

RESEARCH ARTICLE

Evolutionary divergence of induced versus constitutive antiviral gene expression levels between primates and rodents

Lilach Schneor[‡], Dafna Tussia-Cohen[‡], Evgeny Fraimovitch, Sivan Friedman, Tzachi Hagai^{‡*}

The Shmunis School of Biomedicine and Cancer Research, George S Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

‡ Equal contribution.

* tzachiha@tauex.tau.ac.il



Abstract

Hundreds of genes are upregulated in response to pathogen infection. These genes' sequences often diverge across mammals, to counteract rapid pathogen evolution. However, the transcriptional divergence of these genes, their relative levels before and after infection in different host species, remains poorly understood. We studied this divergence by comparing gene expression before and after viral stimulation in cells from primates and rodents. We developed a method to identify orthologs strongly upregulated in one species that are unchanged in response to stimulus in another species. Using human and mouse data, we detected 578 transcriptionally divergent orthologous genes. For example, genes related to the NF κ B complex are only upregulated in mouse. While most divergent genes do not belong to the same cellular process, several pathways and protein complexes are enriched in this set, suggesting that divergence in immune responses between closely related mammals is limited to specific modules rather than involving entire pathways. Transcriptional divergence between human and mouse orthologs was also observed when ortholog expression from different primates and rodents were compared, when responses were studied in several other cell types, and was recapitulated at the chromatin level, using histone mark patterns that denote active promoter regions. Surprisingly, these transcriptional changes were associated with evolutionary changes in coding sequences only when the genes are lowly expressed. In summary, we found genes whose orthologs diverge between primates and rodents in response to immune stimulation. Some of these genes are constitutively expressed in one species even before infection, potentially facilitating rapid antiviral activity that may be linked to clade-specific adaptation to confer greater resistance against pathogens. Further studies are required to test which of these transcriptional changes are adaptive, and what are their functional consequences. Moreover, comparative studies on diverse infections can point to additional species-specific responses and how they enable different species to overcome infection.

OPEN ACCESS

Citation: Schneor L, Tussia-Cohen D, Fraimovitch E, Friedman S, Hagai T (2025) Evolutionary divergence of induced versus constitutive antiviral gene expression levels between primates and rodents. *PLoS Comput Biol* 21(6): e1013165. <https://doi.org/10.1371/journal.pcbi.1013165>

Editor: Claude Loverdo, Sorbonne University, FRANCE

Received: June 13, 2024

Accepted: May 26, 2025

Published: June 24, 2025

Copyright: © 2025 Schneor 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: The manuscript contains all raw data or relevant references and accession numbers to public depositories, required to replicate the results of this study. Major scripts with relevant details and input files can

be found at: https://github.com/HagaiLab/transcriptional_evolution_of_antiviral_genes.

Funding: This work was supported by funding from the The Israel Science Foundation (ISF, grant No. 435/20) (to TH), a joint QBI/UCSF-TAU research grant in computational biology and drug discovery (to TH), and by a grant from the Zimin Institute for Engineering Solutions Advancing Better Lives (to TH). The funders 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.

Author summary

To limit viral replication, infected cells mount an innate immune response where many genes encoding for proteins with antiviral activities are upregulated. Some of these genes are continuously expressed in cells even without infection, to provide a rapid defense against incoming invaders. The identity of these genes and their level of expression before and after infection can vary between host species, in a manner that can impact the ability of different species to inhibit infection but is poorly understood. In this work, we develop an approach to detect these divergent genes, and characterize the genes that differ in expression between human and mouse cells. We show that these genes also vary in transcription between other primates and rodents and that transcriptional changes are in agreement with differences in chromatin accessibility. The divergent genes belong to specific pathways, suggesting that divergence in these pathways may contribute to primate- or rodent-specific antiviral resistance. Our work thus highlights potential differences between human, primates and rodents in their antiviral immunity, and can point to limitations of animal models in infection studies.

Introduction

The antiviral response is a cell-intrinsic innate immune program where hundreds of genes are rapidly upregulated in response to pathogen infection [1,2]. During the first wave of response, following sensing of pathogen-associated molecular patterns (PAMPs), such as dsRNA, numerous cytokines and chemokines are secreted from infected cells, to alert surrounding cells and immune cells of the infection [1]. Interferon (IFN) is often one of the cytokines secreted and its binding to IFN receptors on nearby cells leads to a second signaling cascade that results in the expression of hundreds of IFN-stimulated genes (ISGs) and to an antiviral state in both infected and uninfected cells [3–5]. Because many of the upregulated genes in both waves of response function directly or indirectly to inhibit infection, they are engaged in an evolutionary arms race with numerous viral proteins, leading to rapid evolutionary changes in the coding sequences of many antiviral genes that can alter their activity and the outcome of infection [6–9]. Unlike coding sequence evolution that has been characterized and mechanistically studied in numerous antiviral genes, the transcriptional evolution of antiviral genes and its potential role in host adaptation against viruses is less well-understood. Previous work that compared transcriptional responses to various PAMPs across species showed that genes that differ in their level of response between species have specific promoter architecture [10–12]. Several focused comparisons of this transcriptional response within clades, pointed to lineage-specific regulatory characteristics. For example, comparison of the antiviral response of skin fibroblasts between two bat species [12] and across mouse strains during lung infection [13] showed that within-species adaptation can upregulate either genes associated with disease resistance or disease tolerance, reflecting different

adaptive strategies against pathogens. Another study that compared antiviral and antibacterial responses in great apes and Old world monkeys showed that apes tend to upregulate a broader array of defenses in comparison to Old world monkeys [13].

Here, we focus on a particular type of transcriptional divergence of antiviral genes between species, where orthologous antiviral genes are induced following immune stimulation in one species but remain transcriptionally unchanged in another species. In these cases, there are two general scenarios: (a) in one species the gene is upregulated in expression from a lower to a higher level following immune stimulation, while in the other species the orthologous gene remains constitutively low in expression, and (b) in one species the gene is upregulated, while in the other species the orthologous gene remains constitutively high before and after stimulation (see Fig 1 for examples where the induced and constitutively expressed genes are human and mouse orthologs, respectively). The two scenarios can be related to two physiologically different states, depending on the level of expression of the orthologous gene that its expression remains unchanged: In the first scenario, the gene is not expressed or is lowly expressed in a given cell type and species before and after stimulation, and thus may not be part of the innate immune defense against pathogens in this species. In the second scenario, the gene is expressed in high levels before and after immune stimuli in one of the species, suggesting that this gene acts as a sentinel in this species, with its constitutively high expression level allowing for a rapid inhibition of viruses. The second scenario may thus represent an evolutionary adaptation where a gene evolved to be highly expressed in unstimulated cells and tissues to provide greater protection against an attack. This scenario was observed in *Pteropus Alecto*, the black flying fox, that its uninfected cells constitutively express IFNs or IFN regulatory factors, which is thought to confer a stronger protection against various viruses [14,15]. A similar “evolutionary switch” in gene transcriptional behavior was observed in the antiviral gene RNase L, that is strongly upregulated in response to IFN in black flying fox cells but is not upregulated in human cells [16]. Epithelial cells from a different bat species, the Egyptian fruit bat, showed significantly increased expression of many complement system genes with respect to human and mouse analogous cells [17]. Finally, few examples of such transcriptional divergence in immune responses were also reported between human and mouse orthologs [10,11], but a global analysis of this pattern of divergence is still lacking.

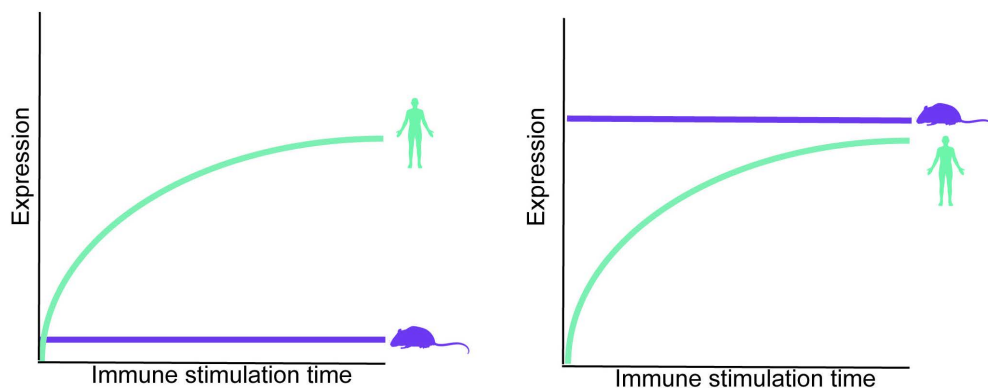


Fig 1. Schematics of the two scenarios of transcriptional divergence between species. In each panel we show the expression level over time of human and mouse orthologs, representing transcriptionally divergent genes we study in this work. In each of these panels, the human orthologous genes is induced following immune stimulation (such as following dsRNA or IFN treatment), while the mouse orthologous gene remains transcriptionally unchanged. The difference between the two scenarios is in the expression level of the mouse gene: In the left panel, the mouse gene level remains low or the gene is not expressed before and after stimulation. In the right panel, the mouse ortholog is already expressed in high levels before stimulation, and remains in this level following stimulation. The two panels represent functionally different scenarios: In the left panel, the mouse ortholog is not expressed and is not part of the immune defense program, unlike its human ortholog. In the right panel, the mouse ortholog is constitutively highly expressed, such that it can immediately function against invading viruses without the need to be induced, unlike the human ortholog that its levels are increased only after infection. The silhouettes were created in BioRender.(2025) <https://BioRender.com/i59u303>.

<https://doi.org/10.1371/journal.pcbi.1013165.g001>

To comprehensively identify cases in which antiviral genes transcriptionally diverge and, specifically, switch between constitutive and induced expression across clades, we chose to use RNA-seq data from comparative cross-species *in vitro* systems, where homologous cells from different species belonging to several clades are stimulated to trigger an antiviral gene expression program: we focused on primate and rodent dermal fibroblasts stimulated with polyinosinic:polycytidylic acid (poly(I:C)), a synthetic dsRNA recognized by intracellular sensors, resulting in a strong upregulation of the antiviral response [18]. The upregulated genes include, among others, a diverse array of cytokines and chemokines, restriction factors that act to suppress viral replication, apoptotic factors and numerous genes related to the regulation of this response. Our analysis focuses on these genes' relative expression before and after stimulation across a set of primate and rodent cells. We used two different, previously published, dsRNA-stimulated cross-species dermal fibroblast systems (Fig 2A): (1) A dataset that includes fibroblasts of two primates (human and rhesus macaque) and two rodents

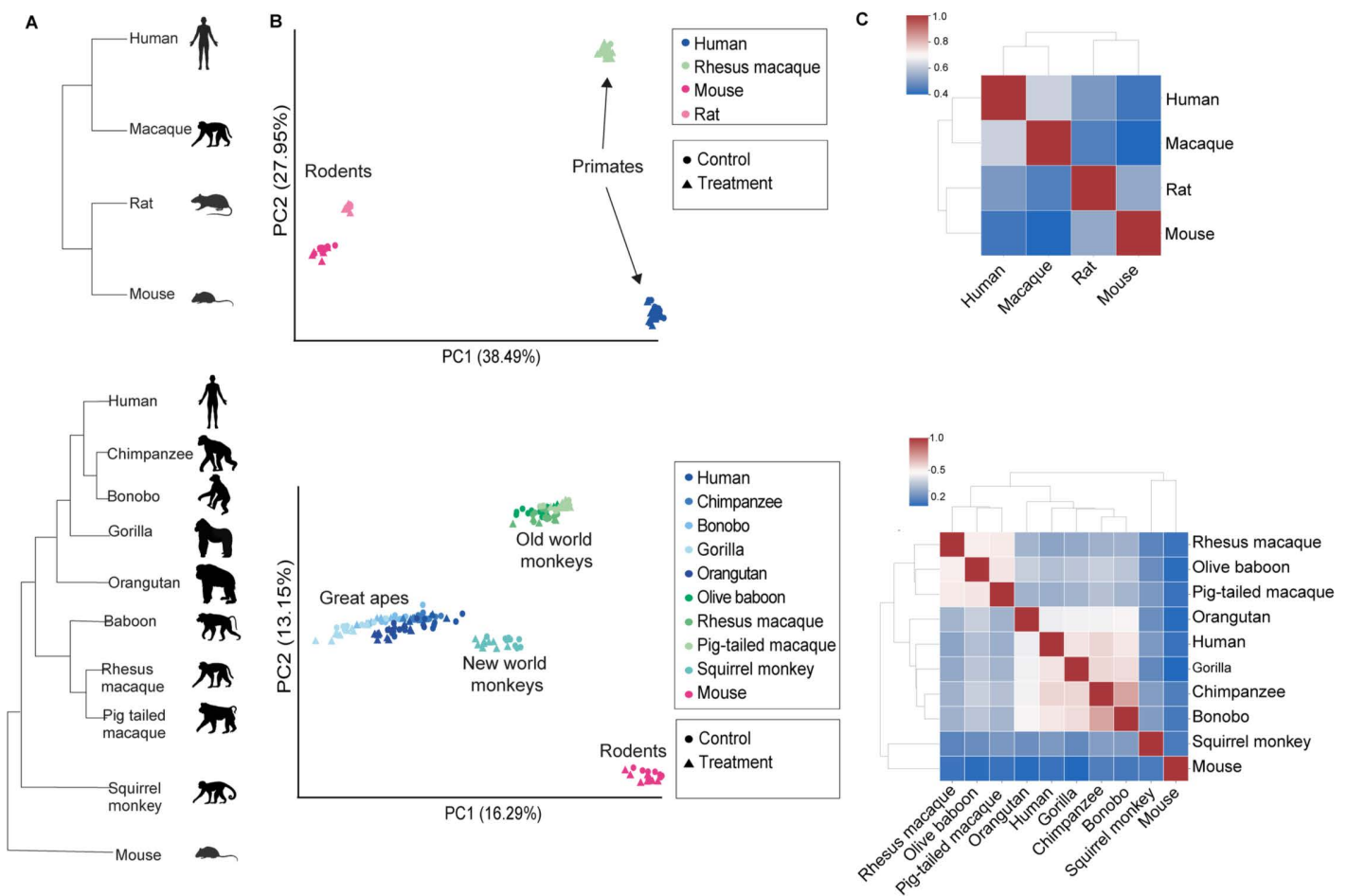


Fig 2. A cross-species *in vitro* immune stimulation analysis shows an overall similarity of transcriptional response across primate and rodent cells with accumulative divergence. (A) Phylogenies of the two cross-species fibroblast dsRNA-stimulation systems used in this analysis. **Top:** 4-species stimulation system: Dermal fibroblasts from two primates and two rodents were profiled with and without dsRNA stimulation using RNA-seq, along with ChIP-seq of histone marks and IFNB stimulation profiling [11]. **Bottom:** 10-species stimulation system: Dermal fibroblasts from nine primates and mouse were profiled with and without dsRNA stimulation using RNA-seq [19]. (B) Scatter plot of principal component analysis (PCA) of expression levels (count values from Salmon mapping [55]). The proportion of variance explained by the principal components is indicated in parentheses. (C) Hierarchical clustering of pairwise Spearman's rank correlations between fold change in response to dsRNA-stimulation values in 1-to-1 orthologous genes that are DE in at least one of the species (2,771 and 7,711 genes, top and bottom, respectively). The silhouettes were created in BioRender.(2025) <https://BioRender.com/i59u303>.

<https://doi.org/10.1371/journal.pcbi.1013165.g002>

(mouse and rat) stimulated with dsRNA for 4 hours [11], hereafter termed the **4-species stimulation system**. The data includes two additional datasets useful for our current analysis: analogous ChIP-seq data of H3K27ac histone marks, that denote active chromatin regions [20,21], in both control and dsRNA stimulation across cells from the four species, and IFN β stimulation of fibroblast cells from the same four species. (2) A second dataset of cells from nine primates (five great apes – human, chimpanzee, bonobo, gorilla, orangutan; three Old World monkeys – rhesus and pig-tailed macaques and baboon; and one New World monkey - squirrel monkey) and one rodent – mouse, unstimulated or dsRNA-stimulated for 24 hours [19], hereafter termed the **10-species stimulation system**. As we demonstrate in the following analysis, in both 4- and 10-species stimulation systems the whole transcriptome and the transcriptional response to dsRNA tend to be more similar between closely related species, suggesting that regulatory changes gradually accumulate over evolutionary time, as observed in other expression programs [20–23]. This allows us to identify specific genes whose primate and rodent orthologs display the distinctive transcriptional divergence we focus on in this work: induced expression following dsRNA-stimulation in one clade and constitutive expression in the second clade. We then study the characteristics of these genes, including their cellular functions and coding sequence evolution, their expression in different human and mouse tissues and cells, and their transcriptional patterns in different species of primates and rodents.

Previous studies employed different methodologies to detect transcriptional divergence in immune response or in other dynamical processes, such as developmental pathways [10–13,22–24]. One of the most common methods is to directly compare the response (or change in expression between conditions) of orthologous genes across species. This “direct” method is easy to use, however it often leads to biases with regards to which genes are truly species-specific. This is because it relies on cutoffs that determine whether a gene is considered responsive or not in each of the species. In cases where one ortholog is above the threshold and another ortholog is slightly below it, may wrongly define this gene as species-specific in its response. The “direct” method may also suffer from caveats regarding technical differences between the species that are often not considered in the analysis but may impact the DE analysis in each species (including genome quality, number of annotated or expressed genes, number and quality of samples, and relative level of reactivity of cells from different species to the stimulus in question). Other approaches include methods that incorporates multiple timepoints and species, however, their results are often not trivial to interpret. Finally, approaches that consider the phylogenetic tree of the species, such as those using Ornstein–Uhlenbeck processes, are often useful to detect evolutionary patterns and selection in specific species or clades, however the usability of these in modelling gene expression evolution has recently been debated [25,26]. The approach developed in this work aims to circumvent some of these limitations, focusing on only two species at a time, and comparing the same condition between these two species. This allows to minimize the reliance on arbitrary cutoffs, overcomes some of the issues involved in species-specific genome quality, and yields results that can be directly interpreted. Furthermore, this method directly points to genes that change between conditions in expression level in one species, while remaining constitutive in their expression in a different species. This constitutive expression is further divided into two classes – of relatively low expression and relatively high levels, that have different physiological meanings. Detection of such genes can be important in various dynamical processes, beyond innate immune response.

Results

Global cross-species analysis of gene expression and transcriptional response to dsRNA

To investigate the global patterns of gene expression similarity across species, we performed a principal-component analysis (PCA) based on the expression of one-to-one ortholog genes in RNA-seq libraries from the 4-species and 10-species datasets (12,811 and 11,350 orthologs, respectively). In each case we used all samples from this system (all individuals, samples from both stimulated and unstimulated conditions and including all species, Fig 2B). In both systems, the PCA plots show a separation by species rather than by condition. In each plot, we observed a separation that largely follows the phylogenetic relationship between the species: a separation between primates and rodents in both studies, and a separation between the major clades of primates in the 10-species dataset.

Next, we performed differential expression (DE) analysis between control and dsRNA stimulation samples within each species in each of the two studies. To study the overall similarity in response to dsRNA, we used the set of DE genes, defined as genes with an FDR-corrected P-value < 0.01 in the DE analysis, in at least one of the species - 2,771 and 7,711 genes in the 4-species and the 10-species dataset, respectively. We compared the Spearman's rank correlation coefficients of the $\text{Log}_2(\text{fold change})$ (logFC, for simplicity we be referred to as "fold change") in response to dsRNA between species: For each pair of species, we computed the correlation between the FC values across the orthologous genes from the DE gene set, using the same set of genes for all pairwise comparisons. We then performed hierarchical clustering using the correlation values across all pairs of species (Fig 2C). We observed that species cluster in a manner that largely recapitulates their phylogenetic relationship. For example, in the 10-species dataset, the FC correlation values of the Old World monkeys first cluster with each other, similarly to the FC values of the great apes. These two clusters form one mega-cluster of all Catarrhini, leaving as outgroups the New World monkey and the rodent, squirrel monkey and mouse, as expected from their phylogeny.

Finally, the numbers of upregulated genes in response to dsRNA (with $\text{FC} > 1$ and $q\text{-value} < 0.01$) are similar between human and mouse - 429 and 419, respectively, out of all 16,534 expressed one-to-one orthologous genes. Furthermore, the correlation between the FC values in all 1-to-1 orthologs, upregulated in at least one of the species, was shown to be high and significant when comparing all paired species [11]. This suggests that the responses are similar across species, allowing us to compare the orthologous gene sets and find specific genes that significantly differ in their responses between pairs of species.

Taken together, the data from these analyses, suggests that the overall gene expression (tested with PCA of orthologous gene expression) as well as the transcriptional response to dsRNA (tested with hierarchical clustering of FC values) are largely conserved between the set of studied species, and that they diverged in a similar manner to their known phylogeny, in agreement with previous analyses of transcriptional divergence data [19,21–23]. This overall similarity with accumulative divergence over evolutionary time, will allow us to focus on genes with particular patterns of divergence in dsRNA-response between the primate and rodent clades. Developing an approach to find these genes and their characterization will be the subject of the following sections.

Identifying primate-rodent divergence between constitutive and induced antiviral gene expression

We next sought to identify genes whose orthologs' expression differ between primates and rodents, such that in one clade the gene is induced following dsRNA stimulation, while in the other clade its expression remains largely unchanged. For this, we first focused on 1-to-1 orthologs between human and mouse from the 4-species dataset (16,534 genes). Using one-to-one ortholog gene expression data in dsRNA stimulation and control of human and mouse we devised the following approach to identify such genes (see **Methods** for details):

- (1) First, we performed a DE analysis between species. Unlike "regular" DE analyses, as in the previous section, where the DE is done between conditions within the same species, here we performed it by comparing gene expression between the human and mouse orthologs within the same condition: i.e., between human and mouse orthologs in stimulated conditions and, separately, between human and mouse orthologs in unstimulated control (Fig 3A). This procedure yielded genes that differ in expression between human and mouse, in either dsRNA-stimulation or in unstimulated control (basal state). DE genes are here defined as those with an FDR-corrected P-value < 0.001 and with a $|\text{Fold Change}| > 1$ between the human and mouse orthologs in the tested condition (Fig 3B).
- (2) Next, we excluded genes that differ between human and mouse in both dsRNA-stimulation and in control. In this manner, we filtered genes that are always higher in one species with respect to the other species, regardless of the antiviral stimulation (e.g., genes that are higher in human versus mouse cells, both in control and in dsRNA stimulation). This was achieved by keeping only genes that were found to be DE between human and mouse ortholog in one

condition (as defined in Stage 1), but whose expression is not significantly different between human and mouse in the other condition (Fig 3C). We thus removed in this stage the set of genes that are less relevant to our study, such as those housekeeping genes that significantly differ between human and mouse, regardless of viral infection.

- (3) Finally, from the resulting filtered set, we focused only on genes that are significantly upregulated in response to stimulation in either human or mouse. This was done based on DE analysis between stimulation and control, within the same species (using the analysis in the previous section). These DE genes are defined as those with an FDR-corrected P-value < 0.01 and whose FC > 0). In this manner, we ensure that the final set of genes includes 246 genes whose expression responds to viral stimulus in at least one species (Fig 3C).

We note that our filtering criteria do not exclude the possibility of detecting a gene that is upregulated in both species but that the FC values significantly differ between the two species. This is because in our dataset we detected only 57 such genes (~10% of the divergent genes), and since we aimed for a minimal set of criteria. However, future users of this method should be aware of this and may choose to add this criterion.

Four groups of human-mouse transcriptionally divergent antiviral genes

The above-mentioned procedure yielded four groups of genes (Fig 4): Two sets of genes that differ between human and mouse in the stimulated conditions, but not in the unstimulated conditions, and two sets that differ between human and mouse only in unstimulated controls. Interestingly, each of the two DE analyses produced a set of genes that is only induced in human and a set of genes that is only induced in mouse, such that we have two sets with orthologs induced

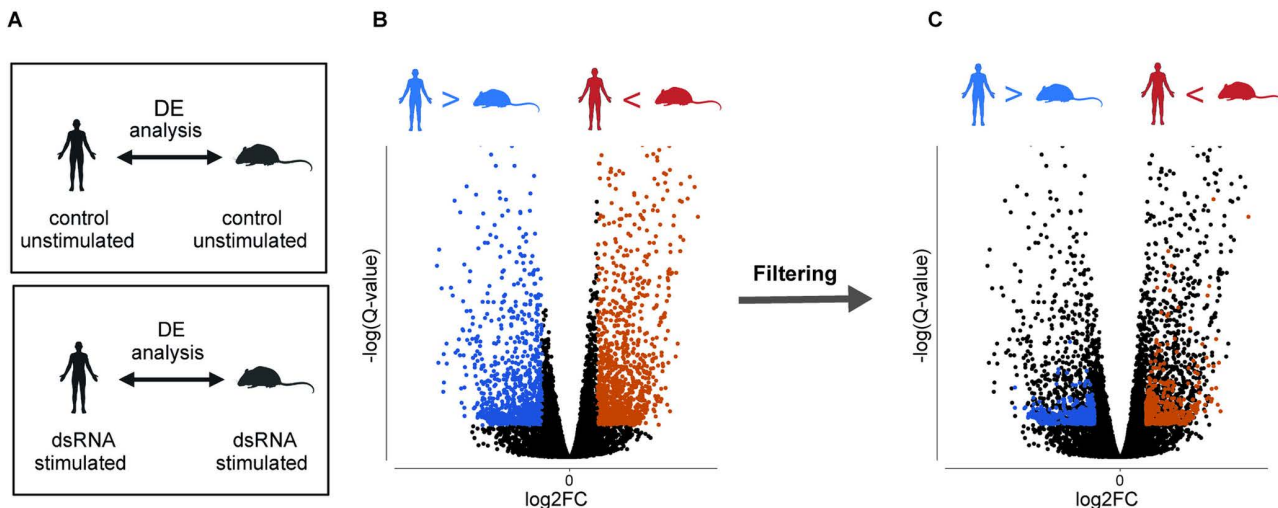


Fig 3. A method to identify divergent genes induced in one species and constitutive in the other. (A) Two separate DE analyses between human and mouse samples were performed in stage 1: One in the control-unstimulated conditions (top) and another in the dsRNA-stimulated conditions (bottom). (B) A representative volcano plot of DE analysis between ortholog gene expression of human and mouse in samples taken in the same condition, as described in stage 1. Each dot represents DE values (Fold Change and Q-value) of a one-to-one orthologous gene between human and mouse. Colored dots represent genes that differ in expression between human and mouse (FDR-corrected P-value < 0.001 and with |Fold Change| > 1), either in the control-unstimulated condition or in the dsRNA-stimulated condition. Orange dots are genes more highly expressed in mouse than in human, and blue are genes expressed more highly in human than in mouse in this specific condition. (C) We employed a filtering procedure to identify genes uniquely upregulated only in one species from the blue and orange sets: First, we filtered genes that differ between human and mouse in both dsRNA-stimulation and control-unstimulated conditions (these genes are genes whose divergence is unrelated to the antiviral response since they are higher in one species regardless of immune stimulation). Next, we excluded genes that are not significantly upregulated in response to dsRNA stimulation in either human or mouse (this ensures removal of genes that are insignificantly upregulated in response to dsRNA treatment in at least one species). The silhouettes were created in BioRender.(2025) <https://BioRender.com/i59u303>.

<https://doi.org/10.1371/journal.pcbi.1013165.g003>

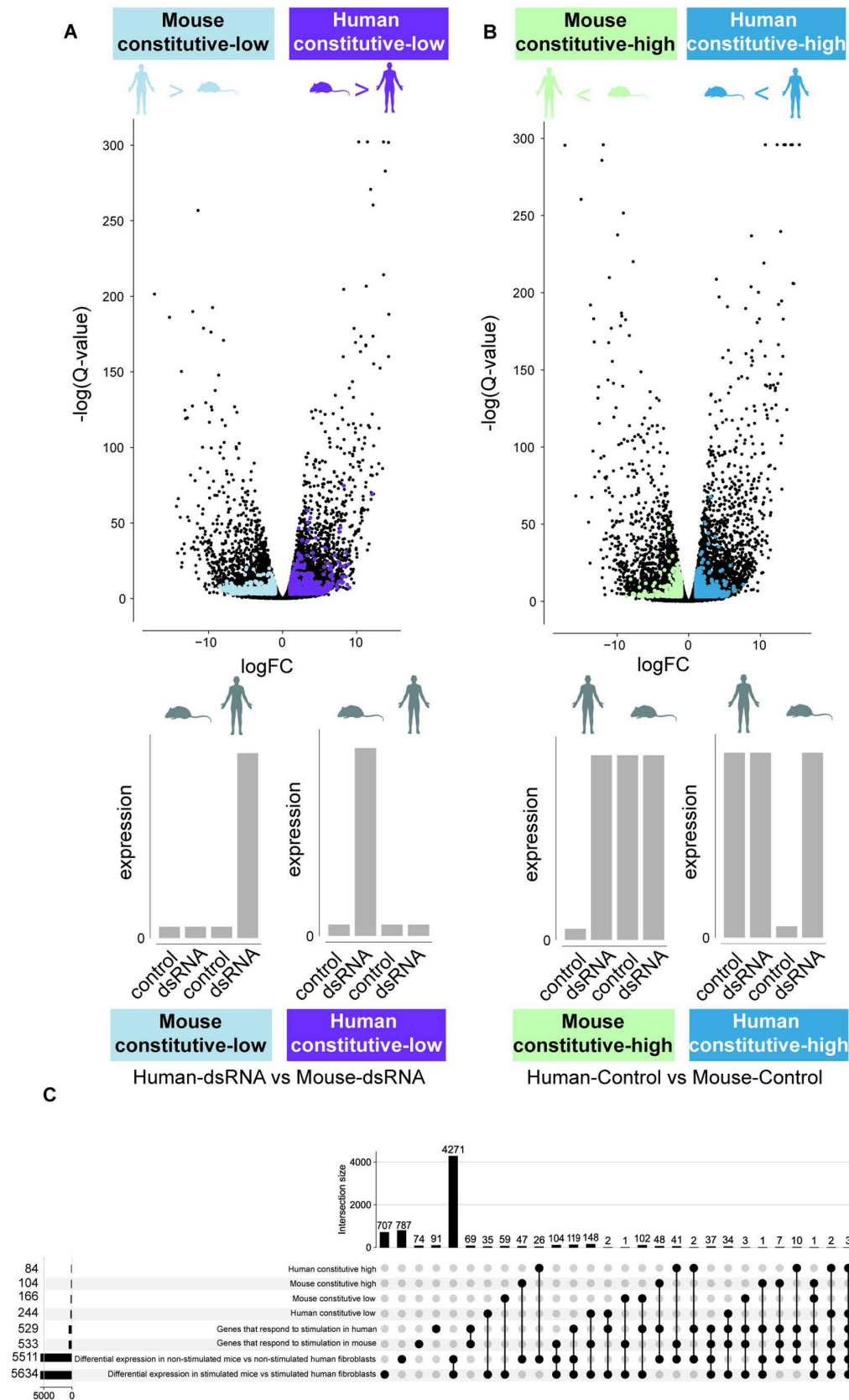


Fig 4. Four groups of human-mouse transcriptionally divergent genes and their patterns of expression. (A) A volcano plot of DE analysis between human and mouse in dsRNA-stimulated conditions. Colored dots show gene sets after the filtering procedure described in Fig 3. The right group (in purple) includes 224 genes induced in stimulation in mouse, but remain constitutively low in human ('**human constitutive-low**'). The left group (light blue) includes 166 genes induced in stimulation in human, but remain constitutively low in mouse ('**mouse constitutive-low**'). **Bottom:** Gray bars illustrate each group's pattern of expression in human and mouse orthologs in both stimulated and unstimulated conditions. (B) The same as in A but comparing human and mouse in the control-unstimulated condition. The right group (turquoise) includes 104 genes induced in stimulation in human, but remain constitutively high in mouse ('**mouse constitutive-high**'). The left group (light turquoise) includes 84 genes induced in stimulation in mouse, but remain constitutively high in human ('**human constitutive-high**'). **Bottom:** Gray bar plots illustrate each group's pattern of expression in human and mouse orthologs in both conditions. (C) An upset plot showing intersection between various gene groups used in this work. The silhouettes were created in BioRender.(2025) <https://BioRender.com/i59u303>.

<https://doi.org/10.1371/journal.pcbi.1013165.g004>

only in human and two with orthologs only induced in mouse. The differences between the two DE analyses lie in the constitutive expression of the orthologous gene: in one DE analysis, the constitutively expressed ortholog has high levels of expression relatively to its induced ortholog, whereas in the other DE analysis the constitutively expressed ortholog has low levels of expression (see Figs 1 and 4, **bottom**). There is an important physiological difference between constitutively high or constitutively low expression of an antiviral gene: If an antiviral gene is expressed highly before infection, it can confer an increased degree of protection against future pathogen infection, without the requirement to be induced. In contrast, constitutively low expression of an antiviral gene in a given species, both before and after stimulation, may suggest that in this species and tissue, the host defense has evolved to resist pathogens without this particular antiviral gene (unlike in the other species, where this gene is induced and may play a role in that defense).

The human-mouse DE analysis in stimulated conditions resulted in two groups of genes: (1) a group of 166 genes induced in stimulation in human, but in mouse remain constitutively low (we term this group '**mouse constitutive-low**'), including several transcription factors, like ETS1 and KLF6, and (2) a group of 224 genes induced in stimulation in mouse, but remain constitutively low in human ('**human constitutive-low**'), including the antiviral factor DAXX (Fig 4A). Similarly, the human-mouse DE analysis in unstimulated control resulted in two additional groups: (3) a group of 104 genes induced in human, whose expression levels in mouse are constitutively high (we term this group '**mouse constitutive-high**'), including the receptor IFNGR2, and (4) a group of 84 genes induced in mouse, that remain constitutively high in human before and after stimulation ('**human constitutive-high**'), including several important restriction factors such as ZC3H7A, ADAR and SLFN5 (Fig 4B). Detailed gene lists appear in S1 Table. In Fig 4C, we added an upset plot, to show the overlap between the different groups included in this analysis. We note that the use of several tests in this method is more restrictive than a single-stage test. This can lead to a smaller and more restrict set of detected genes. We also note that the terms "high" and "low", refer to relative gene levels in the two compared species before and after induction. For example, "low" refers to a scenario where the gene levels are induced to a much higher level than observed in the non-induced condition in one species. In the second species the levels remain similar between the two conditions and are also similar to the non-induced conditions of the ortholog in the first species. However, since these are relative levels, the lower gene expression levels may still be absolutely high. We note that in certain borderline cases, genes can be assigned to two groups (6 genes, as can be seen in Fig 4C). This can happen, for example, when an ortholog in one species is upregulated from very low to very high levels, whereas the ortholog in the second species is upregulated by shifting between two intermediate levels. In the current analysis, we chose not add an additional criterion to remove these genes, since we aimed to minimize the gene selection criteria and since this group was small in the current dataset. In future analyses of different datasets, it is possible to filter out such genes, for example by excluding genes that are upregulated in both species.

Overall, these four sets of genes include 578 genes, out of 1,473 genes that are DE in either human, mouse or both. Thus, 39.2% of the DE genes in this cell system are induced only in one species and remain constitutively high or low in the other species. In Fig 4C, we show the relationship between various groups of genes relevant for these analyses, in terms of overall size and intersection between the different groups. S1 Fig provides additional details on these groups.

To provide an orthogonal approach to test whether the four groups are indeed significantly different in expression across the two species and the two conditions, we used a linear model that includes an interaction term between species and conditions. Indeed, we observe that the distributions of DE values of the four groups found using the new approach, (both fold change values and q-values) differ significantly than the overall set of all DE genes in the interaction term, based on the linear model (see [S2 Fig](#)). In other words, the four groups of genes that are found to transcriptionally diverge in response between human and mouse based on the new approach developed in this manuscript (human constitutive-high and -low and mouse constitutive-high and -low), also have significant values of differential expression in the unrelated linear model, in the specific term that tests whether the two species differ in response to stimulation (the interaction term, species:condition). Thus, the two approaches – the new approach developed here and the linear model, despite being based on different assumptions and using different mathematical tests, are largely in agreement. This orthogonal analysis, using a different model that gives results in agreement with the new approach, provides further evidence for the ability of the new approach to detect relevant genes that transcriptionally diverge between species in their response to stimulation. Additionally, using permutation tests (permuting the conditions within species and repeating the analysis to find divergent genes), we observe that the false positive rate (the fraction of divergent genes detected in the permuted datasets divided by the genes detected in the non-permuted dataset) is below 1% in the approach applied in this work, similar to the rate detected using the interaction test.

Expression levels of antiviral genes across human and mouse individual samples

Our procedure allows for identification of genes that are either up- or downregulated in one species in response to stimulation, while their expression in the other species is not significantly different between stimulation and control. To visualize the expression of these genes across species and conditions, we plotted heatmaps of their normalized expression ($\log_{10}(\text{TPM})$ - transcript per million) across all mouse and human individuals in both dsRNA stimulation and control conditions ([Fig 5](#)). Using hierarchical clustering, we observed that all individual from both species cluster together based on the observed pattern of expression rather than by species: For example, in “human constitutive-low” genes, we observed that the human samples, from both stimulated and unstimulated conditions, cluster with the mouse unstimulated samples, while the mouse stimulated samples are separated ([Fig 5A](#)). This is consistent with our characterization of this set as genes that are constitutively low in mouse while being induced in humans, based on the cross-species DE approach described above. Thus, the unstimulated level of expression in human is similar to the level of expression of their mouse orthologs in both stimulated and unstimulated conditions.

We also observed that the genes within each group are separated into two or three clusters. Each of these clusters represents a different level of relative expression. For example, in “human constitutive-low” genes, we observed two clusters of genes: In both clusters the level of expression in stimulated conditions in mouse samples is higher than the other three types of samples (unstimulated mouse samples, stimulated and unstimulated human samples). However, the higher cluster (in shades of blue to white) represents a set of genes that are lowly expressed and in stimulation of mouse cells increase in expression roughly by two folds. The lower cluster (in shades of light and dark red) has a similar pattern, but its gene levels are significantly higher (the basal expression levels are of $\log_{10}(\text{TPM})$ of ~ 1 , while the induced mouse gene levels are of $\log_{10}(\text{TPM})$ of ~ 2). Thus, within each of the four sets of divergent human-mouse genes, we obtained several subsets of genes with different relative levels of expression in both unstimulated and stimulated conditions.

Antiviral genes switching between induced and constitutive expression between primates and rodents are enriched in specific pathways and protein complexes

In the previous sections we identified groups of antiviral genes that switch between constitutive and induced expression when compared between human and mouse. We next asked whether these genes are enriched in particular pathways or

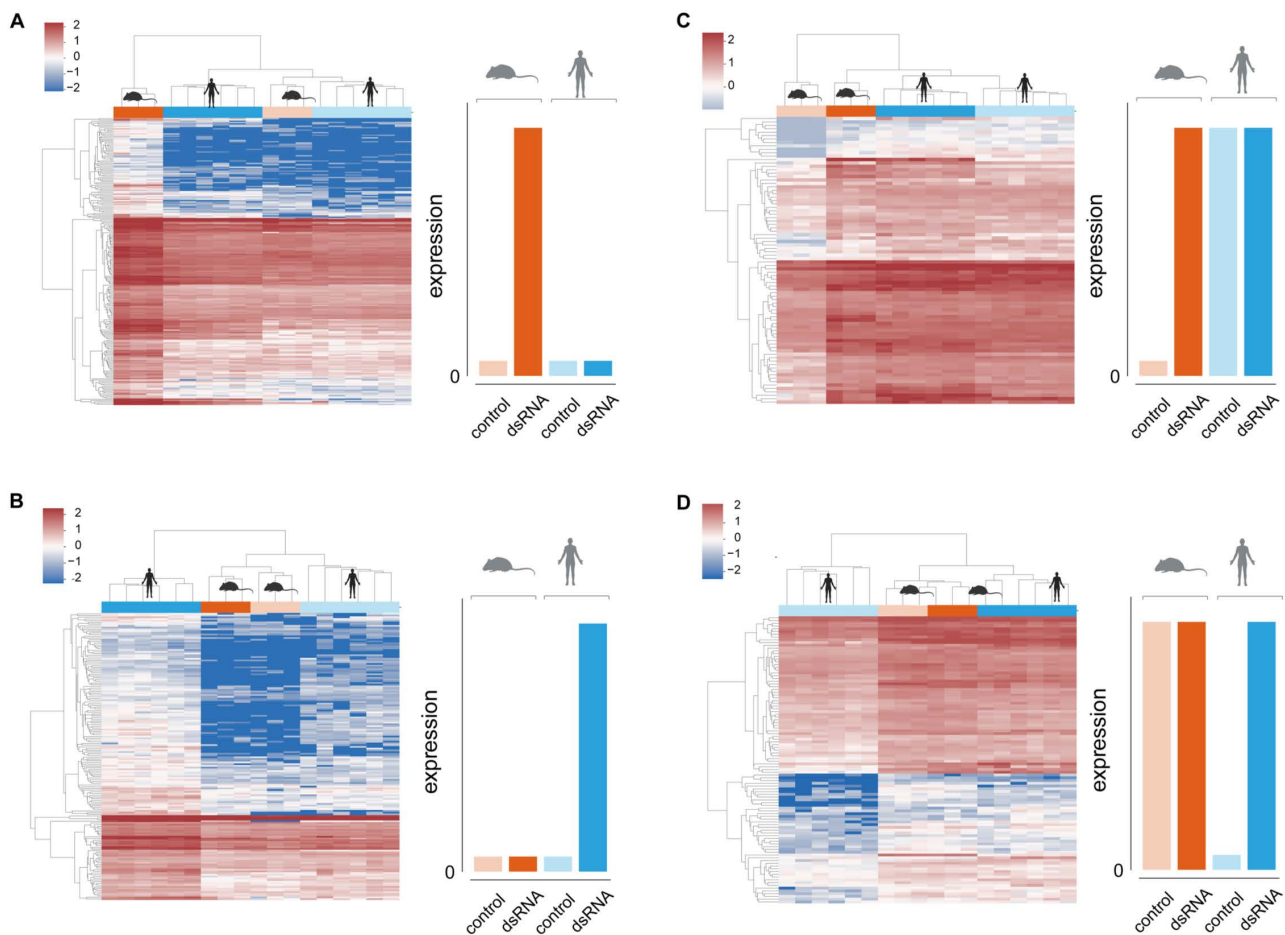


Fig 5. Hierarchically clustered heatmaps based on gene expression levels. Hierarchical clustering of gene expression data from all human and mouse individuals in the 4-species system in both stimulated and unstimulated conditions. In each panel, one the four previously defined divergent genes groups is shown, with expression levels in the human and mouse orthologs: **(A)** 'human constitutive-low', **(B)** 'mouse constitutive-low', **(C)** 'human constitutive-high' and **(D)** 'mouse constitutive-high'. Columns are colored by species (blue and orange for human and mouse, respectively) and by conditions (dark and light for stimulated and unstimulated conditions, respectively) and represent one individual each. Gene levels are shown on a scale based on $\log(\text{TPM})$. The clustering indicates that genes have similar levels in three 'types' of data, while the fourth is significantly different. For example, in A, both stimulated and unstimulated human samples cluster with the mouse unstimulated samples, while the stimulated mouse is the outgroup, with significantly higher levels of this gene set. This agrees with the patterns observed in Fig 4. The silhouettes were created in BioRender.(2025) <https://BioRender.com/i59u303>.

<https://doi.org/10.1371/journal.pcbi.1013165.g005>

functions. To this end, we employed GO term enrichment analysis, using the g:Profiler program [27]. As expected, the four sets are enriched in 'general' immune-related terms, such as those related to cytokine regulation and to response to virus. This is expected given their upregulated in response to dsRNA stimulation in at least one species.

In addition to these terms, in each of the four sets we also found enrichments that refer to more specific pathways or cellular functions, as well as genes whose protein products belong to the same protein complex, as defined by the CORUM database [28], a collection of experimentally verified mammalian protein complexes. In the 'human constitutive-high' group, the genes are associated with "response to cytokine". Another enriched term is "RNA polymerase I transcription regulator complex" that refers to two genes - TAF1A, TAF1D, members of the SL1 complex that is important for the assembly of the RNA polymerase I preinitiation complex [29,30]. Furthermore, the same set is enriched with genes that are downstream to TNF signaling pathway (see S2 Table). In contrast, the 'mouse constitutive-high' group is enriched

with genes associated with the “circulatory system development” pathway (including BMP7, KLF4, TNFAIP2 and FGF18) and with genes related to “cell-cell signalling” (these include the complement system gene C3, the cytokine-related gene TNFAIP6, and other genes, not only related to immune processes, such as growth factors and transcription factors). In the ‘mouse constitutive-low’ set, the enriched terms include “Cell communication” as well as more specific terms associated with leukocytes and their regulation, e.g., “leukocyte cell-cell adhesion”. We also found two members of the OAS pathway (a central pathway in viral RNA sensing [31]) - OAS2 and OAS3. In ‘human constitutive-low’ we found at the top of the enriched pathways genes related to ubiquitin ligation, and genes related to up- and downstream cytokine signaling (either “cytokine production” and “response to cytokine”). In addition, we found various viral sensors and genes involved in upstream regulation of the innate immune response, the inflammatory response and IFN upregulation, including several NOD-like receptors and genes belonging to the inflammasome (NOD1, NLRC5, NLRP3) [32], the NFκB complex (NFκB1, NFκBIE, IKBKE), as well as TRAF2, TBK1 and TANK that belong to the kinase maturation complex and are involved in NFκB signaling and cell survival [33,34]. We note that the fold change values of a set of NFκB target genes are indeed higher in mouse than in human, as expected based on this finding (S2 Fig). Altogether, this suggests that there is no major antiviral gene pathway that all of its gene components shift in regulation between induced and constitutive gene expression across primates and rodents. However, such transcriptional changes occur within smaller sets of genes belonging to specific signaling or regulatory pathways and in particular modules and protein complexes. We note that when testing functional enrichment against the set of DE genes, we do not obtain enriched terms, likely due to the limited size of the background set

Coding sequence evolution and duplication rate of transcriptionally divergent antiviral genes

We next asked whether the identified antiviral genes that switch in regulation between constitutive and induced expression, also evolve rapidly in other evolutionary mechanisms including coding sequence evolution and fast rate of gene duplication and loss. For this, we compared the coding sequence evolution of the four sets of genes with the entire set of antiviral genes (defined as all genes that are DE in either mouse, human or both in response to dsRNA). When looking at sequence similarity between the human and mouse orthologs or at their dN/dS values (the ratio of non-synonymous to synonymous substitutions), we observed that the mouse constitutive-low group tend to evolve faster in sequence evolution, displaying lower sequence identity and greater dN/dS values (Fig 6A and 6B). This trend was also observed in the human constitutive-low group, but was not statistically significant. The human and mouse constitutive-high groups showed an opposite trend of higher sequence conservation than the group of all DE genes, that was stronger in the mouse group of genes.

When looking at the inferred evolutionary age of genes in the various sets, using age estimations from ProteinHistorian [35], we observed that the mouse constitutive-low gene set was enriched with evolutionarily young genes while the group of mouse constitutive-high genes was enriched with evolutionarily ancient genes (Fig 6C). The two groups of human constitutive high and low did not show significant differences in age distribution from the set of all DE genes. Finally, we did not observe significant differences in the rates of gene duplication and loss in the four sets versus the DE set (Fig 6D).

Thus, antiviral genes that exhibit a switch between constitutive and induced gene expression between human and mouse, show higher sequence divergence when the species-specific constitutive expression is low, and the opposite is true when the species-specific constitutive expression is high. These patterns are in agreement with previous findings on the relationship between gene expression and sequence conservation [36].

Human-mouse divergence in dsRNA response is recapitulated at the chromatin level

We next tested whether the expression patterns we observed at the transcriptional level are also observed at the chromatin level. For this, we used ChIP-seq data from the same system where the human and mouse dermal fibroblast response to dsRNA was profiled using RNA-seq. We focused on the histone mark H3K27ac, that is associated with higher activation of transcription [11,37,38], and used H3K27ac ChIP-seq peaks in the vicinity of gene’s transcription start site (TSS)

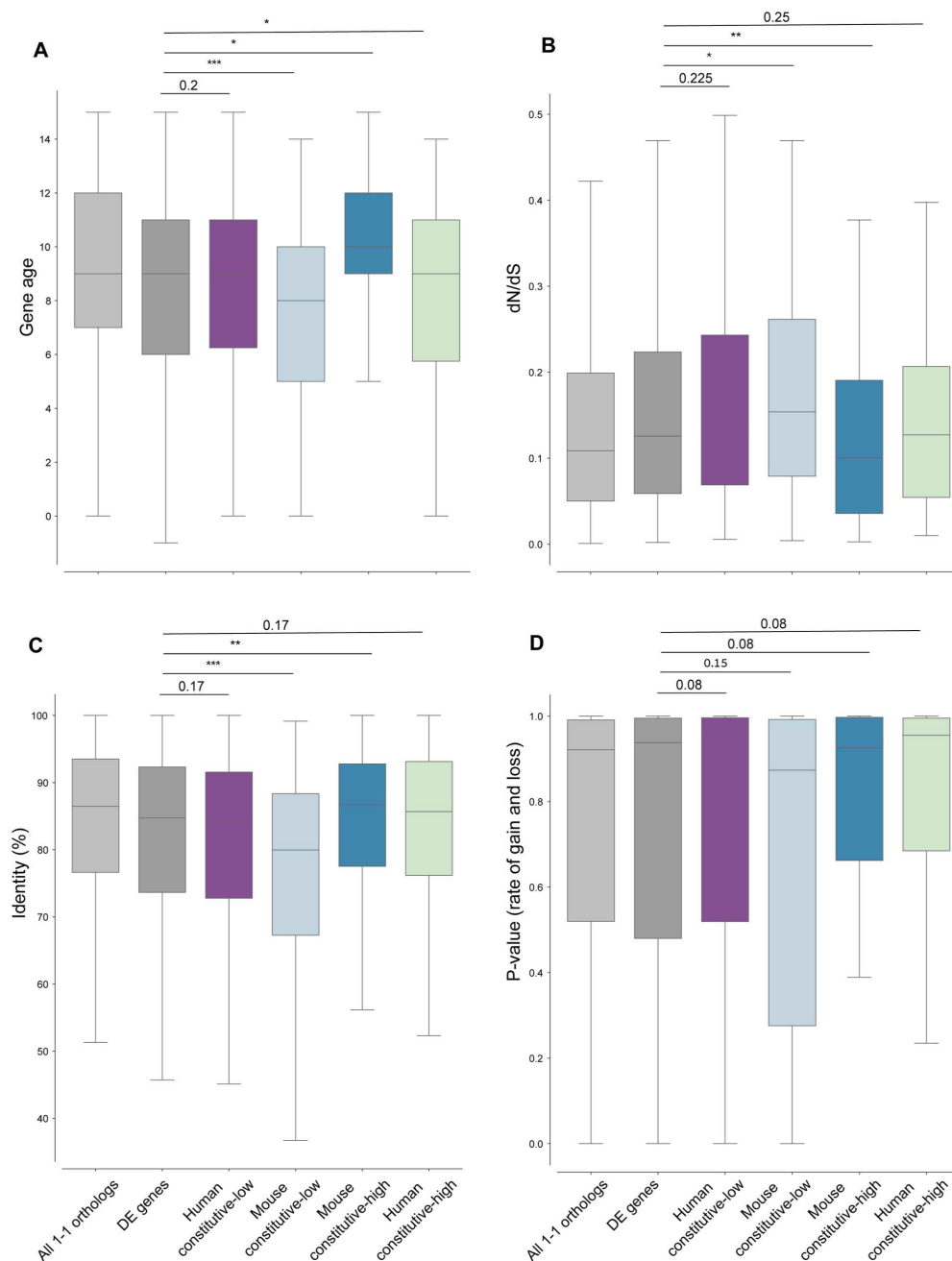


Fig 6. Evolutionary characteristics of genes that transcriptionally diverge between human-mouse in dsRNA response. Distributions of (A) percentage of sequence identity between human and mouse orthologs, (B) ratio of non-synonymous to synonymous substitutions (dN/dS), (C) gene evolutionary age (higher values denotes older age), and (D) rate of gene gain and loss across vertebrates (p-values for each gene's rate are shown), for (from left to right): the set of DE genes (in grey) that includes all genes significantly upregulated in either human and/or mouse in response to dsRNA stimulation (FDR-corrected P-value<0.01 and FC>0), human constitutive-low, mouse constitutive-low, mouse-constitutive-high, human-constitutive high. Group colors, sizes and genes are as in Fig 4. FDR-corrected P-values are shown, one-sided Mann–Whitney tests were performed for each of the 4 groups against all DE genes. For comparison, the values of the group of all 1-to-1 orthologs are also shown.

<https://doi.org/10.1371/journal.pcbi.1013165.g006>

(peaks overlapping with a region 2,000 bp upstream to 500 bp downstream of the TSS), to define active promoter regions (see **Methods**). We compared the presence of these active histone marks in human and mouse cells in unstimulated and dsRNA-stimulated conditions in the four groups of genes we identified.

We reasoned that transcriptional activity across species and conditions should be reflected in patterns of active chromatin marks. For example, in the group of ‘mouse constitutive-high’ genes, the genes are highly transcribed in mouse in both stimulated and unstimulated conditions, but in human cells they are only highly transcribed during stimulation. In this case, histone marks should be present in both conditions in mouse, but only in the stimulated conditions in human cells, if the chromatin marks reflect the transcriptional behavior. To this end, we indeed observed this expected pattern in promoter regions of this gene set: an enrichment in the presence of the histone marks in mouse in both stimulated and unstimulated conditions and only following stimulation in human (FDR-corrected P-value=0.0066, Fisher’s exact test, [Fig 7A](#)). This enrichment was observed in all four groups of genes and was statistically significant in three of the four (in ‘mouse constitutive-low’ the P-value is 0.44, Fisher’s exact test) ([Fig 7A](#)).

We note that a fraction of the genes in these sets has ChIP-seq peaks in all four states or in none of them (human and mouse cells, stimulated and unstimulated conditions). This is also true for many of the DE genes in general, and either reflects smaller dynamics at the chromatin level than at the transcriptional level (where the chromatin remains accessible even without transcription), or the detection and quantification limits using ChIP-seq. In summary, the divergence observed between human and mouse in transcriptional response to dsRNA is also reflected to a large and significant extent at the chromatin level. We note that comparative promoter sequence analyses, between human and primate orthologous genes, to discern specific changes in *cis*-regulatory changes that impact transcriptional changes between species, are hampered by the numerous changes that these promoters have undergone following the split between rodents and primates (see [S3 Fig](#)). We also tested active enhancer conservation and their linkage to transcriptional conservation in a similar manner (see **Methods**). Unlike in the above-described promoter analysis, we did not find a significant enrichment, possibly due to the challenges of linking enhancers with target genes and the rapid turnover of enhancers.

Human-mouse divergence in dsRNA response is reflected in secondary response induced by interferon

Next, we compared human and mouse dermal fibroblast response to interferon (IFN), using stimulation data obtained in parallel to the dsRNA stimulation in the original study [11]. IFN is rapidly secreted following viral or bacterial infection, leading to a strong secondary wave of response against pathogens, including upregulation of hundreds of interferon stimulated genes (ISGs), some of which are also induced in response to dsRNA. Similarly to the analysis of the human-mouse dsRNA stimulation system (shown in [Figs 3 and 4](#)), we performed DE analysis to compare between human and mouse cells in unstimulated conditions and, separately, comparing IFN-stimulated human and mouse cells. We then tested whether genes previously detected to diverge between human and mouse in dsRNA, diverge in a similar manner in IFN response (e.g., genes that were defined as “mouse constitutive-low” in the dsRNA stimulation analysis would behave in a similar manner in unstimulated and IFN-stimulated cells). For this, we tested whether the resulting fold change in ortholog expression in the IFN system is similar to what was found between human and mouse in the dsRNA system. We observed that transcriptional trends in response to IFN in human and mouse largely and significantly match those found in the dsRNA response ([Figs 7B, 7C and S4](#)).

For example, when splitting the genes in the IFN response to four classes, as done previously with the dsRNA stimulation data, we observed that genes originally defined as “human constitutive-low” based on the dsRNA system have the highest FC values in the IFN system while the “mouse constitutive-low” genes have the most negative FC values ([Fig 7B](#)). These trends suggest similarities in transcriptional divergence between the dsRNA and IFN systems that were also observed in the sets defined as “mouse constitutive-high” and “human constitutive-high” ([Fig 7C](#)). Additionally, when separately using the TPM values in human and mouse replicates, we observed that the expression patterns observed in the dsRNA stimulation in the four-divergent groups between human and mouse was largely reflected in the IFN stimulation ([S4 Fig](#)).

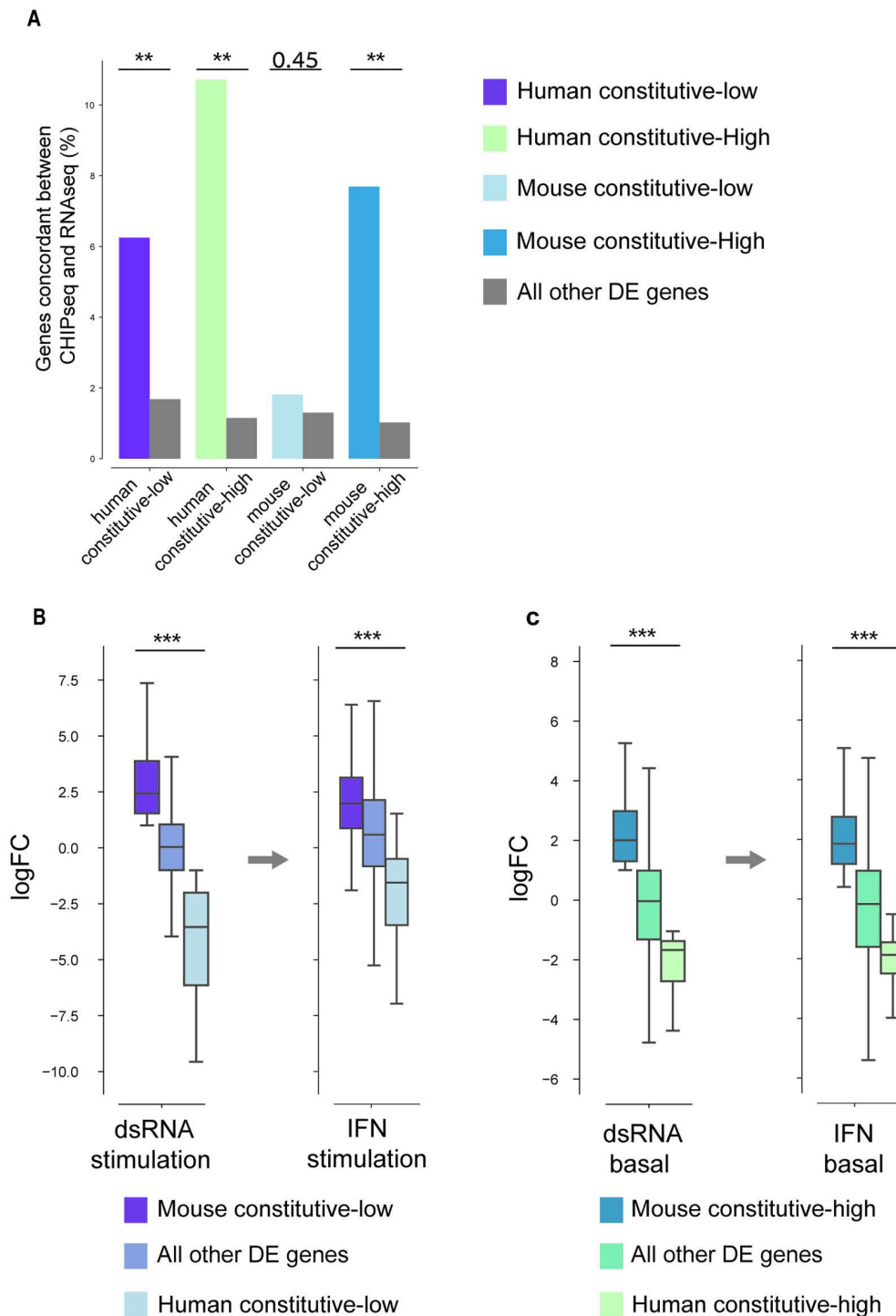


Fig 7. Human-mouse divergence in dsRNA response is recapitulated at the chromatin level, in IFN response and across different cell types. (A) Fraction of genes whose pattern of promoter activity agrees with gene expression, matching one of the 4 divergent groups (colored as in the legend) versus the fraction of genes among the rest of DE genes that display this pattern (in grey). For example, ~6% of human constitutive-low genes show the expected pattern at the chromatin level, and ~2% of all other DE genes show the same pattern (two left-most bars). FDR-corrected Fisher's exact test P-values are shown. These tests show that the pattern of transcriptional response is recapitulated at the chromatin level in all four groups since the fraction of genes whose chromatin accessibility pattern agrees with the transcriptional pattern is higher than that observed in the set of all other DE genes (this is significant in three of four groups tested). (B) logFC distribution values based on DE analyses between human and mouse in dsRNA stimulation

(before the arrow) and in IFN stimulation (after the arrow). In both cases the genes are partitioned into three boxplots based on three groups defined previously: “human constitutive-low”, “mouse constitutive-low” and “all other DE genes”. Thus, the same three groups of genes are shown before and after the arrow, but their FC values varies based on either dsRNA- or IFN-stimulation DE values. The analysis shows that the separation in logFC values in the dsRNA-stimulation between the three groups is also observed in the IFN-stimulation, suggesting similar transcriptional divergence between dsRNA response and IFN response. **(C)** The same as in B, but with basal conditions used in the DE analyses and showing: “mouse constitutive-high”, “human constitutive-high”, “all other DE genes”. In both B and C, FDR-corrected P-values are shown for one-sided Mann–Whitney tests, performed under the hypothesis that the FC distribution of the left group is higher than the right group.

<https://doi.org/10.1371/journal.pcbi.1013165.g007>

From these results we conclude that transcriptional divergence between human and mouse in IFN stimulation largely follows the same patterns as those originally observed in dsRNA stimulation. This serves as a further validation for the divergence we originally found in these genes transcriptional responses dsRNA between human and mouse.

Human-mouse divergence in dsRNA response is recapitulated in other primate and rodent species

Next, we asked whether the divergence observed between human and mouse genes in dsRNA-response represents an overall divergence in dsRNA transcriptional responses between the primate and the rodent clades. To test this, we first repeated our analysis with dsRNA stimulation using the rhesus macaque and rat data from the same 4-species system we took the human and mouse data from. As in the human-mouse dsRNA stimulation analysis, we contrasted the macaque and rat 1-to-1 orthologs in the same condition (where samples from both species are either stimulated with dsRNA, or unstimulated). We then tested whether the resulting fold change in ortholog expression (between macaque and rat) is similar to what we found between human and mouse: In the comparison of macaque-rat dsRNA stimulated samples, we again divided the genes based on their orthologous human-mouse gene divergence and compared their FC in response to dsRNA in the macaque versus rat comparisons. We observed that in the comparison of FC between rat and macaque, the group of genes originally defined as ‘human constitutive-low’ has the highest FC values between rat and macaque while the group of ‘mouse constitutive-low’ genes has the most negative FC values (**Fig 8A Top**). We observed the same trend when contrasting rat and macaque samples in control conditions and using the gene sets originally defined as ‘mouse constitutive-high’ and ‘human constitutive-high’ (**Fig 8A Bottom**). This suggests that the differences in constitutive versus induced expression of antiviral genes originally found between human and mouse are consistent with the expression patterns of their orthologous genes in macaque and rat.

We further tested the transcriptional human-mouse divergence in the dataset that includes nine primates and mouse stimulation (the 10-species system). We repeated the same analysis as done with the rhesus and rat (**Fig 8A**) with the 10-species system (**Fig 8B**). In this system, we compared the mouse data to each of the 9 primates. In each one of the primate-mouse comparisons we again observed the expected trend, of higher FC values in the ‘human constitutive-low’ group in the dsRNA stimulation condition, and in the ‘mouse constitutive-high’ group in the control condition. This trend was also observed when looking at the lower FC values that were in the ‘mouse constitutive-low’ in the dsRNA condition and in the ‘human constitutive-high’ in the control condition (**Fig 8B**). This suggests that the transcriptional differences originally observed between human and mouse in response to dsRNA stimulation are conserved in other primates compared to rodents.

Finally, we asked whether the genes found to be divergent between primates and rodents, also change between different primate species based on the primate phylogeny. For this, we clustered the logFC values of the orthologous genes in each of the 10 species (nine primates and mouse), in the four groups of divergence (**S5 Fig**). We did not find groups of genes that show strong change in behaviour across the primates (for example, that they are only highly upregulated in Great apes and not in other primates). However, we do see that the FC values in response to dsRNA stimulation for each of the four human-mouse divergent groups are clustered in a manner that largely recapitulates the evolutionary relationship between primates (**S5 Fig**). This suggests that while overall the induced versus constitutive expression behavior is consistent across primates, the levels of induction diverge to some extent between primates, and this follows the primate phylogeny. In other words, genes found to be induced following dsRNA stimulation in human but not in mouse

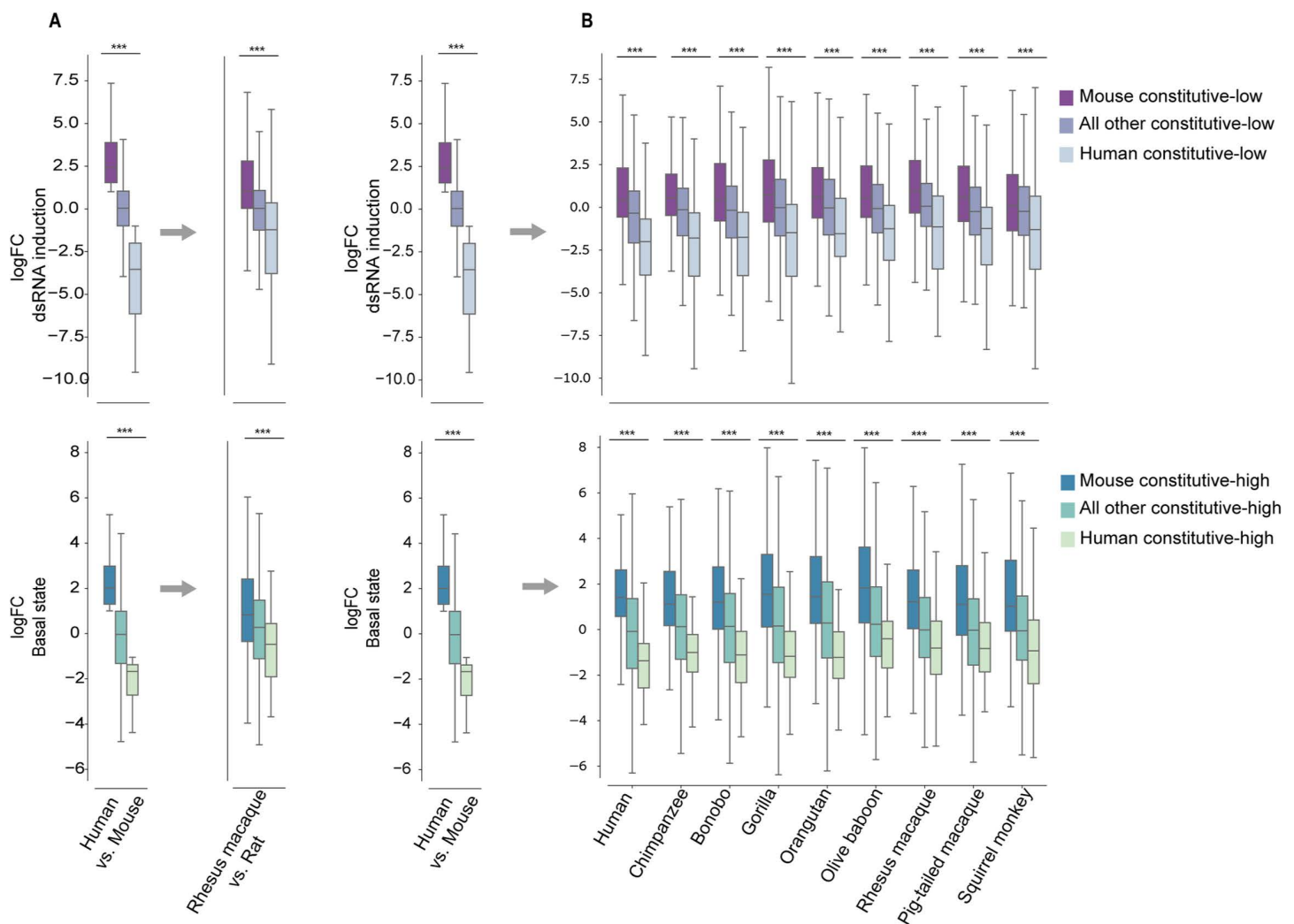


Fig 8. Divergence in dsRNA response across different primate and rodent species. (A) Top: logFC distribution of differential expression analysis in dsRNA-stimulation condition between human and mouse (left) and rhesus macaque and rat (right) from the 4-species system. Separation in both boxplots is according to the groups originally defined from human-mouse divergence: “mouse constitutive-low”, “human constitutive-low”, “All other DE genes” (as defined in Figs 3 and 4). The right panel, which is based on rhesus versus rat data shows that the orthologs of two species diverge in a similar manner to the human and mouse orthologs. Thus, both left and right panels use the same three groups of orthologous genes, but the FC values are based on either human-mouse (left) or macaque-rat (right) DE analysis. **Bottom:** The same as in Top, but with DE performed between human and mouse in basal conditions and showing “human constitutive-high”, “mouse constitutive-high” and “all other DE genes”. **(B)** The same as in A, but comparing the human-mouse divergence in response to dsRNA from the 4-species system to each one of the 9 primates-mouse divergence in response to dsRNA from the 10-species system (the primate species used is shown in the bottom, and its data is always compared to mouse). In both (A) and (B) we observed that the patterns of relative FC between the compared groups recapitulate the patterns from the human-mouse dsRNA comparison. In all analyses, FDR-corrected P-values are shown for one-sided Mann–Whitney test, performed under the hypothesis that the FC distribution of the left group is higher than the right group in each of the analyses.

<https://doi.org/10.1371/journal.pcbi.1013165.g008>

are generally induced also in other primates, and the level of upregulation may be reduced as a function of the primate’s phylogenetic distance from human.

We also used our new approach, to compare the nine primate species response and to obtain groups of constitutively high and constitutively low genes in one primate species versus another, in each pair of primates. Following this, we compared the resulting total numbers of constitutively high and constitutively low genes (S6A and S6B Fig, respectively),

for each pair of primates with respect to all other pairs. We observe that the most closely related pairs of primates – chimpanzee and bonobo, and the two macaque species have the smallest number of constitutively high and constitutively low genes, following expectations that divergence between closely related species would be low. Additionally, we observe that pairs within primate families, i.e., within Old World monkeys and within Great apes, the total numbers of constitutively high and constitutively low genes are by and large lower than those observed between families. This, again, follows the notion that more distant pairs of species would display greater transcriptional divergence. However, this is not always the case, and there are cases that the numbers are lower or higher than expected by phylogeny. These cases are likely influenced by various technical factors, such as differential level of response of cells from different species, intrinsic cell growth, etc. Thus, this analysis shows the utility of the new approach in finding groups of genes that are divergent in response, but also points to the limitations in using some of these cross-species transcriptomics datasets that are also affected by various technical biases.

Human-mouse transcriptional divergence is consistent across different cell types

All previous analyses were performed on dermal fibroblasts that provide a unique system as they are a homogenous and comparative cell system across species, where data of dsRNA stimulation exists for numerous species. We next sought to expand the findings to other cell systems. While other cell systems are more limited in their data comprehensiveness, they can still offer comparisons for specific aspects in our study. We focused on two aspects of our findings: (1) the level of gene expression in non-induced cells and tissues, and (2) the innate immune induction following stimulation in various cells and tissues.

First, since for each species we have genes denoted as ‘constitutive-low’ or ‘constitutive-high’ based on basal expression in fibroblasts, we tested the basal expression levels of these two groups of genes across tissues, using expression data from a large set of mouse and human tissues (taken from the GTEx [39] and the BodyMap [40] datasets). We observed that genes defined as ‘constitutive-high’ based on the fibroblast data are significantly higher than ‘constitutive-low’ genes across tissues. This is true both for most human tissues (35 of 40 tested tissues, based on GTEx) (S7 Fig) and for all mouse tissues (18 mouse tissues, based on BodyMap dataset) (S7 Fig). This suggests that the basal expression of the identified genes in fibroblasts is largely consistent with relative basal expression levels of these genes across tissues.

Secondly, we asked whether transcriptional response patterns observed in the fibroblasts system, are also observed in other cell types. For this we utilized the Interferome database, that includes data from a wide range of IFN stimulation studies of various human and mouse cell types and tissues stimulated in various conditions and timepoints [41]. We note that such a comprehensive data does not exist for dsRNA stimulation. Using the human-mouse IFN analysis (described in Fig 7B), which is based on dermal fibroblast IFN stimulation, we tested the generality of our results, in terms of divergence in response to IFN between human and mouse, in various cells and tissues appearing in the Interferome database. We took 27 human and 10 mouse stimulation experiments and tested for each gene in how many of these experiments it was significantly induced following Type-I IFN treatment. We divided the genes into the four transcriptionally divergent groups previously defined based on our human-mouse fibroblast IFN stimulation dataset. In all four transcriptionally divergent gene sets we observed: (1) a higher fraction of induction in one species than the other, and (2) the induction was higher in the species expected to show an induction based on the fibroblast system (S8 Fig). For example, in the set of genes we originally defined as ‘mouse constitutive-high’ and ‘mouse constitutive-low’, we expected to see that more genes would be induced by IFN in human rather than mouse cells, if the patterns observed in fibroblasts are consistent across other cell types. This is indeed the case, as observed in the higher fraction of IFN stimulation studies that detected these genes as IFN-induced. Additionally, in both ‘human constitutive-high’ and ‘human constitutive-low’ genes, we observed a higher fraction of induction in mouse cell systems, again suggesting a conserved pattern of IFN induction in a species-specific manner across cell types.

Thus, the divergence in induction following IFN stimulation between human and mouse fibroblasts is also observed in other human and mouse cell types. This suggests that species-specific induction is consistent across different cells and tissues, at least to some extent. We note however that the fraction of observed induction across cell types is lower than 30% (as can be observed by the median values of the distributions in [S8 Fig](#)). Out of ~16,000 1-to-1 orthologs assayed in these studies, 7,136 genes are upregulated in at least one study, while 103 are upregulated in at least half of the assays in both human and mouse systems. Thus, while species-specific induction is more conserved than expected, as reflected in the significant similarity to our fibroblast results across the four sets of genes defined in (P-values are always significant), there are still many genes whose induction following immune stimulation varies between different cell types and tissues. Moreover, our analysis suggests that there is a small core of ~100 IFN-induced genes that are strongly upregulated across numerous tissues in both human and mouse, in agreement with previous results regarding conserved ISGs across mammals [\[4\]](#).

Discussion

In this work we aimed to identify cases in which antiviral genes switch between constitutive and induced expression across clades and species. Previous analyses used various methods to identify transcriptional changes between species in homologous tissues, in developmental pathways and in immune responses [\[10,12,22,24,42,43\]](#). These works used various linear models, models utilizing the Ornstein–Uhlenbeck process or direct comparisons of fold change between orthologous genes to analyze trends between groups of genes with varying degree of transcriptional divergence or to find specific cases of divergent genes. In here, we developed a method that particularly focuses on finding orthologous genes upregulated in one species while being constitutive in the other species. The approach further partitions the constitutive expression to relatively “low” and “high” levels with respect to the other species. This partition can potentially denote physiologically different scenarios: lack of significant expression in the first case, or expression in high levels, even before infection, in the latter case. Importantly, some of the observed transcriptional changes may be a result of a drift and further functional studies are needed to determine the functional consequences of such changes.

The expression level of antiviral genes is often increased following immune stimulation or infection, to enable better protection against and restriction of invading pathogens [\[2,3,44\]](#). However, most antiviral genes are also expressed in ‘basal conditions’, without infection or immune stimulation. Basal expression is thought to be important in providing an immediate protection, or readiness, against viruses. Both basal and induced levels of antiviral genes are under strict regulation, since excessive antiviral gene expression can lead to immune pathologies while reduced levels can compromise host defenses [\[45–49\]](#). Evolutionary changes where antiviral genes are strongly induced in one species but are constitutively expressed in another species, may thus impact different host species abilities to inhibit various virus replication.

To study such “evolutionary switches” in antiviral gene expression, we focused on the comparison between primates and rodents – two biomedically important clades that have relatively large amount of comparative data both between and within the clades. We note that our analysis can be done at the species-, rather than the clade-, level. However, comparing between clades allowed us to test whether orthologous genes from different pairs of species display similar patterns (e.g., human vs mouse, and macaque vs rat). We used two comparative *in vitro* systems of dermal fibroblasts including cells from several primates and rodents (4-species and 10-species systems) that were stimulated with dsRNA and include gene expression of both dsRNA-stimulated and control-unstimulated conditions. We first analyzed the whole-transcriptome and the overall response similarity between species, showing that the transcriptional divergence is overall similar to the phylogenetic relationship between these species, suggesting that most genes are conserved in expression and that divergence is accumulated over evolutionary time. This gradual divergence in the majority of gene expression across species, enabled us to focus on specific genes that significantly change in transcriptional response between primates and rodents.

Using a new approach to identify genes that differ in expression between primates and rodents, and first focusing on the comparison between human and mouse from the 4-species stimulation system, we generated four groups of

human-mouse transcriptionally divergent genes. In two of those sets the genes are induced only in mouse, whereas in the other two, the genes are only induced in human. The difference between the two sets depends on whether the group of constitutive genes remains transcriptionally low or stays at relatively high expression levels in both control and dsRNA-stimulation conditions (see Fig 1). Our approach can be extended to studying other clades or other gene expression programs, and can be useful to detect such inter-species switches in transcriptional behavior in various cellular contexts. This can be particularly interesting for studying clades that are known to display differences in their resistance to pathogens, such as bats and reptiles, and can be done at various time scales when such comparative data becomes available.

We next characterized the functional characteristics of these four sets of genes. We found that each of the four sets is enriched with genes belonging to specific pathways or whose protein products are part of the same protein complex. However, our results did not point to a major pathway whose expression was completely altered between primates and rodents. This suggests that the divergence in innate immune response between these two closely related clades has occurred in specific regulatory changes of one or few genes, rather than involving a shift of an entire pathway. This may mirror findings of positive selection in antiviral genes, that showed that evolutionary arms races between host and their viruses are often associated with changes in specific residues of a few antiviral genes related to the restriction of the studied virus [50,51].

Our evolutionary analysis of coding sequences of the transcriptionally divergent antiviral genes, showed that some of these groups tend to more conserved in sequence than the whole set of antiviral genes whereas others tend to be less conserved. These patterns may be related to the expression levels of these genes across tissues in basal conditions. In addition, they may reflect the different functions of the genes in these groups. This is in agreement with previous findings that showed that different classes of antiviral genes have different evolutionary rates and that only certain classes of antiviral genes are enriched with genes with signatures of positive selection [11,52].

To test the generality of our results of the human-mouse dsRNA stimulation, we first compared the response to dsRNA and IFN stimulation. We found that the transcriptional divergence between human and mouse response to dsRNA was similar to the one in the IFN stimulation. This suggests a similarity in the divergence between the primary and secondary innate immune responses. The generality of our findings was also reflected to some extent when we compared the response to IFN of human and mouse fibroblasts to other human and mouse cell types, as observed in the analysis using the Interferome database. However, as expected, many genes vary in their induction between different cell types, reflecting cell-type specific response to infection. Looking at the conservation of response across different primates, we found that the majority of genes induced in human but not in mouse are also induced in other primates. We also found that the response within primates gradually diverge with evolutionary distance. We note that detecting subset of genes that are differentially regulated in a specific group of primates is difficult using the available data given technical differences in the responsiveness between species and the use of the late time point to collect the data.

Taken together, our results point to a group of genes that diverge between primates and rodents in basal and induced expression in response to viral stimulation. The induction of these genes is largely recapitulated across the primate and rodent clade, and is shown to be in agreement with chromatin accessible regions in nearby promoter regions and, to some extent, in different cell types and tissues. While our work suggests that some of the genes that switch in transcriptional behavior in fibroblasts between primates and rodents also seem to behave similarly in other cell types, further comparative studies would be needed to fully establish this.

Together with previous studies focusing on identification of conserved immune-stimulated genes in mammals or primates [4,19] and on different aspects of transcriptional divergence in immune response to infection [11–13,53], our study contributes to understanding of the evolutionary landscape of the innate immune response and its regulation. The transcriptionally divergent genes found in this work can contribute to species-specific adaptation to viral infection and provide directions to understanding the mechanisms behind this adaptation.

Methods

Dataset availability and data assembly

All data used in this work is publicly available, as detailed below. For comparing innate immune responses across species, we searched for comparative cell systems where similar cells grown in similar conditions were treated with the same immune stimulant to elicit an antiviral response across species. The two systems we used were both dermal fibroblasts, grown *in vitro* and stimulated with dsRNA (poly I:C) from primates and rodents: the 4-species system [11] (**Arrayexpress accession - E-MTAB-5919**), and the 10-species system [19] (**SRA accession: SRP120495**). The first system included 4 species - 2 primates (human and rhesus macaque) and 2 rodents (mouse and rat), stimulated with dsRNA for 4h, while the second included 9 primates (5 Great apes – human, chimpanzee, bonobo, gorilla, orangutan; 3 Old world monkeys – rhesus and pig-tailed macaques and baboon; and one New world monkey - squirrel monkey) and mouse, stimulated with dsRNA for 24h. Each of these systems included stimulation of cells from several individuals per species (only females in the 4-species system and both males and females in the 10-species system). Samples from each individual included treated (dsRNA) and mock treated control, profiled using RNA-seq.

In the case of the 4-species system, we also used IFNB stimulation and control for the same species and individuals, stimulated and processed in parallel to the dsRNA transfection. The 4-species system also included ChIP-seq data of histone marks, from which we used the H3K27ac data that marks active chromatin regions, to study active promoters (**ArrayExpress accession: E-MTAB-5918**).

We chose not to use several other cross-species fibroblast system studies [4,12], since they lack more than a single primates and/or rodent species, required for our analysis.

Major scripts with relevant details and input files can be found at: https://github.com/HagaiLab/transcriptional_evolution_of_antiviral_genes

Read mapping and gene expression quantification

Reads were mapped and gene expression was quantified using Salmon (version 0.13.1) [54] with the following commands:

- For paired end libraries: “salmon quant -i {index_file_directory} -l A -p 8 --validateMappings --gcBias --seqBias --numBootstraps 100 -g {transcript_to_gene_file} -1 {first_read} -2 {second_read} -o {output_directory}”
- For single end libraries: “salmon quant -i {index_file_directory} -l A -p 8 --validateMappings --gcBias --seqBias --numBootstraps 100 -g {transcript_to_gene_file} -r {read} -o {output_directory}”

Each sample was mapped to its respective species' annotated transcriptome (coding genes only, downloaded from ENSEMBL version 99 [55]: Human - GRCh38, Chimpanzee - Pan_tro_3.0, Bonobo - panpan1.1, Gorilla - gorGor4, Orangutan - PPYG2, Olive baboon - Panu_3.0, Rhesus macaque - MMUL_1, Pig-tailed macaque - Mnem_1.0, Bolivian squirrel monkey - SaiBol1.0, Mouse - GRCm38, Rat - Rnor_6.0). We removed annotated secondary haplotypes of human genes by removing genes with 'CHR_HSCHR'.

Quantifying differential gene expression in response to dsRNA and to IFNB

To quantify differential gene expression between treatment and control for each species and for each treatment separately, we used edgeR (version 3.32.1) [56], using R version 4.0.3, using the rounded estimated counts from Salmon. Differential expression analysis was performed using the edgeR exact test, and P-values were adjusted for multiple testing by estimating the false discovery rate (FDR), similarly to our previous work [11,12].

Principal component analysis (PCA)

For PCA analysis we used the `prcomp()` function in R version 4.0.3, and as input the rounded estimated counts from Salmon. The analysis included only 1-to-1 orthologous genes in all of the species in the study (12,811 and 11,350 orthologs, for the 4-species and 10-species studies, respectively). Before performing the analysis, we transformed the data with \log_{10} transformation on the counts values, and scaled it using the `scale()` function in R (`center=True`, `scale=False`).

Hierarchical clustering of gene expression and fold change values

We used `clustermap()` function from `seaborn` package in Python version 3.9.7. Clustering parameters: `method='average'`, `metric='euclidean'`.

Identification of genes dsRNA-upregulated in one species and constitutively expressed in another species

We performed the following procedure to identify genes that are upregulated in response to dsRNA in one species, but are constitutively expressed (do not significantly change in expression before and after treatment) in the other species.:

1. Differential expression analysis between the human-mouse orthologs within the same condition (dsRNA-stimulation conditions between human and mouse samples, and separately, the control-unstimulated conditions between human and mouse samples). DE analysis was performed using `edgeR` (version 3.32.1), with the rounded estimated counts obtained from Salmon. This was done only for genes that had a significant level of expression in at least 3 replicates ($\text{TPM} > 0$, transcripts per million).

We note that we used this threshold for genes to be considered as expressed, to ensure that if these genes are lowly expressed and are not detected in the NGS, their expression is below levels that can be considered as functional [57].

We note that a choice of a cutoff to remove genes that are lowly expressed and that may be technically challenging to detect is a common approach used in comparative bulk and single-cell transcriptomics analyses [11,12,17], however any choice of such a cutoff may still result in genes that are wrongly assigned as false negatives or false positives.

Differential expression analysis was performed using the `edgeR` exact test, and P-values were adjusted for multiple testing by estimating the false discovery rate (FDR).

2. Identification of differentially expressed genes between human and mouse in either dsRNA-stimulation, control or both conditions: These DE genes are defined as those with an FDR-corrected $P\text{-value} < 0.001$ and with a $|\text{Fold Change}| > 1$ between the human and mouse orthologs in the tested condition (thus, we obtained genes that are higher in expression in this condition in either human or mouse, depending on the sign of the FC).
3. For each of the groups of DE genes identified (with $\text{FC} > 1$ or $\text{FC} < -1$ and FDR-corrected $P\text{-value} < 0.001$, and in each of the two conditions), we removed genes that displayed the same behavior in the other condition. For example, genes that were significantly higher in mouse than in human cells ($\text{FC} > 1$ and FDR-corrected $P\text{-value} < 0.001$) in both the dsRNA-stimulation as well as in control-unstimulated conditions were removed (since they likely represent a set of genes that is higher in mouse versus human regardless of the innate immune response).
4. We further removed any gene that is not DE in response to treatment in at least one species. That is, if a gene was not found to have a $q\text{-value} < 0.01$ and $\text{FC} > 0$, in either human and/or mouse in response to dsRNA, it was excluded from the analysis (in this case, we used DE values obtained from a "classical DE analysis" where samples from the same species from two different conditions were compared to find DE genes in response to dsRNA-treatment in either human or mouse).

This resulted in 4 groups of species-specific responding genes (detailed lists appear in [S1 Table](#)):

1. **'Mouse constitutive-low'**: Generated from the group of genes that were significantly higher in human than in mouse in the dsRNA-stimulation conditions ($FC < -1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in dsRNA-stimulation conditions) (Stage 2). From this group we excluded genes that were also significantly higher in human than in mouse in the control-unstimulated conditions ($FC < -1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the control-unstimulated condition) (Stage 3). We then removed genes that were not DE in response to treatment in at least one species (Stage 4). This resulted in 166 genes that are induced in dsRNA-stimulation only in human but in mouse remain constitutively low in both unstimulated and stimulated conditions.
2. **'Human constitutive-low'**- Generated from the group of genes that were significantly higher in mouse than in human in the dsRNA-stimulation conditions ($FC > 1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the dsRNA stimulation conditions) (Stage 2). From this group we excluded genes that were also significantly higher in mouse than in human in the control-unstimulated conditions ($FC > 1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the control-unstimulated condition) (Stage 3). We then removed genes that were not DE in response to treatment in at least one species (Stage 4). This resulted in 224 genes that are induced in stimulation in mouse but in human remain constitutively low in both unstimulated and stimulated conditions.
3. **'Mouse constitutive-high'**- Generated from the group of genes that were significantly higher in mouse than in human in the control-unstimulated conditions ($FC > 1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the control-unstimulated conditions) (Stage 2). From this group we excluded genes that were also significantly higher in mouse than in human in the dsRNA-unstimulated conditions ($FC > 1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the dsRNA-unstimulated conditions) (Stage 3). We then removed genes that were not DE in response to treatment in at least one species (Stage 4). This resulted in 104 genes that are induced in stimulation in human but in mouse remains constitutively high in both unstimulated and stimulated conditions.
4. **'Human constitutive-high'**- Generated from the group of genes that were significantly higher in human than in mouse in the control-unstimulated conditions ($FC < -1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the control-unstimulated condition) (Stage 2). From that group we excluded genes that were also significantly higher in human than in mouse in the dsRNA-unstimulated conditions ($FC < -1$ and FDR-corrected $P\text{-value} < 0.001$ in the DE analysis in the dsRNA-unstimulated conditions) (Stage 3). We then removed genes that were not DE in response to treatment in at least one species (Stage 4). This resulted in 84 genes that are induced in stimulation in mouse but in human remain constitutively high in both unstimulated and stimulated conditions.

Similar procedures, as described above, were performed to detect constitutive-high and constitutive-low genes when using different stimulation data and different species, including rat versus macaque and between 9 primate species.

We implemented a linear model with an extra interaction term (Species:Condition) that enable finding the difference in expression based on both expression in the two species and in the two conditions. We tested the obtained DE values (Q-value and Fold Change) of each of the four groups we defined above, to test whether these groups are enriched with genes whose expression levels differ significantly in the interaction term.

We used the python 'upset' package to analyze the intersection between the various sets and subsets used in this work.

Analysis of basal gene expression across tissues in human and mouse

To characterize relevant gene expression across human and mouse tissues, we extracted the data from Genotype-Tissue Expression (GTEx) project [39], version 8, and from BodyMap dataset [40], respectively. We processed the data, to achieve a more comparable dataset between human and mouse, as we previously did [58]. Briefly, from the human tissues data we excluded all the pseudo-autosomal expression records and the tissues of 'Cells EBV-transformed

lymphocytes' and 'skin sun exposed' and merged various brain tissues by computing the mean for each gene across them, to compare them with the available mouse brain from the BodyMap data. As for the mouse tissue data, which included 17 tissues, we merged the tissues by computing the mean for each gene in both male and female (if both existed). Using this human-mouse cross-tissue comparative data, we tested whether genes found in the *in vitro* cultured fibroblasts to be more highly expressed, are also more highly expressed in human and/or mouse tissues. To test this, we performed a one-sided Mann-Whitney test. This was done separately for human and mouse.

Functional enrichment analysis

To study the functional enrichment of genes against the background of the entire human proteome we used g:Profiler program [27]. NFkB targets were obtained from (<https://esbl.nhlbi.nih.gov/Signaling-Pathways/NF-kB-Targets/>, 960 genes overall) and their fold change values were compared between human and mouse.

Functional analysis with the CORUM database

To study whether genes' protein products belong to the same protein complex we used the CORUM database that is a collection of experimentally verified mammalian protein complexes [28].

Gene age analysis

Gene age estimations were obtained from ProteinHistorian [35]. Specifically for our analysis, we used the PPODv4-OrthoMCL and wagner1.0 parameters. For each group of genes (all DE genes in either human or mouse, and the four divergent groups of genes between human and mouse in response to dsRNA stimulation) we plotted the distribution of gene age values. A one-sided Mann-Whitney test was conducted between each one of the four groups against the DE genes to the hypothesis where the age of the DE genes is higher than the one of diverging gene groups.

Coding sequence evolutionary analysis

Human-mouse ortholog's percentage of identity was calculated as the mean of both the percentage of identity of the human ortholog to the mouse ortholog, and the percentage of identity of the mouse ortholog to the human ortholog. Both values were obtained from ENSEMBL [55,59]. For each group of genes (all DE genes in either human or mouse and the four divergent groups of genes between human and mouse in response to dsRNA stimulation) we plotted the distribution percentage of identity. One-sided Mann-Whitney test was conducted between each one of the four groups against the DE genes to the hypothesis where the percentage of the DE genes is higher than the one of diverging gene groups.

dN/dS ratio (non-synonymous to synonymous codon substitutions) for human and mouse orthologs was also obtained from ENSEMBL. For each group of genes (all DE genes in either human or mouse and the four divergent groups of genes between human and mouse in response to dsRNA stimulation) we plotted the distribution of dN/dS ratio. A one-sided Mann-Whitney test was conducted between each one of the four groups against the DE genes to the hypothesis where the ratio of the DE genes is lower than the one of diverging gene groups.

Rate of gene gain and loss analysis

The significance at which a gene's family has experienced a higher rate of gene gain and loss in the course of vertebrate evolution in comparison with other gene families ("P-value of a fast rate of gain and loss") was obtained from ENSEMBL [59]. This statistic was calculated using the CAFE method [60], which estimates the global birth and death rate of gene families and identifies gene families with accelerated rates of gain and loss. For each group of genes (all DE genes in either human or mouse and the four divergent groups of genes between human and mouse in response to dsRNA stimulation) we plotted the distribution of their gene gain and loss P-values. A one-sided Mann-Whitney test was conducted

between each one of the four groups against the DE genes to the hypothesis where the p-value of the DE genes is higher than the one of diverging gene groups.

Interferome analysis

From the Interferome database [41], we downloaded the data of the differential expression analysis for relevant sets of human and mouse cell types - a total of 27 experimental sets in humans and 10 in mouse (the datasets are described below). For each human and mouse 1-to-1 ortholog, we calculated the fraction of the datasets of human, and those of mouse where the ortholog is upregulated ($FC > 1$). We excluded from this analysis genes that were completely missing from the database (that did not appear in any of the datasets of both human and mouse).

The datasets used are as follows:

Mouse:

- one dataset of wild type murine embryo fibroblasts (MEFs) treated with 2500U IFN beta for 5h
- 5 datasets of NIH-3T3 cells treated with 100U IFN alpha, where one of them was treated for 0.5h, two replicates for 1h and the remaining two replicates for 3h
- 4 datasets of lung-whole tissue cells of BALB/c mice carrying a defective allele of the Mx1 resistance gene, treated with 10,000U of recombinant human IFN alpha A/D, one of them for 12h, one for 72h and the other two for 24h.

Human:

- two datasets of Human A549 cell line treated with IFN alpha, where one of them was treated for 6h and the other for 24h.
- one dataset of Huh7 hepatoma cells treated with IFN alpha (50 IU/ml) for 72h.
- 10 datasets of Human bronchial epithelial cells (HBECs) treated with 15-minute pulse of 1000U/ml IFN beta that each one harvested in different time points (0.25, 0.5, 1, 1.5, 2, 4, 6, 8, 12, 18 hours).
- two replicates' datasets of Blood monocyte-derived macrophages treated with 25 ng/ml IFN alpha for 3h.
- one dataset of Human fibroblast cells isolated from umbilical veins treated with 1000U IFN alpha for 5h
- two datasets of Human endothelial cells isolated from umbilical veins, one of them treated with 1000U IFN beta for 5h, and the other with 1000U IFN alpha for 5h
- 4 datasets of Human BE(2)-C neuronal cells stimulated with IFN alpha, where two of them are replicates for 12h and the other two are replicates for 6h
- one dataset of PDOVCA#1- side population of ovarian cancer cells treated with IFN alpha2b for 5h
- one dataset of PDOVCA#1- Hoechst high cells (non-Side Population) treated with IFN alpha2b for 5h
- one dataset of monocyte derived dendritic cells treated with IFN alpha for 24h
- one dataset of monocyte derived macrophages treated with IFN alpha for 72h
- one dataset of Human hepatocyte cells treated with IFN alpha for 6h

Promoter and enhancer definition and ChIP-seq analysis of promoters of upregulated genes

Using gene annotations from ENSEMBL version 99 [59], we selected for each gene the representative transcript as the transcript with the best TSL (lowest) and if there was more than one such transcripts, we selected the longest transcript.

After finding a representative transcript per gene, we defined the promoter region as the region 2,000bp upstream of the TSS and 500bp downstream of it, following our previous work where we investigate promoter characteristics [58].

In the ChIP-seq analysis, we aimed to quantify the extent of agreement between the RNA and chromatin levels: between the levels of expression of genes upregulated in dsRNA-response in a species-specific manner (the four gene groups defined above), and chromatin active regions in their promoters before and after stimulation in each of the two species, as probed by ChIP-seq of H3K27ac. For example, we asked whether genes that are constitutively expressed in mouse cells, in both control and treatment and were only upregulated in response to dsRNA in human cells, will have an active H3K27 mark in both conditions in mouse but in human cells only after dsRNA-treatment.

To define promoter region as “active”, we overlapped called peaks from the ChIP-seq data (peak processing, including calling and merging, was done as described previously [11]), with promoter regions as described above, using the bed-tools [61] `com mand`:

```
intersectBed -a {processed peaks bed file} -b {promoters bed file} -wo
```

To define enhancer region as “active”, we used a similar approach to the above-described promoter analysis. First, we filtered overlapping peaks of H3K27ac with H3K4me3. We then used the remaining peaks, to find active enhancers, by looking at a region 1M up- and downstream of the TSS, using the `intersectBed` command.

For each gene category (e.g., ‘mouse constitutive-high’) we computed the fraction of genes that the patterns of promoter (or enhancer) activity agrees with gene expression across the two species in both conditions (e.g., there is a called peak in the promoters of human and mouse dsRNA condition and in the control condition of mouse, but not in the control condition of human in ‘mouse constitutive-high’ genes). To compute the enrichment of this fraction, we compared it with the fraction of all DE genes in mouse or human using Fisher’s exact test.

Statistical analysis

Statistical analyses (Mann-Whitney test, Fisher’s exact test, Spearman’s rank-order correlation and FDR correction based on Benjamini-Hochberg procedure [62]) were performed using the SciPy package in Python (version 3.9.7). Data in boxplots represent the median, first quartile and third quartile with lines extending to the furthest value within 1.5 of the interquartile range. Plots were created using matplotlib and seaborn packages in Python, except for the PCA plots that were created using ggplot package in R.

Supporting information

S1 Fig. Overlaps between various groups of genes, relevant for this study. (A) An enlarged version of Fig 4C - An upset plot including the four groups of transcriptionally divergent genes in response to dsRNA between species, genes that respond to dsRNA stimulation in each species, and differentially expressed genes between human and mouse in non-stimulated conditions, and separately, in stimulated conditions. The total number of genes is shown to the left of each row, and the intersection between various groups is shown on top of each of the bars. Note that these sets show only the mutually exclusive genes, for example – the group “Genes that respond to stimulation in human” has a total of 529 genes, but most of them are part of intersections with other groups, and thus only 91 of them are shown in the fourth column. (B) Venn diagrams showing how the two groups of constitutive-high genes are part of the group of differential expression in non-stimulated mouse versus human. (C) Venn diagrams showing how the two groups of constitutive-high genes are part of the group of differential expression in non-stimulated mouse versus human. B-C demonstrate four cases of overlaps from the upset plot in A. (PNG)

S2 Fig. Testing differential values of the four groups of transcriptionally divergent genes using a linear model. We implemented a linear model with an interaction model of species:condition. We then tested the DE values of this

interaction term, of the four groups detected using the new approach described in the main text. This provides a measure to test the agreement between the new approach developed in this manuscript and a different approach to find genes that diverge in transcriptional response between species. (A) log(Fold Change) values for each of the four groups and all DE genes from the interaction term (Species:Conditions). FDR-corrected P-values are shown for one-sided Mann–Whitney test that was performed under the hypothesis that the distribution of values in human is higher and in mouse is lower than the overall DE set. (B) Q-value distribution of genes from all four groups, all DE genes, and all one-to-one human-mouse orthologs from the interaction term. FDR-corrected P-values are shown for one-sided Mann–Whitney tests.

(PNG)

S3 Fig. A comparison between fold change in response to dsRNA in human and mouse ortholog genes that are known NFkB targets. The sets were compared using a one-sided Mann-Whitney test.

(PNG)

S4 Fig. Sequence similarity between promoters of human-mouse orthologs. A Sequence similarity scores between promoter regions of human and mouse orthologs (the four groups defined in the manuscript and all orthologs), as well as human paralogs that have duplicated after the split between human and chimpanzee (see Fraimovitch and Hagai, BMC Biology (2023)1). This Fig demonstrates that the sequence similarity between human, and mouse orthologs in all four groups tested and in general, is very low, much lower than recent human duplicates where significant promotor sequence similarity remains.

(PNG)

S5 Fig. Human-mouse divergence in response to dsRNA stimulation is similar to the divergence observed in response to IFN. Averaged TPM levels of human and mouse genes in control and in stimulation conditions in both dsRNA and IFN systems for genes in (A) ‘human constitutive-low’, (B) mouse constitutive-low’, (C) ‘human constitutive-high’ and (D) ‘mouse constitutive-high’. FDR-corrected one- or two-sided Mann–Whitney tests were performed according to the expected pattern of expression matching to the specific group and statistical significance is shown. The observed results in A-D suggest an overall similarity in the transcriptional behavior of the human-mouse orthologs, originally identified based on their divergence in response to dsRNA, to behave similarly also in response to IFN.

(PNG)

S6 Fig. Hierarchical clustering across the primate clade in the four groups of human-mouse transcriptionally divergent genes in response to dsRNA. Hierarchically clustered heatmaps on logFC values in response to dsRNA for orthologous genes in cells from 9 primates and mouse, for each one of the 4 divergent gene groups, defined in Fig 3: (A) ‘human constitutive-low’, (B) mouse constitutive-low’, (C) ‘human constitutive-high’ and (D) ‘mouse constitutive-high’. The logFC values are from differential expression analysis between control and stimulation with dsRNA conditions for each of the species from the 10-species system (in each case, the DE values are within the same species).

(PNG)

S7 Fig. Comparison of constitutive-high and constitutive-low gene numbers between pairs of primate species. Numbers of (A) constitutive-high and (B) constitutive-low genes in each of the 9 primate pairs. Primate transcriptomics data is based on Gaska et al. [19], and the detection of the genes follows the same procedure described in the manuscript for human versus mouse data. Primates are ordered by phylogenetic distance to human.

(PNG)

S8 Fig. Basal expression levels of constitutively low and constitutively high genes in human and mouse across tissues. Distributions of TPM values are shown in boxplots, of constitutive-low and high genes as defined (A) for human,

in 40 human tissues from the GTEx dataset [3], and (B) for mouse, in 17 mouse tissues from the BodyMap dataset [4]. FDR-corrected P-values are shown for one-sided Mann–Whitney test, performed between the basal TPM values of constitutive-high and low genes in human or mouse, for each tissue within the species. In the majority of cases (tissues/species), the constitutive-low genes are significantly lower in expression than the constitutive-high genes.

(PNG)

S9 Fig. Fraction of datasets from Interferome in which genes are upregulated in response to Interferon from the four groups of human-mouse divergent genes. Each of the 4 divergent groups is shown, partitioned by Interferome [5] datasets of human and mouse transcriptional response data (Each dataset is based on transcriptional response to IFN in a different cell or tissue in human or mouse). FDR-corrected P-values are shown for one-sided Mann–Whitney test that was performed under the hypothesis that the fraction in human is higher (in mouse constitutive high or mouse constitutive low groups) or lower (in human constitutive high or human constitutive low groups) than in mouse. We observe that two distributions (human versus mouse) are always significantly different and that they follow our expectations based on the stimulation data from human and mouse fibroblasts. For example, in the group of genes that were identified as “human constitutive low” in human-mouse fibroblast data, we observe that the distribution of human is lower than that of mouse orthologous genes, as expected given the fact that this group is induced in mouse fibroblasts, and not in human fibroblasts, following IFN stimulation.

(PNG)

S1 Table. A - list of human constitutive-low genes. B - list of mouse constitutive-low genes. C - list of mouse constitutive-high genes. D - list of human constitutive-high genes.

(XLSX)

S2 Table. A - enrichment terms of human constitutive-low genes. B - enrichment terms of mouse constitutive-low genes. C - enrichment terms of mouse constitutive-high genes. D - enrichment terms of human constitutive-high genes.

(XLSX)

Acknowledgments

We would like to thank Michael Lässig, Irit Gat-Viks, Manuela Sironi and Peter Sudmant for helpful discussions during the project and on the manuscript, and Maya Lischinski for assistance with creating the initial dataset.

Author contributions

Conceptualization: Lilach Schneor, Tzachi Hagai.

Data curation: Lilach Schneor, Dafna Tussia-Cohen, Sivan Friedman.

Formal analysis: Lilach Schneor, Dafna Tussia-Cohen, Evgeny Fraimovitch, Sivan Friedman.

Funding acquisition: Tzachi Hagai.

Investigation: Lilach Schneor, Dafna Tussia-Cohen, Evgeny Fraimovitch, Sivan Friedman.

Methodology: Lilach Schneor, Dafna Tussia-Cohen, Tzachi Hagai.

Supervision: Tzachi Hagai.

Validation: Tzachi Hagai.

Visualization: Lilach Schneor, Dafna Tussia-Cohen, Sivan Friedman.

Writing – original draft: Lilach Schneor, Sivan Friedman, Tzachi Hagai.

Writing – review & editing: Dafna Tussia-Cohen, Sivan Friedman, Tzachi Hagai.

References

1. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. *Int Rev Immunol*. 2011;30(1):16–34. <https://doi.org/10.3109/08830185.2010.529976> PMID: [21235323](https://pubmed.ncbi.nlm.nih.gov/21235323/)
2. Iwasaki A. A virological view of innate immune recognition. *Annu Rev Microbiol*. 2012;66:177–96. <https://doi.org/10.1146/annurev-micro-092611-150203> PMID: [22994491](https://pubmed.ncbi.nlm.nih.gov/22994491/)
3. Porritt RA, Hertzog PJ. Dynamic control of type I IFN signalling by an integrated network of negative regulators. *Trends Immunol*. 2015;36(3):150–60. <https://doi.org/10.1016/j.it.2015.02.002> PMID: [25725583](https://pubmed.ncbi.nlm.nih.gov/25725583/)
4. Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol*. 2017;15(12):e2004086. <https://doi.org/10.1371/journal.pbio.2004086> PMID: [29253856](https://pubmed.ncbi.nlm.nih.gov/29253856/)
5. 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*. 2011;472(7344):481–5. <https://doi.org/10.1038/nature09907> PMID: [21478870](https://pubmed.ncbi.nlm.nih.gov/21478870/)
6. Tenthorey JL, Emerman M, Malik HS. Evolutionary Landscapes of Host-Virus Arms Races. *Annu Rev Immunol*. 2022;40:271–94. <https://doi.org/10.1146/annurev-immunol-072621-084422> PMID: [35080919](https://pubmed.ncbi.nlm.nih.gov/35080919/)
7. Tsu BV, Fay EJ, Nguyen KT, Corley MR, Hosuru B, Dominguez VA, et al. Running With Scissors: Evolutionary Conflicts Between Viral Proteases and the Host Immune System. *Front Immunol*. 2021;12:769543. <https://doi.org/10.3389/fimmu.2021.769543> PMID: [34790204](https://pubmed.ncbi.nlm.nih.gov/34790204/)
8. Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. *Elife*. 2016;5:e12469. <https://doi.org/10.7554/eLife.12469> PMID: [27187613](https://pubmed.ncbi.nlm.nih.gov/27187613/)
9. Quintana-Murci L. Human Immunology through the Lens of Evolutionary Genetics. *Cell*. 2019;177(1):184–99. <https://doi.org/10.1016/j.cell.2019.02.033> PMID: [30901539](https://pubmed.ncbi.nlm.nih.gov/30901539/)
10. Schroder K, Irvine KM, Taylor MS, Bokil NJ, Le Cao K-A, Masterman K-A, et al. Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. *Proc Natl Acad Sci U S A*. 2012;109(16):E944–53. <https://doi.org/10.1073/pnas.1110156109> PMID: [22451944](https://pubmed.ncbi.nlm.nih.gov/22451944/)
11. Hagai T, Chen X, Miragaia RJ, Rostom R, Gomes T, Kunowska N, et al. Gene expression variability across cells and species shapes innate immunity. *Nature*. 2018;563(7730):197–202. <https://doi.org/10.1038/s41586-018-0657-2> PMID: [30356220](https://pubmed.ncbi.nlm.nih.gov/30356220/)
12. Schneur L, Kaltenbach S, Friedman S, Tussia-Cohen D, Nissan Y, Shuler G, et al. Comparison of antiviral responses in two bat species reveals conserved and divergent innate immune pathways. *iScience*. 2023;26(8):107435. <https://doi.org/10.1016/j.isci.2023.107435> PMID: [37575178](https://pubmed.ncbi.nlm.nih.gov/37575178/)
13. Hawash MBF, Sanz-Remón J, Grenier J-C, Kohn J, Yotova V, Johnson Z, et al. Primate innate immune responses to bacterial and viral pathogens reveals an evolutionary trade-off between strength and specificity. *Proc Natl Acad Sci U S A*. 2021;118(13):e2015855118. <https://doi.org/10.1073/pnas.2015855118> PMID: [33771921](https://pubmed.ncbi.nlm.nih.gov/33771921/)
14. 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- α in bats. *Proc Natl Acad Sci U S A*. 2016;113(10):2696–701. <https://doi.org/10.1073/pnas.1518240113> PMID: [26903655](https://pubmed.ncbi.nlm.nih.gov/26903655/)
15. Zhou P, Cowled C, Mansell A, Monaghan P, Green D, Wu L, et al. IRF7 in the Australian black flying fox, *Pteropus alecto*: evidence for a unique expression pattern and functional conservation. *PLoS One*. 2014;9(8):e103875. <https://doi.org/10.1371/journal.pone.0103875> PMID: [25100081](https://pubmed.ncbi.nlm.nih.gov/25100081/)
16. De La Cruz-Rivera PC, Kanchwala M, Liang H, Kumar A, Wang L-F, Xing C, et al. The IFN Response in Bats Displays Distinctive IFN-Stimulated Gene Expression Kinetics with Atypical RNASEL Induction. *J Immunol*. 2018;200(1):209–17. <https://doi.org/10.4049/jimmunol.1701214> PMID: [29180486](https://pubmed.ncbi.nlm.nih.gov/29180486/)
17. Levinger R, Tussia-Cohen D, Friedman S, Lender Y, Nissan Y, Fraimovitch E, et al. Single-cell and Spatial Transcriptomics Illuminate Bat Immunity and Barrier Tissue Evolution. *Mol Biol Evol*. 2025;42(2):msaf017. <https://doi.org/10.1093/molbev/msaf017> PMID: [39836373](https://pubmed.ncbi.nlm.nih.gov/39836373/)
18. Komal A, Noreen M, El-Kott AF. TLR3 agonists: RGC100, ARNAX, and poly-IC: a comparative review. *Immunol Res*. 2021;69(4):312–22. <https://doi.org/10.1007/s12026-021-09203-6> PMID: [34145551](https://pubmed.ncbi.nlm.nih.gov/34145551/)
19. Gaska JM, Parsons L, Balev M, Cirincione A, Wang W, Schwartz RE, et al. Conservation of cell-intrinsic immune responses in diverse nonhuman primate species. *Life Sci Alliance*. 2019;2(5):e201900495. <https://doi.org/10.26508/lsa.201900495> PMID: [31649152](https://pubmed.ncbi.nlm.nih.gov/31649152/)
20. Khaitovich P, Enard W, Lachmann M, Pääbo S. Evolution of primate gene expression. *Nat Rev Genet*. 2006;7(9):693–702. <https://doi.org/10.1038/nrg1940> PMID: [16921347](https://pubmed.ncbi.nlm.nih.gov/16921347/)
21. Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, et al. The molecular evolution of spermatogenesis across mammals. *Nature*. 2023;613(7943):308–16. <https://doi.org/10.1038/s41586-022-05547-7> PMID: [36544022](https://pubmed.ncbi.nlm.nih.gov/36544022/)
22. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental transition and the evolution of animal body plans. *Nature*. 2016;531(7596):637–41. <https://doi.org/10.1038/nature16994> PMID: [26886793](https://pubmed.ncbi.nlm.nih.gov/26886793/)
23. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011;478(7369):343–8. <https://doi.org/10.1038/nature10532> PMID: [22012392](https://pubmed.ncbi.nlm.nih.gov/22012392/)
24. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene expression divergence recapitulates the developmental hour-glass model. *Nature*. 2010;468(7325):811–4. <https://doi.org/10.1038/nature09634> PMID: [21150996](https://pubmed.ncbi.nlm.nih.gov/21150996/)

25. Cooper N, Thomas GH, Venditti C, Meade A, Freckleton RP. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol J Linn Soc Lond*. 2016;118(1):64–77. <https://doi.org/10.1111/bij.12701> PMID: [27478249](https://pubmed.ncbi.nlm.nih.gov/27478249/)
26. Grabowski M, Pienaar J, Voje KL, Andersson S, Fuentes-González J, Kopperud BT, et al. A Cautionary Note on “A Cautionary Note on the Use of Ornstein Uhlenbeck Models in Macroevolutionary Studies”. *Syst Biol*. 2023;72(4):955–63. <https://doi.org/10.1093/sysbio/syad012> PMID: [37229537](https://pubmed.ncbi.nlm.nih.gov/37229537/)
27. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–8. <https://doi.org/10.1093/nar/gkz369> PMID: [31066453](https://pubmed.ncbi.nlm.nih.gov/31066453/)
28. Giurgiu M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, et al. CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res*. 2019;47(D1):D559–63. <https://doi.org/10.1093/nar/gky973> PMID: [30357367](https://pubmed.ncbi.nlm.nih.gov/30357367/)
29. Pijnappel WP, Kolkman A, Baltissen MP, Heck AJ, Timmers HM. Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. *Proteome Sci*. 2009;7:46. <https://doi.org/10.1186/1477-5956-7-46> PMID: [20034391](https://pubmed.ncbi.nlm.nih.gov/20034391/)
30. Di Pietro C, Rapisarda A, Amico V, Bonaiuto C, Viola A, Scalia M, et al. Genomic localization of the human genes TAF1A, TAF1B and TAF1C, encoding TAF(I)48, TAF(I)63 and TAF(I)110 subunits of class I general transcription initiation factor SL1. *Cytogenet Cell Genet*. 2000;89(1–2):133–6. <https://doi.org/10.1159/000015592> PMID: [10894955](https://pubmed.ncbi.nlm.nih.gov/10894955/)
31. Silverman RH. Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. *J Virol*. 2007;81(23):12720–9. <https://doi.org/10.1128/JVI.01471-07> PMID: [17804500](https://pubmed.ncbi.nlm.nih.gov/17804500/)
32. He Y, Hara H, Núñez G. Mechanism and Regulation of NLRP3 Inflammasome Activation. *Trends Biochem Sci*. 2016;41(12):1012–21. <https://doi.org/10.1016/j.tibs.2016.09.002> PMID: [27669650](https://pubmed.ncbi.nlm.nih.gov/27669650/)
33. Koop A, Lepenies I, Braum O, Davarnia P, Scherer G, Fickenscher H, et al. Novel splice variants of human IKKε negatively regulate IKKε-induced IRF3 and NF-κB activation. *Eur J Immunol*. 2011;41(1):224–34. <https://doi.org/10.1002/eji.201040814> PMID: [21182093](https://pubmed.ncbi.nlm.nih.gov/21182093/)
34. Cheng G, Baltimore D. TANK, a co-inducer with TRAF2 of TNF- and CD 40L-mediated NF-kappaB activation. *Genes Dev*. 1996;10(8):963–73. <https://doi.org/10.1101/gad.10.8.963> PMID: [8608943](https://pubmed.ncbi.nlm.nih.gov/8608943/)
35. Capra JA, Williams AG, Pollard KS. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput Biol*. 2012;8(6):e1002567. <https://doi.org/10.1371/journal.pcbi.1002567> PMID: [22761559](https://pubmed.ncbi.nlm.nih.gov/22761559/)
36. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A*. 2005;102(40):14338–43. <https://doi.org/10.1073/pnas.0504070102> PMID: [16176987](https://pubmed.ncbi.nlm.nih.gov/16176987/)
37. Herrera-Urbe J, Liu H, Byrne KA, Bond ZF, Loving CL, Tuggle CK. Changes in H3K27ac at Gene Regulatory Regions in Porcine Alveolar Macrophages Following LPS or Poly(I:C) Exposure. *Front Genet*. 2020;11:817. <https://doi.org/10.3389/fgene.2020.00817> PMID: [32973863](https://pubmed.ncbi.nlm.nih.gov/32973863/)
38. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. *Cell*. 2015;160(3):554–66. <https://doi.org/10.1016/j.cell.2015.01.006> PMID: [25635462](https://pubmed.ncbi.nlm.nih.gov/25635462/)
39. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648–60. <https://doi.org/10.1126/science.1262110> PMID: [25954001](https://pubmed.ncbi.nlm.nih.gov/25954001/)
40. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep*. 2017;7(1):4200. <https://doi.org/10.1038/s41598-017-04520-z> PMID: [28646208](https://pubmed.ncbi.nlm.nih.gov/28646208/)
41. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferon v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res*. 2013;41(Database issue):D1040–6. <https://doi.org/10.1093/nar/gks1215> PMID: [23203888](https://pubmed.ncbi.nlm.nih.gov/23203888/)
42. Martinez-Jimenez CP, Eling N, Chen H-C, Vallejos CA, Kolodziejczyk AA, Connor F, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science*. 2017;355(6332):1433–6. <https://doi.org/10.1126/science.aah4115> PMID: [28360329](https://pubmed.ncbi.nlm.nih.gov/28360329/)
43. Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lässig M. Adaptive Evolution of Gene Expression in Drosophila. *Cell Rep*. 2017;20(6):1385–95. <https://doi.org/10.1016/j.celrep.2017.07.033> PMID: [28793262](https://pubmed.ncbi.nlm.nih.gov/28793262/)
44. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol*. 2011;1(6):519–25. <https://doi.org/10.1016/j.coviro.2011.10.008> PMID: [22328912](https://pubmed.ncbi.nlm.nih.gov/22328912/)
45. Tisoncik JR, Korth MJ, Simmons CP, Farrar J, Martin TR, Katze MG. Into the eye of the cytokine storm. *Microbiol Mol Biol Rev*. 2012;76(1):16–32. <https://doi.org/10.1128/MMBR.05015-11> PMID: [22390970](https://pubmed.ncbi.nlm.nih.gov/22390970/)
46. Hall JC, Rosen A. Type I interferons: crucial participants in disease amplification in autoimmunity. *Nat Rev Rheumatol*. 2010;6(1):40–9. <https://doi.org/10.1038/nrrheum.2009.237> PMID: [20046205](https://pubmed.ncbi.nlm.nih.gov/20046205/)
47. Crow YJ, Stetson DB. The type I interferonopathies: 10 years on. *Nat Rev Immunol*. 2022;22(8):471–83. <https://doi.org/10.1038/s41577-021-00633-9> PMID: [34671122](https://pubmed.ncbi.nlm.nih.gov/34671122/)
48. Bastard P, Zhang Q, Zhang S-Y, Jouanguy E, Casanova J-L. Type I interferons and SARS-CoV-2: from cells to organisms. *Curr Opin Immunol*. 2022;74:172–82. <https://doi.org/10.1016/j.coi.2022.01.003> PMID: [35149239](https://pubmed.ncbi.nlm.nih.gov/35149239/)
49. Kerner G, Neehus A-L, Philippot Q, Bohlen J, Rinchai D, Kerrouche N, et al. Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell Genom*. 2023;3(2):100248. <https://doi.org/10.1016/j.xgen.2022.100248> PMID: [36819665](https://pubmed.ncbi.nlm.nih.gov/36819665/)
50. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet*. 2012;46:677–700. <https://doi.org/10.1146/annurev-genet-110711-155522> PMID: [23145935](https://pubmed.ncbi.nlm.nih.gov/23145935/)

51. Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol*. 2012;12(10):687–95. <https://doi.org/10.1038/nri3295> PMID: [22976433](https://pubmed.ncbi.nlm.nih.gov/22976433/)
52. Judd EN, Gilchrist AR, Meyerson NR, Sawyer SL. Positive natural selection in primate genes of the type I interferon response. *BMC Ecol Evo*. 2021;21:65. <https://doi.org/10.1186/s12862-021-01783-z>
53. Cohn O, Yankovitz G, Peshes-Yaloz N, Steuerma Y, Frishberg A, Brandes R, et al. Distinct gene programs underpinning disease tolerance and resistance in influenza virus infection. *Cell Syst*. 2022;13(12):1002-1015.e9. <https://doi.org/10.1016/j.cels.2022.11.004> PMID: [36516834](https://pubmed.ncbi.nlm.nih.gov/36516834/)
54. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197> PMID: [28263959](https://pubmed.ncbi.nlm.nih.gov/28263959/)
55. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. *Database (Oxford)*. 2016;2016:bav096. <https://doi.org/10.1093/database/bav096> PMID: [26896847](https://pubmed.ncbi.nlm.nih.gov/26896847/)
56. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: [19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/)
57. Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A, Teichmann SA. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol Syst Biol*. 2011;7:497. <https://doi.org/10.1038/msb.2011.28> PMID: [21654674](https://pubmed.ncbi.nlm.nih.gov/21654674/)
58. Fraimovitch E, Hagai T. Promoter evolution of mammalian gene duplicates. *BMC Biol*. 2023;21(1):80. <https://doi.org/10.1186/s12915-023-01590-6> PMID: [37055747](https://pubmed.ncbi.nlm.nih.gov/37055747/)
59. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amodè MR, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49(D1):D884–91. <https://doi.org/10.1093/nar/gkaa942> PMID: [33137190](https://pubmed.ncbi.nlm.nih.gov/33137190/)
60. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006;22(10):1269–71. <https://doi.org/10.1093/bioinformatics/btl097> PMID: [16543274](https://pubmed.ncbi.nlm.nih.gov/16543274/)
61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
62. 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: Statistical Methodology*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>