

## A hypothetical new role for single-stranded DNA binding proteins in the immune system

Nagarjun Vijay<sup>1</sup>, Ajit Chande<sup>2</sup>

<sup>1</sup>Computational Evolutionary Genomics Lab, <sup>2</sup>Virology and Genome Editing Laboratory, Department of Biological Sciences, IISER Bhopal, Madhya Pradesh, India

\*Correspondence: [nagarjun@iiserb.ac.in](mailto:nagarjun@iiserb.ac.in)

**Abstract:** The breadth of the host range of single-stranded DNA (ssDNA) viruses is roughly comparable to the host range of double-stranded DNA viruses (dsDNA). Yet, general ssDNA sensing receptors that activate the immune system have not been unequivocally identified while numerous dsDNA sensing receptors are known. Here, we hypothesize that some of the Single-Stranded DNA Binding (SSB) proteins may act as receptors that detect single-stranded DNA from pathogens and activate the innate immune system. As the first test of our hypothesis, we checked whether human genes that are known to bind to ssDNA are potentially interferon-regulated. Out of the 102 human genes that are known to have ssDNA binding ability 23 genes show a more than two-fold increase in gene expression upon interferon treatment. Single-stranded DNA viruses are pathogens of not only animals but also of plants and protozoans. We used this information to further prioritize our candidate list to ssDNA binding genes that are common between the model plant *Arabidopsis thaliana* and humans. Based on these strategies, we shortlist several promising candidate genes including the HMGB1 gene which could act as a ssDNA sensor that activates the immune system. Agreeably though we cannot establish a definitive role for these genes as ssDNA sensors of the immune system as yet, our preliminary analysis suggests the potential existence of ssDNA binding protein-like receptors (SLR's) that are worth investigating further.

**Keywords:** innate sensors; DNA sensing; DNA virus infection; interferon response; DNA binding proteins

### Introduction

The innate immune system has been shown to consist of numerous receptors that are capable of recognizing pathogen associated molecules. These receptors are called pattern recognition receptors (PRRs) and the non-self-microbial molecules that these receptors recognize are

called pathogen-associated molecular patterns (PAMPs). PRRs can detect a diverse set of ligands such as nucleic acids, proteins, carbohydrates, peptides, compound molecules, lipopolysaccharides, peptidoglycan, lipopeptides, phospholipids and glycoproteins (1, 2). After recognising the ligand molecule the receptors initiate downstream signalling to activate the immune response.

Within nucleic acid recognizing receptors, specific genes that detect ssRNA (single-stranded RNA), dsRNA (double-stranded RNA), CpG DNA, poly-nucleotide stretches, and DNA have been identified (3, 4). A considerable number of DNA sensing immune receptors have been identified in the past two decades (5). Receptors that specifically recognize ssDNA in a sequence independent manner have not been reported. However, certain ssDNA molecules that have an AT-rich stem-loop structure are thought to directly activate the immune response (6). Although it has been suggested that detection of ssDNA viruses might be possible through TLR9 (7) and IFI16 (8), it is unclear whether this is a general mechanism (9).

### **The hypothesis**

The Baltimore classification of viruses is based on the nucleic acid (RNA or DNA) used as the genetic material, presence of single or double strandedness, coding strand and mode of replication (10). Based on this classification system seven groups of viruses are recognised. Pathogen recognition receptors that are able to specifically identify dsDNA, dsRNA, ssRNA (both sense and antisense) have been experimentally characterised (3, 4). The gene TREX1 is able to degrade reverse transcribed DNA and the lack of the gene leads to build up of ssDNA generated by endogenous retro-elements (11). Infact, mutations in the TREX1 gene have been linked to initiation of autoimmune response (12, 13). The ability of TREX1 to degrade reverse transcribed DNA has been attributed to its exonuclease activity (14). Although TREX1 is the most highly up-regulated gene upon interferon treatment that also has ssDNA binding it cannot be considered to be an ssDNA sensor as it does not initiate a downstream response. The closest candidate for a true ssDNA sensor has been reported to be able to detect AT-rich ssDNA molecules (6). The well-known DNA sensing receptor IFI16 has also been proposed as a ssDNA sensor due to its ability to detect ssDNA stem loop structure (8). The existing literature on nucleic acid sensors has not been able to conclusively establish the existence of a true generic ssDNA sensor (15). It is possible that DNA sensors that specifically recognise the single stranded form of the DNA molecule don't exist and the dsDNA sensors themselves are able to recognise ssDNA (16). However, this potential ability

of dsDNA sensors to identify ssDNA molecules and initiating a immune response has also not been established (9).

Large scale profiling of meta-genomes has started revealing the true diversity of various classes of viruses by profiling previously inaccessible niches (17, 18). This has bolstered the idea that ssDNA viruses are probably more widespread among animal and plant species than previously estimated (19). New ssDNA viruses are also being discovered in host species such as archaea (20) and diatoms (21). The niches occupied by ssDNA viruses overlaps with that of other types of viruses enough for ssDNA viruses to have acquired genes from these other viruses through horizontal gene transfer to form a new family of ssDNA viruses (22). Since pathogen recognition receptors exist for every other class of viruses defined by the Baltimore classification, it seems unexpected that ssDNA sensing receptors have not been identified. If the ssDNA viruses were extremely restricted in their host range and/or ability to infect it could have explained this anomaly. However, the breadth of the host range of ssDNA viruses seems comparable to that of dsDNA viruses. This lack of general ssDNA sensors motivated us to hypothesize the existence of another class of PRRs that could act as receptors to specifically recognize ssDNA from pathogens such as ssDNA viruses. We call this hypothetical new class of receptors SLR's (ssDNA binding protein-Like Receptors). Based on this hypothesis we make four predictions, (1) These receptors should possess ability to bind single stranded DNA, (2) Be regulated or responsive to the interferons or pro-inflammatory gene products, (3) These genes should be evolutionarily conserved at least across species that are permissive to the ssDNA viruses, and (4) Potentially be under episodic positive selection.

### **Evaluation of the hypothesis – preliminary data analysis**

To evaluate whether our hypothesis is a feasible scenario we searched public datasets and performed preliminary computational data analysis based on the four predictions from our hypothesis. We reasoned that an ssDNA sensor should probably have the ability to bind ssDNA. Genes with ssDNA binding ability have been assigned the Gene Ontology (GO) accession GO: 0003697. A total of 102 genes in the human genome come under this GO classification (see **Supplementary Table S1**). However, only 19 of these 102 genes have an oligonucleotide binding motif, or the OB fold. The OB fold (Interpro # IPR012340) is a five-stranded beta-barrel which is found in ss nucleic acid binding proteins. This OB fold structure

is found in proteins with a wide range of functions such as DNA replication & repair genes (23).

The OB fold structure was absent from a majority of the genes with ssDNA binding activity we looked for other domains that are common among these genes. The P-loop NTPase fold (Interpro# IPR027417) domain was found in 20 of the genes. The P-loop NTPase fold catalyses NTP hydrolysis and utilises the energy released to induce conformational changes in other molecules. This catalytic activity makes genes with P-loop NTPase fold potential candidates for acting as receptors of ssDNA that possibly signal downstream molecules through conformational changes.

In addition to the ssDNA binding ability, pathogen recognition receptors should have the ability to increase their abundance upon a pathogenic attack to combat the greater number of pathogenic molecules. A vast majority of immune genes that are required to combat infection are regulated by a group of signalling proteins called interferons that are released in response to the presence of pathogens. These signalling proteins activate a group of transcription factor genes called signal transducer and activator of transcription (STAT) complexes and Interferon regulatory factors (IRF) to regulate the expression of various genes associated with the immune system (24). The complete set of genes that are directly induced by interferons or regulated indirectly is still being explored through genome-wide screens (25, 26). Publicly available gene expression datasets of interferon induced samples have been compiled and annotated to identify genes whose expression changes in response to interferon treatment. The Interferome database is an actively maintained comprehensive annotated collection of IFN-regulated genes (27, 28). We searched the Interferome database for the genes from our list of ssDNA binding proteins. However, presence of a gene in the Interferome database is just suggestive and not conclusive evidence for the gene being interferon regulated or not. Out of the 102 genes that are thought to have ssDNA binding ability, we found that 23 showed more than two-fold increase in gene expression upon interferon treatment. The type-1 interferon only regulation was seen for 4 genes, type-2 interferon only regulation was seen for 8 genes and 11 genes were found to be up-regulated independently by type-1 or type-2 interferon treatment (see **Supplementary Table S1** for details).

Pattern recognition receptors (PRRs) that detect DNA as well as those that detect RNA have been shown to be under positive selection (29–33), some have undergone recurrent whole gene or exon level losses and duplications (34–36). Prevalence of different types of

pathogens has been a strong driver of evolutionary change in immune genes. The vast majority of genes that show strong signatures of adaptive evolution happen to be involved in the immune system in diverse species (37, 38) including humans (39). The hypersensitive response (HR) in plants is activated by resistance genes that code for receptor like proteins. Even these resistance genes found in host plants have been the targets of positive selection more often than other genes (40) and have been used to prioritise candidate resistance genes (41). Hence, genes that have experienced strong positive selection are likely to be involved in immune-related functions. This same line of reasoning has been used by previous studies to short-list immune related genes by looking at signatures of selection (42, 43). Only one of the 102 genes that have ssDNA binding ability is detected to be under positive selection in primates after using very conservative data analysis criteria (33). Surprisingly, we found that this gene contains a leucine-rich PPR motif (LR PPR Containing). The toll-like receptors that are involved in pathogen recognition for the innate immune system also have numerous successive leucine-rich repeat motifs but do not include the LRPPRC gene. Similarly, in plants, nucleotide binding site (NBS) - LRR gene family members are involved in activating the hypersensitive response (41, 44). The function of LRPPRC is only beginning to be understood and it has been implicated in suppressing autophagy. This gene has been implicated in suppressing autophagy and mitophagy by sequestering autophagy regulators Beclin 1 and Bcl-2 and the mitophagy regulator Parkin (45, 46). It has also been shown that the viral restriction factor Tetherin interacts with LRPPRC to activate autophagy (47).

Circular ssDNA viruses of the family Geminiviridae are considered to form a large fraction of plant pathogens (48). The flood of new data that is being generated from metagenomics screens has shown that circular ssDNA viruses have much greater diversity than previously estimated (19). However, plant genomes don't have the homologous gene of LRPPRC. This could suggest the existence of other LRPPRC like genes as an alternative mechanism for ssDNA sensing in plants or that LRPPRC is not really involved in ssDNA sensing at all. Hence, as a fourth criteria we shortlisted genes with ssDNA binding ability in humans and the model plant *Arabidopsis thaliana*. We preferred to use a model plant species as it is likely to have more reliable annotation and would probably have been used for assigning GO terms in other species (49). Further restricting the gene set of 23 genes to those present in the plant *Arabidopsis thaliana* we are left with 2 genes (see **Supplementary Table S1**). We also choose few top candidates based on their similarity to known dsDNA sensors and ability to initiate a downstream immune response. In addition to these top candidates we also discuss

few groups of genes that could also be potentially good candidates based on their conservation across animals and plants.

While experimental validation of the candidates would be a definitive test of the hypothesis proposed here, such experimentation is challenging following the existence of various other immune mechanisms for pathogen detection. Many of the candidates that we identify are no doubt involved in sensing Damage Associated Molecular Patterns (DAMP's), however, the more important point that needs to be established is the ability of these sensors in being able to distinguish self from non-self-molecules. Whether the sensors are able to utilise the differences in base compositions of host vs pathogen DNA content in a spatio-temporal manner is an interesting idea worth exploring. Disentangling the role played by SLR's if any is an arduous task. We simply provide a list of candidate genes which we prioritised based on the criteria described above. This gene list is not intended as definitive evidence for the existence of SLR's. However, we hope that our hypothesis and preliminary reasoning based data analysis would motivate a more long-term search for the existence of SLRs.

### **Empirical data – top candidate genes for single stranded DNA sensing**

1. **HMGB1:** The HMGB1 gene has been the focus of numerous papers (50–54) and it has even been called the “nuclear weapon in the immune arsenal” (55). The role of HMGB1 in autoimmune diseases is also established (56). In addition to its role in the immune system, HMGB1 is also involved in regulation of gene expression (57). Studies in animals and plants have established the HMGB family of genes as damage-associated molecular pattern (DAMP) molecules (58, 59). It should be also be noted however that the HMGB1 gene has been proposed as a nucleic acid sensor that mediates an immune response (60, 61). However, a specific ability to detect various forms in which ssDNA could exist has not been explicitly discussed (62).

HMGB1 has also been shown to have affinity for different DNA structures (52). Interestingly it has been shown to bind DNA mini-circles (63), looped structures (64) as well as hemi-catenated DNA (65) in addition to four-way junctions (66) and triplex DNA (67). HMGB1 is also one of the five genes in the human genome that has been assigned the GO term of bubble DNA binding (GO: 0000405). This broad range of DNA binding ability of HMGB1 and its established role in the immune response have not been linked with each other. Since the discovery of HMGB1's role in DNA

damage response has largely relegated it to the status of a DAMP (68). Single-stranded DNA viruses have many forms such as circular ssDNA found in many plant viruses as well as secondary-structures of linear ssDNA (69). It could be argued that the ability of HMGB1 to bind such a broad range of DNA structures makes it an ideal candidate to be an ssDNA sensor. In addition to being present in Interferome database, it has experimentally been shown that HMGB1 is actually interferon induced (70). Being a widely conserved protein it is present across animals, plants and protozoans. Given all these attributes HMGB1 would probably be the best candidate for being a SLR.

2. **DDX11 & DHX9**: Both these genes are part of the DEAD box protein family that is mostly involved in anti-viral innate immune response (71, 72). Contradictory to the role of other DEAD box genes, the DHX9 gene as well as DDX1 and DDX3 are known to be required for replication of numerous viruses (5). The retention of these genes that assists pathogens in the host genome is hard to explain. DDX3 is known to activate the immune system through MAVS (73). It has been suggested that these genes are potentially the target of the ongoing ‘arms race’ and have a role in innate immunity that might be exploited by pathogens (5). However, despite being part of the DEAD box protein family DDX11 has not been clearly linked to the innate immune response pathway.
3. **RPA1, RPA2, RPA3 & RPA4**: The Replication Protein A (RPA) is a protein complex involved in sensing DNA damage through association with ssDNA (74). It has already been proposed to activate a downstream immune response (75). All four RPA’s have ssDNA binding ability, but none of them are upregulated upon interferon treatment (see **Supplementary Table 1**).

### **Other single stranded DNA sensor candidates**

In addition to the top candidate genes discussed in the previous section, we believe the following set of genes could also be involved in innate immune sensing of single stranded DNA from pathogens. Based on their function and evolutionary relationships we have grouped these genes together.

1. The Mini chromosome Maintenance protein complex (MCM): The MCM complex in eukaryotes has six genes, four of which are thought to be able to bind to ssDNA with only three of them having annotated OB fold domains. Only two of the six MCM's with ssDNA binding ability showed a greater than two-fold increase in expression upon interferon treatment and also have annotated orthologs in *A. thaliana*. However, two other MCM genes show a more than two-fold decrease in gene expression. One of these two genes, MCM7 has been shown to play a role in the innate immune system through phagocytosis (76).
2. Cold shock domain containing proteins: The cold shock domain containing proteins YBX1 and YBX3 have the ability to bind ssDNA. While YBX3 gene showed a greater than two-fold increase in gene expression upon interferon treatment, YBX1 gene showed decrease in gene expression. However, a recent report has suggested that YBX1 might play the role of a receptor to detect bacterial cell wall components and activate the innate immune system (77).
3. Nucleic Acid Binding Proteins (NABP): The human genome has two genes in the NABP family. We see that NABP1 shows a greater than two-fold increase in expression upon interferon treatment while NABP2 does not. However, both genes contain the OB fold domain and are part of the Sensor of Single-Strand DNA Complex (SOSS Complex). The SOSS complex is a multiprotein complex involved in DNA repair and cell checkpoint activation and maintenance of genomic stability. NABP1 is also known to play an important role in regulating thymopoiesis and is expressed in immune cells (78). Since NABP's can't bind dsDNA they would provide specificity for ssDNA sensing.

### **Consequences of the hypothesis**

The evaluation of the hypothesis proposed here has far reaching implications. First, if the existence and identity of SLR's are confirmed, this would greatly improve our understanding of the functioning of the innate immune system that integrates ssDNA sensing with immune activation. Signal initiators that act downstream from these receptors might reveal the existence of additional layers of regulation of the innate immune pathways. Many of the candidate genes that we identify as single stranded DNA sensors are part of the DNA damage response pathway. A similar pattern has been seen for dsDNA sensors that have been

identified. Experimental validation of the hypothesised SLR's could further reinforce the idea of co-option of the DNA damage response genes by the immune system (75).

One of our candidates for being an ssDNA sensor is the LRPPRC gene. A recently published study has identified 3 amino-acid residues (located at position 478, 889 and 1139) from the gene to be under positive selection among primates. All three of the positively selected residues are located outside the predicted LRR domains. The presence of positively selected residues outside the LRR domains in plant resistance genes has been interpreted as changes in pathogen recognition regions (41). Further investigation of the functional role of these positively selected residues would help clarify this. Another leucine rich repeat motif containing gene LRRFIP1 (Leucine-rich repeat flightless-interacting protein 1) has been shown to be a cytosolic nucleic acid sensor (79). LRRFIP1 activates the interferon response through its downstream partner beta-catenin. Similar to this dsDNA sensor our candidate gene LRPPRC could act as a receptor that activates downstream signalling through transcriptional regulation of MDR1 and MVP (80). The MVP gene codes for the Major Vault Protein. Multiple molecules of this protein come together to form larger sub-cellular structures called vaults. The function of these vaults has been a mystery despite the various apoptosis related functions that have been attributed to it (81). Recent work has established an important role for the Major Vault Protein in virus induced interferon response (82, 83). Hence, the interaction of LRRPC with MVP provides a potential new pathway for the activation of innate immunity.

The HMGB1 gene is well known for its role as a non-histone chromatin protein involved in gene regulation (84) and also as a mediator of immune response (85). Our preliminary analysis has suggested a yet another role for this HMGB1 gene as a single stranded DNA sensor. The downstream signalling cascade and initiation of immune response through the Toll-like receptor 4 as well as cGAS pathway have been characterised in detail (86–88). The fact that we find the HMGB1 gene from our screening of the databases could even be taken as evidence that our approach has probably been successful in identifying true SLR's.

Among the various viral sensing mechanisms that have been discovered, Dicers acting through the RNA interference (RNAi) pathway are considered ancient enough to have existed in the LECA (Last eukaryotic common ancestor). The RIG-I-like receptors (RLRs) are a more recent innovation that is known to be present only in vertebrates. The Toll-like receptors (TLRs) potentially originated in the early Bilateria and are older than RLR's (89).

Interestingly, many plants have a greater diversity of genes in the Dicer pathway that are able to respond to a diverse set of viral infections (90). Hence, the ancestral Dicer mediated anti-viral response of plants is much more powerful than that found in other eukaryotes.

If our hypothesis is wrong, it suggests the intriguing possibility that ssDNA sensing might be done by means other than a dedicated ssDNA sensing receptor. One possibility is that the immune system is directly activated by the recognition of the DNA damage response complex which is activated by ssDNA. This would require some way to distinguish self vs non-self ssDNA mediated DNA damage response. Another possible mode of ssDNA sensing could be through the formation of dsDNA by complementation of the ssDNA of the pathogen followed by recognition by dsDNA sensors. It has been shown that ssDNA viruses have used various mechanisms for insertion into their host genomes (91, 92). Transcription of such endogenous viral elements could lead to complementation of ssDNA and formation of DNA-RNA hybrids that would get detected by TLR9 followed by activation of the cGAS-STING pathway (93, 94). Single stranded DNA is always protected by the binding of ssDNA binding proteins. Hence, direct recognition of ssDNA might not be the best strategy for the detection of ssDNA from pathogens. Hence, it is possible that the single stranded DNA binding proteins or their complexes with ssDNA can be activators of the immune system. The Replication Protein A genes are promising candidates in this regard. In any case, this hypothesis provides interesting candidate single stranded DNA binding protein like receptor genes by novel thinking about pathogen recognition receptors (PRRs).

### **Conflict of Interest**

The authors declare no conflicts of interest.

### **Acknowledgement**

NJ would like to acknowledge funding from IISER Bhopal under grant # INST/BIO/2017/019. AC is a recipient of Ramanujan Fellowship from the SERB, Department of Science and Technology, and an Innovative Young Biotechnologist Award from the Department of Biotechnology, (both Government of India). AC thanks IISER Bhopal for funding under grant # INST/BIO/2017/012.

## References

1. Lata S, Raghava G. 2008. PRRDB: A comprehensive database of pattern-recognition receptors and their ligands. *BMC Genomics* 9.
2. Kumar H, Kawai T, Akira S. 2011. Pathogen recognition by the innate immune system. *Int Rev Immunol* 30:16–34.
3. Takeuchi O, Akira S. 2010. Pattern Recognition Receptors and Inflammation. *Cell*.
4. Kawasaki T, Kawai T, Akira S. 2011. Recognition of nucleic acids by pattern-recognition receptors and its relevance in autoimmunity. *Immunol Rev*.
5. Dempsey A, Bowie AG. 2015. Innate immune recognition of DNA: A recent history. *Virology*.
6. Sharma S, DeOliveira RB, Kalantari P, Parroche P, Goutagny N, Jiang Z, Chan J, Bartholomeu DC, Lauw F, Hall JP, Barber GN, Gazzinelli RT, Fitzgerald KA, Golenbock DT. 2011. Innate Immune Recognition of an AT-Rich Stem-Loop DNA Motif in the *Plasmodium falciparum* Genome. *Immunity* 35:194–207.
7. Pohar J, Lainšček D, Ivičak-Kocjan K, Cajnko M-M, Jerala R, Benčina M. 2017. Short single-stranded DNA degradation products augment the activation of Toll-like receptor 9. *Nat Commun* 8:15363.
8. Christensen MH, Paludan SR. 2017. Viral evasion of DNA-stimulated innate immune responses. *Cell Mol Immunol*.
9. Mattei LM, Cotmore SF, Li L, Tattersall P, Iwasaki A. 2013. Toll-Like Receptor 9 in Plasmacytoid Dendritic Cells Fails To Detect Parvoviruses. *J Virol* 87:3605–3608.
10. Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol Rev* 35:235–241.
11. Stetson DB, Ko JS, Heidmann T, Medzhitov R. 2008. Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* 134:587–598.
12. Peschke K, Achleitner M, Frenzel K, Gerbaulet A, Ada SR, Zeller N, Lienenklaus S, Lesche M, Poulet C, Naumann R, Dahl A, Ravens U, Günther C, Müller W, Knobloch K-P, Prinz M, Roers A, Behrendt R. 2016. Loss of Trex1 in Dendritic Cells Is Sufficient To Trigger Systemic Autoimmunity. *J Immunol* 197:2157–2166.
13. Ablasser A, Hemmerling I, Schmid-Burgk JL, Behrendt R, Roers A, Hornung V. 2014. TREX1 Deficiency Triggers Cell-Autonomous Immunity in a cGAS-Dependent

- Manner. *J Immunol* 192:5993–5997.
14. Yang YG, Lindahl T, Barnes DE. 2007. Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. *Cell* 131:873–886.
  15. Mogensen TH, Melchjorsen J, Larsen CS, Paludan SR. 2010. Innate immune recognition and activation during HIV infection. *Retrovirology*.
  16. Paludan SR. 2015. Activation and Regulation of DNA-Driven Immune Responses. *Microbiol Mol Biol Rev* 79:225–241.
  17. Labonté JM, Suttle CA. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* 7:2169–2177.
  18. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Eiprocta). *J Gen Virol* 93:2668–2681.
  19. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch Virol*.
  20. Mochizuki T, Krupovic M, Pehau-Arnaudet G, Sako Y, Forterre P, Prangishvili D. 2012. Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc Natl Acad Sci* 109:13386–13391.
  21. Tomaru Y, Toyoda K, Suzuki H, Nagumo T, Kimura K, Takao Y. 2013. New single-stranded DNA virus with a unique genomic structure that infects marine diatom *Chaetoceros setoensis*. *Sci Rep* 3.
  22. Krupovic M, Koonin E V. 2014. Evolution of eukaryotic single-stranded DNA viruses of the Bidnaviridae family from genes of four other groups of widely different viruses. *Sci Rep* 4.
  23. Flynn RL, Zou L. 2010. Oligonucleotide/oligosaccharide-binding fold proteins: A growing family of genome guardians. *Crit Rev Biochem Mol Biol*.
  24. Schneider WM, Chevillotte MD, Rice CM. 2014. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu Rev Immunol* 32:513–545.
  25. Schoggins JW, Rice CM. 2011. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol* 1:519–525.
  26. Varela M, Diaz-Rosales P, Pereiro P, Forn-Cuní G, Costa MM, Dios S, Romero A,

- Figueras A, Novoa B. 2014. Interferon-induced genes of the expanded IFIT family show conserved antiviral activities in non-mammalian species. *PLoS One* 9.
27. Samarajiwa SA, Forster S, Auchetl K, Hertzog PJ. 2009. INTERFEROME: The database of interferon regulated genes. *Nucleic Acids Res* 37.
28. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, Chapman R, Hertzog PJ. 2013. INTERFEROME v2.0: An updated database of annotated interferon-regulated genes. *Nucleic Acids Res* 41.
29. Cagliani R, Forni D, Biasin M, Comabella M, Guerini FR, Riva S, Pozzoli U, Agliardi C, Caputo D, Malhotra S, Montalban X, Bresolin N, Clerici M, Sironi M. 2014. Ancient and recent selective pressures shaped genetic diversity at AIM2-like nucleic acid sensors. *Genome Biol Evol* 6:830–845.
30. Vasseur E, Patin E, Laval G, Pajon S, Fornarino S, Crouau-Roy B, Quintana-Murci L. 2011. The selective footprints of viral pressures at the human RIG-I-like receptor family. *Hum Mol Genet* 20:4462–4474.
31. Fumagalli M, Cagliani R, Riva S, Pozzoli U, Biasin M, Piacentini L, Comi GP, Bresolin N, Clerici M, Sironi M. 2010. Population genetics of IFIH1: Ancient population structure, local selection, and implications for susceptibility to type 1 diabetes. *Mol Biol Evol* 27:2555–2566.
32. Quach H, Wilson D, Laval G, Patin E, Manry J, Guibert J, Barreiro LB, Nerrienet E, Verschoor E, Gessain A, Przeworski M, Quintana-Murci L. 2013. Different selective pressures shape the evolution of toll-like receptors in human and African great ape populations. *Hum Mol Genet* 22:4829–4840.
33. Van Der Lee R, Wiel L, Van Dam TJP, Huynen MA. 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res* 45:10634–10648.
34. Ahn M, Cui J, Irving AT, Wang LF. 2016. Unique loss of the PYHIN gene family in bats amongst mammals: Implications for inflammasome sensing. *Sci Rep* 6:21722.
35. Brunette RL, Young JM, Whitley DG, Brodsky IE, Malik HS, Stetson DB. 2012. Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. *J Exp Med* 209:1969–1983.
36. Cridland JA, Curley EZ, Wykes MN, Schroder K, Sweet MJ, Roberts TL, Ragan MA,

- Kassahn KS, Stacey KJ. 2012. The mammalian PYHIN gene family: Phylogeny, evolution and expression. *BMC Evol Biol* 12.
37. Jiggins FM, Kim KW. 2007. A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol* 20:965–970.
38. McTaggart SJ, Obbard DJ, Conlon C, Little TJ. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. *BMC Evol Biol* 12.
39. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet* 7.
40. Endo T, Ikeo K, Gojobori T. 1996. Large-scale search for genes on which positive selection may operate. *Mol Biol Evol* 13:685–690.
41. Mondragón-Palomino M, Meyers BC, Michelmore RW, Gaut BS. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res*.
42. Meyerson NR, Rowley PA, Swan CH, Le DT, Wilkerson GK, Sawyer SL. 2014. Positive selection of primate genes that promote HIV-1 replication. *Virology* 454–455:291–298.
43. Ortiz M, Guex N, Patin E, Martin O, Xenarios I, Ciuffi A, Quintana-Murci L, Telenti A. 2009. Evolutionary trajectories of primate genes involved in HIV pathogenesis. *Mol Biol Evol* 26:2865–2875.
44. Ruggieri V, Nunziata A, Barone A. 2014. Positive selection in the leucine-rich repeat domain of Gro1 genes in *Solanum* species. *J Genet* 93:755–765.
45. Zou J, Yue F, Jiang X, Li W, Yi J, Liu L. 2013. Mitochondrion-associated protein LRPPRC suppresses the initiation of basal levels of autophagy via enhancing Bcl-2 stability. *Biochem J* 454:447–457.
46. Zou J, Yue F, Li W, Song K, Jiang X, Yi J, Liu L. 2014. Autophagy inhibitor LRPPRC suppresses mitophagy through interaction with mitophagy initiator Parkin. *PLoS One* 9.
47. Zou J, Li W, Misra A, Yue F, Song K, Chen Q, Guo G, Yi J, Kimata JT, Liu L. 2015. The viral restriction factor tetherin prevents leucine-rich pentatricopeptide repeat-

- containing protein (LRPPRC) from association with Beclin 1 and B-cell CLL/lymphoma 2 (Bcl-2) and enhances autophagy and mitophagy. *J Biol Chem* 290:7269–7279.
48. Harrison BD, Barker H, Bock KR, Guthrie EJ, Meredith G, Atkinson M. 1977. Plant viruses with circular single-stranded DNA. *Nature* 270:760–762.
  49. Primmer CR, Papakostas S, Leder EH, Davis MJ, Ragan MA. 2013. Annotated genes and nonannotated genomes: Cross-species use of Gene Ontology in ecology and evolution research. *Mol Ecol*.
  50. Kang R, Chen R, Zhang Q, Hou W, Wu S, Cao L, Huang J, Yu Y, Fan XG, Yan Z, Sun X, Wang H, Wang Q, Tsung A, Billiar TR, Zeh HJ, Lotze MT, Tang D. 1998. HMGB1 in health and disease. *Mol Aspects Med*.
  51. Yang H. 2005. The cytokine activity of HMGB1. *J Leukoc Biol* 78:1–8.
  52. Lange SS, Vasquez KM. 2009. HMGB1: The Jack-of-all-trades protein is a master DNA repair mechanic. *Mol Carcinog* 48:571–580.
  53. Lee S, Kwak M, Kim S, Shin J. 2014. The role of high mobility group box 1 in innate immunity. *Yonsei Med J* 55:1165–1176.
  54. Ranzato E, Martinotti S, Patrone M. 2015. Emerging roles for HMGB1 protein in immunity, inflammation, and cancer. *ImmunoTargets Ther* 101.
  55. Lotze MT, Tracey KJ. 2005. High-mobility group box 1 protein (HMGB1): Nuclear weapon in the immune arsenal. *Nat Rev Immunol*.
  56. Harris HE, Andersson U, Pisetsky DS. 2012. HMGB1: A multifunctional alarmin driving autoimmune and inflammatory disease. *Nat Rev Rheumatol*.
  57. Yang H, Wang H, Chavan SS, Andersson U. 2015. High Mobility Group Box Protein 1 (HMGB1): The Prototypical Endogenous Danger Molecule. *Mol Med* 21 Suppl 1:S6–S12.
  58. Rosin DL, Okusa MD. 2011. Dangers Within: DAMP Responses to Damage and Cell Death in Kidney Disease. *J Am Soc Nephrol* 22:416–425.
  59. Choi HW, Manohar M, Manosalva P, Tian M, Moreau M, Klessig DF. 2016. Activation of Plant Innate Immunity by Extracellular High Mobility Group Box 3 and Its Inhibition by Salicylic Acid. *PLoS Pathog* 12.

60. Yanai H, Ban T, Wang Z, Choi MK, Kawamura T, Negishi H, Nakasato M, Lu Y, Hangai S, Koshihara R, Savitsky D, Ronfani L, Akira S, Bianchi ME, Honda K, Tamura T, Kodama T, Taniguchi T. 2009. HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* 462:99–103.
61. Yanai H, Ban T, Taniguchi T. 2012. High-mobility group box family of proteins: Ligand and sensor for innate immunity. *Trends Immunol.*
62. Yanai H, Taniguchi T. 2014. Nucleic acid sensing and beyond: Virtues and vices of high-mobility group box 1. *J Intern Med.*
63. Webb M, Payet D, Lee KB, Travers AA, Thomas JO. 2001. Structural requirements for cooperative binding of HMG1 to DNA minicircles. *J Mol Biol* 309:79–88.
64. Gibb CL, Cheng W, Morozov VN, Kallenbach NR. 1997. Effect of nuclear protein HMG1 on in vitro slippage synthesis of the tandem repeat dTG-dCA. *Biochemistry* 36:5418–5424.
65. Gaillard C, Strauss F. 2000. High affinity binding of proteins HMG1 and HMG2 to semicatenated DNA loops. *BMC Mol Biol* 1.
66. Webb M, Thomas JO. 1999. Structure-specific binding of the two tandem HMG boxes of HMG1 to four-way junction DNA is mediated by the A domain. *J Mol Biol* 294:373–387.
67. Reddy MC, Christensen J, Vasquez KM. 2005. Interplay between human high mobility group protein 1 and replication protein A on psoralen-cross-linked DNA. *Biochemistry* 44:4188–4195.
68. Malina J, Kasparikova J, Natile G, Brabec V. 2002. Recognition of major DNA adducts of enantiomeric cisplatin analogs by HMG box proteins and nucleotide excision repair of these adducts. *Chem Biol* 9:629–638.
69. Muhire BM, Golden M, Murrell B, Lefeuvre P, Lett J-M, Gray A, Poon A YF, Ngandu NK, Semegni Y, Tanov EP, Monjane AL, Harkins GW, Varsani A, Shepherd DN, Martin DP. 2014. Evidence of Pervasive Biologically Functional Secondary Structures within the Genomes of Eukaryotic Single-Stranded DNA Viruses. *J Virol* 88:1972–1989.
70. Rendon-Mitchell B, Ochani M, Li J, Han J, Wang H, Yang H, Susarla S, Czura C, Mitchell R a, Chen G, Sama AE, Tracey KJ, Wang H. 2003. IFN-gamma induces high

- mobility group box 1 protein release partly through a TNF-dependent mechanism. *J Immunol* 170:3890–7.
71. Ranji A, Boris-Lawrie K. 2010. RNA helicases: Emerging roles in viral replication and the host innate response. *RNA Biol*.
  72. Fullam A, Schröder M. 2013. DExD/H-box RNA helicases as mediators of anti-viral innate immunity and essential host factors for viral replication. *Biochim Biophys Acta - Gene Regul Mech*.
  73. Gringhuis SI, Hertoghs N, Kaptein TM, Zijlstra-Willems EM, Sarrami-Fooroshani R, Sprokholt JK, Van Teijlingen NH, Kootstra NA, Booiman T, Van Dort KA, Ribeiro CMS, Drewniak A, Geijtenbeek TBH. 2017. HIV-1 blocks the signaling adaptor MAVS to evade antiviral host defense after sensing of abortive HIV-1 RNA by the host helicase DDX3. *Nat Immunol* 18:225–235.
  74. Zou L, Elledge SJ. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science (80- )* 300:1542–1548.
  75. Nakad R, Schumacher B. 2016. DNA damage response and immune defense: Links and mechanisms. *Front Genet*.
  76. Wang Z, Zhu F. 2016. Minichromosome maintenance protein 7 regulates phagocytosis in kuruma shrimp *Marsupenaeus japonicas* against white spot syndrome virus. *Fish Shellfish Immunol* 55:293–303.
  77. Laman AG, Lathe R, Shepelyakovskaya AO, Gartseva A, Brovko FA, Guryanova S, Alekseeva L, Meshcheryakova EA, Ivanov VT. 2016. Muramyl peptides activate innate immunity conjointly via YB1 and NOD2. *Innate Immun* 22:666–673.
  78. Kang HS, Beak JY, Kim Y-S, Petrovich RM, Collins JB, Grissom SF, Jetten AM. 2006. NABP1, a novel RORgamma-regulated gene encoding a single-stranded nucleic-acid-binding protein. *Biochem J* 397:89–99.
  79. Yang P, An H, Liu X, Wen M, Zheng Y, Rui Y, Cao X. 2010. The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a B-catenin-dependent pathway. *Nat Immunol* 11:487–494.
  80. Labialle S, Dayan G, Gayet L, Rigal D, Gambrelle J, Baggetto LG. 2004. New invMED1 element cis-activates human multidrug-related MDR1 and MVP genes, involving the LRP130 protein. *Nucleic Acids Res* 32:3864–3876.

81. Ryu SJ, An HJ, Oh YS, Choi HR, Ha MK, Park SC. 2008. On the role of major vault protein in the resistance of senescent human diploid fibroblasts to apoptosis. *Cell Death Differ* 15:1673–1680.
82. Peng N, Liu S, Xia Z, Ren S, Feng J, Jing M, Gao X, Wiemer EAC, Zhu Y. 2016. Inducible Major Vault Protein Plays a Pivotal Role in Double-Stranded RNA- or Virus-Induced Proinflammatory Response. *J Immunol* 196:2753–2766.
83. Liu S, Hao Q, Peng N, Yue X, Wang Y, Chen Y, Wu J, Zhu Y. 2012. Major vault protein: A virus-induced host factor against viral replication through the induction of type-I interferon. *Hepatology* 56:57–66.
84. Bianchi ME, Agresti A. 2005. HMG proteins: Dynamic players in gene regulation and differentiation. *Curr Opin Genet Dev*.
85. Wang H, Bloom O, Zhang M, Vishnubhakat JM, Ombrellino M, Che J, Frazier A, Yang H, Ivanova S, Borovikova L, Manogue KR, Faist E, Abraham E, Andersson J, Andersson U, Molina PE, Abumrad NN, Sama A, Tracey KJ. 1999. HMG-1 as a late mediator of endotoxin lethality in mice. *Science* (80- ) 285:248–251.
86. Yang H, Hreggvidsdottir HS, Palmblad K, Wang H, Ochani M, Li J, Lu B, Chavan S, Rosas-Ballina M, Al-Abed Y, Akira S, Bierhaus A, Erlandsson-Harris H, Andersson U, Tracey KJ. 2010. A critical cysteine is required for HMGB1 binding to Toll-like receptor 4 and activation of macrophage cytokine release. *Proc Natl Acad Sci* 107:11942–11947.
87. Yu M, Wang H, Ding A, Golenbock DT, Latz E, Czura CJ, Fenton MJ, Tracey KJ, Yang H. 2006. HMGB1 signals through toll-like receptor (TLR) 4 and TLR2. *Shock* 26:174–179.
88. Andreeva L, Hiller B, Kostrewa D, Lässig C, De Oliveira Mann CC, Jan Drexler D, Maiser A, Gaidt M, Leonhardt H, Hornung V, Hopfner KP. 2017. CGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* 549:394–398.
89. Lewis SH, Obbard DJ. 2014. Recent insights into the evolution of innate viral sensing in animals. *Curr Opin Microbiol*.
90. Blevins T, Rajeswaran R, Shivaprasad P V., Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F, Hohn T, Pooggin MM. 2006. Four plant Dicers

mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res* 34:6233–6246.

91. Krupovic M, Forterre P. 2015. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann N Y Acad Sci* 1341:41–53.
92. Liu H, Fu Y, Li B, Yu X, Xie J, Cheng J, Ghabrial SA, Li G, Yi X, Jiang D. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol Biol* 11.
93. Rigby RE, Webb LM, Mackenzie KJ, Li Y, Leitch A, Reijns MAM, Lundie RJ, Revuelta A, Davidson DJ, Diebold S, Modis Y, MacDonald AS, Jackson AP. 2014. RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J* 33:542–558.
94. Mankan AK, Schmidt T, Chauhan D, Goldeck M, Höning K, Gaidt M, Kubarenko A V, Andreeva L, Hopfner K, Hornung V. 2014. Cytosolic RNA:DNA hybrids activate the cGAS – STING axis. *EMBO J* 33:2937–2946.