

RESEARCH ARTICLE

# Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses

Andrew E. Shaw, Joseph Hughes, Quan Gu, Abdelkader Behdenna<sup>‡</sup>, Joshua B. Singer, Tristan Dennis, Richard J. Orton, Mariana Varela, Robert J. Gifford, Sam J. Wilson\*, Massimo Palmarini\*

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

<sup>‡</sup> Current address: Boyd Orr Centre for Population and Ecosystem Health, Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

\* [sam.wilson@glasgow.ac.uk](mailto:sam.wilson@glasgow.ac.uk) (SJW); [massimo.palmarini@glasgow.ac.uk](mailto:massimo.palmarini@glasgow.ac.uk) (MP)



**OPEN ACCESS**

**Citation:** Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. (2017) Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol* 15(12): e2004086. <https://doi.org/10.1371/journal.pbio.2004086>

**Academic Editor:** Harmit Malik, Fred Hutchinson Cancer Research Center, United States of America

**Received:** August 29, 2017

**Accepted:** November 22, 2017

**Published:** December 18, 2017

**Copyright:** © 2017 Shaw et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Fastq sequence data files are available from the European Bioinformatics Institute (EBI, project accession number PRJEB21332). Alternatively, differentially expressed genes can be queried via a webserver at <http://isg.data.cvr.ac.uk/>. All other relevant data are within the paper and its Supporting Information files.

**Funding:** Medical Research Council <https://www.mrc.ac.uk/> (grant number G0801822 and

## Abstract

The host innate immune response mediated by type I interferon (IFN) and the resulting up-regulation of hundreds of interferon-stimulated genes (ISGs) provide an immediate barrier to virus infection. Studies of the type I ‘interferome’ have mainly been carried out at a single species level, often lacking the power necessary to understand key evolutionary features of this pathway. Here, using a single experimental platform, we determined the properties of the interferomes of multiple vertebrate species and developed a webserver to mine the dataset. This approach revealed a conserved ‘core’ of 62 ISGs, including genes not previously associated with IFN, underscoring the ancestral functions associated with this antiviral host response. We show that gene expansion contributes to the evolution of the IFN system and that interferomes are shaped by lineage-specific pressures. Consequently, each mammal possesses a unique repertoire of ISGs, including genes common to all mammals and others unique to their specific species or phylogenetic lineages. An analysis of genes commonly down-regulated by IFN suggests that epigenetic regulation of transcription is a fundamental aspect of the IFN response. Our study provides a resource for the scientific community highlighting key paradigms of the type I IFN response.

## Author summary

The type I interferon (IFN) response is triggered upon sensing of an incoming pathogen in an infected cell and results in the expression of hundreds of IFN-stimulated genes (ISGs, collectively referred to as ‘the interferome’). Studies on the interferome have been carried out mainly in human cells and therefore often lack the power to understand comparative evolutionary aspects of this critical pathway. In this study, we characterized the interferome in several animal species (including humans) using a single experimental

MR/K024752/1). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AIV, avian influenza virus; DIGS, database-integrated genome screening; DMEM, Dulbecco's modified Eagle medium; EBI, European Bioinformatics Institute; ECACC, European Collection of Authenticated Cell Cultures; ERCC, External RNA Controls Consortium; FBS, fetal bovine serum; FDR, false discovery rate; FPKM, fragments per kilobase mapped values; HGNC, HUGO Gene Nomenclature Committee; IFN, interferon; IFN- $\alpha$ , interferon alpha; IFN- $\beta$ , interferon beta; IFN- $\gamma$ , interferon gamma; IFN- $\lambda$ , interferon lambda; IRGs, interferon-repressed genes; ISGs, interferon-stimulated genes; lincRNAs, long intergenic noncoding RNAs; log<sub>2</sub>FC, log<sub>2</sub> fold change; MDS, multidimensional scaling; MHC, major histocompatibility complex; ncRNAs, noncoding RNAs; PAMPs, pathogen-associated molecular patterns; p/s, penicillin/streptomycin; RIN, RNA integrity number; RNAseq, RNA sequencing; RSV, respiratory syncytial virus; SNP, single nucleotide polymorphism; UIFN, Universal interferon.

framework. This approach allowed us to identify fundamental properties of the innate immune system. In particular, we revealed 62 'core' ISGs, up-regulated in response to IFN in all vertebrates, highlighting the ancestral functions of the IFN system. In addition, we show that many genes repressed by the IFN response normally function as regulators of cell transcription. ISGs shared by multiple species have a higher propensity than other genes to exist as multiple copies in the genome. Importantly, we observed that genes have arisen as ISGs throughout evolution. Hence, every animal species possesses a unique repertoire of ISGs that includes core and lineage-specific genes. Collectively, our data provide a framework on which it will be possible to test the role of the IFN response in pathogen emergence and cross-species transmission.

## Introduction

Most emerging human viruses have an animal origin [1]. The increase in the global human population, international travel, and ecological changes, in addition to changes in agricultural practices, has led to complicated interactions between wildlife, domestic species, and humans that has enhanced the opportunities for cross-species transmission of known, as well as newly discovered, viruses [1,2]. Physical and molecular components of the innate immune system represent early barriers to incoming viruses that must be overcome in order for an infection to prevail. In vertebrates, one of the key innate immune defences against virus infection is the interferon (IFN) system. Type I interferons (including IFN- $\beta$  and IFN- $\alpha$  among others; here referred to simply as IFN), type II interferons (IFN- $\gamma$ ), and type III interferons (IFN- $\lambda$ ) are cytokines with antipathogen, immunomodulatory, and proinflammatory properties. The IFN system is usually stimulated by the detection of pathogen signatures, known as pathogen-associated molecular patterns (PAMPs), resulting in the secretion of IFN. In turn, IFN signalling results in the up-regulation of hundreds of interferon-stimulated genes (ISGs), collectively referred to as the type I 'interferome' (here simply the 'interferome') [3,4]. Unsurprisingly, given the importance of IFN in combatting pathogen invasion, there are numerous examples of coevolutionary arms races between ISGs and invading pathogens [5,6]. However, previous studies investigating ISG transcription have focused on the interferomes of single species [7,8]. Despite resources such as the Interferome database [9], variations in experimental and bioinformatic approaches make comparing interferomes derived from divergent species and collected from different studies a difficult prospect if significant technical caveats and confounding factors are to be avoided. Here, we used the same RNA sequencing (RNAseq) approach on 10 animal species to deliver a snapshot of the genes that are differentially expressed in cells (fibroblasts) in a type I IFN-induced antiviral state. This snapshot of the interferome from a single cell type at one point in time cannot capture the entire temporal and tissue-specific complexity of the interferome. Nonetheless, using this comparative approach, we have uncovered fundamental paradigms of the IFN system.

## Results

### Interferomes differ among evolutionary lineages

We first determined the individual interferomes of *Homo sapiens* (human), *Rattus norvegicus* (rat), *Bos taurus* (cow), *Ovis aries* (sheep), *Sus scrofa* (pig), *Equus caballus* (horse), *Canis lupus familiaris* (dog), *Myotis lucifugus* (little brown bat, microbat), *Pteropus vampyrus* (large flying fox, fruit bat), and *Gallus gallus* (chicken) cells. Interferomes were obtained from cells

stimulated with type I IFN and experimentally confirmed as being in an antiviral state as described in the Materials and methods (S1 Fig). In our study, we defined an ISG as a gene up-regulated by IFN with a false discovery rate (FDR) of  $<0.05$ , regardless of the extent to which it was up-regulated. To facilitate mining of the data, we also developed an open access webserver (<http://isg.data.cvr.ac.uk>) capable of filtering the dataset based upon user-defined criteria.

The absolute number of ISGs differentially expressed in each species varied (S1 Table), but their pattern of differential expression in response to type I IFN was remarkably similar (Fig 1A). The presence of shared ISGs at specific nodes on a schematic phylogeny provided evidence that interferomes have been sculpted over time by lineage-specific pressures possibly exerted by different pathogens (Fig 1B). As expected, we observed that the most closely related species in our dataset, cows and sheep, showed the greatest similarity in the genes they up-regulate (Fig 1C). However, we also observed substantial levels of similarity in the interferomes of some species that are more distantly related phylogenetically, most notably pigs and humans (Fig 1C). Interestingly, this finding was reflected in a principal component analysis of the 35 one-to-one (i.e., single copy) ISG orthologs up-regulated by every mammalian species in our study (see below), whereby the patterns of differential expression were again similar between humans and pigs (Fig 1D).

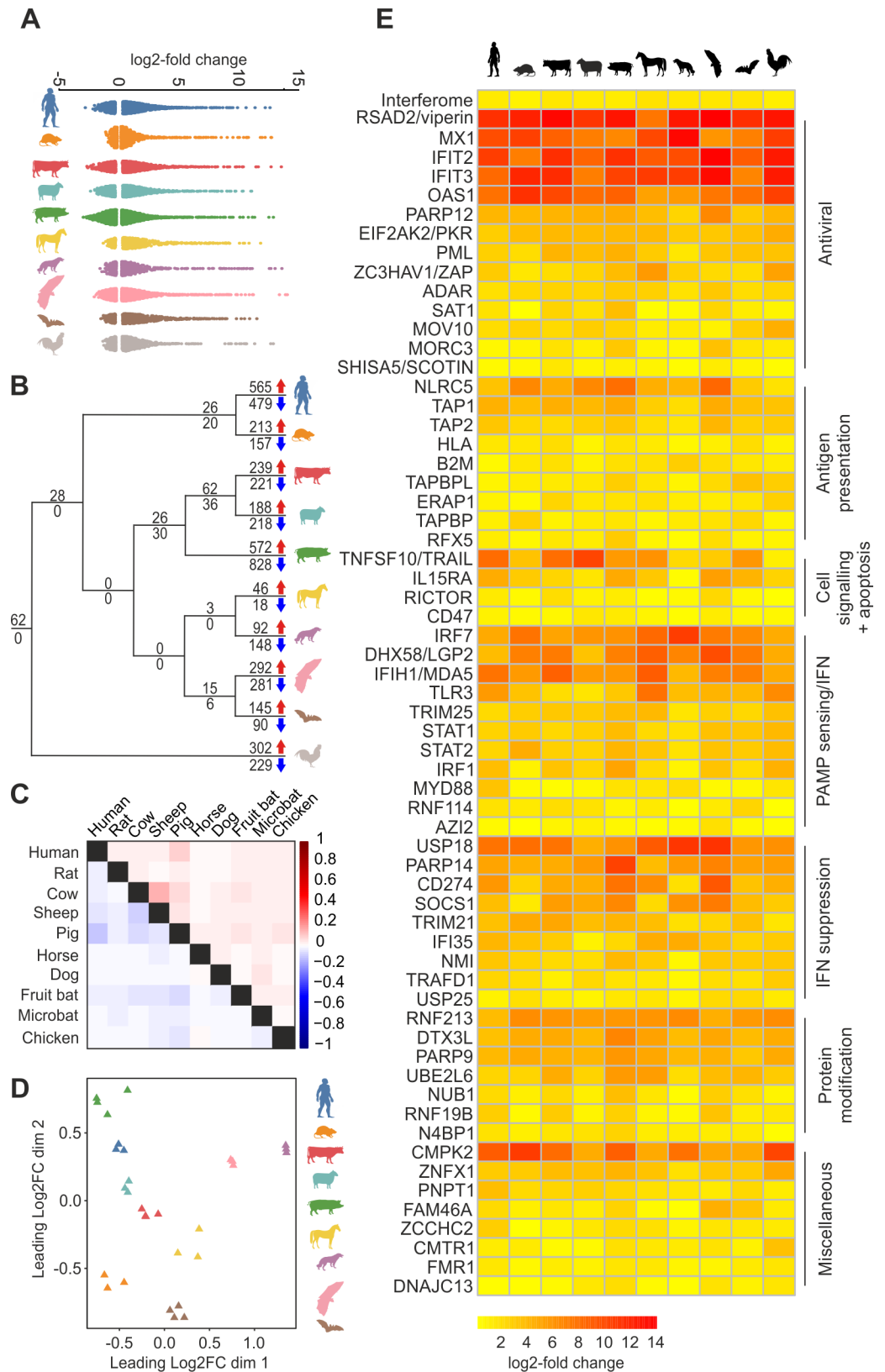
Every species possessed unique ISGs that were not up-regulated by IFN in any of the other nine species. Furthermore, certain ISGs present in our dataset (despite being up-regulated by IFN  $>2 \log_2$  fold change [ $\log_2FC$ ]) had few (if any) orthologs in the other genomes in the Ensembl database. Examples include a gene (RGD1561157) that is annotated on chromosome 10 of the rat genome and two chicken genes (Ensembl IDs ENSGALG00000019325 and ENSGALG00000020899).

## Core ISGs

We identified a core set of 62 genes (“core vertebrate ISGs”, hereinafter core<sup>vert</sup> ISGs) that were up-regulated by IFN by all 10 species analysed in this study, with an additional 28 genes up-regulated specifically in the nine mammalian species (“core mammalian ISGs”, hereinafter core<sup>mamm</sup> ISGs) (Table 1). The core<sup>vert</sup> ISGs represent the ancestral functions of the IFN system and include genes encoding proteins broadly involved in (i) orchestrating antigen presentation, (ii) IFN induction and response, (iii) IFN suppression, (iv) ubiquitination and protein degradation, (v) cell signalling and apoptosis, and (vi) antiviral responses (Table 1, Fig 1E).

Nine of the 62 core<sup>vert</sup> ISGs (e.g., various HLA genes, TAP1, ERAP1, etc.) are involved in the generation, trimming, loading, and presentation of MHC-I-restricted antigens, thus providing a direct link between the IFN response and the adaptive immune response via the CD8 T-cell response (Table 1). Additionally, components of the immunoproteasome (e.g., PSMB8 and PSMB9) were specifically up-regulated in the mammalian core (Table 1), reflecting previous studies reporting the absence of the immunoproteasome in birds [10].

We found that various core<sup>vert</sup> ISGs are involved in IFN induction and response, including pattern recognition receptors, the key adaptor molecule MyD88, and transcription factors. The classical sensors for RNA PAMPs, including IFIH1/MDA5, DHX58/LGP2, and TLR3 (and RIG-I/DDX58 in the mammals) were among the core<sup>vert</sup> ISGs (Table 1, Fig 2). RIG-I is known to be absent in chickens (and other galliformes) but is present and active in other birds, including ducks and geese [11–13]. Furthermore, TRIM25, a ubiquitin E3 ligase responsible for ubiquitinating RIG-I, was also a core<sup>vert</sup> ISG [14]. DAI/ZBP1, which was originally classified as a DNA sensor but is now thought to be an RNA sensor [15], was also highly up-regulated by IFN in every mammalian species except humans (Fig 2). No DNA sensors were found among the core<sup>vert</sup> ISGs. However, cGAS was found to be an ISG in every mammalian species



**Fig 1. Patterns of differential gene expression in response to type I IFN among cells from the 10 vertebrate species used in this study. (A)** The patterns of differential expression of ISGs and IRGs are

broadly similar in cells derived from 10 different animal species. Each dot in the panel represents a gene that is differentially expressed in response to type I IFN treatment (S1 Data). (B) Number of ISGs (above line, red arrow) and IRGs (below line, blue arrow) are plotted on the branches of a simplified phylogenetic tree (branch lengths are not shown to scale). ISGs common to every species ( $n = 62$ ) are located at the root of the tree with an additional 28 ISGs up-regulated by all mammalian species used in this study. At the tips of the tree lie genes that are only up- or down-regulated in an individual species in our study. (C) Normalised correlation matrix showing pairwise comparisons between ISGs (red) and IRGs (blue) of the indicated animal species (S1 Data). (D) A PCA of the log<sub>2</sub>FC data of the one-to-one orthologs up-regulated by all nine mammalian species used in this study (S1 Data). Each point represents an animal or experiment, coloured according to species. The distribution of the samples reflects expression patterns. Separate animals/experiments cluster according to species, with the pig and human showing similar patterns. (E) A heatmap of the relative expression of the 62 vertebrate core ISGs. The first row (labelled as 'Interferome') represents the average log<sub>2</sub>FC of all up-regulated ISGs for each animal species. Up-regulated paralogs have been averaged in the case of genes for which there are expansions; for example, the rat has two copies of MX1 compared to the single copy in the remaining species (S1 Data). IFN, interferon; IRGs, interferon-repressed genes; ISGs, interferon-stimulated genes; log<sub>2</sub>FC, log<sub>2</sub> fold change; PCA, principal component analysis.

<https://doi.org/10.1371/journal.pbio.2004086.g001>

in this study (Fig 2). By contrast, the DNA sensor AIM2 was only up-regulated in human cells. Interestingly, whilst the basal expression (defined in terms of fragments per kilobase mapped values [FPKM] in the absence of IFN treatment) of RNA sensors was very low, many of the genes in the literature associated with DNA sensing are constitutively transcribed (Fig 2).

With the exception of MyD88, a key adaptor involved in both the RNA- and DNA-sensing pathways [16], we observed limited up-regulation among genes involved downstream of

**Table 1. Core ISGs.**

Antigen presentation	Antiviral	Ubiquitin + protein modification	PAMP sensing + IFN pathway	IFN suppression	Cell signalling and apoptosis	Miscellaneous
B2M	ADAR	DTX3L	AZI2	CD274	CASP8*	C2*
ERAP1	APOL1,2,3,4*	HERC6*	cGAS (MB21D1)*	IFI35	CD47	CMPK2
HLA <sup>#</sup>	C19orf66 (IRAV)*	N4BP1	IRF1	NMI	IL15RA	CMTR1
NLRC5	IFIT2	NUB1	IRF7	PARP14	LGALS9*	DNAJA1*
PSMA5*	IFIT3	PARP9	IRF9*	SOCS1	RICTOR	DNAJC13
PSMB8*	ISG15*	RBCK1*	LGP2 (DHX58)	TRAFD1	TRAIL (TNFSF10)	EHD4*
PSMB9*	ISG20*	RNF19B	MDA5 (IFIH1)	TRIM21		FAM46A
PSMB10*	MORC3	RNF213	MYD88	USP18		FMR1
PSME1*	MOV10	RNF31*	RIG-I (DDX58)*	USP25		PNPT1
PSME2*	MX1	UBA7*	RNF114			SERTAD1*
RFX5	OAS1	UBE2L6	STAT1			SLC25A28*
TAP1	PARP12		STAT2			SP110*
TAP2	PKR (EIF2AK2)		TLR3			TDRD7*
TAPBP	PML		TRIM25			WARS*
TAPBPL	RSAD2 (viperin)					XAF1*
	SAT1					ZCCHC2
	SCOTIN (SHISA5)					ZNFX1
	ZAP (ZC3HAV1)					

\* = ISGs shared only among the mammalian species analysed in this study (core<sup>mamm</sup> ISGs)

<sup>#</sup> = Includes HLA-A, -B, -C, -E, -F, -G

**Abbreviations:** IFN, interferon; ISGs, interferon-stimulated genes; PAMP, pathogen-associated molecular pattern

<https://doi.org/10.1371/journal.pbio.2004086.t001>



**Fig 2. Basal transcription levels and IFN-induced expression of genes related to PAMP sensing and IFN induction and response.** Boxplots showing differential expression (log<sub>2</sub>FC) in response to IFN and basal transcription levels (expressed as FPKM) of genes associated with pattern recognition (sensors), downstream signal transduction (adapters), and transcription factors related to either IFN induction or response (transcription factors). Every ortholog for each gene is indicated with a dot coloured according to their presence in the DNA-, RNA-, or both DNA- and RNA-sensing pathways (S1 Data). FPKM, fragments per kilobase mapped values; IFN, interferon; log<sub>2</sub>FC, log<sub>2</sub> fold change; PAMP, pathogen-associated molecular pattern.

<https://doi.org/10.1371/journal.pbio.2004086.g002>

nucleic acid detection (Fig 2). On the other hand, core ISGs included key transcription factors involved in IFN induction and response (IRF1, IRF7, STAT1 and STAT2 are all core<sup>vert</sup> ISGs

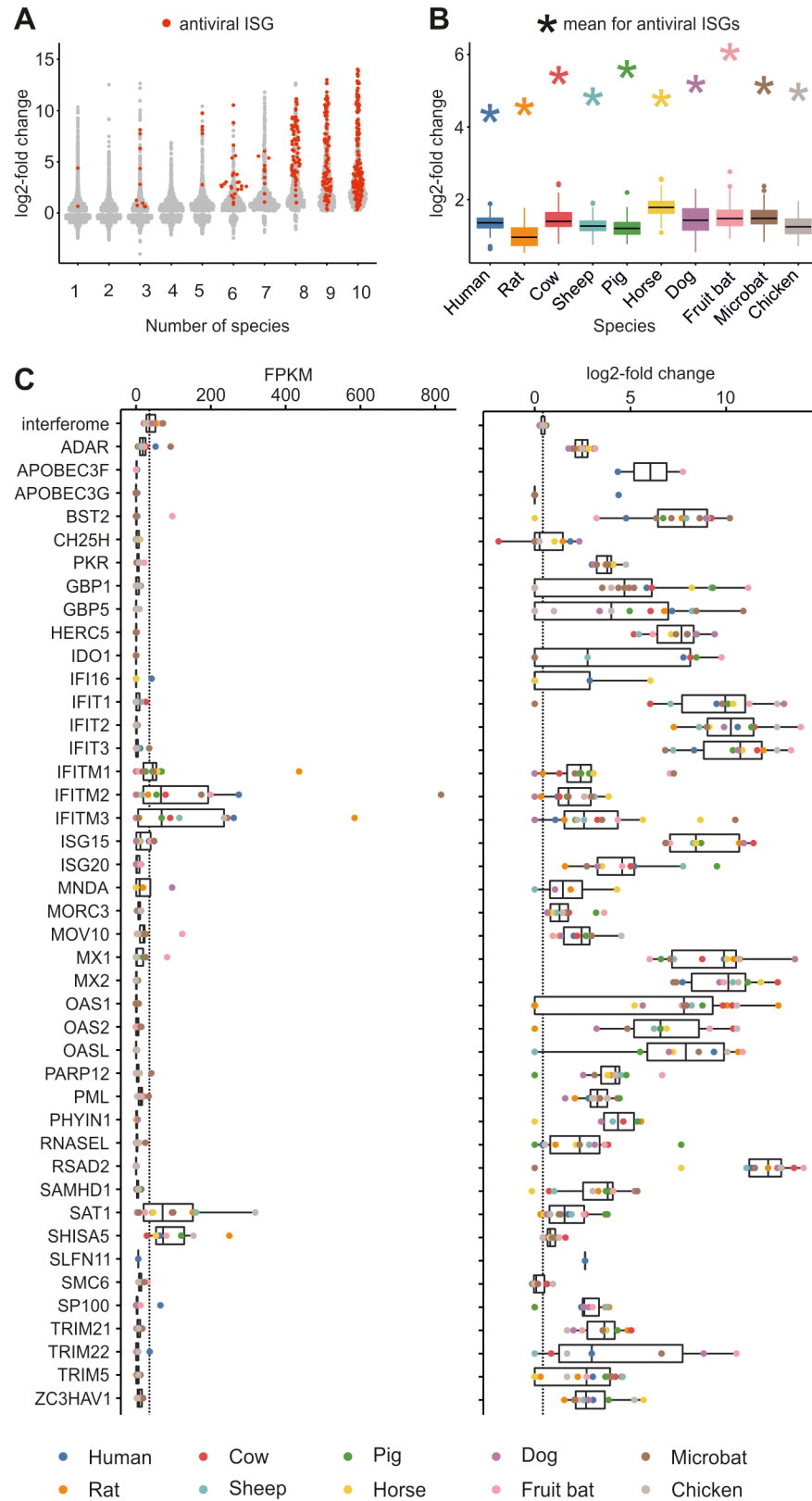
in addition to IRF9 among the core<sup>mamm</sup> ISGs (Fig 2). Importantly, several ISGs that play a role in the suppression of the IFN system, including USP18, USP25, IFI35, and SOCS1 were up-regulated in all species under examination. The encoded proteins of these genes target different points in the IFN response. Thus, negative regulation of the IFN response is multifaceted and a fundamental, ancestral failsafe necessary to avoid excessive/perpetual up-regulation of IFN-induced pathways. Among the core<sup>vert</sup> ISGs, we found several genes relating to ubiquitination, such as the ring finger proteins RNF213 and RNF19B (Table 1, Fig 1E), highlighting protein modification as part of the IFN response. Interestingly, N4BP1, originally identified as a target of Nedd4-mediated ubiquitination, has not previously been directly linked to the IFN response.

The core<sup>vert</sup> ISGs contained 14 IFN-induced antiviral factors such as MX1, IFIT2, and viperin (Table 1). Interestingly, when we assembled a list of 40 genes that either create a cellular environment hostile to or act directly upon the virus lifecycle (based upon the scientific literature; S2 Table), we noticed that 75% of these antiviral genes were ISGs in at least eight of the 10 species analysed in this study (Fig 3A). In addition, antiviral ISGs were up-regulated to a significantly higher extent than randomly sampled ISGs ( $P < 0.01$  for each species, Fig 3B). Furthermore, some well-studied antiviral ISGs were not up-regulated by IFN in certain species. For example, OAS2 was not up-regulated in the rat, SAMHD1 was not up-regulated in the horse, OASL was not up-regulated in either the cow or the sheep, and the IFITM genes were not up-regulated in the dog (Fig 3C). In general, genes encoding antiviral factors were transcribed to minimal levels in the absence of IFN. Indeed, the median FPKM level for antiviral genes was lower than that of the overall interferome for every antiviral factor except SAT1, SHISA5, and the IFITM genes. Of interest, we noticed particularly high FPKM values for IFITM1 and 3 in the rat and IFITM2 in the microbat (Fig 3C).

Two core<sup>vert</sup> genes, CD47 and IL15RA, encode proteins involved in signalling to components of the adaptive immune system. CD47 is involved in a variety of biological roles, including leukocyte and dendritic cell migration, the development of antigen presenting cells, and immune apoptosis, and it also provides a 'don't eat me' signal [17]. The IL15–IL15R $\alpha$  axis is well characterised as being important in the promotion of both natural killer cells and a variety of T cell populations, including activated CD8 T cells [18].

## Genes previously unrelated to the IFN response

Among the core<sup>vert</sup> ISGs, we identified a number of genes with few or no reported associations with the type I IFN response (Table 1). Several of these genes have been studied extensively, but not always in the context of the IFN response. DNAJC13, for example, is reported to be involved in endosome trafficking, and it has been closely linked to Parkinson's disease [19]. Zinc Finger CCHC-Type Containing 2 (ZCCHC2) has nucleic acid-binding properties and, interestingly, contains a single nucleotide polymorphism (SNP) associated with insect bite hypersensitivity in Exmoor ponies [20]. The fragile X mental retardation (FMR1) gene encodes an RNA-binding protein that plays a role in intracellular RNA transport and in the regulation of translation of target mRNAs. FMR1 was not previously linked to the IFN response, although it has recently been shown to be a proviral factor for influenza virus and previously was shown to induce mild restriction of HIV-1 [21,22]. Cap methyltransferase 1 (CMTR1), also known as ISG95, binds to RNA pol II and is a 2'-O-ribose methyltransferase that participates in the conversion of cap0 to cap1 type transcripts [23]. Interestingly CMTR1 is also known as an important component of IFIT-mediated antiviral activity [24]. We hypothesise that these genes, as core<sup>vert</sup> ISGs, likely play fundamental roles in host immunity that are underappreciated or have yet to be fully determined.



**Fig 3. Properties of antiviral ISGs.** (A) Sinaplot showing the differential expression values (log<sub>2</sub>FC) of 40 genes previously published as exerting antiviral activity (red dots) (S2 Table) as opposed to the rest of the



ISGs (grey dots). ISGs (including antiviral ISGs) were allocated to bins according to the number of species in which they were found to be up- or down-regulated (i.e., the core<sup>vert</sup> ISGs are bin 10). The majority of antiviral ISGs (red) were found to be up-regulated in at least eight species (S1 Data). (B) Graph showing the extent of up-regulation of antiviral ISGs compared to nonantiviral ISGs. The mean log<sub>2</sub>FC of 40 known antiviral ISGs (indicated with an asterisk on the plot) is compared to 100 samplings of 40 randomly selected ISGs from the interferome of each species (box and whiskers). In all cases, antiviral ISGs are up-regulated to a significantly greater extent as compared to nonantiviral ISGs ( $P < 0.01$  for each species). The code used for random sampling and the generation of Fig 3B is available in S2 Data, with the required input files available as S5 Data and S6 Data. (C) Boxplots showing basal transcription levels (expressed as FPKM) and differential expression (log<sub>2</sub>FC) in response to IFN of known antiviral ISGs as in panels A and B. Every ortholog for each gene is indicated with a dot coloured according to species. The median FPKM value for the entire interferome is indicated with a broken line (S1 Data). FPKM, fragments per kilobase mapped values; IFN, interferon; ISGs, interferon-stimulated genes; log<sub>2</sub>FC, log<sub>2</sub> fold change.

<https://doi.org/10.1371/journal.pbio.2004086.g003>

Our data also indicate a relatively underappreciated link between local synthesis of early components of the complement system and the type I IFN response [25,26]. C2 was among our core<sup>mamm</sup> ISGs. In addition, we found that C1r and C1s, essential components of the C1 complex, were up-regulated by IFN in cells from all species analysed in our study with the exception of the cow. Interestingly, this was also the case for a negative regulator of C1r and C1s (SERPING1/C1-INH).

### Interferon-repressed genes and epigenetic controls of the IFN response

Our data enabled an unprecedented opportunity to investigate the interferon-repressed genes (IRGs), which, to date, have received comparatively little attention with regards to their involvement in the innate immune system. Unlike ISGs, the extent of down-regulation of IRGs across all the 10 species used in this study was relatively modest (overall average  $-0.56$  log<sub>2</sub>FC for IRGs as compared to  $1.64$  log<sub>2</sub>FC for ISGs). We found no IRGs shared by all species, although this may reflect the low fold change in expression and/or greater variability in the response of individual genes. This result could also imply that there has been less conservation of the down-regulated genes over time. The most consistently down-regulated genes were FAM117B and KDM5B, which were both significantly down-regulated in eight of the 10 species analysed in this study. Relatively little is known about the function of FAM117B with the exception that it is a risk factor for sarcoidosis [27]. On the other hand, it has been shown that suppression of the KDM5B gene product, a H3K4 demethylase causing transcriptional repression, results in increased expression of IFN- $\beta$  and other inflammatory cytokines following infection with respiratory syncytial virus (RSV) [28]. We also noticed that, with the exception of the rat, every species analysed down-regulated at least one KDM gene in response to IFN.

It is already established that another form of epigenetic control, acetylation, is also required for robust ISG transcription [29]. ANP32A, a protein involved in acetyltransferase inhibition [30], was down-regulated in five species. Interestingly, ANP32A has been shown to be a host component necessary for influenza virus replication and influences the ability of the virus to replicate in a given animal species [31].

Our data have thus revealed that key epigenetic factors regulating ISG transcription are themselves frequently responsive to IFN treatment. Along these lines, we also investigated the presence of noncoding RNAs (ncRNAs), a class of RNA molecule increasingly recognised as being important in the antiviral response [32,33]. We found that in human cells 75 long intergenic noncoding RNAs (lincRNAs) were differentially expressed (38 up-regulated, 37 down-regulated) in response to IFN (S2 Fig), including NEAT1, a lincRNA that has previously been associated with viral infections [34–36]. As other genomes become increasingly well annotated, it will be possible to resolve a more in-depth understanding of the impact that ncRNAs play in the control of the innate immune system.

## Evolution of ISGs

Virus–host coevolution has shaped the innate immune system, most frequently by placing antiviral genes under positive selection. We assessed the dN/dS ratios (as compared to the human) for one-to-one ISG orthologs. We observed that the overall distributions of dN/dS values of ISGs were significantly higher than those of randomly selected non-ISGs (Fig 4A). In addition, we assessed the copy number of each ISG across the different species studied here. Strikingly, for each species—with the exception of sheep and, to a lesser extent, cow—the proportion of ISGs with gene expansions was significantly above that of the genome as a whole (Fig 4B). Interestingly, sheep and cows are the only species with multiple copies of the IFN- $\beta$  gene (generally a single copy gene in mammals).

The data described above suggest that, in general, expanded ISGs (compared to the rest of the genome) have an increased likelihood of conferring a selective advantage to the host species. Indeed, we observed that ISGs that are shared between multiple species have a higher likelihood of being expanded in the genome compared to other genes ( $P < 0.001$ , Fig 4C). Furthermore, among the core<sup>vert</sup> ISGs, one-to-one orthologs are induced by IFN to a significantly higher level than genes present as paralogs (two-way ANOVA,  $F = 2.284$ ,  $P < 0.05$ ).

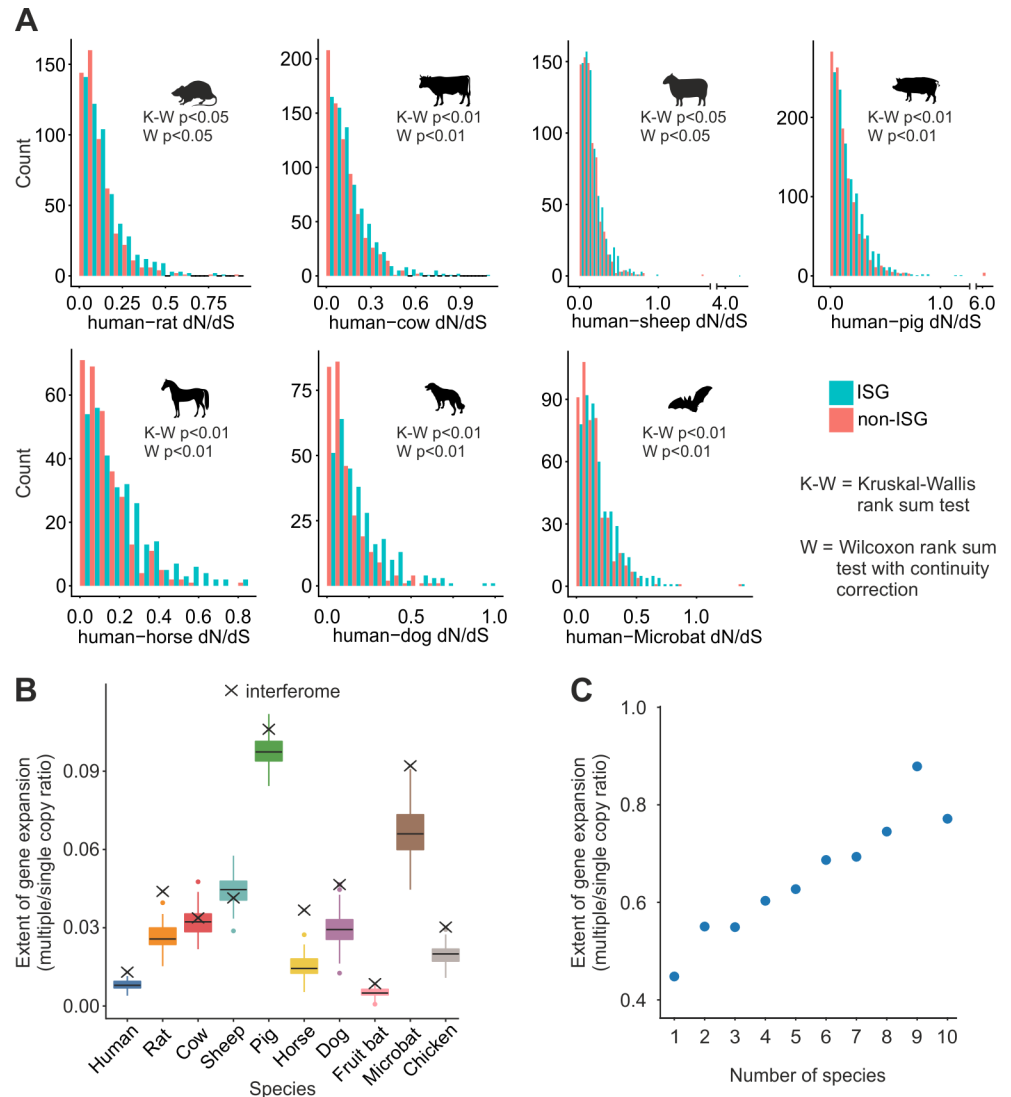
We further analysed gene expansions and deletions among the core<sup>mamm</sup> ISGs in the genomes of 111 mammalian species using an in silico sequence-similarity screening approach [37]. Although the uneven quality of the genomes used in the analysis make this approach prone to artefacts, we were able to detect core<sup>mamm</sup> ISG deletions. For example, we observed that XAF1, which is a negative regulator of inhibitors of apoptosis, is deleted in cats (Felidae) (Fig 5A). In addition, we confirmed previously published deletions of IFIT3 among the Scandentia (tree shrews), Cetacea (whales and dolphins), and marsupials (Fig 5A) [38,39]. Furthermore, we observed that IFIT2 exists as a pseudogene in the Cetacea (Fig 5B).

## Discussion

In this study, we devised a systematic approach to unveil fundamental properties of the type I IFN system in vertebrates. We investigated the IFN response in several mammalian species and the chicken using a consistent experimental framework. By grouping ISGs and IRGs according to the number of species in which they were differentially expressed, we were able to reveal key facets related to the evolution and function of the IFN response.

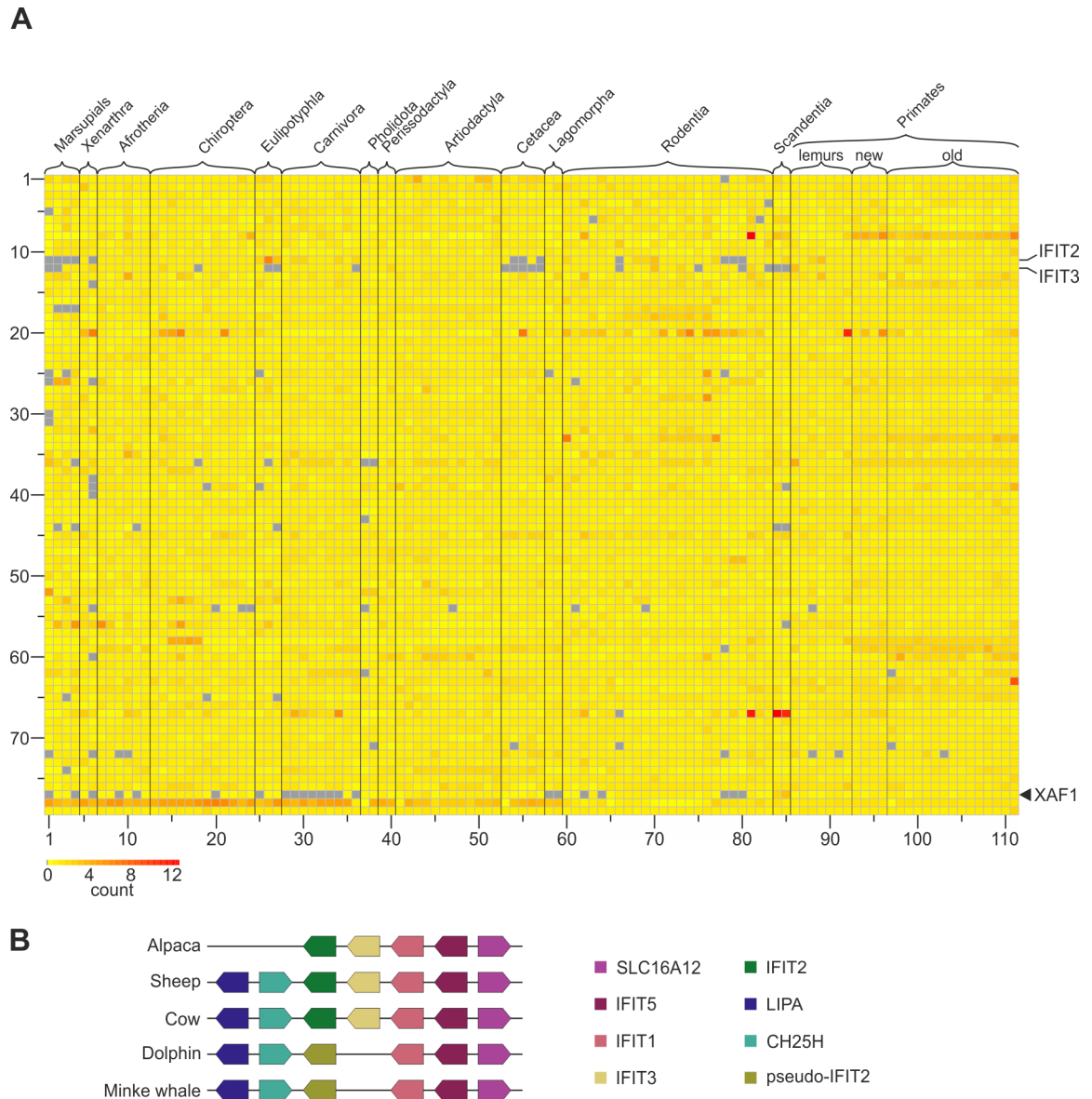
We identified 62 core<sup>vert</sup> ISGs that were up-regulated both in the chicken and nine mammalian species. Similarities between the chicken and mammalian IFN systems likely reflect fundamental functions that were present in the common ancestor of birds and mammals that have remained conserved over the ensuing circa 300 million years. Orchestration of the adaptive immune response by IFN appears to be a fully conserved and prioritised function among vertebrates. Specifically, within the core<sup>vert</sup> ISGs, we found MHC class I components, along with genes involved at all levels of the antigen presentation process and genes involved in cell signalling to diverse cells of the adaptive immune system. Interestingly, we found that the type I IFN response may also facilitate local expression of factors of the complement system.

Only a limited set of genes relating to PAMP detection (MDA5, LGP2, TLR3) and downstream signalling (IRF1, IRF7, STAT1 and STAT2) were within the core<sup>vert</sup> ISGs. In general, we noticed a greater number of IFN-up-regulated RNA sensors (and a greater level of their up-regulation) compared to DNA sensors. These data imply that the type I IFN response biases the sensitivity of surveillance for RNA viruses and may also reflect the difficulty, and danger, of differentiating self from nonself RNA in the cytoplasm.



**Fig 4. Evolutionary properties of ISGs.** (A) For each nonhuman species, ISGs with one-to-one orthologs that were up-regulated in the human interferome and an identical number of random genes not differentially expressed by IFN stimulation were selected. dN and dS values were then retrieved from the Ensembl database. Histograms show dN/dS ratio values for ISGs (blue) and non-ISGs (red). Differences in the distribution of dN/dS values of the non-ISGs compared to ISGs were tested using the Kruskal–Wallis rank sum test and Wilcoxon rank sum test with continuity correction. (B) The extent of gene expansion was compared between ISGs and the genome as a whole. The y-axis represents the ratio between the number of genes for which there are paralogs (multiple) and those which are orthologs (single) as a proxy for gene expansion. Boxes and whiskers represent the values for 500 randomly selected non-ISGs, while ‘x’ represents the mean value for the ISGs for each species. With the exception of the sheep, all ISG values were above the median value. The code used for the random sampling and the generation of Fig 4A and 4B is provided in [S3 Data](#) and [S4 Data](#), respectively, with the input file available as [S5 Data](#). (C) Up- or down-regulated genes were divided into bins according to the number of species in which they were differentially expressed. The extent of gene expansion was calculated as panel B ([S1 Data](#)). A positive trend was identified for up-regulated genes whereby the greater the number of species which up-regulate a gene, the greater the likelihood of copies being retained ( $P < 0.05$ ). IFN, interferon; ISGs, interferon-stimulated genes; K-W, Kruskal-Wallis; W, Wilcoxon rank sum test.

<https://doi.org/10.1371/journal.pbio.2004086.g004>



**Fig 5. In silico screening of core<sup>mamm</sup> ISGs in mammalian genomes.** (A) A heatmap displaying the results of a similarity search-based screen of 111 mammalian genomes for sequences disclosing homology to mammalian core ISGs. Each column represents a distinct mammalian species, while each row represents a core<sup>mamm</sup> ISG. Numbers on the left of each row identify each ISG as listed in S3 Table. Column numbers refer to species (S4 Table). Colours are proportional to the number of matches found in each genome, normalised by the median hit count for that gene across the Mammalia. Because the method is based upon sequence-similarity screening, high count levels for a particular gene do not necessarily reflect gene expansion. Note that the grey boxes indicate that no matches were identified, either due to a bona fide deletion or as a result of relatively poor quality of the genomes. Note that only 79 core<sup>mamm</sup> ISGs are included in the analysis. Some of the core<sup>mamm</sup> ISGs were excluded from this analysis because of their high levels of similarity between each other posing a risk for spurious results. (B) A cartoon of syntenic loci showing the absence of IFIT3 and pseudogenisation of IFIT2 in cetaceans. ISGs, interferon-stimulated genes.

<https://doi.org/10.1371/journal.pbio.2004086.g005>

Known antiviral ISGs are, in general, shared by many species. It therefore appears that a large proportion of the known antiviral capability of the IFN system evolved at an early point, a finding consistent with the presence of antiviral activity among fish ISGs [40,41].

Characterisation of the core<sup>vert</sup> ISGs has also revealed several genes that hitherto have had little, if any, association in previous studies with the IFN response. Our data suggest that these genes play fundamental roles in the innate immune response of vertebrates that remain yet to be discovered. It is possible that genes such as these have been overlooked in previous studies simply as a result of their relatively modest fold up-regulation in response to IFN treatment in cells derived from individual animal species. This dataset therefore provides additional power with which to uncover novel genes central to the IFN system and an alternative approach by which to prioritise their relative biological significance and evolutionary conservation.

We observed that genes have arisen as ISGs throughout evolution, to the extent that certain genes are responsive to IFN only in particular phylogenetic groups. In addition, ISGs shared by multiple species have a higher propensity to be retained in genomes, yet another example of the pressure exerted by invading pathogens shaping vertebrate evolution. Hence, the result of these evolutionary processes is that every species possesses a unique repertoire of ISGs. These findings may help explain the differing sensitivities of certain animal species to specific viruses. For example, it has been widely hypothesised that the bat innate immune system has unique features that allow this species to withstand persistent, asymptomatic infection with viruses that are pathogenic in humans [42–44]. Our data reveal that the overall pattern of the bat interferome is relatively unremarkable: they up-regulate the core ISGs, have similar distributions of up-regulated and down-regulated genes, and up-regulate lineage (order Chiroptera), as well as species-specific, ISGs. However, we found that the basal transcription level of the type I interferome (including the known antiviral ISGs) to be higher in both the megachiropteran and microchiropteran cells compared to cells from other species (S5 Fig), a finding consistent with the previous observation that the interferon alpha (IFN- $\alpha$ ) locus is constitutively active in bats [44]. Hence, bat cells might exist in a relatively constitutively active antiviral state.

The IRGs were differentially expressed to a conspicuously lower extent than ISGs, although the overall pattern was largely uniform across the species. The ability to assemble lists of shared IRGs allowed us to suggest that epigenetic control via down-regulation of genes involved in acetylation and methylation may be a relatively underappreciated function of the IFN response. For example, we found that ANP32A, a protein involved in acetyltransferase inhibition [30], was down-regulated by IFN in human, rat, sheep, cow, and pig cells. ANP32A was recently identified as a key cellular cofactor for avian influenza virus (AIV). Indeed, the avian influenza virus polymerase functions relatively poorly in mammalian cells, and this is due, at least in part, to the inability of AIV polymerase to bind efficiently to mammalian ANP32A [31]. It is intriguing that, in our data, chicken ANP32A is not significantly down-regulated by IFN.

Our study, like many of a similar nature, relies on the quality of the annotations of the genomes used. Indeed, many ISGs have been shown to have complex orthologies, and it is possible that some genes are misannotated or not yet annotated in the Ensembl Compara database. In order to decrease the impact of this factor in our data, we manually curated all ISGs that were initially identified in at least eight animal species. In addition, the use of primary cells for most species (in most cases derived from different individuals) reduced the possibility of artefacts deriving from cells that were passaged extensively *in vitro*. We also ensured that all RNAseq experiments were carried out in cells in which IFN stimulation resulted in an antiviral state (see [Materials and methods](#)). The system-level nature of RNAseq experiments and downstream bioinformatic analyses complicates direct comparisons between distinct studies. MORC3, for example, was previously suggested to be specifically up-regulated by the megabats, as it was not up-regulated in human A549 cells [45]. By contrast, in our study, we observed MORC3 among the core<sup>vert</sup> genes, albeit robustly up-regulated in the fruit bat cells (> 12-fold) and minimally in human cells (< 2-fold). Similarly, the scope of this study is

limited to type I IFN and a single cell type per species. It will be interesting in future studies to observe the differences in interferomes generated by IFN- $\gamma$  and IFN- $\lambda$  and, additionally, the variation that exists between cell types.

It is notable that the list of species for which we generated interferomes includes wild (rat, microbat, and fruit bat), as well as domestic companion (dog and horse) and livestock (pigs, chicken, cow, and sheep), species. We observed clear species- and lineage-specific ISGs for every species examined, which, as more genomes become sequenced, can be explored for evidence of how, for example, the domestication process has influenced the evolution of ISGs. Overall, the dataset described here represents the most comprehensive, cross-species ‘snapshot’ of the IFN response published to date. Our data provide a framework with which it will be possible to test hypotheses pertaining to the role of host innate immunity on virus emergence, cross-species transmission and pathogenesis.

## Materials and methods

### Ethics statement

Ex vivo tissue samples were collected postmortem either at commercial slaughterhouses or at the University of Glasgow School of Veterinary Medicine. In all cases, animals had been euthanized according to protocols approved by the local ethical committee and in accordance with the Council of the European Communities Directive of 24 November 1986 (86/609/EEC).

### Primary cell isolation and cell culture

Ex vivo skin samples were collected from chickens ( $n = 3$ ), cows ( $n = 4$ ), sheep ( $n = 3$ ), horses ( $n = 3$ ), a dog ( $n = 1$ ), and pigs ( $n = 4$ ) and primary fibroblasts isolated following an explant procedure. Briefly, the hair or feathers were removed from the skin prior to disinfection by soaking in 70% ethanol. After rinsing in PBS, the skin was removed from the underlying tissues, cut into circa 3-mm square explants, and added to cell culture dishes (without media, squamous surface uppermost) for one hour at 37°C before adding Dulbecco’s modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), 1% penicillin/streptomycin (p/s) (Sigma) and 100 U/ml nystatin (Sigma). Human primary dermal fibroblasts were purchased from PromoCell (catalogue number C-12302). Rat primary dermal fibroblasts were purchased from the European Collection of Authenticated Cell Cultures (ECACC) (catalogue number 06090769). *M. lucifugus* (little brown bat, representative of the microbats) primary dermal fibroblast cells, isolated from individuals caught in Oregon, United States of America, were kindly provided by William Kohler [46]. *P. vampyrus* (the large flying fox fruit bat, representative of the megabats) cells (PVK4) are an immortalised kidney cell line kindly provided by Megan Shaw [45]. The origin of each cell used in this study is summarized in S5 Table.

Human, rat, and dog cells were cultured in fibroblast growth medium 2 (PromoCell) supplemented with 10% FBS and p/s. 293T and PVK4 were all cultured in DMEM supplemented with 10% FBS and p/s. All cell cultures were incubated at 37°C with 5% CO<sub>2</sub> in a humidified atmosphere.

### Virus infectivity assays

IFN- or mock-treated cells were challenged with infectious VLPs of envelope minus vesicular stomatitis virus (VSV- $\Delta$ G-GFP) decorated with a VSV-G envelope (provided in trans during VLP production) in order to confirm the antiviral state of the cells at the time of RNA harvest essentially as already described [47,48]. Briefly, cells were harvested by trypsinization and fixed

in 5% formaldehyde. The number of infected cells in IFN- and mock-treated cells was assessed by flow cytometry.

### IFN treatment of cells

Parallel sets of cells were plated in multiwell plates 24 or 48 hours prior to IFN treatment and incubated at 37°C. Cells were treated with either 1,000 U/ml Universal interferon (UIFN, PBL InterferonSource), 200 ng/ml canine IFN $\alpha$  (Kingfisher), 1,000 U/ml porcine IFN $\alpha$ 1 (Stratech), or 200 ng/ml chicken IFN $\alpha$  (AbD Serotec). Mock treatment was performed in parallel using DMEM lacking IFN. Cells were incubated for the indicated time period at 37°C, washed with PBS, and either lysed in Trizol (Thermo Fisher) for RNA extraction or challenged with VSV- $\Delta$ G-GFP to assess the antiviral state. Cells were only further processed for RNAseq analyses when they were in an antiviral state. In this study, cells were considered in an antiviral state when IFN stimulation induced at least 75% inhibition of VSV- $\Delta$ G-GFP infectivity (value chosen as average of three independent experiments) (S1 Fig). Pilot experiments were performed for each cell type in order to optimise conditions necessary for cells to reach an antiviral state. With the exception of dog cells, all cells reached an antiviral state after four hours of IFN treatment. The antiviral state in the primary canine cells isolated in these experiments required 24 hours of treatment with canine IFN $\alpha$  (S1 Fig). For comparative purposes, a complete set of experiments ranging from IFN stimulation to interferome analysis was performed in parallel in pig cells stimulated with either UIFN or porcine IFN- $\alpha$  (S3 Fig).

### RNA extraction and sequencing

RNA was extracted using Trizol (Thermo Fisher) and RNeasy (Qiagen) protocols. Briefly, chloroform was added to the RNA-containing phase of the Trizol sample and centrifuged. The aqueous phase was then mixed with ethanol and purified using RNeasy columns, incorporating an on-column DNase step (Qiagen) to ensure the complete removal of genomic DNA. Total RNA samples were quantified using the Qubit (Thermo Fisher) and were assessed for integrity using the Bioanalyser pico eukaryotic II RNA chip (Agilent). Only samples with an RNA integrity number (RIN) value > 9 were taken forward for library preparation.

Libraries of mock- and IFN-treated cell RNA were assembled using equal masses of total RNA. The External RNA Controls Consortium (ERCC) spike control was added to the total RNA sample in order to assess library quality following sequencing. RNA samples were first enriched for mRNA by selecting poly(A) RNA using the Dynabeads mRNA DIRECT Micro purification kit (Thermo Fisher). The eluted RNA was used to generate RNAseq libraries using the Ion Total RNA-seq Kit v2 (Thermo Fisher) following the manufacturer's instructions, with the exception that RNA samples were sheared for just 1.5 minutes. Amplified and purified libraries were checked for quality and quantity using the Agilent TapeStation (D1000 tape) and Qubit (hsDNA assay). Libraries were run on the Ion Proton (Thermo Fisher) according to the manufacturer's instructions.

### Data QC

Raw data were trimmed and assessed for quality using FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). We performed multidimensional scaling (MDS) of normalised counts per million data using EdgeR (Bioconductor) in order to assess the impact of IFN and assessed biological covariance as a further control for data quality.

In order to assess the presence of cell culture contaminants on the transcriptomic data, we used Kraken to assign taxonomic labels to the reads using the MiniKraken database, which contains all complete bacterial, archaeal, and viral genomes in RefSeq [49]. To assess

mycoplasma levels in the cell cultures, we used Bowtie2 to map the data to six different strains of mycoplasma known as frequent contaminants of cell cultures (S4 Fig).

## Data processing and differential expression analysis

Reads were aligned to host genomes (S6 Table) using a two-step procedure. A first round of mapping used TopHat2, followed by a second round of mapping using Bowtie2 in an attempt to map the remaining unmapped reads [50]. HTSeq-Count [51] was used to count reads mapping to genes annotated in .gtf files. Genes with <1 read mapping in at least half of the samples were removed prior to differential gene expression analysis using the EdgeR package (Bioconductor) [52,53]. FDR values were calculated using the Benjamini–Hochberg method. MDS and statistical analyses of the data were performed in R.

## Generation of orthologous gene clusters

We utilised the Ensembl Compara database [54], combined with our expression data, to generate a table of orthologs with the following associated data for each species: Ensembl ID, Gene name, log<sub>2</sub>FC, and FDR following IFN treatment. The Compara database provides a thorough account of gene orthology based upon whole genomes available in Ensembl and thus provided us with a standardised approach by which to define phylogenetically the clusters of orthologous genes relative to the chicken taken as an outgroup in the orthology assignment. However, certain gene families relating to innate immunity (for example, the IFITM genes) have undergone lineage-specific expansion, potentially resulting in genes not being annotated and clustered within the database [55]. To account for the misannotation and absence of genes in Compara, we improved the table by manually checking (and annotating, if necessary) genes initially found to be ‘missing’ in either one, two, or three of the 10 species analysed in this study. For this subset of genes, we searched for the presence of an as-yet-unannotated ortholog in the Ensembl genomes using blastn. In cases where a clear ortholog was detected, we included this gene within the appropriate orthogroup. In total, we identified an additional 18 genes that were added to the final table. In cases where a gene was not annotated in the Ensembl genome but a sequence (or predicted sequence) for the homologous species was available in NCBI, we mapped the RNAseq reads to the gene of the homologous species using Bowtie2. The number of reads mapping from each sample was then counted. In cases where an ortholog was present in NCBI from a closely related species, a relaxed Bowtie2 algorithm was first used to map reads to the sequence. The consensus sequence of the resulting contig was then used to count the reads in individual samples using Bowtie2 as above. Differential expression values were then determined using EdgeR having appended the results to the HTseq file. In total, 64 orthologs were added to the table as a result of the orthologous sequence mapping approach. Finally, we modified the .gtf file of sheep to reannotate STAT4 (to become STAT1 and STAT4) and SOCS1, and manually annotated the ZCCHC2 gene in the cow .gtf file.

## Verifying species-unique genes

Using the human genome as a gold standard for annotation, we next assessed the authenticity of each “species-specific” ISG (i.e., an ISG present only in one of the 10 species analysed in this study) based upon its current annotation within the genomes. We first generated a subset of species-specific genes by applying an arbitrary cutoff of their differential expression of  $\geq 2$  log<sub>2</sub>FC. This cutoff resulted in a total of 102 genes across eight species. A total of 33 genes (32%) were in this category as a result of misannotation. In particular, the current little brown bat and pig genomes appear to be currently less well annotated compared to other genomes, as



a large proportion (73% and 79%, respectively) of seemingly species-unique genes were in reality misannotations.

### In silico genome screening

We used the database-integrated genome screening (DIGS) tool [37] to systematically screen the genomes of 111 mammalian species for sequences disclosing homology to 79 of the 90 core<sup>mamm</sup> ISGs. The peptide sequences of the human copies of these 79 genes (obtained from BioMart [56]) were used as probes for tBLASTn-based searches of each species genome. Sequences disclosing above-threshold similarity to peptide probes were extracted and classified by BLASTx-based comparison to a reference library. This library contained, for each ISG, peptide sequences of human paralogs and orthologs from selected additional mammal species. The DIGS tool captures screening results in a relational database, wherein they could be interrogated using structured query language. The number of significant matches for each gene was determined using a gene-specific bitscore cutoff. These counts were normalised to the median value across the mammalian genomes screened to account for variation in exon numbers. Because DIGS is based upon sequence-similarity screening, high counts for a particular gene do not necessarily reflect bona fide gene expansion. Where no matches to a given gene were identified and no ortholog had been annotated in Ensembl, we attempted to confirm that deletion had occurred by viewing the corresponding genomic region in the UCSC and Genomicus genome browsers [57,58] and by comparing alignments of the orthologous genomic regions derived from species with and without the deleted gene. Deletions could not always be confirmed mainly due to low coverage or relatively poor quality assembly of the available genomes.

### dN/dS ratios for vertebrate interferomes

Clusters comprising one-to-one orthologs present in the interferome of each species were extracted and filtered to check for the presence of a gene in species X matching a human Ensembl ID. The human Ensembl ID was then used to query BioMart to extract the corresponding pairwise dN and dS values for species X against the human. Values for differential expression ( $\log_2FC$ ) and FDR values for species X were merged with the dN/dS values and histograms of both significant (ISG) and nonsignificant (non-ISG) clusters plotted. dN and dS values were not available in Ensembl for the Large flying fox or the chicken. The overall distributions of dN/dS values for ISGs and non-ISGs were compared using the Kruskal–Wallis rank sum test and Wilcoxon rank sum test.

### Webserver

In order to allow mining of our data by the wider research community, we created a web-based public interactive data server, accessible at <http://isg.data.cvr.ac.uk>. The server hosts a database containing the orthologous ISG clusters studied in this paper. This web tool allows users to search for and download orthologous clusters and the associated experimental results. The search may be based on various search criteria:

1. Search clusters by HUGO Gene Nomenclature Committee (HGNC) gene name or Ensembl ID.
2. Select clusters based on the presence and expression characteristics in particular species.
3. Select clusters by presence or expression characteristics in a specific number of species.

It should be noted that genes can have different aliases to those in Ensembl, and these must be checked if a gene that is not initially present, e.g., MDA5, is present as IFIH1. These aliases are stated in Ensembl.

## URLs

The webserver for querying the dataset is available at <http://isg.data.cvr.ac.uk/>.

The DIGS blast framework is available at <http://giffordlabcvr.github.io/DIGS-tool/>.

## Accession numbers

The raw fastq files generated during this project have been submitted to the European Bioinformatics Institute (EBI) under project accession number PRJEB21332.

## Supporting information

**S1 Fig. Generation of transcriptomes used in this study.** (A) Schematic outline of the experimental approach used to generate the interferomes used in this study. Parallel sets of cell cultures were used simultaneously. One set was used to determine the antiviral status of each cell line, while the other set was used to extract RNA and prepare libraries for sequencing. Cells were treated with either type I IFN or mock treated with cell culture media. Cells used for RNA extraction and library preparation were lysed in Trizol and stored at  $-80^{\circ}\text{C}$ . The remaining samples were washed with PBS and challenged with VSV- $\Delta\text{G}$ -GFP. The infection status was assessed using flow cytometry. The extent of the antiviral state was determined by comparing infected (green) and uninfected (grey) cells using flow cytometry. (B) A graph showing the levels of infection by VSV- $\Delta\text{G}$ -GFP of mock-treated (grey bars) and IFN-treated (blue bars) cells. Results were normalised to the level of infection by VSV- $\Delta\text{G}$ -GFP of mock-treated cells (taken as 100%) and are available in [S1 Data](#). GFP, green fluorescent protein; IFN, interferon; PBS, phosphate buffered saline; VSV, vesicular stomatitis virus. (TIF)

**S2 Fig. ISG encoded transcript types.** Barplots showing the types of transcripts generated by human genes significantly up-regulated (left panel) or down-regulated (right panel) in response to IFN. Transcript types were retrieved using BioMart from the Ensembl database. Some genes are annotated with  $> 1$  transcript type; if one of the types was protein coding, this was assumed to be the canonical product. A total of 90% and 92% of up-regulated and down-regulated genes, respectively, are classified as protein coding genes. ncRNAs were present among both up-regulated and down-regulated genes ([S1 Data](#)). IFN, interferon; ISG, interferon-stimulated gene; lincRNA, long intergenic noncoding RNA; miRNA, microRNA; ncRNAs, noncoding RNAs. (TIF)

**S3 Fig. A comparison of pig interferomes resulting from treatment with either UIFN or porcine IFN $\alpha$ .** Interferomes were generated in primary pig fibroblasts using either porcine IFN $\alpha$  ( $n = 3$ ) or UIFN ( $n = 3$ ) ([S1 Data](#)). (A) A heatmap shows the log<sub>2</sub>FC values obtained for the core<sup>vert</sup> ISGs in response to either UIFN or porcine IFN $\alpha$ . A high level of concordance was observed for the core genes in response to the different IFNs. Porcine IFN $\alpha$  up-regulated orthologs of every core gene with the exception of NLRC5. (B) Venn diagrams showing the overlap in the genes that are up-regulated (top) or down-regulated (bottom) by UIFN or porcine IFN $\alpha$ . A total of 89.1% of genes up-regulated by porcine IFN $\alpha$  were also up-regulated by UIFN. Similarly, 82.1% of genes down-regulated in response to porcine IFN $\alpha$  were also down-regulated by UIFN. (C) A scatterplot showing the log<sub>2</sub>FC in expression of ISGs up-

regulated by either UIFN (X axis) or porcine IFN $\alpha$  (Y axis), where every gene is plotted as a point. The goodness of fit was assessed using the coefficient of determination, showing that genes were up-regulated to similar extents by both UIFN and porcine IFN $\alpha$ . IFN, interferon; log<sub>2</sub>FC, log<sub>2</sub> fold change; UIFN, Universal interferon.  
(TIF)

**S4 Fig. Reads generated for each dataset and mycoplasma screening.** Mycoplasma can be detected in every deep sequencing experiment as either environmental or cell culture contaminants (S1 Data). Stacked barplots show the reads generated as part of this study where every bar represents a sample. Reads were assessed using FASTQC, trimmed, and mapped to the appropriate genome. Reads mapping to the host genome were then counted. Dark blue bars are mock-treated samples, and sky blue bars are those treated with IFN. Unmapped reads were further screened using genomes from six species of mycoplasma using Bowtie2 and the counts stacked on to the blue bars (coloured according to mycoplasma species). Only the fruit bat samples had appreciable levels of mycoplasma present.  
(TIF)

**S5 Fig. Constitutive expression of ISGs in cells used in this study.** Barplots showing the median basal FPKM levels for the whole interferome (left) or specific ISGs encoding known antiviral factors (right) for cells derived from each species analysed in this study in the absence of type I IFN treatment. Antiviral ISGs used in these graphs are provided in S2 Table. Both microbats and megabats show statistically higher basal FPKM levels compared to every non-Chiropteran species (Kruskal–Wallis, < 0.05). The data used to derive this figure are accessible in S1 Data. FPKM, fragments per kilobase mapped values; IFN, interferon; ISGs, interferon-stimulated genes.  
(TIF)

**S1 Table. Number of differentially expressed orthologous clusters/genes in response to type I IFN.**  
(DOCX)

**S2 Table. Antiviral genes analysed in this study.**  
(DOCX)

**S3 Table. Core mammalian ISGs shown in Fig 5A.**  
(DOCX)

**S4 Table. Species and genome versions used for the DIGS analysis.**  
(DOCX)

**S5 Table. Origin of cells used in this study.**  
(DOCX)

**S6 Table. Genome statistics.**  
(DOCX)

**S1 Data. The raw values relative to the figures displayed in this study.** This file contains the data used to plot Figures displayed in the manuscript. R code and separate files used as inputs for the R code are present as additional files.  
(XLSX)

**S2 Data. R code used to generate Fig 3B.** This R code requires Data S5 as an input, referred to in the code as Big\_table\_v3.6.1.  
(TXT)

**S3 Data. R code used to generate Fig 4A.** This R code requires Data S5 and Data S6 as input files, referred to in the code as Big\_table\_v3.6.1 and ASEPs, respectively.  
(TXT)

**S4 Data. R code used to generate Fig 4B.** This R code requires Data S5 as an input, referred to in the code as Big\_table\_v3.6.1.  
(TXT)

**S5 Data. Input file for R code used to generate Figures Fig 3B, Fig 4A and Fig 4B.** This file, referred to as Big\_table\_v3.6.1 in the R code, contains data required to run the code present in S2 Data, S3 Data and S4 Data that is used to generate Fig 3B, Fig 4A and Fig 4B, respectively.  
(XLSX)

**S6 Data. List of confirmed antiviral genes required as input for S2 Data.** This file contains data required to run the code present in S2 Data that is used to generate Fig 3B. This file is referred to as ASEPs in the R code.  
(XLSX)

## Acknowledgments

The authors wish to thank the University of Glasgow School of Veterinary Medicine and Pablo Murcia for sources of clinical material, William Kohler, Michael Whitt and Megan Shaw for useful reagents. We additionally thank Megan MacLeod, Paul Garside, Ian Goodfellow, Rick Randall, Simon Babayan, and Chris Boutell for helpful discussions.

## Author Contributions

**Conceptualization:** Andrew E. Shaw, Sam J. Wilson, Massimo Palmarini.

**Data curation:** Andrew E. Shaw, Joseph Hughes, Quan Gu, Abdelkader Behdenna, Richard J. Orton.

**Formal analysis:** Andrew E. Shaw, Joseph Hughes, Quan Gu, Abdelkader Behdenna, Tristan Dennis.

**Funding acquisition:** Sam J. Wilson, Massimo Palmarini.

**Investigation:** Andrew E. Shaw, Joseph Hughes, Abdelkader Behdenna, Tristan Dennis, Richard J. Orton, Mariana Varela.

**Methodology:** Andrew E. Shaw, Joseph Hughes, Quan Gu, Tristan Dennis, Richard J. Orton, Robert J. Gifford, Sam J. Wilson.

**Project administration:** Sam J. Wilson, Massimo Palmarini.

**Resources:** Robert J. Gifford, Sam J. Wilson.

**Software:** Joseph Hughes, Abdelkader Behdenna, Joshua B. Singer, Robert J. Gifford.

**Supervision:** Sam J. Wilson, Massimo Palmarini.

**Validation:** Abdelkader Behdenna.

**Visualization:** Andrew E. Shaw, Joseph Hughes, Tristan Dennis.

**Writing – original draft:** Andrew E. Shaw, Sam J. Wilson, Massimo Palmarini.

**Writing – review & editing:** Andrew E. Shaw, Joseph Hughes, Abdelkader Behdenna, Mariana Varela, Robert J. Gifford, Sam J. Wilson, Massimo Palmarini.

## References

1. Woolhouse MEJ, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens. *Emerg Infect Dis*. Centers for Disease Control and Prevention; 2005; 11: 1842–7. <https://doi.org/10.3201/eid1112.050997> PMID: 16485468
2. Reperant LA, Brown IH, Haenen OL, de Jong MD, Osterhaus ADME, Papa A, et al. Companion Animals as a Source of Viruses for Human Beings and Food Production Animals. *J Comp Pathol*. 2016; 155: S41–S53. <https://doi.org/10.1016/j.jcpa.2016.07.006> PMID: 27522300
3. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*. NIH Public Access; 2011; 472: 481–5. <https://doi.org/10.1038/nature09907> PMID: 21478870
4. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol*. NIH Public Access; 2011; 1: 519–25. <https://doi.org/10.1016/j.coviro.2011.10.008> PMID: 22328912
5. Daugherty MD, Malik HS. Rules of Engagement: Molecular Insights from Host-Virus Arms Races. *Annu Rev Genet*. Annual Reviews; 2012; 46: 677–700. <https://doi.org/10.1146/annurev-genet-110711-155522> PMID: 23145935
6. Meyerson NR, Sawyer SL. Two-stepping through time: mammals and viruses. *Trends Microbiol*. NIH Public Access; 2011; 19: 286–94. <https://doi.org/10.1016/j.tim.2011.03.006> PMID: 21531564
7. Der SD, Zhou A, Williams BRG, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci*. 1998; 95: 15623–8. <https://doi.org/10.1073/pnas.95.26.15623> PMID: 9861020
8. Giotis ES, Robey RC, Skinner NG, Tomlinson CD, Goodbourn S, Skinner MA. Chicken interferome: Avian interferon-stimulated genes identified by microarray and RNA-seq of primary chick embryo fibroblasts treated with a chicken type I interferon (IFN- $\alpha$ ). *Vet Res*. BioMed Central; 2016; 47: 1–12. <https://doi.org/10.1186/s13567-016-0363-8>
9. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res*. Oxford University Press; 2013; 41: D1040–6. <https://doi.org/10.1093/nar/gks1215> PMID: 23203888
10. Erath S, Groettrup M. No evidence for immunoproteasomes in chicken lymphoid organs and activated lymphocytes. *Immunogenetics*. Springer Berlin Heidelberg; 2015; 67: 51–60. <https://doi.org/10.1007/s00251-014-0814-1> PMID: 25403261
11. Magor KE, Navarro DM, Barber MRW, Petkau K, Fleming-Canepa X, Blyth GAD, et al. Defense genes missing from the flight division. 2013; <https://doi.org/10.1016/j.dci.2013.04.010> PMID: 23624185
12. Barber MRW, Aldridge JR, Webster RG, Magor KE, Magor KE. Association of RIG-I with innate immunity of ducks to influenza. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2010; 107: 5913–8. <https://doi.org/10.1073/pnas.1001755107> PMID: 20308570
13. Chen S, Zhang W, Wu Z, Zhang J, Wang M, Jia R, et al. Goose Mx and OASL Play Vital Roles in the Antiviral Effects of Type I, II, and III Interferon against Newly Emerging Avian Flavivirus. *Front Immunol*. Frontiers Media SA; 2017; 8: 1006. <https://doi.org/10.3389/fimmu.2017.01006> PMID: 28878774
14. Gack MU, Shin YC, Joo C-H, Urano T, Liang C, Sun L, et al. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature*. Nature Publishing Group; 2007; 446: 916–920. <https://doi.org/10.1038/nature05732> PMID: 17392790
15. Kuriakose T, Man SM, Malireddi RKS, Karki R, Kesavardhana S, Place DE, et al. ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci Immunol*. NIH Public Access; 2016; 1. <https://doi.org/10.1126/sciimmunol.aag2045> PMID: 27917412
16. Wesche H, Henzel WJ, Shillinglaw W, Li S, Cao Z. MyD88: An Adapter That Recruits IRAK to the IL-1 Receptor Complex. *Immunity*. 1997; 7: 837–847. [https://doi.org/10.1016/S1074-7613\(00\)80402-1](https://doi.org/10.1016/S1074-7613(00)80402-1) PMID: 9430229
17. Barclay AN, van den Berg TK. The Interaction Between Signal Regulatory Protein Alpha (SIRP  $\alpha$ ) and CD47: Structure, Function, and Therapeutic Target. *Annu Rev Immunol*. Annual Reviews; 2014; 32: 25–50. <https://doi.org/10.1146/annurev-immunol-032713-120142> PMID: 24215318
18. Patidar M, Yadav N, Dalai SK. Interleukin 15: A key cytokine for immunotherapy. *Cytokine Growth Factor Rev*. 2016; 31: 49–59. <https://doi.org/10.1016/j.cytogfr.2016.06.001> PMID: 27325459
19. Perrett RM, Alexopoulou Z, Tofaris GK. The endosomal pathway in Parkinson's disease. *Mol Cell Neurosci*. 2015; 66: 21–28. <https://doi.org/10.1016/j.mcn.2015.02.009> PMID: 25701813
20. Velie BD, Shrestha M, François L, Schurink A, Tesfayonas YG, Stinckens A, et al. Using an Inbred Horse Breed in a High Density Genome-Wide Scan for Genetic Risk Factors of Insect Bite

- Hypersensitivity (IBH). PLoS ONE. Public Library of Science; 2016; 11: e0152966. <https://doi.org/10.1371/journal.pone.0152966> PMID: 27070818
21. Pan Q, Rong L, Zhao X, Liang C. Fragile X mental retardation protein restricts replication of human immunodeficiency virus type 1. *Virology*. Elsevier Inc.; 2009; 387: 127–135. <https://doi.org/10.1016/j.virol.2009.02.006> PMID: 19249802
  22. Zhou Z, Cao M, Guo Y, Zhao L, Wang J, Jia X, et al. Fragile X mental retardation protein stimulates ribonucleoprotein assembly of influenza A virus. *Nat Commun*. 2014; 5: 3259. <https://doi.org/10.1038/ncomms4259> PMID: 24514761
  23. Bélanger F, Stepinski J, Darzynkiewicz E, Pelletier J. Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *J Biol Chem*. American Society for Biochemistry and Molecular Biology; 2010; 285: 33037–44. <https://doi.org/10.1074/jbc.M110.155283> PMID: 20713356
  24. Daffis S, Szretter KJ, Schriewer J, Li J, Youn S, Errett J, et al. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*. NIH Public Access; 2010; 468: 452–6. <https://doi.org/10.1038/nature09489> PMID: 21085181
  25. Bajic G, Degn SE, Thiel S, Andersen GR. Complement activation, regulation, and molecular basis for complement-related diseases. *EMBO J*. European Molecular Biology Organization; 2015; 34: 2735–57. <https://doi.org/10.15252/embj.201591881> PMID: 26489954
  26. LAPPIN DF, BIRNIE GD, WHALEY K. Interferon-mediated transcriptional and post-transcriptional modulation of complement gene expression in human monocytes. *Eur J Biochem*. Blackwell Publishing Ltd; 1990; 194: 177–184. <https://doi.org/10.1111/j.1432-1033.1990.tb19443.x> PMID: 1701385
  27. Fischer A, Ellinghaus D, Nutsua M, Hofmann S, Montgomery CG, Iannuzzi MC, et al. Identification of Immune-Relevant Factors Conferring Sarcoidosis Genetic Risk. *Am J Respir Crit Care Med*. American Thoracic Society; 2015; 192: 727–36. <https://doi.org/10.1164/rccm.201503-0418OC> PMID: 26051272
  28. Ptaschinski C, Mukherjee S, Moore ML, Albert M, Helin K, Kunkel SL, et al. RSV-Induced H3K4 Demethylase KDM5B Leads to Regulation of Dendritic Cell-Derived Innate Cytokines and Exacerbates Pathogenesis In Vivo. Heise MT, editor. *PLOS Pathog*. Public Library of Science; 2015; 11: e1004978. <https://doi.org/10.1371/journal.ppat.1004978> PMID: 26083387
  29. Chang H-M, Paulson M, Holko M, Rice CM, Williams BRG, Marié I, et al. Induction of interferon-stimulated gene expression and antiviral responses require protein deacetylase activity. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2004; 101: 9578–83. <https://doi.org/10.1073/pnas.0400567101> PMID: 15210966
  30. Kadota S, Nagata K. pp32, an INHAT component, is a transcription machinery recruiter for maximal induction of IFN-stimulated genes. *J Cell Sci*. 2011; 124. Available from: <http://jcs.biologists.org/content/124/6/892.long>
  31. Long JS, Giotis ES, Moncorgé O, Frise R, Mistry B, James J, et al. Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature*. Europe PMC Funders; 2016; 529: 101–4. <https://doi.org/10.1038/nature16474> PMID: 26738596
  32. Josset L, Tchitchek N, Gralinski LE, Ferris MT, Eisfeld AJ, Green RR, et al. Annotation of long non-coding RNAs expressed in collaborative cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts. *RNA Biol*. Taylor & Francis; 2014; 11: 875–90. <https://doi.org/10.4161/rna.29442> PMID: 24922324
  33. Kambara H, Niazi F, Kostadinova L, Moonka DK, Siegel CT, Post AB, et al. Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res*. Oxford University Press; 2014; 42: 10668–80. <https://doi.org/10.1093/nar/gku713> PMID: 25122750
  34. Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K, et al. Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli. *Mol Cell*. 2014; 53: 393–406. <https://doi.org/10.1016/j.molcel.2014.01.009> PMID: 24507715
  35. Zhang Q, Chen C-Y, Yedavalli VSRK, Jeang K-T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio*. American Society for Microbiology (ASM); 2013; 4: e00596–12. <https://doi.org/10.1128/mBio.00596-12> PMID: 23362321
  36. Saha S, Murthy S, Rangarajan PN. Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J Gen Virol*. Microbiology Society; 2006; 87: 1991–1995. <https://doi.org/10.1099/vir.0.81768-0> PMID: 16760401
  37. Gifford R. The DIGS tool [Internet]. Available from: <http://giffordlabcvr.github.io/DIGS-tool/> (accessed on 11/11/2017)
  38. Daugherty MD, Schaller AM, Geballe AP, Malik HS. Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals. *Elife*. eLife Sciences Publications, Ltd; 2016; 5. <https://doi.org/10.7554/eLife.14228> PMID: 27240734

39. Zhou X, Michal JJ, Zhang L, Ding B, Lunney JK, Liu B, et al. Interferon induced IFIT family genes in host antiviral defense. *Int J Biol Sci*. Ivyspring International Publisher; 2013; 9: 200–8. <https://doi.org/10.7150/ijbs.5613> PMID: 23459883
40. Larsen R, Røkenes TP, Robertsen B. Inhibition of infectious pancreatic necrosis virus replication by atlantic salmon Mx1 protein. *J Virol*. American Society for Microbiology; 2004; 78: 7938–44. <https://doi.org/10.1128/JVI.78.15.7938-7944.2004> PMID: 15254166
41. Wang B, Zhang Y-B, Liu T-K, Shi J, Sun F, Gui J-F. Fish viperin exerts a conserved antiviral function through RLR-triggered IFN signaling pathway. *Dev Comp Immunol*. 2014; 47: 140–149. <https://doi.org/10.1016/j.dci.2014.07.006> PMID: 25058853
42. Papenfuss AT, Baker ML, Feng Z-P, Tachedjian M, Cramer G, Cowled C, et al. The immune gene repertoire of an important viral reservoir, the Australian black flying fox. *BMC Genomics*. BioMed Central; 2012; 13: 261. <https://doi.org/10.1186/1471-2164-13-261> PMID: 22716473
43. Wang L-F, Walker PJ, Poon LLM. Mass extinctions, biodiversity and mitochondrial function: are bats “special” as reservoirs for emerging viruses? *Curr Opin Virol*. 2011; 1: 649–657. <https://doi.org/10.1016/j.coviro.2011.10.013> PMID: 22440923
44. Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN- $\alpha$  in bats. *Proc Natl Acad Sci*. 2016; 113: 2696–2701. <https://doi.org/10.1073/pnas.1518240113> PMID: 26903655
45. Glennon NB, Jabado O, Lo MK, Shaw ML. Transcriptome Profiling of the Virus-Induced Innate Immune Response in *Pteropus vampyrus* and Its Attenuation by Nipah Virus Interferon Antagonist Functions. *J Virol*. American Society for Microbiology (ASM); 2015; 89: 7550–66. <https://doi.org/10.1128/JVI.00302-15> PMID: 25972557
46. Pickering AM, Lehr M, Kohler WJ, Han ML, Miller RA. Fibroblasts From Longer-Lived Species of Primates, Rodents, Bats, Carnivores, and Birds Resist Protein Damage. *J Gerontol A Biol Sci Med Sci*. Oxford University Press; 2015; 70: 791–9. <https://doi.org/10.1093/gerona/glu115> PMID: 25070662
47. Rihn SJ, Foster TL, Busnadiego I, Aziz MA, Hughes J, Neil SJD, et al. The Envelope Gene of Transmitted HIV-1 Resists a Late Interferon Gamma-Induced Block. *J Virol*. American Society for Microbiology (ASM); 2017; 91. <https://doi.org/10.1128/jvi.02254-16> PMID: 28100611
48. Whitt MA. Generation of VSV pseudotypes using recombinant  $\Delta$ G-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines. *J Virol Methods*. NIH Public Access; 2010; 169: 365–74. <https://doi.org/10.1016/j.jviromet.2010.08.006> PMID: 20709108
49. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. BioMed Central; 2014; 15: R46. <https://doi.org/10.1186/gb-2014-15-3-r46> PMID: 24580807
50. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. BioMed Central; 2013; 14: R36. <https://doi.org/10.1186/gb-2013-14-4-r36> PMID: 23618408
51. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. Oxford University Press; 2015; 31: 166–9. <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700
52. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. Oxford University Press; 2012; 40: 4288–97. <https://doi.org/10.1093/nar/gks042> PMID: 22287627
53. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. Oxford University Press; 2010; 26: 139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
54. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. *Database (Oxford)*. Oxford University Press; 2016; 2016. <https://doi.org/10.1093/database/bav096> PMID: 26896847
55. Zhang Z, Liu J, Li M, Yang H, Zhang C. Evolutionary dynamics of the interferon-induced transmembrane gene family in vertebrates. *PLoS ONE*. Public Library of Science; 2012; 7: e49265. <https://doi.org/10.1371/journal.pone.0049265> PMID: 23166625
56. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*. Oxford University Press; 2015; 43: W589–98. <https://doi.org/10.1093/nar/gkv350> PMID: 25897122
57. Kent WJ, Sugnet CW, Furey TS, Roshkin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002; 12: 996–1006. <https://doi.org/10.1101/gr.229102> PMID: 12045153
58. Louis A, Muffato M, Roest Crolius H. Genomicus: five genome browsers for comparative genomics in eukaryota. *Nucleic Acids Res*. 2012; 41: D700–D705. <https://doi.org/10.1093/nar/gks1156> PMID: 23193262