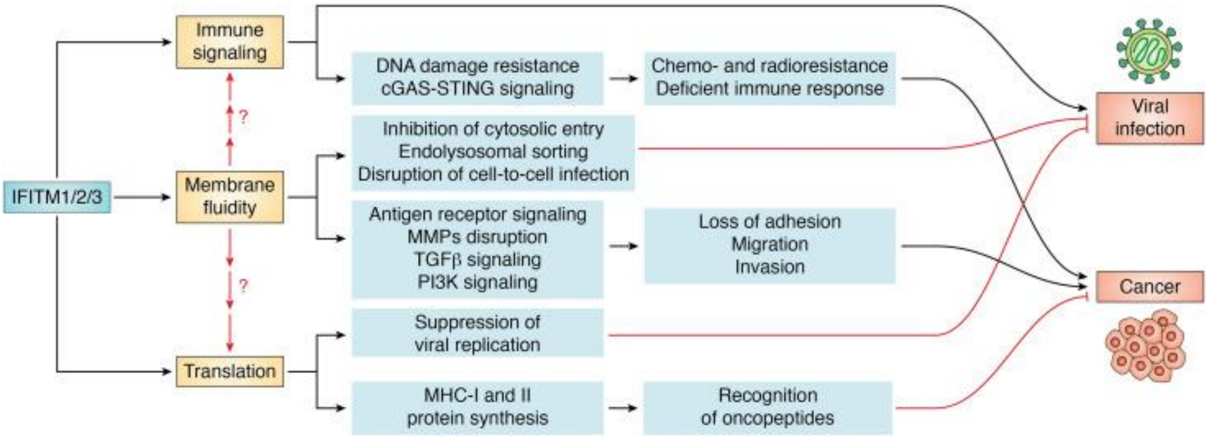


Figure 62 Diverse roles of IFITMs in viral infections and cancer in humans IFITMs mediate various processes which may be interdependent as part of general immune response, characteristic of viral infection and cancer. It is possible the immune strategies overlap as part of redundancy inbuilt into the system. Ostensibly the commonality could be IFITM-mediated manipulation of membrane structure and composition which in turn has secondary effects on many subsequent processes. Taken from Gomez et al [181].

In summary, in this section we show that *chIFITM* KO confers the DF-1 clonal cell lines the ability to support higher viral titre tan the WT in equal measure. Clonal cell lines have been characterised in terms their permissiveness to influenza virus, cytologic profile, viral yield and gene expression. Transcriptomics was used with aim of quantifying that variability and provide guideposts as to what regulatory mechanisms are taking place that support the phenotype observed in the laboratory. The analysis revealed great similarity among clones. Based on the findings in the pilot experiment described in the previous chapter, this RNA-seq study was better powered and the differential expression analysis resulted in more than 10 000 differentially expressed genes in every comparison conducted. All differentially expressed genes were employed in the pathways analysis which served to uncover biological processes that govern immune response in WT and chIFITM KO DF-1 clonal cell lines. KO cell lines were intrinsically more different than the WT clonal cell lines, as determined by the PCA which showed the drop of contribution of the infection status to the overall variance in the data drop by approximately 20% (from 70% for the WT to 56% for the KO). I investigated this variability by performing differential expression analysis between the WT and KO DF-1 fibroblasts and discussing differentially regulated pathways between the two groups. This knowledge will be valuable to understand better the role IFITMs may have outside of the immune system and aid in conceptualisation of future experiments studying chicken IFITMs.

5

chIFITM KO in PGC-derived fibroblasts and their innate immune response to influenza

5.1 Introduction

Despite the long-term efforts to reduce reliance of influenza vaccine production on embryonated chicken eggs, they remain a mainstay in the industry [182] due to the high antigen yield relative to other platforms and the wealth of information the community (academia, manufacturers, medical profession, and regulatory bodies) accumulated in the 80-year-old history of influenza vaccination [183], vaccine production and antigen calibration [184] . Therefore, research conducted to optimise further this platform, alongside the development of new ones, remains very relevant.

Chicken primordial germ cells (chPGC) have a wide array of applications, among which the gene therapy research, genetic conservation and the production of genetically modified organisms are just a few [185–188]. This work aims to provide preliminary insights whether a genetically modified chicken, missing the *chIFITM* locus, could be a useful tool for virus vaccine production. A model was developed in form of chicken PGC derived fibroblasts by Dr McGrew at the Roslin Institute. It approximates the behaviour of the chicken innate immune system modelling a more natural response to infection and simultaneously facilitates the comparison with DF-1 fibroblasts that serve as a model of a cell culture-based platform. Two primary PGC-derived fibroblast cell lines were produced; a CRISPR-edited *chIFITM* KO and a corresponding WT PGC which were subsequently differentiated into fibroblasts, from here onwards referred to as PGC-KO and PGC-WT for convenience.

This approach was chosen because of its flexibility. chPGCs are easily retrieved from the chicken embryo, cultured in vitro in suitable conditions[189,190], they tolerate editing and can be subsequently manipulated in a desired way: to generate offspring [191] or to be differentiated into a specific cell type [192]. Fibroblast was chosen to facilitate the comparability with the established avian cell lines used in the chicken immunology research, most notably DF-1, which are also fibroblasts.

5.2 Objectives

- Establish stable cell culture conditions for growth and cell culture maintenance for PGC-derived fibroblasts
- Confirm the deletion of the *chIFITM* locus
- Confirm that PGC-derived fibroblasts are successfully infected by the influenza strains used with DF-1 cells and the suitability of H5N3 as the model virus in PGC-derived fibroblasts
- Characterise the transcriptome of the WT PGC-derived fibroblasts upon infection with the influenza virus

- Investigate the alterations in the basal transcriptome of PGC-derived fibroblasts upon *chIFITM* knock-out
- Compare the transcriptomic response to influenza at 6hpi between the WT and *chIFITM* KO PGC-derived fibroblasts

5.3 Results

5.3.1 Characterisation of PGC-derived fibroblasts

5.3.1.1 Locus architecture

Chicken *IFITM* locus is clustered on the chromosome 5. It consists of 4 genes: *chIFITM1, chIFITM2*, *chIFITM3* and *chIFITM5*, flanked by telomeric *B4GALNT4* and centromeric *ATHL1. chIFITMs* are all transcribed in the same direction from the reverse strand.

First efforts to characterise the locus were published by Smith *et al* [42] pinpointing the location of the flanking genes. The identification procedure of individual *chIFITM* genes was anchored in the sequence similarity of CIL (cytoplasmatic intracellular loop) region among them, meaning that in the absence of exhaustive functional data for *chIFITMs* as well as poor sequence coverage at the time, a syntenic approach was used to annotate the genes. Furthermore, correct annotation is still difficult to ascertain today due in part to the low sequence identity between human and chicken *IFITM*s as well as the change in the direction of transcription for *chIFITM1* and *chIFITM2* between human and chicken putative orthologues, suggesting a gene transversion event, which renders the orthology assignment less straightforward. Therefore, the nomenclature was guided by gene position in the locus and the conservation of specific amino acids.

There is a relatively low level of sequence conservation between human and chicken *IFITMs,* specifically for *chIFITM1* and *chIFITM3*, at least in part due to the strong positive selection pressure they are under [193]. The degree of accuracy of the nomenclature still remains unclear today. Bassano *et al* [193] propose the following order based on the long-read sequencing data: centromeric – *B4GALNT4 – chIFITM3 – chIFITM2 – chIFITM1 – chIFITM5 – ATHL1* – telomeric (**Figure 65**). However, they also state that their own immunofluorescence data are in agreement with Smith *et al*. [42], reiterating the complexities involved in the derivation of an accurate annotation for the *chIFITM* gene family. The nomenclature introduced by Smith *et al* is used in this thesis, given there has been no substantial new functional evidence generated since to warrant a reconsideration.

A schematic representation of the locus, as hypothesised by Smith *et al*., is shown in **Figure 63**. Coloured blocks in the schematic indicate genes annotated before 2013 and their order within the locus. The only officially annotated chicken *IFITM* gene within the *chIFITM* locus is *chIFITM5* (**Figure 64**). White coloured blocks denote the putative genes that have not been annotated, but whose proposed nomenclature is supported by a collection of suggestive evidence [42,193].

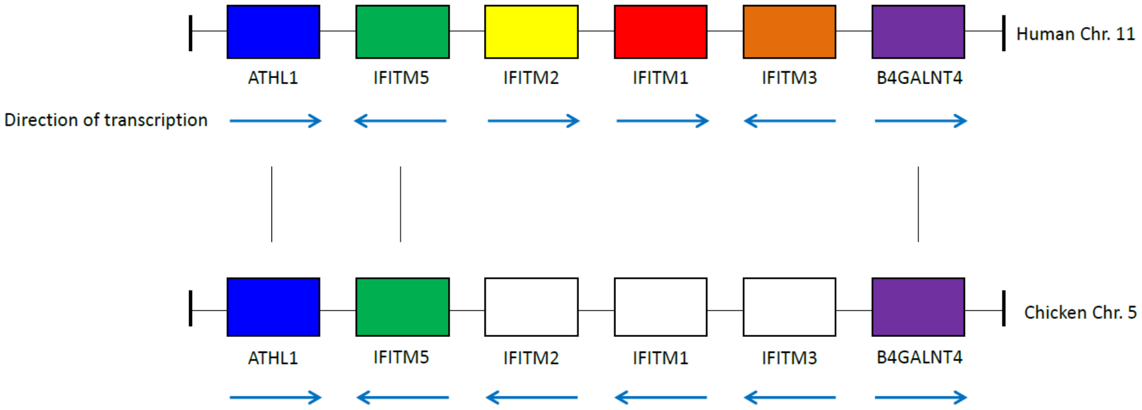


Figure 63 The chIFITM locus architecture Human *IFITM* locus is found on the chromosome 11 and flanked by *ATHL1* and *B4GALNT4*. Corresponding syntenic region in chicken genome is located on chromosome 5. Direction of transcription is indicated with blue arrows. Note that the diagram is depicting the reverse strand from which the genes are transcribed. Taken from (Smith et al., 2013).

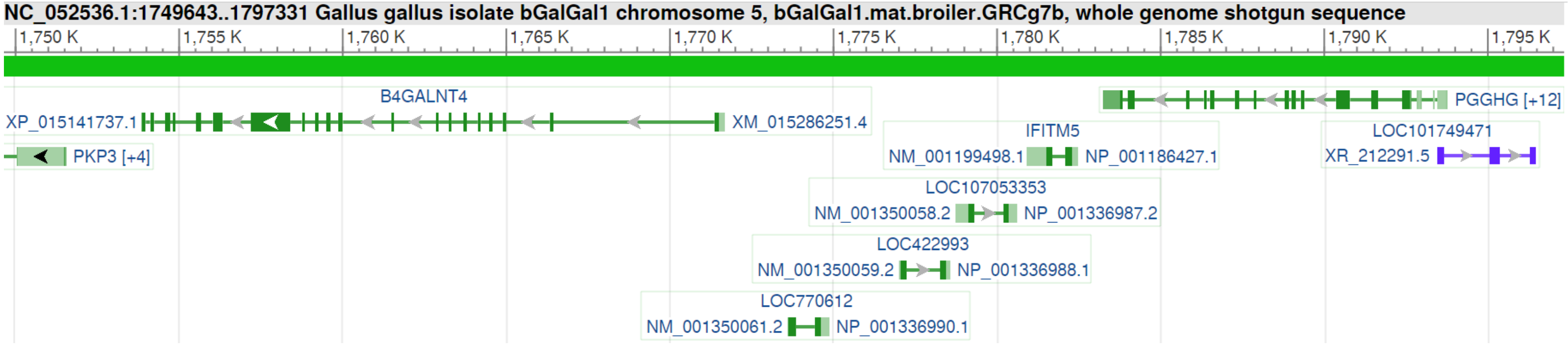


Figure 64 The chIFITM locus from current reference genome The newest (minor) release of chicken reference genome from 2021: GRCg7b, does not contain any alterations to the previous releases in the region of the *IFITM* locus. The sole annotated gene remains *chIFITM5.* Note the alternative name for *ATHL1*: *PGGHG* (protein -glucosylgalactosylhydroxylysine glucosidase). Retrieved from NCBI, Sequence Reference: NC_052536.1.

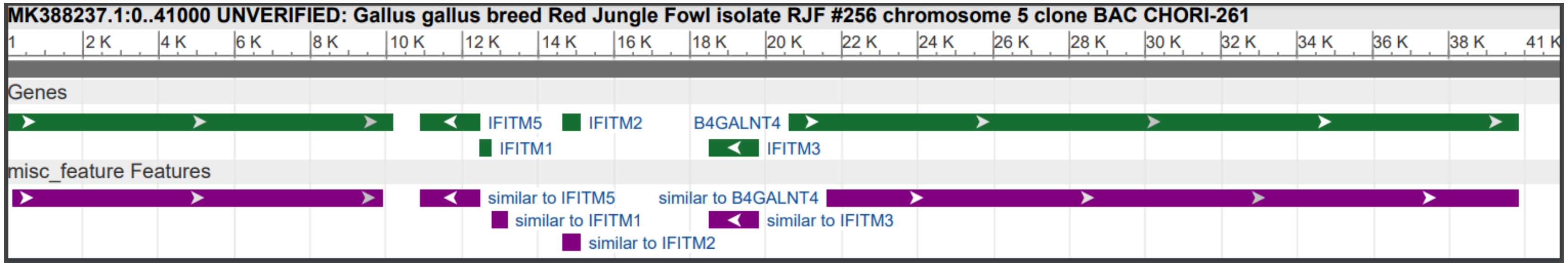


Figure 65 Graphical representation of the chIFITM locus A BAC clone was isolated to re-sequence the *chIFITM* locus to improve coverage and the accuracy of its structure (Bassano et al., 2017). Note the inversed order of *chIFITM1* and *chIFITM2.* Retrieved from NCBI, Sequence Reference: MK388237.1.

5.3.1.2 Confirmation of the chIFITM knock-out in the PGC-derived fibroblasts

A combination of transcriptomic and PCR techniques was deployed to characterise the *chIIFITM* deletions in KO PGC-derived fibroblasts.

The region of the *chIFITM* locus was shown to be very difficult to sequence using traditional sequencing methods, for reasons that are not clear. Genome features which typically render sequencing a challenge, such as high proportion of GC content, are not characteristic of the *chIFITM* locus. The lack of accurate sequences makes any genetic manipulation of the locus very complicated. It also contributes to the resulting paucity of information about the genetic variation of *chIFTM* locus.

Substantial progress has been made in the area of genetic variation by Bassano *et al.* [193]. They used Illumina MiSeq and PacBio sequencing technologies to analyse the *chIFITM* locus of a selection of commercially available breeds, indigenous Nigerian and Ethiopian chickens, European breeds and inbred chicken lines from The Pirbright Institute. They showed that the structure of the locus is preserved across diverse populations in addition to identifying a number of SNVs (single-nucleotide variants) of interest in *chIFITM1* and *chIFITM3*, in addition to highlighting that the *chIFITM2* exhibits the highest degree of variability among all the *chIFITM* genes.

chIFITM KO PGC-derived fibroblasts used in this research were a kind gift from Dr Michael McGrew at the Roslin Institute. Their *chIFITM* locus sequence has not been published. Based on the findings of Bassano *et al*., described in the previous paragraph, it was reasonable to assume comparatively low sequence divergence and use henceforth the sequence MK388237.1 as retrieved from NCBI [49] as a reference sequence for the purposes of confirming the *chIFITM* KO.

The objective in this section is to confirm the presence of the *chIFITM* genes in the WT PGC-derived fibroblasts and to obtain information about the locus knock-out in the KO PGC-derived fibroblasts.

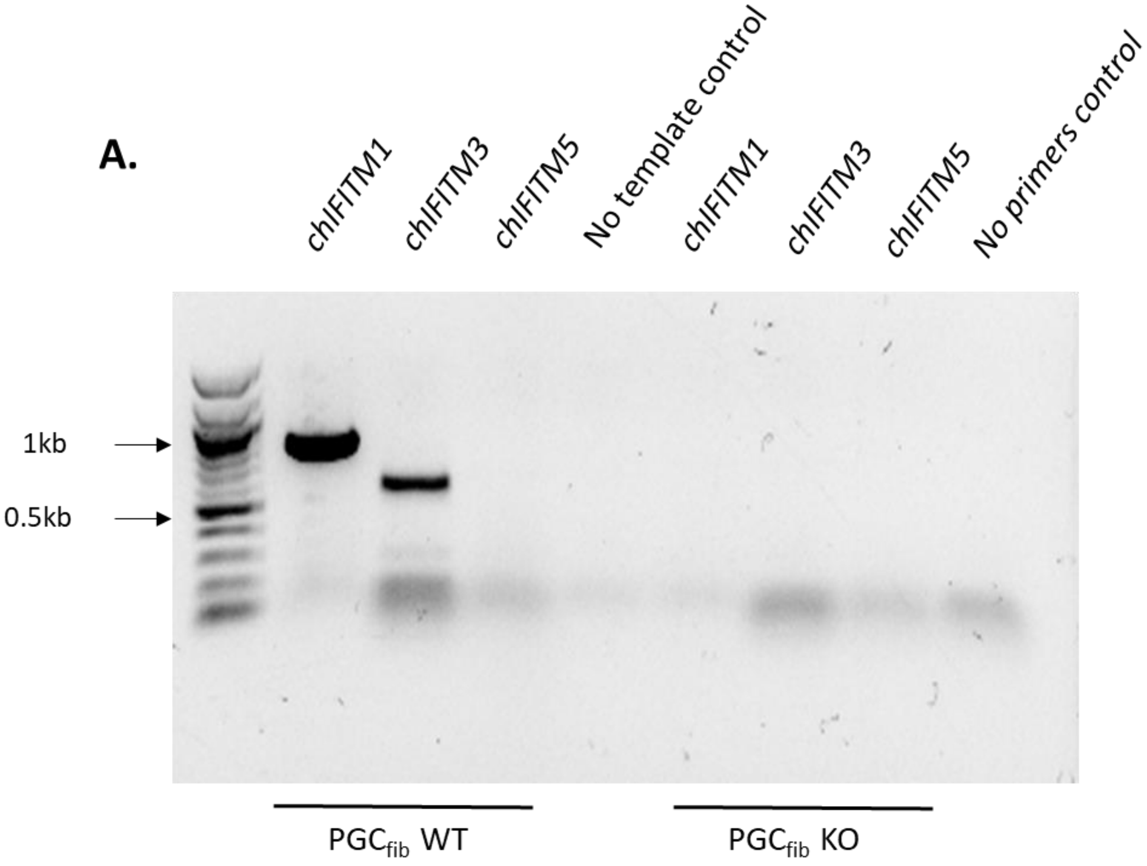
Same end-point PCR design that was used for the confirmation of the knock-out in the DF-1 cell lines, described in section 4.3.1.1 was used here.

Initially, primers based on DF-1 sequence of *chIFITM1, chIFITM3* and *chIFITM5* (see 2.1.3.3, and **Table 8**) were tested in an end-point PCR assay using a Q5 high-fidelity polymerase from NEB (as described in Materials and Methods section 2.2.3.2) on genomic DNA isolated from the cell culture.

Table 8 End-point PCR primers for chIFITM1, chIFITM3 and chIFITM5

Primer name	Primer sequence	5'-3' Orientation	Product size
IFITM1_F	GAGAGGGCCTCAAAACACAAAG	Forward	868 bp
IFITM1_R	GGATGAGATAGAATGGGACGGAG	Reverse	
IFITM3_F	ACCAGTTCCCTCCGTCCTCCC	Forward	615 bp
IFITM3_R	TCCACAGGATTCTGTGGGGTCCACG	Reverse	
IFITM5_F	CTGAGGCTGGGCTGGAGA	Forward	645 bp
IFITM5_R	ACCCCATGGGCACCCCTA	Reverse	

PCR analysis confirmed, albeit only partially, a successful knock-out of the *chIFITM* locus in KO PGC-derived fibroblasts (**Figure 66 A.**), as demonstrated by the absence of bands for *chIFITM*1 and *chIFITM*3 genes. Binding of the *chIFITM5* primers was unsuccessful for both templates.



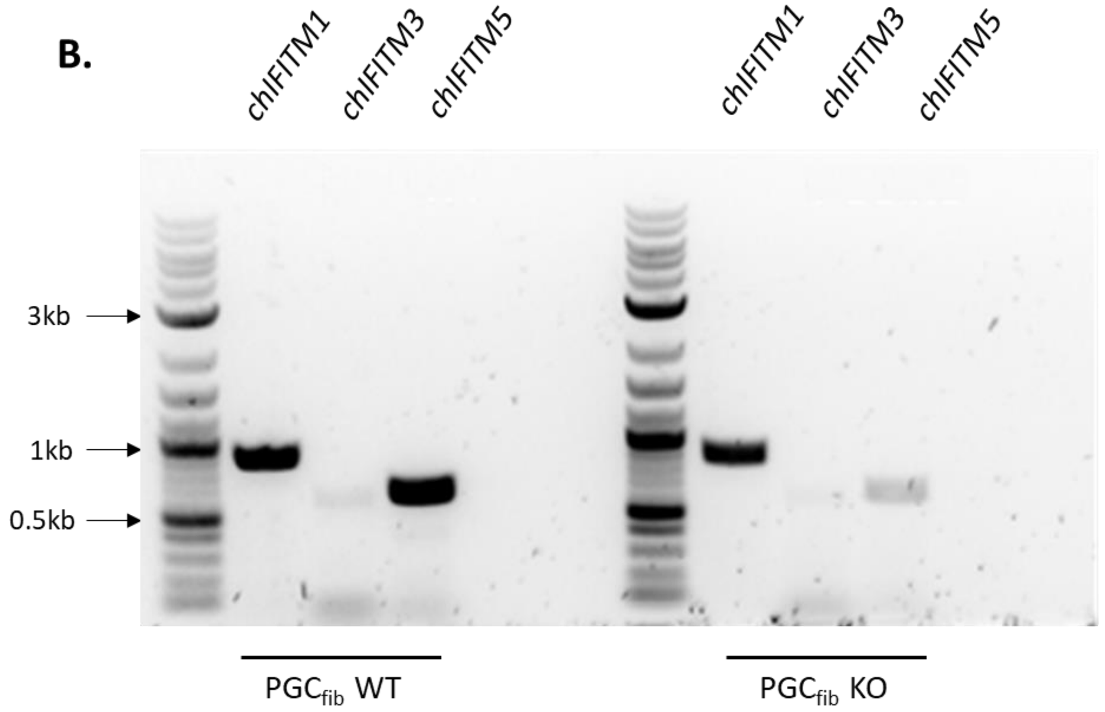


Figure 66 Detection of chIFITM genes in PGC-derived fibroblasts Genomic DNA was extracted from the PGC-derived fibroblasts and regions of interest were amplified in an end-point PCR assay with Q5 High Fideity DNA Polymerase. A. Agarose gel shows *chIFITM1* and *chIFITM3* products were detected in the WT and absent in the KO. Primer binding for chIFITM5 was unsuccessful and required further optimisation. B. End point PCR with same primers using KOD Xtreme Hot start DNA polymerase.

The amplification of *chIFITM5* was successful using KOD Xtreme Hot start DNA polymerase (**Figure 66. B.**) It is a high-fidelity polymerase specifically developed for the amplification of GC-rich regions, AT-stretches and long targets. *chIFITM1, chIFITM2, chIFITM3* and *chIFITM5* have 53,8%, 58.8%, 53.8% and 58.2% GC-content respectively. These values fall in the typical range of %GC content for an average genome, which is known to achieve good coverage in sequencing; it is not clear what is causing the observed difficulties in amplification of this region. This protocol generated bands of expected sizes (**Table 8**) in the wild type for *chIFITM1 and chIFITM5,* but not *chIFITM3* (**Figure 66**). More interestingly, it also produced conspicuous bands for same genes in the KO.

To investigate further those results, an additional endpoint PCR was conducted with an alternative experimental design, illustrated in **Figure 67**.

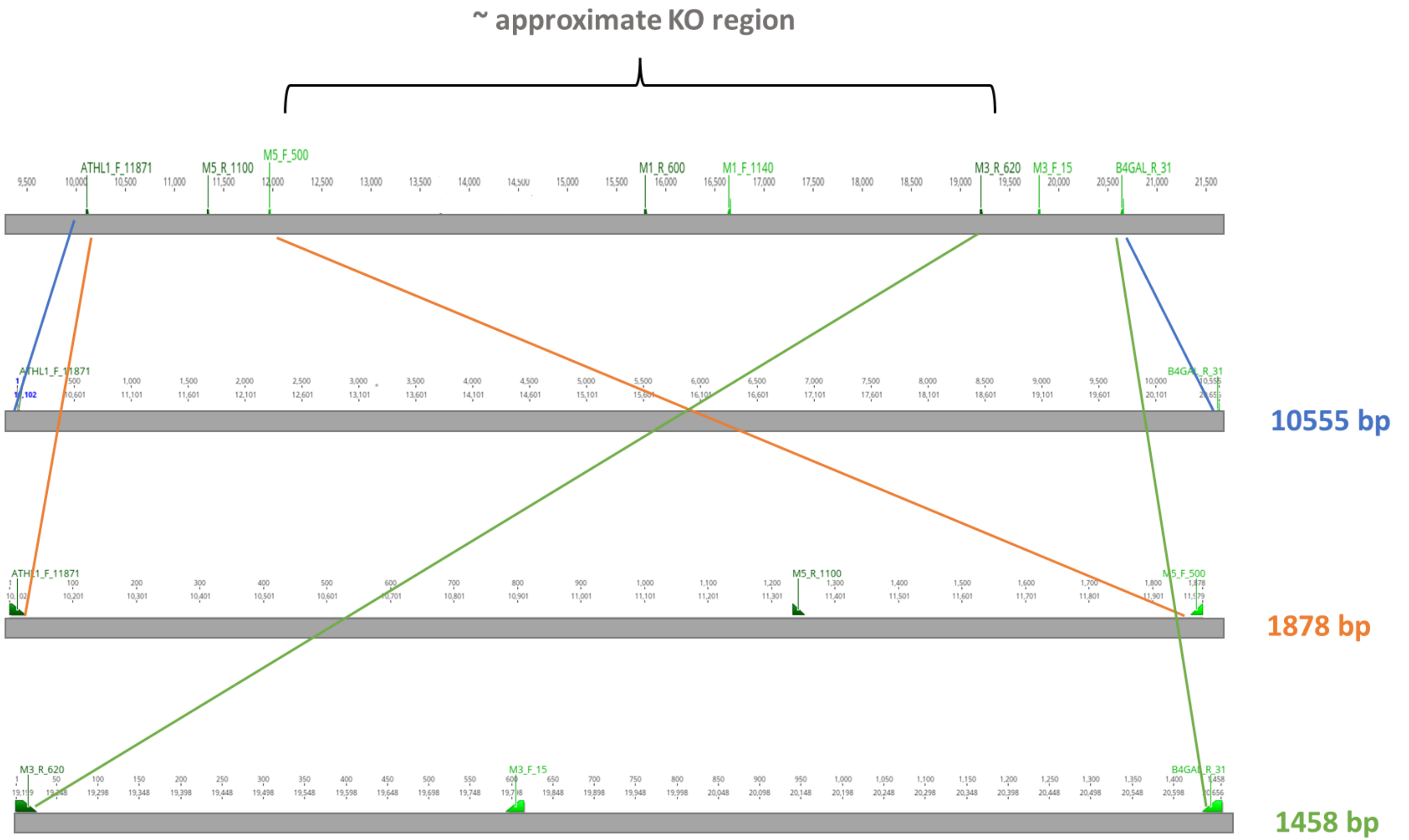


Figure 67 Alternative end-point PCR design for confirmation of the KO strategy in PGC-derived fibroblasts Four primers, ATHL1_F_11870, IFITM5_F_500, IFITM3_R_620 and B4GALNT4_R_31, in three combinations give rise to the following products: ATHL1_F_11870 and IFITM5_F_500 in orange (1878 bp), IFITM3_R_620 and B4GALNT4_R_31 in green (1485 bp), ATHL1_F_11870 and B4GALNT4_R_31 in blue (KO strategy dependent).

Evidence of a deletion in the *chIFITM* locus is shown in **Figure 68**.

PCR reaction with primers binding in *ATHL1* and ch*IFITM5* yielded a product of approximately 2kb in both WT cell lines tested: PGC-derived fibroblasts and DF-1 fibroblasts, whereby suggesting the centromeric end of the locus is identical between the two cell lines.

Analogous reaction designed to provide information about the telomeric end of the locus, with the primers targeting *B4GALNT4* and ch*IFITM3*, was not productive. This was most likely due to the difference between the melting temperatures for each primer being greater than 5°C. Optimisation of this reaction proved problematic. I was unable to identify a primer sequence that would bind in the desired region and have a satisfactory melting temperature as well as no predicted self-complementary tertiary structures.

PCR reaction between primers targeting the flanking genes generated a product of approximately 700-800 base pairs for the KO PGC-derived fibroblasts indicating that a deletion greater than 4kb took place. To identify the precise locations of the deletion the PCR product was sequenced by Sanger sequencing and the sequences were aligned with Geneious v11.1.5. [194] using ClustalW algorithm with default parameters or using MAFFT[195] also with default parameters (**Figure 69**).

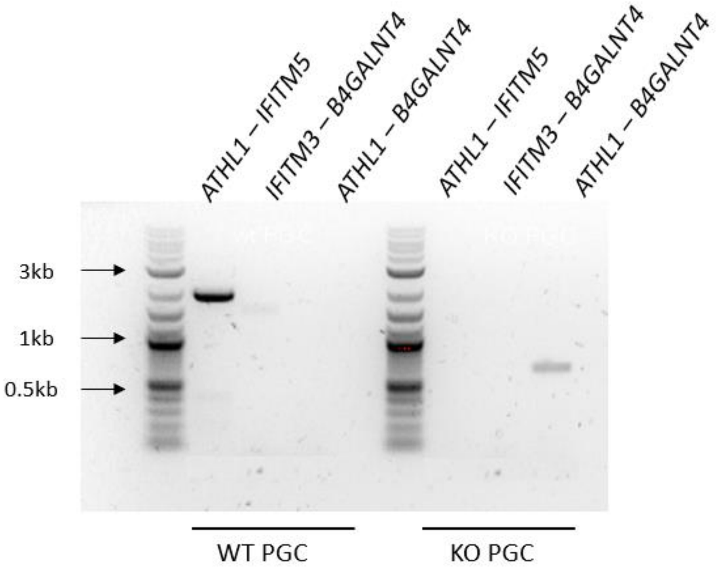


Figure 68 Confirmation of the chIFITM locus KO in PGC-derived fibroblasts Primers targeting *ATHL1, chIFITM5, chIFITM3 and B4GALNT4* used in different combinations. Resulting bands are indicative of the KO-strategy applied in each clone. Combination 2 failed to produce amplicons due to the suboptimal melting temperature in the PCR protocol.

Figure 69 Alignment of PCR products to the chIFITM locus reference Multiple alignment of the PCR products generated from by primers binding in *B4GALNT4* and *ATHL1* was performed with Geneious. Track 1 – *chIFITM* locus reference (NCBI, Sequence Reference: MK388237.1), Track 2 – *ATHL1* reference (NCBI, Sequence Reference: NC_052536.1_c1795013-1783230, Track 3 – *B4GALNT4* reference (NCBI, Sequence Reference: NC_052536.1_c1772658-1753033), Track 4 – WT PGC-derived fibroblast sequenced from 3' end, Track 5 - KO PGC-derived fibroblast sequenced from 3' end, , Track 6 - KO PGC-derived fibroblast sequenced from 3' end; solid black bar – consensus sequence, solid gray bar – sequence different to consensus, connector – gap introduced by the alignment. Identity bar has three colours – green indicates complete identity, yellow non-complete identity, and red very low identity.

In addition to the results confirming the knock-out presented here, we detect the IFITM genes in the KO PGC-derived fibroblasts using same endpoint PCR protocol, opening the possibility that the knock-out is not bi-allelic.

To further confirm the knockout, RNA-Seq data were inspected for the expression of the *chIFITM* genes using the quantitation feature of the RNA-Seq pipeline in SeqMonk v1.48.0 [140]. This measure is not evaluating differentially expressed genes using statistical methods. However, it is practical for QC purposes and provides a general overview of the expression data. Software counts primary alignments from the supplied BAM file over a chosen "probe" region, in this case mRNA. Output is $\log_2$RPM that has not been corrected for transcript length to facilitate high-level comparison of the reads mapping to the same gene among sample groups. By not expressing transcript abundance in kilobases per transcript we avoid confounding the length of the transcript with the level of observation and the associated error in measurement which are always dependent individual data set.

When comparing the WT with the KO we can observe that relative abundance estimates show drastic down-regulation in detected reads for all *IFITM* genes in the locus. Basal expression is the highest for *chIFITM5* among the *chIFITM* genes and almost absent for *chIFITM1* and *chIFITM3*. Very low expression levels of *chIFITM1* and *chIFITM3* in WT PGC-derived fibroblasts have been almost entirely ablated in the KO derivative, while transcript levels for *chIFITM2* and *chIFITM5* suffered a severe reduction. Active transcription quantitation, an approach that counts the reads aligning to the intronic regions, ostensibly coming from unspliced moieties of mRNA, thus representing newly transcribed molecules, produced the same results (data not shown).

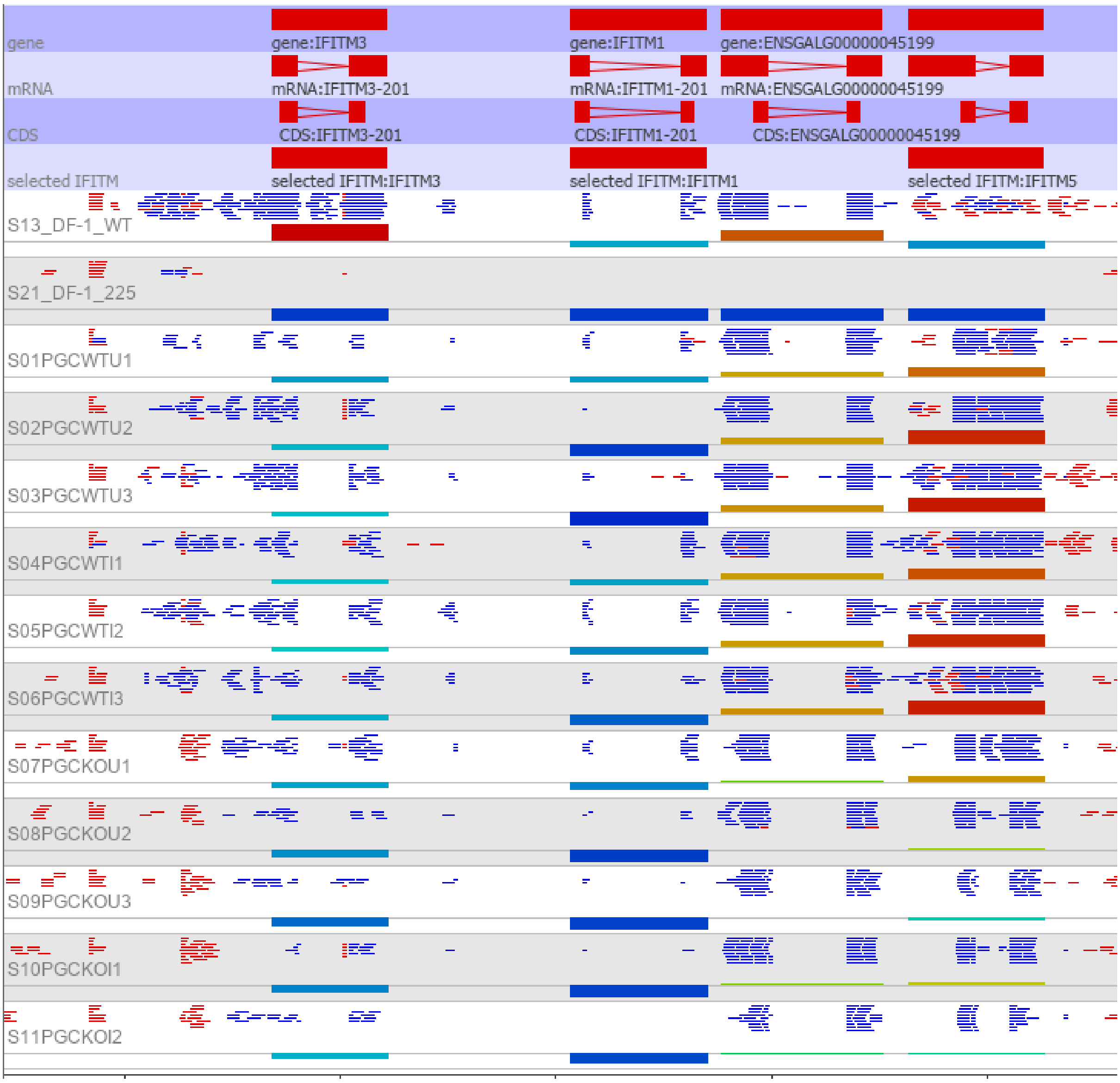


Figure 70 Visualisation of RNA-Seq reads for chIFITM genes Bulk RNA-Seq was conducted on Illumina NovaSeq and reads were aligned to Gallus gallus GRCg6a_v95 using STAR vSTAR 2.7.2a. Visualisation was produced with SeqMonk v1.48.0, using only the uniquely mapped reads. Raw reads are coloured red or blue indicating they map to forward or reverse strand, respectively. Tracks show the *chIFITM* locus with individual genes from left to right: *IFITM3, IFITM1, IFITM2* and *IFITM5* and corresponding features from top to bottom: gene, mRNA, CDS and custom track (*chIFITM* locus). *IFITM2* is missing from the "selected IFITMs" track because the annotation information that was available in the reference did not include the gene name. The height of the bars underneath the reads reflects relative expression levels for the genes, as does their colour: ranging from red (hot) to blue (cold). Quantitation was conducted over exons, assuming a directional reverse library and represents read count normalised for library depth and subsequently log transformed. Samples 13 and 21 are included for reference of level of basal expression of IFITMs in a WT DF-1 and KO DF-1 cell line (225), respectively.

5.3.1.3 *Establishement of the infection system for study of the innate immune response of PGC-derived fibroblasts to influenza infection*

Both KO and WT PGC-derived fibroblasts are finite adherent cell lines exhibiting typical fibroblast phenotype (**Figure 71**) when seeded on fibronectin coated plates. A range of concentrations from 0.1ug/cm^2 to 5ug/cm^2 were tested for their ability to aid cellular adherence and it was established that concentration as low as 0.1ug/cm^2 is sufficient to support a healthy morphology. Commonly for finite cell lines, their growth is arrested upon contact inhibition. They are approximately three times bigger in surface than DF-1 fibroblasts, evidenced by total number of cells at 100% confluency which consistently fluctuated around one third of that of DF-1 fibroblasts in the same plating conditions.

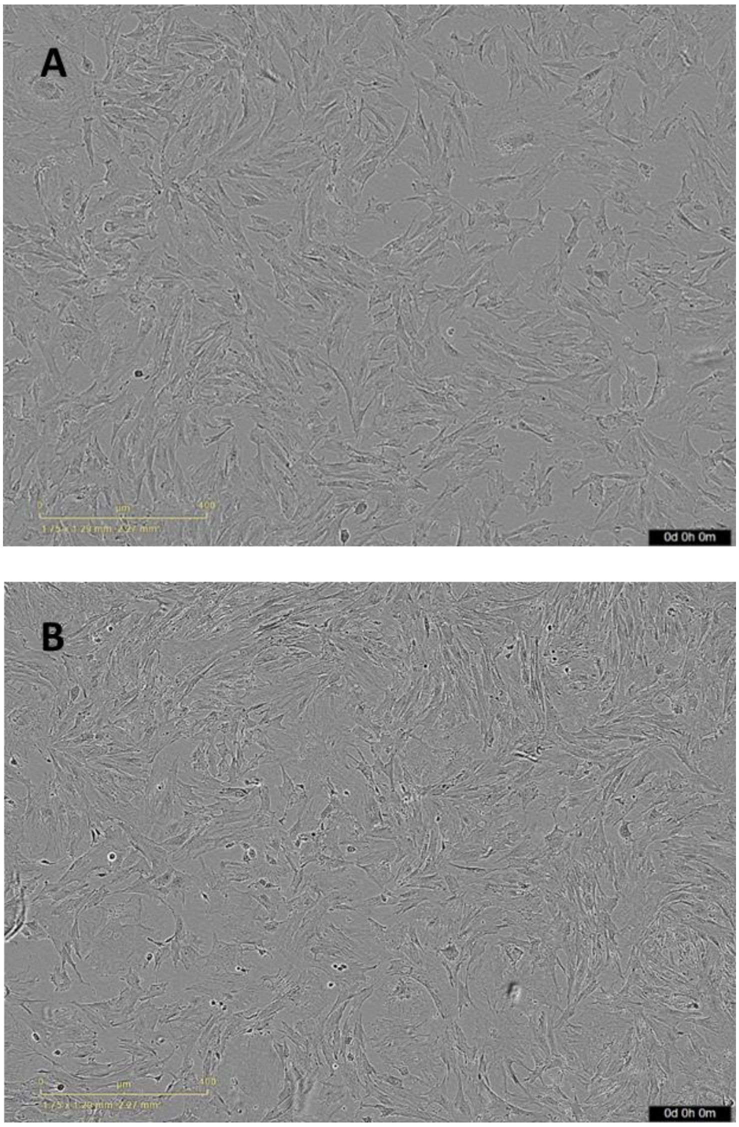


Figure 71 PGC-derived fibroblasts *A. WT B. chIFITM KO*

5.3.1.4 Growth kinetics of PGC-derived fibroblasts

Viral yield is influenced by many factors. Studies looking to quantify the generation of virus particles by infected cells frequently evaluate the productivity of the system on a big scale by studying parameters such as temperature, pressure, medium content and rate of replacement, oxygenation efficiencies, MOI, pH, etc. Ongoing optimisation efforts of biomanufacturing processes often use Design of Experiments (DOE) approaches to ensure maximal efficiency and cost reduction [196,197]. Briefly, DOE is a statistical framework that allows for simultaneous investigation of various parameters at various levels as opposed to single parameter at a time: a standard approach in academic settings. This enables the researcher to study the interactions among the variables to ensure the optimal output. However, on a smaller scale, parameters such as cell permissiveness, clonal diversity, host metabolic activity and inherent stochasticity associated with the early viral replication in the host have a greater impact on viral yield. Timm and Yin found that cell-to-cell differences can account for as much as 300-fold in virus production rates [198]. In this section I describe the approach taken to discern the direct effects of the knock-out on morphological level.

To ensure that the measurements of differential viral yields are not a result of different doubling times, a growth curve was constructed using a live cell imaging system (**Figure 72**). Both cell lines take approximately 24h to double, seeded at all tested densities, with those seeded at 10% never quite reaching the stationary phase. It would appear there needs to be a critical minimal concentration of neighbouring cells that facilitate favourable environment for doubling. Both WT and *chIFITM* KO PGC-derived fibroblasts exhibit optimal growth seeded at 25% confluency with longest time spent in log phase as compared to cells seeded at approximately 40% and 70%. Thus, the split ratio of ¼ was used throughout this work to bring the stress caused by trypsinisation to a minimum while doubling kinetics remain stable resulting in most predictable growth behaviour required for efficient execution of experiments.

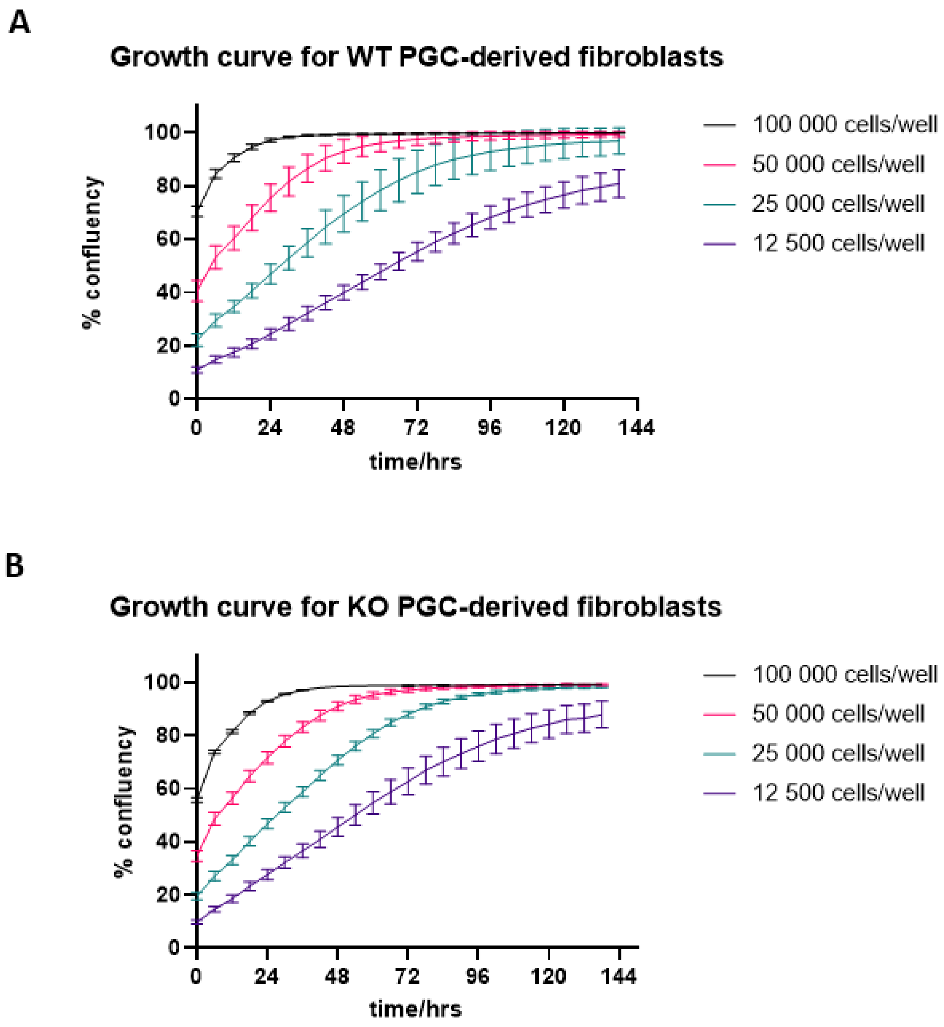


Figure 72 Growth curve for PGC-derived fibroblasts Cells were plated in 24-well plates in split ratio 1:5 and 1:3 and left in the IncuCyte incubator at 37°C, 5%CO2 for 5 days. Growth curve was constructed by measuring the expanding confluence area over time using in-built image analysis software IncuCyte S3 v.2018C. Plotted values are means of 6 replicate wells with error ±1 SD.

5.3.1.5 Setting up the infection system

To ensure a meaningful comparison of the two systems proposed in this work, the experimental conditions were same in both. Thus, the experimental system developed for DF-1 was used as a starting point. To investigate the effects of the IFITM knock-out on host defence system against influenza, firstly the ability of the virus to produce a robust infection was evaluated and secondly the ability of the host to sustain long-term viral production. Following parameters were optimised: virus strain, time post infection for transcriptomic studies, infection media composition and cellular permissiveness to the virus.

Influenza strains used in this work are dependent on exogenous proteases for HA cleavage, so the infection protocols include the addition of TPCK-trypsin to the infection medium. Titration showed that PGC-derived fibroblasts are highly sensitive to TPCK-trypsin and suffer adverse effects at all concentrations above 0.01 ug/ml (data not shown). This was a familiar behaviour, seen already in DF-1 fibroblasts, hence the same infection protocol was used (see 2.2.2.4).

A single-cycle infection study was set-up to evaluate the permissiveness of PGC-derived fibroblasts by flow cytometry. To establish the MOI level required to achieve a 100% infection rate, WT equivalents of DF-1 and PGC-derived fibroblasts were infected with H5N3 A/duck/Sing-Q/3/97 at MOI 1, 3 and 5 as indicated in 2.2.2.4 and collected for analysis at 6h post infection. WT DF-1 was included as a comparator cell line. Despite DF-1 data exhibiting some variability at MOI of 1, permissiveness of both cell lines to influenza was significant; more than 90% of cells were detected as infected, even at the lowest MOI tested (**Figure 73**).

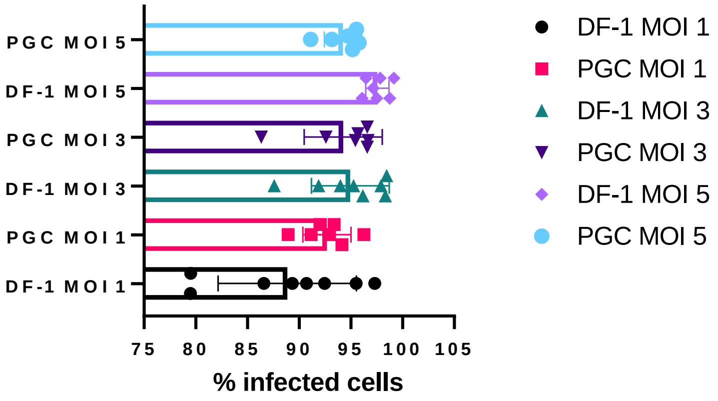


Figure 73 Permissiveness of PGC-derived fibroblasts to influenza Parental DF-1 and WT PGC-derived fibroblasts were infected with H5N3 A/duck/Sin-Q/3/97 at MOI 1, 3 and 5 for 6h. Cells were collected, fixed, and analysed by flow cytometry for influenza virus NP staining. Data show the percentage of infected cells collected during three independent experiments in duplicates. Bars represent ±1SD.

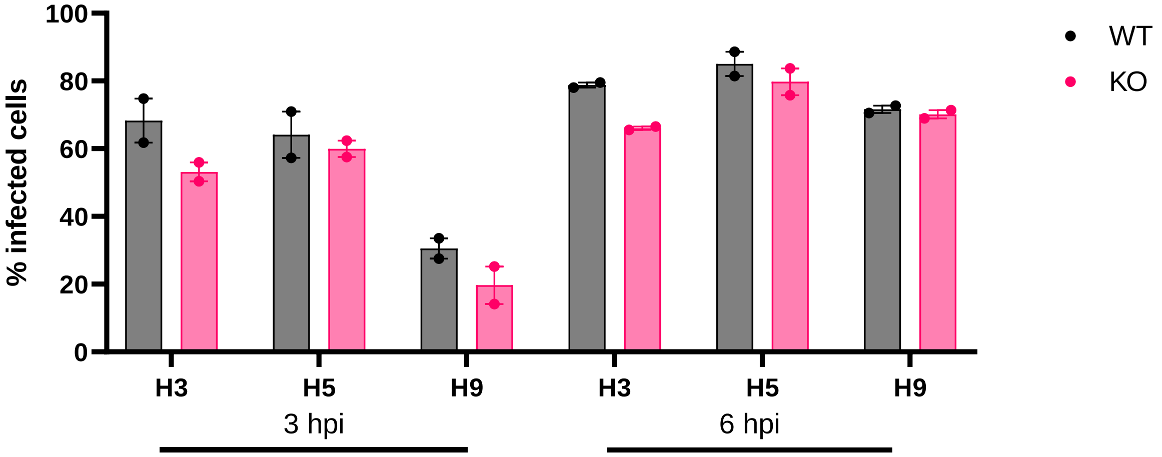


Figure 74 Comparison of permissiveness to influenza virus between WT and chIFITM KO PGC-derived fibroblasts WT and *chIFITM* KO PGC-derived fibroblasts were infected with A/Chicken/Pakistan/UDL-01/08 (H9N2), A/duck/Ukraine/63 (H3N8) and A/duck/Sing-Q/119/97 (H5N3), labelled H9, H3 and H5, at MOI 1. Cells were collected at 3h and 6h post infection (hpi), fixed, and analysed by flow cytometry for influenza virus NP staining. Experiment was conducted once in duplicates. Bars represent ±1SEM.

The capacity of WT and KO PGC-derived fibroblasts to sustain long-term infection was evaluated next. Suitable system would be capable of harbouring a robust viral replication, potentially with some moderate CPE that does not infringe on the host's capacity to sustain the infection until the viral growth reaches the plateau phase. Three strains of influenza of different subtypes of epidemiological relevance were considered: A/Chicken/Pakistan/UDL-01/08 (H9N2), A/duck/Ukraine/63 (H3N8) and A/duck/Sing-Q/119/97 (H5N3). All strains readily infected the PGC-derived fibroblasts at MOI 1 (**Figure 74**), as detected by anti-NP staining. In contrast to *chIFITM* KO DF-1 cell lines (**Figure 39**), *chIFITM* KO PGC-derived fibroblasts showed greater refraction to infection, for all time-points and all influenza strains tested (**Figure 74**).. The phenomenon was not very pronounced, but it was consistent. Similar phenotype was observed in LMH cell line, whose basal expression was 100-fold lower than that of DF-1 (personal communication with Dr. Ibrahim). Subsequent virus yield studies on the LMH cell line did not demonstrate any significant effect of the chIFITM knock-out on the virus yield opening the possibility the same may be the case with PGC-derived fibroblasts.

Following the confirmation of suitable permissiveness to influenza, PGC-derived fibroblasts were subjected to a multi-cycle infection study. The cells were infected with the aforementioned strains at MOI 0.01, applying the infection protocol outlined in 2.2.2.4 also used with DF-1 fibroblasts**.** The cells were monitored daily for CPE for 48h and the supernatant was collected at intervals of 8h. Cells exhibited changes in morphology typical of an infection, such as rounding and detachment from the flask surface (data not shown). Certain percentage of cells remained alive for the duration of the experiment with all influenza subtypes used, indicating moderate pathogenicity suitable for experimentation. However, the plaque assay results for the collected supernatant were negative. It is possible that the observed behaviour was due to an additional source of stress to the cells, rendering the host system non-functional, i.e. unable to yield virus.

To investigate if the amount of TPCK-trypsin used is excessive and to potentially mitigate the negative effects of it, two conditions were tested next: addition of 1% chicken serum vs no serum, and low vs high MOI (0.001 and 1, respectively). Adding chicken serum might seem counterintuitive given the variety of protease inhibitors found in blood serums. However, the Genetics and Genomics group at the Pirbright Institute previously had some success with this strategy. Anecdotal evidence suggests it is possible to strike a balance between supplying enough growth factors enabling more prolific cell expansion while limiting the number of inhibitors that would inactivate all available TPCK trypsin. Cells were infected with H5N3 strain only and the concentration of the TPCK-trypsin in the infection medium was reduced ten-fold to 0.01 ug/ml. Supernatant was collected at 24h time intervals and viral titres were measured by plaque assay (**Figure 75**).

Given the virus is capable of successfully infecting the cells, as shown by flow cytometry in **Figure 73**, it was clear the virus lifecycle is obstructed at a later step. I hypothesised that the low titres of the H5N3 influenza virus could be attributed to the low levels of TPCK-trypsin available to activate the released virus particles. To make testing higher concentrations of TPCK-trypsin, i.e. 0.1%, possible, 0.1% chicken serum was added to the medium to mitigate the CPE. Results demonstrated that H5N3 replicates very poorly in PGC-derived fibroblasts, reaching titres between 10^2 and 10^3 pfu/ml, regardless of the MOI applied (**Figure 75**). Addition of the chicken serum to the WT PGC-derived fibroblasts resulted in the absence of detectable virus entirely, presumably due to the deactivation of the available TPCK-trypsin required for the HA cleavage. None of the conditions tested produced a dramatic difference in viral titre, nor between the WT and the KO PGC-derived fibroblasts, nor between the low and high MOI.

To investigate if this behaviour is strain specific, a human, mouse-adapted influenza strain PR8 (A/Puerto Rico/8/1934 (H1N1)) was tested. Concentration of chicken serum was kept to 0.1% to maximally limit the proportion of inactivated TPCK-trypsin. The PR8 stock was sourced from the Influenza group at the Pirbright Institute and expanded in cell culture of MDCKs according to the protocol described in section 2.2.2.2. As previously, 80-90% confluent layers of WT and KO PGC-derived fibroblasts were infected with low and high (i.e. 0.0001 and 1) MOI, in presence or absence of 0.1% chicken serum in the infection medium containing 0.1% TPCK-trypsin. Supernatant was collected at various time points during the multi-cycle infection experiment and frozen to be ultimately titred using the plaque assay.

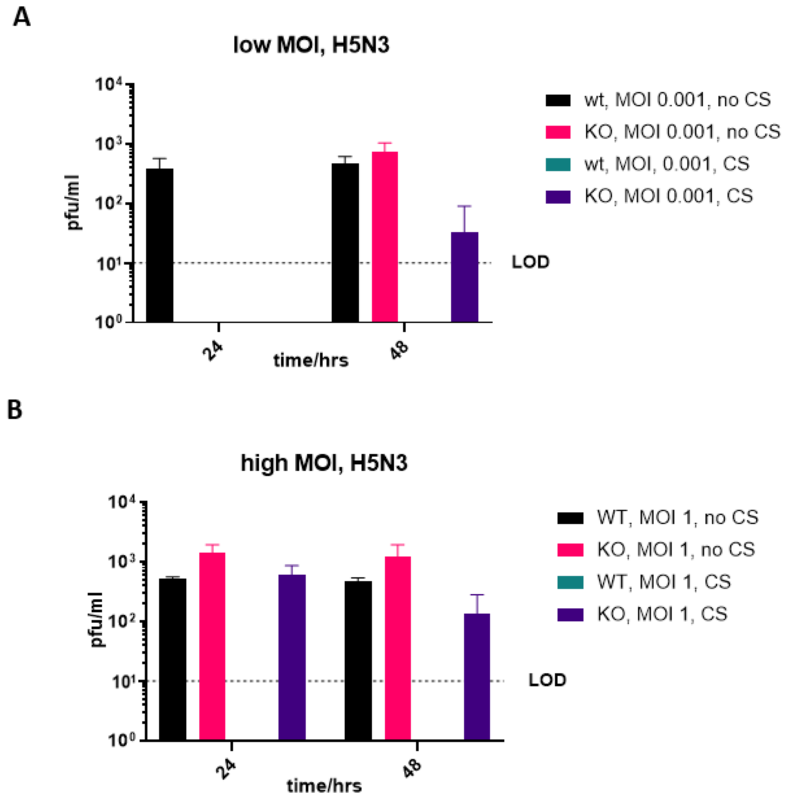


Figure 75 Infection of PGC-derived fibroblasts with H5N3 influenza virus WT and KO PGC-derived fibroblasts were infected with H5N3 A/duck/Sing-Q/3/97 at multiplicity of infection (MOI) 0.001 (**A**) and 1 (**B**) in presence of 0.1% TPCK-trypsin for 2 days, with and without the addition of 0.1% chicken serum (CS) in the infection medium. Virus titres in the supernatant were detected by plaque assay. Experiment was conducted once in triplicates. The bars show ±1SEM. The dotted line indicates the limit of detection (LOD).

High viral titres, in the range between 10^5 and 10^6 pfu/ml, were achieved with both MOI used (**Figure 76**), confirming the results previously observed with PGC-derived fibroblasts by Long *et al* [199]. With high MOI, the system was immediately saturated and no increase in viral titre was observed for the duration of the experiment. In the case of the low MOI, the stationary phase of the growth curve was reached in 48h. Addition of the chicken serum does not seem to improve the virus yield, though it is possible that in the case of high viral titres, the resolution afforded by the plaque assay is not sufficiently high to detect marginal improvements.

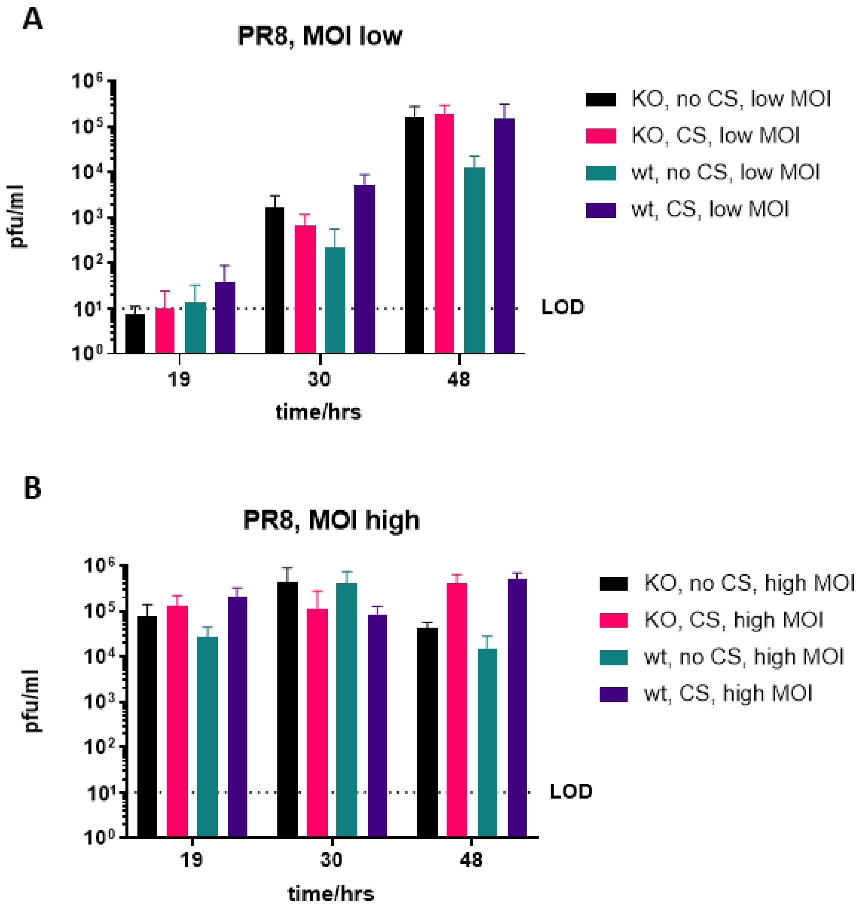


Figure 76 Infection of PGC-derived fibroblasts with PR8 influenza virus WT and KO PGC-derived fibroblasts were infected with H1N1 PR8 influenza virus at multiplicity of infection (MOI) 1 and 0.001 in presence of 0.1% TPCK-trypsin for 2 days, with and without the addition of chicken serum (CS). Virus yield was measured by plaque assay. Experiment was conducted once in triplicates. The bars show ±1SEM. The dotted line indicates the limit of detection (LOD).

Taken together, all the results presented here show that the addition of chicken serum does not have any added benefit thus it was not considered further. H5N3 subtype of influenza was selected as the model virus due to its moderate pathogenicity which allows for monitoring of infection that is progressing at a rate conducive to its investigation. This setup also maintains continuity with the work conducted on DF-1 cell lines and facilitates comparability. As PGC-derived fibroblasts showed great sensitivity to TPCK-trypsin, it was omitted from the medium as was the chicken serum. This approach provided a two-pronged benefit: it eliminated the need to mitigate negative effects of TPCK-trypsin (via addition of serum or otherwise) as well as ensured that all the stress the cells are experiencing has a single origin: viral infection.

5.3.2 Infection of PGC-derived fibroblasts with H5N3 influenza virus

5.3.2.1 Virus stock calibration

To secure a sufficient quantity of A/duck/Sing-Q/119/97 (H5N3) virus stock for the infection studies, a new batch was produced in embryonated chicken eggs using methods described in 2.2.2.1. Considering retrospectively, the old stock of virus was not back titrated during experiments and likely was no longer MOI of 1. Subsequently, upon producing the new stock, the titration of the old and new stock in two independent plaque assays, conducted in duplicates, indicated that the titre of the new stock is three times higher than the old stock (data not shown).To validate the results, DF-1 parental cell line was infected with both stocks in a combination of MOIs: old stock at MOI 1 and new stock at MOIs 0.1, 0.3, 1 and 3. Flow cytometry results showed that equivalent number of infected cells is achieved by using the old stock at MOI 1 as the new at MOI 0.3 (**Figure 77**). All the calculations for subsequent infection experiments were adjusted accordingly to maintain comparability of results.

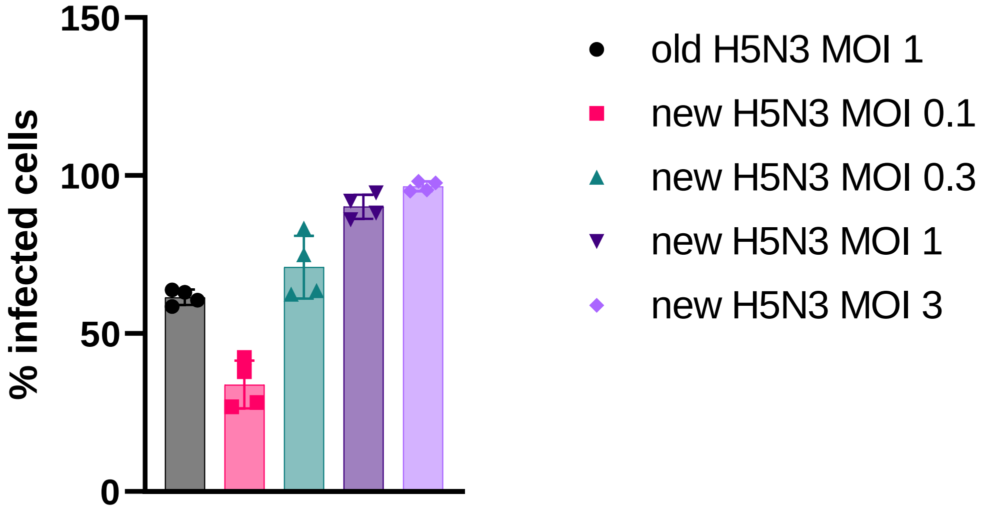


Figure 77 Comparison of the old and new stock of H5N3 A/duck/Sing-Q/119/97 WT DF-1 parental cell line was infected with the old and new stock of A/duck/Sing-Q/119/97 (H5N3) influenza virus. Plaque assay indicated the old stock of virus lost 65% of its potency compared to the new one, thus MOIs used for confirmation were: MOI 1 for the old stock and 0.1, 0.3, 1 and 3 of the new stock. Cells were collected at 6h post infection (hpi) and analysed by flow cytometry for influenza virus NP staining. Experiment was conducted twice in duplicates. Bars represent ±1SD.

By way of additional validation, an RT-qPCR was conducted to investigate the transcriptomic response of the cells to different amount of virus. RNA was extracted (as per section 2.2.3.5) from parental DF-1 cells infected with following amounts of H5N3 virus: old stock in MOI of 1, and new stock in MOI of 0.1, 0.3 and 1. Quantitative RT-PCR was performed as described in 2.2.3.9 with primers and probes listed in **Table 2** and **Table 3**. Ct values were averaged, and those values were used in the fold change calculation. Experiment was conducted only once, disallowing the testing for statistical significance. Nevertheless, similar levels of expression were induced by the old stock in MOI 1 and the new stock MOI 0.3, in agreement with the flow cytometry and plaque assay results. The equivalent infectious dose of MOI 1 of the old viral stock is 0.3 MOI of the new stock.

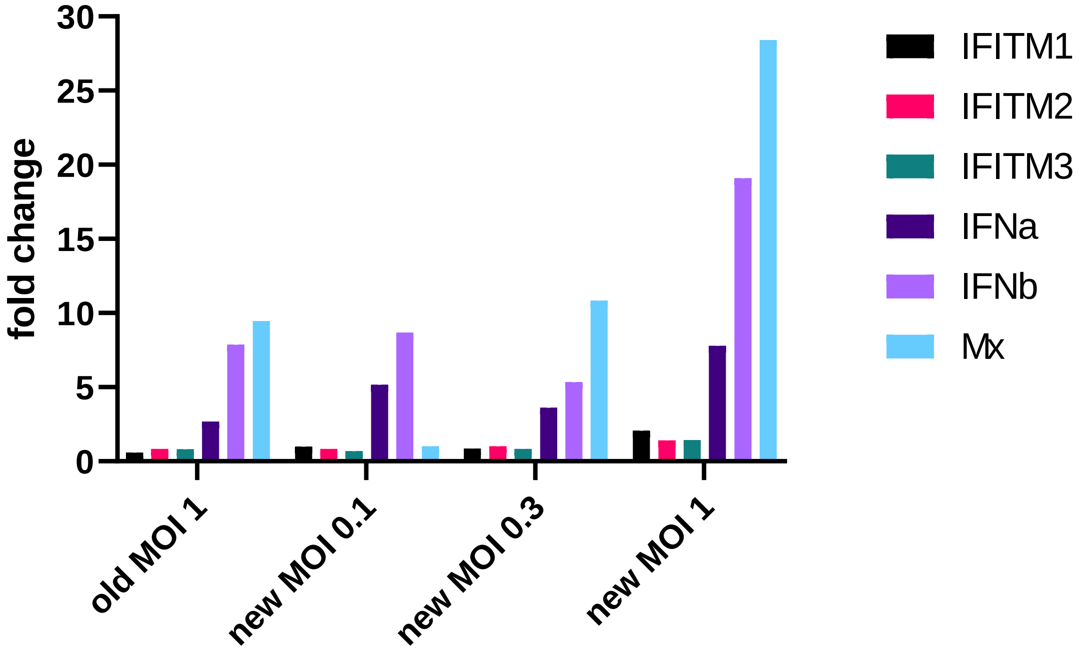


Figure 78 Expression levels for genes of interest induced by different virus stocks RNA was extracted from DF-1 cells infected with varying amounts of new and old stock of H5N3 influenza virus, as indicated in the figure. cDNA was synthesised and RT-qPCR conducted for genes of interest to compare levels of expression induced by different amounts of virus. Experiment was conducted once, and Ct values of technical triplicates were averaged for calculation of fold change in expression between infected and mock wells.

5.3.2.2 PGC-derived fibroblasts are readily infected by influenza virus

Direct negative effects of IFITMs on the progression of the viral life cycle have been extensively demonstrated in various systems [40,44,164,200]. Even with many mechanistic details still obscured, the consensus is that they hinder the release of viral genetic material into the cytoplasm. Thus, the overarching hypothesis of this work is that in the absence of IFITM genes viral replication within the cell increases. Previously we showed that WT PGC-derived fibroblasts are slightly more permissive to all tested influenza subtypes at both 3h and 6h post infection when percentage of infected cells in measured (**Figure 74**). To investigate what effect this would have on viral replication, differential quantification of the mean fluorescence intensity was conducted as a proxy. In addition to that, to investigate the transcriptomic response to influenza infection in the absence of *chIFITM* genes, three independent infection experiments were set up. WT and KO PGC-derived fibroblasts were grown to

80-90% confluency and infected with H5N3 A/duck/Sing-Q/119/97 at the equivalent MOI as the DF-1. Half of the cells were collected for RNA-Seq (using the protocol 2.2.5.2) and the other half were processed for flow cytometry to confirm the infection status as previously (see section 2.2.3.11).

With no statistically significant difference in permissiveness to influenza between WT and KO PGC-derived fibroblasts in all tested conditions, WT PGC-derived fibroblasts seem to be slightly more conducive to viral replication than the KO derivative, as judged by the MFI of the anti-NP staining (**Figure 79**, panel D). Experiment shown in panels A and C was conducted in triplicates after the phenomenon was observed in the statistically weaker experiment where 3 different subtypes of the influenza virus (H3, H5, H9) and two different time points (3hpi and 6hpi) (panels B and D) were tested. The effect that WT PGC-derived fibroblast is more permissive than the chIFITM KO PGC-derived fibroblast is observable regardless of the strain or time point tested (panel B) or the MOI used (panel A). The difference between the cell lines in permissiveness is small and not statistically significant, but interestingly the difference MFI is observable, especially at 6h post infection (panel C and D). The 10x difference in MOI applied also does not seem to alter the effect and we can see that regardless of the MOI, the MFI for the WT is greater (panel C).

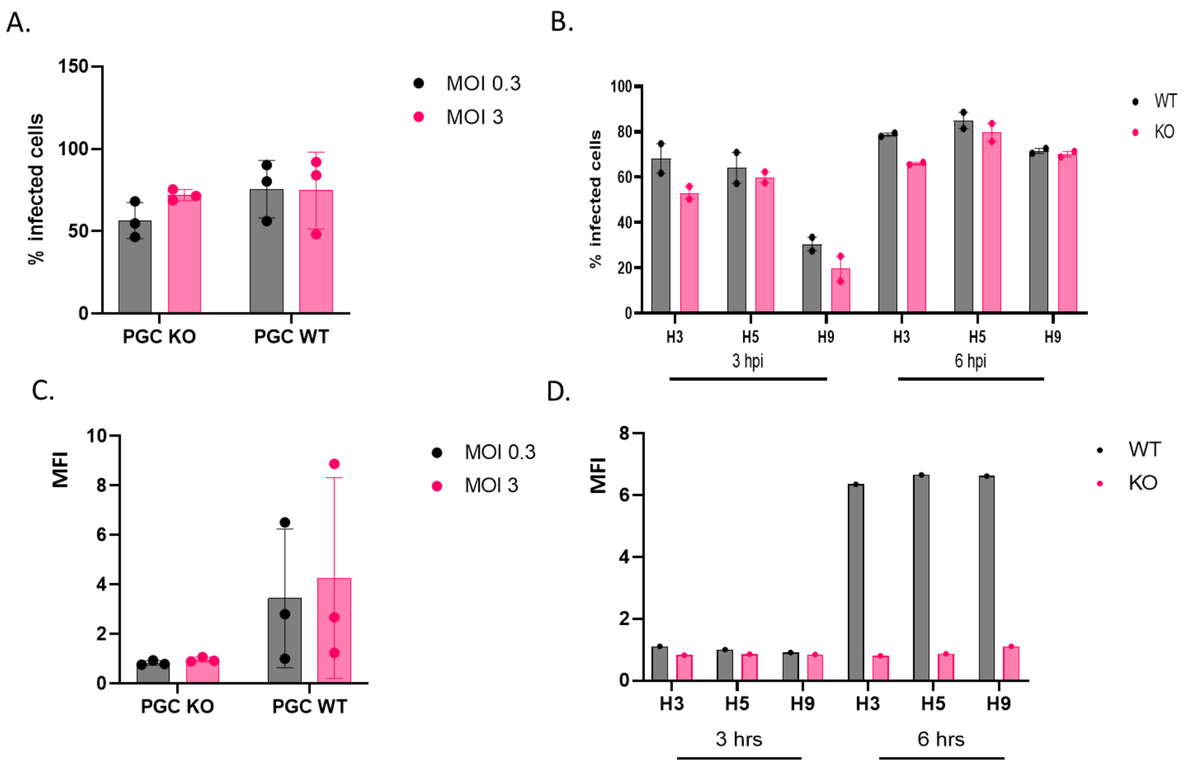


Figure 79 Knock-out PGC-derived fibroblasts are more refractive to influenza infection than the wild type equivalent WT and *chIFITM* KO PGC-derived fibroblasts were infected with influenza and stained with anti-NP antibody to detect the virus by flow cytometry. Experiments in A. and C. were conducted in biological triplicates and the bars indicate ±1SD. Cells were infected with H5 subtype and collected at 6 hrs post infection. Experiment in B. and D. was conducted once in technical

duplicates and the bars represent ±1SEM. Cells were infected with H3, H5 and H9 influenza subtypes at MOI 1 and collected at 3 and 6 hrs post infection. Graphs in A. and B. show percentage of infected cells as a measure of permissiveness, and graphs in C. and D. show normalised mean fluorescence intensity as a proxy for viral replication.

Overall, the KO PGC-derived fibroblasts were more refractory to influenza infection, opening questions about what mechanism is producing this behaviour and whether it is caused by an inadvertent downstream effect of the knock-out.

5.3.3 Transcriptomic analysis of acutely infected PGC-derived fibroblasts

5.3.3.1 Whole transcriptome sequencing

Whole transcriptome sequencing for PGC-derived fibroblasts was conducted simultaneously with KO DF-1 cells thus the processing steps were identical to ones described in section 4.3.1.4. Quality assessments conducted using fastQC confirmed a very high read quality, so no trimming was performed prior to alignment with STAR [85]. As with the DF-1 data set, alignment to only chicken reference genome GRCg6a did not reach >90% so the same bespoke genome reference described in 4.3.1.4 was used to improve alignment. It consisted of the chicken reference genome and the corresponding H5N3 influenza sequences (**Table 5**). Aligning the raw reads to the concatenated sequences resulted in all samples having more than 90% of reads successfully mapped (**Figure 80**, panel B). Note that when the RNA-Seq reads were aligned only to the chicken reference genome GRCg6a the lowest percentage was <65% in contrast to the DF-1 data where that percentage was around 40% (**Figure 41**, panel A), indicating less viral replication is taking place in the PGC-derived fibroblasts compared to DF-1 cells.

As with the DF-1 data, PGC data set was also processed with SeqMonk [140] to assess the library size and the mapping quality (**Figure 80**). Results of the QC analysis confirmed the mapping was successful and no contamination was detected. The smallest library belonged to an uninfected WT PGC sample and was approximately 40% of the size of the biggest data set, also an uninfected WT PGC-derived fibroblast.

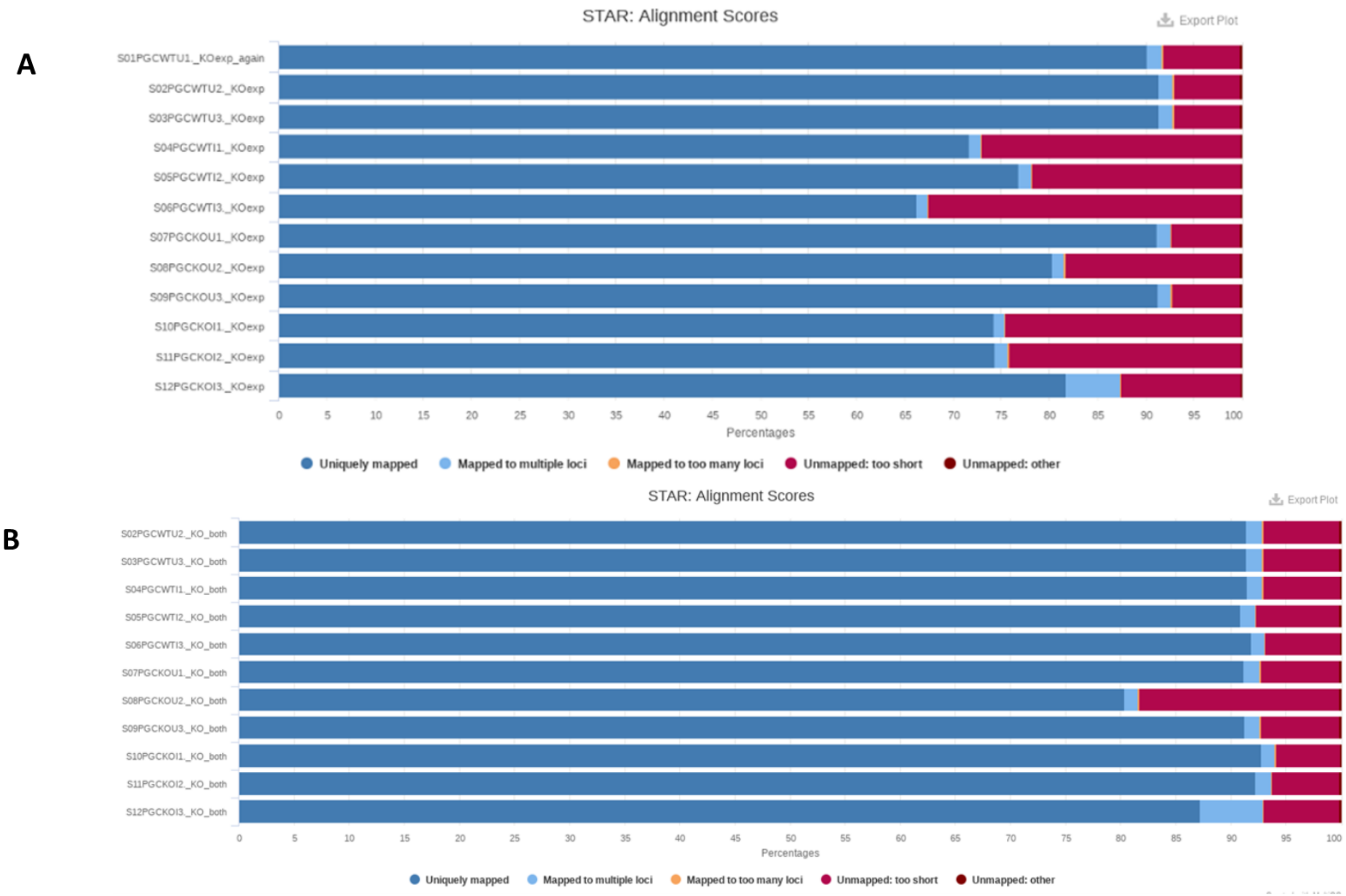


Figure 80 Mapping quality for PGC-derived data set *RNA-Seq data for PGC-derived fibroblasts were aligned to the chicken reference genome GRCg6ae (panel A) using STAR and to the bespoke genome reference which consisted of chicken genome to which influenza sequences were concatenated in panel B. Percentages of sequences mapped to the reference are plotted on the x-axis. Uniquely mapped reads are represented by the blue block and unmapped by the red. Summarisation of mapping reports generated by STAR was produced by MultiQC.*

Figure 81 Mapping QC for the PGC-derived fibroblast RNA-seq data *WT parental and KO DF-1 clonal cell lines 225, 250 and 600 were infected with the influenza virus. 6h post infection RNA was extracted and subjected to RNA-Seq. Reads were mapped to GRCg6a reference genome using STAR. The resulting BAM files were assessed using SeqMonk RNA-Seq QC pipeline.*

Exploratory data data analysis

To render the data homoskedastic *vst* and *rlog* transformations were tested (**Figure 82**). Both transformations worked well for this data set. *Vst* performed slightly better and was subsequently used to prepare the data for the PCA.

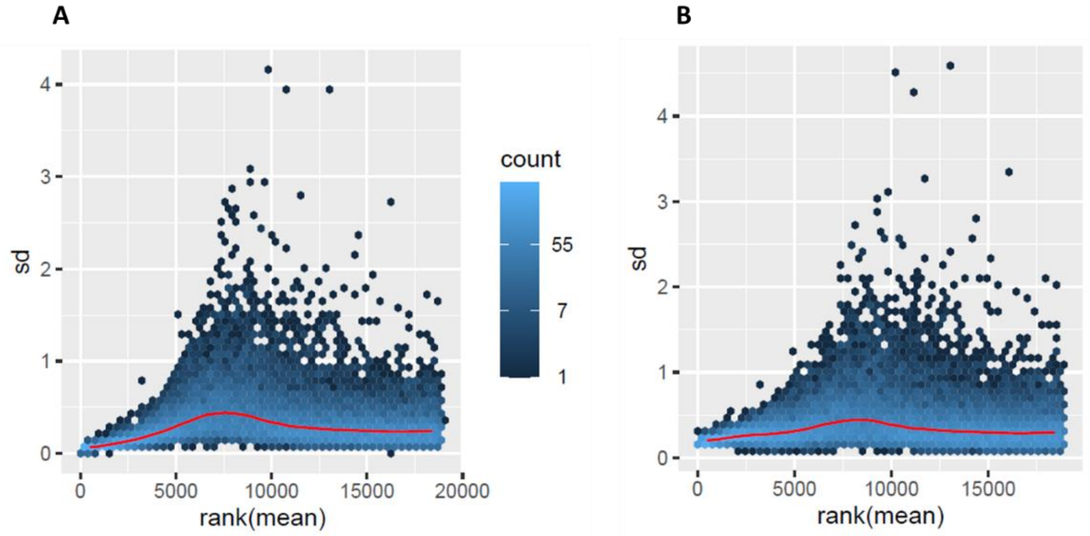


Figure 82 Transformation of the PGC-derived fibroblast data RNA-Seq data were normalised and transformed using *rlog* (panel A) and *vst* (panel B) functions in DESeq2. Mean values of gene counts across all sample groups are plotted on x-axis and standard deviation on y-axis. Red line denotes the trend of mean-variance dependency.

PCA was produced with *DESeq2* package and plotted with *ggplot* in R (**Figure 83**). Analaysis showed that the *chIFITM KO* has much greater effect on the transcription than the infection. PC1 separated WT and *chIFITM* KO PGC-derived fibroblasts and explained 70% of variance. 17% of variability in the data set was explained by the PC2 which separated infected from uninfected samples. Clear separation of the four sample groups was also produced by plotting sample distance matirx (not shown) and no outliers were detected.

To look into the variability described by PC1 and PC2 top 20% of loadings of each were plotted. Variables contained within PC1 were *CAPN2, ANXA6, SYDE1, BKJ, LMX1B, CACFD1, FBLN1, FIBIN, FEZ1, MEOX2 and SELENBP1* and for PC2 are *G0S2, ID2, ID1, CCN2, RASL11B, RIMS2, ENSGALG00000047514, NEDD9, CYR61, FGFRL1, WNT9A, ALS2L, CPED1, PACSIN3, ENSGALG00000007080, RRAD, IL11RA, SHROOM2, EGLN3, MC5R, PTH2R, NARF, KCTD12B, PPP1R3B, FZD8, LMOD1, NGFR, MCM5, FSCN1, RNF145, PIK3R5, CX3CL1, LPL, DDIT3, CCL4, LOC107049603, PCDH8, DUSP6, EGR3, SERPINB2, HIST1H111L, AQP1, FSTL3, PTGS2, PHLDA2, NR4A2, PTCH2, ETS2, LOC101748577, NAB2, CYP1A1, FRMD4B, IL1R2, BCL6, MYB, EVA1A, HMGA2, AGPAT1, FAS, and LAT2. (***Figure 84**).

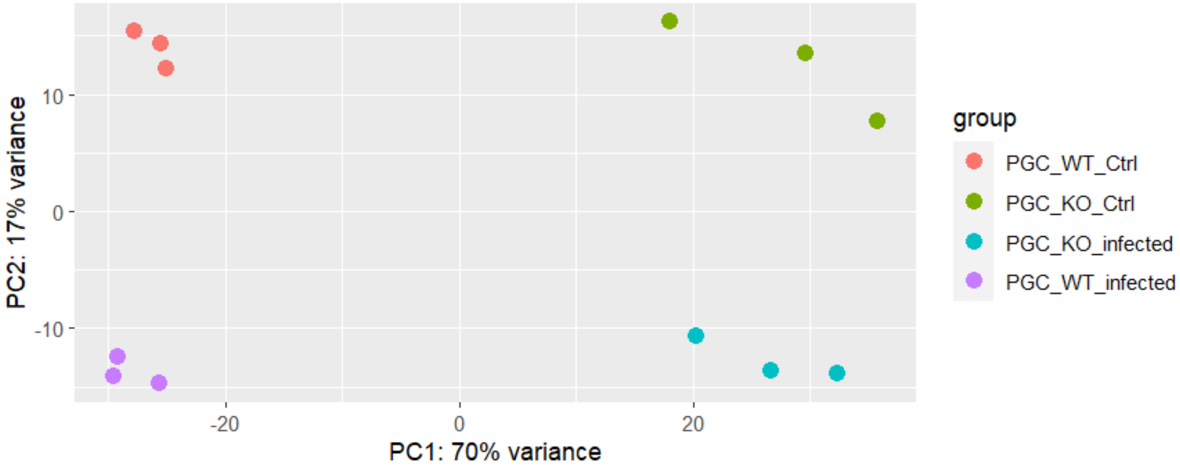


Figure 83 Genotype is a greater cause of change in transcription profile than condition in PGC-derived fibroblasts PCA was conducted on *vst* transformed data in DESeq2 package in R. Biplot shows PC1 and PC2 plotted on the X and Y axes respectively. Datapoints are representing samples belonging to samples groups denoted in the legend.

To look further into differences in the transcriptomic profile of the WT and *chIFITM* KO PGC-derived fibroblasts, the most variable genes were extracted. Variance was calculated from vst transformed values for every gene and the fifty most variable genes were clustered using Euclidean distances (**Figure 85**). Two clusters emerged. Cluster characterised by upregulation in the knock-out cell line contained following genes: *BKJ, SHOX, ITGA4, SFRP2, CA9, SYNC, CLMN, CRIP1, ANXA6, EML3, ENSGALG00000026613, CAPN2, SYDE1, TAPBP, PTPRVP, LMX1B and CACFD1*. The expression levels of these genes were more similar among replicates (experiments conducted on different days) regardless of the condition, while in the WT PGC-derived fibroblast clustering grouped the samples by condition, rather than the replicate. Cluster containing *IFITM5, ALDH1L2, PHGDH, PCDH7, OSTN, FBN3, NTN1, MT3, WNT5B, LOC418108, ENSGALG00000037349, EDN3, AAED1, GLRX, ZIC3, ENSGALG00000042045, FBLN1, FIBIN, SELENBP1, HAND2, FEZ1, MEOX2, PDK4, LOC101749287, NRN, ICOS, FAT3, IGFBP2, GCHFR, C1QTNF5* and *PDGFD* were down regulated in the KO PGC-derived fibroblasts.

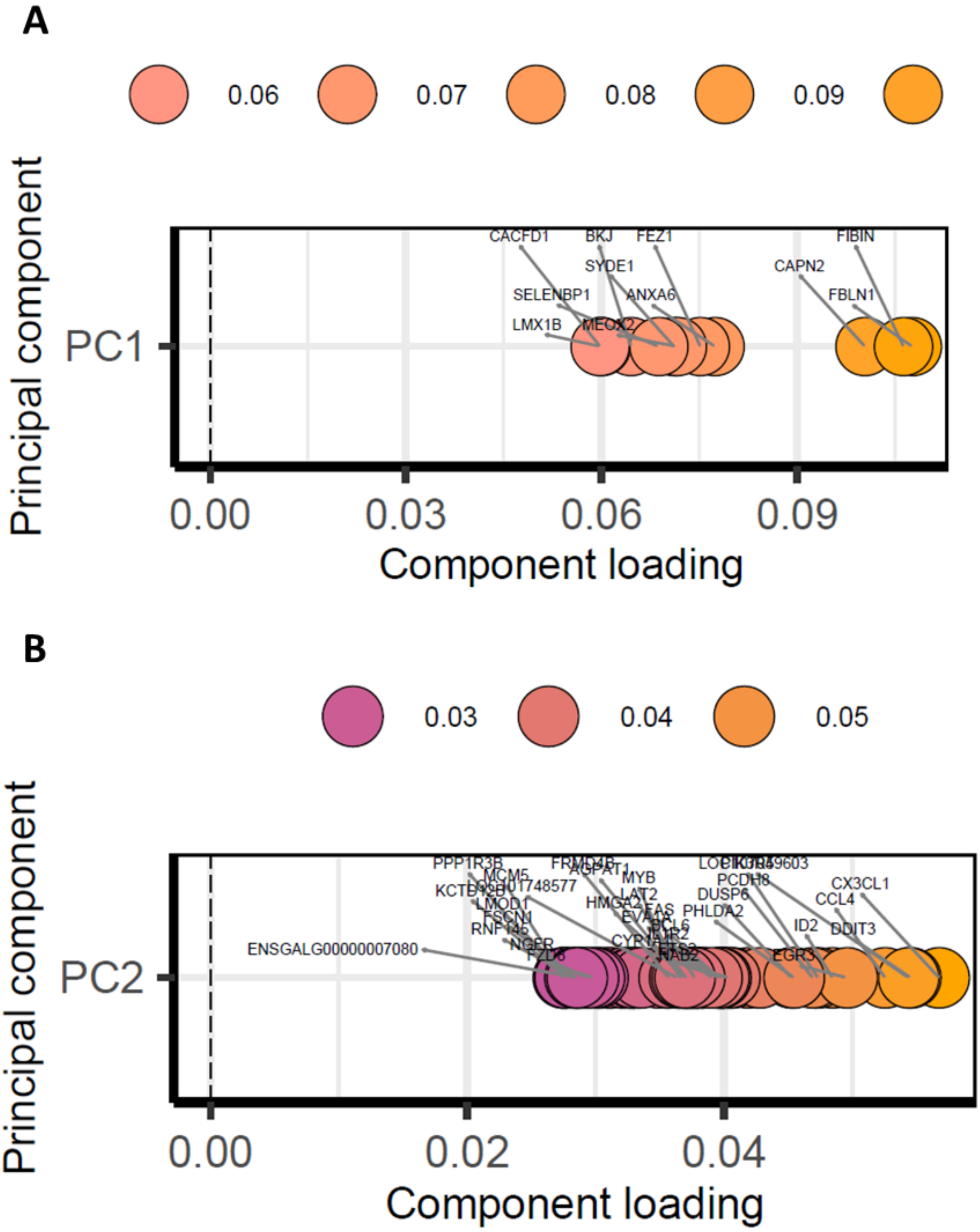


Figure 84 Loadings for PC1 and PC2 Top 20% of loadings for PC1 and PC2 are plotted in panel A and B respectively. Retained variables for PC1 are *CAPN2, ANXA6, SYDE1, BKJ, LMX1B, CACFD1, FBLN1, FIBIN, FEZ1, MEOX2, SELENBP1* and for PC2 are *G0S2, ID2, ID1, CCN2, RASL11B, RIMS2, ENSGALG00000047514, NEDD9, CYR61, FGFRL1, WNT9A, ALS2L, CPED1, PACSIN3, ENSGALG00000007080, RRAD, IL11RA, SHROOM2, EGLN3, MC5R, PTH2R, NARF, KCTD12B, PPP1R3B, FZD8, LMOD1, NGFR, MCM5, FSCN1, RNF145, PIK3R5, CX3CL1, LPL, DDIT3, CCL4, LOC107049603, PCDH8, DUSP6, EGR3, SERPINB2, HIST1H111L, AQP1, FSTL3, PTGS2, PHLDA2, NR4A2, PTCH2, ETS2, LOC101748577, NAB2, CYP1A1, FRMD4B, IL1R2, BCL6, MYB, EVA1A, HMGA2, AGPAT1, FAS, LAT2.*

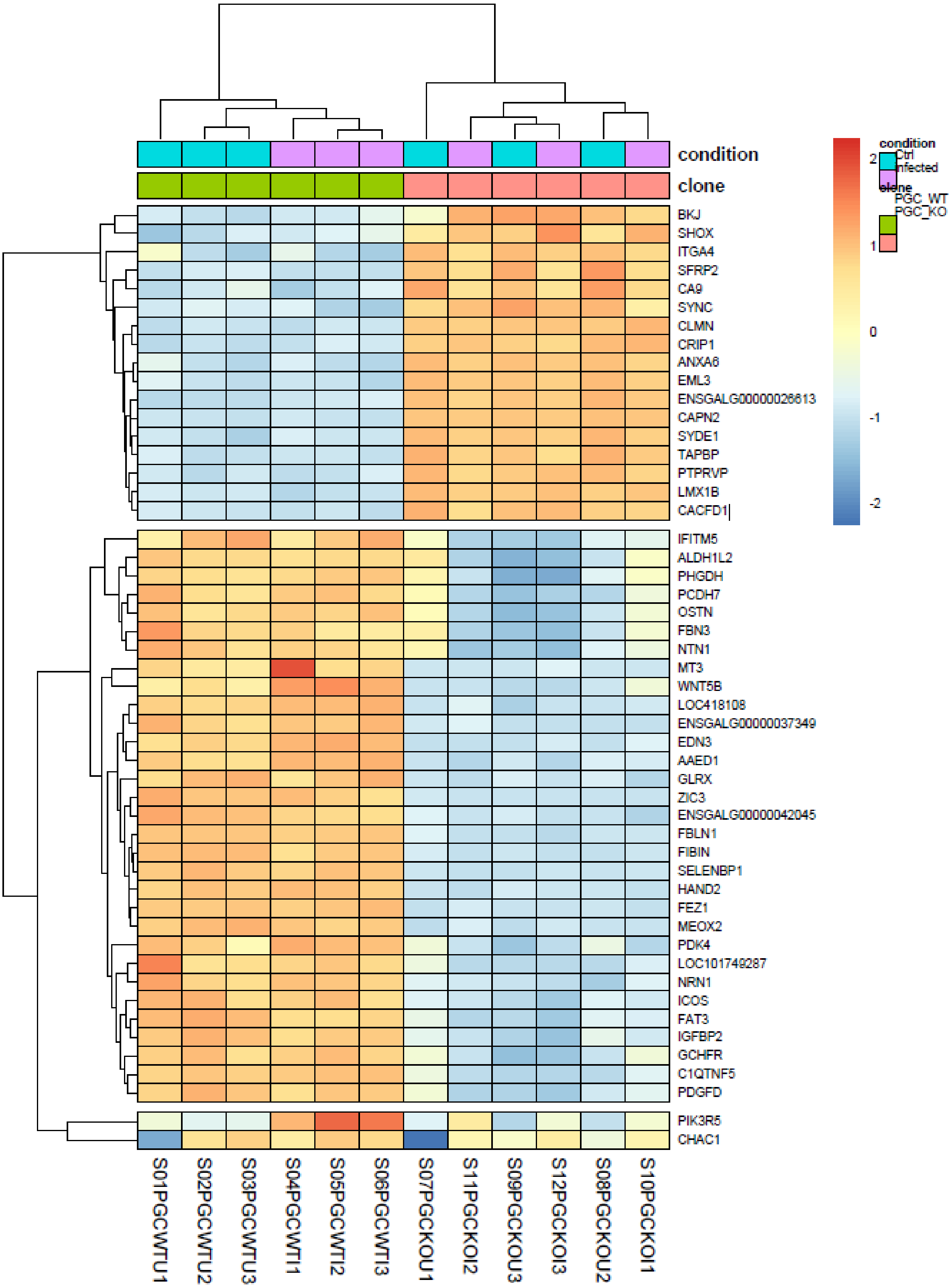


Figure 85 Fifty most "differently" expressed genes in PGC-derived fibroblast dataset Variance was calculated for *vst* values and all the genes were sorted by variance in increasing manner. Top 50 genes were Z-scored and clustered using Euclidean distances.

5.3.3.2 Evaluation of the impact of chIFITM KO on the transcriptome in PGC-derived fibroblasts

Differential expression

RNA-Seq experiment consisted of 4 different sample groups in triplicates: two genotypes (WT and KO) and two conditions (infected and uninfected). PCA showed that both parameters influence the variability in gene expression significantly and must be included in the statistical modelling. To enable comparison of all desired contrasts, a variable “group” was created and a model without intercept was employed to conduct the Wald test, i.e. ~ 0 + group. DESeq2 [94] was used for the analysis; FDR of 0.05 and $\log_2$ fold change of 1 were chosen as the cut off for statistical significance. As before, *ashr* [141] was used to adjust the effect sizes and the Benjamini-Hochberg method [201] for multiple testing correction.

Firstly, the effect of the *chIFITM* knock-out on the transcriptome was evaluated by comparing basal levels of gene expression in the WT and the KO PGC-derived fibroblasts in uninfected condition.

1466 genes were discarded from the analysis due to low count and 19 genes were outliers (0.1%). 1204 (6.4%) genes were upregulated and 1545 (8.7%) were downregulated at the designated p-value of 0.05. Differentially expressed genes that were statistically significant were sorted by fold change in expression and the 100 most upregulated and the most downregulated were further analysed by clustering; the results are presented in **Figure 86** and **Figure 87**, respectively.

The TPM values were Z-scored and the Euclidean distances were used to cluster the genes to relate their levels of expression across all groups in the experiment. Heatmaps revealed very noticeable intergroup biological variability. Group mean and the standard deviation were frequently skewed by one or two extreme values. Genes that are upregulated in the KO PGC-derived fibroblasts are not showing any coordinated patterns of expression that could be uncovered by clustering. On the other hand, downregulated genes did organise into clusters of genes that are potentially co-regulated.

ACE, MFSD4A, RAB11B, VWA1, MARCH2, HS3ST3A1, ENSGALG00000000314, RPL36, PANX2, TUB4A, TRAF1, MVB12B, PHF19, RALGPS1, REN, LOC416147, NDUFA7, SPPL2C, RDM1, PTPRVP, ETNK2, MFSD12, NKAIN1, RAB7B, IL12B, CUTAL, LOC419390, MAPT, RPL23, KLHDC8A, NUDCD2, SLC25A37, ADRA1B, ST6GALNAC1, TEK and *ENSGALG00000001856* are more highly expressed in the KO PGC-derived fibroblasts when compared across all sample groups (**Figure 87**).

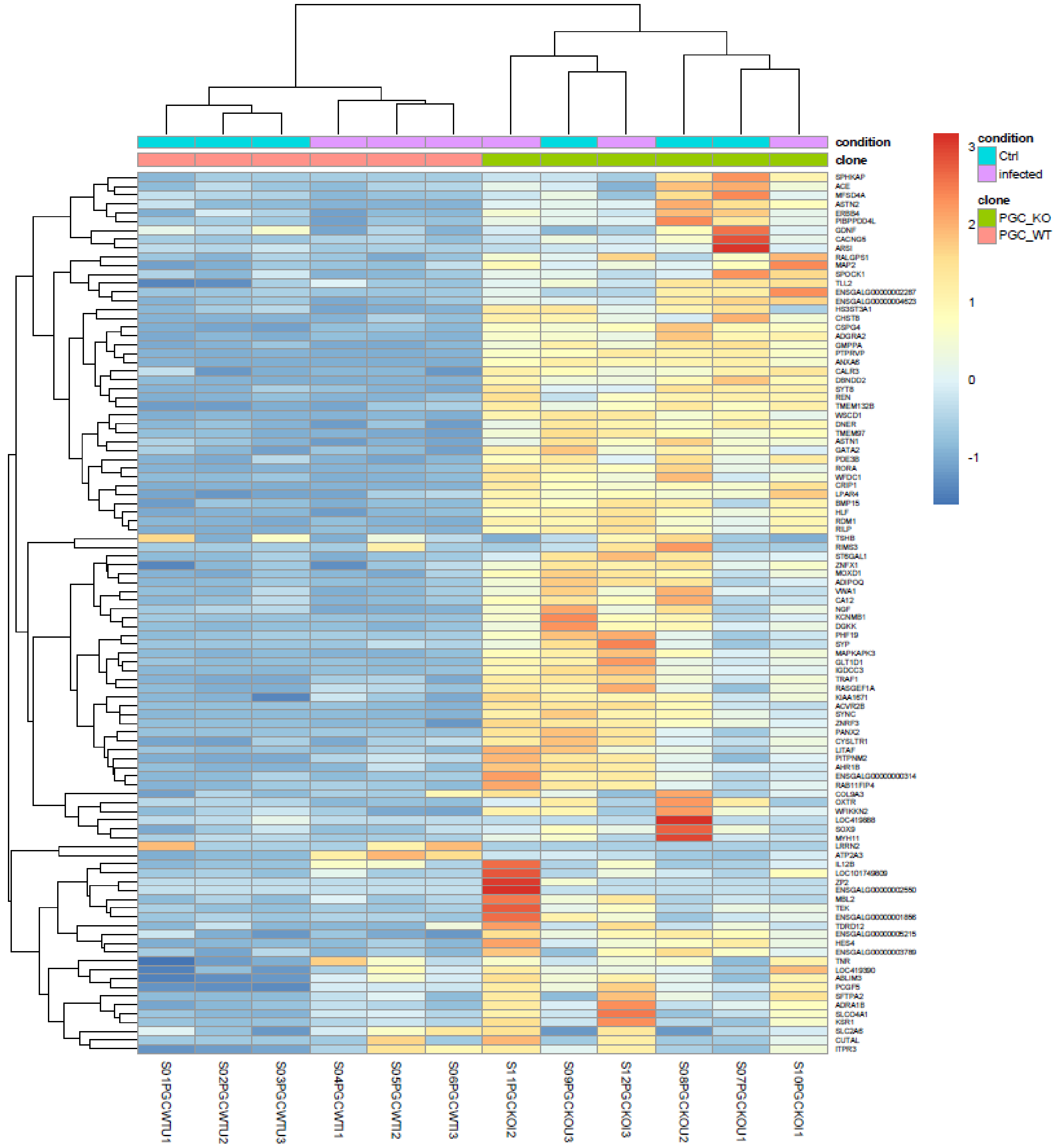


Figure 86 100 most upregulated genes in differential expression analysis between WT and chIFITM KO PGC-derived fibroblasts Differentially expressed genes that were statistically significant between the WT and KO PGC-derived fibroblasts in uninfected condition were sorted by increasing fold change. TPM values of the 100 top genes were Z-scored and clustered with *pheatmap* package in R using Euclidean distance as the metric of similarity.

ADAT1, SSU72, ATP2A3, EFNA5, PCNX1, RSPRY1, SPIRE2, TERF2IP, TERF2, CITED4, PKP1, MEOX1, HERPUD1, NLRC5, GAS8, SPR, CLDN5, NOB1, POLR2C, EXOC3L1, CYB5B, KARS, ST14, PRNP, CPNE2, ADAMTS8, TMEM52, RASA4B, LRRN2, MAN1C1, PSMD7, TENM2, TCF25, GBP1, ENSGALG00000000044, COQ9, SLC16A4, FAM132A, NR6A1, STX12, ZFHX3, DHX38, ZNF319, TNNC1, RNF122, NEO1, RANBP10, SELENBP1, SLIT3, PLXNA2, JAM3, ZNF276, PHLPP2, NEK6, GFOD2, SLC27A6, MYLKSML, ROGDI, AP1G1, UTS2R, EPB41L4A, DUSP26 and *GUCY2D* show a clearer pattern of down-regulation in the KO regardless of the condition (**Figure 87**).

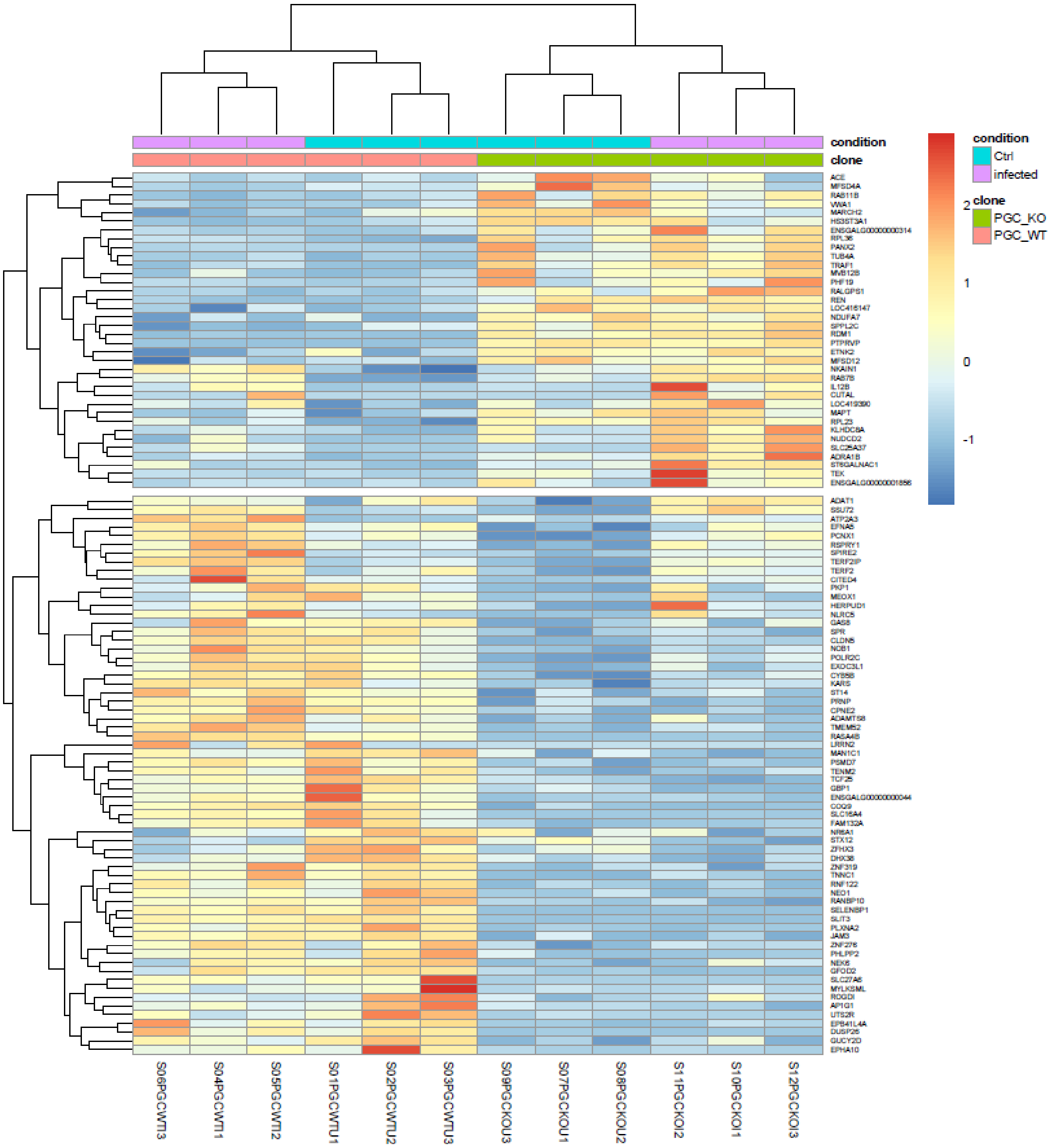


Figure 87 Most downregulated genes in comparison WT vs chIFITM KO PGC-derived fibroblasts TPM values of the hundred most downregulated genes that were statistically significant in comparison WT vs KO PGC-derived fibroblasts in control condition were normalised across all groups using the Z-score method. Those values were clustered in *pheatmap* package in R by their Euclidean distances.

Functional analysis

GSEA in clusterProfiler did not yield any results with the parameters for statistical significance used throughout this work. A reason may be that the level of noise in the data was too high. The intra-group variability in this dataset is non-negligible, reflecting a true biological system. This makes discernment between true differential expression and general biological heterogeneity difficult.

PADOG[102] was conducted with ReactomeGSA [112]. 2387 pathways were identified as significantly regulated.

Most notable upregulated pathways, as judged by the FDR, extent of the fold change and the number of identified genes belonging to the pathway were: PKB-mediated events (R-HSA-109703), Cellular response to hypoxia (R-HSA-1234174), Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha (R-HSA-1234176), Reversible hydration of carbon dioxide (R-HSA-1475029), Synthesis of CL (R-HSA-1483076), Cytosolic sulfonation of small molecules (R-HSA-156584), Eukaryotic Translation Elongation (R-HSA-156842), Peptide chain elongation (R-HSA-156902), PKA-mediated phosphorylation of key metabolic factors (R-HSA-163358), PP2A-mediated dephosphorylation of key metabolic factors (R-HSA-163767), PDE3B signalling (R-HSA-165160), Membrane binding and targeting of GAG proteins (R-HSA-174490), Synthesis And Processing Of GAG, GAGPOL Polyproteins (R-HSA-174495), Assembly Of The HIV Virion (R-HSA-175474), Retrograde neurotrophin signalling (R-HSA-177504), SRP-dependent co-translational protein targeting to membrane (R-HSA-1799339), Synthesis of IP3 and IP4 in the cytosol (R-HSA-1855204) and Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell (R-HSA-198933).

Most downregulated pathways were: Recognition and association of DNA glycosylase with site containing an affected purine (RHSA-110330), Cleavage of the damaged purine (R-HSA-110331), PKA-mediated phosphorylation of CREB (R-HSA-111931), Signaling by ERBB4 (R-HSA-1236394), ER-Phagosome pathway (R-HSA-1236974), Antigen processing-Cross presentation (R-HSA-1236975), Cross-presentation of soluble exogenous antigens (endosomes) (R-HSA-1236978), Nuclear signaling by ERBB4 (R-HSA-1251985), Developmental Biology (R-HSA-1266738), Acrosome Reaction and Sperm:Oocyte Membrane Binding (R-HSA-1300645), Extracellular matrix organization (R-HSA-1474244), Synthesis of PI (R-HSA-1483226), PKA activation (R-HSA-163615), PKA activation in glucagon signalling (R-HSA-164378), Regulation of Apoptosis (R-HSA-169911, Adenylate cyclase activating pathway (R-HSA-170660), Adenylate cyclase inhibitory pathway (R-HSA-170670) Packaging Of Telomere Ends (R-HSA-171306) and Vif-mediated degradation of APOBEC3G (R-HSA-180585).

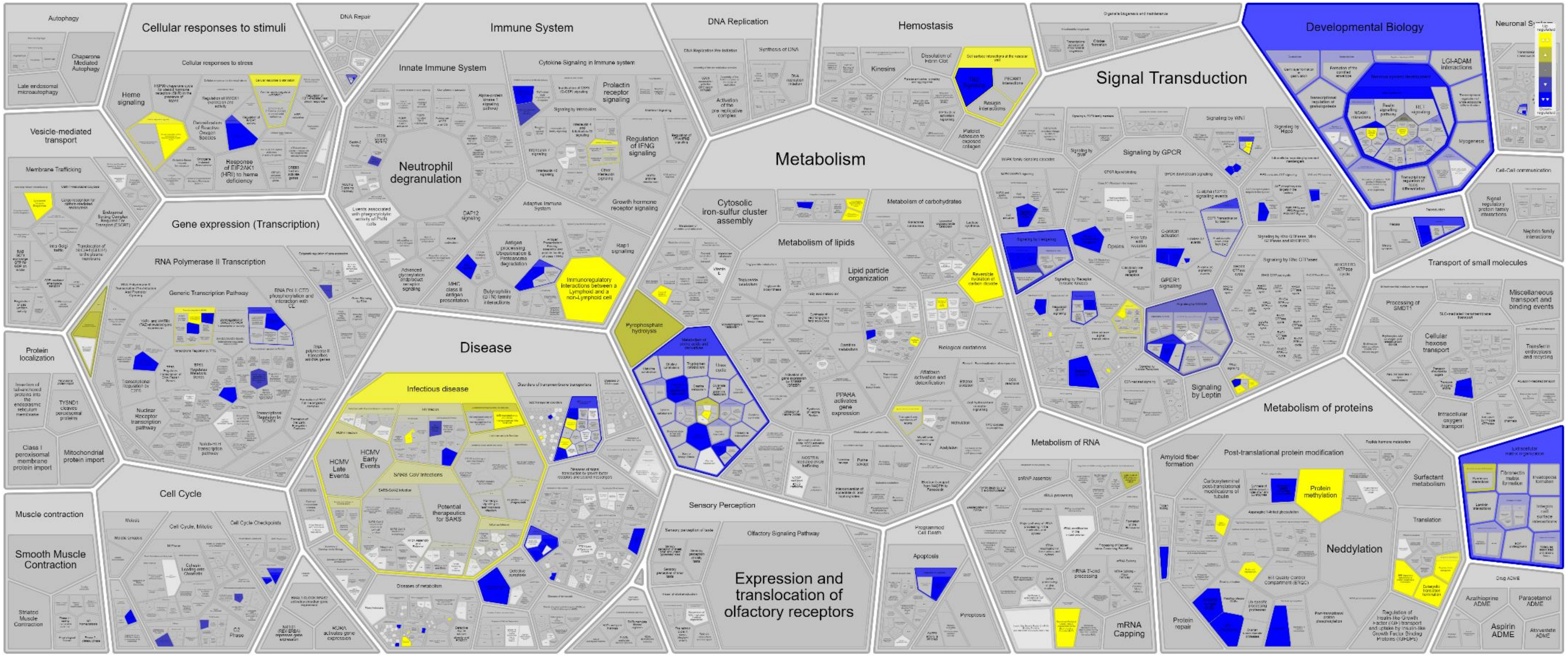


Figure 88 PADOG analysis of comparison of gene expression levels in WT vs KO PGC-derived fibroblasts PADOG analysis was conducted using ReactomeGSA for comparison WT vs KO PGC-derived fibroblasts from raw RNA-Seq counts. Differential expression analysis is conducted on the pathway level and FDR < 0.05 was considered significant. Upregulated pathways are highlighted in yellow and downregulated in blue.

5.3.3.3 Profiling of the transcriptomic response to infection with influenza virus in WT PGC-derived fibroblasts

Differential expression

We next compared the transcriptomic profile of the WT PGC-derived fibroblasts in infected condition against the uninfected. Differential expression was conducted in DESeq2 [94] with the same parameters as section 5.3.3.2. 1816 (9.6%) genes were upregulated at the p value 0.05 cut-off. 1427 (7.6%) were downregulated, while 1466 (7.8%) were discarded due to low count number. General linear model was the same as before ~0 + group. Because contrasts were used for the Wald test *ashr* [141] method was used to shrink the $\log_2$FC.

Differentially expressed genes filtered for statistical significance (FDR <0.05) and ordered by $\log_2$FC. Top 100 most upregulated genes were visualised in a heatmap (**Figure 89**). Their TPM values were Z-scored and clustered using Euclidean distances. Upregulated genes clustered primarily according to the condition, pointing towards a shared immune response between the two genotypes. The extent of upregulation varied meaningfully among replicates and between genotypes making precise calculation difficult with triplicate sample group per condition.

Downregulated genes (**Figure 90**) also confirm that the response to the influenza infection for the two genotypes is communal a the segregation of samples is foremostly by the infection status. Extent of downregulation was different between the WT and the KO PGC-derived fibroblasts, but consistent among triplicates, confirming a tighter regulation of the downregulated genes in comparison to the upregulated.

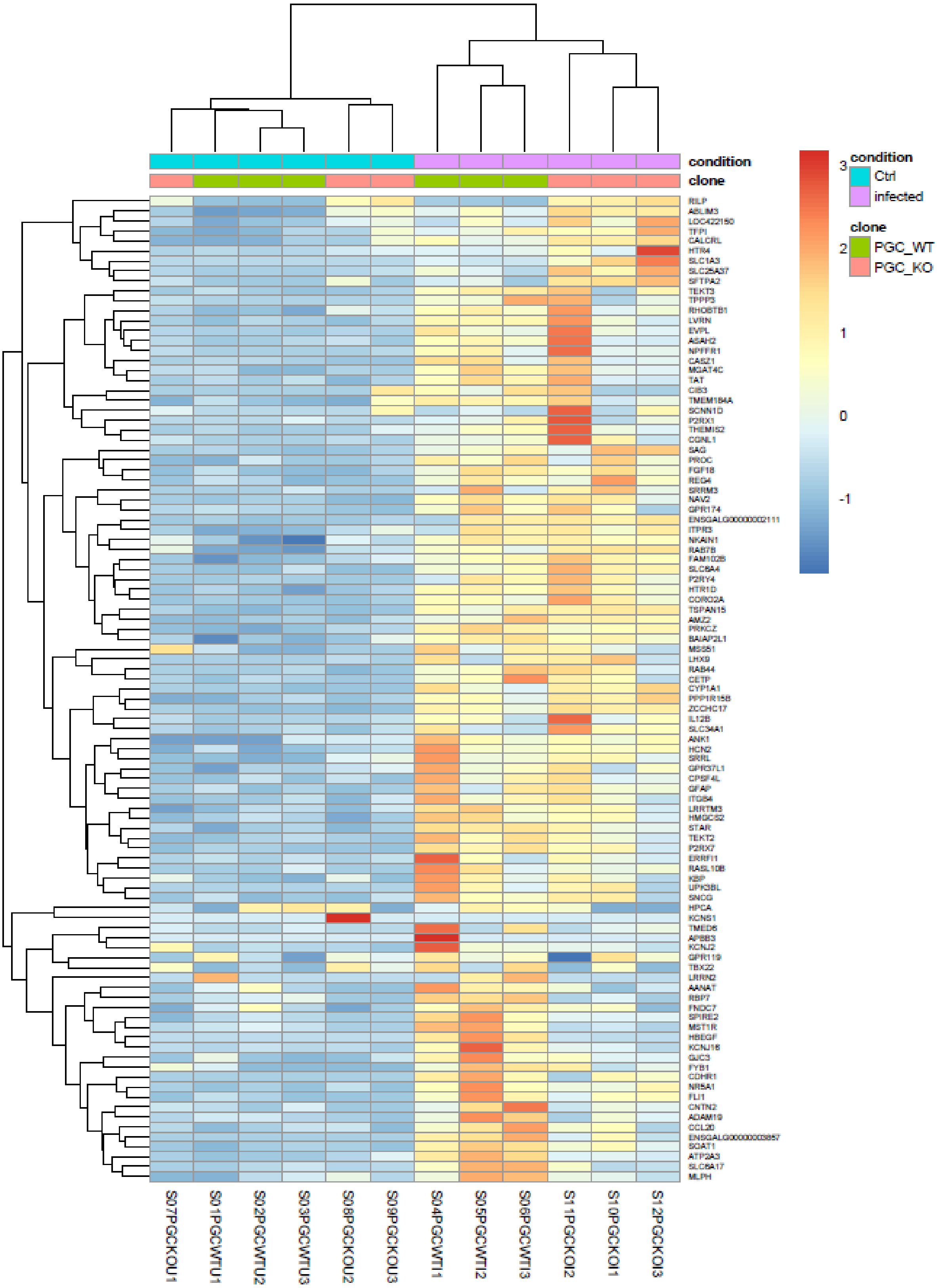


Figure 89 Most upregulated genes in WT PGC-derived fibroblasts TPM values for hundred genes with highest upregulation that was statistically significant were Z-scored and clustered in *pheatmap in* R using Euclidean distances.

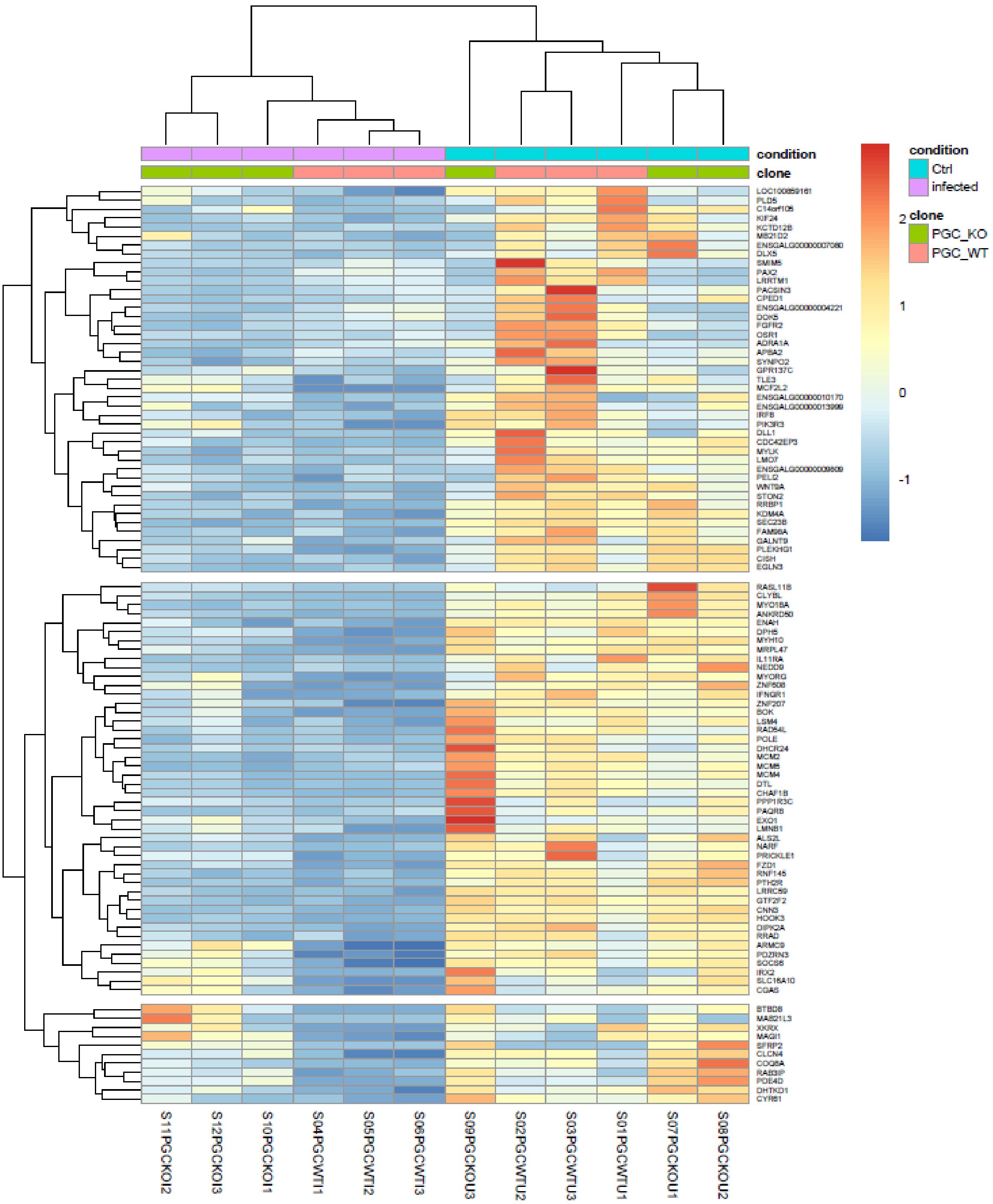
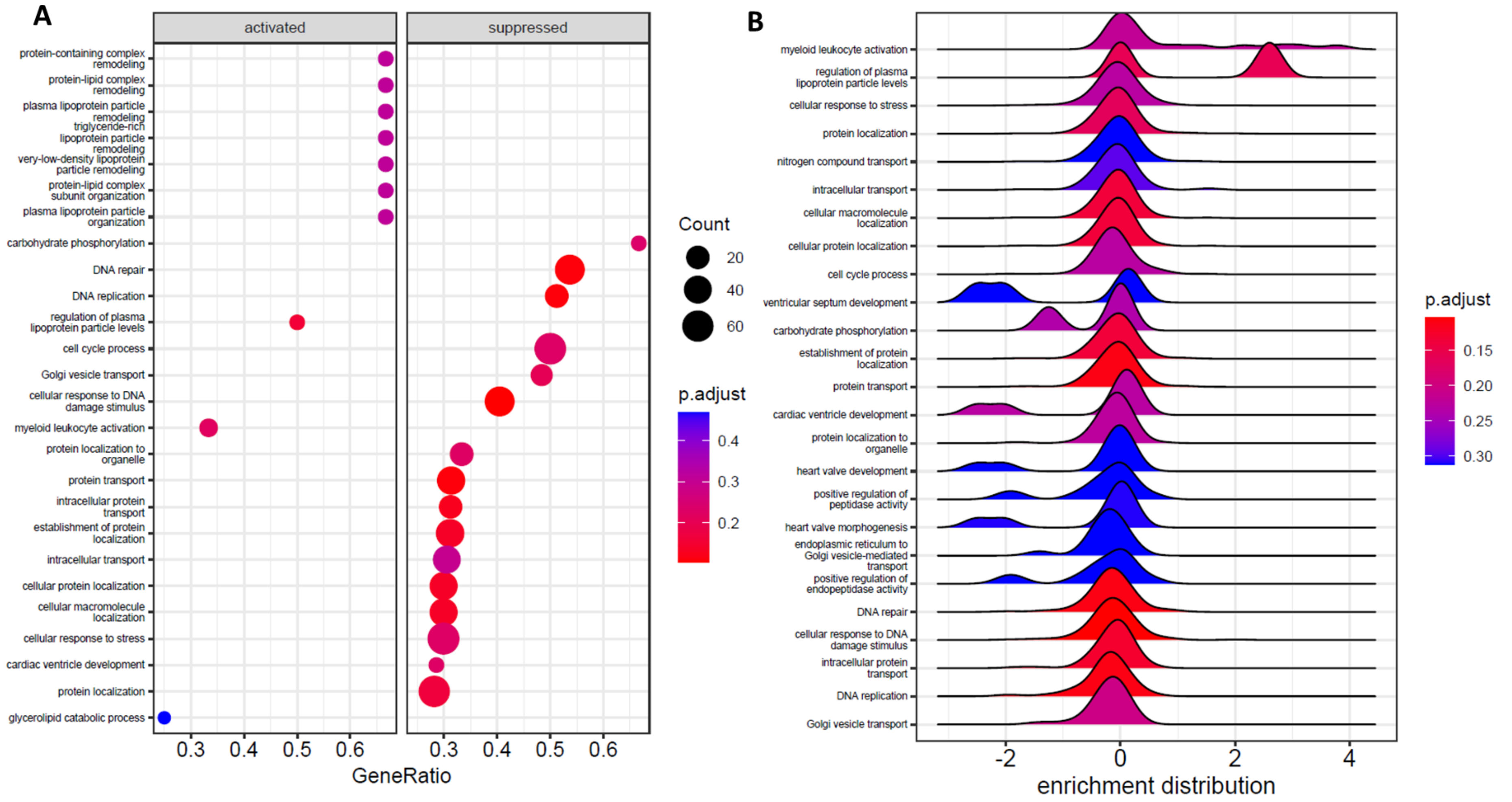


Figure 90 Most downregulated genes in WT PGC-derived fibroblasts at 6h post infection with H5N3 influenza virus

Functional analysis

Differentially expressed genes with their corresponding logFoldChanges shrunken with *ashr* function were supplied to the *gsea* function in clusterProfiler to perform the gene set enrichment analysis. Dataset list was tested against Gene Ontology database, specifically Biological Process category with the following parameters: minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.5, pAdjustMethod = "BH", eps = 0.


A
activated
suppressed
protein-containing complex remodeling
protein-lipid complex remodeling
plasma lipoprotein particle remodeling
triglyceride-rich lipoprotein particle remodeling
very-low-density lipoprotein particle remodeling
protein-lipid complex subunit organization
plasma lipoprotein particle organization
carbohydrate phosphorylation
DNA repair
DNA replication
regulation of plasma lipoprotein particle levels
cell cycle process
Golgi vesicle transport
cellular response to DNA damage stimulus
myeloid leukocyte activation
protein localization to organelle
protein transport
intracellular protein transport
establishment of protein localization
intracellular transport
cellular protein localization
cellular macromolecule localization
cellular response to stress
cardiac ventricle development
protein localization
glycerolipid catabolic process
0.3
0.4
0.5
0.6
0.3
0.4
0.5
0.6
GeneRatio
Count
20
40
60
p.adjust
0.4
0.3
0.2
B
myeloid leukocyte activation
regulation of plasma lipoprotein particle levels
cellular response to stress
protein localization
nitrogen compound transport
intracellular transport
cellular macromolecule localization
cellular protein localization
cell cycle process
ventricular septum development
carbohydrate phosphorylation
establishment of protein localization
protein transport
cardiac ventricle development
protein localization to organelle
heart valve development
positive regulation of peptidase activity
heart valve morphogenesis
endoplasmic reticulum to Golgi vesicle-mediated transport
positive regulation of endopeptidase activity
DNA repair
cellular response to DNA damage stimulus
intracellular protein transport
DNA replication
Golgi vesicle transport
-2
0
2
4
enrichment distribution
p.adjust
0.15
0.20
0.25
0.30

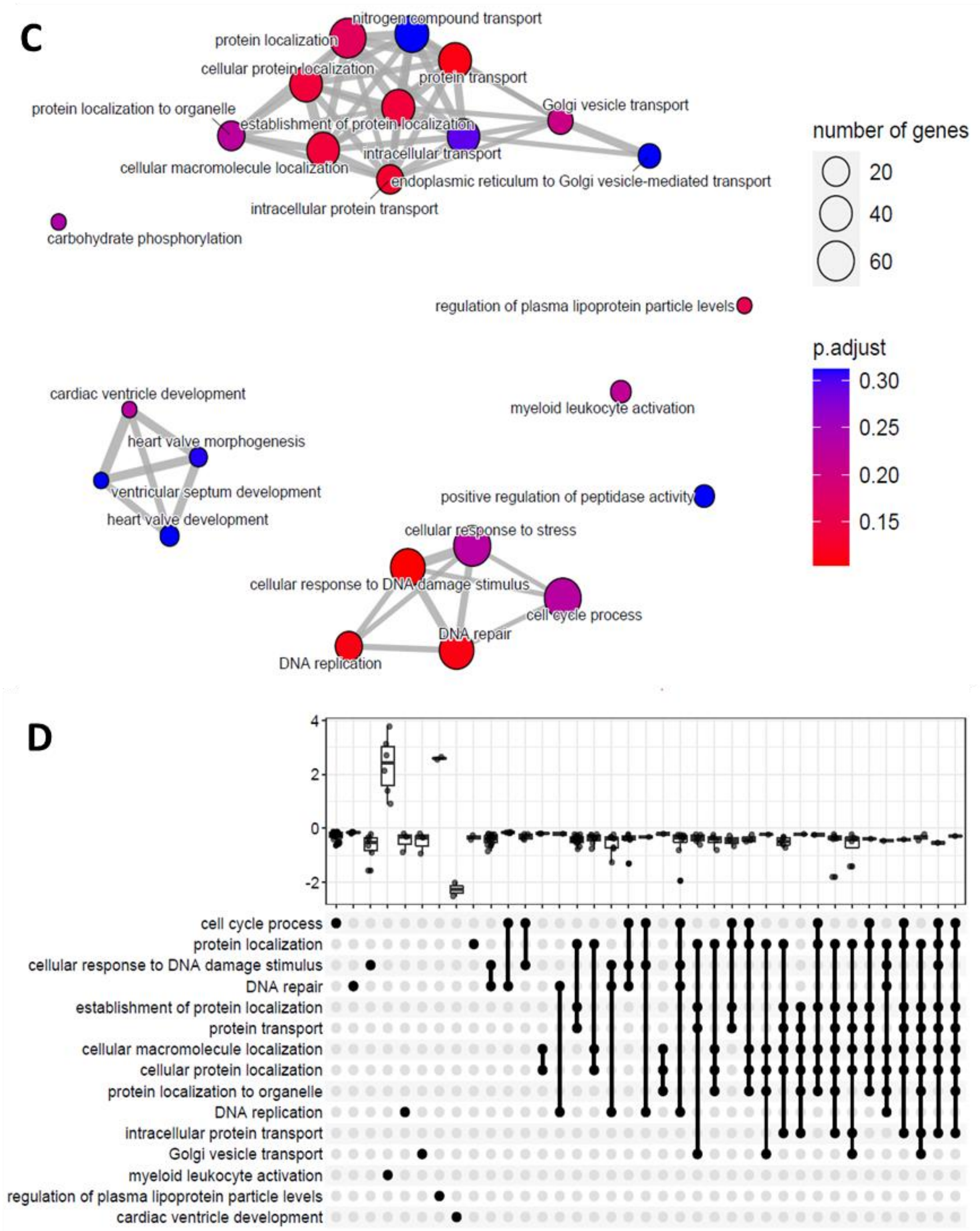


Figure 91 GSEA for comparison infected vs uninfected WT PGC-derived fibroblasts Differentially expressed genes were passed to the gsea function in clusterProfiler in R to test for enrichment against GO database, Biological Process category. FDR cutoff was set to 0.5 and p-values were adjusted for multiple testing with Benjamini-Hochberg correction. A) Dotplot of enriched processes split between activated and suppressed. Size of the circle reflects the number of genes from the set recognised in the differentially expressed list. Gene ratio plotted on the x-axis represent the proportion of those two numbers. B) Ridgeplot conveys the distribution of enrichment scores of the genes contained in the pathway. C) Emapplot identifies emergent biological themes in which enriched process participate. D) Upset plot displays distribution of fold changes for genes belonging to each gene set/pathway. Black dot indicates all the pathways where a particular gene is identified.

PADOG analysis

ReactomeGSA [112] was used to conduct Pathway Analysis with Downweighing of Overlapping Genes. 2416 pathways were enriched. Most significant pathways as judged by statistical significance, the extent of the upregulation and number of genes matched were: Intrinsic Pathway for Apoptosis (R-HSA-109606), GPVI-mediated activation cascade (R-HSA-114604), Signalling by ERBB2 (R-HSA-1227986), Cytokine Signalling in Immune system (R-HSA-1280215), TWIK-related alkaline pH activated K+ channel (TALK) (R-HSA-1299361), betaKlotho-mediated ligand binding (R-HSA-1307965), NR1D1 (REV-ERBA) represses gene expression (R-HSA-1368071), Signalling by EGFR in Cancer (R-HSA-1643713), Signalling by EGFR (R-HSA-177929), GRB2 events in EGFR signalling (R-HSA-179812), SHC1 events in EGFR signalling (R-HSA-180336), EGFR downregulation (R-HSA-182971), Signalling by activated point mutants of FGFR3 (R-HSA-1839130), FGFR2 ligand binding and activation (R-HSA-190241), FGFR4 ligand binding and activation (R-HSA-190322), FGFR2c ligand binding and activation (R-HSA-190375), Signalling by Rho GTPases (R-HSA-194315), FGFR3 mutant receptor activation (R-HSA-2033514), PECAM1 interactions (R-HSA-210990), TGF-beta receptor signalling in EMT (epithelial to mesenchymal transition) (R-HSA-2173791).

Downregulated pathways included: Recognition of DNA damage by PCNA-containing replication complex (R-HSA-110314), Translesion Synthesis by POLH (R-HSA-110320), Activation and oligomerization of BAK protein (R-HSA-111452), Inhibition of replication initiation of damaged DNA by RB1/E2F1 (R-HSA-113501), Activation, translocation and oligomerization of BAX (R-HSA-114294), Transport of Mature mRNA Derived from an Intronless Transcript (R-HSA-159231), Late Phase of HIV Life Cycle (R-HSA-162599), Synthesis of dolichyl-phosphate mannose (R-HSA-162699), HIV Transcription Initiation (R-HSA-167161), RNA Polymerase II HIV Promoter Escape (R-HSA-167162), Inhibition of Host mRNA Processing and RNA Silencing (R-HSA-168315), Unwinding of DNA (R-HSA-176974), p75NTR signals via NF-kB (R-HSA-193639) and Ceramide signalling (R-HSA-193681), Membrane Trafficking (R-HSA-199991), NFG and proNGF binds to p75NTR (R-HSA-205017), NADE modulates death signalling (R-HSA-205025) and NF-kB is activated and signals survival (R-HSA-209560).

General themes that were downregulated involved developmental biology and metabolism of amino acids and their derivatives, highlighted in blue in **Figure 92**. Wider theme that was upregulated infectious disease, highlighted in yellow.

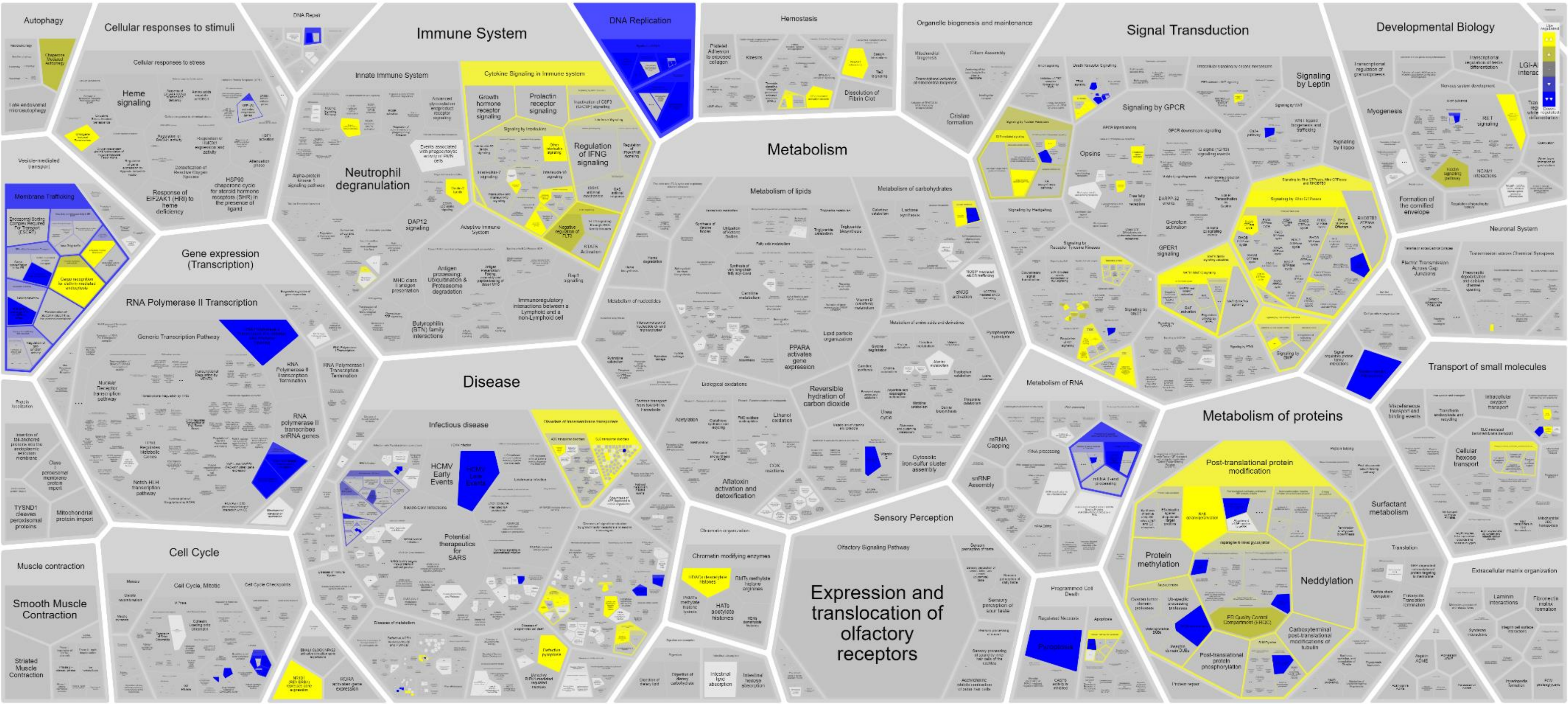


Figure 92 Pathway analysis for comparison WT PGC-derived fibroblasts infected vs uninfected Raw RNA-Seq data were passed to the ReactomeGSA function and differential expression was conducted on the pathway level using Reactome API using PADOG algorithm. FDR <0.05 was the cut-off for significance.

5.3.3.4 Analysis of the transcriptomic response to infection with influenza virus in chIFITM KO PGC-derived fibroblasts

Differential expression

Lastly, we compared levels of expression between infected and uninfected condition for the KO PGC-derived fibroblasts. As previously, DESeq2 was used to perform the Wald test with design formula ~ 0 + group. Contrasts were used to set up the comparison of interest and *ashr* shrinkage was employed to adjust the effect sizes. FDR < 0.05 was considered significant after Benjamini-Hochberg correction. 2319 (12%) genes were upregulated, 1892 (10%) were downregulated and 1100 (5,5%) were discarded due to low counts. 19 genes were considered outliers (0.1%) and were not shrunken sing the mean across all samples.

Statistically insignificant genes were filtered out; remaining genes were sorted by LogFoldChange. For the hundred most extreme genes in both directions TPM values were Z-scored and Euclidian distances calculated to cluster the genes in *pheatmap* package in R. At the highest level, the upregulated genes clustered according to the condition (**Figure 93**) and following that according to the genotype. Replicates clustered together suggesting that the regulation of gene expression as a response to infection is robust despite varying levels of upregulation in individual samples. No genetic clusters that could suggest co-regulation were observed.

Downregulated genes exhibited same overall characteristics as the upregulated in terms of clustering. At the highest level, two clusters are obvious: infected and uninfected. At the subsequent level we can observe the separation according to the genotype (**Figure 94**). No clusters were identified at a higher level of resolution.

Functional analysis

The absence of a more granular clustering of genes that would suggest co-regulated pathways and metabolic processes is not uncommon when rough measures such as Z-scores are used. Thus, a functional analysis was conducted to uncover biological themes present in this experimental system. GSEA was performed in *clusterProfiler* in R interrogating the Biological Process category of the GO database using following parameters: minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.05, pAdjustMethod = "BH", eps = 0).

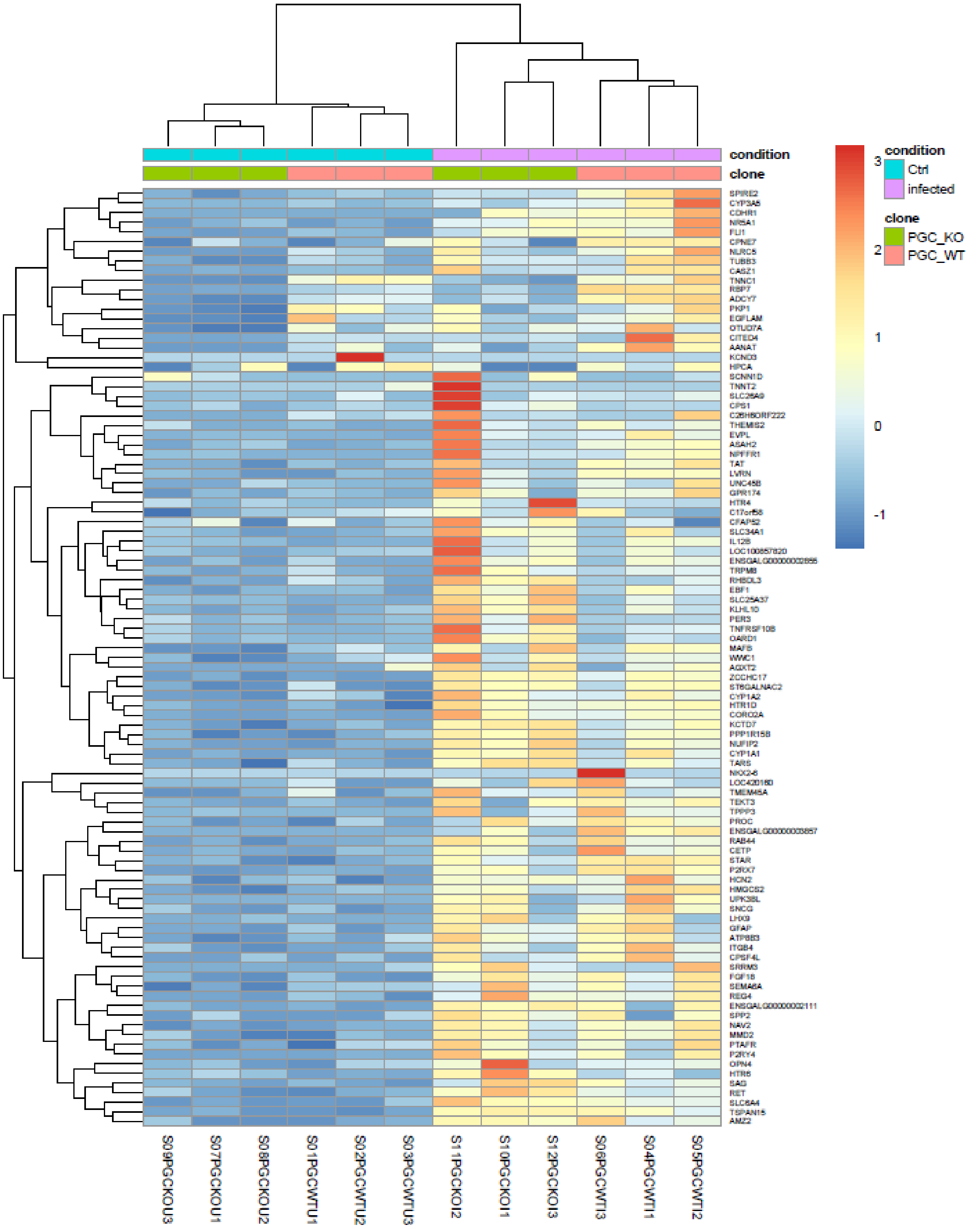


Figure 93 Upregulated genes in comparison of infected and uninfected samples of KO PGC-derived fibroblasts 100 most upregulated genes with FDR<0.05 were extracted and their expression levels standardised by Z-scoring their TPM values. Euclidean distances were used as the metric of distance for clustering. Calculations were performed in baseR and visualisations produced with pheatmap package in R.

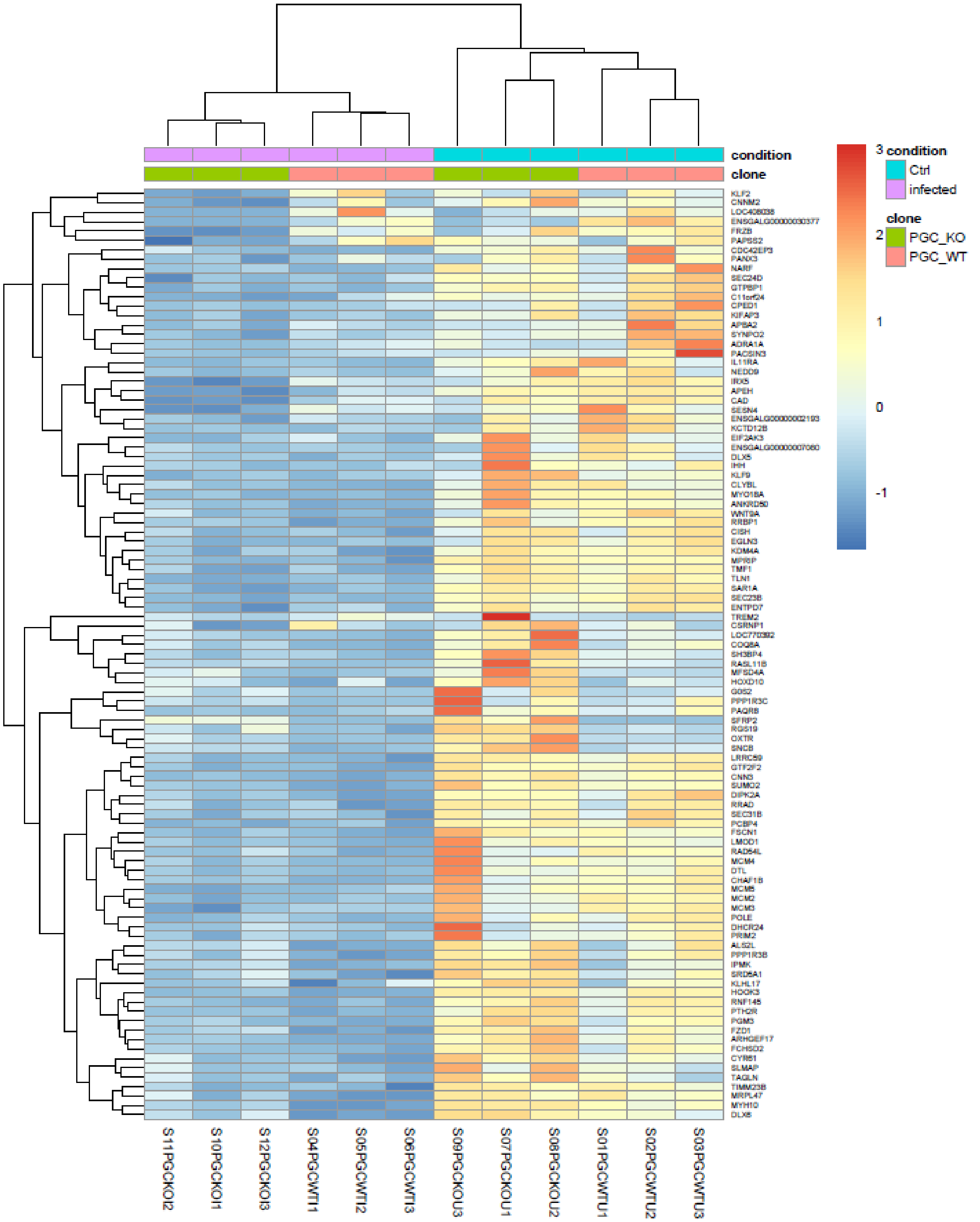
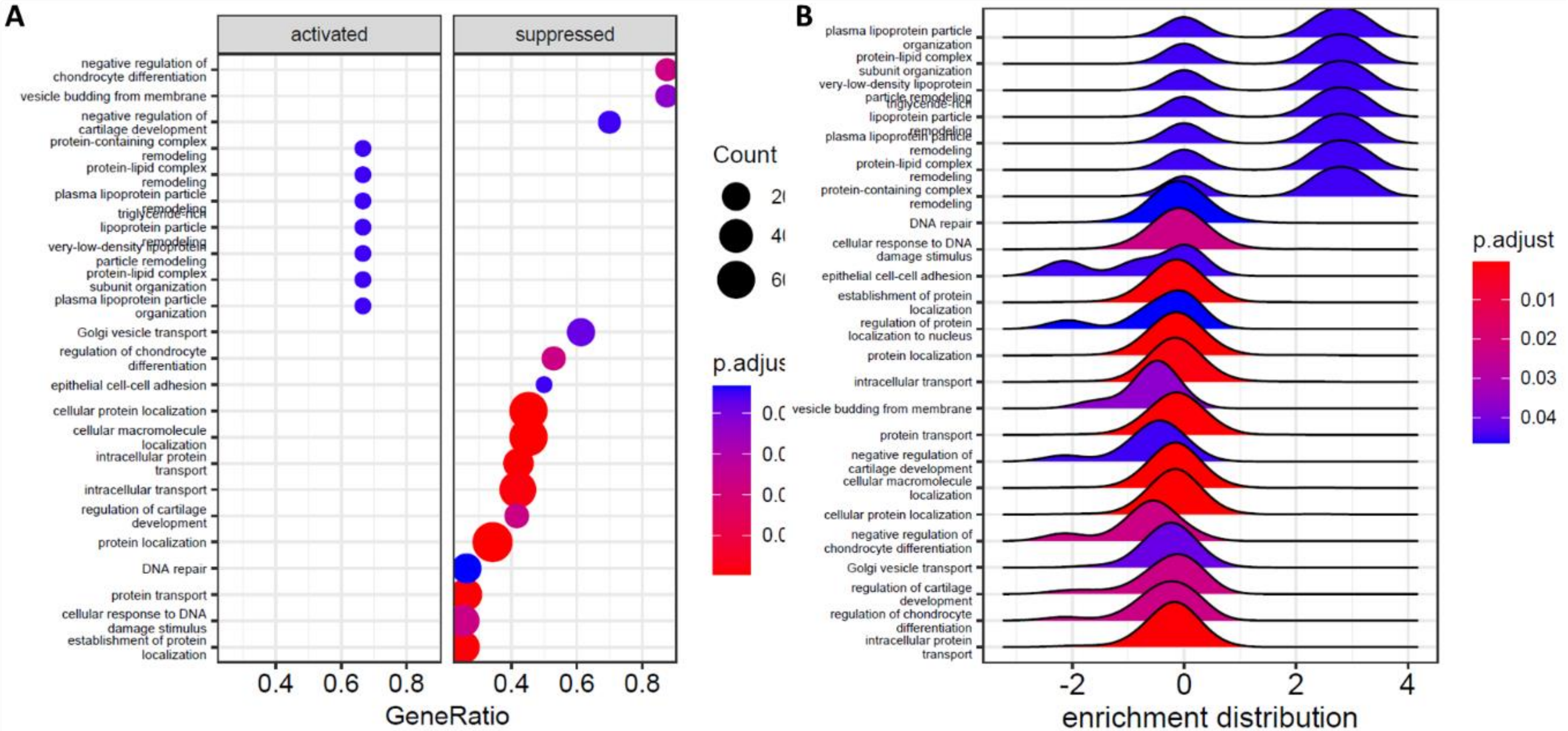


Figure 94 Most downregulated genes in KO PGC-derived fibroblasts at 6h post infection with influenza TPM values of most downregulated differentially expressed genes with FDR < 0.05 were normalised across samples using Z-score method and clustered using Euclidean distances in *pheatmap* in R.


A
activated
suppressed
negative regulation of chondrocyte differentiation
vesicle budding from membrane
negative regulation of cartilage development
protein-containing complex remodeling
protein-lipid complex remodeling
plasma lipoprotein particle remodeling
triglyceride-rich lipoprotein particle remodeling
very-low-density lipoprotein particle remodeling
protein-lipid complex subunit organization
plasma lipoprotein particle organization
Golgi vesicle transport
regulation of chondrocyte differentiation
epithelial cell-cell adhesion
cellular protein localization
cellular macromolecule localization
intracellular protein transport
intracellular transport
regulation of cartilage development
protein localization
DNA repair
protein transport
cellular response to DNA damage stimulus
establishment of protein localization
0.4
0.6
0.8
0.4
0.6
0.8
GeneRatio
Count
2(
4(
6(
p.adjus
0.(
0.(
0.(
0.(
B
plasma lipoprotein particle organization
protein-lipid complex subunit organization
very-low-density lipoprotein particle remodeling
triglyceride-rich lipoprotein particle remodeling
plasma lipoprotein particle remodeling
protein-lipid complex remodeling
protein-containing complex remodeling
DNA repair
cellular response to DNA damage stimulus
epithelial cell-cell adhesion
establishment of protein localization
regulation of protein localization to nucleus
protein localization
intracellular transport
vesicle budding from membrane
protein transport
negative regulation of cartilage development
cellular macromolecule localization
cellular protein localization
negative regulation of chondrocyte differentiation
Golgi vesicle transport
regulation of cartilage development
regulation of chondrocyte differentiation
intracellular protein transport
-2
0
2
4
enrichment distribution
p.adjust
0.01
0.02
0.03
0.04

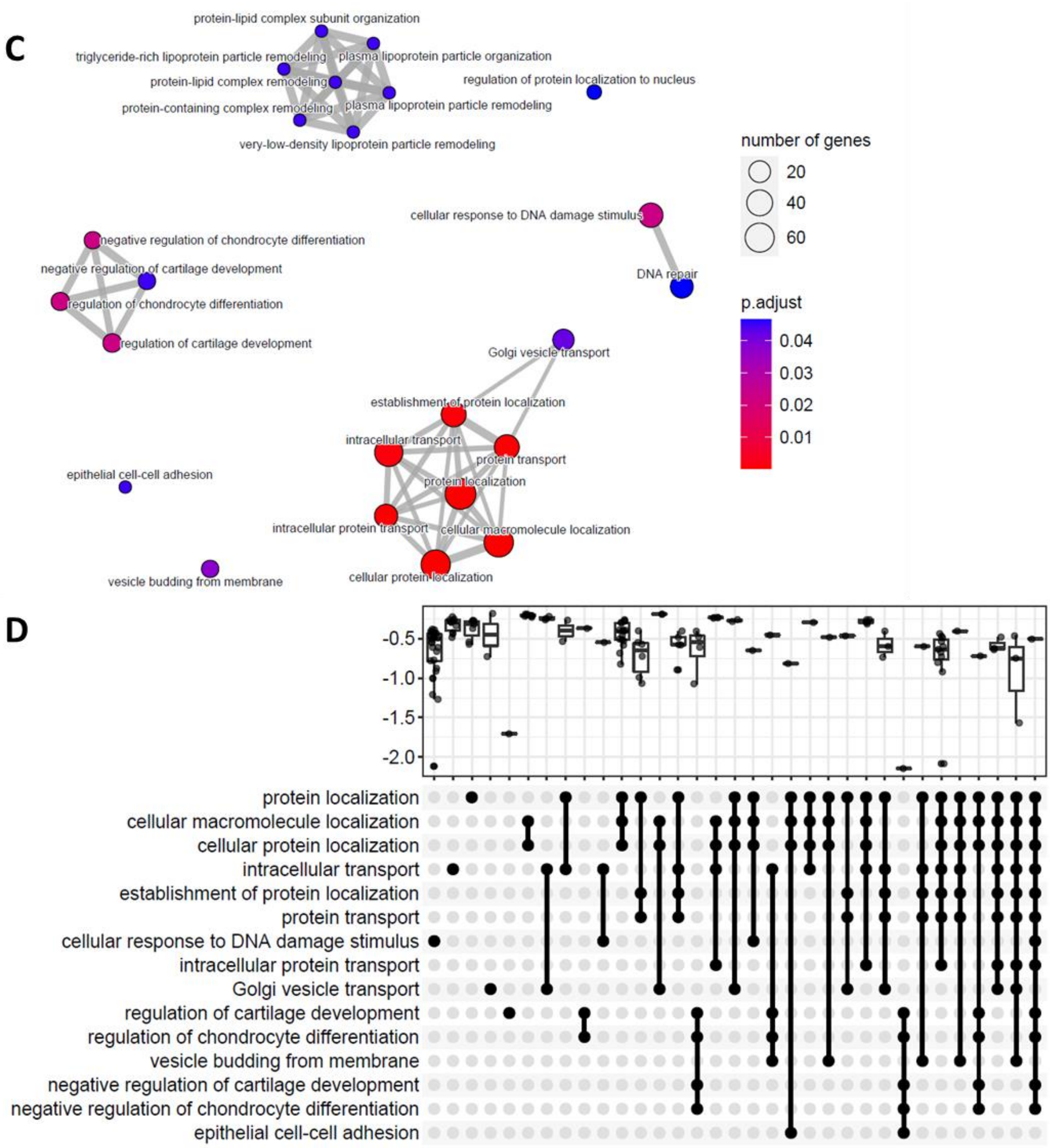


Figure 95 GSEA analysis for KO PGC-derived fibroblasts at 6h post infection with influenza Differentially expressed genes, as determined by DESeq2 had their logFoldChanges shrunken using ashr procedure. Resuting values were analuysed wit cluerProfiler using GSEA enrichment approach. A) Dotplot. Size indicated number of genes from the gene set identified in the dataset,; that ratio is plotted on the x-axis. B)Ridgeplot showcases the distribution of enrichment scores for all genes identified in a given pathway. C) Emapplot. Enriched sets that overlap form a network and suggest a common wider biological theme. Size of the node indicates number of matched genes and colour the FDR of the hypergeometric test. D)Upsetplot Top panel shows the distribution of fold changes of genes that are part of a pathway plotted on the y-axis/left side of the panel. If a gene from a genet set is identified as part a pathway the intersection is marked with a black dot.

Gene set enrichement analysis showed the most statistically significant pathways are almost exclusively suppressed. The supressed pathways included establishment of protein localization and transport, response to DNA damage, regulation of cartilage development. Transport more broadly was negatively regulated extending to processes such as vesicle budding from membrane and Golgi vesicle transport. Activated processes were concerned with protein-lipid interface such as complex formation and remodelling. However, those gene sets were small and not as statistically significant because identified genes were members of various overlapping terms.

PADOG analysis

PADOG analysis was next employed to validate the GSEA results using a more statistically sophisticated approach. As previously, raw RNA-Seq data were passed to the ReactomeGSA and differential expression analysis was done on the pathway level with FDR of <0.05. Enrichment was tested against Reactome database based on protein-protein interactions. Chicken Ensembl IDs were automatically translated into human and projected onto common interaction space. Analysis identified 2427 differentially regulated pathways.

Upregulated pathways included: Formation of the Early Elongation Complex (R-HSA-113418), Signalling by ERBB2 in Cancer (R-HSA-1227990), SHC1 events in ERBB2 signalling (R-HSA-1250196), SHC1 events in ERBB4 signalling (R-HSA-1250347), PIP3 activates AKT signalling (R-HSA-1257604), Formation of the HIV-1 Early Elongation Complex (R-HSA-167158), RNA Pol II CTD phosphorylation and interaction with CE during HIV infection (R-HSA-, HIV Transcription Initiation (R-HSA-167161), RNA Polymerase II HIV Promoter Escape (R-HSA-167162), Transcription of the HIV genome (R-HSA-167172), Conjugation of salicylate with glycine (R-HSA-177128), Cytosolic sensors of pathogen-associated DNA (R-HSA-1834949), GRB2 events in ERBB2 signaling (R-HSA-1963640), Generic Transcription Pathway (R-HSA-212436), POU5F1 (OCT4), SOX2, NANOG activate genes related to proliferation (R-HSA-2892247) and Regulation of innate immune responses to cytosolic DNA (R-HSA-3134975).

Downregulated pathways that were identified from Reactome were: Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template (RHSA-110313), Displacement of DNA glycosylase by APEX1 (R-HSA-110357), Synthesis of PE (R-HSA-1483213), Cytosolic sulfonation of small molecules (R-HSA-156584), Entry of Influenza Virion into Host Cell via Endocytosis (R-HSA-168275), Synthesis of IPs in the nucleus (R-HSA-1855191), Synthesis of IPs in the ER lumen (R-HSA-1855231), ER to Golgi Anterograde Transport (R-HSA-199977), trans-Golgi Network Vesicle Budding (R-HSA-199992), COPII-mediated vesicle transport (R-HSA-204005), EGFR Transactivation by Gastrin (R-HSA-2179392), Electron transport from NADPH to Ferredoxin (R-HSA-2395516), Regulation of ornithine decarboxylase (ODC) (R-HSA-350562), IRE1alpha activates chaperones (R-HSA-381070), Golgi Associated Vesicle Biogenesis (R-HSA-432722) and Asparagine N-linked glycosylation (R-HSA-446203).

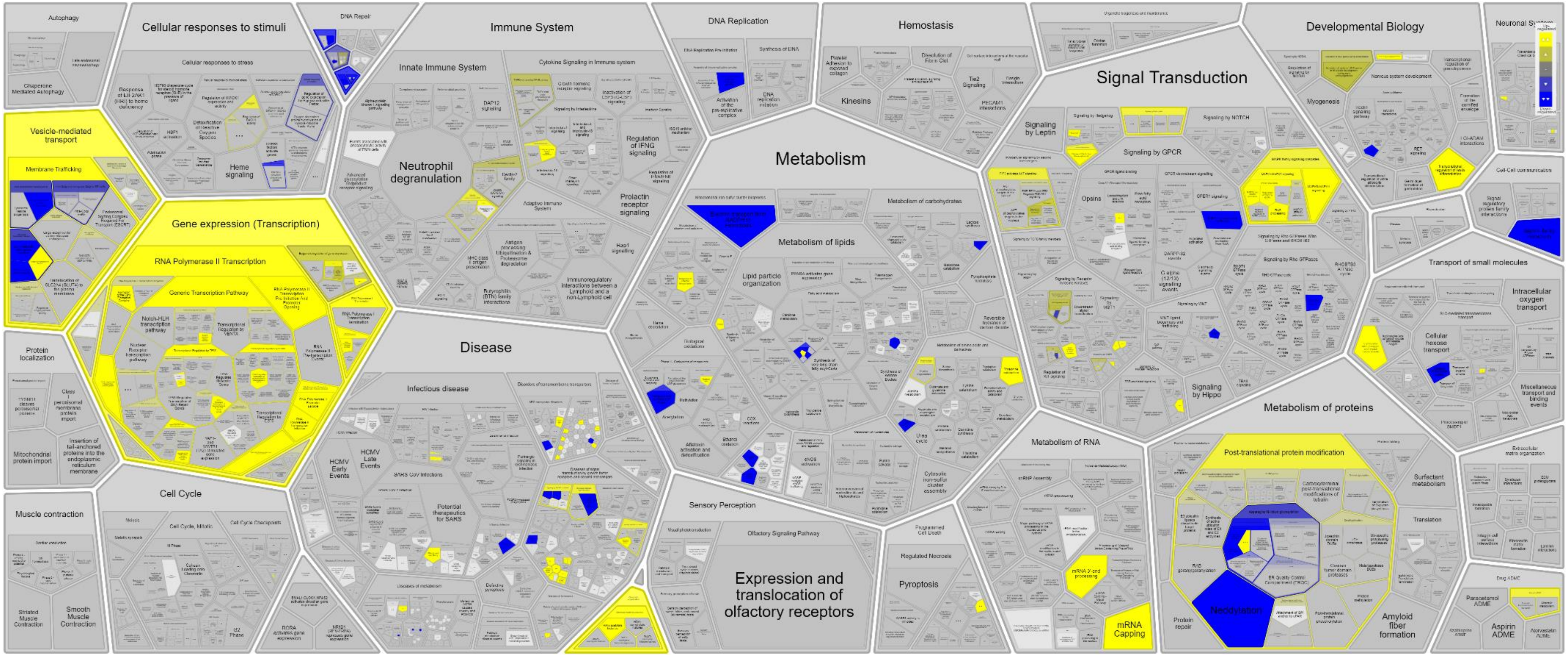


Figure 96 Pathway analysis for differentially expressed genes in KO PGC-derived fibroblasts at 6 hours post infection with influenza Raw RNA-Seq data were analysed for differential expression on the pathway level with Reactome API using the PADOG algorithm and FDR <0.05. Upregulated processes are coloured in yellow and downregulated in blue. Saturated colours indicate the term is statistically significant, while muted tones depict terms which were found to be differentially expressed but did not pass the statistical significance threshold.

5.4 Discussion

By virtue of their localisation, chIFITM have a direct or indirect role in processes involving membranes. To expand the knowledge of IFITM function in chicken we endeavoured to make a model more akin to what can be found in the field. Continuous cell lines have various characterised and uncharacterised changes that alter gene regulation networks and signalling pathways, potentially lessening the relevance and utility of acquired information when such systems are used in research. Using primary cells not only provides fundamental knowledge but also a reference for comparison with other cell lines. Chicken PGCs are easily retrieved, maintained in culture [189,190,202], edited [199,203] and differentiated into other cell types [204] making them an extremely flexible model for research in avian immunology and developmental biology. PGCs were differentiated into fibroblasts to ensure comparability with DF-1 fibroblasts, both readily infected by IAV.

CEFs are also frequently used as a primary model. However, they have a short life span and do not lend themselves to genetic modifications, a central premise of this work. PGCs were, therefore, the cell type of choice. They were edited using CRISPR/Cas9 and differentiated into fibroblasts using serum induction (**Figure 97**). Knock-out was confirmed by PCR, Sanger sequencing and RNA-seq. We could not irrefutably demonstrate bi-allelic deletion in PGC-derived fibroblasts, however the levels of expression of *chIFITM* genes were extremely reduced in the KO (**Figure 70**), therefore we proceeded with the analysis.

IFITMs are readily expressed in mammalian PGCs where they ostensibly influence homing [177,205] through an unknown mechanism. Subsequently however, Lang *et al* have proven IFITMs are not indispensable for that function when they successfully generated *IFITM* KO mice who reached maturity with no defects in development [179]. Nevertheless, this research is still used to inform databases of signalling pathways and ontological terms assigning to IFITMs a functional role in organ development. It is currently unknown whether chicken IFITMs would have the same or any role in PGC migration or organ development.

WT chicken PGC-derived fibroblasts used in this work had low relative levels of expression of *chIFITM3* in comparison with WT DF-1 (**Figure 70**). This is not a rarity; same phenomenon was observed in LMH cells (personal communication with Dr Ahmed Ibrahim). Levels of *chIFITM2* and 5 were relatively higher than *chIFITM3* in WT PGC-derived fibroblasts. *chIFITM1* expression was very low and did not change significantly in infection. Contrary to our hypothesis that ablating *chIFITM* locus would make the cells more permissive to the influenza virus, WT PGC-derived fibroblasts were consistently more permissive at all time points tested, albeit not significantly. This suggests that a, at present unknown, downstream effect of the *chIFITM* KO is rendering the cells more refractory to infection. At 3hpi expression levels of influenza NP protein, as measured by the MFI, were the same for WT and KO (**Figure 79**). At 6hpi, however, there was a remarkable increase in the MFI in the WT, indicating influenza NP protein being successfully produced in the WT but not so in the KO PGC-derived fibroblasts.

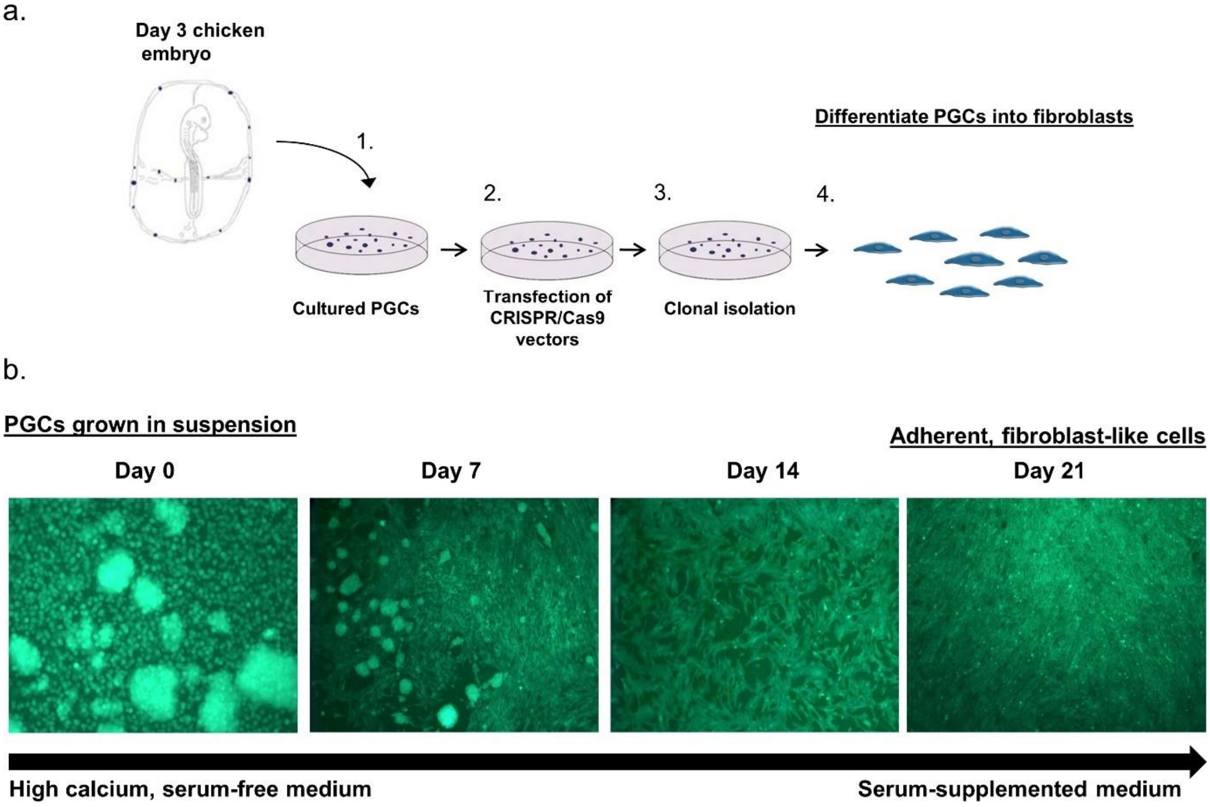


Figure 97 Generation of chicken PGC-derived fibroblasts a) PGCs were retrieved from Hyline cross chicken embryos, stabilised in vitro and edited to delete *chIFITM* locus. Correct clones were subsequently differentiated into fibroblasts. b) Fluorescent images of cells showing the progression from PGCs to fibroblast-like cells. Curtesy of Dr Idoko-Akoh and Dr McGrew at the Roslin Institute. Figure borrowed from [199].

To elucidate the mechanism getting activated by the *chIFITM* KO that hinders the same processes in the KO cells themselves will require further research.

Viral restriction is not the only function of *IFITMs*. There is information in the literature, albeit in mammalian models, documenting infection-promoting activities of IFITMs, specifically in coronaviruses [60,61] as well as IFITMs having alternative roles in cell signalling [172] and cancer [181,206]. Nevertheless, restrictive capabilities of IFITMs, namely IFITM3, are a common theme when influenza was used as the model virus, making these results surprising. We did expect a lesser effect of the *chIFITM* KO on the PGCs' ability to support viral replication than what was observed in the DF-

1. DF-1 are known to yield high titres because of high levels SOCS1 expression [120]. Aligned RNA-Seq data affirmed that; mapping the reads to concatenated host and virus genome, as opposed to host only, improved the mapping by 40-50% (**Figure 41**) in DF-1 infected samples, while the improvement in PGC-derived fibroblasts was only 20% (**Figure 80**). To relate the mapping data to the flow cytometry results, in the future viral reads could be mapped to the influenza genome to determine which genes exactly are being translated. This would provide information on the kinetics of the virus life cycle as well as validate the MFI results that indicated that virus is replicating in the WT more than in the KO. It is known however that the correlation of mRNA levels and corresponding protein levels is poor [207,208]. Nevertheless, in the context of more data becoming available with time, this piece of information could contribute to the overall knowledge of how virus replicates in PGC-derived fibroblasts.

To our knowledge, the effects of *chIFITM* knock-out on cellular transcriptome have never been studied in chicken PGC-derived fibroblasts.

Intergroup variability on the transcriptomic level was significant, exacerbated in the *chIFITM* KO (**Figure 83**.). The gene edit disturbed the transcriptome dramatically; PCA determined that 71% of variability in the data could be explained by the knock-out of the *chIFITM* locus. More than 2700 differentially expressed genes indicate that many processes are disturbed as a consequence. To look into processes affected, GSEA and PADOG were employed and three comparisons were conducted for PGC-derived fibroblasts:

1. WT vs KO in basal condition
2. WT Infected vs control
3. KO infected vs control.

Most significant differentially regulated pathways are summarised in **Table 9**.

WT vs KO in basal condition

GSEA of differentially expressed genes between WT and KO under specified FDR 0.05 did not identify any enriched terms. With the more sophisticated statistical approach of PADOG I was able to extract information despite the observed intrinsic variability. PADOG downweighs genes that participate in more than one process accordingly and subsequently calculates enrichment and the statistical significance. This approach increases the probability of correct identification of truly enriched terms.

Similarly to *chIFITM* KO vs WT comparison in DF-1, biological themes of developmental biology such as nervous system development, axon guidance and fertilization were downregulated. Leading edge analysis would provide information which genes and consequently pathways contribute the most to the enrichment these pathways. It would be interesting to investigate if chIFITM5 could have a role in this. IFITM5 in mammals has been implicated in bone mineralisation [209,210] and the exact mechanism remains unclear. In mammals IFITM5 is not an ISG, while in the chicken we have some preliminary evidence it may be (data not shown). But regardless, it could be that its expression is inducible with different triggers and its function adapted accordingly. chIFITM5 expression was high in WT PGC-fibroblasts, relative to other IFITMs thus the impact of its deletion may be more profound. It is notable that this process/biological theme of developmental biology surfaced in the comparison of WT vs KO DF-1 as well, pointing to a common downstream dysregulation of expression of genes when the system is not under stress of infection.

Extra cellular matrix organisation was also downregulated in the *chIFITM* KO PGC-derived fibroblasts. IFITM3 has been shown to participate in control of raft formation, acting as a PIP-scaffold and amplifying PI3K signalling [172]. Impaired raft formation affects proper distribution and concentration of receptors on the cell surface which in turn has a destabilising effect on cell adhesion and signalling. It is a challenge to singularly interpret these results in the context of chIFITM because the research is scarce and the information from human and mouse data is conflicting. The results often depend on the experimental setup. This ambiguity testifies of the flexibility of IFITMs to adapt to signalling from the environment and exert different functions depending on need.

Upregulated pathways were smaller and metabolic, and not part of big biological themes indicating potentially a short-term response to local stimulus rather than an organised process involving many coordinated pathways. In cell culture in native conditions both WT and KO PGC-derived fibroblasts grew equally fast and had similar morphology (see 5.3.1). It would appear that true effect of the knock-out comes through under stress such as infection.

Table 9 Short summary of differentially regulated pathways in PGC-derived fibroblasts Three comparisons were conducted from raw RNA-seq data using PADOG analysis in ReactomeGSA. Summary includes meaningful biological themes that were supported by more than one pathway or by other evidence in this thesis. Bolded terms were statistically significant.

	Upregulated	Downregulated
PGC Ctrl_ KO_vs_WT	**Lysososme vesicle trafficking** Cellular response to hypoxia and starvation, Protein metylation Sialic acid metabolism Eukaryotic translation termination Peptide chain elongation Signalling by NOTCH Signalling by Hedgehog	Disorders of ABS transporters Metabolism of amino acid derivatives E3 ubiquitination ligases **Extracellular matrix organisation** **Nervous system development** Axon guidance Developmental biology Fertilization
PGC WT_Inf_vs_Ctrl	**Cytokine signalling** **Disorders of transmembrane transporters** **MAPK signalling** Rho Signalling EGFP signalling Post translational protein modifications **Apoptosis**	**Membrane trafficking** **DNA replication** **Processing of capped pre-mRNA**
PGC KO_Inf_vs_Ctrl	**Transcription** **Cargo concentration in the ER** **Vesicle mediated transport** **Chromatin organisation** **mRNA capping** **mRNA 3'end processing** **PIP3 activates AKT signalling** **MAPK signalling** Post-translational protein modification	Exit to Golgi Antergrade transport TransGolgi vesicle budding Retrograde trafficking to ER Response to hypoxia DNA repair

WT vs KO in infected conditions

When influenza virus infects a cell, an intricate and dynamic network of signalling pathways gets set in motion by the host in the attempt to mitigate the attack. In turn, virus encodes its own proteins that interact with the host's network and alter the pathways for its own benefit. Exact strategy depends upon the type of the cell, genetic background of the host and the virus, site of the infection, ecological niche etc. PGC-derived fibroblasts of WT and *chIFITM* KO genetic background were used to describe the transcriptomic response at 6hpi with the H5N3 influenza virus.

Similarities in the response and ultimately the differences that explain, at least in part, the differential permissiveness of the two cell lines to the influenza virus and the higher levels of expression of viral NP protein in the WT, are described in the following section.

Themes that were identified in both systems were signalling pathways controlling proliferation and cell division, cell survival and cell death, differentiation, and immune response: all ubiquitously functioning pathways in many conditions, viral infection notwithstanding. Some of such pathways that were upregulated here are MAPK signalling, PI3K/Akt pathway, ERBB network and signalling via Rho GTP-ase. These pathways intersect and overlap giving rise to multiple checkpoints to ensure that a desirable response to the external stimulus is produced.

MAPK signalling facilitates transmission of an extracellular stimulus into intracellular output, usually via transcription machinery. This pathway is ubiquitously used by cells to respond to the environment and can be easily employed in response to viral attacks as secondary signalling that controls proliferation, immune response and cell death [211,212]. Influenza can both re-direct already activated pathway or induce a latent one. Interestingly, Raf/MERK/ERK branch of MAPK signalling can be activated anew by accumulation of the HA at the plasma membrane later in the viral life cycle to facilitate RNP export from nucleus after all other influenza segments are transcribed and viral progeny is ready to bud [213].

PI3K-Akt can be induced by NS1 to support viral replication by manipulating regular cellular function, e.g. suppression of cell death in the early and intermediate phases of the virus life cycle [214] or inducing apoptosis in later stages of life cycle (around 20hpi), mediated by upregulated expression of p53, to facilitate progeny release [215]. Note that apoptosis is upregulated in the WT PGC-derived fibroblasts, further supporting viral expansion, even at 6hpi. None of the programmed cell death pathways (apoptosis, necroptosis or pyroptosis) were activated in the chIFIITM KO PGC-derived fibroblasts.

Connected with PI3K was Rho signalling, upregulated in WT PGC-derived fibroblasts. PI3K/AKT can inactivate Rho-GTPase via GRAF1 which ultimately leads to detachment of actin from the plasma membrane, facilitating virus entry. SV40 activates integrin signalling to make use of PI3K signalling in this way [216]. It is not clear that this would be the mechanism resulting in slight increase in virus yield in WT PGC- fibroblasts which we observed given SV40 is not an enveloped virus like influenza. However, it is not inconceivable that influenza could benefit from destabilised host membrane for its own entry via endosomes. In case of parainfluenza, Rho signalling deactivation was also correlated with viral growth [217]. Rho participates in complex signalling networks that influence an array of cellular processes involving the actin cytoskeleton making it difficult to unambiguously determine the mechanism by which it confers advantage to the virus.

Notably, pathways “Entry of Influenza virus into the host cell via Endocytosis” (R-HSA-168275.4) was significantly downregulated in KO relative to WT PGC-derived fiborblasts.. 20 molecules from the set containing 198 were identified in KO PGC-derived fibroblasts as contributing to enrichment of this biological process. Further analysis and research are needed to elucidate if downregulation of any of those molecules singularly confer the benefit to fight against the infection or it is a case of cumulative effect of all the molecules on aggregate that render the KO more resistant. It would also need to be clarified if the downregulation is in any way a consequence of the *chIFITM* KO or an independent event. Note that “Influenza” (R-HSA-168255) biological process was also only detected in DF-1 *chIFITM* KO PADOG analysis (4.3.2.2).

In agreement with the previous data, a stark difference is also membrane trafficking which is upregulated in KO PGC-derived fibroblasts but downregulated in the WT. “Vesicle biogenesis” is regulated in the opposite directions – down in the KO and up in the WT PGC-derived fibroblasts. It has been known that influenza uses cellular trafficking system to shuttle its genetic material between the nucleus and the plasma membrane. Yadav et al [156] established that it is the intermediate compartment Golgi-ER where the disruption happens because the virus provoked Golgi fragmentation. “Cargo concentration in the ER” and “Recognition for clathrin-mediated endocytosis” biological processes were upregulated in both the WT and the KO, indicating that the normal transport has been disturbed. A more granular analysis of these processes in the PGC-derived fibroblasts will provide information on the subtleties in differential regulation of membrane trafficking in PGC derived fibroblasts results in their different ability to support influenza infection.

To conclude, in this section I have shown that chIFITM KO PGC-derived fibroblasts can be successfully generated and grown in culture. They have low relative level of *chIFITM1* and *chIFITM3* mRNA in comparison to DF-1 as judged by RNA-seq read quantitation. Their transcriptomic profile is extremely affected by the *chIFITM* KO giving rise to changes in transcript composition that account for up to 70% of variability in the dataset compared to the WT PGC-derived fibroblasts. They can be readily infected with the influenza virus, but their ability to support viral replication is heavily dependent on the virus strain as shown by lack of success in growing the H5N3 subtype of influenza over a period of days to measure virus titre in PGC-derived fibroblasts.

Transcriptomic study demonstrated dramatic difference in gene expression profile between the WT and KO in basal condition, as proven by the differential expression analysis which returned more than 10 000 differentially expressed genes. Same was the case for every conducted comparison. To maintain objectivity in assessment what that means biologically, I focussed on pathway analysis and chose tests that preserve as much information as possible by not imposing thresholds for effect size or FDR value for genes to be included in the analysis and cross-checked the results in an attempt to extract the most robust pathways while benefitting from advantages of method/software. To deem a term enriched , p-value of 0.05 was used. This value had to be relaxed, except in the case of KO infected vs uninfected comparison to obtain any results. Untangling biological heterogeneity from change in expression caused by the infection was not straightforward and requires a higher number of replicates to ensure informative statistical power. Nevertheless, I highlight some pathways that were significant biologically and statistically for consideration in future studies.

6

Discussion

6.1 Concluding remarks

IFITMs are potent restriction factors that garner a lot of research interest due to their capacity to counteract the early effects of infection by many dangerous viral pathogens. Avian research into chIFITMs is heavily handicapped by the lack of suitable antibodies against the proteins of the IFITM family. To mitigate this hindrance, here I take the KO approach and evaluate the impact on the transcriptome level.

The only similar report is by Shen *et al* where they identify differentially regulated processes after knock-down of endogenous *IFITM3* in HeLa using microarrays [161]. I conducted an RNA-Seq study on the DF-1 and PGC-derived fibroblasts of the WT and *chIFTM* KO background infected with H5N3 subtype of influenza virus at 6h post infection, i.e. a comparison is not like-for-like. Nevertheless, some parallels could be drawn between differential expression analyses in the control condition. Both studies identified dysregulation in MAPK pathway, immune-regulatory interactions, calcium signalling in non-infected condition.

Here we generated unique reagents for the research community in form of characterised stable cell lines available for further research. Overarching goal was to compare the effect of the chIFITM knock-out in two genetically different backgrounds. Aided by the recently sequenced chicken IFITM locus [49] this work shows it is possible to genetically engineer using CRISPR two very distinct chicken systems: an established, widely used laboratory cell line and a primordial germ cell. Covering these extremes, I capture changes brought upon by the loss of chIFITM expression in a very metabolically altered system which originated by transformation of primary cells: DF-1, and a more natural, arguably more relevant system that has not gone through the stresses of a long term cell culture and which serves as a prototype for an animal model potentially used in future research: PGC.

I was able to identify typical hallmarks of WT- DF-1 such as high levels of basal expression of SOCS1 [120] and its downregulation in infection. Individual characterisation for each WT DF-1 clone was challenging because statistical inference from a duplicate sample is not reliable. Even though the effect we observed in the laboratory was pronounced, reasons for it were not unambiguously detectable on the transcriptomic level. It is possible that a statistically stronger experiment (bigger sample size) would allow detection of more informative changes in gene expression. Nevertheless, with a combination of analytical approaches, I detected *CMPK2* and *ERBB3* as potential targets to investigate further.

I have shown that the knock-out only marginally affects permissiveness to the influenza virus of cell lines tested. This indicated that when a high viral yield is observed, the causal phenomena must be additionally searched downstream of the virus entry. It is likely that the high viral titre is a result of various cellular and metabolic behaviours that are disturbed by the virus and act in synergy for its benefit [218–220].

Wang et al [221] established that the library size of 30M of 75bp reads is enough to detect all chicken annotated genes. Our experiments demonstrate the benefit of accounting for infection during initial experimental planning and ensuring that the remaining sequencing real-estate when viral mRNA reads are subtracted from total be above that threshold. Number of replicates should also be carefully considered. Normally, in research three replicates are considered an absolute minimum. RNA-seq is able to collect information on thousands of genes simultaneously from, habitually, a very small number of samples. This results in a high signal-to-noise ratio. This challenge becomes even greater when the task is to identify differences in transcript levels among very similar conditions, or when the replicates have very heterogeneous patterns of expression (e.g. patient data). In such experimental conditions discerning biological noise (heterogeneity) and technical noise (imprecision of measurement) from authentic signal becomes a non-trivial task. Ensuring a sufficient number of replicates is a better approach toward strengthening the experimental design than adding sequencing depth.

Given the size of the datasets produced and the time limitations of this thesis, I took an exploratory approach and conducted a high-level analysis to investigate changes in metabolism and the immune response in the absence of *chIFITM* locus at the pathway level.

In summary – this work set out to investigate the downstream effects of the *chIFITM* KO in two systems: an established cell line and a primary cell line; and on two levels: on the granular level of transcriptomic changes and on the high level of the cellular permissivity as well as virus yield. We show robustly that the consequences of a complete or partial ablation of chIFITMs has significant and permanent consequences for the measured parameters, especially so for DF-1 fibroblasts. This finding could be further confirmed by additional growth kinetics experiments for DF-1 chIFITM KO clonal cell lines.

Despite the inability to confirm full KO in the PGC-derived fibroblasts, we show the levels of mRNA were severely reduced. That perturbation in gene expression caused a downstream effect on the physiological level making the KO more refractory to infection as judged by the flow cytometry. However, unambiguous proof of loss of function would require a generation of a complete knock-out as well as subsequent rescue experiments. Nevertheless, these results do point towards alternative functions and interactions of IFITMs that are still to be elucidated.

We also indirectly quantify the inherent heterogeneity of the parental DF-1 cell line via transcriptomic analysis of its clonal populations. The study generated a well powered transcriptomic dataset that should be mined for clonal differences among the cell lines to generate hypotheses of the origins of phenotypic differences observed in the laboratory and more quantitively assess the transcriptomic changes. Albeit small, they could be important. Due to time limitations of this project, the analysis was done on the pathway level to determine more wide-spanning downstream changes in biological processes. Detailed statistical analysis using two different approaches was conducted to uncover which pathways are disrupted either by the KO or by the infection. Thorough and flexible experimental design will allow for more questions to be answered than what was possible during this project.

6.2 Future directions

In the pilot experiment described in Chapter 3 it was shown that transcriptional profile of WT DF-1 608 cell line was very similar to the overall WT DF-1 profile (see **Figure 109**). To uncover which genes enable the support of the high virus titre, a more thorough analysis would be required. This knowledge would be very useful in cell line engineering for development of cell lines used for candidate vaccine production.

Even though PGC-derived fibroblasts were permissive to the influenza virus (see 5.3.2.2) they were not able to support multi-cycle replication of the H5N3 virus (**Figure 75**) due to in part their sensitivity to TPCK-trypsin. Detailed investigation into H5N3 NS1 may provide insight what strategy is the virus employing and why is it successful in DF-1 fibroblasts and not in the PGC-derived fibroblasts. Level of expression of *chIFITM* genes did not rise dramatically in PGC-derived fibroblast upon infection with H5N3 (**Figure 70**). PR8 (H1N1) grew to high titres in both WT and KO PGC-derived fibroblasts (**Figure 76**) giving rise to the hypothesis that chIFITMs potentially do not contribute significantly to the immune response in these cells. It would be informative to test more viruses and/or viral strains to verify whether the phenomenon we observed here is ubiquitous for PGC-derived fibroblasts. It would also be beneficial to check if their IFN pathway is functional by stimulating them artificially and measuring ISG expression. Given that I was not able to unambiguously prove bi allelic deletion of the locus in PGC-derived fibroblasts, it is possible that leaky expression, albeit undetectable, impedes the observation of true KO effect. A new attempt at a knockout could be a possible avenue to address this problem.

To unambiguously prove the function of IFITM in their capacity of restriction factors or otherwise, it would also be necessary to restore the function via use of knock-in clones, of individual genes or the whole chIFITM family. This has proven difficult in the past due to the difficulties to sequence the region as one of many reasons.

Another informative comparison would be to do an overlap of differentially regulated genes in both models in corresponding conditions, be that infection or native conditions to uncover how similar is the response on the deletion of the knock-out and how the differentially expressed genes translate into differentially regulated pathways in those conditions.

Transcriptomic studies frequently face difficulty in interpreting the results uniformly and cite the heterogeneity of the cell line as one of the limiting factors [222]. This work arguably produced more questions than answers and the analysis uncovered many candidate processes to investigate further with functional assays.

7

Appendix

7.1 RNA QC

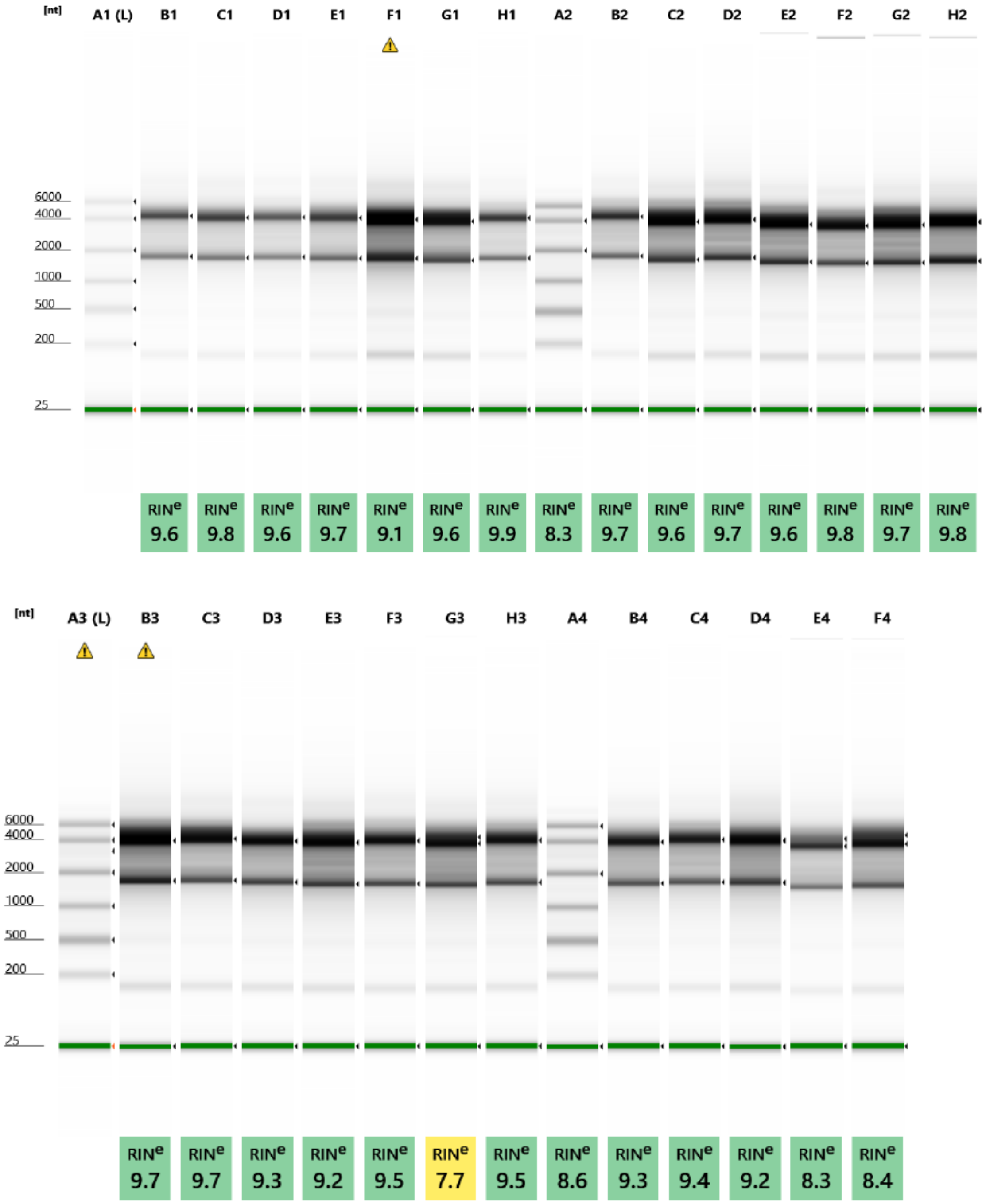


Figure 98 RNA quality as assessed by TapeStation 2200 RNA samples were mixed with buffer and processed on the TapeStation 2200 using RNA Screen tape. The RNA quantitation and image generation was done using TapeStation Analysis Software A.02.02

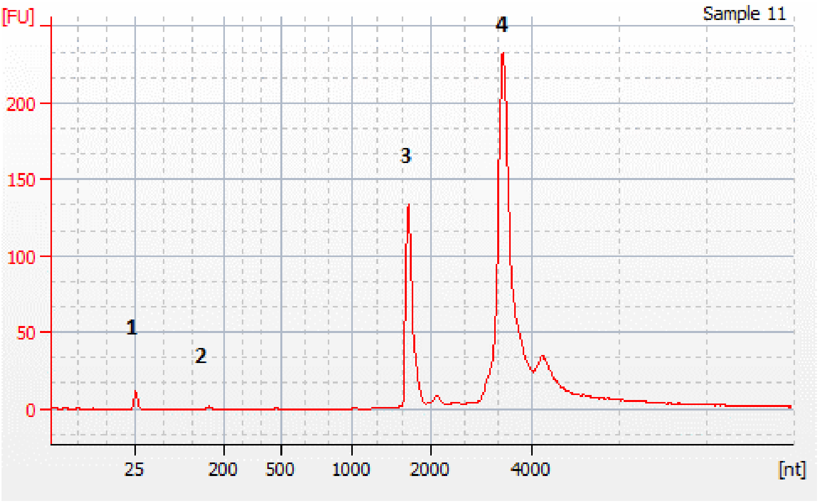


Figure 99 Representative example of electrophoretic histogram generated by Bioanalyzer 2100 Relative abundance is assessed using fluorescence intensity and represents relative abundance; plotted on y axis. Time is plotted on x axis. Numbers denote characteristic bands. **1.** 25 nucleotide marker; **2.** mRNA; **3.** 18S ribosomal RNA; **4.** 28S ribosomal RNA

7.2 Sequencing QC

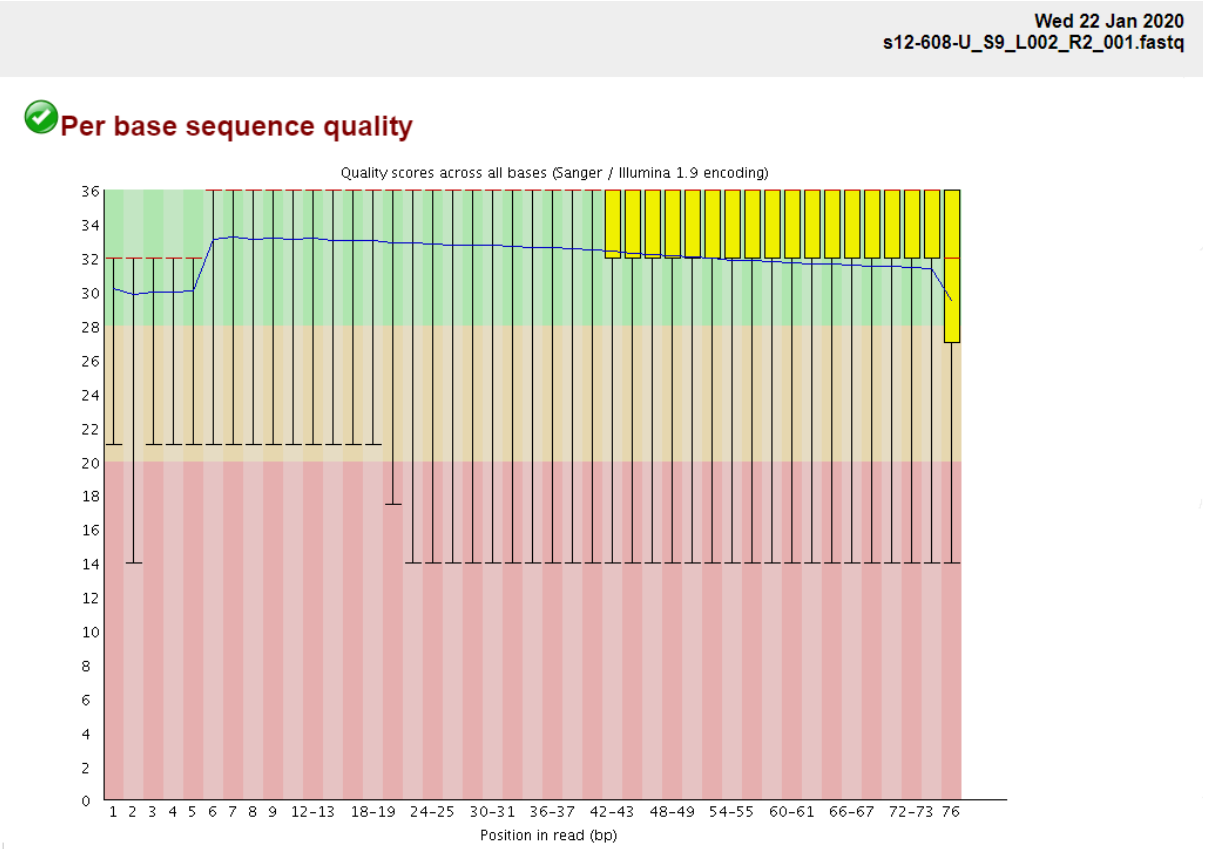


Figure 100 Example of a QC analysis by FastQC Population of calls aligning to each position in a read is plotted on the x-axis, with a box whisker plot. Median value is marked by the red line; inter-quartile range is represented by the yellow box; and upper and lower whiskers are demarcating the 10% and 90% points. Blue line represents the mean quality. The interpretation is helped by partitioning the y-axis, showing the quality scores, into green part (very good quality calls), orange (reasonable quality calls) and red (calls of poor quality).

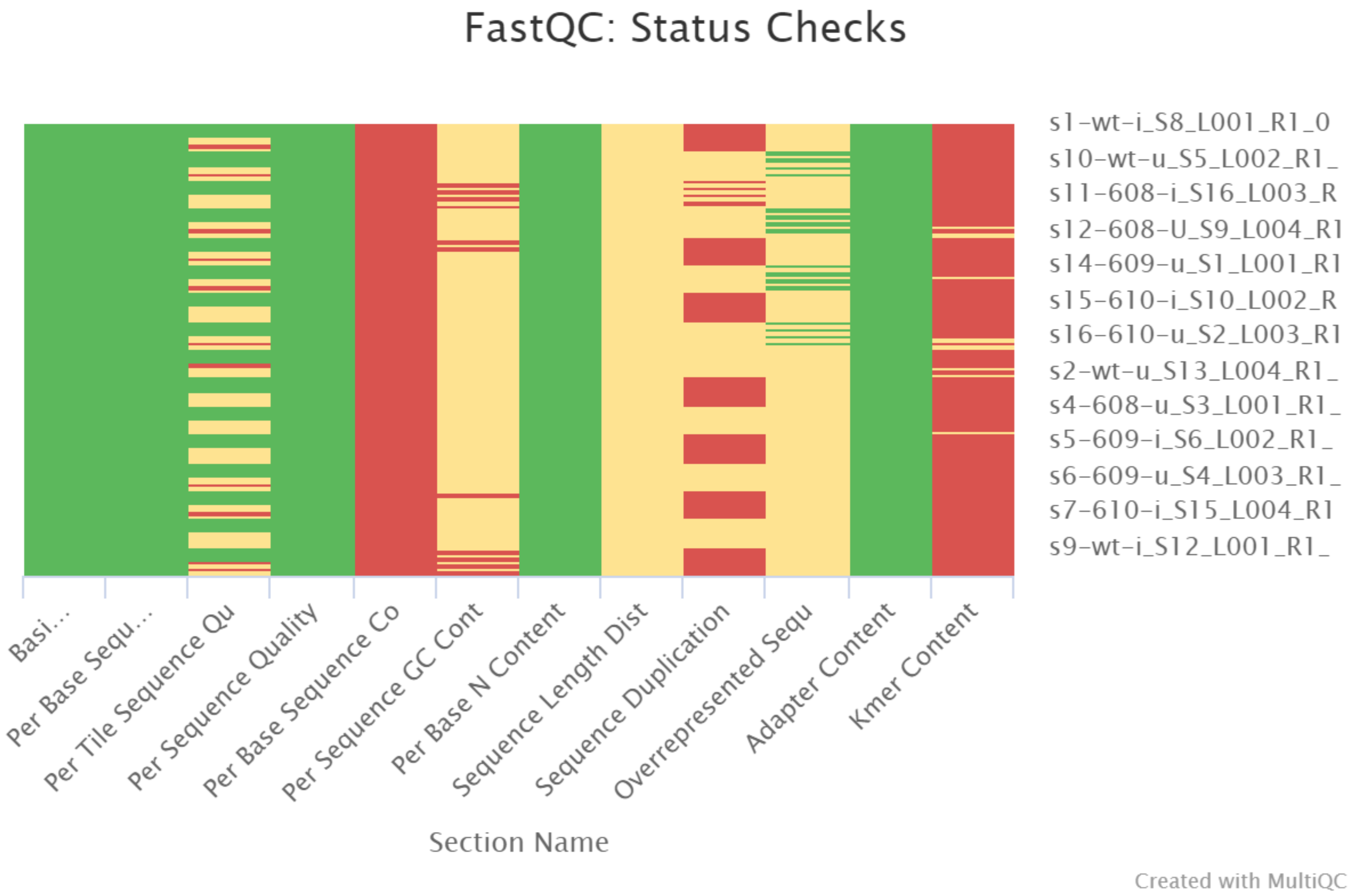


Figure 101 MultiQC summarisation of FastQC reports MultiQC [92]was used to summarise all FastQC [84] reports into one graphic. Note that the FastQC software is optimised for assessment of quality of NGS data, not RNA-Seq data. Thus, the parameters of concern, flagged red, such as Sequence duplication or Overrepresented sequences are an expected phenomenon in RNA-Seq data. Repetitive pattern observed is a reflection of reads stemming from viral mRNA detected in infected samples, which show extremely high level of duplication because they are preferentially amplified by the cellular machinery as well as the sequencing process.

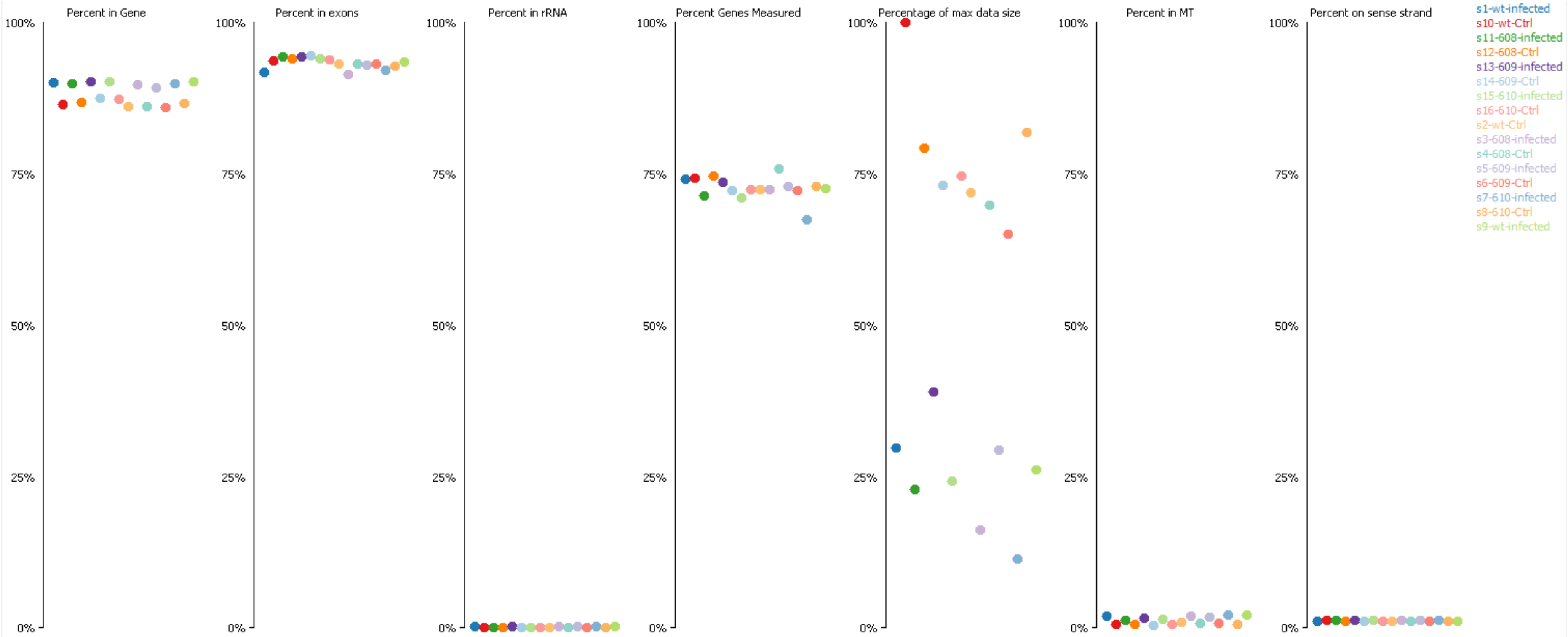


Figure 102 WT DF-1 RNA-Seq quality check as assessed by SeqMonk BAM files generated by STAR for the WT DF-1 data set were used as input for SeqMonk. RNA-Seq QC pipeline was used on uniquely mapped reads (as indicated by STAR with value 255). Library sizes are normalised relative to the sample with highest number of reads which was then set to 100%.

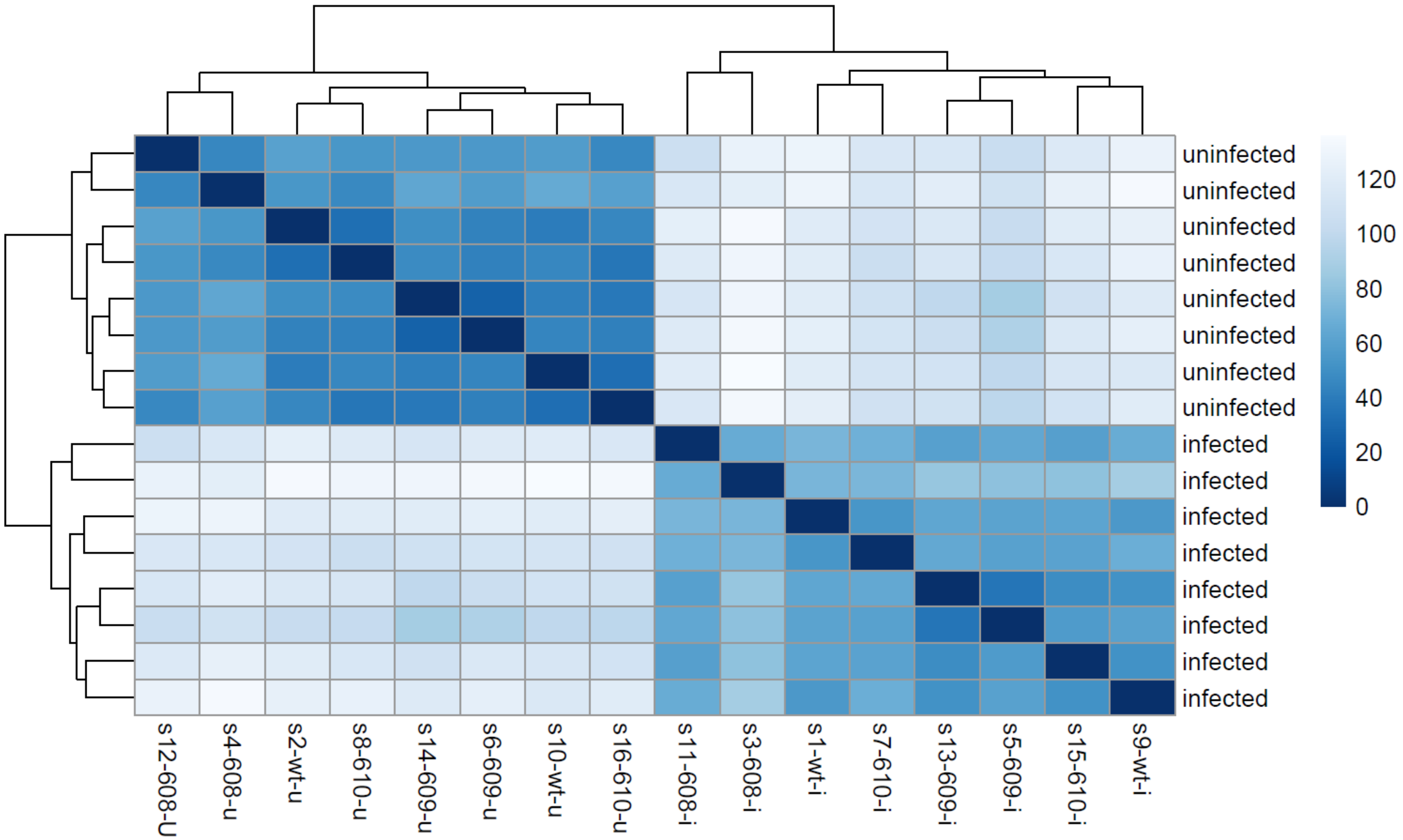


Figure 103 Sample distances clustering for WT DF-1 dataset Prepared RNA-seq data were normalised using DESeq2 and transformed using vst (variance stabilising transformation). Resulting values were used as Euclidian distances for unsupervised hierarchical clustering. Legend shows increasing measure of disimilarity, from dark (most similar) to light (most dissimilar).

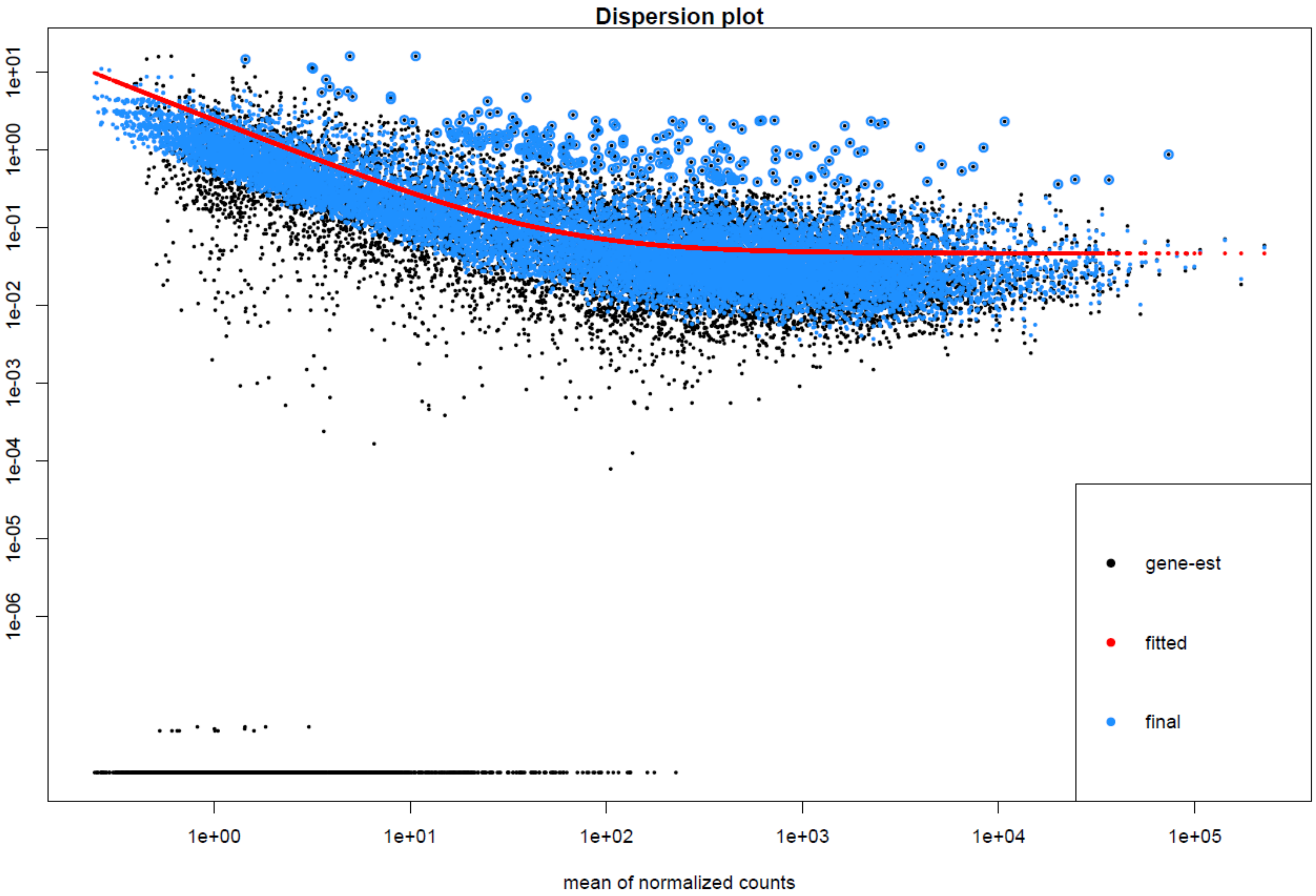


Figure 104 Gene-wise dispersion estimates for WT DF-1 data set Mean of normalised counts for each gene is plotted on the x-axis and the corresponding dispersion on the y-axis. Gene-wise estimates of dispersion (in black) are shrunk towards the fitted (red line) dispersion estimates, resulting in final estimates (in blue). Circled values have not been shrunk towards the fitted value (when the value deviates a lot from the mean of normalised counts due to high variance. This trend is typical demonstrating high variance/dispersion in the region of low counts and its stabilisation towards the region of high counts.

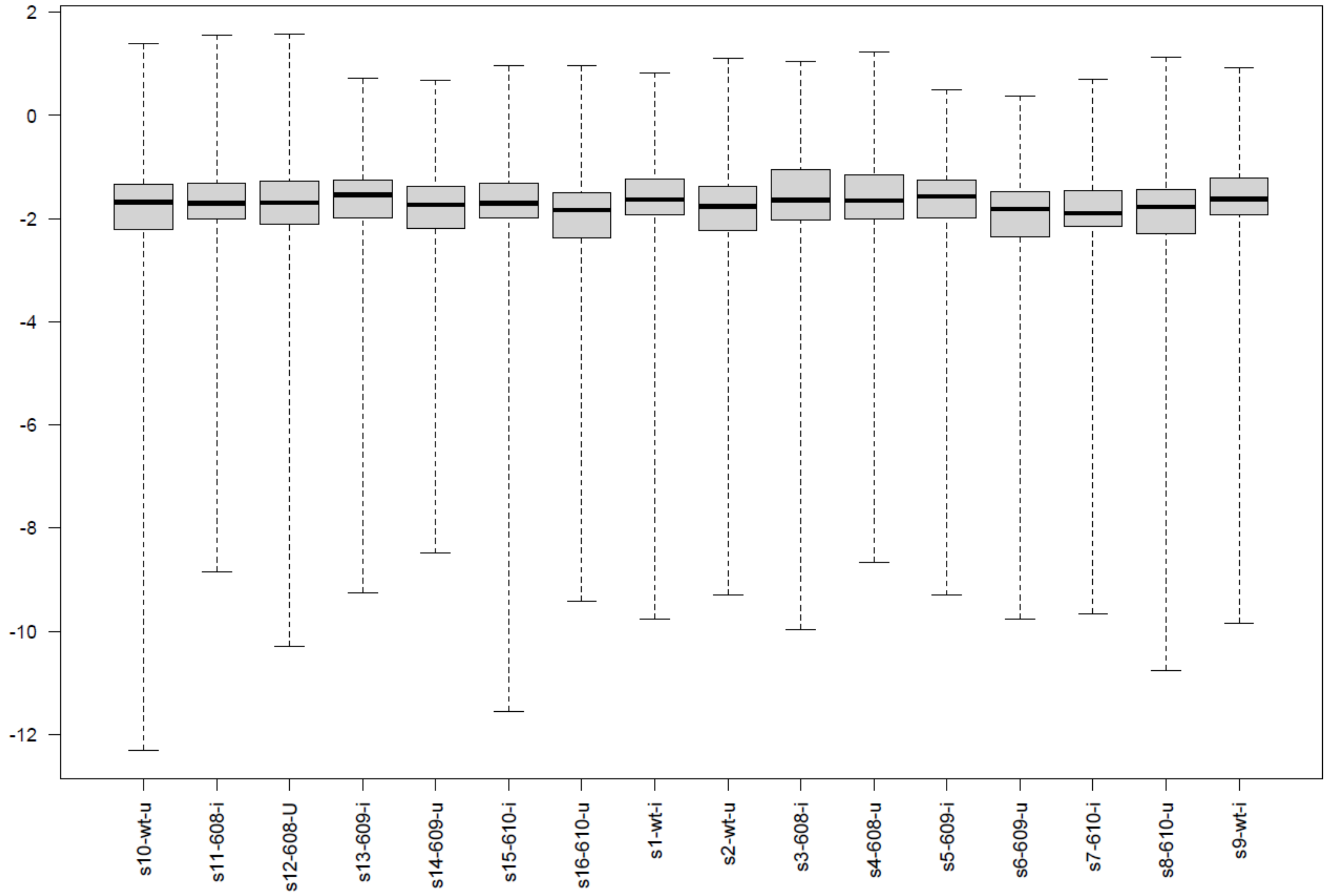


Figure 105 No sample outliers are detected by Cook's distance for WT DF-1 data set

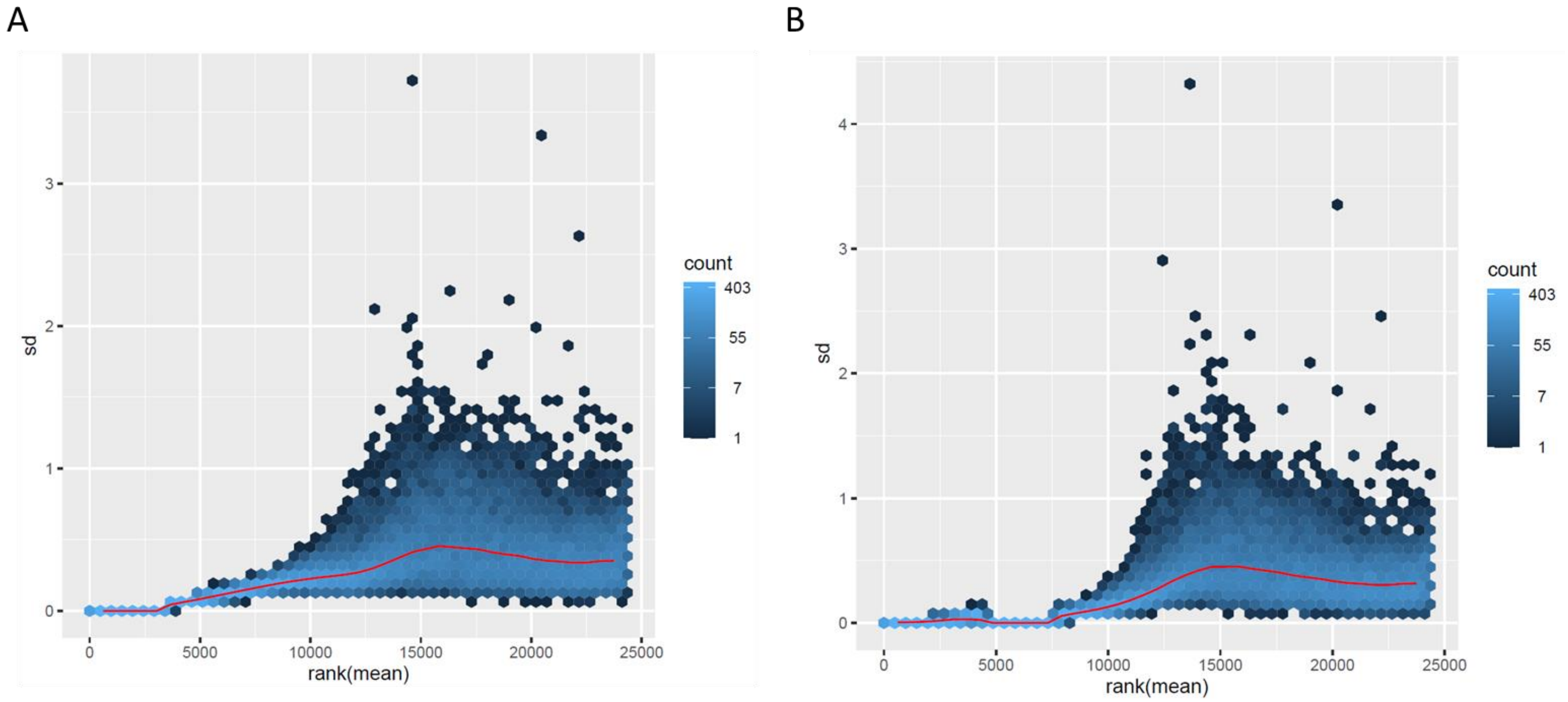


Figure 106 KO DF-1 RNA-Seq data transformation Raw RNA-Seq counts were transformed using vst (panel A) and rlog (panel B) transformations to stabilise the variance. Estimator is depicted by the red line. Counts are represented by ranks and plotted on the x-axis. Standard deviation is plotted on the y-axis.

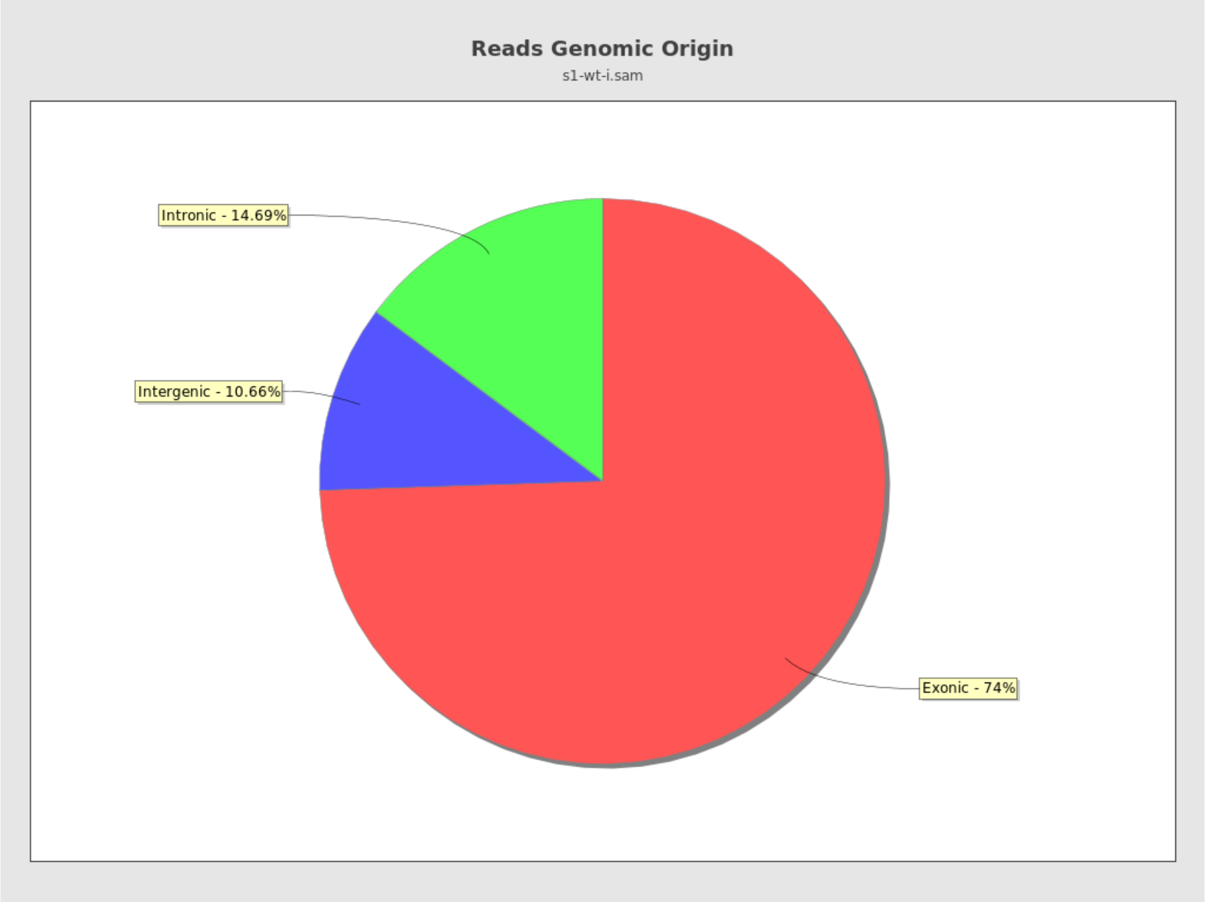


Figure 107 Quality control of the mapping step conducted by STAR as assessed by Qualimap BAM_QC feature of Qualimap v.2.2.1 (2016) was used to assess the quality of the mapping performed by STAR. Note that for RNA-seq data, >60% and above mapping to exonic regions is considered an acceptable accuracy for a reasonably well annotated genome (model organisms).

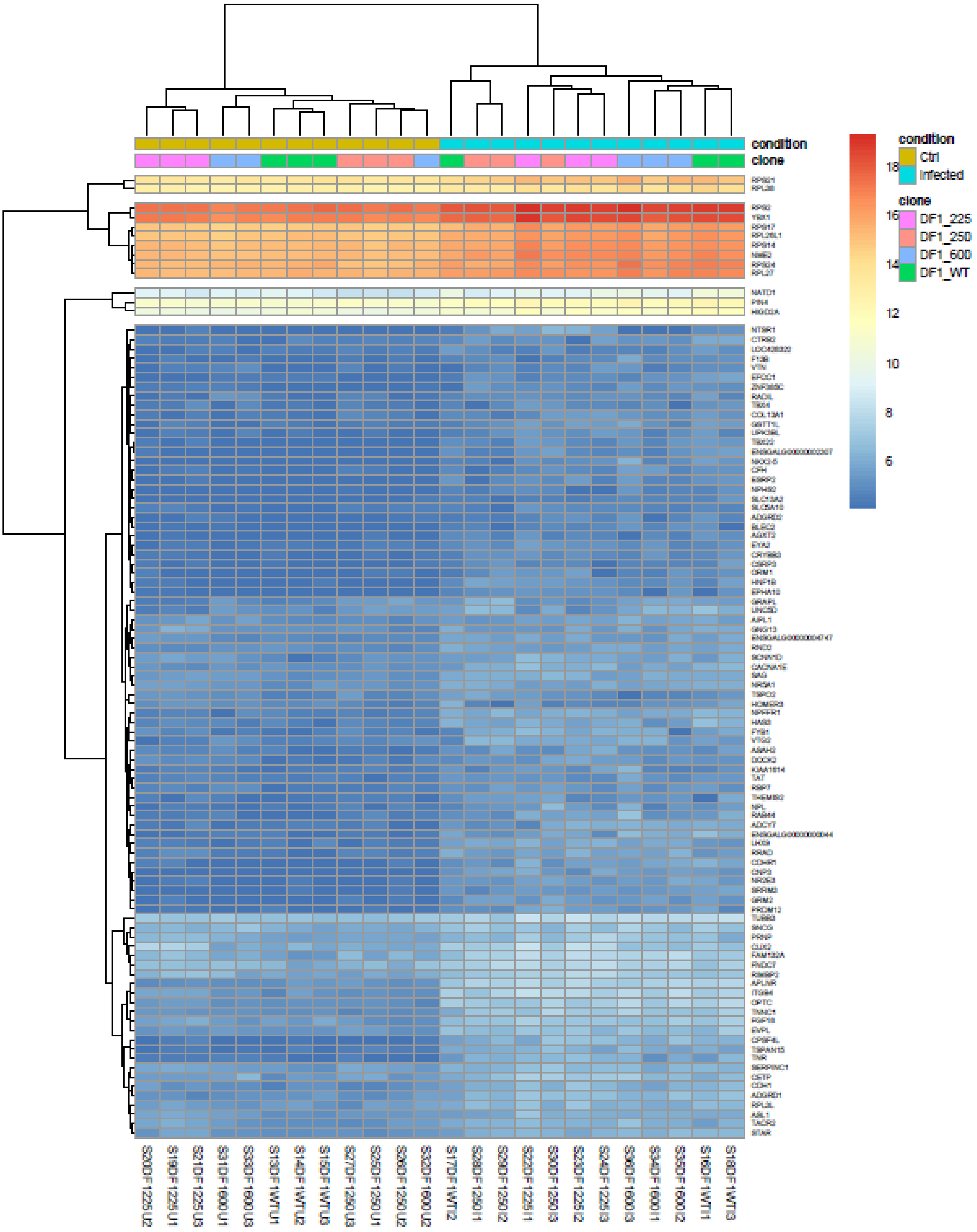


Figure 108 Heatmap of vst transformed counts for 100 most upregulated genes in WT-DF1 pairwise comparison infected vs uninfected Raw RNA-Seq counts for the 100 most upregulated genes by fold change were *vst* transformed and clustered using "manhattan" distances in pheatmap packages in R. Cells are coloured in a range of colours from blue to red indicating increasing values.

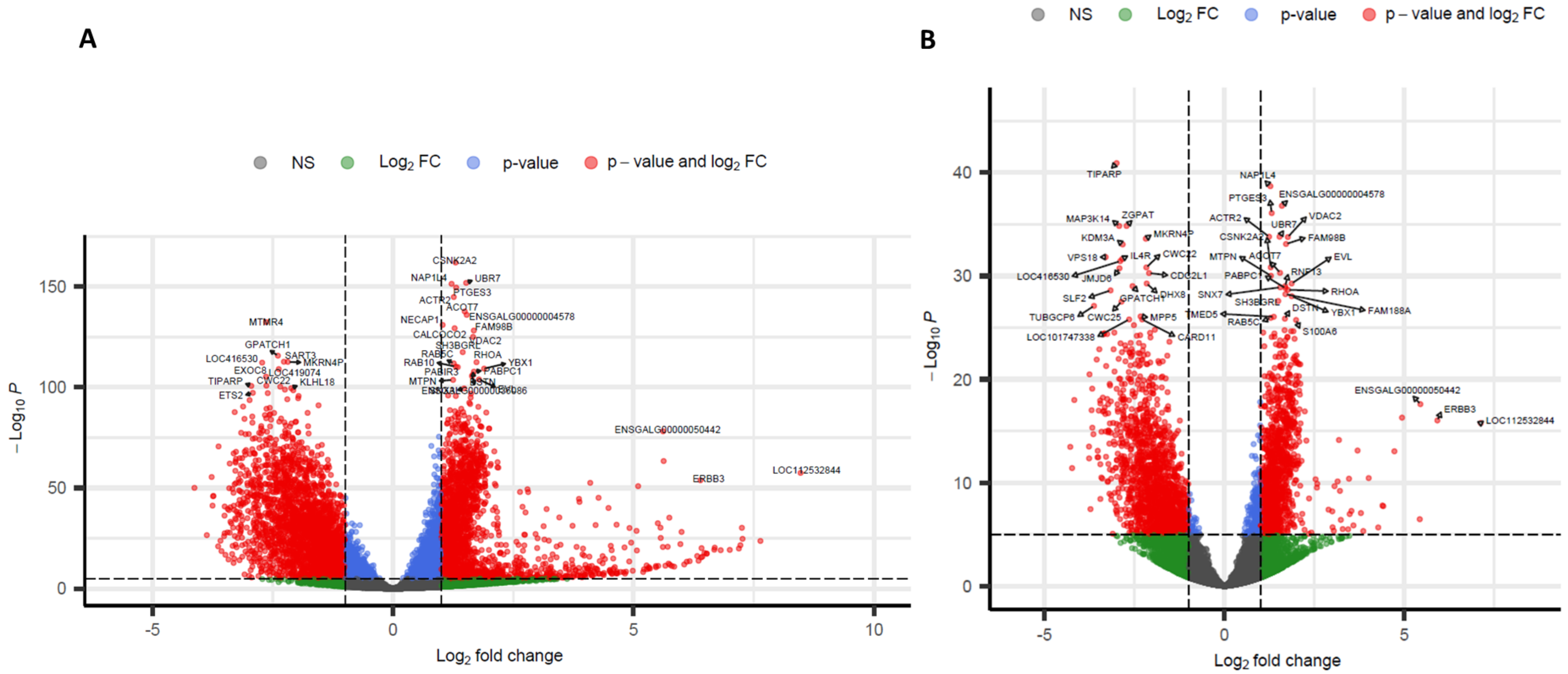


Figure 109 Volcano plots of differentially expressed genes in WT DF1 cell lines Differential expression was conducted in R with package *DESeq2* to compare infected and uninfected conditions in WT DF-1 cell lines. Panel A shows comparison between the two conditions when all WT clones are pooled together. Panel B shows individual comparison of uninfected vs infected for WT DF1 clonal cell line 608 only. Marked in red are genes which are deemed statistically significant at FDR 0.05 and with a FoldChange greater than 2 ($\log_2$ >=1). Visualisation was produced with *Enhanced Volcano* in R.

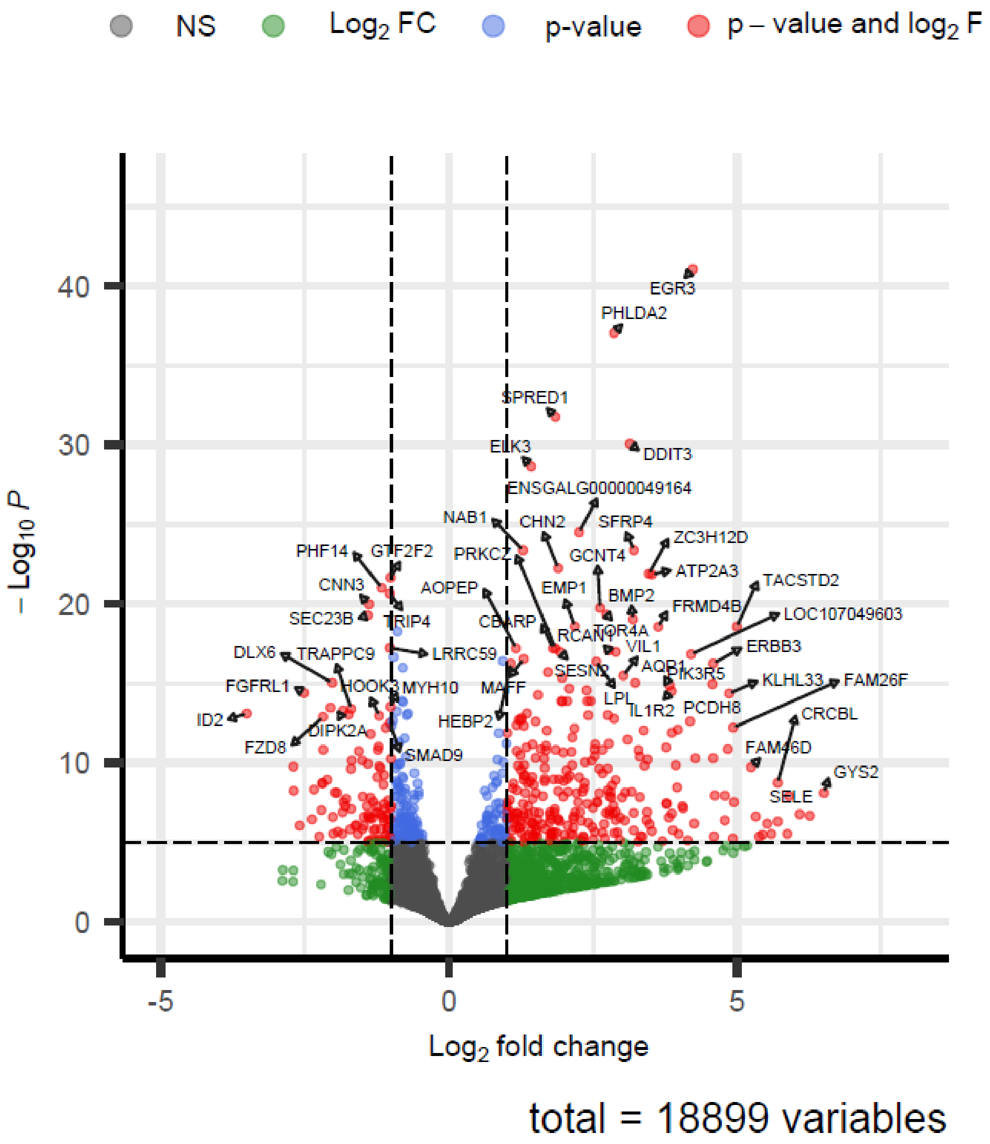


Figure 110 Volcano plot of differentially expressed genes in WT PGC-derived fibroblasts in influenza infection Marked in red are genes which are deemed statistically significant at FDR 0.05 and with a FoldChange greater than 2 (log$_2$ >=1). Visualisation was produced with *Enhanced Volcano* in R.

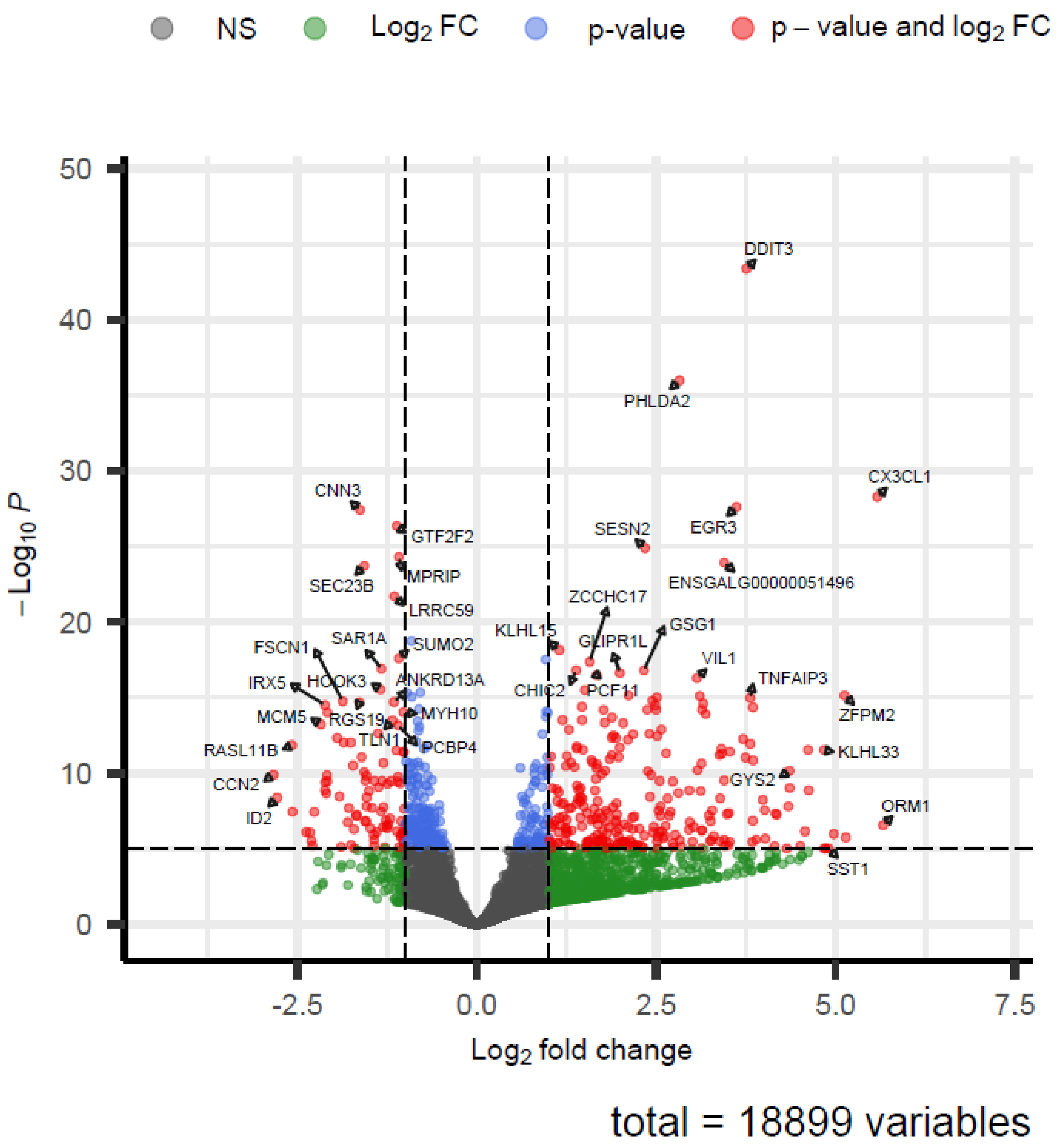


Figure 111 Differentially expressed gene in chIFITM KO PGC-derived fibroblasts in influenza infection Marked in red are genes which are deemed statistically significant at FDR 0.05 and with a FoldChange greater than 2 (log_2 >=1). Visualisation was produced with Enhanced Volcano in R

7.3 R packages

```
sessionInfo()

R version 4.1.1 (2021-08-10)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19044)


Matrix products: default

 [1] LC_COLLATE=English_United Kingdom.1252  LC_CTYPE=English_United Kingdom.1252

[3] LC_MONETARY=English_United Kingdom.1252 LC_NUMERIC=C

[5] LC_TIME=English_United Kingdom.1252


attached base packages:

 [1] stats4    stats     graphics  grDevices utils     datasets  methods   base


other attached packages:

 [1] EnhancedVolcano_1.12.0     ggrepel_0.9.1             RColorBrewer_1.1-3

[4] gplots_3.1.3              genefilter_1.76.0         org.Gg.eg.db_3.14.0

[7] AnnotationDbi_1.56.2       pheatmap_1.0.12           ashr_2.2-54

[10] vsn_3.62.0                ggplot2_3.3.6             dplyr_1.0.9

[13] DESeq2_1.34.0             SummarizedExperiment_1.24.0 Biobase_2.54.0

[16] MatrixGenerics_1.6.0       matrixStats_0.61.0        GenomicRanges_1.46.1

[19] GenomeInfoDb_1.30.0        IRanges_2.28.0            S4Vectors_0.32.3

[22] BiocGenerics_0.40.0


loaded via a namespace (and not attached):

 [1] ggbeeswarm_0.6.0        colorspace_2.0-2        ellipsis_0.3.2          XVector_0.34.0

[5] rstudioapi_0.14         farver_2.1.1            affyio_1.64.0           bit64_4.0.5

[9] fansi_1.0.3             xml2_1.3.3              splines_4.1.1           sparseMatrixStats_1.6.0

[13] extrafont_0.18          cachem_1.0.6            geneplotter_1.72.0      Rttf2pt1_1.3.10

[17] annotate_1.72.0         dbplyr_2.2.1            png_0.1-7               BiocManager_1.30.18
```

```
[21] compiler_4.1.1        httr_1.4.4             dqrng_0.3.0            assertthat_0.2.1
[25] Matrix_1.3-4          fastmap_1.1.0          limma_3.50.3           cli_3.3.0
[29] BiocSingular_1.10.0   prettyunits_1.1.1      tools_4.1.1            rsvd_1.0.5
[33] gtable_0.3.0          glue_1.6.2             GenomeInfoDbData_1.2.7 affy_1.72.0
[37] reshape2_1.4.4        maps_3.4.0             rappdirs_0.3.3         Rcpp_1.0.7
[41] vctrs_0.4.1           Biostrings_2.62.0      ggalt_0.4.0            preprocessCore_1.56.0
[45] extrafontdb_1.0       DelayedMatrixStats_1.16.0 stringr_1.4.1         beachmat_2.10.0
[49] lifecycle_1.0.1       irlba_2.3.5            gtools_3.9.3           PCAtools_2.6.0
[53] XML_3.99-0.8          MASS_7.3-54            zlibbioc_1.40.0        scales_1.2.1
[57] hms_1.1.2             proj4_1.0-11           parallel_4.1.1         curl_4.3.2
[61] memoise_2.0.1         ggrastr_1.0.1          biomaRt_2.50.3         stringi_1.7.6
[65] RSQLite_2.2.9         SQUAREM_2021.1         ScaledMatrix_1.2.0     caTools_1.18.2
[69] filelock_1.0.2        BiocParallel_1.28.2    truncnorm_1.0-8        rlang_1.0.4
[73] pkgconfig_2.0.3       bitops_1.0-7           lattice_0.20-45        invgamma_1.1
[77] purrr_0.3.4           labeling_0.4.2         cowplot_1.1.1          bit_4.0.4
[81] tidyselect_1.1.2      plyr_1.8.7             magrittr_2.0.1         R6_2.5.1
[85] generics_0.1.3        DelayedArray_0.20.0    DBI_1.1.3              pillar_1.8.1
[89] withr_2.5.0           ash_1.0-15             survival_3.2-11        KEGGREST_1.34.0
[93] RCurl_1.98-1.5        mixsqp_0.3-43          tibble_3.1.8           crayon_1.5.1
[97] KernSmooth_2.23-20    utf8_1.2.2             BiocFileCache_2.2.1    progress_1.2.2
[101] locfit_1.5-9.6       grid_4.1.1             blob_1.2.3             digest_0.6.29
[105] xtable_1.8-4         munsell_0.5.0          beeswarm_0.4.0         vipor_0.4.5
```

8 Bibliography

1. Bieniasz PD. Intrinsic immunity: a front-line defense against viral attack. *Nat Immunol* 2004;**5**:1109–15.

2. Yan N, Chen ZJ. Intrinsic antiviral immunity. *Nature Immunology 2012 13:3* 2012;**13**:214–22.

3. Bruce Alberts Julian Lewis Martin Raff Keith Roberts Peter Walter AJ. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science, 2002.

4. Takeuchi O, Akira S. Innate immunity to virus infection. *Immunol Rev* 2009;**227**:75–86.

5. Kumar H, Kawai T, Akira S. Pathogen Recognition by the Innate Immune System. *Int Rev Immunol* 2011;**30**:16–34.

6. Loo Y-M, Gale Jr. M. Immune signaling by RIG-I-like receptors. *Immunity* 2011;**34**:680–92.

7. Sabbah A, Chang TH, Harnack R *et al.* Activation of innate immune antiviral responses by Nod2. *Nat Immunol* 2009;**10**:1073–80.

8. Satoh T, Kato H, Kumagai Y *et al.* LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proc Natl Acad Sci U S A* 2010;**107**:1512–7.

9. Gack MU, Shin YC, Joo CH *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* 2007;**446**:916–20.

10. Oshiumi H, Miyashita M, Inoue N *et al.* The ubiquitin ligase Riplet is essential for RIG-I-dependent innate immune responses to RNA virus infection. *Cell Host Microbe* 2010;**8**:496–509.

11. Kuniyoshi K, Takeuchi O, Pandey S *et al.* Pivotal role of RNA-binding E3 ubiquitin ligase MEX3C in RIG-I-mediated antiviral innate immunity. *Proc Natl Acad Sci U S A* 2014;**111**:5646–51.

12. da Costa LS, Outlioua A, Anginot A *et al.* RNA viruses promote activation of the NLRP3 inflammasome through cytopathogenic effect-induced potassium efflux. *Cell Death Dis* 2019;**10**:346.

13. Karpala AJ, Stewart C, McKay J *et al.* Characterization of chicken Mda5 activity: regulation of IFN-beta in the absence of RIG-I functionality. *J Immunol* 2011;**186**:5397–405.

14. Hayashi T, Watanabe C, Suzuki Y *et al.* Chicken MDA5 senses short double-stranded RNA with implications for antiviral response against avian influenza viruses in chicken. *J Innate Immun* 2014;**6**:58–71.

15. Santhakumar D, Rubbenstroth D, Martinez-Sobrido L *et al.* Avian Interferons and Their Antiviral Effectors. *Front Immunol* 2017;**8**:49.

16. Magor KE, Miranzo Navarro D, Barber MRW *et al.* Defense genes missing from the flight division. *Dev Comp Immunol* 2013;**41**:377–88.

17. Goossens KE, Ward AC, Lowenthal JW *et al.* Chicken interferons, their receptors and interferon-stimulated genes. *Dev Comp Immunol* 2013;**41**:370–6.

18. Au-Yeung N, Horvath CM. Transcriptional and chromatin regulation in interferon and innate antiviral gene expression. *Cytokine Growth Factor Rev* 2018;**44**:11–7.

19. Marié I, Durbin JE, Levy DE. Differential viral induction of distinct interferon-alpha genes by positive feedback through interferon regulatory factor-7. *EMBO J* 1998;**17**:6660–9.

20. Schoggins JW. Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol* 2019;**6**:567–84.

21. Rajsbaum R, Albrecht RA, Wang MK *et al.* Species-Specific Inhibition of RIG-I Ubiquitination and IFN Induction by the Influenza A Virus NS1 Protein. *PLoS Pathog* 2012;**8**:e1003059.

22. Ludwig S, Wang X, Ehrhardt C *et al.* The influenza A virus NS1 protein inhibits activation of Jun N-terminal kinase and AP-1 transcription factors. *J Virol* 2002;**76**:11166–71.

23. Gaucherand L, Porter BK, Levene RE *et al.* The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression. *Cell Rep* 2019;**27**:776-792.e7.

24. Varga ZT, Grant A, Manicassamy B *et al.* Influenza Virus Protein PB1-F2 Inhibits the Induction of Type I Interferon by Binding to MAVS and Decreasing Mitochondrial Membrane Potential. *J Virol* 2012;**86**:8359.

25. JIGGINS FM, KIM KW. A screen for immunity genes evolving under positive selection in Drosophila. *J Evol Biol* 2007;**20**:965–70.

26. Jopling CL, Yi M, Lancaster AM *et al.* Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science (1979)* 2005;**309**:1577–81.

27. Takata MA, Goncalves-Carneiro D, Zang TM *et al.* CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* 2017;**550**:124–7.

28. Mao R, Nie H, Cai D *et al.* Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog* 2013;**9**:e1003494.

29. Goossens KE, Karpala AJ, Ward A *et al.* Characterisation of chicken ZAP. *Dev Comp Immunol* 2014;**46**:373–81.

30. Pindel A, Sadler A. The role of protein kinase R in the interferon response. *J Interferon Cytokine Res* 2011;**31**:59–70.

31. Ko JH, Asano A, Kon Y *et al.* Characterization of the chicken PKR: polymorphism of the gene and antiviral activity against vesicular stomatitis virus. *Jpn J Vet Res* 2004;**51**:123–33.

32. Daviet S, van Borm S, Habyarimana A *et al.* Induction of Mx and PKR Failed to Protect Chickens from H5N1 Infection. *Viral Immunol* 2009;**22**:467–72.

33. Choi UY, Kang J-S, Hwang YS *et al.* Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp Mol Med* 2015;**47**:e144–e144.

34. Rong E, Wang X, Chen H *et al.* Molecular Mechanisms for the Adaptive Switching Between the OAS/RNase L and OASL/RIG-I Pathways in Birds and Mammals. *Front Immunol* 2018;**9**, DOI: 10.3389/FIMMU.2018.01398.

35. Tag-El-Din-Hassan HT, Sasaki N, Moritoh K *et al.* The chicken 2’-5’ oligoadenylate synthetase A inhibits the replication of West Nile virus. *Jpn J Vet Res* 2012;**60**:95–103.

36. Gizzi AS, Grove TL, Arnold JJ *et al.* A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature* 2018;**558**:610–4.

37. Wang X, Hinson ER, Cresswell P. The interferon-inducible protein viperin inhibits influenza virus release by perturbing lipid rafts. *Cell Host Microbe* 2007;**2**:96–105.

38. Goossens KE, Karpala AJ, Rohringer A *et al.* Characterisation of chicken viperin. *Mol Immunol* 2015;**63**:373–80.

39. Friedman RL, Manly SP, McMahon M *et al.* Transcriptional and posttranscriptional regulation of interferon-induced gene expression in human cells. *Cell* 1984;**38**:745–55.

40. Brass AL, Huang IC, Benita Y *et al.* IFITM Proteins Mediate the Innate Immune Response to Influenza A H1N1 Virus, West Nile Virus and Dengue Virus. *Cell* 2009;**139**:1243–54.

41. Everitt AR, Clare S, Pertel T *et al.* IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* 2012;**484**:519–23.

42. Smith SE, Gibson MS, Wash RS *et al.* Chicken Interferon-Inducible Transmembrane Protein 3 Restricts Influenza Viruses and Lyssaviruses In Vitro. *J Virol* 2013;**87**:12957–66.

43. Smith SE, Busse DC, Binter S *et al.* Interferon-induced Transmembrane Protein 1 restricts replication of virus that enter cells via the plasma membrane. *J Virol* 2018, DOI: 10.1128/jvi.02003-18.

44. Huang IC, Bailey CC, Weyer JL *et al.* Distinct Patterns of IFITM-Mediated Restriction of Filoviruses, SARS Coronavirus, and Influenza A Virus. *PLoS Pathog* 2011;**7**:e1001258.

45. Anafu AA, Bowen CH, Chin CR *et al.* Interferon-inducible transmembrane protein 3 (IFITM3) restricts reovirus cell entry. *J Biol Chem* 2013;**288**:17261–71.

46. Sällman Almén M, Bringeland N, Fredriksson R *et al.* The Dispanins: A Novel Gene Family of Ancient Origin That Contains 14 Human Members. *PLoS One* 2012;**7**:e31961.

47. Yount JS, Karssemeijer RA, Hang HC. S-palmitoylation and ubiquitination differentially regulate interferon-induced transmembrane protein 3 (IFITM3)-mediated resistance to influenza virus. *J Biol Chem* 2012;**287**:19631–41.

48. Yount JS, Moltedo B, Yang YY *et al.* Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nat Chem Biol* 2010;**6**:610–4.

49. Bassano I, Ong SH, Lawless N *et al.* Accurate characterization of the IFITM locus using MiSeq and PacBio sequencing shows genetic variation in Galliformes. *BMC Genomics* 2017;**18**:419.

50. Bailey CC, Kondur HR, Huang IC *et al.* Interferon-induced transmembrane protein 3 is a type II transmembrane protein. *J Biol Chem* 2013;**288**:32184–93.

51. Ling S, Zhang C, Wang W *et al.* Combined approaches of EPR and NMR illustrate only one transmembrane helix in the human IFITM3. *Sci Rep* 2016;**6**:24029.

52. Desai TM, Marin M, Chin CR *et al.* IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. *PLoS Pathog* 2014;**10**:e1004048.

53. Amini-Bavil-Olyaee S, Choi YJ, Lee JH *et al.* The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. *Cell Host Microbe* 2013;**13**:452–64.

54. Perreira JM, Chin CR, Feeley EM *et al.* IFITMs restrict the replication of multiple pathogenic viruses. *J Mol Biol* 2013;**425**:4937–55.

55. John SP, Chin CR, Perreira JM *et al.* The CD225 domain of IFITM3 is required for both IFITM protein association and inhibition of influenza A virus and dengue virus replication. *J Virol* 2013;**87**:7837–52.

56. Moulin C, Crupi MJF, Ilkow CS *et al.* Extracellular Vesicles and Viruses: Two Intertwined Entities. *Int J Mol Sci* 2023;**24**, DOI: 10.3390/IJMS24021036.

57. Kelemen A, Carmi I, Oszvald Á *et al.* IFITM1 expression determines extracellular vesicle uptake in colorectal cancer. *Cellular and Molecular Life Sciences* 2021;**78**:7009–24.

58. Zou X, Yuan M, Zhang T *et al.* EVs Containing Host Restriction Factor IFITM3 Inhibited ZIKV Infection of Fetuses in Pregnant Mice through Trans-placenta Delivery. *Molecular Therapy* 2021;**29**:176–90.

59. Wrensch F, Winkler M, Pöhlmann S. IFITM proteins inhibit entry driven by the MERS-coronavirus spike protein: evidence for cholesterol-independent mechanisms. *Viruses* 2014;**6**:3683–98.

60. Zhao X, Guo F, Liu F *et al.* Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. *Proc Natl Acad Sci U S A* 2014;**111**:6756–61.

61. Zhao X, Sehgal M, Hou Z *et al.* Identification of Residues Controlling Restriction versus Enhancing Activities of IFITM Proteins on Entry of Human Coronaviruses. Williams BRG (ed.). *J Virol* 2018;**92**:e01535-17.

62. Shi G, Schwartz O, Compton AA. More than meets the I: the diverse antiviral and cellular functions of interferon-induced transmembrane proteins. *Retrovirology* 2017;**14**:53.

63. Okuzaki Y, Kidani S, Kaneoka H *et al.* Characterization of chicken interferon-inducible transmembrane protein-10. *Biosci Biotechnol Biochem* 2017;**81**:914–21.

64. Jang H-J, Lee H-J, Kang KS *et al.* Molecular responses to the influenza A virus in chicken trachea-derived cells. *Poult Sci* 2015;**94**:1190–201.

65. Chauhan RP, Gordon ML. An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates. *Virus Genes 2022 58:4* 2022;**58**:255–69.

66. Firth AE, Brierley I. Non-canonical translation in RNA viruses. *J Gen Virol* 2012;**93**:1385–409.

67. Hu J, Ma C, Liu X. PA-X: a key regulator of influenza A virus pathogenicity and host immune responses. *Medical Microbiology and Immunology 2018 207:5* 2018;**207**:255–69.

68. Barrett T, Wolstenholme AJ, Mahy BWJ. Transcription and replication of influenza virus RNA. *Virology* 1979;**98**:211–25.

69. Pan Y, Yang P, Dong T *et al.* IFITM3 Rs12252-C Variant Increases Potential Risk for Severe Influenza Virus Infection in Chinese Population. *Front Cell Infect Microbiol* 2017;**7**:294.

70. Chen T, Xiao M, Yang J *et al.* Association between rs12252 and influenza susceptibility and severity: an updated meta-analysis. *Epidemiol Infect* 2018:1–9.

71. Makvandi-Nejad S, Laurenson-Schafer H, Wang L *et al.* Lack of Truncated IFITM3 Transcripts in Cells Homozygous for the rs12252-C Variant That is Associated With Severe Influenza Infection. *J Infect Dis* 2018;**217**:257–62.

72. Kim Y-C, Jeong B-H. No Correlation of the Disease Severity of Influenza A Virus Infection with the rs12252 Polymorphism of the Interferon-Induced Transmembrane Protein 3 Gene. *Intervirology* 2017;**60**:69–74.

73. Blyth GAD, Chan WF, Webster RG *et al.* Duck Interferon-Inducible Transmembrane Protein 3 Mediates Restriction of Influenza Viruses. *J Virol* 2016;**90**:103–16.

74. Smith J, Smith N, Yu L *et al.* A comparative analysis of host responses to avian influenza infection in ducks and chickens highlights a role for the interferon-induced transmembrane proteins in viral resistance. *BMC Genomics* 2015;**16**:574.

75. Allen EK, Randolph AG, Bhangale T *et al.* SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat Med* 2017;**23**:975–83.

76. Wu S, Zhang J, Huang J *et al.* Immune-Related Gene Expression in Ducks Infected With Waterfowl-Origin H5N6 Highly Pathogenic Avian Influenza Viruses. *Front Microbiol* 2019;**10**, DOI: 10.3389/FMICB.2019.01782.

77. Kumar A, Vijayakumar P, Gandhale PN *et al.* Genome-wide gene expression pattern underlying differential host response to high or low pathogenic H5N1 avian influenza virus in ducks. *Acta Virol* 2017;**61**:66–76.

78. Ranaware PB, Mishra A, Vijayakumar P *et al.* Genome Wide Host Gene Expression Analysis in Chicken Lungs Infected with Avian Influenza Viruses. 2016;**11**:e0153671.

79. Evseev D, Magor KE. Innate Immune Responses to Avian Influenza Viruses in Ducks and Chickens. *Vet Sci* 2019;**6**, DOI: 10.3390/vetsci6010005.

80. Rohaim MA, Al-Natour MQ, Abdelsabour MA *et al.* Transgenic Chicks Expressing Interferon-Inducible Transmembrane Protein 1 (IFITM1) Restrict Highly Pathogenic H5N1 Influenza Viruses. *International Journal of Molecular Sciences Article* 2021, DOI: 10.3390/ijms22168456.

81. Staines K, Batra A, Mwangi W *et al.* A Versatile Panel of Reference Gene Assays for the Measurement of Chicken mRNA by Quantitative PCR. *PLoS One* 2016;**11**:e0160173.

82. Vandesompele J, de Preter K, Pattyn F *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:1–12.

83. GraphPad Software. GraphPad Prism version 8.0.0 for Windows. 2019.

84. Andrews S et al. FastQC: a quality control tool for high throughput sequence data. 2021.

85. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.

86. Howe KL, Achuthan P, Allen J *et al.* Ensembl 2021. *Nucleic Acids Res* 2021;**49**, DOI: 10.1093/nar/gkaa942.

87. Yates AD, Achuthan P, Akanni W *et al.* Ensembl 2020. *Nucleic Acids Res* 2020;**48**:D682–8.

88. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**:323.

89. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;**30**:923–30.

90. García-Alcalde F, Okonechnikov K, Carbonell J *et al.* Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 2012;**28**:2678–9.

91. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016;**32**:292–4.

92. Ewels P, Magnusson M, Lundin S *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**:3047–8.

93. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2014.

94. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.

95. Stephens M. False discovery rates: a new deal. *Biostatistics* 2017;**18**:275–94.

96. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019;**47**:e47–e47.

97. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;**11**:1–12.

98. Bayerlová M, Jung K, Kramer F *et al.* Comparative study on gene set and pathway topology-based enrichment methods. *BMC Bioinformatics* 2015;**16**:1–15.

99. Subramanian A, Tamayo P, Mootha VK *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.

100. Geistlinger L, Csaba G, Küffner R *et al.* From sets to graphs: towards a realistic enrichment analysis of transcriptomic systems. *Bioinformatics* 2011;**27**:i366–73.

101. Tarca AL, Draghici S, Khatri P *et al.* A novel signaling pathway impact analysis. *Bioinformatics* 2009;**25**:75–82.

102. Tarca AL, Draghici S, Bhatti G *et al.* Down-weighting overlapping genes improves gene set analysis. *BMC Bioinformatics* 2012;**13**:136.

103. Mootha VK, Lindgren CM, Eriksson KF *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics 2003 34:3* 2003;**34**:267–73.

104. Subramanian A, Tamayo P, Mootha VK *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.

105. Ashburner M, Ball CA, Blake JA *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–9.

106. Yu G, Wang LG, Yan GR *et al.* DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;**31**:608–9.

107. Carbon S, Douglass E, Good BM *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* 2021;**49**:D325–34.

108. Gillespie M, Jassal B, Stephan R *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* 2022;**50**:D687–92.

109. Jassal B, Matthews L, Viteri G *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* 2020;**48**:D498–503.

110. Fabregat A, Sidiropoulos K, Viteri G *et al.* Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinformatics* 2017;**18**, DOI: 10.1186/s12859-017-1559-2.

111. Griss J, Viteri G, Sidiropoulos K *et al.* ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Molecular & Cellular Proteomics* 2020;**19**:2115–25.

112. Griss J, Viteri G, Sidiropoulos K *et al.* ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol Cell Proteomics* 2020;**19**:2115–24.

113. Köppl C, Lingg N, Fischer A *et al.* Fusion Tag Design Influences Soluble Recombinant Protein Production in Escherichia coli. *Int J Mol Sci* 2022;**23**, DOI: 10.3390/IJMS23147678.

114. Tanz SK, Castleden I, Small ID *et al.* Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Front Plant Sci* 2013;**4**, DOI: 10.3389/FPLS.2013.00214/ABSTRACT.

115. Weill U, Krieger G, Avihou Z *et al.* Assessment of GFP tag position on protein localization and growth fitness in yeast. *J Mol Biol* 2019;**431**:636.

116. Majorek KA, Kuhn ML, Chruszcz M *et al.* Double trouble—Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci* 2014;**23**:1359.

117. Weston S, Czieso S, White IJ *et al.* A Membrane Topology Model for Human Interferon Inducible Transmembrane Protein 1. *PLoS One* 2014;**9**:e104341.

118. Kim H, You S, Kim IJ *et al.* Increased mitochondrial-encoded gene transcription in immortal DF-1 cells. *Exp Cell Res* 2001;**265**:339–47.

119. Kong BW, Lee JY, Bottje WG *et al.* Genome-wide differential gene expression in immortalized DF-1 chicken embryo fibroblast cell line. *BMC Genomics* 2011;**12**:571.

120. Giotis ES, Ross CS, Robey RC *et al.* Constitutively elevated levels of SOCS1 suppress innate responses in DF-1 immortalised chicken fibroblast cells. *Sci Rep* 2017;**7**:17485.

121. Foster DN, Foster LK. US5672485A - Immortalized cell lines for virus growth - Google Patents. *Immortalized cell lines for virus growth* 1997.

122. You S, Kong BW, Jeon SY *et al.* Deregulation of catalase, not MnSOD, is associated with necrotic death of p53-defective DF-1 cells under antimycin A-induced oxidative stress. *Mol Cells* 2004;**18**:220–9.

123. Kim H, You S, Kong BW *et al.* Necrotic cell death by hydrogen peroxide in immortal DF-1 chicken embryo fibroblast cells expressing deregulated MnSOD and catalase. *Biochim Biophys Acta Mol Cell Res* 2001;**1540**:137–46.

124. Lee CW, Jung K, Jadhao SJ *et al.* Evaluation of chicken-origin (DF-1) and quail-origin (QT-6) fibroblast cell lines for replication of avian influenza viruses. *J Virol Methods* 2008;**153**:22–8.

125. Chungu K, Park YH, Woo SJ *et al.* Establishment of a genetically engineered chicken DF-1 cell line for efficient amplification of influenza viruses in the absence of trypsin. *BMC Biotechnol* 2021;**21**:1–9.

126. Schaefer-Klein J, Givol I, Barsov E v *et al.* The EV-O-Derived Cell Line DF-1 Supports the Efficient Replication of Avian Leukosis-Sarcoma Viruses and Vectors. 1998.

127. Levy AM, Gilad O, Xia L *et al.* Marek's disease virus Meq transforms chicken cells via the v-Jun transcriptional cascade: a converging transforming pathway for avian oncoviruses. *Proc Natl Acad Sci U S A* 2005;**102**:14831–6.

128. Buttigieg K, Laidlaw SM, Ross C *et al.* Genetic Screen of a Library of Chimeric Poxviruses Identifies an Ankyrin Repeat Protein Involved in Resistance to the Avian Type I Interferon Response. *J Virol* 2013;**87**:5028.

129. Kint J, Fernandez-Gutierrez M, Maier HJ *et al.* Activation of the Chicken Type I Interferon Response by Infectious Bronchitis Coronavirus. *J Virol* 2015;**89**:1156–67.

130. Tiwari A, Patnayak DP, Chander Y *et al.* Permissibility of different cell types for the growth of avian metapneumovirus. *J Virol Methods* 2006;**138**:80–4.

131. Himly M, Foster DN, Bottoli I *et al.* The DF-1 chicken fibroblast cell line: Transformation induced by diverse oncogenes and cell death resulting from infection by avian leukosis viruses. *Virology* 1998;**248**:295–304.

132. O'Hare TH, Delany ME. Genetic variation exists for telomeric array organization within and among the genomes of normal, immortalized, and transformed chicken systems. *Chromosome Research* 2009;**17**:947–64.

133. Kim H, You S, Kong BW *et al.* Necrotic cell death by hydrogen peroxide in immortal DF-1 chicken embryo fibroblast cells expressing deregulated MnSOD and catalase. *Biochim Biophys Acta Mol Cell Res* 2001;**1540**:137–46.

134. Karpala AJ, Lowenthal JW, Bean AG. Activation of the TLR3 pathway regulates IFNβ production in chickens. *Dev Comp Immunol* 2008;**32**:435–44.

135. Characterization and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications | FDA.

136. Ko P, Misaghi S, Hu Z *et al.* Probing the importance of clonality: Single cell subcloning of clonally derived CHO cell lines yields widely diverse clones differing in growth, productivity, and product quality. *Biotechnol Prog* 2018;**34**:624–34.

137. Altschuler SJ, Wu LF. Cellular heterogeneity: when do differences make a difference? *Cell* 2010;**141**:559.

138. Hillier LW, Miller W, Birney E *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature 2005 432:7018* 2004;**432**:695–716.

139. Martins C, Trifonov V, Houben A. Addressing Long-Standing Questions with Advanced Approaches: The 4th B Chromosome Conference. *Cytogenet Genome Res* 2020;**160**:111–7.

140. Andrews S. SeqMonk. *http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/* 2007.

141. Stephens M. False discovery rates: a new deal. *Biostatistics* 2017;**18**:275–94.

142. Yu G, Wang LG, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7.

143. Wu T, Hu E, Xu S *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2021;**2**:100141.

144. Lai JH, Wu DW, Wu CH *et al.* Mitochondrial CMPK2 mediates immunomodulatory and antiviral activities through IFN-dependent and IFN-independent pathways. *iScience* 2021;**24**, DOI: 10.1016/J.ISCI.2021.102498.

145. Li X, Feng Y, Liu W *et al.* A Role for the Chicken Interferon-Stimulated Gene CMPK2 in the Host Response Against Virus Infection. *Front Microbiol* 2022;**13**:874331.

146. Kiavue N, Cabel L, Melaabi S *et al.* ERBB3 mutations in cancer: biological aspects, prevalence and therapeutics. *Oncogene 2019 39:3* 2019;**39**:487–502.

147. Kramer IjM. Regulation of Cell Proliferation by Receptor Tyrosine Protein Kinases. *Signal Transduct* 2016:589–654.

148. Rahimi AM, Cai M, Hoyer-Fender S. Heterogeneity of the NIH3T3 Fibroblast Cell Line. *Cells* 2022;**11**, DOI: 10.3390/CELLS11172677.

149. Westermann L, Li Y, Göcmen B *et al.* Wildtype heterogeneity contributes to clonal variability in genome edited cells. *Scientific Reports 2022 12:1* 2022;**12**:1–13.

150. Delany ME, Krupkin AB, Miller MM. Organization of telomere sequences in birds: evidence for arrays of extreme length and for in vivo shortening. *Cytogenet Cell Genet* 2000;**90**:139–45.

151. Wang Y, Ghaffari N, Johnson CD *et al.* Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. *BMC Bioinformatics* 2011;**12**:1–7.

152. Chambers BS, Heaton BE, Rausch K *et al.* DNA mismatch repair is required for the host innate response and controls cellular fate after influenza virus infection. *Nature Microbiology 2019 4:11* 2019;**4**:1964–77.

153. Downey J, Pernet E, Coulombe F *et al.* Dissecting host cell death programs in the pathogenesis of influenza. *Microbes Infect* 2018;**20**:560–9.

154. Tisoncik JR, Billharz R, Burmakina S *et al.* The NS1 protein of influenza A virus suppresses interferon-regulated activation of antigen-presentation and immune-proteasome pathways. *J Gen Virol* 2011;**92**:2093.

155. Ayllon J, García-Sastre A. The NS1 protein: a multitasking virulence factor. *Curr Top Microbiol Immunol* 2015;**386**:73–107.

156. Yadav V, Panganiban AT, Honer Zu Bentrup K *et al.* Influenza infection modulates vesicular trafficking and induces Golgi complex disruption. *Virusdisease* 2016;**27**:357.

157. Jo S, Kawaguchi A, Takizawa N *et al.* Involvement of vesicular trafficking system in membrane targeting of the progeny influenza virus genome. *Microbes Infect* 2010;**12**:1079–84.

158. Raghavendar Reddy T, Suhasini M, Xu W *et al.* A role for KH domain proteins (Sam68-like mammalian proteins and quaking proteins) in the post-transcriptional regulation of HIV replication. *J Biol Chem* 2002;**277**:5778–84.

159. Hartl M, Reiter F, Bader AG *et al.* JAC, a direct target of oncogenic transcription factor Jun, is involved in cell transformation and tumorigenesis. *undefined* 2001;**98**:13601–6.

160. Rossman JS, Jing X, Leser GP *et al.* Influenza Virus M2 Protein Mediates ESCRT-Independent Membrane Scission. *Cell* 2010;**142**:902–13.

161. Shen C, Li YJ, Yin QQ *et al.* Identification of differentially expressed transcripts targeted by the knockdown of endogenous IFITM3. *Mol Med Rep* 2016;**14**:4367–73.

162. Rayner E, Durin M-A, Thomas R *et al.* CRISPR-Cas9 Causes Chromosomal Instability and Rearrangements in Cancer Cell Lines, Detectable by Cytogenetic Methods. *CRISPR J* 2019;**2**:406.

163. Aggarwal CC, Hinneburg A, Keim DA. On the Surprising Behavior of Distance Metrics in High Dimensional Space.

164. Shapira SD, Gat-Viks I, Shum BO v *et al.* A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 2009;**139**:1255–67.

165. Kim Y-C, Jeong M-J, Jeong B-H. Development of a chicken interferon-induced transmembrane protein 3 (IFITM3)-specific monoclonal antibody using phage display. *Acta Vet Hung* 2022;**70**:121–6.

166. Thomas Whitehead. Characterisation of chicken interferon-inducible transmembrane proteins: locus architecture, gene expression and viral restriction - UCL Discovery. *PhD Thesis* 2018.

167. Li N, Parrish M, Chan TK *et al.* Influenza infection induces host DNA damage and dynamic DNA damage responses during tissue regeneration. *Cell Mol Life Sci* 2015;**72**:2973.

168. Chambers BS, Heaton BE, Rausch K *et al.* DNA mismatch repair is required for the host innate response and controls cellular fate after influenza virus infection. *Nature Microbiology 2019 4:11* 2019;**4**:1964–77.

169. Ayllon J, García-Sastre A. The NS1 protein: a multitasking virulence factor. *Curr Top Microbiol Immunol* 2015;**386**:73–107.

170. Zhong L, Song Y, Marziali F *et al.* A novel domain within the CIL regulates egress of IFITM3 from the Golgi and reveals a regulatory role of IFITM3 on the secretory pathway. 2022, DOI: 10.26508/lsa.202101174.

171. Buchrieser J, Degrelle SA, Couderc T *et al.* IFITM proteins inhibit placental syncytiotrophoblast formation and promote fetal demise. *Science (1979)* 2019;**365**:176–80.

172. Lee J, Robinson ME, Ma N *et al.* IFITM3 functions as a PIP3 scaffold to amplify PI3K signalling in B cells. *Nature 2020 588:7838* 2020;**588**:491–7.

173. Sakamoto S, Inoue H, Kohda Y *et al.* Interferon-Induced Transmembrane Protein 1 (IFITM1) Promotes Distant Metastasis of Small Cell Lung Cancer. *International Journal of Molecular Sciences 2020, Vol 21, Page 4934* 2020;**21**:4934.

174. Min J, Feng Q, Liao W *et al.* IFITM3 promotes hepatocellular carcinoma invasion and metastasis by regulating MMP9 through p38/MAPK signaling. *FEBS Open Bio* 2018;**8**:1299.

175. He JD, Luo HL, Li J *et al.* Influences of the interferon induced transmembrane protein 1 on the proliferation, invasion, and metastasis of the colorectal cancer SW480 cell lines. *Chin Med J (Engl)* 2012;**125**:517–22.

176. Xu L, Zhou R, Yuan L *et al.* IGF1/IGF1R/STAT3 signaling-inducible IFITM2 promotes gastric cancer growth and metastasis. *Cancer Lett* 2017;**393**:76–85.

177. Tanaka SS, Yamaguchi YL, Tsoi B *et al.* IFITM/Mil/fragilis family proteins IFITM1 and IFITM3 play distinct roles in mouse primordial germ cell homing and repulsion. *Dev Cell* 2005;**9**:745–56.

178. Wylie C. IFITM1-mediated cell repulsion controls the initial steps of germ cell migration in the mouse. *Dev Cell* 2005;**9**:723–4.

179. Lange UC, Adams DJ, Lee C *et al.* Normal Germ Line Establishment in Mice Carrying a Deletion of the Ifitm/Fragilis Gene Family Cluster . *Mol Cell Biol* 2008;**28**:4688–96.

180. Klymiuk I, Kenner L, Adler T *et al.* In Vivo Functional Requirement of the Mouse Ifitm1 Gene for Germ Cell Development, Interferon Mediated Immune Response and Somitogenesis. *PLoS One* 2012;**7**:e44609.

181. Gómez-Herranz M, Taylor J, Sloan RD. IFITM proteins: understanding their diverse roles in viral infection, cancer, and immunity. *Journal of Biological Chemistry* 2022;**0**:102741.

182. Sparrow E, Wood JG, Chadwick C *et al.* Global production capacity of seasonal and pandemic influenza vaccines in 2019. *Vaccine* 2021;**39**:512–20.

183. Francis T, Salk JE, Pearson HE *et al.* PROTECTIVE EFFECT OF VACCINATION AGAINST INDUCED INFLUENZA A. *J Clin Invest* 1945;**24**:536–46.

184. Wood JM, Weir JP. Standardisation of inactivated influenza vaccines-Learning from history. *Influenza Other Respi Viruses* 2018;**12**:195–201.

185. Sun Y, Li Y, Zong Y *et al.* Poultry genetic heritage cryopreservation and reconstruction: advancement and future challenges. *Journal of Animal Science and Biotechnology 2022 13:1* 2022;**13**:1–18.

186. Nakamura Y. Poultry genetic resource conservation using primordial germ cells. *J Reprod Dev* 2016;**62**:431.

187. Cooper CA, Doran TJ, Challagulla A *et al.* Innovative approaches to genome editing in avian species. *J Anim Sci Biotechnol* 2018;**9**:1–7.

188. Vick L, Li Y, Simkiss K. Transgenic birds from transformed primordial germ cells. *Proc Biol Sci* 1993;**251**:179–82.

189. Kong L, Qiu L, Guo Q *et al.* Long-term in vitro culture and preliminary establishment of chicken primordial germ cell lines. *PLoS One* 2018;**13**:e0196459.

190. van de Lavoir MC, Diamond JH, Leighton PA *et al.* Germline transmission of genetically modified primordial germ cells. *Nature 2006 441:7094* 2006;**441**:766–9.

191. Altgilbers S, Klein S, Dierks C *et al.* Cultivation and characterization of primordial germ cells from blue layer hybrids (Araucana crossbreeds) and generation of germline chimeric chickens. *Scientific Reports |* 2021;**11**:12923.

192. Li BC, Tian ZQ, Sun M *et al.* Directional Differentiation of Chicken Primordial Germ Cells Into Adipocytes, Neuron-Like Cells, and Osteoblasts. *Mol Reprod Dev* 2010;**77**:795–801.

193. Bassano I, Ong SH, Sanz-Hernandez M *et al.* Comparative analysis of the chicken IFITM locus by targeted genome sequencing reveals evolution of the locus and positive selection in IFITM1 and IFITM3. *BMC Genomics* 2019;**20**:1–14.

194. Biomatters. Geneious. *www.geneious.com* 2018.

195. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2018;**20**:1160–6.

196. Trabelsi K, Rourou S, Loukil H *et al.* Optimization of virus yield as a strategy to improve rabies vaccine production by Vero cells in a bioreactor. *J Biotechnol* 2006;**121**:261–71.

197. Chen K-D, Wu X-X, Yu D-S *et al.* Process optimization for the rapid production of adenoviral vectors for clinical trials in a disposable bioreactor system. *Appl Microbiol Biotechnol* 2018;**102**:6469–77.

198. Timm A, Yin J. Kinetics of virus production from single cells. *Virology* 2012;**424**:11–7.

199. Long JS, Idoko-Akoh A, Mistry B *et al.* Species specific differences in use of ANP32 proteins by influenza A virus. *Elife* 2019;**8**:e45066.

200. Mudhasani R, Tran JP, Retterer C *et al.* IFITM-2 and IFITM-3 but not IFITM-1 restrict Rift Valley fever virus. *J Virol* 2013;**87**:8451–64.

201. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995;**57**:289–300.

202. Whyte J, Glover JD, Woodcock M *et al.* FGF, Insulin, and SMAD Signaling Cooperate for Avian Primordial Germ Cell Self-Renewal. *Stem Cell Reports* 2015;**5**:1171–82.

203. Idoko-Akoh A, Taylor L, Sang HM *et al.* High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells. *Scientific Reports 2018 8:1* 2018;**8**:1–14.

204. Li BC, Tian ZQ, Sun M *et al.* Directional differentiation of chicken primordial germ cells into adipocytes, neuron-like cells, and osteoblasts. *Mol Reprod Dev* 2010;**77**:795–801.

205. Hickford D, Frankenberg S, Shaw G *et al.* Evolution of vertebrate interferon inducible transmembrane proteins. *BMC Genomics* 2012;**13**:155.

206. Zhang D, Wang H, He H *et al.* Interferon induced transmembrane protein 3 regulates the growth and invasion of human lung adenocarcinoma. *Thorac Cancer* 2017;**8**:337–43.

207. Edfors F, Danielsson F, Hallström BM *et al.* Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* 2016;**12**:883.

208. Wegler C, Ölander M, Wiśniewski JR *et al.* Global variability analysis of mRNA and protein concentrations across and within human tissues. *NAR Genom Bioinform* 2020;**2**, DOI: 10.1093/NARGAB/LQZ010.

209. Zhytnik L, Maasalu K, Duy BH *et al.* IFITM5 pathogenic variant causes osteogenesis imperfecta V with various phenotype severity in Ukrainian and Vietnamese patients. *Hum Genomics* 2019;**13**:25.

210. Hanagata N. IFITM5 mutations and osteogenesis imperfecta. *J Bone Miner Metab* 2016;**34**:123–31.

211. Yu J, Sun X, Goie JYG *et al.* Regulation of Host Immune Responses against Influenza A Virus Infection by Mitogen-Activated Protein Kinases (MAPKs). *Microorganisms* 2020;**8**:1–17.

212. Ludwig S, Hrincius ER, Boergeling Y. The Two Sides of the Same Coin—Influenza Virus and Intracellular Signal Transduction. *Cold Spring Harb Perspect Med* 2021;**11**:a038513.

213. Marjuki H, Alam MI, Ehrhardt C *et al.* Membrane accumulation of influenza A virus hemagglutinin triggers nuclear export of the viral genome via protein kinase Calpha-mediated activation of ERK signaling. *J Biol Chem* 2006;**281**:16707–15.

214. Ehrhardt C, Wolff T, Pleschka S *et al.* Influenza A Virus NS1 Protein Activates the PI3K/Akt Pathway To Mediate Antiapoptotic Signaling Responses. *J Virol* 2007;**81**:3058.

215. Zhirnov OP, Klenk HD. Control of apoptosis in influenza virus-infected cells by up-regulation of Akt and p53 signaling. *Apoptosis* 2007;**12**:1419–32.

216. Stergiou L, Bauer M, Mair W *et al.* Integrin-mediated signaling induced by simian virus 40 leads to transient uncoupling of cortical actin and the plasma membrane. *PLoS One* 2013;**8**, DOI: 10.1371/JOURNAL.PONE.0055799.

217. Ohta K, Goto H, Matsumoto Y *et al.* Graf1 Controls the Growth of Human Parainfluenza Virus Type 2 through Inactivation of RhoA Signaling. *J Virol* 2016;**90**:9394.

218. Thaker SK, Ch’ng J, Christofk HR. Viral hijacking of cellular metabolism. *BMC Biology 2019 17:1* 2019;**17**:1–15.

219. Sadri Nahand J, Shojaie L, Akhlagh SA *et al.* Cell death pathways and viruses: Role of microRNAs. *Mol Ther Nucleic Acids* 2021;**24**:487–511.

220. Fan Y, Sanyal S, Bruzzone R. Breaking Bad: How Viruses Subvert the Cell Cycle. *Front Cell Infect Microbiol* 2018;**8**:407205.

221. Wang Y, Lupiani B, Reddy SM *et al.* RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poult Sci* 2014;**93**:485–93.

222. Giotis ES, Montillet G, Pain B *et al.* Chicken Embryonic-Stem Cells Are Permissive to Poxvirus Recombinant Vaccine Vectors. *Genes 2019, Vol 10, Page 237* 2019;**10**:237.