

Published in final edited form as:

Nature. 2018 November ; 563(7730): 197–202. doi:10.1038/s41586-018-0657-2.

Gene expression variability across cells and species shapes innate immunity

Tzachi Hagai^{1,2,*}, Xi Chen¹, Ricardo J Miragaia^{1,3}, Raghd Rostom^{1,2}, Tomás Gomes¹, Natalia Kunowska¹, Johan Henriksson¹, Jong-Eun Park¹, Valentina Proserpio^{4,5}, Giacomo Donati^{4,6}, Lara Bossini-Castillo¹, Felipe A Vieira Braga^{1,7}, Guy Naamati², James Fletcher⁸, Emily Stephenson⁸, Peter Vegh⁸, Gosia Trynka¹, Ivanela Kondova⁹, Mike Dennis¹⁰, Muzlifah Haniffa^{8,11}, Armita Nourmohammad^{12,13}, Michael Lässig¹⁴, and Sarah A Teichmann^{1,2,15,*}

¹Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

²EMBL- European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK ³Centre of Biological Engineering, University of Minho, Braga, Portugal

⁴Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Torino, Italy ⁵Italian Institute for Genomic Medicine (IIGM), Via Nizza 52, 10126 Torino, Italy ⁶Molecular Biotechnology Center, University of Turin, Via Nizza 52, 10126, Torino, Italy

⁷Open Targets, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK ⁸Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK ⁹Division of Pathology and Microbiology, Animal Science Department, Biomedical

Primate Research Centre, 2280 GH Rijswijk, the Netherlands ¹⁰Research Department, Public Health England, National Infection Service, Porton Down, Wiltshire, SP4 0JG, UK ¹¹Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS

Foundation Trust, Newcastle upon Tyne, NE2 4LP, UK ¹²Max Planck Institute for Dynamics and Self-Organization, am Faßberg 17, 37077 Göttingen, Germany ¹³Department of physics, University of Washington, 3910 15th Avenue Northeast, Seattle, WA 98195, USA ¹⁴Institute for Biological Physics, University of Cologne, Zùlpicher Str. 77, 50937 Cologne, Germany ¹⁵Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, UK

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests should be addressed to Sarah Teichmann - st9@sanger.ac.uk; Tzachi Hagai – tzachi@ebi.ac.uk.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability

Scripts for major analyses are available in: https://github.com/Teichlab/innate_evo.

Data availability

Sequencing data are deposited in ArrayExpress with the following accessions: E-MTAB-5918, E-MTAB-5919, E-MTAB-5920, E-MTAB-6754, E-MTAB-6773, E-MTAB-5988, E-MTAB-5989, E-MTAB-6831, E-MTAB-6066, E-MTAB-7032, E-MTAB-7037, E-MTAB-7051, E-MTAB-7052.

Author contributions:

T.H. and S.A.T. designed the project, T.H., X.C., R.J.M., R.R., N.K., and J.P. performed the experiments with help from V.P., G.D. and F.A.V.B.; T.H., X.C., R.J.M., R.R., T.G. and J.H. analysed the data with help from G.N., L.B., G.T., A.N. and M.L.; J.F., E.S., P.V., I.K., M.D. and M.H. provided samples; S.A.T. supervised the project; T.H., R.R., N.K. and S.A.T. wrote the manuscript with inputs from all authors.

Competing interests: The authors declare no competing interests.

Summary

As the first line of defence against pathogens, cells mount an innate immune response, which is highly variable from cell to cell. The response must be potent yet carefully controlled to avoid self-damage. How these constraints have shaped the evolution of innate immunity remains poorly understood. Here, we characterise this programme's transcriptional divergence between species and expression variability across cells. Using bulk and single-cell transcriptomics in fibroblasts and mononuclear phagocytes from different species, challenged with immune stimuli, we reveal a striking architecture of the innate immune response. Transcriptionally diverging genes, including cytokines and chemokines, vary across cells and have distinct promoter structures. Conversely, genes involved in response regulation, such as transcription factors and kinases, are conserved between species and display low cell-to-cell expression variability. We suggest that this unique expression pattern, observed across species and conditions, has evolved as a mechanism for fine-tuned regulation, achieving an effective but balanced response.

The innate immune response is a cell-intrinsic defence program that is rapidly upregulated upon infection in most cell types. It acts to inhibit pathogen replication while signalling the pathogen's presence to other cells. This programme involves modulation of several cellular pathways, including production of antiviral and inflammatory cytokines, upregulation of genes functioning in pathogen restriction, and induction of cell death^{1,2}.

An important characteristic of the innate immune response is the rapid evolution that many of its genes have undergone along the vertebrate lineage^{3,4}. This is often attributed to pathogen-driven selection^{5–7}.

Another hallmark of this response is its high level of heterogeneity among responding cells: various studies have shown that cells display extensive cell-to-cell variability in response to pathogen infection^{8,9} or to pathogen-associated molecular patterns (PAMPs)^{10,11}. The functional importance of this cell-to-cell variability is unclear.

These two characteristics – rapid divergence in the course of evolution and high cell-to-cell variability – seem to be at odds with the strong regulatory constraint imposed on the host immune response: the need to execute a well-coordinated and carefully balanced programme to avoid tissue damage and pathological immune conditions^{12–15}. How this tight regulation is maintained despite rapid evolutionary divergence and high cell-to-cell variability remains an open question, central to our understanding of the innate immune response and its evolution.

Here, we study the evolution of this programme using two different cell types – fibroblasts and mononuclear phagocytes – across different mammalian clades challenged with several immune stimuli (Fig. 1a).

Our main experimental system uses primary dermal fibroblasts, which are commonly employed in immunological studies^{8,13}. We compare the response of fibroblasts from primates (human and macaque) and rodents (mouse and rat) to poly I:C, a synthetic dsRNA (Fig. 1a, left). Poly I:C is frequently used to mimic viral infection as it rapidly elicits an antiviral response¹⁶.

We comprehensively characterize the transcriptional changes between species and among individual cells in their innate immune response. We use population (bulk) transcriptomics to investigate transcriptional divergence between species, and single-cell transcriptomics to estimate cell-to-cell variability in gene expression. Using promoter sequence analyses along with ChIP-seq, we study how changes in the expression of each gene between species and across cells relate to the architecture of its promoter. Furthermore, we examine the relationship between cross-species divergence in gene sequence and expression and constraints imposed by host-pathogen interactions.

Additionally, we use a second system – bone marrow-derived mononuclear phagocytes from mouse, rat, rabbit and pig challenged with lipopolysaccharide (LPS), a commonly used PAMP of bacterial origin (Fig. 1a, right).

Taken together, these two systems provide insights into the architecture of the immune response across species, cell types and immune challenges.

Results

Transcriptional divergence in immune response

First, we studied the transcriptional response of fibroblasts to dsRNA (poly I:C) stimulation across the four species (human, macaque, rat and mouse). We generated bulk RNA-sequencing data for each species after 4 hours of stimulation, along with respective controls (see Fig. 1a and Methods for detailed experimental design).

In all species, we observe rapid upregulation of expected antiviral and inflammatory genes, including IFNB, TNF, IL1A and CCL5, following dsRNA treatment (see also Supplementary Table 3). Focusing on one-to-one orthologs, we performed correlation analysis between species and observed a similar transcriptional response (Spearman correlation, p -value $< 10^{-10}$ in all comparisons; Extended Data Fig. 1), as reported in other immune contexts^{17–19}. Furthermore, the response tends to be more strongly correlated between closely related species, similar to other expression programmes^{20–24}.

We characterized the differences in response to dsRNA between species for each gene, using this cross-species bulk transcriptomics data. While some genes, such as the NF- κ B subunits RELB and NFKB2, have a similar response across species, other genes respond differently between the primate and the rodent clades (Fig. 1b, left). For example, IFI27 (a restriction factor against numerous viruses) is strongly upregulated in primates but not in rodents, while DAXX (an antiviral transcriptional repressor) exhibits the opposite behaviour.

Similarly, in our second experimental system, the LPS-stimulated mononuclear phagocytes from mouse, rat, rabbit, and pig (Fig. 1b, right), we observe genes with comparable response across species (e.g. NFKB2), while others are highly upregulated only in specific clades (e.g. PHLDA1).

To quantify transcriptional divergence in immune response between species, we focused on genes that are differentially expressed during the stimulation (see Methods). For simplicity, we refer to these genes as “responsive genes” (Fig. 1c). In this analysis, we study the subset

of these genes with one-to-one orthologs across the studied species. There are 955 such genes in dsRNA-stimulated human fibroblasts and 2,336 genes in mouse LPS-stimulated phagocytes. We define a measure of response divergence by calculating the differences between the fold change estimates while taking the phylogenetic relationship into account (Methods, Supplementary Figures 1-7 and Supplementary Table 4).

For subsequent analyses, we split the 955 genes responsive in fibroblasts into three groups based on their level of response divergence: (1) highly divergent dsRNA-responsive genes (the top 25% genes with the highest divergence values in response to dsRNA across the four studied species), (2) lowly divergent dsRNA-responsive genes (the bottom 25%), and (3) genes with medium divergence across species (the middle 50%) (Fig 1c). We performed an analogous procedure for the 2,336 LPS-responsive genes in phagocytes.

Promoter architecture of diverging genes

Next, we tested whether divergence in transcriptional response is reflected in the conservation of promoter function and sequence. Using ChIP-seq, we profiled the active histone marks in the fibroblasts of all species. We observe that the presence of H3K4me3 mark in promoter regions of highly divergent genes is significantly less conserved between human and rodents than in promoters of lowly divergent genes (Extended Data Fig. 2).

We then used the human H3K4me3 ChIP-seq peaks to define active promoter regions of the responsive genes in human fibroblasts. We show that the density of transcription factor binding motifs (TFBMs) in the active promoter regions of highly divergent genes is significantly higher than in lowly divergent genes (Fig. 2a). Interestingly, when comparing the conservation of the core promoter regions in highly *versus* lowly divergent dsRNA-responsive genes, we observe that genes that highly diverge in response show higher sequence conservation in this region (Fig. 2b).

This surprising discordance may be related to the fact that promoters of highly and lowly divergent genes have distinctive architectures, associated with different constraints on promoter sequence evolution^{18,25,26}. Notably, promoters containing TATA-box elements tend to have most of their regulatory elements in regions immediately upstream of the transcription start site (TSS). These promoters are thus expected to be more conserved. The opposite is true for CpG island (CGI)^{26,27} promoters. Indeed, we found that TATA-boxes are associated with higher transcriptional divergence, while genes with CGIs diverge more slowly, both in fibroblasts and phagocytes (Fig. 2c and Extended Data Fig. 3). Thus, a promoter architecture enriched with TATA-boxes and depleted of CGIs is associated with higher transcriptional divergence, while entailing higher sequence conservation upstream of these genes^{18,26,27}.

Cytokines diverge rapidly in immune response

We next asked whether different functional classes among responsive genes are characterised by varying levels of transcriptional divergence. For this, we divided them into categories based on their function (such as cytokines, transcriptional factors, kinases) or the processes in which they are known to be involved (such as apoptosis or inflammation).

We observed that genes related to cellular defence and inflammation – most notably cytokines, chemokines and their receptors (hereafter “cytokines”) – tend to diverge in response significantly faster than genes involved in apoptosis or immune regulation (chromatin modulators, transcription factors, kinases and ligases) (Figs. 2d-e, Extended Data Fig. 4, and Supplementary Fig. 1).

We note that cytokines also have a higher transcriptional range in response to immune challenge (a higher fold change). Regressing the fold change from the divergence estimates results in reduction of the relative divergence of cytokines *versus* other responsive genes, but the difference still remains significant (Supplementary Fig. 2). Cytokine promoters are enriched in TATA-boxes (17% *versus* 2.5%, $p=1.1 \times 10^{-3}$, Fisher's exact test) and depleted of CGIs (14% *versus* 69%, $p=1.6 \times 10^{-9}$), suggesting that this promoter architecture is associated both with greater differences between species (response divergence) and larger changes between conditions (transcriptional range).

Cell-to-cell variability in immune response

Previous studies have shown that the innate immune response displays high variability across responding cells^{28,29}. However, the relationship between cell-to-cell transcriptional variability and response divergence between species is not well-understood.

To study gene expression heterogeneity across individual cells, we performed single-cell RNA-seq in all species in a time course following immune stimulation. We estimated cell-to-cell variability quantitatively using an established measure for variability: “distance to median” (DM)³⁰.

We found a clear trend where genes that are highly divergent in response between species are also more variable in expression across individual cells within a species (Fig. 3a). The relationship between rapid divergence and high cell-to-cell variability holds true in both the 955 dsRNA-responsive genes in fibroblasts and the 2,336 LPS-responsive genes in phagocytes. This can be observed across the stimulation time points and in different species (Extended data Figs. 5-6). We have analysed in depth the relationship between transcriptional divergence and cell-to-cell variability by using additional immune stimulations (Supplementary Figs. 8-9), and different experimental and computational approaches (Extended Data Fig 7 and Supplementary Figs. 10-13). Importantly, the trends we observed are not a result of technical biases due to low expression levels in either the bulk or the single-cell RNA-seq data (Supplementary Figs. 14-15).

Next, we examined the relationship between the presence of promoter elements – CGIs and TATA-boxes - and a gene's cell-to-cell variability. We observed that genes that are predicted to have a TATA-box in their promoter have significantly higher transcriptional variability, while CGI-genes tend to have lower variability (Fig. 3b), in agreement with previous findings³¹. Thus, both transcriptional variability between cells (Fig. 3b) and transcriptional divergence between species (Fig. 2c) are associated with the presence of specific promoter elements.

Transcriptional variability of cytokines

We subsequently compared the response divergence across species with the transcriptional cell-to-cell variability of three groups of responsive genes with different functions: cytokines, transcription factors, and kinases and phosphatases (hereafter “kinases”) (Fig. 3c and Extended Data Fig. 8). In contrast to kinases and transcription factors, many cytokines display relatively high levels of cell-to-cell variability (Extended Data Fig. 9), being expressed only in a small subset of responding cells (Extended Data Fig. 10). This has previously been reported for several cytokines²⁹. For example, IFNB is expressed in only a small fraction of cells infected with viruses or challenged with various stimuli^{8,11,32}. Here, we observe high levels of expression variability between cells for cytokines from several different families (e.g. IFNB, CXCL10, CCL2).

Cell-to-cell variability of cytokines remains relatively high in comparison with kinases and transcription factors during a time course of 2, 4 and 8hrs following dsRNA stimulation of fibroblasts (Extended Data Fig. 9). This pattern is similar across species, and can also be observed in LPS-stimulated phagocytes (Extended Data Fig. 9). Thus, the high variability of cytokines and their expression in a small fraction of stimulated cells across all time points is evolutionarily conserved.

We observed that cytokines tend to be co-expressed in the same cells, raising the possibility that their expression is coordinated (See Supplementary Information and Supplementary Fig 16). We also identified genes whose expression is correlated with cytokines in human fibroblasts and showed that their orthologs are significantly co-expressed with cytokines in the other species. This set is enriched with genes known to be involved in cytokine regulation (Supplementary Table 5).

As an example, we focused on the genes whose expression is positively correlated with the chemokine CXCL10 in at least two species (Fig. 3d). This set includes four cytokines co-expressed with CXCL10 (in purple), as well as known positive regulators of the innate immune response and cytokine production (in blue), such as the viral sensors MDA5 and RIG-I. This is in agreement with previous findings that suggest that IFNB expression is limited to cells where important upstream regulators are expressed at high enough levels^{8,11,32}. Here, we showed that this phenomenon of co-expression with upstream regulators applies to a wide set of cytokines and is conserved across species. Importantly, we found that cytokines are not only co-expressed with their positive regulators but also with genes that are known to act as negative regulators of cytokine expression or cytokine signalling (in red), suggesting how cytokine expression and function is tightly controlled at the level of individual cells.

The Evolutionary landscape of innate immunity

Previously, many immune genes, including several cytokines and their receptors, have been shown to evolve rapidly in coding sequence^{3,33}. However, it is not known how divergence in coding sequence relates to transcriptional divergence in innate immune genes. Using the set of 955 dsRNA-responsive genes in fibroblasts, we assessed the coding sequence

evolution in the three subsets of lowly, medium and highly divergent genes (as defined in Fig 1c).

We compared the rate at which genes evolved in their coding sequences with their response divergence by considering the ratio between non-synonymous (dN) to synonymous substitutions (dS). We observed that genes that rapidly evolve in transcriptional response have higher coding sequence divergence (higher dN/dS values) than dsRNA-responsive genes with low response divergence (Fig 4a).

Rapid gene duplication and gene loss have been observed in several important immune genes^{34–39} and are thought to be a result of pathogen-driven pressure^{40,41}. We have thus tested the relationship between a gene's divergence in response and the rate at which the gene's family has expanded and contracted in the course of vertebrate evolution. We found that transcriptionally divergent dsRNA-responsive genes have higher rates of gene gain and loss (Fig 4b) and consequently are also evolutionarily younger (Fig. 4c and Supplementary Fig. 17).

Previous reports suggest that younger genes tend to have fewer protein-protein interactions within cells⁴². Indeed, we observe that rapidly diverging genes tend to have fewer protein-protein interactions (Fig 4d). Together, these results suggest that transcriptionally divergent dsRNA-responsive genes evolve rapidly through various mechanisms, including fast coding sequence evolution and higher rates of gene loss and duplication events, and have fewer protein interactions with other cellular proteins.

The interaction between pathogens and the host immune system is thought to be an important driving force in the evolution of both sides. We thus investigated the relationship between transcriptional divergence and interactions with viral proteins by compiling a dataset of known host-virus interactions in human^{6,43,44}. Interestingly, genes with no known viral interactions have higher response divergence than genes with viral interactions (Fig 4e). Furthermore, the transcriptional divergence of genes targeted by viral immunomodulators⁴⁵ - viral proteins that subvert the host immune system - is lower still (Fig 4e). These observations suggest that viruses have evolved to modulate the immune system by interacting with immune proteins that are relatively conserved in their response. Presumably, these genes cannot evolve away from viral interactions, unlike host genes that are less constrained⁴⁶.

The summary of our results in Fig 4f highlights the difference in both the regulatory and evolutionary characteristics between cytokines and other representative dsRNA-responsive genes. Cytokines evolve rapidly through various evolutionary mechanisms and have higher transcriptional variability across cells. In contrast, genes that are involved in immune response regulation, such as transcription factors and kinases, are more conserved and less heterogeneous across cells. These genes have higher numbers of cellular protein-protein interactions, suggesting higher constraints imposed on their evolution. This group of conserved genes is more often targeted by viruses, revealing a relationship between host-pathogen dynamics and the evolutionary landscape of the innate immune response.

Discussion

In this work, we have charted the evolutionary architecture of the innate immune response. We show that genes rapidly diverging between species exhibit higher levels of variability in their transcription across individual cells. Both of these characteristics are associated with a similar promoter architecture, enriched with TATA-boxes and depleted of CGIs. Importantly, such promoter architecture is also associated with the high transcriptional range of genes during the immune response. Thus, transcriptional changes between conditions (stimulated *versus* unstimulated), species (transcriptional divergence), and individual cells (cell-to-cell variability) may all be mechanistically related to the same promoter characteristics. Interestingly, in yeast, TATA-boxes are enriched in promoters of stress-related genes, displaying rapid transcriptional divergence between species and high variability in expression^{30,47}. This suggests intriguing analogies between the mammalian immune and yeast stress responses - two systems that were exposed to continuous changes in external stimuli during evolution.

In addition, we show that genes involved in regulation of the immune response – such as transcription factors and kinases, are relatively conserved in their transcriptional response. These genes might have stronger functional and regulatory constraints, due to their roles in multiple contexts and pathways, limiting their ability to evolve. This can be an Achilles' heel used by pathogens to subvert the immune system. Indeed, we observe that viruses preferentially interact with conserved proteins of the innate immune response. Cytokines, on the other hand, diverge rapidly between species, stemming from their promoter architecture and from having fewer constraints imposed by intracellular interactions or additional non-immune functions. We thus suggest that cytokines represent a successful host strategy to counteract rapidly evolving pathogens as part of the host-pathogen evolutionary arms race.

Cytokines also display high cell-to-cell variability and tend to be co-expressed together with other cytokines and cytokine's regulators in a small subset of cells - a conserved pattern across species. Since prolonged or increased cytokine expression can result in tissue damage^{48–50}, restricting cytokine production to only a few cells may enable a rapid, yet controlled, response across the tissue to avoid long-lasting and potentially damaging effects.

Methods

Ethical compliance

This project has been approved by the Wellcome Sanger Institute Animal Welfare and Ethical Review Body, and complied with all relevant ethical regulations regarding animal research and human studies. Human cells were obtained from the Hipsci project⁵¹, where they were collected from consented research volunteers recruited from the NIHR Cambridge BioResource (written consent was given). Human skin profiling was performed in accordance with protocols approved by the Newcastle Research Ethics Committee (REC approval 08/H0906/95+5). Macaque skin samples were obtained from animals assigned to unrelated non-infectious studies, provided by Public Health England's National Infection Service in accordance with Home Office (UK) guidelines and approved by the Public Health England Ethical Review Committee under an appropriate UK Home Office project license.

(1) Cross-species dermal fibroblast stimulation with dsRNA and IFNB

Experimental methods

Tissue culture: We cultured primary dermal fibroblasts from low passage cells (below 10) that originated from females from four different species (human (European ancestry), rhesus macaque, black 6 mouse and brown Norway rat). All skin samples originated from shoulders. Stimulation experiments and library preparations were done in identical conditions across all species and for all genomics techniques. Details on the numbers of individuals used in each technique are listed in each technique's section and in Supplementary Table 1.

Human cells were obtained from the Hipsci project (<http://www.hipsci.org/51>). Rhesus macaque cells were extracted from skin tissues that were incubated for 2hrs with 0.5% collagenase B (Roche; 11088815001) after mechanical processing, and then filtered through 100µm strainers before being plated and passaged prior to cryo-banking. Rodent cells were obtained from PelloBiotech where they were extracted using a similar protocol. *In vitro* cultured fibroblasts from all four species resemble a particular *in vivo* cluster of dermal fibroblasts (see Supplementary Information).

Prior to stimulation, cells were thawed and grown for several days in ATCC fibroblast growth medium (Fibroblast Basal Medium (ATCC, ATCC-PCS-201-030) with Fibroblast Growth Kit-Low serum (ATCC, PCS-201-041) (supplemented with Primocin (Invivogen, ant-pm-1) and Pen/Strep (Life Technologies, Cat. Code: 15140122))– a controlled medium that has proven to provide good growing conditions for fibroblasts from all species, having slightly less than 24h doubling times. ~18hrs before stimulation, cells were trypsinized, counted and cells were seeded into 6-well plates (100,000 cells per well). Cells were stimulated with either: (1) 1µg/mL High-Molecular Weight poly I:C (Invivogen, Cat. Code: tlr-pic) transfected with 2µL/mL Lipofectamin 2,000 (ThermoFisher, Cat Number 11668027); (2) mock transfected with Lipofectamin 2,000; (3) stimulated with 1,000 IU of IFNB for 8 hours (human IFNB - 11410-2 (for human and macaque cells); rat IFNB - 13400-1; mouse IFNB – 12401-1; all IFNs were obtained from PBL, and had activity units based on similar virological assays); or (4) left untreated. Interferon stimulations were used as a control, to study how the genes upregulated in the secondary wave of innate immune response diverge between species.

Additional samples in human and mouse were stimulated with 1,000 IU of Cross-mammalian IFN (CMI, or Universal Type I IFN Alpha, PBL, Cat Number 11200-1). The latter stimulation was done to assess the effects of species-specific and batch-specific IFNB.

In all the above-mentioned stimulations, we used a longer time course for single-cell RNA-seq than for the bulk RNA-seq. This was done for two main reasons: (1) In the bulk, we chose to focus on one main stimulation time point for simplicity and to obtain an intuitive fold change between stimulated and unstimulated conditions. (2) In single-cells, when studying cell-to-cell variability, we chose to profile, in addition to the main stimulation time point, cells in earlier and later stages of the response. This is important for studying how the

dynamics and magnitude of the response affect gene expression variability between responding cells.

We note that the poly I:C we used was tested negative for the presence of bacterial beta-endotoxin using a coagulation test (PYROGENT Plus, 0.06 EU/ml Sensitivity, Cat No N283-06).

Bulk RNA sequencing

Library preparation and sequencing: For bulk transcriptomics analysis, individuals from different species were grown in parallel and stimulated with dsRNA, IFNB (and cross-mammalian IFN) and with their respective controls. In total, samples from 6 humans, 6 macaques, 3 mice and 3 rats were used. Total RNA was extracted using the RNeasy Plus Mini kit (Qiagen, Cat Number 74136), using QIAcube (Qiagen). RNA was then measured using a Bioanalyzer 2100 (Agilent Technologies), and samples with RIN<9 were excluded from further analysis (one macaque sample stimulated with poly I:C and its control).

Libraries were produced using the Kapa Stranded mRNA-seq Kit (Kapa Biosystems, Kit code: KK8421). The Kapa library construction protocol was modified for automated library preparation by Bravo (Agilent Technologies). cDNA was amplified in 13 PCR cycles, and purified by Ampure XP beads (Beckman Coulter, Cat Number A63882) (1.8x volume) using Zephyr (Perkin Elmer). Pooled samples were sequenced on an Illumina HiSeq 2500 instrument, using paired-end 125-bp reads.

ChIP sequencing

Library preparation and sequencing: Samples from three individuals from each of the four species were grown and stimulated (with poly I:C for 4 hours or left untreated, as described above) in parallel to samples collected for bulk RNA-seq. Following stimulation, sample were crosslinked in 1% HCHO (prepared in 1X DPBS) at room temperature for 10 minutes, and HCHO was quenched by the addition of glycine at a final concentration of 0.125M. Cells were pelleted at 4°C at 2000xg, washed with ice-cold 1X DPBS twice, and snapped frozen in liquid nitrogen. Cell pellets were stored in -80°C until further stages were performed. ChIPmentation was performed according to version 1.0 of the published protocol⁵² with a few modifications (see additional details in Supplementary Methods).

Library preparation reactions contained the following reagents: 10ul purified DNA (from the above procedure), 2.5ul PCR Primer Cocktails (Nextera kit, Illumina, Cat Number FC-121-1030), 2.5ul N5xx (Nextera index kit, Illumina Cat Number FC-121-1012), 2.5ul N7xx (Nextera index kit, Illumina, Cat Number FC-121-1012), 7.5ul NPM PCR Master Mix (Nextera kit, Illumina, Cat Number FC-121-1030). PCR cycles were as follows: 72°C, 5 mins; 98°C, 2 mins; 98°C, 10 secs, 63°C, 30 secs, 72°C, 20 secs] X 12; 10°C hold.

Amplified libraries were purified by double AmpureXP beads purification: first with 0.5X bead ratio, keep supernatant, second with 1.4X bead ratio, keep bound DNA. Elution was done in 20ul Buffer EB (QIAGEN).

1ul of library was run on a Bioanalyzer (Agilent Technologies) to verify normal size distribution. Pooled samples were sequenced on an Illumina HiSeq 2000 instrument, using paired-end 75-bp.

Single-Cell RNA sequencing

Flow cytometry: For scRNA-seq, we performed two different biological replicates, with each replicate having one individual from each of the four studied species. A time course of dsRNA stimulation of 0, 4, and 8hrs was used in one replicate (divided into two technical replicates), while a second replicate included a time course of 0, 2, 4, and 8hrs. Poly I:C transfection was done as described above. In the case of sorting with INFLUX, we used rhodamine-labelled poly I:C.

Cells were sorted with either Beckman Coulter MoFlo XDP (first replicate) or Becton Dickinson INFLUX (second replicate) into wells containing 2uL of Lysis Buffer (1:20 solution of RNase Inhibitor (Clontech, Cat Number 2313A) in 0.2% v/v Triton X-100 (Sigma-Aldrich, Cat Number T9284), spun down and immediately frozen at -80°.

When sorting with MoFlo, a pressure of 15psi was used with a 150µm nozzle, using the 'Single' sort purity mode. Dead or late-apoptosis cells were excluded using Propidium Iodide at 1µg/ml (Sigma, Cat Number P4170) and single cells were selected using FSC W vs FSC H. When sorting with INFLUX, a pressure of 3psi was used with a 200µm nozzle, with the "single" sort mode. Dead or late-apoptosis cells were excluded using 100ng/ml DAPI (4', 6-diamidino-2-phenylindole) (Sigma, Cat Number D9542). DAPI was detected using the 355nm laser (50mW) using a 460/50nm bandpass filter. Rhodamine was detected using the 561nm laser (50mW), using a 585/29nm bandpass filter. Single cells were collected using the FSC W vs FSC H.

Library preparation from full-length RNA from single cells and sequencing: Sorted plates were processed according to the Smart-seq2 protocol⁵³: Oligo-dT primer (IDT), dNTPs (ThermoFisher, Cat Number 10319879) and ERCC RNA Spike-In Mix (1:25,000,000 final dilution, Ambion, Cat Number 4456740) were added to each well, and Reverse Transcription (using 50U SmartScribe, Clontech Cat Number 639538) and PCR were performed following the original protocol with 25 PCR cycles. cDNA libraries were prepared using Nextera XT DNA Sample Preparation Kit (Illumina, Cat Number FC-131-1096), according to the protocol supplied by Fluidigm (PN 100-5950 B1). Quality Checks on cDNA were done using a Bioanalyser 2100 (Agilent Technologies). Libraries were quantified using the LightCycler 480 (Roche), pooled and purified using AMPure XP beads (Beckman Coulter) with Hamilton 384 head robot (Hamilton Robotics). Pooled samples were sequenced on an Illumina HiSeq 2500 instrument, using paired-end 125-bp reads.

Computational methods

Cross-species response divergence

Read mapping to annotated transcriptome: For bulk RNA-Seq samples, adaptor sequences and low-quality score bases were first trimmed using Trim Galore (version 0.4.1) (with the

parameters “--paired --quality 20 --length 20 -e 0.1 --adapter AGATCGGAAGAGC”). Trimmed reads were mapped and gene expression was quantified using Salmon (version 0.6.0)⁵⁴ with the following command: "salmon quant -i [index_file_directory] -l ISR -p 8 --biasCorrect --sensitive --extraSensitive -o [output_directory] -l -g [ENSEMBL_transcript_to_gene_file] --useFSPD --numBootstraps 100". Each sample was mapped to its respective species' annotated transcriptome (downloaded from ENSEMBL, version 84: GRCh38 for human, MMUL_1 for macaque, GRCm38 for mouse, Rnor_6.0 for rat). We only included the set of coding genes (*.cdna.all.fa files). We removed annotated secondary haplotypes of human genes by removing genes with “CHR_HSCHR”.

Quantifying differential gene expression in response to dsRNA: To quantify differential gene expression between treatment and control for each species and for each treatment separately, we used edgeR (version 3.12.1)⁵⁵ using the rounded estimated counts from Salmon. This was done only for genes that had a significant level of expression in at least one of the four species (TPM>3 in at least N-1 libraries, where N is the number of different individuals we have for this species with libraries that passed quality control). 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).

Conservation and divergence in response

Fold change-based phylogeny: We compared the overall change in response to treatment (dsRNA or IFNB) between pair of species, by computing the Spearman correlation of the fold change in response to treatment across all one-to-one orthologs that were expressed in at least one species (Extended Data Fig. 1a-h). Fold change was calculated with edgeR, as described above. Spearman correlations of all expressed genes appear in grey. Correlations of the subset of differentially expressed genes (genes with an FDR-corrected p-value<0.01 in at least one of the compared species) appear in black.

In panels a-c, we show comparisons in response to dsRNA. In panels d-f, we show comparisons in response to IFNB, which we use here to study the similarity of the secondary immune response between species.

We constructed a tree based on gene's change in expression in response to dsRNA and to IFNB, using expressed genes that had one-to-one orthologs across all four species and were expressed in at least one species in at least one condition (Extended Data Fig 1i). We used hierarchical clustering, with the hclust command from the stats R package, with the distance between samples computed as $1-\rho$, where ρ is the pairwise Spearman correlation between each pair of species mentioned above (a greater similarity, reflected in a higher correlation, results in a smaller distance) and “average” as the clustering method.

The above-mentioned analyses focus on one-to-one orthologs between the compared species. In Supplementary Table 6, we quantify the similarity in response between species (based on Spearman correlations) when adding genes with one-to-many orthologs.

Quantifying gene expression divergence in response to immune challenge: To quantify transcriptional divergence in immune response between species, we focus on genes that have

annotated one-to-one orthologs across the studied species (human, macaque, mouse and rat). 9,753 of the expressed genes have annotated one-to-one orthologs in all four species, out of which 955 genes are differentially expressed in human in response to dsRNA treatment (genes with an FDR-corrected p-value<0.01).

We define a measure of response divergence (based on a previous study⁵⁶) by calculating the differences between the fold change estimates across the orthologs:

$$\text{Response divergence} = \log[1/4 \times \sum_{i,j} (\log\text{Fold_Change}_{\text{Primate}_i} - \log\text{Fold_Change}_{\text{Rodent}_j})^2]$$

This measure takes into account the structure of the phylogeny, and gives a relative measure of divergence in response across all genes with one-to-one orthologs.

In order to consider differences between species, we here focus on between-clade differences (primates - rodents), rather than on within-clade differences. In this way, we map the most significant macro-evolutionary differences along the longest branches of our four-species phylogeny. In addition, averaging within clades acts as a reduction of noise⁵⁶.

We compared this divergence measure to two other measures that use models (and incorporate both between- and within-clade divergence) and show strong correlation between the divergence estimates across the three approaches (Supplementary Figures 3-4).

In most of the subsequent analyses, we focus on the 955 “dsRNA-responsive genes” – genes that are differentially expressed in response to dsRNA (genes that have an FDR-corrected p-value<0.01 in human, and have annotated one-to-one orthologs in the other three species). For some of the analyses, we split these 955 genes based on quartiles, into genes with high, medium and low divergence (Figure 1c).

We also studied how imprecisions in the fold change estimates affect the response divergence estimates and subsequent analyses (Supplementary Figures 5-6).

Comparison of response divergence between different functional groups: To compare the divergence rates between sets of dsRNA-responsive genes that have different functions in the innate immune response, we split these 955 genes into the following functional groups (all groups are mutually-exclusive, and any gene that belongs to two groups was excluded from the latter group; human gene annotations were used):

We first grouped genes by annotated molecular functions: Viral sensors (genes that belong to one of the GO categories: GO:0003725 - double-stranded RNA binding, GO:0009597 - detection of virus; GO:0038187 - pattern recognition receptor activity), Cytokines, chemokines & their receptors (GO:0005125 cytokine activity; GO:0008009 chemokine activity, GO:0004896 - cytokine receptor activity; GO:0004950 - chemokine receptor activity), Transcription factors (taken from the Animal Transcription Factor DataBase (version 2.0)⁵⁷), Chromatin modulators (GO:0016568 - chromatin modification, GO:0006338 - chromatin remodelling, GO:0003682 - chromatin binding, GO:0042393- histone binding), Kinases and phosphatases (GO:0004672 - protein kinase activity; GO:0004721 -

phosphoprotein phosphatase activity), Ligases and deubiquitinases (GO:0016579 protein deubiquitination; GO:0004842 - ubiquitin-protein transferase activity GO:0016874 ligase activity), Other enzymes (mostly involved in metabolism rather than regulation - GO:0003824 - catalytic activity). The divergence response values of these functional subsets were compared to the entire group of 955 dsRNA-responsive genes (Figures 2d-e).

Next, we grouped genes by biological processes that are known to be important in the innate immune response: Antiviral defence (GO:0051607 - defence response to virus), Inflammation (GO:0006954 - inflammatory response), Apoptosis (GO:0006915 - apoptotic process), and Regulation (GO annotations related to regulation of innate immune response pathways include only few genes. We thus used as the group of genes related to regulation, the merged group of genes that are annotated as transcription factors, chromatin modulators, kinases and phosphates or ligases and deubiquitinases, since all these groups include many genes that are known to regulate the innate immune response.)

Gene lists belonging to the mentioned GO annotations were downloaded using QuickGo58. The distribution of response divergence values for each of the functional groups was compared with the distribution of response divergence of the entire set of dsRNA-response genes. Cytokines, chemokines and their receptors are merged in Figs 2d-e & 3c. Analogous comparisons of functional groups in IFNB response (with 841 IFNB-responsive genes) are shown in Supplementary Figure 1. See additional analyses in Supplementary Information.

Conservation of active chromatin marks analysis

Alignment and peak calling of ChIP-seq reads: ChIP-seq reads were trimmed using trim_galore (version 0.4.1) with "--paired --trim1 --nextera" flags. The trimmed reads were aligned to the corresponding reference genome (hg38 for human, rheMac2 for macaque, mm10 for mouse, rn6 for rat – all these genomes correspond to the transcriptomes used for RNA-Seq mapping) from the UCSC Genome Browser59 using bowtie2 (version 2.2.3) with default settings60. In all four species, we removed the Y chromosome. In the case of human, we also removed all alternative haplotype chromosomes. Following alignment, low-confidence mapped and improperly-paired reads were removed by samtools61 with "-q 30 -f 2" flags.

Enriched regions (peaks) were called using MACS2 (v 2.1.1)62 with a corrected p-value cutoff of 0.01 with "-f BAMPE -q 0.01 -B --SPMR" flags, using input DNA as control. The genome sizes (the argument for "-g" flag) used were "hs" for human, "mm" for mouse, 3.0e9 for macaque and 2.5e9 for rat. Peaks were considered reproducible when they were identified in at least two of the three biological replicates and overlapped by at least 50% of their length (non-reproducible peaks were excluded from subsequent analyses). Reproducible peaks were then merged to create consensus peaks from overlapping regions of peaks from the three replicates by using mergeBed from the bedtools suite63.

Gene assignment and conservation of active promoters and enhancers: We subsequently linked human peaks with the genes they might be regulating as follows: H3K4me3 consensus peak was considered the promoter region of a given gene if its centre was

between 2kb upstream and 500bp downstream of the annotated TSS of the most abundantly-expressed transcript of that gene.

Similarly, H3K27ac was considered the enhancer region of a given gene if its centre was in a distance above 1kb and below 1Mb, and there was no overlap (of 1bp or more) with any H3K4me3 peak.

In each case where, based on the distance criteria, more than a single peak was linked to a gene (or more than a single gene was linked to a peak), we took only the closest peak-gene pair (ensuring that each peak will have up to one gene and vice versa).

To compare active promoters and enhancers between species, we excluded any human peak that could not be uniquely mapped to the respective region in the other species. This was done by looking for syntenic regions of human peaks in the other three species by using liftOver64, and removing peaks that have either unmapped regions or more than one mapped region in the compared species. We considered syntenic regions with at least 10% sequence similarity between the species (minMatch=0.1), with a minimal length (minSizeQ and minSizeT) corresponding to the length of the shortest peak (128bp in H3K4 and 142bp in H3K27).

We defined an active human promoter or enhancer as conserved if a peak was identified in the corresponding region of the other species (we repeated this analysis by comparing human with each of the other three species separately). We compared the occurrence of conserved promoters and enhancers in genes that are highly divergent in response to dsRNA with lowly divergent genes, and used Fisher's exact test to determine statistical significance of the observed differences between highly and lowly divergent genes. Results are shown in Extended Data Figure 2.

Promoter sequence analysis: To calculate the total number of transcription factors bindings motifs in a gene's active promoter region, we downloaded the non-redundant JASPAR core motif matrix (pfm Vertebrates.txt) from the JASPAR 2016 server⁶⁵ and searched for significant matches for these motifs using FIMO66 in human H3K4me3 peaks. Peak's TFBM density was calculated by dividing the total number of motif matches in a peak by the peak's length. TFBM density values in human H3K4me3 peaks linked with highly and lowly divergent genes were compared (Fig. 2a).

PhyloP7 values were used to assess promoter sequence conservation⁶⁷. Sequence conservation quantification was performed by taking the estimated nucleotide substitution rate for each nucleotide along the promoter sequence (500bp upstream of the TSS of the relevant human gene). When several annotated transcripts exist, the TSS of the most abundantly expressed transcript was used (based on bulk RNA data). The substitution rate values from all genes were aligned, based on their TSS position, and a mean for each of the 500 positions was calculated separately for the group of genes with high, medium and low response divergence. Two-sample Kolmogorov-Smirnov test was used to compare the paired distribution of rates between the mean of the highly divergent and the mean of the lowly divergent sets of genes. To plot the mean values of the three sets of divergent genes,

geom_smooth function from the ggplot2 R package was used with default parameters (with loess as the smoothing method) (see Fig. 2b).

Human CpG island (CGI) annotations were downloaded from the UCSC genome table browser (hg38), and CGI genes were defined as those with a CGI overlapping their core promoter (300bp upstream of the TSS reference position, and 100bp downstream of it, as suggested previously¹⁸). Genes were defined as having a TATA box if they had a significant match to the Jaspar TATA box matrix (MA0108.1) in the 100bp upstream of their TSS by FIMO66 with default settings (we used a 100bp window due to possible inaccuracies in TSS annotations). (We note that only 28 out of 955 dsRNA-response genes had a matching TATA-box motif in this region). For both TATA and CGI analyses the promoter sequences of the human orthologs were used.

Cell-to-cell transcriptional variability analysis

Read mapping and quality control of single-cell RNA-seq (full-length RNA): Gene expression was quantified in a manner similar to the quantification for bulk transcriptomics libraries described above. Low quality cells were filtered using quality control criteria (cells with at least 100,000 mapped reads, with at least 2,000 expressed genes with TPM>3, with ERCC<10% and MT<40%). This QC filtering resulted in 240 cells from a first biological replicate, including two technical replicates (with a time course of 0, 4, 8 hours). In a second larger biological replicate (with a dsRNA stimulation time course of 0, 2, 4, 8 hours), 728 cells passed QC. Results throughout the manuscript relate to the second cross-species biological replicate where a higher proportion of cells passed QC, and the lower quality first replicate data was not considered further.

Cell-to-cell variability analysis: To quantify gene's biological cell-to-cell variability we have applied the DM (Distance to Median) approach - an established method, which calculates the cell-to-cell variability in gene expression, while accounting for confounding factors such as gene expression level³⁰. This is done by first filtering out genes that are lowly expressed: For Smart-seq2 data we only included genes that have an average expression of at least 10 size-factor normalized reads (except for Extended Data Figure 9a, where we reduced the threshold to 5, to allow a larger number of genes to be included in the comparisons). This procedure was done to filter genes that display higher levels of technical variability between samples due to low expression. Secondly, to account for gene expression level, the observed cell-to-cell variability of each gene is compared with its expected variability, based on its mean expression across all samples and in comparison with a group of genes with similar levels of mean expression. This distance from mean value (DM), is also corrected by gene length (in the case of Smart-seq2 data), yielding a value of variability that can be compared across genes regardless of their length and mean expression values⁶⁸. As a second approach, we used BASiCS^{69,70} (see details in Supplementary Information).

We note that the relationship observed in Figure 3a between response divergence and cell-to-cell variability are not an artefact, stemming from differences in expression levels: (A) With respect to cell-to-cell variability, gene's expression level is controlled for by DM calculations, where expression level is regressed by using a running median (Supplementary

Figure 14). (B) Similarly, we can regress the expression level measured in bulk RNA-seq from the quantified response divergence by subtracting the running median of expression from the divergence estimates. When repeating the analysis comparing cell-to-cell variability *versus* regressed response divergence, the relationship between the two is maintained (Supplementary Figure 15).

Cytokine co-expression analysis: For the chemokine CXCL10, we built a network (using CytoScape71) of genes that correlate with CXCL10 in dsRNA-stimulated human fibroblasts and in at least one more species, using genes with a Spearman correlation value above 0.3 (see Fig 3d and additional analyses in Supplementary Information).

Analysis of different evolutionary modes

Coding sequence evolution analysis: The ratio dN/dS of non-synonymous to synonymous codon substitutions, of human genes across the mammalian clade was obtained from a previous study that used orthologous genes from 29 mammals72. Distributions of dN/dS values were computed for each of the three groups of genes with low, medium and high divergence in response to dsRNA, and are plotted in Figs 4a.

Rate of gene gain and loss analysis: The significance at which gene's family has experienced higher rate of gene gain and loss in the course of vertebrate evolution, in comparison with other gene families, was retrieved from ENSEMBL73. The statistics provided by ENSEMBL is calculated based on the CAFE method74, which estimates the global birth and death rate of gene families and identifies gene families that have accelerated rates of gain and loss. Distributions of the p-values from this statistic were computed for each of the three groups of genes with low, medium and high divergence in response to dsRNA and are plotted as the negative logarithm values in Fig 4b.

Gene age analysis: Gene age estimations were obtained from ProteinHistorian75. To ensure that the results are not biased by a particular method of ancestral protein family reconstruction or by specific gene family assignments, we used eleven different estimates for mammalian genes (combining five different databases of protein families with two different reconstruction algorithms for age estimation, as well as an estimate from the phylostratigraphic approach). For each gene, age is defined with respect to the species tree, where a gene's age corresponds to the branch in which its family is estimated to have appeared (thus, larger numbers indicate evolutionarily older genes).

Data for gene age in comparison with divergence in response to dsRNA is shown in Fig 4c (using Panther7 phylogeny and Wagner reconstruction algorithm) and in Supplementary Figure 17a (for all 11 combinations of gene family assignments and ancestral family reconstructions). See additional analyses in Supplementary Information.

Cellular protein-protein interactions analysis: Data on number of experimentally-validated protein-protein interactions (PPIs) for human genes was obtained from STRING (version 10)76. Distributions of PPIs for genes with low, medium and high divergence in response to dsRNA are plotted in Fig 4d.

Host-virus interactions analysis: Data on host-virus protein-protein interactions was downloaded from the VirusMentha database⁴³, and combined with two additional studies that have annotated host-virus protein-protein interactions^{6,44}. We split the 955 dsRNA-response genes into genes with known viral interactions (genes whose protein products were reported to interact with at least one viral protein), and genes with no known viral interactions: “Viral Interactors” and “No viral Interactions” respectively, in Figure 4e. In addition, we define a subset of genes within the viral interactors set: those known to interact with viral proteins that are immunomodulators (proteins known to target the host immune system and modulate its response, obtained from Pichlmair, A. *et al.*⁴⁵).

We note that the results presented in Figure 4e are in agreement with previous analyses that are based on all human genes and on coding sequence evolution⁴⁶. However, the overlap in the sets of genes between the previous analyses and the one presented here is small (e.g. in Dyer et al., 2009⁴⁶ there are 535 human genes with known interactions with pathogens, 57 of which overlap with the 955 genes that are the basis of the current analysis).

(2) Additional experiments with human fibroblasts and human skin tissue

Additional experiments were performed with human dermal fibroblasts and with cells extracted from human skin tissues to study in greater detail the relationship between response divergence across species and cell-to-cell variability. See Supplementary Methods and Supplementary Discussion for details.

(3) Cross-species bone marrow-derived phagocyte stimulation with LPS and dsRNA

Experimental procedures

Tissue culture: Primary bone marrow-derived mononuclear phagocytes originating from females of four different species (black 6 mouse, brown Norway rat, rabbit and pig) were obtained from PeloBiotech. 24 hours before the start of the stimulation time course, cells were thawed and split into 12-well plates (500,000 cells per well). Cells were stimulated with either: (1) 100ng/mL lipopolysaccharide (LPS) (Invivogen, Cat. Code: tlr1-smlps), or with (2) 1ug/mL High-Molecular Weight poly I:C (Invivogen, Cat. Code: tlr1-pic) transfected with 2uL/mL Lipofectamin 2,000 (ThermoFisher, Cat Number 11668027). LPS stimulation time courses of 0, 2, 4, 6 hours were performed for all species. Poly I:C stimulations were performed for rodents for 0, 2, 4, 6 hours. We also processed cells for bulk RNA-Seq for 0 and 4 hours stimulation time points. Details on the individuals used in each technique are listed in Supplementary Table 2.

Library preparation for single cells using microfluidic droplet cell capture: Following stimulation, cells were collected using Cell Dissociation Solution Non-enzymatic (Sigma-Aldrich, Cat No C5914), washed and resuspended in 1xPBS with 0.5% (w/v) BSA. Cells were then counted and loaded on the 10x Chromium machine aiming for a targeted cell recovery of 5,000 cells according to the manual. Libraries were prepared following the Chromium™ Single Cell 3' v2 Reagent Kit Manual⁷⁷. Libraries were sequenced on Illumina HiSeq 4000 instrument with 26bp for read 1 and 98bp for read 2.

Library preparation and sequencing for Bulk RNA-Sequencing: Total RNA was extracted and libraries were prepared as described in the fibroblasts section. Pooled samples were sequenced on an Illumina HiSeq 4000 instrument, using paired-end 75-bp reads.

Computational methods

Quantifying gene expression in bulk RNA transcriptomics: Adaptor sequences and low-quality score bases were trimmed using Trim Galore (version 0.4.1). Trimmed reads were mapped and gene expression was quantified using Salmon: (version 0.9.1)⁵⁴ with the following command: "salmon quant -i [index_file_directory] / -l ISR -p 8 --seqBias --gcBias --posBias -q -o [output_directory] -l -g [ENSEMBL_transcript_to_gene_file] --useVBOpt --numBootstraps 100". Mouse samples were mapped to mouse transcriptome (ENSEMBL, version 84). We note that we used the bulk data only for TSS analysis. For differential expression analysis, we used an *in silico* bulk from the single-cell data (see below).

Quantifying gene expression in microfluidic droplet cell capture data: Microfluidic droplet cell capture data was first quantified using 10X Genomics' Cell Ranger Single-Cell Software Suite (version 2.0, 10x Genomics Inc)⁷⁷ against the relevant genome (ENSEMBL, version 84). We removed cells with less than 200 genes or more than 10% mitochondrial reads. To remove potential doublets, we excluded the top 10% of cells expressing the highest numbers of genes. Genes expressed in less than 0.5% of the cells were excluded from the calculations. We then filtered cells that express less than 10% of the total number of filtered genes.

Since bone marrow-derived phagocytes may include secondary cell populations, we focused our analysis on the major cell population. We identified clusters within each dataset, using the Seurat⁷⁸ functions RunPCA, followed by FindClusters (using 20 dimensions from the PCA, default perplexity and a resolution of 0.1) and have taken the cells belonging to the largest cluster for further analysis, resulting in a less heterogeneous population of cells. Lower resolution of 0.03 was used for rabbit-LPS4, rabbit-LPS2, mouse-PIC2, mouse-PIC4; and 0.01 for rabbit-LPS6.

Quantifying gene expression divergence in response to immune challenge: We created an *in silico* bulk table, by summing up the UMIs of the post-QC single cells belonging to the largest cluster of cells, in each of the samples. We then used the three replicates in unstimulated conditions and in 4 hours LPS stimulation to perform a differential expression analysis using DESeq²⁷⁹ Wald test, and p-values were adjusted for multiple testing by estimating the false discovery rate (FDR). A similar procedure was performed with mouse and rat dsRNA stimulation (with 4 hours dsRNA stimulation *versus* unstimulated conditions).

To quantify transcriptional divergence in immune response between species, we focus on genes that have annotated one-to-one orthologs across the studied species.

We define a measure of response divergence by calculating the differences between the fold change estimates across the orthologs:

$$\text{Response divergence} = \log\left[\frac{1}{3} \times \sum_j (\log\text{Fold_Change Pig} - \log\text{Fold_Change Glire}_j)^2\right]$$

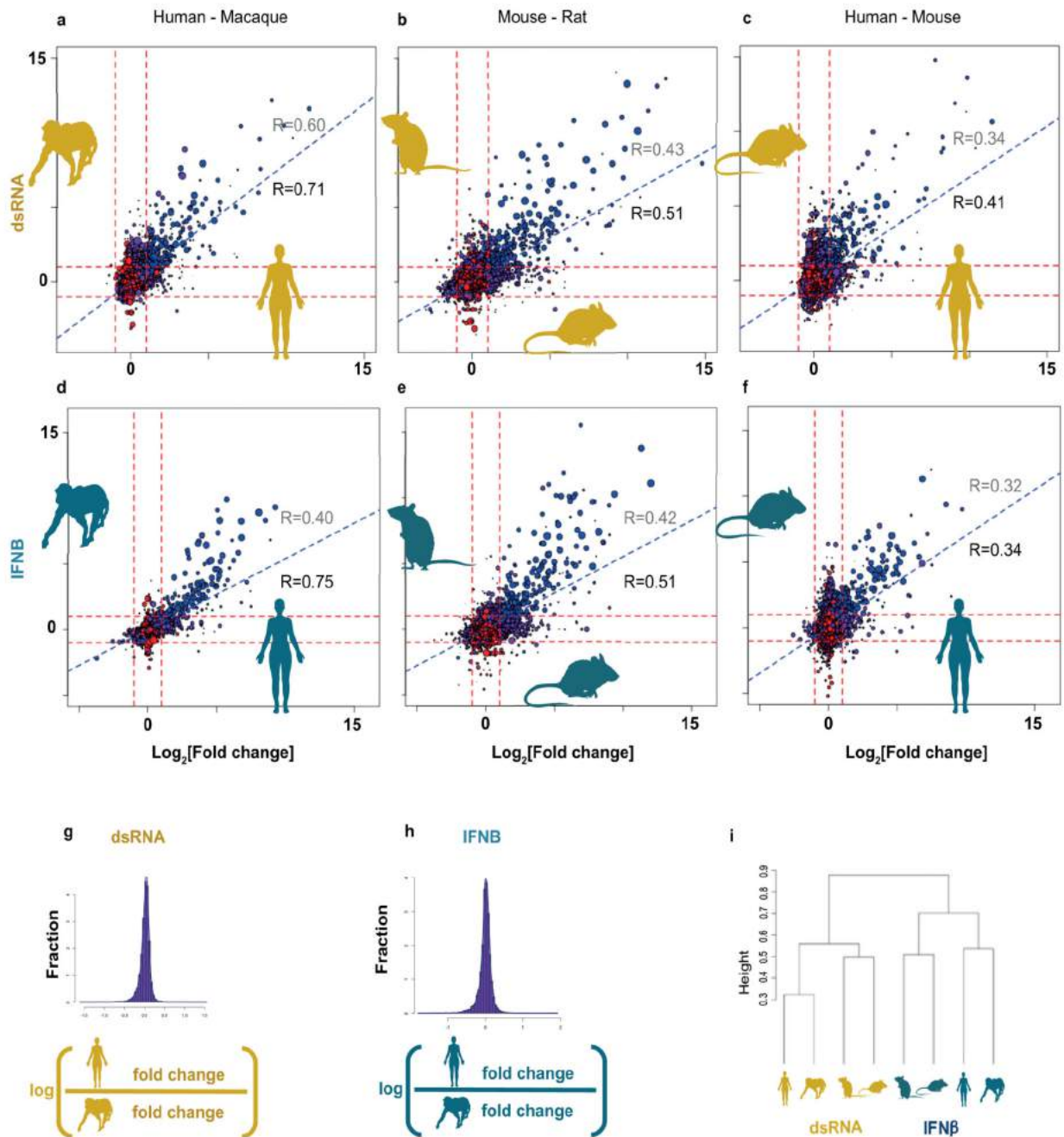
For each gene, the fold change in the outer group (pig), is subtracted from the fold change in the orthologs of the three Glires (mouse, rat and rabbit), and the average of the square values of these subtractions is taken as the response divergence measure. In most of the analyses, we focus on the 2,336 “LPS-responsive genes” – genes that are differentially expressed in response to LPS (genes that have an FDR-corrected p-value < 0.01 in mouse, and have annotated one-to-one orthologs in the other three species).

Promoter elements, gene function and cell-to-cell variability analyses: Promoter elements (TATA and CGIs), gene function and cell-to-cell variability analyses were performed as described in the fibroblasts section. Mouse genes were used as the reference for gene function and TSS annotations. For variability analysis, we used one representative replicate out of three.

Statistical analysis and reproducibility: Statistical analyses were done with R version 3.3.2 for Fisher's exact test, Two-sample Kolmogorov-Smirnov test and Mann-Whitney test. Data in boxplots (in all relevant panels) represent the median, first quartile and third quartile with lines extending to the furthest value within 1.5 of the inter-quartile range (as implemented by the R function `geom_boxplot`). Violin plots show the kernel probability density of the data (as implemented by the R function `geom_violin`).

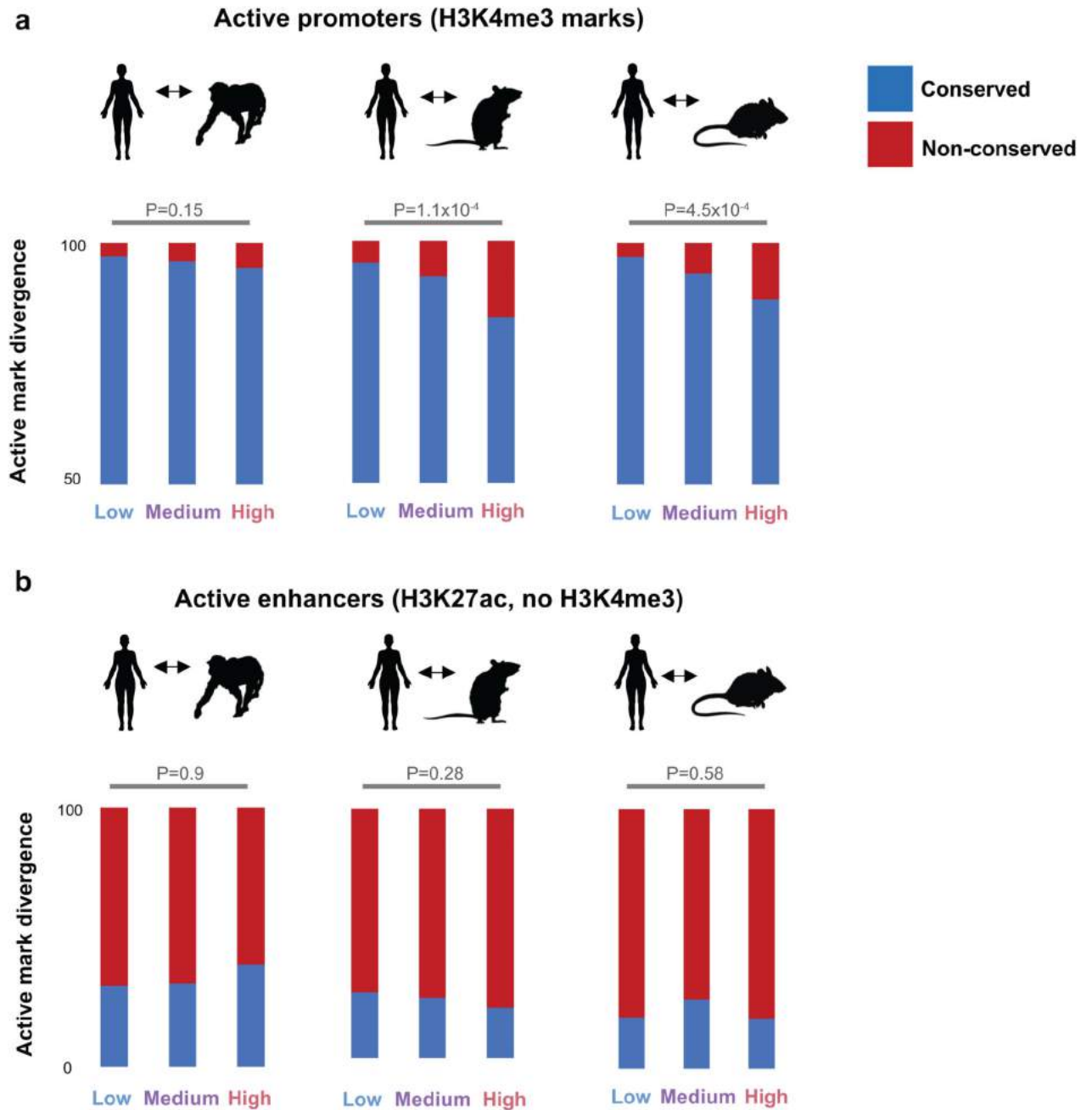
All cross-species bulk RNA-Seq replicates were successful, except for one macaque individual where the treated sample had a low RNA quality and was removed from the analysis (along with the matching control). All cross-species ChIP-Seq replicates were successful. Cross-species scRNA-seq of fibroblasts was performed in two biological replicates. Results throughout the manuscript relate to the second cross-species biological replicate where a higher proportion of cells passed technical QC. Three out of three replicates for each species and condition were successful when preparing single-cell libraries for mononuclear phagocytes, except for two libraries that failed at the emulsion preparation stage. Two out of two replicates of single-cell *in situ* RNA hybridization assay were performed and both are shown.

Extended Data

**Extended Data Figure 1. Fibroblast response to dsRNA and IFN β across species**

To study the similarity in response to treatment across species, we plotted the fold change values of all expressed genes (with one-to-one orthologs) between pairs of species (human-macaque, mouse-rat and human-mouse) in response to dsRNA (poly I:C) (a-c). As a control, we performed the same procedure with IFN β stimulations (d-f). Fold changes were inferred from differential expression analyses, determined by the exact test in the edgeR package⁶

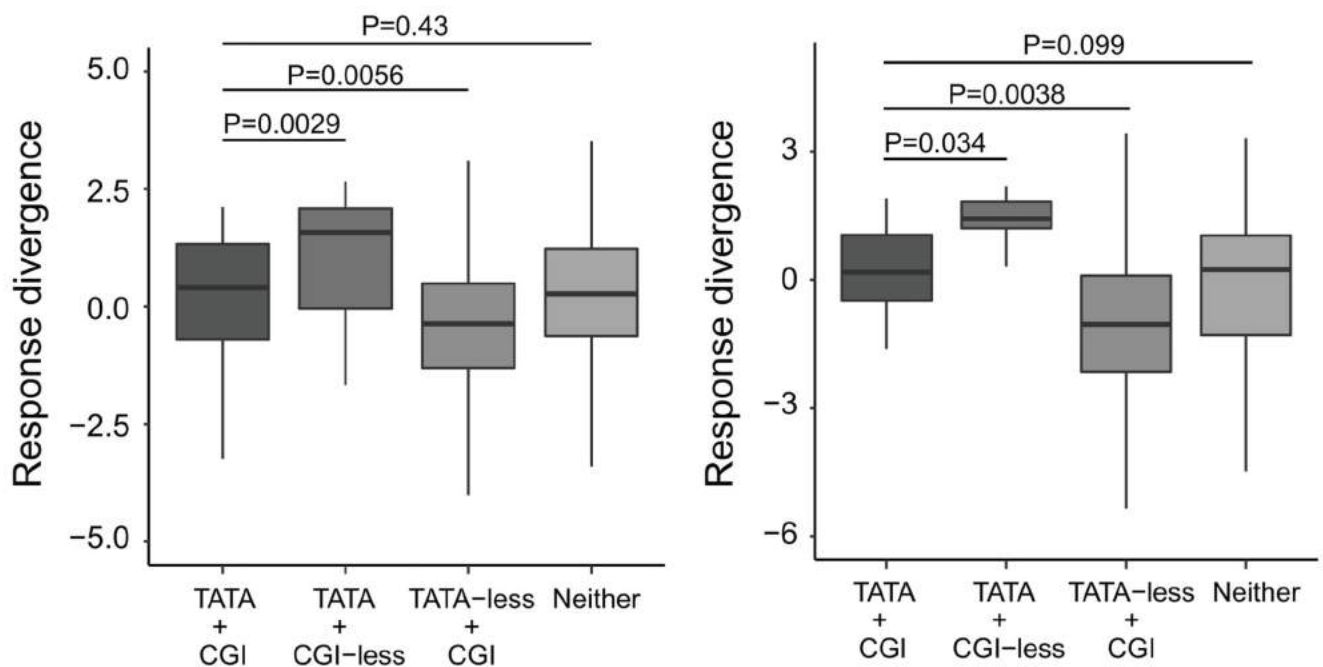
and based on n=6, 5, 3, 3 individuals from human, macaque, rat and mouse, respectively. Spearman correlations between all expressed one-to-one orthologs are shown in grey, Spearman correlations between the subset of differentially expressed genes (FDR-corrected p-value<0.01 in at least one species) appear in black. Number of genes shown is n=11,035, 11,005, 11,137, 10,851, 10,826, 10,957 in a-f, respectively. Genes are coloured in blue if they were differentially expressed (FDR-corrected p-value <0.01) in both species, in purple if they were differentially expressed in only one species, or in red if they were not differentially expressed. **(g-h)** Density plots of ratio of fold change in response to dsRNA or to IFNB: **(g)** comparison between human and macaque orthologs in dsRNA response; **(h)** comparison between human and mouse orthologs in IFNB response. **(i)** A dendrogram based on the fold change in response to dsRNA or to IFNB across 9,835 one-to-one orthologs in human, macaque, rat and mouse.



Extended Data Figure 2. Correspondence of transcriptional divergence and divergence of active promoters and enhancers

A comparison of divergence in transcriptional response to dsRNA with divergence of active chromatin marks– (a) active promoters (profiled using H3K4me3 in proximity to gene’s TSS) and (b) enhancers (H3K27ac without overlapping H3K4me3). Chromatin marks were linked to genes based on their proximity to the gene’s transcriptional start site (TSS). Chromatin marks were obtained from $n=3$ individuals in each of the four species, from fibroblasts stimulated with dsRNA or left untreated. The statistics are based on $n=855, 818$,

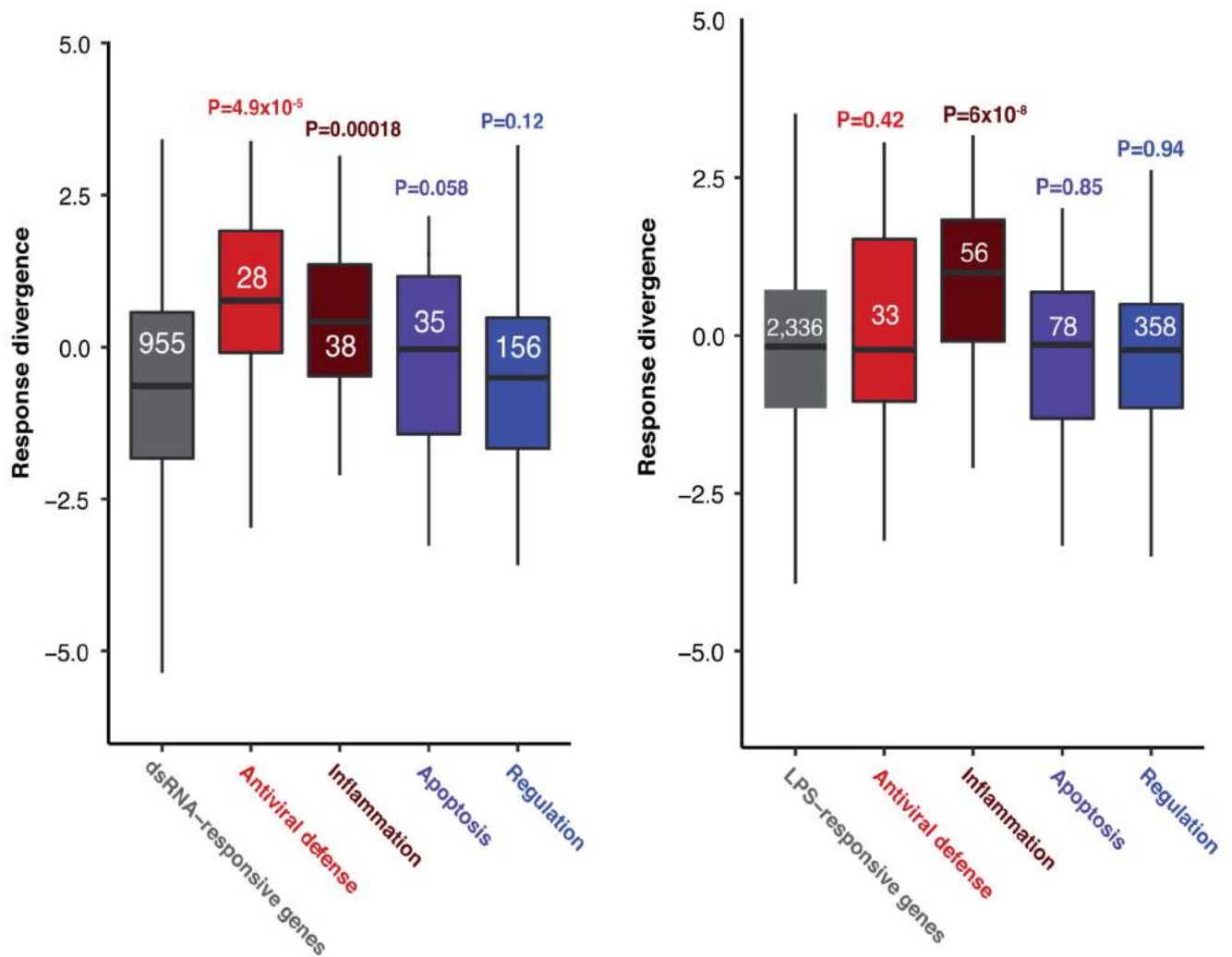
813 human genes that have a linked H3K4me3 mark with a syntenic region in macaque, rat and mouse, respectively (a); and on n=326, 241, 242 human genes that have a linked H3K27ac mark with a syntenic region in macaque, rat and mouse, respectively (b). In each panel, we show the fraction of conserved marks between human and macaque, rat or mouse, in genes that have high, medium and low divergence in their transcriptional response. In each column, the histone mark's signal was compared between human and the syntenic region in one of the three other species. If an active mark was found in the corresponding syntenic region, the linked gene was considered to have a conserved active mark (promoter or enhancer, depending on the tested mark). The fractions of genes with conserved promoters (or enhancers) in each pair of species were compared between highly and lowly divergent genes using a one-sided Fisher's exact test. When comparing active promoter regions of highly *versus* lowly divergent genes, we observe that lowly divergent genes have a significantly higher fraction of conserved marks in rodents. This suggests an agreement between divergence at the transcriptional and the chromatin levels in active promoter regions. In active enhancer regions, we do not observe these patterns, suggesting that the major contribution to divergence comes from the promoters.



Extended Data Figure 3. Comparison of response divergence of genes containing various promoter elements

Comparing response divergence between genes with and without a TATA-box and CGI: Left – fibroblasts (n=14, 14, 633, 294 differentially expressed genes with only TATA-box element, with both CGI and TATA-box elements, with only CGI, and with neither element in their promoters, respectively); right – phagocytes (n=13, 29, 1,718, 576 differentially expressed genes with only a TATA-box element, with both CGI and TATA-box elements, with only a CGI, and with neither element in their promoters, respectively). Genes with

TATA without a CGI have higher response divergence than genes with both elements. Genes with a CGI but without TATA diverge more slowly than genes with both elements. Genes with both elements do not significantly differ in their divergence than genes lacking both elements (one-sided Mann-Whitney test). 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.

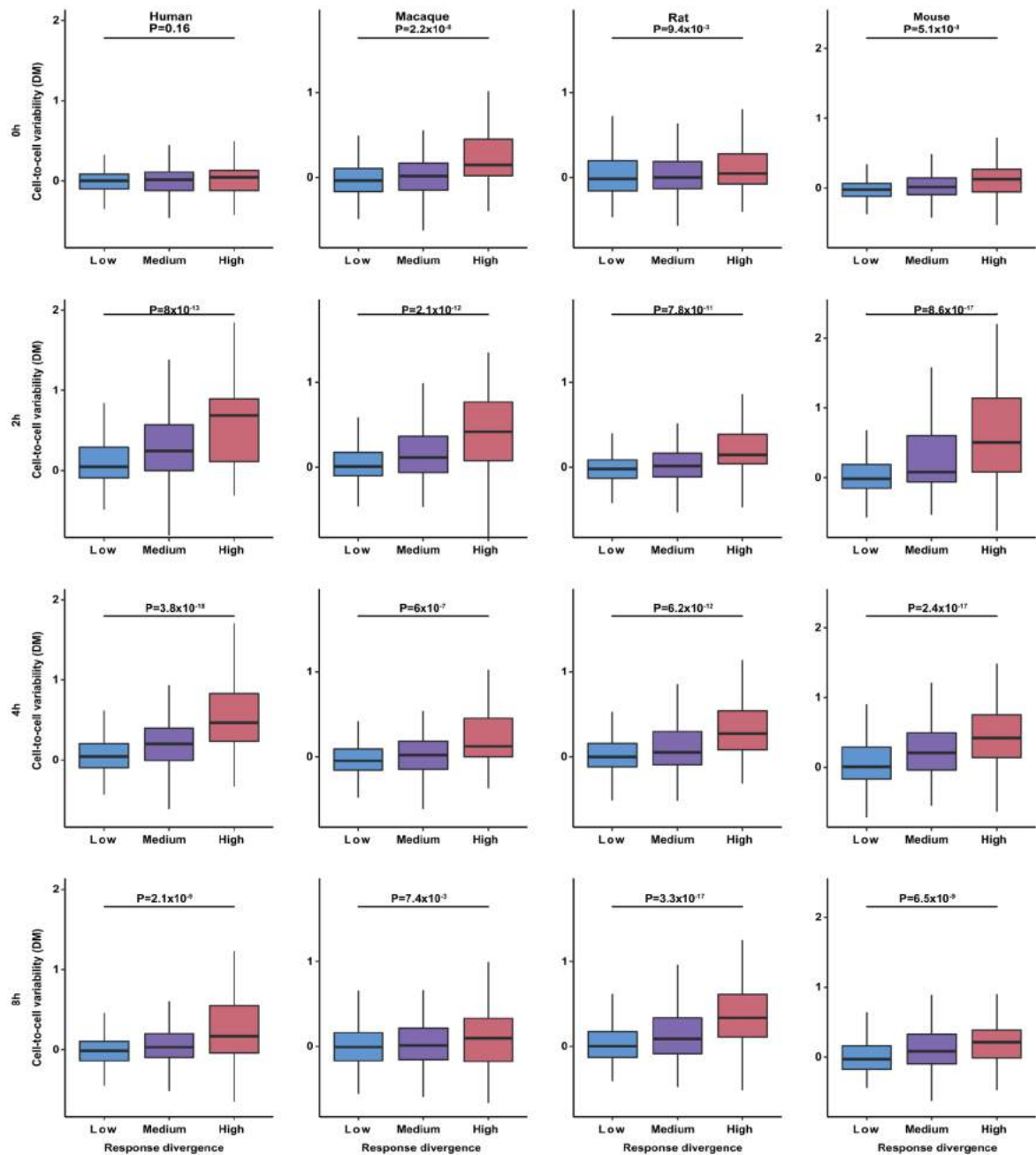


Extended Data Figure 4. Response divergence of molecular processes upregulated in immune response

(a) Distributions of divergence values of $n=955$ dsRNA-responsive genes in fibroblasts and subsets of this group belonging to different biological processes. For each functional subset, the distribution of divergence values is compared with the set of 955 dsRNA-responsive genes using a one-sided Mann-Whitney test. FDR-corrected p-values are shown above each group, group size is shown inside each box.

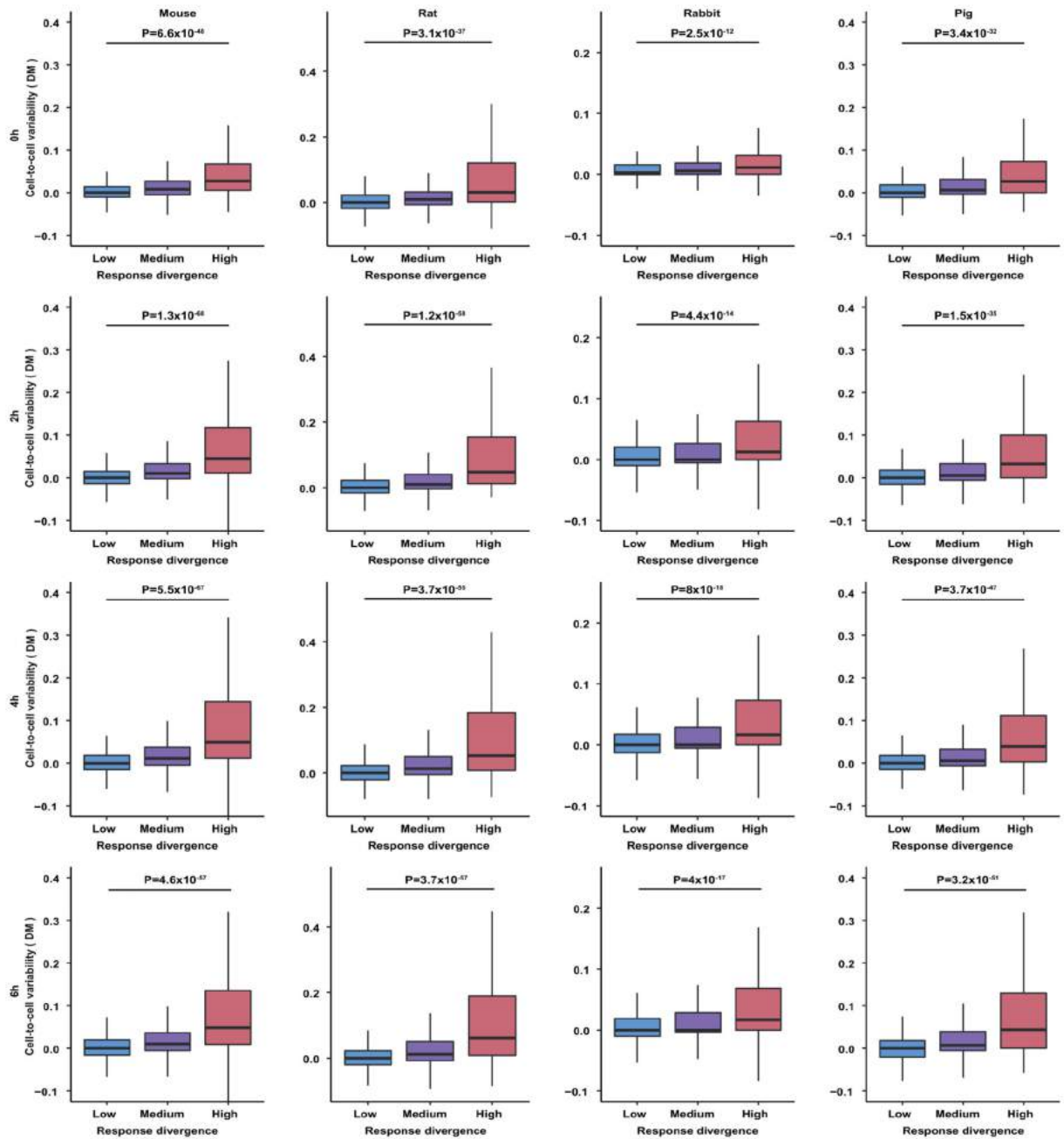
(b) Distributions of divergence values of $n=2,336$ LPS-responsive genes in mononuclear phagocytes and subsets of this group belonging to different biological processes. For each

functional subset, the distribution of divergence values is compared with the set of 2,336 LPS-responsive genes. FDR-corrected p-values (one-sided Mann-Whitney test) are shown above each group, group size is shown inside each box. Data in boxplots represent the median, first quartile and third quartile with lines extending to the furthest value within 1.5 of the inter-quartile range.



Extended Data Figure 5. Cell-to-cell variability *versus* response divergence across species and conditions in fibroblasts dsRNA stimulation

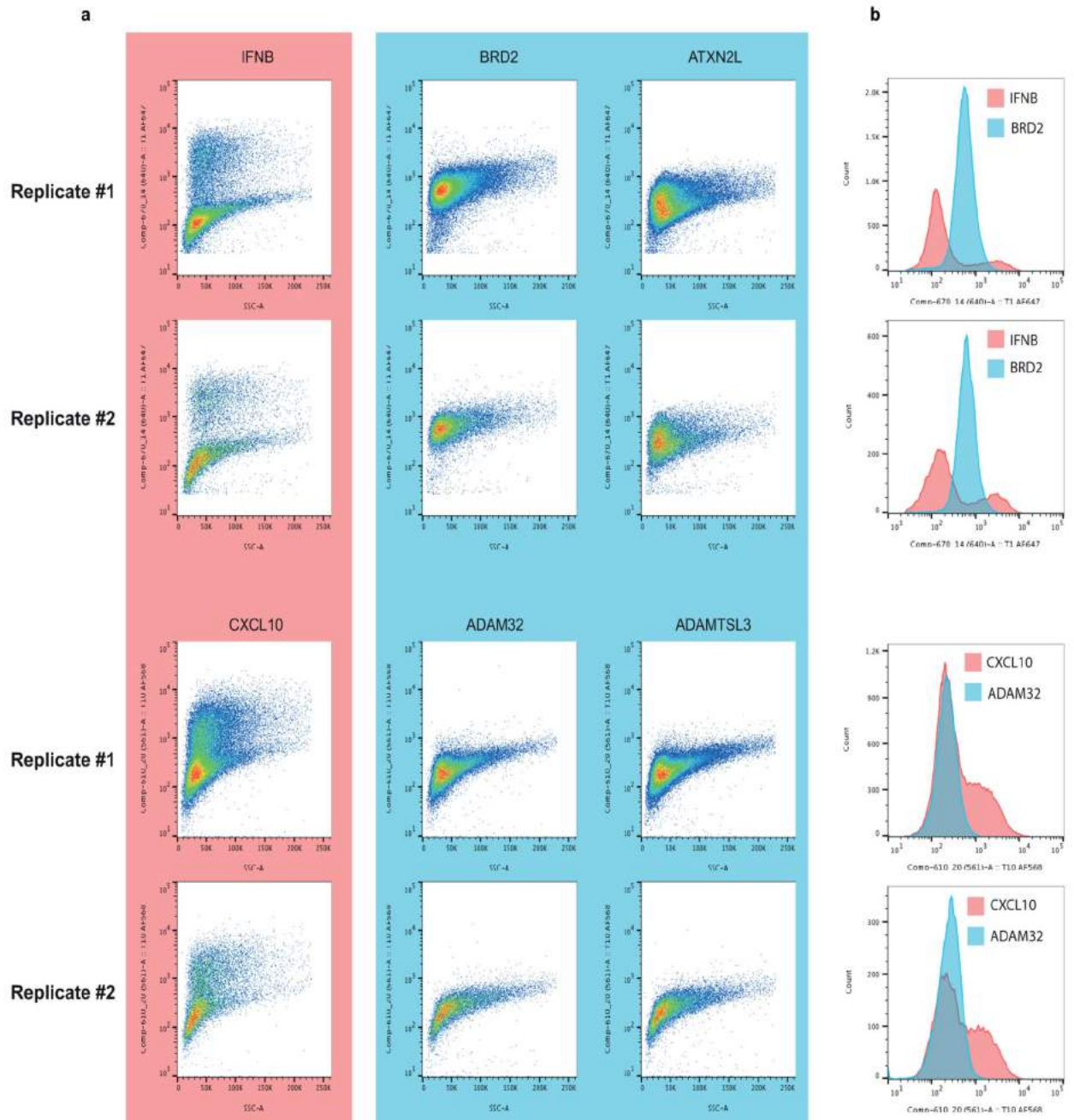
Cell-to-cell variability values, as measured with DM across individual cells, in comparison with response divergence between species (grouped into low, medium and high divergence). Variability values are based on n=29, 56, 55, 35 human cells, n=20, 32, 29, 13 rhesus cells, n=33, 70, 65, 40 rat cells, and n= 53, 81, 59, 30 mouse cells, stimulated with dsRNA for 0, 2, 4 and 8 hours, respectively. Rows represent different dsRNA stimulation time points (0, 2, 4 and 8 hours), and columns represent different species: human, macaque, mouse and rat. Highly divergent genes were compared with lowly divergent genes using a one-sided Mann-Whitney test. Data in boxplots represent the median, first quartile and third quartile with lines extending to the furthest value within 1.5 of the inter-quartile range.



Extended Data Figure 6. Cell-to-cell variability *versus* response divergence across species and conditions in mononuclear phagocytes LPS stimulation

Cell-to-cell variability values, as measured with DM across cells, in comparison with response divergence between species (grouped into low, medium and high divergence). Variability values are based on $n=3,519, 4,321, 3,293, 2,126$ mouse cells, $n= 2,266, 2,839, 1,963, 1,607$ rat cells, $n= 3,275, 1,820, 1,522, 1,660$ rabbit cells, and $n= 1,748, 1,614, 1,899, 1,381$ pig cells, stimulated with LPS for 0, 2, 4 and 6 hours, respectively. Rows represent different LPS stimulation time points (0, 2, 4 and 6 hours), and columns represent different species: mouse, rat, rabbit and pig. Highly divergent genes were compared with lowly

divergent genes using a one-sided Mann-Whitney test. Data in boxplots represent the median, first quartile and third quartile with lines extending to the furthest value within 1.5 of the inter-quartile range.



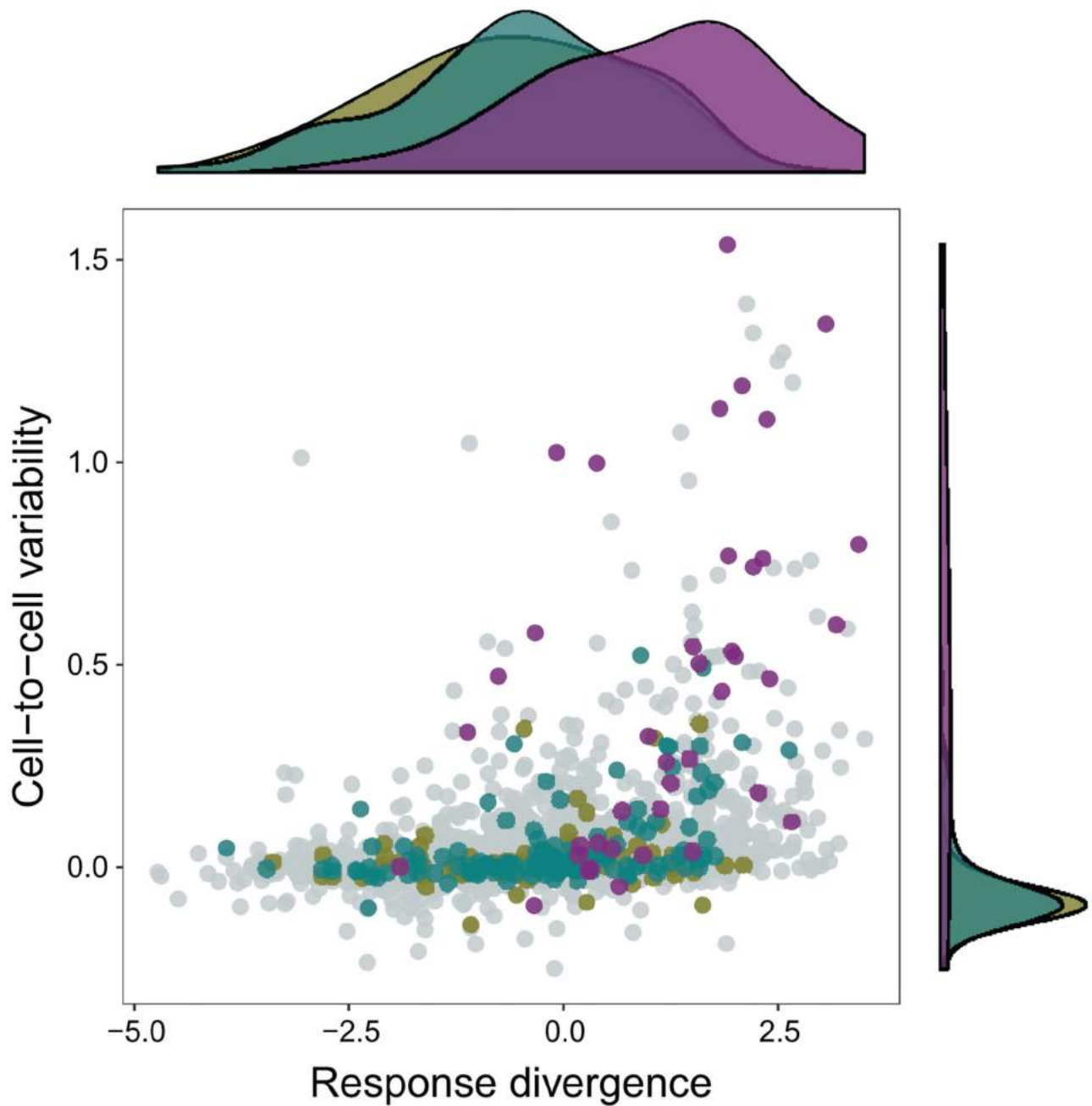
Extended Data Figure 7. Cell-to-cell variability of cytokine expression in single cell *in situ* RNA hybridization assay combined with flow cytometry (PrimeFlow)

PrimeFlow measurement of two cytokine genes displaying high cell-to-cell variability in single-cell RNA-seq – IFNB and CXCL10. As controls, two genes matched on expression levels (ATXN2L and ADAM32) but displaying low cell-to-cell variability in single-cell

RNA-seq data are shown. Since the expression level of cytokines is at the low end of the distribution, we also chose two genes with middle-range expression values (ADAMTSL3 and BRD2) as additional controls. The experiment was performed in n=2 independent replicates, originating from the same individual. Both replicates are shown.

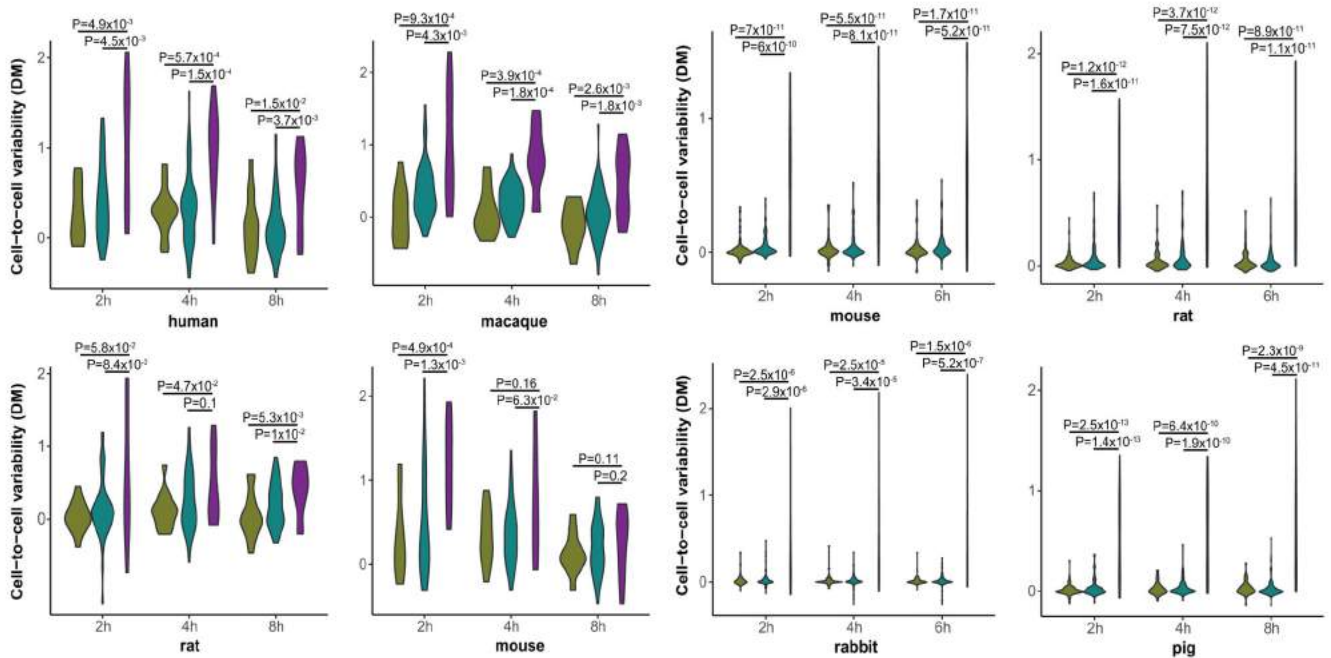
(a) Pseudocolor contour plot for RNA target expression in dsRNA-stimulated human fibroblasts. X-axis shows area of side scatter (SSC-A) and Y-axis shows fluorescent signal for target RNA probes. RNA targets detected by same fluorescent channel are displayed together. Upper panel: IFNB and control genes: BRD2, ATXN2L, Type 1 probe, Alexa Fluor™ 647. Lower panel: CXCL10, and controls: ADAMTSL3, ADAM32, Type 10 probe, Alexa Fluor™ 568. The cytokines display a broader range of fluorescence signal than the controls.

(b) Histograms comparing fluorescence of cytokine and control pairs (IFNB-BRD2 for Type 1 probe and CXCL10-ADAM32 for Type 10 probe). The histograms show a bimodal distribution of expression signal for the two cytokines (IFNB, CXCL10, in red), but not for controls (in blue). This agrees with single-cell RNA-seq data where CXCL10 and IFNB display high levels of cell-to-cell variability.



Extended Data Figure 8. Cell-to-cell variability levels and response divergence of cytokines, transcription factors and kinases in LPS stimulation of phagocytes

A scatter plot showing divergence in response to LPS across species and transcriptional cell-to-cell variability in mouse mononuclear phagocytes following 4H of LPS treatment, in $n=2,262$ LPS-response genes. Genes from three functional groups - cytokines, transcription factors and kinases - are coloured in purple, green and beige, respectively. The distributions of divergence values of each of the three functional groups are shown above the scatter plot (with colours matching the scatter plot). The distributions of cell-to-cell variability values of these groups are shown to the right.

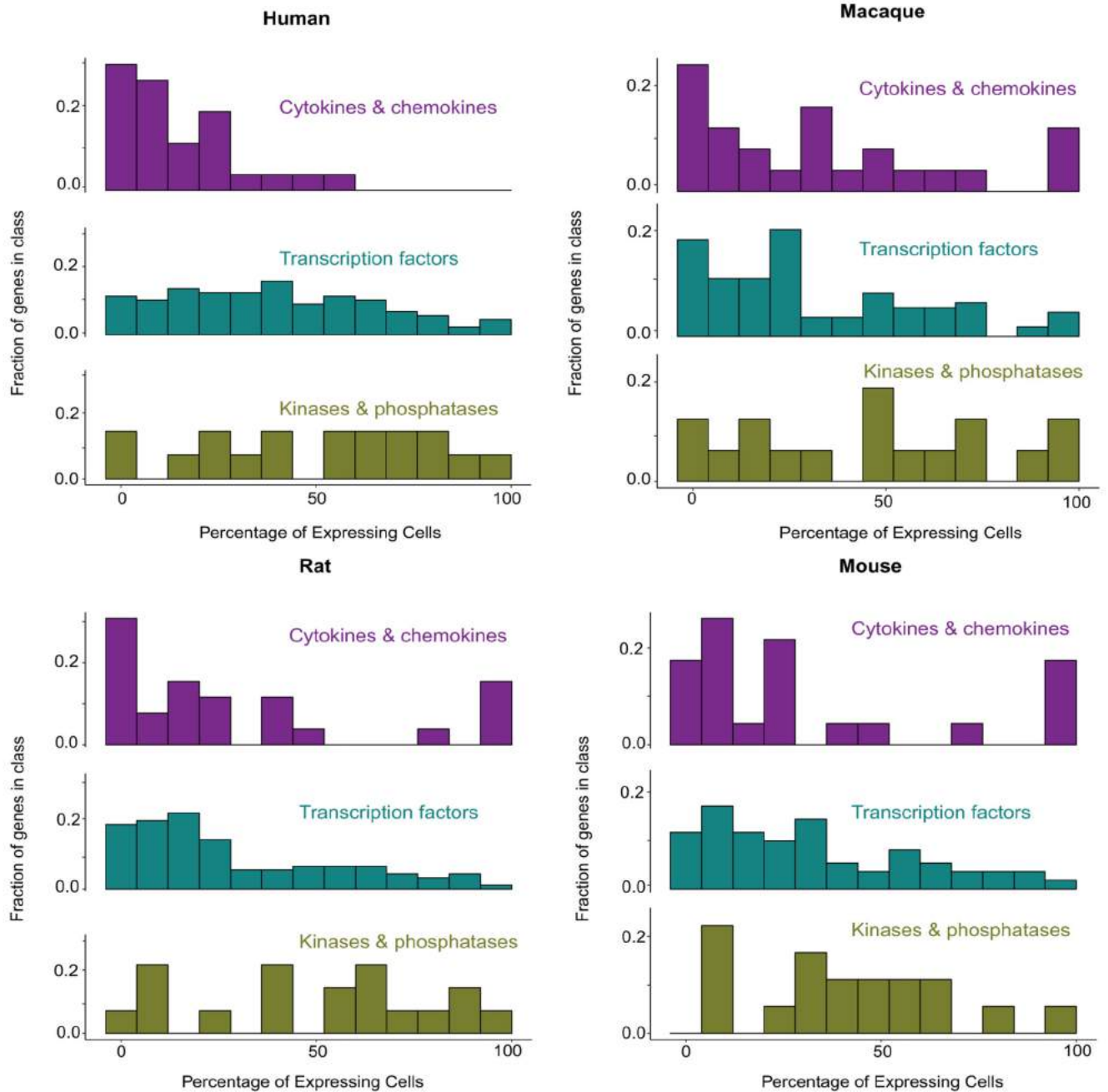


Extended Data Figure 9. Cell-to-cell variability levels in cytokines, transcription factors and kinases across species and stimulation time points

Violin plots showing the distribution of cell-to-cell variability values (DM) of cytokines, transcription factors and kinases during immune stimulation.

Left: Fibroblast dsRNA stimulation time course. Number of cells used in each species (at 2, 4, 8h dsRNA, respectively): human – 56, 55, 35; macaque – 32, 29, 13; rat – 70, 65, 40; mouse – 81, 59, 30. Right: Phagocyte LPS stimulation time course. Number of cells used in each species (at 2, 4, 6h LPS, respectively): mouse – 4321, 3293, 2126; rat – 2839, 1963, 1607; rabbit – 1820, 1522, 1660; pig – 1614, 1899, 1381.

For both panels, colours as in Fig 3c. Comparisons between groups of genes were performed using one-sided Mann-Whitney tests. Violin plots show the kernel probability density of the data.



Extended Data Figure 10. Percentage of expressing cells in cytokines, transcription factors and kinases

Histograms showing the percentage of fibroblasts expressing cytokines (top), transcription factors (middle) and kinases (bottom) following 4H dsRNA stimulation, in human, macaque, rat and mouse cells (based on $n=55, 29, 65, 59$ cells, respectively). The percentage of expressing cells is divided into 13 bins (x-axis). The y-axis represents the fraction of genes from this gene class (e.g. cytokines) that are expressed in each bin (e.g. in human, nearly 30% of the cytokines genes (y-axis) are expressed in the first bin, corresponding to expression in fewer than 8% of the cells).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Nils Eling, Matteo Fumagalli, Yoav Gilad, Orly Laufman, Amir Marcovitz, John Marioni, Kerstin Meyer, Matthieu Muffato, Duncan Odom, Oliver Stegle, Adi Stern, Mike Stubbington, Valentine Svensson and Michelle Ward for helpful discussions; Guy Emerton, Ayesha Jinat, Lira Mamanova, Krzysztof Polanski, Anja Fullgrabe, Nancy George, Sam Barnett, Rachel Boyd, Sharad Patel and Celine Gomez for technical assistance, the Hipsci consortium for human fibroblast lines and members of the Teichmann lab for support in various stages. This project was supported by ERC grants (ThDEFINE, ThSWITCH) and an EU FET-OPEN grant (MRG-GRAMMAR). T.H was supported by an HFSP Long-Term Fellowship and by an EMBO Long-Term and Advanced fellowships. VP is funded by Fondazione Umberto Veronesi.

References

1. Borden EC, et al. Interferons at age 50: past, current and future impact on biomedicine. *Nat Rev Drug Discov.* 2007; 6:975–990. DOI: 10.1038/nrd2422 [PubMed: 18049472]
2. Iwasaki A. A virological view of innate immune recognition. *Annu Rev Microbiol.* 2012; 66:177–196. DOI: 10.1146/annurev-micro-092611-150203 [PubMed: 22994491]
3. Nielsen R, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 2005; 3:e170. doi: 10.1371/journal.pbio.0030170 [PubMed: 15869325]
4. Haygood R, Babbitt CC, Fedrigo O, Wray GA. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc Natl Acad Sci U S A.* 2010; 107:7853–7857. DOI: 10.1073/pnas.0911249107 [PubMed: 20385805]
5. Fumagalli M, et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* 2011; 7:e1002355. doi: 10.1371/journal.pgen.1002355 [PubMed: 22072984]
6. Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. *Elife.* 2016; 5:doi: 10.7554/eLife.12469
7. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 2010; 11:17–30. DOI: 10.1038/nrg2698 [PubMed: 19953080]
8. Zhao M, Zhang J, Phatnani H, Scheu S, Maniatis T. Stochastic expression of the interferon-beta gene. *PLoS Biol.* 2012; 10:e1001249. doi: 10.1371/journal.pbio.1001249 [PubMed: 22291574]
9. Avraham R, et al. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell.* 2015; 162:1309–1321. DOI: 10.1016/j.cell.2015.08.027 [PubMed: 26343579]
10. Shalek AK, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature.* 2014; 510:363–369. DOI: 10.1038/nature13437 [PubMed: 24919153]
11. Hwang SY, et al. Biphasic RLR-IFN-beta response controls the balance between antiviral immunity and cell damage. *J Immunol.* 2013; 190:1192–1200. DOI: 10.4049/jimmunol.1202326 [PubMed: 23284052]
12. Porritt RA, Hertzog PJ. Dynamic control of type I IFN signalling by an integrated network of negative regulators. *Trends Immunol.* 2015; 36:150–160. DOI: 10.1016/j.it.2015.02.002 [PubMed: 25725583]
13. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nat Rev Immunol.* 2014; 14:36–49. DOI: 10.1038/nri3581 [PubMed: 24362405]
14. Brinkworth JF, Barreiro LB. The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. *Curr Opin Immunol.* 2014; 31:66–78. DOI: 10.1016/j.coi.2014.09.008 [PubMed: 25458997]
15. Kobayashi KS, Flavell RA. Shielding the double-edged sword: negative regulation of the innate immune system. *J Leukoc Biol.* 2004; 75:428–433. DOI: 10.1189/jlb.0703321 [PubMed: 14597727]

16. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. *Int Rev Immunol*. 2011; 30:16–34. DOI: 10.3109/08830185.2010.529976 [PubMed: 21235323]
17. Barreiro LB, Marioni JC, Blekhman R, Stephens M, Gilad Y. Functional comparison of innate immune signaling pathways in primates. *PLoS Genet*. 2010; 6:e1001249. doi: 10.1371/journal.pgen.1001249 [PubMed: 21187902]
18. Schroder K, 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:E944–953. DOI: 10.1073/pnas.1110156109 [PubMed: 22451944]
19. Shay T, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci U S A*. 2013; 110:2946–2951. DOI: 10.1073/pnas.1222738110 [PubMed: 23382184]
20. Brawand D, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011; 478:343–348. DOI: 10.1038/nature10532 [PubMed: 22012392]
21. Kalinka AT, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 2010; 468:811–814. DOI: 10.1038/nature09634 [PubMed: 21150996]
22. Khaitovich P, Enard W, Lachmann M, Paabo S. Evolution of primate gene expression. *Nat Rev Genet*. 2006; 7:693–702. DOI: 10.1038/nrg1940 [PubMed: 16921347]
23. Levin M, et al. The mid-developmental transition and the evolution of animal body plans. *Nature*. 2016; 531:637–641. DOI: 10.1038/nature16994 [PubMed: 26886793]
24. Reilly SK, Noonan JP. Evolution of Gene Regulation in Humans. *Annu Rev Genomics Hum Genet*. 2016; 17:45–67. DOI: 10.1146/annurev-genom-090314-045935 [PubMed: 27147089]
25. Tirosh I, Weinberger A, Carmi M, Barkai N. A genetic signature of interspecies variations in gene expression. *Nat Genet*. 2006; 38:830–834. DOI: 10.1038/ng1819 [PubMed: 16783381]
26. Haberle V, Lenhard B. Promoter architectures and developmental gene regulation. *Semin Cell Dev Biol*. 2016; 57:11–23. DOI: 10.1016/j.semcdb.2016.01.014 [PubMed: 26783721]
27. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*. 2012; 13:233–245. DOI: 10.1038/nrg3163 [PubMed: 22392219]
28. Franz KM, Kagan JC. Innate Immune Receptors as Competitive Determinants of Cell Fate. *Mol Cell*. 2017; 66:750–760. DOI: 10.1016/j.molcel.2017.05.009 [PubMed: 28622520]
29. Satija R, Shalek AK. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol*. 2014; 35:219–229. DOI: 10.1016/j.it.2014.03.004 [PubMed: 24746883]
30. Newman JR, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006; 441:840–846. DOI: 10.1038/nature04785 [PubMed: 16699522]
31. Faure AJ, Schmiedel JM, Lehner B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst*. 2017; 5:471–484 e474. DOI: 10.1016/j.cels.2017.10.003 [PubMed: 29102610]
32. Rand U, et al. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Mol Syst Biol*. 2012; 8:584. doi: 10.1038/msb.2012.17 [PubMed: 22617958]
33. Fumagalli M, Sironi M. Human genome variability, natural selection and infectious diseases. *Curr Opin Immunol*. 2014; 30:9–16. DOI: 10.1016/j.coi.2014.05.001 [PubMed: 24880709]
34. Johnson WE, Sawyer SL. Molecular evolution of the antiretroviral TRIM5 gene. *Immunogenetics*. 2009; 61:163–176. DOI: 10.1007/s00251-009-0358-y [PubMed: 19238338]
35. Choo SW, et al. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Res*. 2016; 26:1312–1322. DOI: 10.1101/gr.203521.115 [PubMed: 27510566]
36. Braun BA, Marcovitz A, Camp JG, Jia R, Bejerano G. Mx1 and Mx2 key antiviral proteins are surprisingly lost in toothed whales. *Proc Natl Acad Sci U S A*. 2015; 112:8036–8040. DOI: 10.1073/pnas.1501844112 [PubMed: 26080416]
37. Xu L, et al. Loss of RIG-I leads to a functional replacement with MDA5 in the Chinese tree shrew. *Proc Natl Acad Sci U S A*. 2016; 113:10950–10955. DOI: 10.1073/pnas.1604939113 [PubMed: 27621475]
38. Sackton TB, Lazzaro BP, Clark AG. Rapid expansion of immune-related gene families in the house fly, *Musca domestica*. *Mol Biol Evol*. 2017; doi: 10.1093/molbev/msw285

39. Brunette RL, et al. Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. *J Exp Med.* 2012; 209:1969–1983. DOI: 10.1084/jem.20121960 [PubMed: 23045604]
40. Malfavon-Borja R, Wu LI, Emerman M, Malik HS. Birth, decay, and reconstruction of an ancient TRIMCyp gene fusion in primate genomes. *Proc Natl Acad Sci U S A.* 2013; 110:E583–592. DOI: 10.1073/pnas.1216542110 [PubMed: 23319649]
41. Barber MF, Lee EM, Griffin H, Elde NC. Rapid Evolution of Primate Type 2 Immune Response Factors Linked to Asthma Susceptibility. *Genome Biol Evol.* 2017; 9:1757–1765. DOI: 10.1093/gbe/evx120 [PubMed: 28854632]
42. Saeed R, Deane CM. Protein protein interactions, evolutionary rate, abundance and age. *BMC Bioinformatics.* 2006; 7:128.doi: 10.1186/1471-2105-7-128 [PubMed: 16533385]
43. Calderone A, Licata L, Cesareni G. VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.* 2015; 43:D588–592. DOI: 10.1093/nar/gku830 [PubMed: 25217587]
44. Halehalli RR, Nagarajaram HA. Molecular principles of human virus protein-protein interactions. *Bioinformatics.* 2015; 31:1025–1033. DOI: 10.1093/bioinformatics/btu763 [PubMed: 25417202]
45. Pichlmair A, et al. Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature.* 2012; 487:486–490. DOI: 10.1038/nature11289 [PubMed: 22810585]
46. Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.* 2008; 4:e32.doi: 10.1371/journal.ppat.0040032 [PubMed: 18282095]
47. Tirosch I, Barkai N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 2008; 18:1084–1091. DOI: 10.1101/gr.076059.108 [PubMed: 18448704]
48. Crow YJ, Manel N. Aicardi-Goutieres syndrome and the type I interferonopathies. *Nat Rev Immunol.* 2015; 15:429–440. DOI: 10.1038/nri3850 [PubMed: 26052098]
49. Hall JC, Rosen A. Type I interferons: crucial participants in disease amplification in autoimmunity. *Nat Rev Rheumatol.* 2010; 6:40–49. DOI: 10.1038/nrrheum.2009.237 [PubMed: 20046205]
50. Tisoncik JR, et al. Into the eye of the cytokine storm. *Microbiol Mol Biol Rev.* 2012; 76:16–32. DOI: 10.1128/MMBR.05015-11 [PubMed: 22390970]
51. Kilpinen H, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature.* 2017; 546:370–375. DOI: 10.1038/nature22403 [PubMed: 28489815]
52. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods.* 2015; 12:963–965. DOI: 10.1038/nmeth.3542 [PubMed: 26280331]
53. Picelli S, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* 2014; 9:171–181. DOI: 10.1038/nprot.2014.006 [PubMed: 24385147]
54. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017; doi: 10.1038/nmeth.4197
55. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. DOI: 10.1093/bioinformatics/btp616 [PubMed: 19910308]
56. Nourmohammad A, et al. Adaptive Evolution of Gene Expression in *Drosophila*. *Cell Rep.* 2017; 20:1385–1395. DOI: 10.1016/j.celrep.2017.07.033 [PubMed: 28793262]
57. Zhang HM, et al. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* 2012; 40:D144–149. DOI: 10.1093/nar/gkr965 [PubMed: 22080564]
58. Binns D, et al. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics.* 2009; 25:3045–3046. DOI: 10.1093/bioinformatics/btp536 [PubMed: 19744993]
59. Kent WJ, et al. The human genome browser at UCSC. *Genome Res.* 2002; 12:996–1006. DOI: 10.1101/gr.229102 [PubMed: 12045153]
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9:357–359. DOI: 10.1038/nmeth.1923 [PubMed: 22388286]
61. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. DOI: 10.1093/bioinformatics/btp352 [PubMed: 19505943]

62. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9:R137.doi: 10.1186/gb-2008-9-9-r137 [PubMed: 18798982]
63. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26:841–842. DOI: 10.1093/bioinformatics/btq033 [PubMed: 20110278]
64. Kuhn RM, et al. The UCSC genome browser database: update 2007. *Nucleic Acids Res.* 2007; 35:D668–673. DOI: 10.1093/nar/gkl928 [PubMed: 17142222]
65. Mathelier A, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016; 44:D110–115. DOI: 10.1093/nar/gkv1176 [PubMed: 26531826]
66. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics.* 2011; 27:1017–1018. DOI: 10.1093/bioinformatics/btr064 [PubMed: 21330290]
67. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010; 20:110–121. DOI: 10.1101/gr.097857.109 [PubMed: 19858363]
68. Kolodziejczyk AA, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell.* 2015; 17:471–485. DOI: 10.1016/j.stem.2015.09.011 [PubMed: 26431182]
69. Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLoS Comput Biol.* 2015; 11:e1004333.doi: 10.1371/journal.pcbi.1004333 [PubMed: 26107944]
70. Martinez-Jimenez CP, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science.* 2017; 355:1433–1436. DOI: 10.1126/science.aah4115 [PubMed: 28360329]
71. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* 2011; 27:431–432. DOI: 10.1093/bioinformatics/btq675 [PubMed: 21149340]
72. Lindblad-Toh K, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature.* 2011; 478:476–482. DOI: 10.1038/nature10530 [PubMed: 21993624]
73. Herrero J, et al. Ensembl comparative genomics resources. *Database (Oxford).* 2016; 2016doi: 10.1093/database/baw053
74. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006; 22:1269–1271. DOI: 10.1093/bioinformatics/btl097 [PubMed: 16543274]
75. Capra JA, Williams AG, Pollard KS. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput Biol.* 2012; 8:e1002567.doi: 10.1371/journal.pcbi.1002567 [PubMed: 22761559]
76. Szklarczyk D, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015; 43:D447–452. DOI: 10.1093/nar/gku1003 [PubMed: 25352553]
77. Zheng GX, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017; 8doi: 10.1038/ncomms14049
78. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015; 33:495–502. DOI: 10.1038/nbt.3192 [PubMed: 25867923]
79. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15:550.doi: 10.1186/s13059-014-0550-8 [PubMed: 25516281]

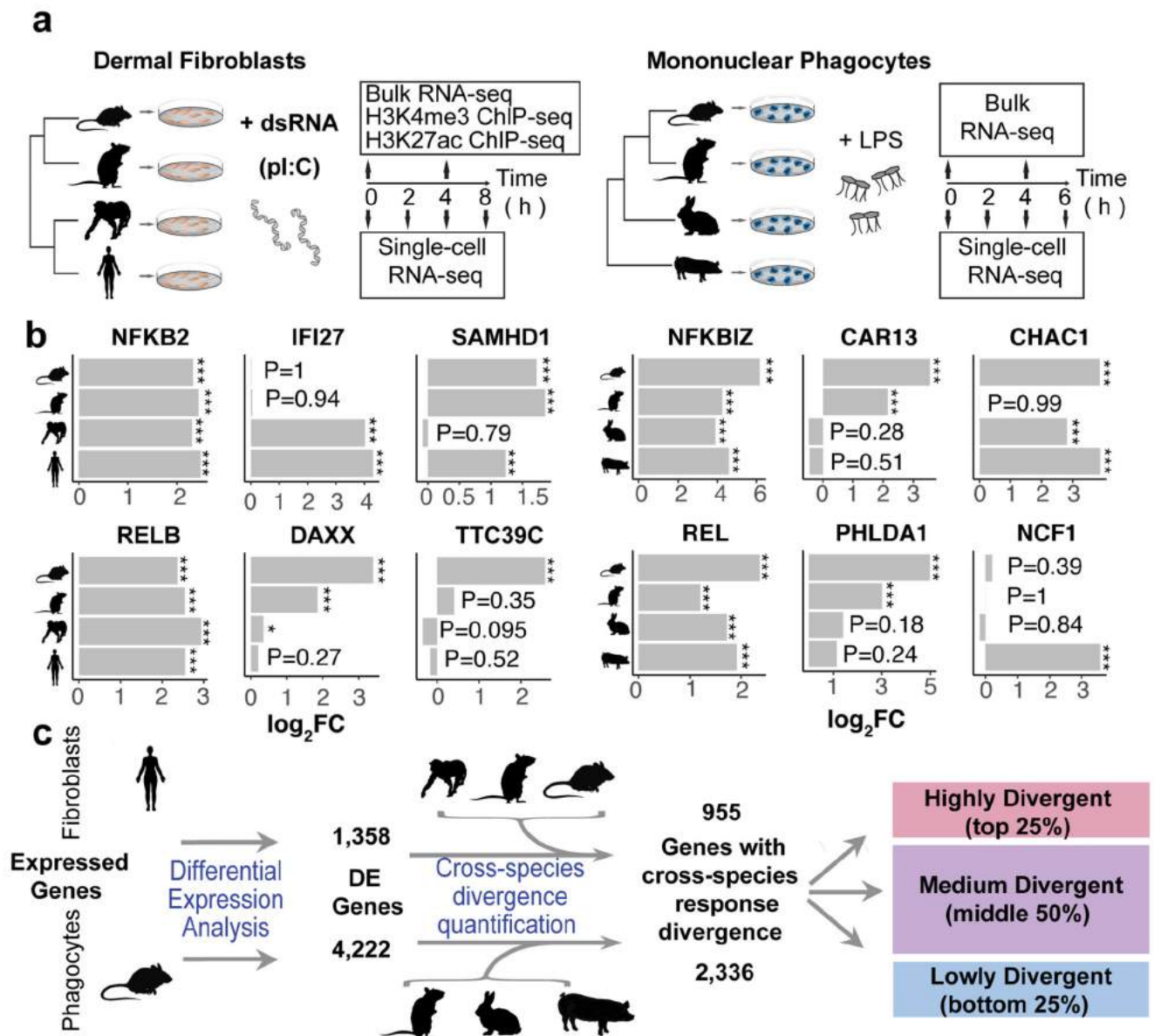


Figure 1. Response divergence across species in innate immune response

(a) Study design: Left: Primary dermal fibroblasts from mouse, rat, human and macaque - stimulated with dsRNA or controls. Samples were collected for bulk and single-cell RNA-seq and ChIP-seq. Right: Primary bone marrow-derived mononuclear phagocytes from mouse, rat, rabbit and pig - stimulated with LPS or controls. Samples were collected for bulk and single-cell RNA-seq.

(b) Left: Fold change in dsRNA stimulation in fibroblasts for example genes across the species (edgeR exact test, based on $n=6, 5, 3, 3$ individuals from human, macaque, rat and mouse, respectively). Right: Fold change in LPS stimulation for phagocytes in example genes across the species (Wald test implemented in DESeq2, based on $n=3$ individuals from each species). FDR-corrected p-values are shown (***) denotes $p\text{-value} < 0.001$, ** $p\text{-value} < 0.01$, * $p\text{-value} < 0.05$.)

(c) Top: Estimating each gene's level of cross-species divergence in transcriptional response to dsRNA stimulation in fibroblasts: (1) Using differential expression analysis, fold change in dsRNA response was assessed for each gene in each species. (2) 1,358 human genes were identified as differentially expressed (FDR-corrected q-value<0.01), of which 955 have one-to-one orthologs across the four studied species. For each gene with one-to-one orthologs across all species, a response divergence measure was estimated using:

$$\text{Response divergence} = \log^{1/4} \times \sum_{i,j} (\log\text{Fold_Change Primate}_i - \log\text{Fold_Change Rodent}_j)^2$$

(3) Genes were grouped into 'low', 'medium' and 'high' divergence based on their relative response divergence values for subsequent analysis.

Bottom: Estimating each gene's level of cross-species divergence in LPS response in mononuclear phagocytes. A response divergence measure was estimated using:

$$\text{Response divergence} = \log^{1/3} \times \sum_j (\log\text{Fold_Change Pig} - \log\text{Fold_Change Glire}_j)^2$$

(Where Glires are mouse, rat and rabbit).

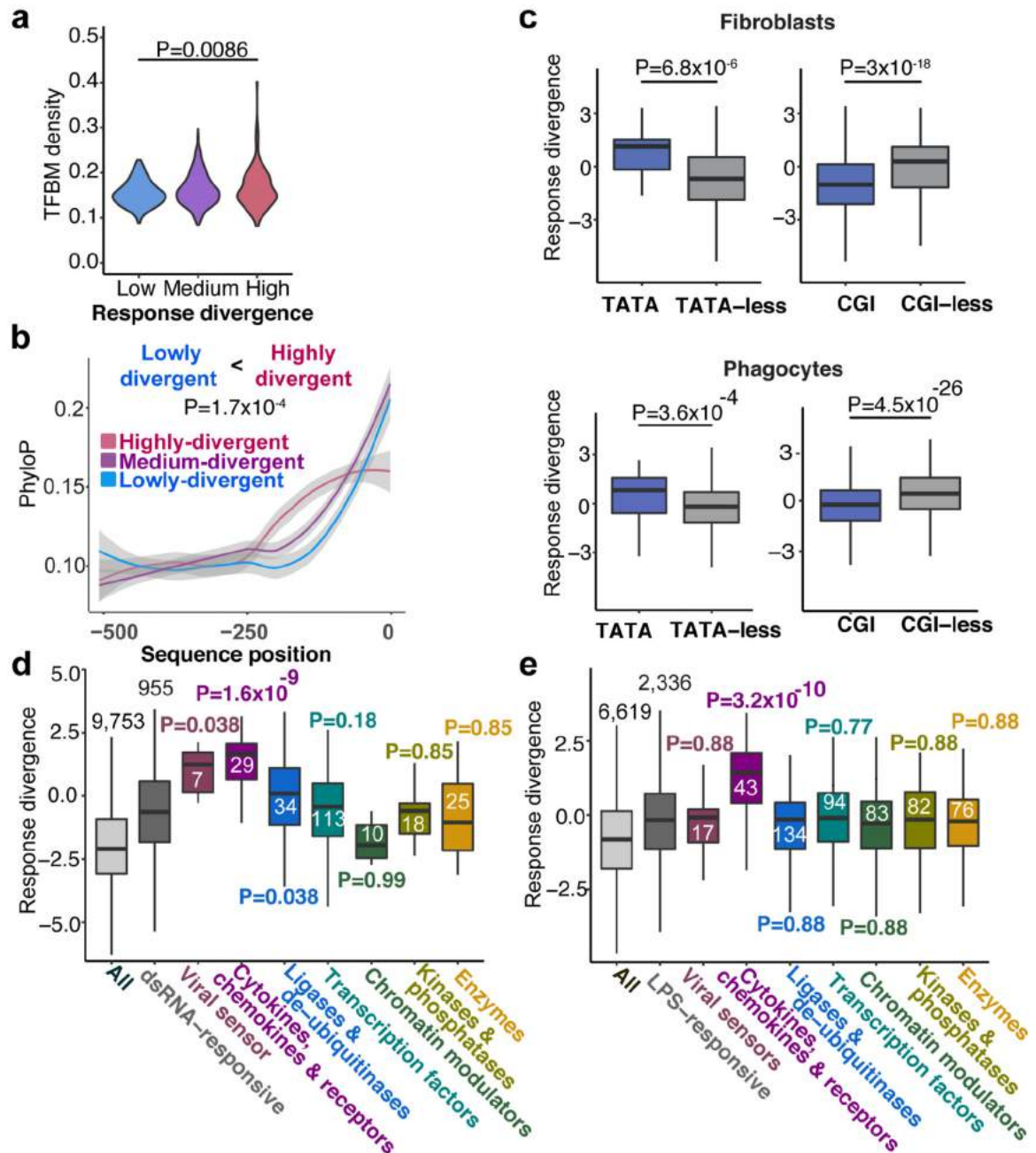


Figure 2. Transcriptionally divergent genes have unique functions and promoter architectures
 (a) TFBM density in active promoters and response divergence: For each gene studied in fibroblast dsRNA stimulation, the total number of transcription factor binding motif (TFBM) matches in its H3K4me3 histone mark was divided by the length of the mark (human marks were used) (n=879 differentially expressed genes with ChIP-seq data). Highly divergent genes have higher TFBM density than lowly divergent genes (one-side Mann-Whitney test).
 (b) Promoter sequence conservation and response divergence in fibroblast dsRNA stimulation: Sequence conservation values are estimated with phyloP7 for 500bp upstream

of the transcription start site (TSS) of the human gene. Mean conservation values of each of the 500 base pairs upstream of the TSS are shown for highly, medium and lowly divergent genes (n=840 genes). Genes that are highly divergent have higher sequence conservation (one-sided Kolmogorov-Smirnov test). The 95% confidence interval for predictions from a linear model computed by `geom_loess` function is shown in grey. (c) Comparison of divergence in response of genes with and without a TATA-box and CpG Islands (CGIs) in fibroblast dsRNA stimulation and phagocyte LPS stimulation. TATA-box matches and CGI overlaps were computed with respect to the TSS of human genes in fibroblasts (n=955 genes), and to the TSS of mouse genes in phagocytes (n=2,336). (d) Distributions of divergence values of 9,753 expressed genes in fibroblasts, 955 dsRNA-responsive genes and different functional subsets of the dsRNA-responsive genes (each subset is compared with the set of 955 genes using a one-sided Mann-Whitney test, FDR corrected p-values are shown). (e) Distributions of divergence values of 6,619 expressed genes in phagocytes, 2,336 LPS-responsive genes and different functional subsets of the LPS-responsive genes (each subset is compared with the set of 2,336 genes using a one-sided Mann-Whitney test, FDR corrected p-values are shown).

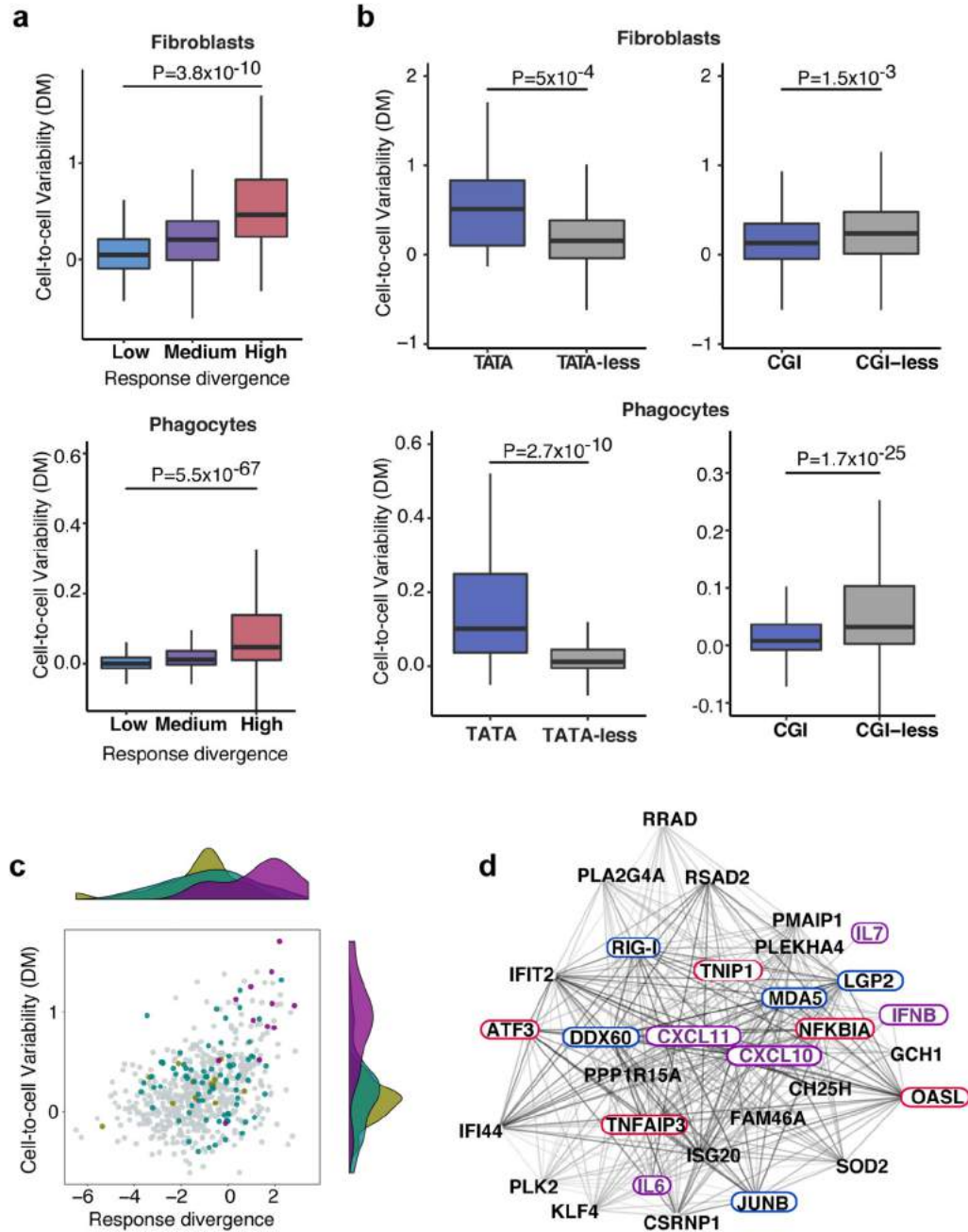


Figure 3. Cell-to-cell variability in immune response corresponds to response divergence
 (a) Comparison of divergence in response across species with transcriptional variability between individual cells. Top: fibroblast dsRNA stimulation (variability is measured in $n=55$ human cells, following 4H dsRNA stimulation). Bottom: Phagocyte LPS stimulation (variability is measured in $n=3,293$ mouse cells, following 4H LPS stimulation). Genes are split into highly, medium and lowly diverging genes based on their level of response divergence. Cell-to-cell variability values of highly divergent genes were compared with lowly divergent genes (one-sided Mann-Whitney test).

(b) Comparison of cell-to-cell variability of genes with and without a TATA-box and CGI, in fibroblast dsRNA stimulation and phagocyte LPS stimulation (one-sided Mann-Whitney test). Cell-to-cell variability values are from DM estimations of human fibroblasts stimulated with dsRNA for 4H (n=55 cells) and from mouse phagocyte stimulated with LPS for 4H (n=3,293 cells).

(c) A scatter plot showing divergence in response to dsRNA in fibroblasts across species and transcriptional cell-to-cell variability in human cells following 4H of dsRNA stimulation, of n=684 dsRNA-response genes. Genes from three functional groups - cytokines, transcription factors and kinases - are in purple, green and beige, respectively. The distributions of divergence values of these groups are shown above the scatter plot. The distributions of their variability values are shown to the right.

(d) A network showing genes that positively correlate in expression with the chemokine CXCL10 across cells (Spearman correlation, $\rho > 0.3$), in at least two species, following dsRNA treatment in fibroblasts (based on n=146, 74, 175, 170 human, macaque, rat and mouse cells, respectively). Cytokines are coloured in purple, positive regulators of cytokine expression are coloured in red, and negative regulators are coloured in blue. Colours of edges, from light to dark grey, reflect the number of species in which this pair of genes was correlated.

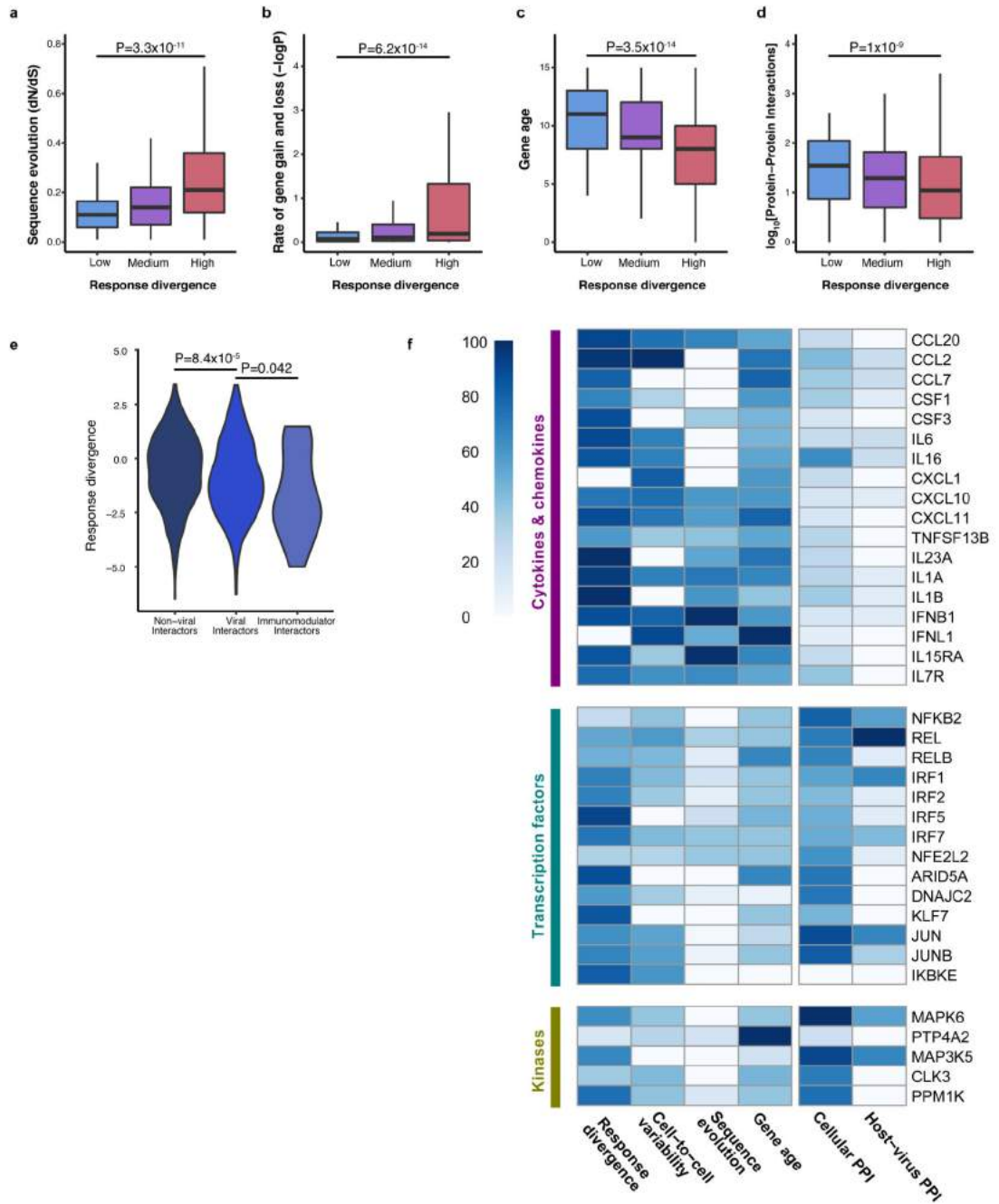


Figure 4. Relationship of response divergence and other evolutionary modes

In a-d, dsRNA-responsive genes in fibroblasts are divided by level of response divergence into three groups - as in Fig 1c. In each panel a different measure is shown in the y-axis for each of these groups.

(a) Coding sequence divergence, as measured using dN/dS values across 29 mammals. Higher dN/dS values denote faster coding sequence evolution (n=567 genes with dN/dS values).

(b) Rate at which genes are gained and lost within the gene family across the vertebrate clade (plotted as $-\log P$). Higher values denote faster gene gain and loss rate (n=955 genes).

(c) Evolutionary age (estimated with Panther7 phylogeny and Wagner reconstruction algorithm). Values denote the branch number with respect to human (distance from human in the phylogenetic tree); higher values indicate older age (n=931 genes).

(d) Number of known physical interactions with other cellular proteins (n=955 genes).

(e) Distribution of transcriptional response divergence values among dsRNA-responsive genes whose protein products do not interact with viruses, that interact with at least one virus, and that interact with viral immunomodulators (n=648, 377, 25 genes, respectively).

One-sided Mann-Whitney tests are used in panels a-e.

(f) A scaled heatmap showing: values of response divergence (as in Fig 1c), cell-to-cell variability (as in Fig 3a), coding sequence divergence (dN/dS values, as in 4a), gene age (as in 3c – younger genes have darker colours), number of cellular interactions (as in 4d) and number of host-virus interactions (as in 4e), for example genes from three functional groups: cytokines, transcription factors, and kinases. Values are shown in a normalized scale between 0-100, with the gene with the highest value assigned a score of 100. Missing values are coloured in white.