A balance between regulatory constraints and pathogen pressure shapes the evolution of innate immunity

Authors
Tzachi Hagai[1,2], Xi Chen[1], Ricardo J Miragaia[1,3], Tomás Gomes[1], Raghd Rostom[1,2], Natalia Kunowska[1], Valentina Proserpio[2], Giacomo Donati[4], Lara Bossini-Castillo[1], Guy Naamati[2], Guy Emerton[1], Gosia Trynka[1], Ivanela Kondova[5], Mike Denis[6] and Sarah A Teichmann[1,2,7]

Affiliations
(1) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
(2) EMBL- European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
(3) Centre of Biological Engineering, University of Minho, Braga, Portugal
(4) Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Torino, Italy
(5) Division of Pathology and Microbiology, Biomedical Primate Research Centre, 2280 GH Rijswijk, the Netherlands
(6) Research Department. Public Health England, National Infection Service, Porton Down, Wiltshire SP4 0JG, UK
(7) Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK

Correspondence:

Sarah Teichmann - st9@sanger.ac.uk (Lead contact)
Tzachi Hagai – Tzachi@ebi.ac.uk

Abstract

The immune system is under constant pressure from pathogens to evolve and mount a sufficiently strong response. At the same time, an overresponsive immune system can lead to autoimmunity and tissue damage. How these conflicting demands have shaped human immunity is not well understood.
Here, we characterize transcriptional divergence in the innate immune response across primates and rodents using bulk and single-cell transcriptomics, combined with chromatin analysis.
We discover that genes that diverge in transcriptional response across species vary in cell-to-cell expression within each species. These divergent genes are evolutionarily younger, experience rapid coding sequence evolution and display a distinct promoter architecture. They have exclusive immune functions, such as cellular defence and inflammation, while pleiotropic genes involved in immunity and other pathways are more conserved.
Analysis of viral interactions and mimicry shows that viruses target conserved elements of the innate immune response, suggesting that regulatory constraints imposed on the host are exploited by viruses. Importantly, innate immune genes implicated in autoimmune diseases show high levels of transcriptional divergence but also more interactions with viruses. This reveals a conflict between pathogens and regulatory constraints, which has likely contributed to genetic architectures driving the pathogenesis of immune disorders.

Introduction

Pathogens exert a strong selective pressure on their host (Enard et al., 2016; Fumagalli and Sironi, 2014; Fumagalli et al., 2011), and as a result, the host immune system must rapidly evolve to face adept and ever-changing pathogens. Changes that confer stronger resistance to pathogens may however come at a cost. Indeed, several studies suggest that some genes that were targets of pathogen-driven selection in human evolution might also be implicated in autoimmune disorders (Fumagalli et al., 2011; Li et al., 2016).

Comparative analyses across the genome have highlighted immune genes as amongst the fastest to evolve in their coding sequences (Nielsen et al., 2005). Evolutionary changes in the coding sequences of several immune genes were shown to affect resistance to viruses, corroborating the link between rapid sequence evolution and pathogen pressure (Daugherty and Malik, 2012; Duggal and Emerman, 2012).

Long term adaptation to pathogens through other evolutionary mechanisms, such as changes in gene expression, are less well understood. A cross-species comparison of coding and non-coding sequences has suggested that long term adaptation of immune genes is dominated by coding sequence evolution(Haygood et al., 2010). Comparative transcriptome studies showed that various innate and adaptive expression programmes are largely conserved between species (Barreiro et al., 2010; Mostafavi et al., 2016; Schroder et al., 2012; Shay et al., 2013). This regulatory conservation might stem from the requirement to maintain a highly coordinated and balanced reaction: resisting pathogens while minimizing disturbance of normal homeostasis and avoiding long-term tissue damage (Ivashkiv and Donlin, 2014; Porritt and Hertzog, 2015). How these conflicting demands have shaped the immune system, and how immune disease pathogenesis has arisen as a result, remain outstanding questions for our understanding of human health and evolution.

Here, we investigate these questions by focusing on the cell-intrinsic innate immune response – the first line of defence initiated in most cell types upon pathogen detection. This program comprises two major transcriptional waves: (1) Recognition of microbial elements, such as cytoplasmic dsRNA, induces rapid upregulation of antiviral and inflammatory genes, including many cytokines and chemokines such as interferon (IFN). (2) Consequently, secreted IFN shifts nearby cells into an antiviral state through upregulation of hundreds of interferon stimulated genes (ISGs) (Borden et al., 2007; Iwasaki, 2012).

To study the evolution of innate immunity in a systematic manner, we compare the response of dermal fibroblasts from primates (human and rhesus macaque) and rodents (mouse and rat) to two stimuli that elicit the primary and

secondary response waves respectively: poly I:C (synthetic dsRNA, commonly used to mimic viral infection) and interferon beta (IFNB, the dominant Type I IFN expressed in fibroblasts in response to infection). By combining three genomics approaches – population (bulk) transcriptomics, single-cell transcriptomics and chromatin-IP sequencing of two histone marks, we comprehensively characterize the transcriptional changes between species in their innate immune response (see Fig 1B-D). Subsequently, we investigate how divergence in expression is related to other evolutionary changes, such as gene age and coding sequence evolution (Fig 1E-F). Finally, we identify correlations between divergence, interaction with viruses and implication in autoimmunity (Fig 1G-H). Our results suggest that conflicts between pathogen pressure and regulatory constraints imposed on the immune system have given rise to pathogenesis of autoimmune diseases.

Results

Skin fibroblasts offer an excellent model to study the evolution of the ubiquitous, cell-intrinsic innate immune response as they are commonly used in immunological and virological studies, and provide a comparable *in vitro* system between species.

To compare the transcriptional response across the four species, we generated bulk RNA-sequencing data from two stimuli, along with respective controls. The early response was profiled 4 hours after dsRNA stimulation. The secondary response was studied with direct IFNB treatment, to avoid complicating factors such as apoptosis and secondary feedback loops that occur in late stages of the dsRNA stimulation.

Gene expression divergence recapitulates the evolutionary relationships between species

First, we characterized the overall similarity between species in response to dsRNA and to IFNB using bulk transcriptomic data. For this, we constructed a dendrogram for dsRNA and IFNB stimulations across species using fold change in expression between treatment and control across orthologous genes (Fig 2A). The resulting tree first separates the two responses (IFNB and dsRNA). Within each response, the two clades (primates and rodents) are correctly split. Thus, divergence in response between species largely recapitulates the known evolutionary relationship as observed in other expression programmes (Brawand et al., 2011; Kalinka et al., 2010; Levin et al., 2016), suggesting that

regulatory changes accumulate gradually over evolutionary time (see also Fig S1).

Inflammatory and cellular-defence response genes diverge most rapidly

Many genes have a conserved response across species, such as the NFkB subunits RELB and NFKB2 (Fig 2B). However, some genes diverge between clades: IFI27 – a restriction factor against numerous viruses - is strongly upregulated in primates but not in rodents (Fig 2B), while DAXX – an important antiviral transcriptional repressor, and IL18 exhibit the opposite behaviour (Fig 2B). Furthermore, some genes have species-specific response patterns: for example, IL7RA – a cytokine receptor - is upregulated in response to dsRNA in human and other species, but not in mouse (in agreement with observations from LPS stimulation of human and mouse macrophages(Schroder et al., 2012)). SAMHD1 – a triphosphohydrolase that restricts various lentiviruses, is highly upregulated in all species but macaque (see additional examples in Fig S2).

To quantify transcriptional response divergence between species, we focused on genes that are differentially expressed in human in response to treatment (with a corrected p-value<0.01, as determined by edgeR exact test). For simplicity, we will refer to these genes as "innate response genes" (see Fig 2C). There are 955 such genes with one-to-one orthologs across the species in our dsRNA dataset, and 841 genes in the IFNB dataset.

For each gene, we fitted Ornstein–Uhlenbeck models - which incorporate selection and genetic drift (Butler and King, 2004) – to the change in expression after stimulation across the four species (see Fig 2C and Methods). This provides a measure of response divergence between species that is comparable across all genes (numbers in Fig 2C represent the set of innate response genes in dsRNA treatment. This analysis was repeated with the IFNB expression data – see Methods).

Following this procedure, we split the innate response genes into three groups: (1) highly divergent innate response genes (the top 25% of genes with the highest divergence values), (2) lowly divergent genes (the bottom 25%), and (3) genes with medium divergence across species (the middle 50%).

Next, we compared the functional enrichment in the set of highly divergent genes versus those with low divergence, using GOrilla enrichment (Eden et al., 2009), with one group serving as the target and the other as the background. We observed functional differences between the two groups: highly divergent genes are enriched with processes such as response to stimulus and defence response, and lowly divergent genes are enriched with processes

related to cellular homeostasis, such as transport and splicing (Fig 2D shows the top five non-redundant GO terms enriched in each of the groups).

Furthermore, when we compare the divergence rate of genes in different functional categories, we observe that genes that are related to cellular defence and inflammation, most notably cytokines, chemokines and their receptors, tend to diverge significantly faster than genes involved in apoptosis, immune regulation (chromatin modulators, transcription factors, kinases and ligases) or metabolism (Fig 2E). Similar findings are observed in response to direct IFNB treatment (Fig S4).

Thus, we have identified a group of genes that have experienced relatively fast expression divergence between the four species, and these genes have specific functions in innate immunity. Importantly, pleiotropic genes involved in immune response regulation and other non-immune pathways are more conserved in their transcriptional divergence.

Highly divergent genes have promoters with conserved sequences and distinct architectures

Next, we analyzed how divergence in response to dsRNA is reflected at the chromatin level. We quantified genomic enrichment of histone 3 lysine 4 trimethylation (H3K4me3) and histone 3 lysine 27 acetylation (H3K27ac) – two histone marks widely used to profile active promoters and enhancers. We linked genes with these marks by proximity (see Methods). We then compared divergence of the histone marks' signal with transcriptional divergence in response to dsRNA. For example, Fig 3A shows the expression level and the levels of the two histone marks around the TSS of IL33 - a diverging cytokine expressed in rodents but not in primates. IL33 is strongly upregulated in mouse early in response to dsRNA, but not in other species. Consistent with gene expression, increased signals of H3K4me3 and H3K27ac are detected around the IL33 promoter in mouse after dsRNA stimulation.

To study this across all innate response genes we assessed conservation of active promoters (H3K4me3-positive) and enhancers (H3K27ac-positive, H3K4me3-negative) for each human gene in the corresponding syntenic regions in macaque, rat and mouse (Fig 3B). When we compare the conservation of active promoters between human and other species, we observe that genes more conserved in their response are enriched with conserved active promoters (larger blue fraction in Fig 3B). This is also the case for enhancers (except for the human-macaque comparison).

We next tested whether divergence in response is reflected in promoter sequence conservation and architecture. Interestingly, the region immediately

upstream of the TSS has higher sequence conservation in genes that diverge in response (Fig 3C). This surprising discordance may be related to the fact that promoters of highly and lowly divergent genes have different architectures, which impose different constraints on promoter sequence evolution (Roider et al., 2009; Schroder et al., 2012; Tirosh et al., 2006).

By analyzing motifs, we find that TATA-box elements tend to occur in promoters of highly divergent innate response genes, while CpG islands (CGIs) occur in more conserved genes (Fig 3D). Furthermore, a higher density of transcription factor binding motifs (TFBMs) is found in highly divergent innate response genes (Fig 3E). Thus, a promoter architecture with a higher density of TFBMs, CGI depletion and TATA-box enrichment supports higher divergence, both in histone marks and in transcriptional response, while entailing higher sequence conservation upstream of these genes. Conversely, genes with CGI promoters tend to be transcriptionally conserved.

Divergent genes are heterogeneous in their expression in individual cells within a species

We next asked whether genes that vary in response between species have high cell-to-cell variability in expression. Using single-cell RNA-seq, we profiled transcriptomes of hundreds of individual cells from all species in a time course of 0,2,4 and 8 hours following dsRNA stimulation. We quantified the transcriptional cell-to-cell variability for each gene, and compared this to cross-species response divergence (based on bulk transcriptomics, as mentioned above). Variability in gene expression can result from technical variation or from truly stochastic gene expression. We thus used BASiCS – an approach that discerns biological variation from technical noise by utilizing observed levels of synthetic spiked-in RNA molecules (Vallejos et al., 2015). These are added in equal amounts to each lysed cell, but their detection may vary due to differences in processing and sequencing, thereby informing on technical variation.

Strikingly, we observe a clear trend for genes that highly diverge in response between species to also be more variable in expression across individual cells within a species (see Fig 4A for human cells after 4hrs of dsRNA stimulation). The divergence and cell-to-cell variation levels for several functional groups of innate response genes are shown in Fig 4B. Some of the most variable genes are cytokines, chemokines and their receptors. For example, IFNB - the gene with the highest cell-to-cell variability in our data - was previously observed to have a high level of stochasticity in cells infected with Sendai virus, with very high IFNB expression in a low proportion of cells (Zhao et al., 2012). In contrast to cytokines and chemokines, most of the genes involved

in regulation, such as kinases and transcription factors, display lower cell-to-cell variability.

When looking at the progression of the response along a time course from 0 to 8hrs, we observe that the transcriptional variation of cytokines increases, peaking at early time points (Fig 4C). This behaviour reflects the known expression profile of several cytokines: initial expression in a few cells increases to a larger proportion of cells as the response progresses (Shalek et al., 2014). Importantly, this behaviour is conserved across all species (Fig S5).

We confirmed that these results are not biased by particular computational or experimental settings by repeating the analyses with a second computational approach - distance from median (DM) (Newman et al., 2006) and by profiling 4,463 human cells that were processed using a different experimental technique - microfluidic droplet capture (Methods and Figs S6-7).

Taken together, these results reveal a global trend in the innate immune response with genes diverging in expression between species also varying between cells within a species. These within- and between-species variable genes tend to have specific functions in the innate response, including cytokines, chemokines and their receptors.

Transcriptionally divergent genes also diverge in protein sequence

So far, we have systematically characterized the divergence and variability in transcriptional response and at the chromatin level. Next, we examined the relationship between divergence in response and divergence in coding sequence. Various studies have found that immune genes exhibit high rates of sequence evolution (Fumagalli and Sironi, 2014; Nielsen et al., 2005). However, whether high divergence in transcriptional response is related to high divergence in coding sequence in immune genes is not known.

To compare the rate at which genes evolved in their coding sequences with divergence in response, we used the ratio between non-synonymous (dN) to synonymous substitutions (dS), which can be used to indicate the level of selection operating on a protein. We observe that genes with high levels of divergence in response tend to have higher levels of divergence in coding sequences (higher dN/dS values) compared with innate response genes with low response divergence (Fig 5A). This is consistent for both the dsRNA and the IFNB response (Fig S8A) and when calculating dN/dS using different sets of mammals (see Methods). Thus, the two evolutionary modes – changes in coding sequence and in expression - occur in tandem in innate response genes.

Transcriptionally divergent genes have higher rates of gene gain and loss in the course of mammalian evolution

We next 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. For this, we used the statistics provided in ENSEMBL where the relative rate of gene duplication and loss across vertebrates is assessed using a model based on global birth-and-death rates (Herrero et al., 2016). We compared these rates in genes that diverged rapidly in innate response with genes that have a more conserved response. We found that transcriptionally divergent response genes have experienced faster rates of gene gain and loss (Figs 5B & S8B).

Rapid expansion of gene families was previously observed in several classes of immune genes (e.g. – MHC class I and APOBEC3). More recently, important immune genes were reported to have been lost in several mammalian lineages: IFNE was lost in pangolins (Choo et al., 2016), MX1 and MX2 were lost in toothed whales (Braun et al., 2015), TRIM5 in dog (Johnson and Sawyer, 2009), RIG-I was lost in tree shrews (Xu et al., 2016), and AIM2 in two mammalian orders (Brunette et al., 2012). Here, we show that those innate response genes that have evolved faster in their expression and coding sequences have experienced gene gain and loss events significantly more rapidly. Interestingly, rapid gene gain and loss was also observed in immune genes in the fly (Sackton et al., 2017), suggesting that this phenomenon is recurrent in different clades. A plausible scenario for the observed high rate of gene gain and loss in important immune genes is the historical rise of genes as a result of encounters with new pathogens, and a rapid decay following changes in pathogen pressure, due to the cost inflicted on the host by expressing these genes (Barber et al., 2016; Malfavon-Borja et al., 2013).

Transcriptionally divergent innate response genes arose more recently in evolution

Higher rates of gene gain and loss in a gene family, as observed above, imply that at least some members of this family are evolutionarily younger. We thus investigated the evolutionary age of the set of innate response genes using eleven different combinations of age estimation parsimonies and vertebrate gene families obtained from ProteinHistorian (Capra et al., 2012). Indeed, genes that diverged more rapidly in their innate response are evolutionarily younger (Fig 5C, Fig S9).

Previous reports suggest that younger genes tend to have fewer cellular interactions (Saeed and Deane, 2006). We indeed observe that diverging genes tend to have fewer experimentally validated protein-protein interactions (PPIs, obtained from STRING (Szklarczyk et al., 2015)), in agreement with their estimated age (Fig 5D & Fig S8C).

Figure 5E shows various evolutionary and regulatory characteristics of different cytokines, chemokines and their receptors, as well as transcription factors and kinases. Genes predominantly functioning in immunity, such as many cytokines and their receptors, evolve more rapidly in different evolutionary modes and have less cellular PPIs. Genes with pleiotropic or regulatory functions, such as kinases and transcription factors are more constrained in their evolution, in line with the higher numbers of cellular interactions they form. The contrasting trends between these groups underline how functional and regulatory constraints have shaped the evolution of innate response genes.

Viruses evolve to preferentially target conserved innate response genes

Pathogens are a major selective pressure in mammalian evolution (Enard et al., 2016; Fumagalli and Sironi, 2014), and are thought to drive the evolution of innate immune genes (although other factors might also be at work). We thus investigated viral antagonistic relationships with highly divergent *versus* conserved innate response genes.

We combined three datasets of host-virus protein-protein interactions from various viruses (Calderone et al., 2015; Enard et al., 2016; Halehalli and Nagarajaram, 2015). By comparing the total number of known interactions with divergence in response (Fig 6A), we observe that lowly divergent innate response genes have a significantly higher number of known viral interactions. This is also true of divergence in sequence (Fig S10). Thus, it seems that viruses have preferentially evolved to interact with innate response proteins that are relatively conserved in their transcriptional response. Presumably, these conserved host genes cannot evolve away from these interactions (Dyer et al., 2008).

Another form of antagonistic interactions can be manifested in molecular mimicry – where viruses mimic a host gene, usually by incorporating this gene into their genome through horizontal gene transfer. Using a dataset of viral genes identified as products of horizontal gene transfer (Palmeira et al., 2011), we observe that viruses tend to incorporate host genes belonging to the class of transcriptionally conserved genes (Fig 6B). In addition, mimicked genes tend to be evolutionarily older and to evolve more slowly in their coding sequences (Fig 6B, inset). Thus, antagonistic interactions of viruses with immune genes, either

through direct interactions or in the form of mimicry, preferentially occur with conserved elements of the innate immune response.

Genes implicated in autoimmunity have high response divergence and many interactions with viruses

The Type-I IFN pathway has been implicated in the pathogenesis of various autoimmune diseases (Crow and Manel, 2015; Hall and Rosen, 2010). We thus sought to characterize the relationship between innate response genes' involvement in autoimmunity and their regulation, evolution and interactions with viruses.

We first focused on a small set of interferonopathy-related genes: these are genes that have known mutations leading to aberrant expression of Type I IFN in pathological conditions such as systemic lupus erythematosus and Aicardi–Goutières syndrome (Crow and Manel, 2015). Twelve of these genes are expressed in fibroblasts, eight of which are upregulated in response to dsRNA or IFNB. When looking at the regulatory, evolutionary and physiological characteristics of these genes, an intriguing picture emerges (Fig 6C).

Interferonopathy-related genes diverge rapidly in expression and in coding sequence, but most of them are evolutionarily old genes and do not display high rates of gene gain and loss in the course of vertebrate evolution in comparison with other innate response genes (Fig 6C). Importantly, despite their high transcriptional divergence, many of them interact with numerous viral proteins – contrary to the overall trend where viruses interact preferentially with lowly divergent innate response genes (Fig 6C *versus* Fig 6A).

To investigate this on a larger set of genes that are associated with autoimmunity, we gathered a set of SNPs from genome-wide association studies (GWAS) of ten autoimmune and chronic inflammatory diseases (see Methods). We found genes co-localizing with SNP-associated genomic regions. These genes, which are likely to be implicated in autoimmune diseases, are referred to as "GWAS genes" here.

When comparing GWAS innate response genes with "non-GWAS" innate response genes, we observe that the GWAS genes have significantly higher transcriptional divergence and cell-to-cell variability (Fig 6D). GWAS innate response genes are not different than other innate response genes in their coding sequence evolution or gene age (Fig 6D). Thus, genes associated with autoimmune diseases tend to diverge more rapidly in transcriptional response but not in other evolutionary modes.

Similarly to the interferonopathy set, GWAS genes have significantly higher numbers of interacting viral proteins (Fig 6D), contrasting with their high

transcriptional divergence. Moreover, when looking at the number of viruses that are inhibited *in vitro* by innate response genes, we observe that GWAS genes inhibit significantly higher numbers of viruses in comparison with other innate response genes (Fig 6D and Methods). These trends are also observed across GWAS gene sets defined with different stringencies (Fig S11), but not in control sets unrelated to autoimmunity (Fig S12).

These results suggest that dsRNA-regulated genes associated with various autoimmune diseases have distinctive regulatory, evolutionary and physiological characteristics in comparison with other innate response genes. These genes are targeted by numerous viruses but are also transcriptionally divergent, unlike the general trend of innate response genes where genes targeted by viruses are mostly conserved.

We next tested whether these GWAS genes are also enriched with specific functions compared to non-GWAS innate response genes (using GOrilla to compare GO term enrichment between these two sets). We observe that the set of GWAS genes is enriched with functions related to regulation of the innate immune response and to signal transduction (Fig 6E). Indeed, this set includes key regulators of the innate immune response, such as TRAF1, TNFAIP3, IRF7, REL, NFKB1, STAT2 and ETS1.

The trends we observe with respect to the regulation, physiology and evolution of genes implicated in autoimmune diseases provide new insights into the pathogenesis of autoimmune diseases and its evolution: Innate response genes that are implicated in autoimmune diseases are involved in extensive antagonistic interactions with pathogens, as suggested by the high numbers of viruses they inhibit or interact with.

However, as these genes are enriched in regulatory functions, and often have additional non-immune functions, their protein sequences are constrained and cannot rapidly evolve in response to pathogen pressure.

Changes in expression, unlike changes in coding sequences, can be tissue- and condition-specific. Therefore, changes in expression in an immune context can act as a means to evolve as part of the arms race with pathogens, without affecting non-immune functions of these genes. However, some changes in expression of these regulatory factors, which operate up- and down-stream of interferon and inflammatory cytokines, can lead to an imbalanced response – resulting in tissue damage and autoimmune diseases.

Various GWAS and expression quantitative trait loci (eQTL) studies have suggested that transcriptional variation to immune challenges between human individuals and between human populations has a large genetic component, which has been at least in part driven by pathogen selection (Fairfax et al., 2014; Lee et al., 2014; Li et al., 2016; Nedelec et al., 2016; Quach et al., 2016). Some

of these studies have found interesting links between genetic loci implicated in autoimmune diseases and signatures of recent human adaptation (Brinkworth and Barreiro, 2014; Fairfax et al., 2014; Fumagalli and Sironi, 2014; Kumar et al., 2014; Lee et al., 2014; Li et al., 2016). Taken together with our results, it seems likely that many transcriptional changes that have occurred both recently and during longer evolutionary periods, since the split between primates and rodents, have been driven by pathogen pressure, at the cost of a predisposition to immune disorders.

Discussion

In this work, we have taken an integrative approach to chart the evolutionary architecture of the innate immune system, combining genomics tools, evolutionary analyses and associations with viruses and autoimmune diseases (data can be visualized in: https://teichlab.sanger.ac.uk/ImmunityEvolution). We have identified a set of genes that rapidly diverge in response between species. These genes are distinguished from lowly-diverging genes in their regulation, evolution and physiology, as well having distinct patterns of interactions with viruses (Fig 7).

Changes in expression between species are often attributed to neutral drift (Khaitovich et al., 2006), and based on our results we cannot categorically determine whether the majority of the changes we observe are selective or neutral. The observation that genes that diverge faster tend be evolutionarily younger and transcriptionally more variable between individual cells may point to the neutral scenario. However, fast evolving genes have direct roles in pathogen restriction, making them plausible targets of strong selection. Furthermore, their rapid evolution occurs in independent evolutionary modes, supporting the notion that at least some of the changes we observe in their transcriptional response are driven by selection.

Transcriptionally divergent genes show higher levels of cell-to-cell variability. This stochasticity in transcription can be adaptive: cytokines and chemokines display the highest transcriptional variation and are limited in their expression to a small subset of cells (Zhao et al., 2012). This mode of expression can balance between the requirements of a strong rapid response while avoiding a fatal immune (over)reaction. Thus, restricting cytokine expression to high levels in only a few cells, a phenomenon conserved across the species studied, can be sufficient to upregulate a timely immune response in nearby cells, while retaining the ability to be rapidly turned off in a controlled manner.

Our observation that viruses evolve to preferentially antagonize conserved immune genes underlines the constraints imposed on some components of the host immune system. Slowly evolving immune genes often function in additional contexts unrelated to the innate immune response (such as those involved in apoptosis, metabolism or signalling in other pathways) as indicated by their higher number of cellular protein-protein interactions. These functions limit genes' ability to evolve - an Achilles' heel used by pathogens to subvert the immune system. Stemming from these constraints, a successful host strategy is manifested through cytokines and other secreted immune proteins. Not only do they act as an efficient means of alerting neighbouring and immune cells of pathogen presence, they also function as an important element in the evolutionary arms race, since they can readily evolve with fewer constraints imposed by intracellular interactions or additional non-immune functions.

Other components of the innate immune response, however, are more constrained. Genes that function as upstream regulators of the innate immune response (from viral sensors to signal transducers and transcription factors) are trapped between antagonistic interactions with viruses and constraints imposed by regulatory restrictions and their pleiotropic functions in other non-immune contexts. Changes in transcription can be context-specific and can thus provide means to evolve against pathogens without affecting non-immune functions. These regulatory changes may, nevertheless, come with a cost, such as predisposition to autoimmune diseases due to an imbalanced immune response.

Our study thus uncovers several underlying characteristics of the evolutionary architecture of innate immunity and gives insights into how pathogen pressure may have led to autoimmune disease pathogenesis.

Acknowledgements

We would like to thank Yoav Gilad, Orly Laufman, Amir Marcovitz, John Marioni, Matthieu Muffato, Oliver Stegle, Adi Stern, Mike Stubbington and Michelle Ward for helpful discussions and advice throughout the project; Lira Mamanova and Sanger Support teams for technical assistance, and members of the Teichmann lab for support in various stages. This project was supported by ERC grants (ThDEFINE and ThSWITCH) and an EU FET-OPEN grant (MRG-GRAMMAR). T.H was supported by a Human Frontier Science Program Long-Term Fellowship and by an EMBO Long-Term fellowship.

Figure Legend:

Fig 1: Study design: An integrative approach to charting the evolutionary architecture of the innate immune response

(A) Dermal fibroblasts from sexually-mature females of four species were stimulated with dsRNA or IFNB (and controls) to initiate the innate immune response. Samples were collected for population (bulk) and single-cell RNA-seq and for ChIP-seq. **(B)** Using bulk transcriptomics, genes were quantified to estimate the level of transcriptional divergence in response to treatment between the species. This measure of divergence was compared with various regulatory, evolutionary and physiological characteristics: **(C)** Divergence in regulatory elements (using ChIP-seq of two histone marks - H3K4me3 and H3K27ac) and promoter analysis, **(D)** Cell-to-cell variability in expression (using single-cell transcriptomics), **(E-F)** Divergence in various evolutionary modes: coding sequence evolution, evolutionary age and rate of gene gain and loss, **(G)** Host-virus protein-protein interactions and viral mimicry, **(H)** Implication in autoimmune diseases.

Fig 2: Conservation and divergence in innate immune response

(A) 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 (see Fig S1 for additional comparisons). **(B)** Fold change in response to dsRNA in example genes. **(C)** Estimating the level of cross-species divergence in response: (1) using edgeR, the fold change in expression in response to dsRNA was assessed for each gene in each species. (2) Differentially expressed genes were identified using edgeR exact test (q-value<0.01), out of which 955 have one-to-one orthologs across species (3) Ornstein-Uhlenbeck models were fitted to the fold change data for each gene, using the expression changes of the four orthologs and the known phylogeny (see Methods). **(D)** GO-term enrichment (negative log of p-values) of highly divergent (pink, left) and lowly divergent (blue, right) genes in response to dsRNA. In each case, the mentioned group was used as the target and the second group as the background. (**E)** Distributions of divergence values of all 955 innate response genes and different functional subsets of this group (see Methods for details on each of the functional subsets). Left: division by molecular function; Right: division by biological process. Values of divergence in response are estimated as $log\frac{\sigma^2}{2\alpha}$, where α and σ are the coefficients for selection and drift (respectively) from a fitted Ornstein-Uhlenbeck model (see Methods). Empirical p-values are shown (*p<0.05, **p<0.01,***p<0.001).

Fig 3: Regulatory elements, promoter architecture and response divergence
(A) H3K4me3 and H3K27ac marks around the IL-33 promoter in unstimulated and dsRNA-treated fibroblasts from human, macaque, rat and mouse. The mean RNA expression level (in transcript per million – (TPM)) for each species and condition is shown in brackets. **(B)** The degree of divergence of active promoters (H3K4me3) and enhancers (H3K27ac without overlapping H3K4me3), 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. Genes are separated into three groups by their degree of transcriptional divergence. The fractions of genes with conserved promoters (or enhancers) were compared between highly and lowly divergent genes using Fisher exact test. **(C)** Promoter sequence conservation and transcriptional divergence: Sequence conservation values are estimated with phyloP7 (higher values indicate higher conservation) for regions around the transcription start site (TSS). Upstream of the TSS, genes that are highly-divergent have higher sequence conservation ($p=3.E-6$), while downstream of the TSS, lowly-divergent genes are more conserved in their sequence ($p<2E-16$) (two-sample Kolmogorov-Smirnov test). **(D)** Comparison of divergence in response of genes with and without TATA-box (top) and CpG Islands (CGIs) (bottom); empirical p-values are shown. **(E)** Density of transcription factor binding motifs (TFBMs) in H3K4me3 marks of genes with high, medium and low response divergence; empirical p-values are shown. (*$p<0.05$, **$p<0.01$, ***$p<0.001$ in all panels).

Fig 4: Cell-to-cell variability and response divergence
(A) A comparison of divergence in response (measured in bulk transcriptomics across four species) and transcriptional heterogeneity between individual cells (measured in single-cell transcriptomics in human cells). In both measures, cells were stimulated with dsRNA for 4 hour. Cell-to-cell variability was estimated using BASiCs (higher values indicate higher levels of variability). Cell-to-cell variability values of highly-divergent genesis were compared with lowly-divergent genes using Mann-Whitney test. **(B)** The same data as in (A), here plotted as a scatter with cross-species divergence and transcriptional heterogeneity per gene. Highlighted genes include cytokines and chemokines (purple), transcription factors (turquoise), and kinases and phosphates (khaki), all other innate response genes are in grey. Top bar – density plots of response divergence in cytokines, transcription factors, kinases and all genes. Right bar – density plots of cell-to-cell variability of the same functional classes. **(C)** Cell-to-cell variability along the dsRNA time course (0,2,4 and 8hrs following stimulation) in human. In each panel the distribution of cell-to-cell variability is

shown for each of the functional classes of genes (colors as in B). A schematic dotted line represents the level of cytokine cell-to-cell variability along the time course.

Fig 5: Response divergence and other evolutionary modes

Comparing divergence in response to dsRNA with: **(A)** Coding sequence divergence, as measured using dN/dS values across a clade of 29 mammals. Higher dN/dS values indicate faster sequence evolution. **(B)** Rate at which genes are gained and lost within the gene family across the vertebrate clade. Values are plotted as –logP, higher values indicate faster rates. **(C)** Evolutionary age (as estimated using Panther7 phylogeny and Wagner reconstruction algorithm). Values denote the branch number with respect to human, higher values indicate older age. **(D)** Number of known physical interactions with other cellular proteins (protein-protein interactions - PPIs) (log scale). **(E)** A heatmap showing the values of response divergence, coding sequence divergence (dN/dS values across a clade of 29 mammals and between human and mouse), gene age (distance from human branch, estimated as in C – older genes have higher values), rate of gene gain and loss (in p-values, lower values denote higher rate), number of cellular PPIs and number of host-virus PPIs for example genes: cytokines, chemokines and their receptors, transcription factors, kinases and phosphates and other divergent genes. Values are shown in a normalized scale between 0-100. In A-D a statistical comparison (Mann-Whitney test) is made between the distribution of the tested measure's values in highly and lowly divergent genes. (*$p<0.05$, **$p<0.01$, ***$p<0.001$.)

Fig 6: Innate immune gene evolution, interactions with viruses and association with autoimmune diseases

(A) Comparison of gene divergence in response to dsRNA with the number of known physical interactions it forms with viral proteins: Lowly divergent genes have higher numbers of interactions with viruses (Mann-Whitney test). **(B)** Comparisons of divergence in response to dsRNA, sequence divergence (inset, top, dN/dS values, as in 5A, higher values indicate faster sequence evolution) and gene age (inset, bottom, branch number with respect to human gene, as in 5C, higher values indicate older age) in innate response genes that are mimicked by viruses with those that were not found to be mimicked. (Empirical p-values, *$p<0.05$, **$p<0.01$, ***$p<0.001$). **(C)** A heatmap of evolutionary, regulatory and physiological characteristics of interferonopathy genes showing if the gene's value in the respective characteristics is above, below or equal to the median of the entire set of innate response genes (e.g. – brown color denotes genes with older than median evolutionary age): Transcriptional divergence in response to

dsRNA, sequence divergence (dN/dS across a clade of 29 mammals), gene age (Panther7 phylogeny, Wagner algorithm, higher values indicate older age), rate of gene gain and loss (in –log10 scale, higher values indicate faster rate), viral PPIs (number of known host-virus interactions), viral inhibition (number of viruses inhibited by this gene *in vitro* – see Methods), cell-to-cell variability (estimated by BASiCs). Data on sequence divergence of ADAR is not available. **(D)** Density plots of 66 autoimmune GWAS genes in comparison with all other innate response genes, showing distributions of response divergence to dsRNA, sequence divergence, gene age, cell-to-cell variability, number of known host-virus interactions (in log2 scale), and number of *in vitro* inhibited viruses. All measurement units are as described in C. P-values shown are empirical; *p<0.05, **p<0.01, ***p<0.001. Data here refers to GWAS genes identified using the most stringent parameters, see additional plots of GWAS gene sets identified based on different cut-offs in Fig S11. **(E)** GO-term enrichment of autoimmune GWAS genes in comparison with all other innate response genes (showing all terms with q-value<0.05).

Fig 7: Regulatory, evolutionary and physiological characteristics of highly and lowly divergent immune genes

Innate immune genes that are highly-divergent in their response (left, in pink) have distinct evolutionary, regulatory and physiological characteristics in comparison with lowly-divergent genes (right, in blue): **Expression and regulation:** Highly divergent genes have higher divergence in their regulatory elements and higher transcriptional heterogeneity between individual cells. **Evolutio**n: Highly divergent genes have conserved promoter sequences, and have evolved faster in independent evolutionary modes - exhibiting faster coding sequence evolution, faster rate of gene gain and loss and a larger fraction of evolutionarily younger genes. **Physiology:** Highly divergent genes are enriched with functions related to cellular defence and inflammation, while lowly-divergent genes tend to be pleiotropic and to function in both immune and non-immune contexts, enriched with functions related to regulation, metabolism and apoptosis. Viruses tend to target conserved innate response genes (by either physical interactions or by molecular mimicry).

References:

Barber, M.F., Lee, E.M., Hayden, G., H., and Elde, N.C. (2016). Rapid evolution of primate type 2 immune response factors linked to asthma susceptibility. bioRxiv.

Barreiro, L.B., Marioni, J.C., Blekhman, R., Stephens, M., and Gilad, Y. (2010). Functional comparison of innate immune signaling pathways in primates. PLoS Genet *6*, e1001249.

Borden, E.C., Sen, G.C., Uze, G., Silverman, R.H., Ransohoff, R.M., Foster, G.R., and Stark, G.R. (2007). Interferons at age 50: past, current and future impact on biomedicine. Nat Rev Drug Discov *6*, 975-990.

Braun, B.A., Marcovitz, A., Camp, J.G., Jia, R., and Bejerano, G. (2015). Mx1 and Mx2 key antiviral proteins are surprisingly lost in toothed whales. Proc Natl Acad Sci U S A *112*, 8036-8040.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Brinkworth, J.F., and Barreiro, L.B. (2014). The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. Curr Opin Immunol *31*, 66-78.

Brunette, R.L., Young, J.M., Whitley, D.G., Brodsky, I.E., Malik, H.S., and Stetson, D.B. (2012). Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. J Exp Med *209*, 1969-1983.

Butler, M.A., and King, A.A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am Nat *164*, 683-695.

Calderone, A., Licata, L., and Cesareni, G. (2015). VirusMentha: a new resource for virus-host protein interactions. Nucleic Acids Res *43*, D588-592.

Capra, J.A., Williams, A.G., and Pollard, K.S. (2012). ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. PLoS Comput Biol *8*, e1002567.

Choo, S.W., Rayko, M., Tan, T.K., Hari, R., Komissarov, A., Wee, W.Y., Yurchenko, A.A., Kliver, S., Tamazian, G., Antunes, A.*, et al.* (2016). Pangolin genomes and the evolution of mammalian scales and immunity. Genome Res *26*, 1312-1322.

Crow, Y.J., and Manel, N. (2015). Aicardi-Goutieres syndrome and the type I interferonopathies. Nat Rev Immunol *15*, 429-440.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet *46*, 677-700.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol *12*, 687-695.

Dyer, M.D., Murali, T.M., and Sobral, B.W. (2008). The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog *4*, e32.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Enard, D., Cai, L., Gwennap, C., and Petrov, D.A. (2016). Viruses are a dominant driver of protein adaptation in mammals. Elife *5*.

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C.*, et al.* (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.

Fumagalli, M., and Sironi, M. (2014). Human genome variability, natural selection and infectious diseases. Curr Opin Immunol *30*, 9-16.
Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., and Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet *7*, e1002355.
Halehalli, R.R., and Nagarajaram, H.A. (2015). Molecular principles of human virus protein-protein interactions. Bioinformatics *31*, 1025-1033.
Hall, J.C., and Rosen, A. (2010). Type I interferons: crucial participants in disease amplification in autoimmunity. Nat Rev Rheumatol *6*, 40-49.
Haygood, R., Babbitt, C.C., Fedrigo, O., and Wray, G.A. (2010). Contrasts between adaptive coding and noncoding changes during human evolution. Proc Natl Acad Sci U S A *107*, 7853-7857.
Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M., Amode, R., Brent, S.*, et al.* (2016). Ensembl comparative genomics resources. Database (Oxford) *2016*.
Ivashkiv, L.B., and Donlin, L.T. (2014). Regulation of type I interferon responses. Nat Rev Immunol *14*, 36-49.
Iwasaki, A. (2012). A virological view of innate immune recognition. Annu Rev Microbiol *66*, 177-196.
Johnson, W.E., and Sawyer, S.L. (2009). Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics *61*, 163-176.
Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. Nature *468*, 811-814.
Khaitovich, P., Enard, W., Lachmann, M., and Paabo, S. (2006). Evolution of primate gene expression. Nat Rev Genet *7*, 693-702.
Kumar, V., Wijmenga, C., and Xavier, R.J. (2014). Genetics of immune-mediated disorders: from genome-wide association to molecular mechanism. Curr Opin Immunol *31*, 51-57.
Lee, M.N., Ye, C., Villani, A.C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K.*, et al.* (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science *343*, 1246980.
Levin, M., Anavy, L., Cole, A.G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D.H., Feder, M.*, et al.* (2016). The mid-developmental transition and the evolution of animal body plans. Nature *531*, 637-641.
Li, Y., Oosting, M., Smeekens, S.P., Jaeger, M., Aguirre-Gamboa, R., Le, K.T., Deelen, P., Ricano-Ponce, I., Schoffelen, T., Jansen, A.F.*, et al.* (2016). A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. Cell *167*, 1099-1110 e1014.
Malfavon-Borja, R., Wu, L.I., Emerman, M., and Malik, H.S. (2013). Birth, decay, and reconstruction of an ancient TRIMCyp gene fusion in primate genomes. Proc Natl Acad Sci U S A *110*, E583-592.
Mostafavi, S., Yoshida, H., Moodley, D., LeBoite, H., Rothamel, K., Raj, T., Ye, C.J., Chevrier, N., Zhang, S.Y., Feng, T.*, et al.* (2016). Parsing the Interferon Transcriptional Network and Its Disease Associations. Cell *164*, 564-578.
Nedelec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S.*, et al.* (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell *167*, 657-669 e621.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840-846.
Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J.*, et al.* (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol *3*, e170.
Palmeira, L., Penel, S., Lotteau, V., Rabourdin-Combe, C., and Gautier, C. (2011). PhEVER: a database for the global exploration of virus-host evolutionary relationships. Nucleic Acids Res *39*, D569-575.
Porritt, R.A., and Hertzog, P.J. (2015). Dynamic control of type I IFN signalling by an integrated network of negative regulators. Trends Immunol *36*, 150-160.
Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M.*, et al.* (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643-656 e617.
Roider, H.G., Lenhard, B., Kanhere, A., Haas, S.A., and Vingron, M. (2009). CpG-depleted promoters harbor tissue-specific transcription factor binding signals--implications for motif overrepresentation analyses. Nucleic Acids Res *37*, 6305-6315.
Sackton, T.B., Lazzaro, B.P., and Clark, A.G. (2017). Rapid expansion of immune-related gene families in the house fly, Musca domestica. Mol Biol Evol.
Saeed, R., and Deane, C.M. (2006). Protein protein interactions, evolutionary rate, abundance and age. BMC Bioinformatics *7*, 128.
Schroder, K., Irvine, K.M., Taylor, M.S., Bokil, N.J., Le Cao, K.A., Masterman, K.A., Labzin, L.I., Semple, C.A., Kapetanovic, R., Fairbairn, L.*, et al.* (2012). Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. Proc Natl Acad Sci U S A *109*, E944-953.
Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N.*, et al.* (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363-369.
Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A.*, et al.* (2013). Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proc Natl Acad Sci U S A *110*, 2946-2951.
Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P.*, et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res *43*, D447-452.
Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. Nat Genet *38*, 830-834.
Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. PLoS Comput Biol *11*, e1004333.
Xu, L., Yu, D., Fan, Y., Peng, L., Wu, Y., and Yao, Y.G. (2016). Loss of RIG-I leads to a functional replacement with MDA5 in the Chinese tree shrew. Proc Natl Acad Sci U S A *113*, 10950-10955.
Zhao, M., Zhang, J., Phatnani, H., Scheu, S., and Maniatis, T. (2012). Stochastic expression of the interferon-beta gene. PLoS Biol *10*, e1001249.

Fig 1

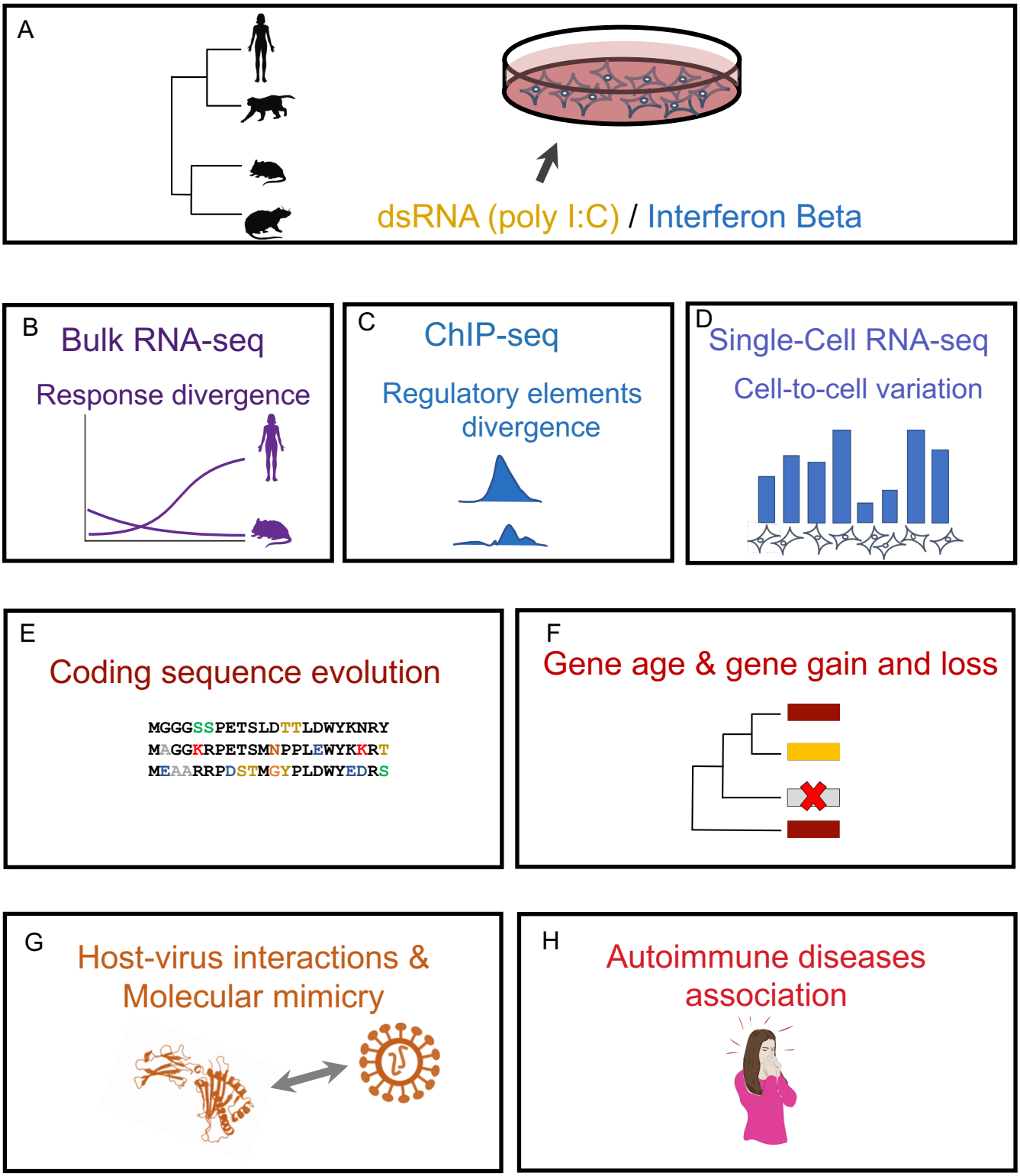

A
dsRNA (poly I:C) / Interferon Beta
B
Bulk RNA-seq
Response divergence
C
ChIP-seq
Regulatory elements divergence
D
Single-Cell RNA-seq
Cell-to-cell variation
E
Coding sequence evolution
MGGGSSPETSLDTTLDWYKNRY
MAGGKRPETSMNPPLEWYKKRT
MEAARRPDSTMGYPLDWYEDRS
F
Gene age & gene gain and loss
G
Host-virus interactions &
Molecular mimicry
H
Autoimmune diseases
association

Fig 2

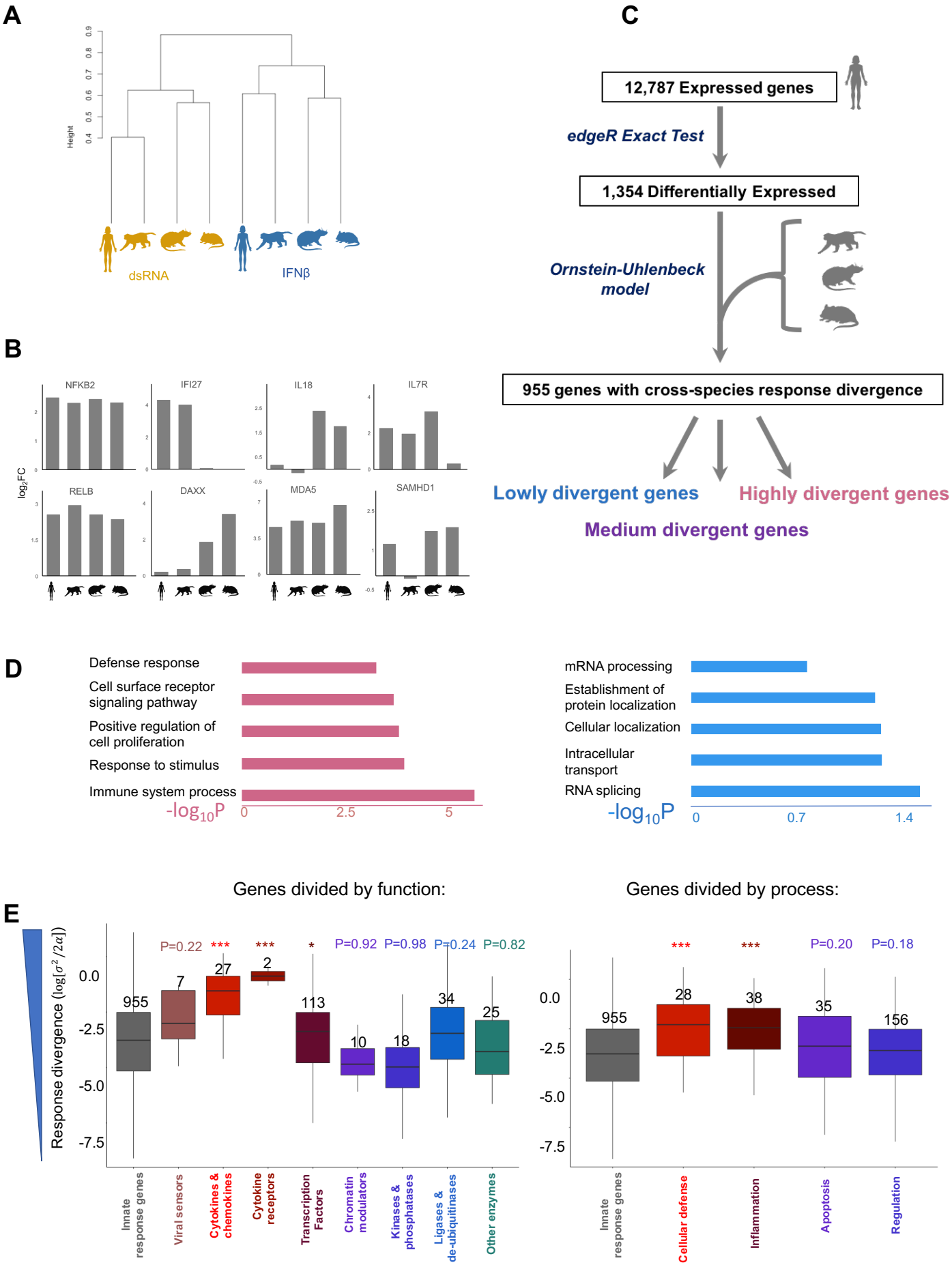
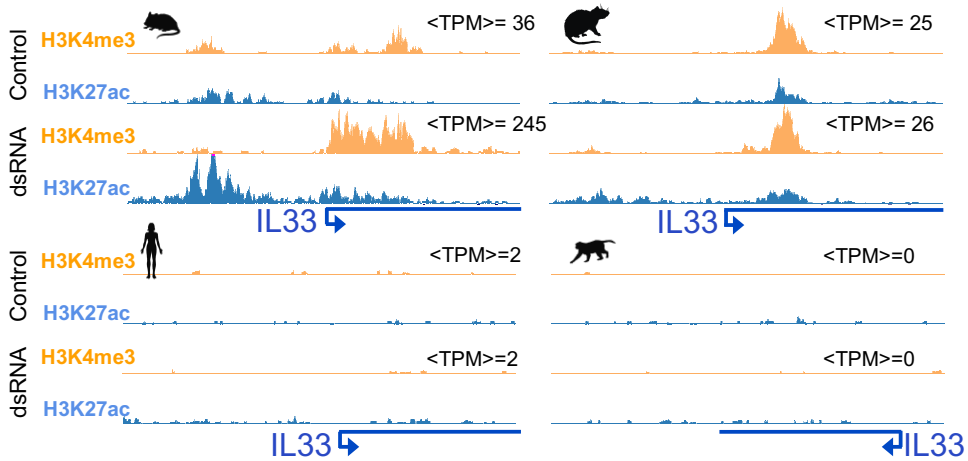
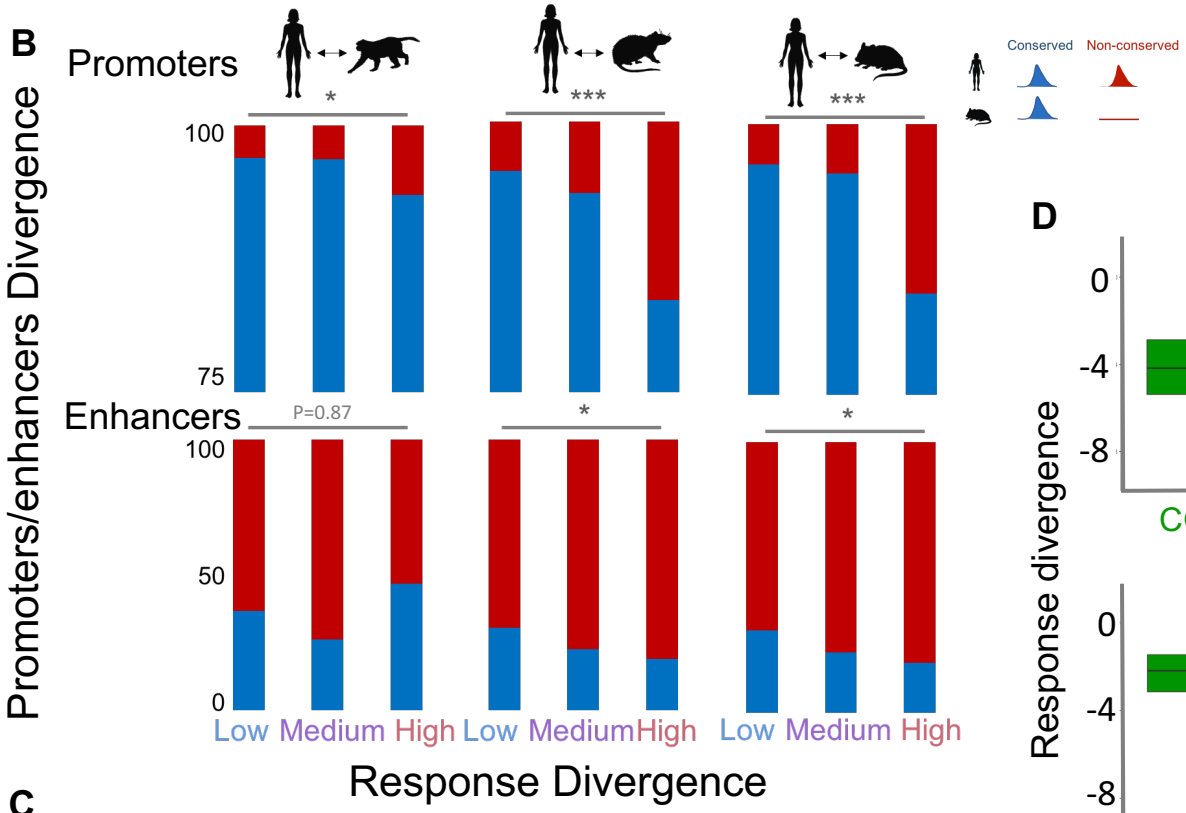

A
Height
dsRNA
IFNβ
B
NFKB2
IFI27
IL18
IL7R
RELB
DAXX
MDA5
SAMHD1
log2FC
C
12,787 Expressed genes
edgeR Exact Test
1,354 Differentially Expressed
Ornstein-Uhlenbeck model
955 genes with cross-species response divergence
Lowly divergent genes
Highly divergent genes
Medium divergent genes
D
Defense response
Cell surface receptor signaling pathway
Positive regulation of cell proliferation
Response to stimulus
Immune system process
-log10P
0
2.5
5
mRNA processing
Establishment of protein localization
Cellular localization
Intracellular transport
RNA splicing
-log10P
0
0.7
1.4
E
Genes divided by function:
Genes divided by process:
Response divergence (log[σ²/2α])
P=0.22
***
***
*
P=0.92
P=0.98
P=0.24
P=0.82
955
7
27
2
113
10
18
34
25
Innate response genes
Viral sensors
Cytokines & chemokines
Cytokine receptors
Transcription Factors
Chromatin modulators
Kinases & phosphatases
Ligases & de-ubiquitinases
Other enzymes
***
***
P=0.20
P=0.18
955
28
38
35
156
Innate response genes
Cellular defense
Inflammation
Apoptosis
Regulation

Fig 3

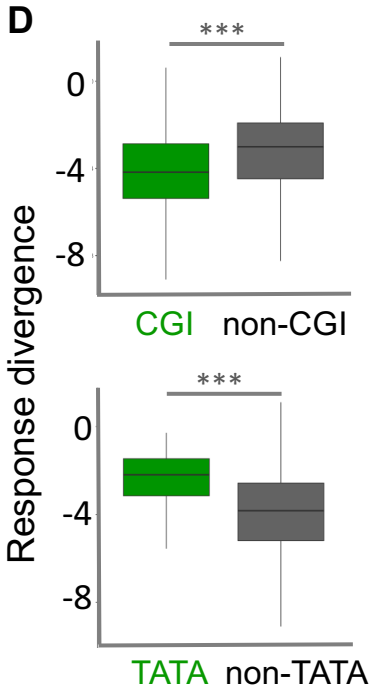
A


Control
dsRNA
H3K4me3
H3K27ac
<TPM>= 36
<TPM>= 25
<TPM>= 245
<TPM>= 26
IL33
<TPM>=2
<TPM>=0
IL33

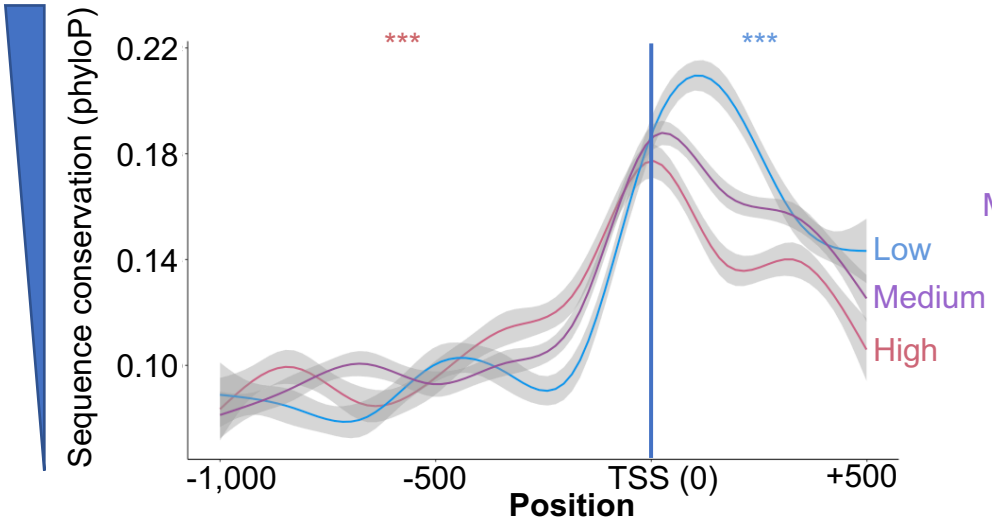
B


Promoters
Enhancers
Promoters/enhancers Divergence
Conserved
Non-conserved
P=0.87
Low Medium High Low Medium High Low Medium High
Response Divergence

D


Response divergence
CGI
non-CGI
TATA
non-TATA

C


Sequence conservation (phyloP)
Low
Medium
High
-1,000
-500
TSS (0)
+500
Position

E

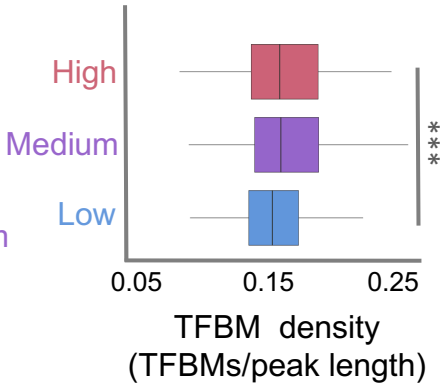
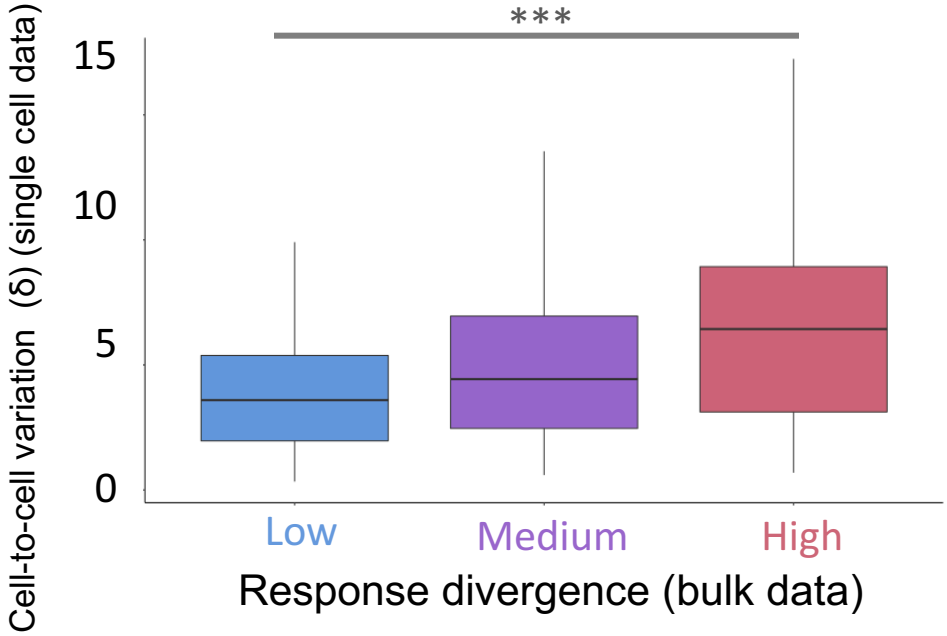
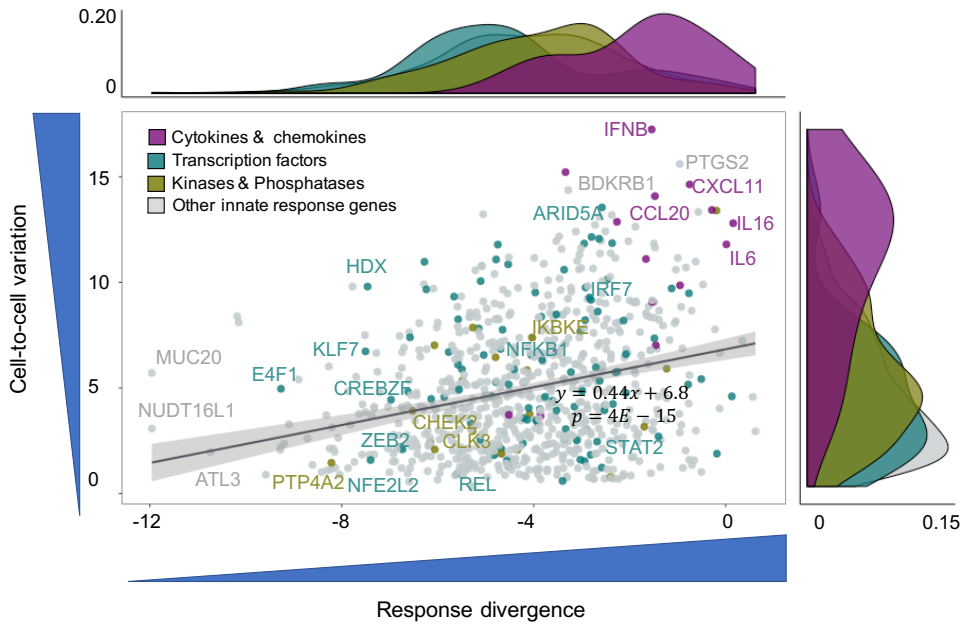

High
Medium
Low
TFBM density
(TFBMs/peak length)

Fig 4

A


***
15
10
5
0
Cell-to-cell variation (δ) (single cell data)
Low
Medium
High
Response divergence (bulk data)

B


0.20
0
Cytokines & chemokines
Transcription factors
Kinases & Phosphatases
Other innate response genes
IFNB
PTGS2
BDKRB1
CXCL11
ARID5A
CCL20
IL16
IL6
HDX
IRF7
IKBKE
NFKB1
KLF7
MUC20
E4F1
CREBZF
NUDT16L1
CHEK2
ZEB2
CLK3
STAT2
ATL3
PTP4A2
NFE2L2
REL
$y = 0.44x + 6.8$
$p = 4E - 15$
Cell-to-cell variation
15
10
5
0
-12
-8
-4
0
0
0.15
Response divergence

C

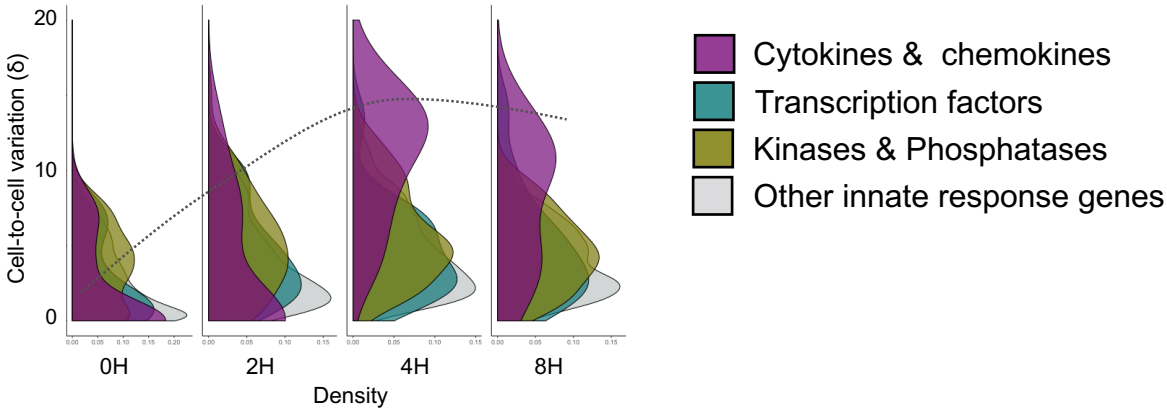

Cell-to-cell variation (δ)
20
10
0
0H
2H
4H
8H
Density
Cytokines & chemokines
Transcription factors
Kinases & Phosphatases
Other innate response genes

Fig 5

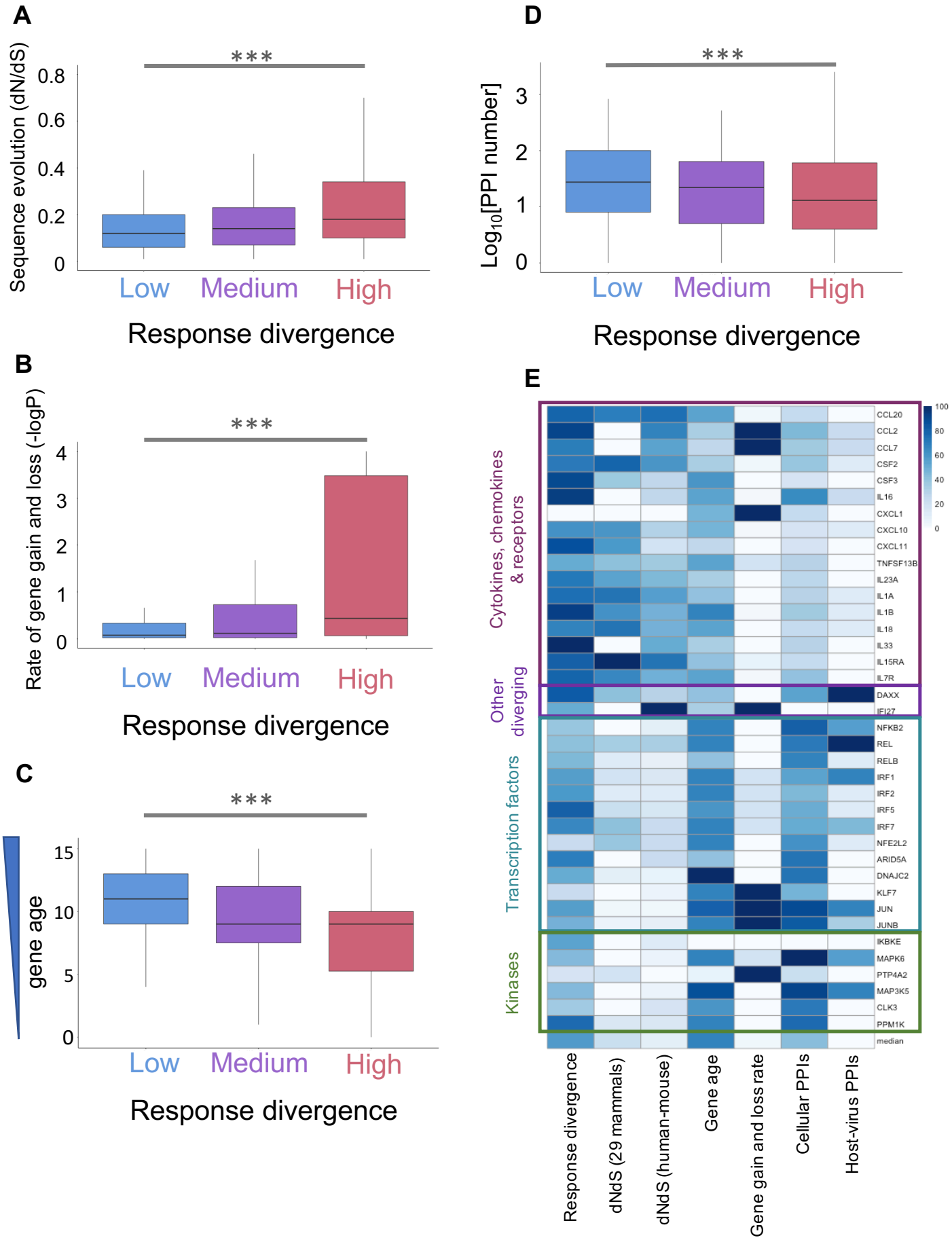

A
***
Sequence evolution (dN/dS)
0.8
0.6
0.4
0.2
0
Low
Medium
High
Response divergence
D
***
Log10[PPI number]
3
2
1
0
Low
Medium
High
Response divergence
B
***
Rate of gene gain and loss (-logP)
4
3
2
1
0
Low
Medium
High
Response divergence
C
***
gene age
15
10
5
0
Low
Medium
High
Response divergence
E
Cytokines, chemokines & receptors
Other diverging
Transcription factors
Kinases
CCL20
CCL2
CCL7
CSF2
CSF3
IL16
CXCL1
CXCL10
CXCL11
TNFSF13B
IL23A
IL1A
IL1B
IL18
IL33
IL15RA
IL7R
DAXX
IFI27
NFKB2
REL
RELB
IRF1
IRF2
IRF5
IRF7
NFE2L2
ARID5A
DNAJC2
KLF7
JUN
JUNB
IKBKE
MAPK6
PTP4A2
MAP3K5
CLK3
PPM1K
median
100
80
60
40
20
0
Response divergence
dNdS (29 mammals)
dNdS (human-mouse)
Gene age
Gene gain and loss rate
Cellular PPIs
Host-virus PPIs

Fig 6

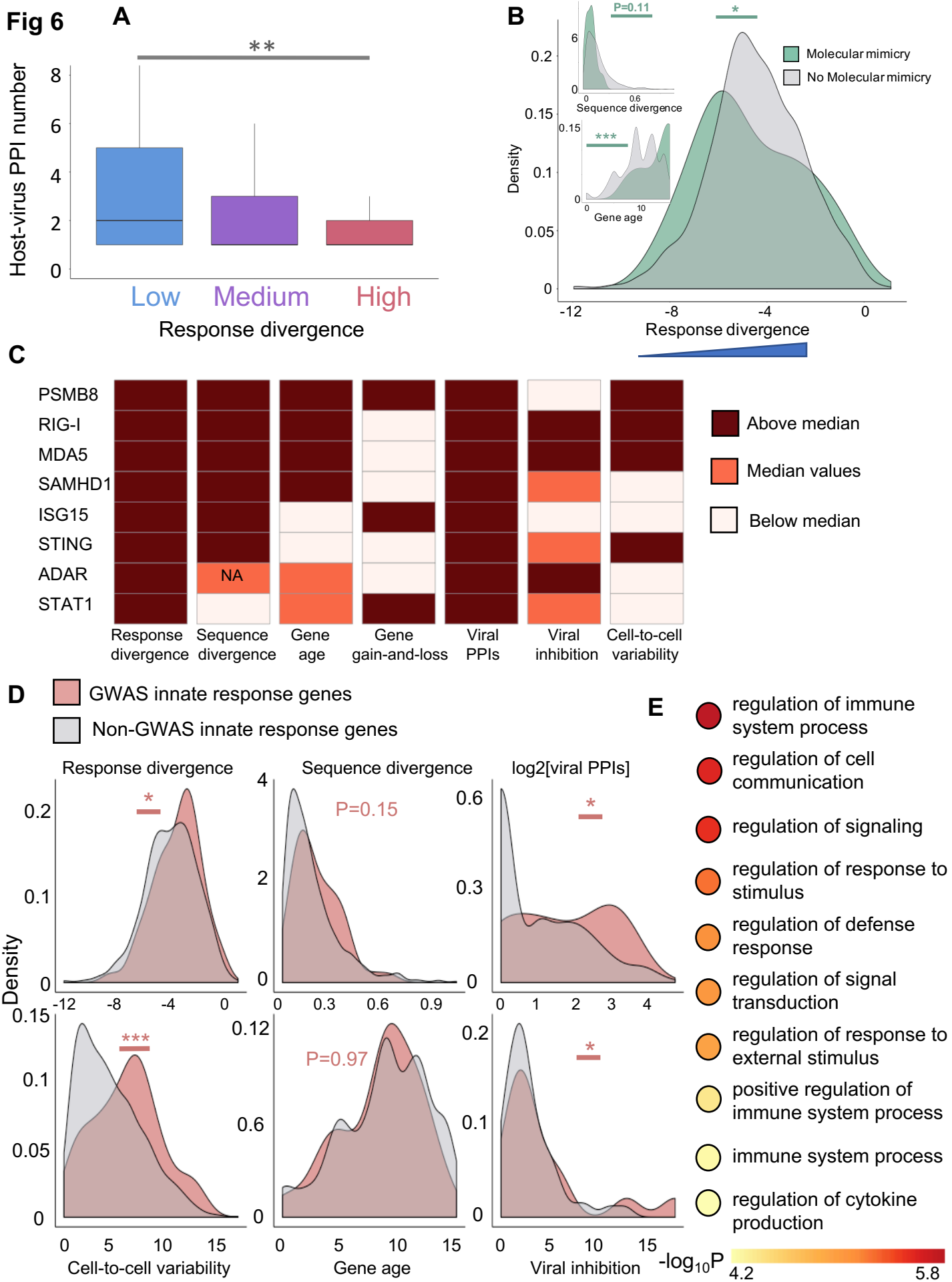

A
**
Host-virus PPI number
8
6
4
2
0
Low
Medium
High
Response divergence
B
P=0.11
*
Molecular mimicry
No Molecular mimicry
Sequence divergence
***
Gene age
Density
0.2
0.15
0.1
0.05
0
-12
-8
-4
0
Response divergence
C
PSMB8
RIG-I
MDA5
SAMHD1
ISG15
STING
ADAR
STAT1
NA
Response divergence
Sequence divergence
Gene age
Gene gain-and-loss
Viral PPIs
Viral inhibition
Cell-to-cell variability
Above median
Median values
Below median
D
GWAS innate response genes
Non-GWAS innate response genes
Response divergence
Sequence divergence
log2[viral PPIs]
P=0.15
P=0.97
Density
Cell-to-cell variability
Gene age
Viral inhibition
E
regulation of immune system process
regulation of cell communication
regulation of signaling
regulation of response to stimulus
regulation of defense response
regulation of signal transduction
regulation of response to external stimulus
positive regulation of immune system process
immune system process
regulation of cytokine production
-log10P
4.2
5.8

Fig 7

Innate Immune Response

High Response divergence

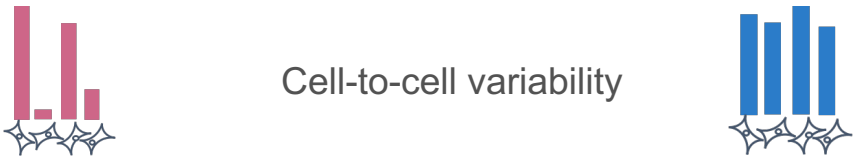
Low Response divergence

Expression & Regulation:

High

Response Divergence &
Divergence in regulatory elements

Low



High

Cell-to-cell variability

Low

Evolution

AAGAGCTTAATAGCGT
ACGAGCTTAATAGCAT

Slow

Promoter Sequence
Evolution

GAGAGGTTACCGGCAT
ACGAGCTTGATAATAT

Fast

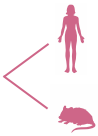
MAGGKRPETSMNPPLE
MEAARRPDSTMGYPLD

Fast

Coding Sequence
Evolution

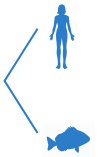
MGGGRRPETSMNPPLD
MGGGRRPDSSMDYPLD

Slow



Younger

Gene Age



Older

Physiology

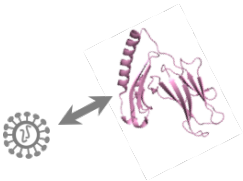
Exclusive immune functions:

- Cellular defense
- Inflammation

Gene Function

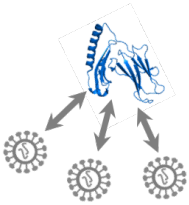
Pleiotropic genes:

- Metabolism
- Apoptosis
- Regulation



Low

Virus Antagonism



High