Systems-level comparison of protein-protein interactions between viruses and the human type I interferon system network

V. Navratil*, B. de Chassey*, L. Meyniel, F. Pradezynski, P. André, C. Rabourdin-Combe, V. Lotteau

* These two authors contributed equally to this work

Correspondence to: vincent.navratil@ens-lyon.fr; vincent.lotteau@inserm.fr

Supplementary file

Literature curation and selection of the 70 VIN proteins

Depending on their PAMPs, viruses are essentially sensed by Toll-like receptors (TLRs) or RIG-like receptors (RLRs). Signal is then transmitted through adapter proteins (MYD88, TICAM1, TICAM2, TIRAP and VISA) and the recruitment or activation of various proteins (TRAFs, IRAKs, MAPKs, IKKs…) leading to activation of transcription factors (NF-KB, IRFs and ATF2/JUN/FOS) and initiating IFN production. Secreted type I IFNs bind heterodimeric IFN receptor, initiate a phosphorylation cascade activating STAT proteins in complex with ISGF3G (IRF9) and induce the expression of ISG proteins. The most essential proteins involved in these pathways have been selected from the following articles to construct VIN:

1. *Kawai T, Akira S. Toll-like receptor and RIG-I-like receptor signaling. Ann N Y Acad Sci. 2008 Nov;1143:1-20.*
2. *Bowie AG, Unterholzner L. Viral evasion and subversion of pattern-recognition receptor signaling. Nat Rev Immunol. 2008 Dec;8(12):911-22.*
3. *Katze MG, Fornek JL, Palermo RE, Walters KA, Korth MJ. Innate immune modulation by RNA viruses: emerging insights from functional genomics. Nat Rev Immunol. 2008 Aug;8(8):644-54.*
4. *O'Neill LA. When signaling pathways collide: positive and negative regulation of toll-like receptor signal transduction. Immunity. 2008 Jul;29(1):12-20.*
5. *Kumagai Y, Takeuchi O, Akira S. Pathogen recognition by innate receptors. J Infect Chemother. 2008 Apr;14(2):86-92.*
6. *Takeuchi O, Akira S. MDA5/RIG-I and virus recognition. Curr Opin Immunol. 2008 Feb;20(1):17-22.*
7. *Borden EC, Sen GC, Uze G, Silverman RH, Ransohoff RM, Foster GR, Stark GR. Interferons at age 50: past, current and future impact on biomedicine. Nat Rev Drug Discov. 2007 Dec;6(12):975-90.*
8. *Takeuchi O, Akira S. Recognition of viruses by innate immunity. Immunol Rev. 2007 Dec;220:214-24.*
9. *Krishnan J, Selvarajoo K, Tsuchiya M, Lee G, Choi S. Toll-like receptor signal transduction. Exp Mol Med. 2007 Aug 31;39(4):421-38.*
10. *Hiscott J. Convergence of the NF-kappaB and IRF pathways in the regulation of the innate antiviral response. Cytokine Growth Factor Rev. 2007 Oct-Dec;18(5-6):483-90.*

11. *Thompson AJ, Locarnini SA. Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response. Immunol Cell Biol. 2007 Aug-Sep;85(6):435-45.*
12. *Baccala R, Hoebe K, Kono DH, Beutler B, Theofilopoulos AN. TLR-dependent and TLR-independent pathways of type I interferon induction in systemic autoimmunity. Nat Med. 2007 May;13(5):543-51.*
13. *Barton GM. Viral recognition by Toll-like receptors. Semin Immunol. 2007 Feb;19(1):33-40.*
14. *Lee MS, Kim YJ. Signaling pathways downstream of pattern-recognition receptors and their cross talk. Annu Rev Biochem. 2007;76:447-80.*
15. *Meylan E, Tschopp J. Toll-like receptors and RNA helicases: two parallel ways to trigger antiviral responses. Mol Cell. 2006 Jun 9;22(5):561-9.*
16. *Seth RB, Sun L, Chen ZJ. Antiviral innate immunity pathways. Cell Res. 2006 Feb;16(2):141-7.*
17. *Kawai T, Akira S. Innate immune recognition of viral infection. Nat Immunol. 2006 Feb;7(2):131-7.*
18. *Katze MG, He Y, Gale M Jr. Viruses and interferon: a fight for supremacy. Nat Rev Immunol. 2002 Sep;2(9):675-87.*
19. *Zak DE, Aderem A. Immunol Rev. Systems biology of innate immunity. 2009 Jan 227(1):264-82.*

VIN construction and visualization.

All protein-protein interactions between the 70 selected proteins and with the viral proteome have been extracted from VirHostNet to reconstruct the VIN. The Guess tool (http://graphexploration.cond.org) was used to graphically represent VIN in Figure 1. Figure 1 is available in a GUESS interactive format (GUESS Data Format) in SF1.tar.gz (VIN_network.properties) available at http://vinavratil.free.fr/navratil_vin/SF1.tar.gz.

Control of Experimental bias.

We carefully controlled that the differential trends identified in our analyses does not results from the variety of experimental technologies (*e.g.* Y2H, Co-IP, Pull-Down, etc.) or from the different tissue or cell type samples used to identify protein-protein interactions. To provide more insights into the robustness of our data against both expression bias introduced by the technology and inspection bias due to small-scale studies, we split the dataset into untargeted automated methods (mainly Y2H screens) and targeted small scale studies (mainly coimmunoprecipitation and pull down). More than 50% of the virus-host protein-protein interactions (n=1100/2097) were identified by Y2H automated method. By considering only protein-protein interactions generated by Y2H screens, the VIN remains significantly highly targeted with 22% of VIN proteins interacting with the viral proteome (Exact Fisher test - P-Value < 2.7 10e-09) (Supplementary Figure 4, SFigure 4). The neighbourhood remains also significantly highly targeted, with 11% of these proteins interacting with the viral proteome (Exact Fisher test - P-Value < 2.2 10e-16). The gradient of targeting that was observed with the full dataset was also conserved with the Y2H-restricted dataset (gradient from receptors (8%; 1/12) to mediators (18%; 7/38) to transcription factors (53%; 8/15)). Finally, based on the Y2H dataset, the global topology of the hierarchical clustering tree that was obtained with the full dataset (Figure 3a) remained unchanged (correlation coefficient, R=86%). Although some variations occurred, the composition of the 6 clusters identified appears globally conserved (Supplementary Figure 5, SFigure 5).

Therefore, the hypothesis of a massive viral attack of VIN and its neighbourhood appears robust and does not results from experimental bias.

Since a significant proportion of the integrated protein-protein interactions were identified from the different tissue or cell type (e.g. Y2H screens were performed on different cDNA library), the dataset could theoretically be influenced by gene expression bias. However, the type I interferon response pathway is known to be constitutively functional in a large panel of cells or tissues. To provide systematic support to this fact, we used millions of human ESTs (Expressed Sequence Tags) extracted from dbEST (www.ncbi.nlm.nih.gov/dbEST) and Unigene (www.ncbi.nlm.nih.gov/unigene) to evaluate the transcriptome of multiple human cell types and tissues. The data show that all the 70 proteins of VIN are expressed in more than 30 different tissues and in average in more than 50% of tissues sampled. Moreover, 97% of the VIN neighbourhood proteins are expressed in more than 10 different tissues and in average in more than 50% of tissues sampled. This suggests that according to the broad expression of proteins related to type I interferon response, bias related to transcript specific expression or abundance can exist but remains weak.

Supplementary Figures Legends

SFigure 1. VIN protein Interconnectivity distribution. The resampled number of protein-protein interactions between the 70 proteins of VIN (interconnectivity) was computed n=10.000 times based on a random resampling procedure (Materials and Methods). The observed interconnectivity (n=375) appears significantly greater than the maximum number obtained after resampling.

SFigure 2. Topological properties of targeted versus not targeted VIN proteins. Mean degree and betweenness were computed for targeted (red bar) and not targeted (blue bar) proteins within VIN subnetwork and the whole human protein interactome network (HPIN). The conventional standard error is given at the 5% threshold (Wilcoxon test - P-Value < 0.01 **).

SFigure 3. Comparative analysis of ISG and not-ISG proteins targeting according to virus families. Red bar: fraction of ISGs targeted. Blue bar: fraction of not-ISGs targeted. Fractions were calculated for *Flaviviridae, Herpesviridae, Papillomaviridae and Retroviridae* in VIN (left) and in VIN neighborhood (right). The conventional standard error is given at the 5% threshold.

SFigure 4. Interactions profiles of viral families onto the human type I IFN system. a. *Flaviviridae* interactions profiles. b. *Herpesviridae* interactions profiles. *c. Papillomaviridae* interactions profiles. d. *Retroviridae* interactions profiles. The same layout than the Figure 1 was applied. VIN proteins targeted by a viral family appear as red nodes (gray if not). *Flaviviridae*, *Herpesviridae*, *Papillomaviridae and Retroviridae* viral families have been selected because quantitatively comparable virus-host protein interaction datasets were available.

SFigure 5. Network-based clustering of viral subversion strategies based on Yeast Two-Hybrid screens. (a) hierarchical clustering of all VIN proteins according to the full dataset, same as Figure 3a. (b) hierarchical clustering of all VIN proteins according to Y2H dataset. VIN proteins (rows) are clustered according to the modulation index computed across viral families (columns). Colours range from red (no modulation) to white (highest modulation). Colour-coded clusters are represented as in Figure 3a. The global topology of the hierarchical clustering trees remained highly similar (correlation coefficient R=86%).

SFigure 6. Network-based clustering of viral subversion strategies at the species level. (a) hierarchical clustering of all VIN proteins at the viral proteome level, same as Figure 3a. (b) hierarchical clustering of all VIN proteins at the viral species level (HCV, HIV-1, HHV4, HPV16). VIN proteins (rows) are clustered according to the modulation index computed across viral species (columns). Colours range from red (no modulation) to white (highest modulation). The global topology of the hierarchical clustering trees remained highly similar (correlation coefficient R=80,4%).

Supplementary Tables Legends

STable 1. VIN proteins. Proteins belonging to VIN are referenced with their Ensembl gene ID (column 1), their NCBI official gene name (column 2) with their alias or synonymous and a short description (column 3).

STable 2. VIN and neighbourhood network metrics. VIN proteins are referenced with their NCBI official gene name (column 1) and are tagged according to their general function (column 2: receptor, adapter, mediator or transcription factor). For each VIN protein are given the number of interacting VIN proteins (VIN_hh_degree, column 3), number of cellular interactors in human network (HN_hh_degree, column 4), number of interacting viral families (targeting degree, column 5), number of cellular interactors outside VIN (neighbours_outside_VIN, column 6), number of neighboors outside VIN targeted by at least one viral protein (targeted_neighbour outside_VIN, column 7), its adjusted betweenness in VIN (VIN_hh_betweenness_adj, column 8) and its adjusted betweenness in the whole human network (HN_hh_betweenness_adj, column 9).

STable 3. Number of virus-host protein interactions in VIN and its neighbourhood. These numbers (lane 1) are given with the corresponding numbers of targeted host proteins (lane 2), distinct viral proteins (lane 3), viral species (lane 4) and viral families (lane 5).

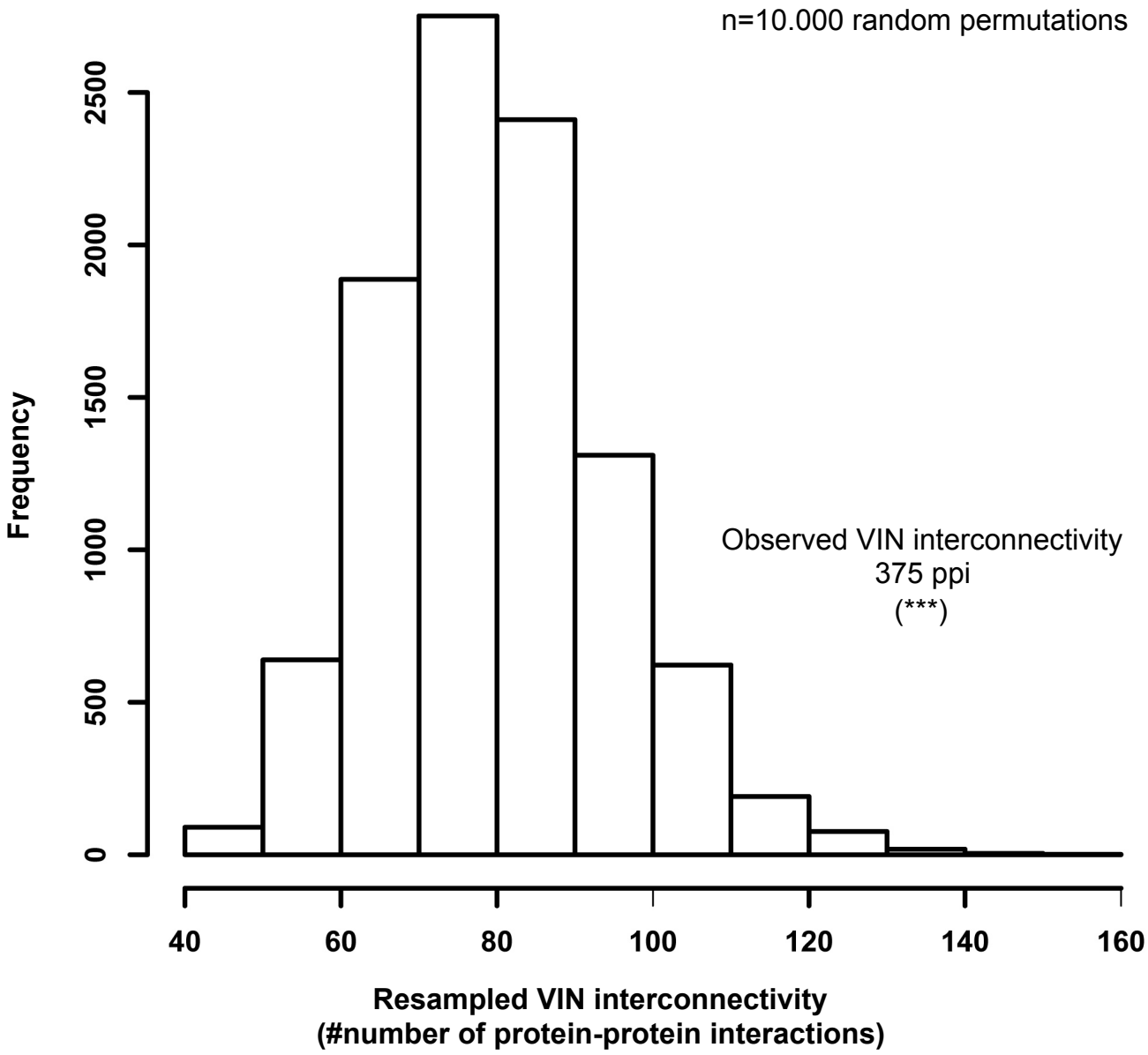
STable 4. List of cellular proteins targeted in the neighbourhood. VIN proteins and their neighbours are referenced with their cognate Ensembl gene ID (column 1 and 2 respectively) or their NCBI official gene name (column 3 and 4). PMIDs from where interactions are extracted are given (column 5) with the information of high throughput screening origin only (column 6) i.e., interaction not functionally validated.

STable 5. Gene Ontology functional enrichment analysis of VIN neighbours. The description of the Gene Ontology biological process, the corresponding Gene Ontology identifier, the number of known proteins involved in the process and the number of VIN neighbours involved in the process are given.

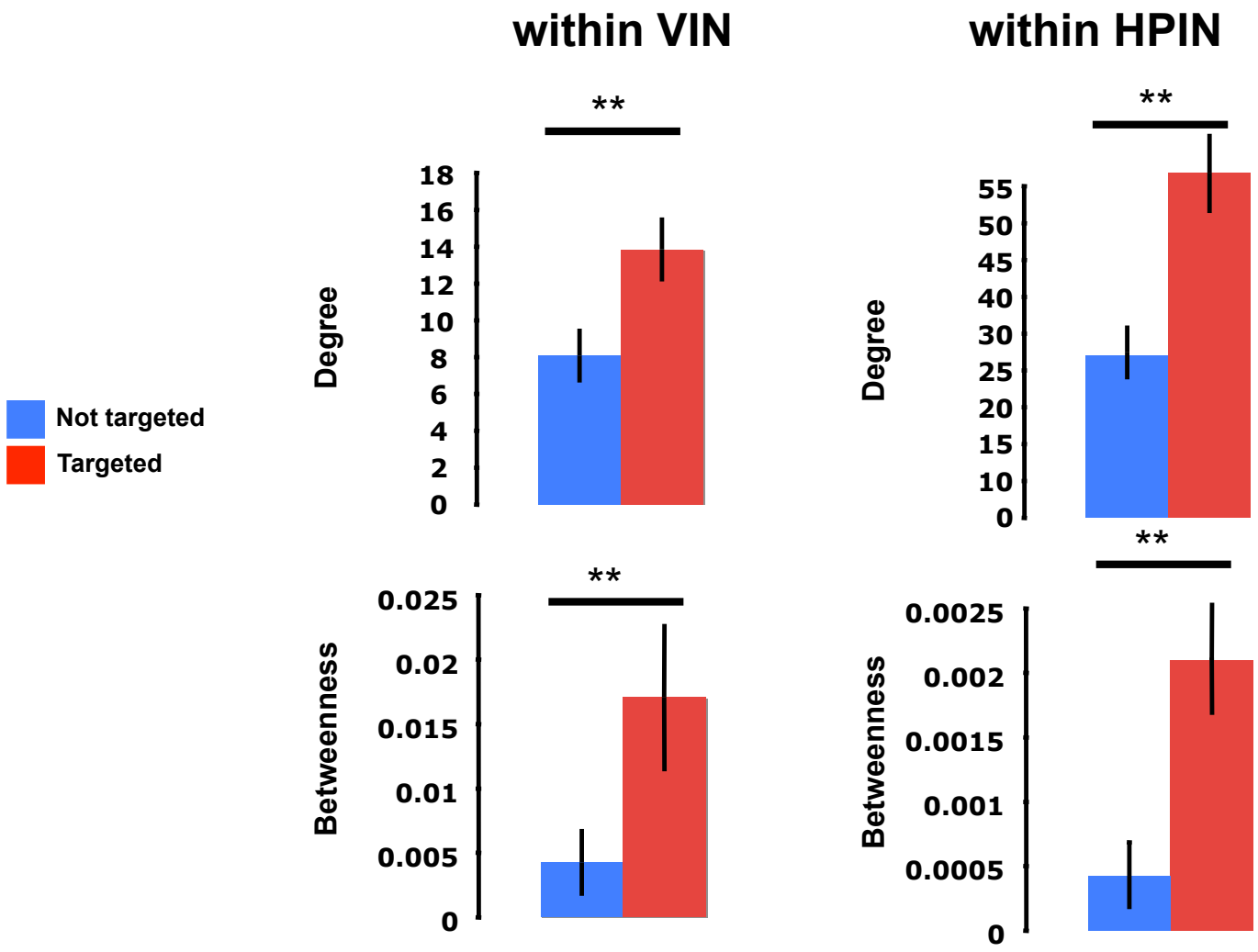
STable 6. List of VIN proteins targeted by viral family. Interactions between viral proteins and VIN proteins are given for *Herpesviridae, Flaviviridae, Retroviridae, Papillomaviridae*.

STable 7. Network metrics by viral families. VIN proteins are referenced with their NCBI official gene name (column 1) and are tagged according to their general function (column 2: receptor, mediator or transcription factor). For each VIN proteins are given number of interacting VIN proteins (VIN_hh_degree, column 3), number of cellular interactors in human network (HN_hh_degree, column 4), adjusted betweenness in VIN (VIN_hh_betweenness_adj, column 5) and adjusted betweenness in the whole human network (HN_hh_betweenness_adj, column 6). Metrics' means are given for each family.

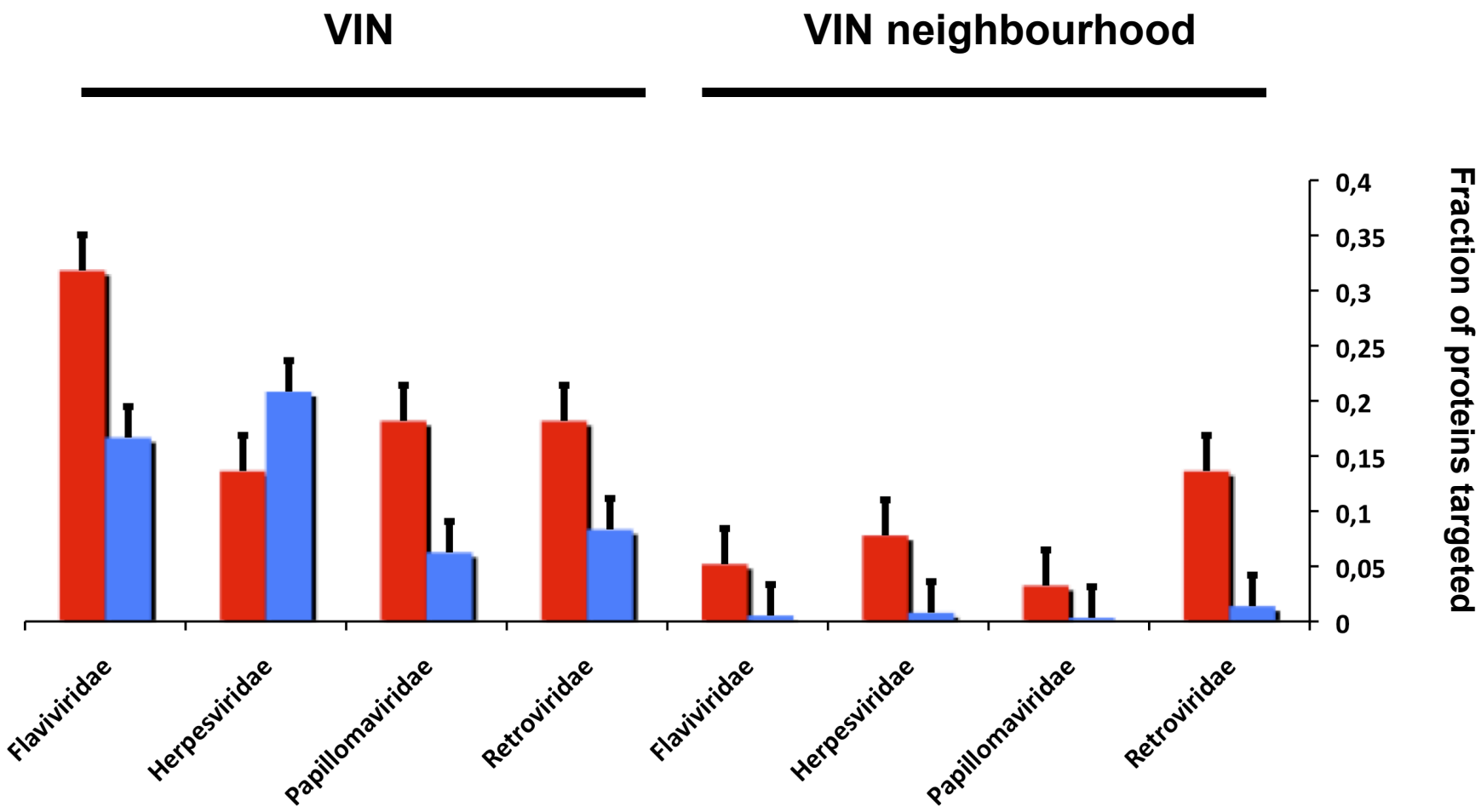
SFigure 1


n=10.000 random permutations
Observed VIN interconnectivity
375 ppi
(***)
Frequency
0
500
1000
1500
2000
2500
40
60
80
100
120
140
160
Resampled VIN interconnectivity
(#number of protein-protein interactions)

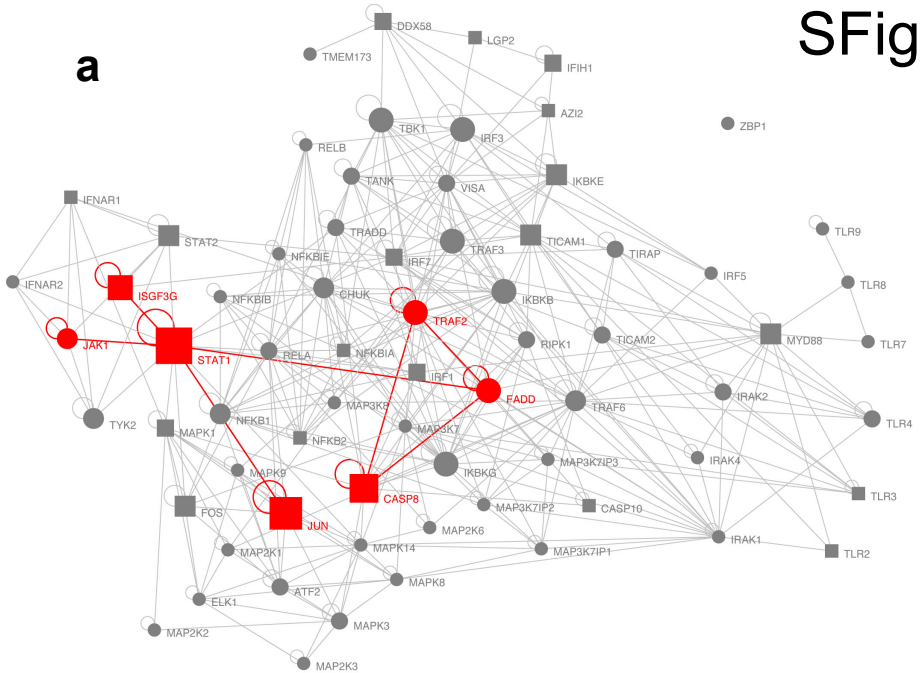
SFigure 2


within VIN
within HPIN
Not targeted
Targeted
Degree
Betweenness
**
18
16
14
12
10
8
6
4
2
0
55
50
45
40
30
25
20
15
10
0.025
0.02
0.015
0.01
0.005
0.0025
0.002
0.0015
0.001
0.0005

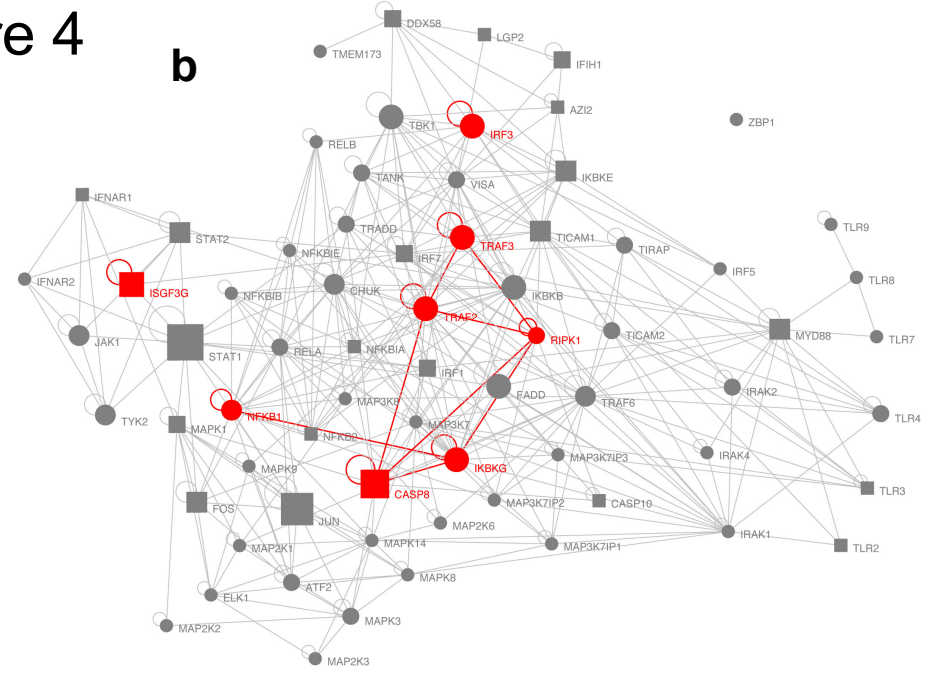
SFigure 3


VIN
VIN neighbourhood
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
Fraction of proteins targeted
Flaviviridae
Herpesviridae
Papillomaviridae
Retroviridae
Flaviviridae
Herpesviridae
Papillomaviridae
Retroviridae

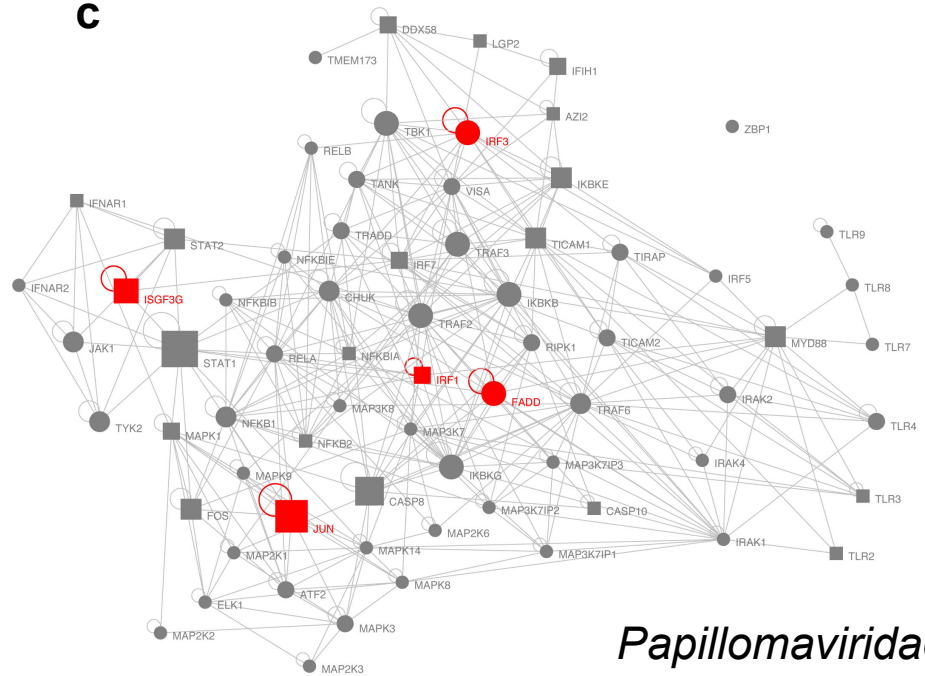
SFigure 4



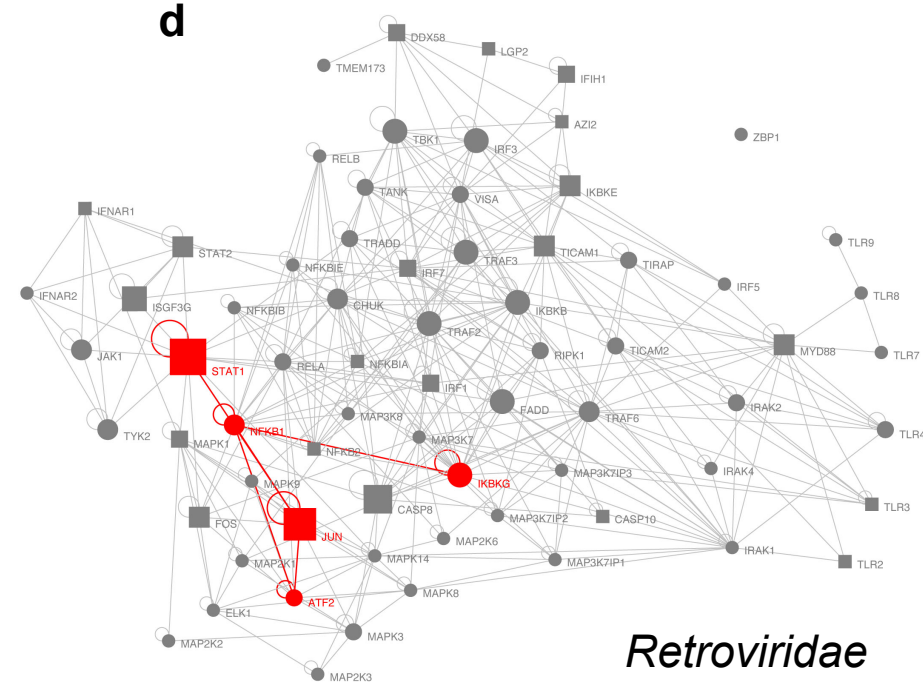
Flaviviridae



Herpesviridae

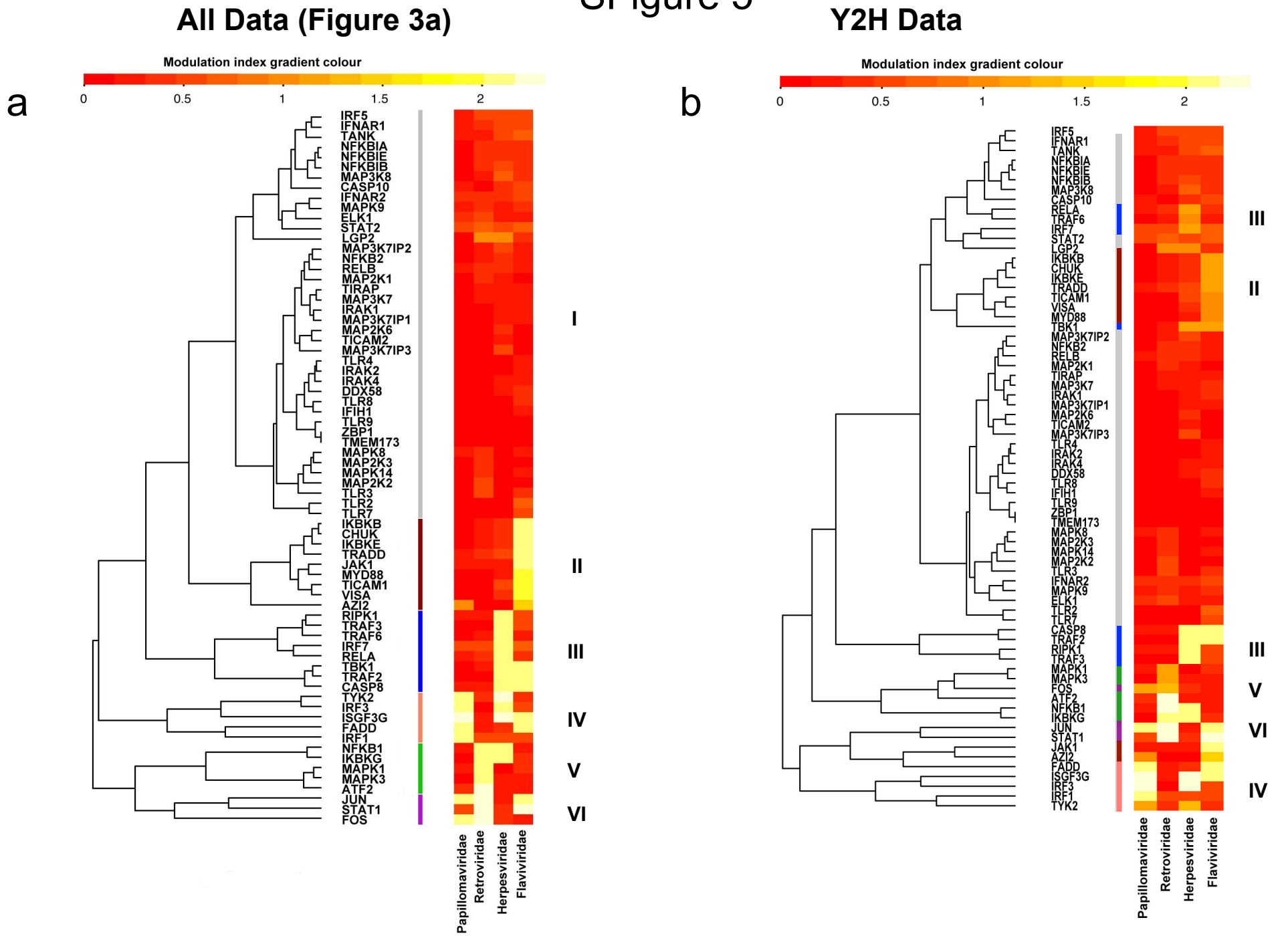


Papillomaviridae

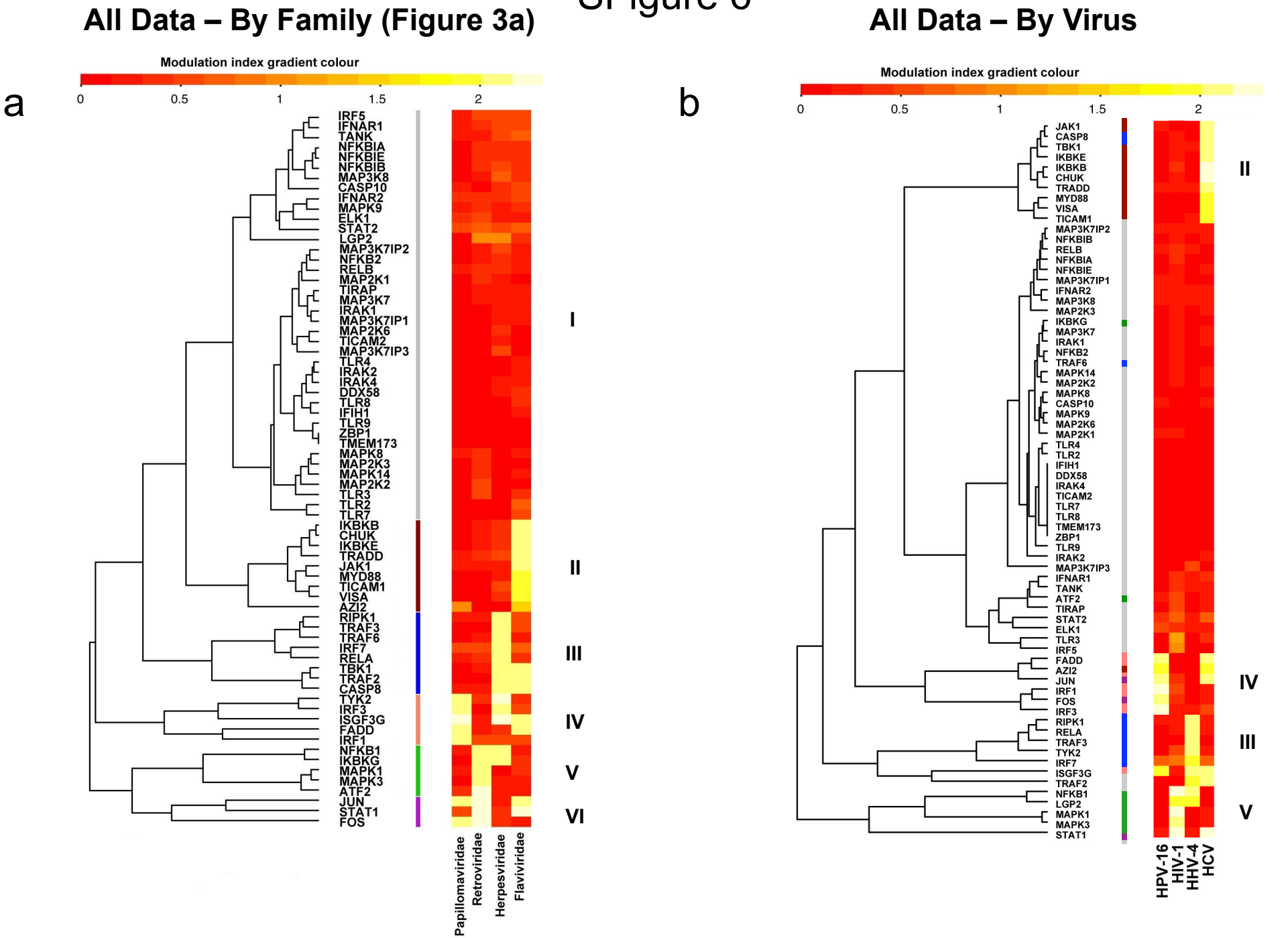


Retroviridae

SFigure 5


a
All Data (Figure 3a)
Modulation index gradient colour
0
0.5
1
1.5
2
IRF5
IFNAR1
TANK
NFKBIA
NFKBIE
NFKBIB
MAP3K8
CASP10
IFNAR2
MAPK9
ELK1
STAT2
LGP2
MAP3K7IP2
NFKB2
RELB
MAP2K1
TIRAP
MAP3K7
IRAK1
MAP3K7IP1
MAP2K6
TICAM2
MAP3K7IP3
TLR4
IRAK2
IRAK4
DDX58
TLR8
IFIH1
TLR9
ZBP1
TMEM173
MAPK8
MAP2K3
MAPK14
MAP2K2
TLR3
TLR2
TLR7
IKBKB
CHUK
IKBKE
TRADD
JAK1
MYD88
TICAM1
VISA
AZI2
RIPK1
TRAF3
TRAF6
IRF7
RELA
TBK1
TRAF2
CASP8
TYK2
IRF3
ISGF3G
FADD
IRF1
NFKB1
IKBKG
MAPK1
MAPK3
ATF2
JUN
STAT1
FOS
I
II
III
IV
V
VI
Papillomaviridae
Retroviridae
Herpesviridae
Flaviviridae
b
Y2H Data
Modulation index gradient colour
0
0.5
1
1.5
2
IRF5
IFNAR1
TANK
NFKBIA
NFKBIE
NFKBIB
MAP3K8
CASP10
RELA
TRAF6
IRF7
STAT2
LGP2
IKBKB
CHUK
IKBKE
TRADD
TICAM1
VISA
MYD88
TBK1
MAP3K7IP2
NFKB2
RELB
MAP2K1
TIRAP
MAP3K7
IRAK1
MAP3K7IP1
MAP2K6
TICAM2
MAP3K7IP3
TLR4
IRAK2
IRAK4
DDX58
TLR8
IFIH1
TLR9
ZBP1
TMEM173
MAPK8
MAP2K3
MAPK14
MAP2K2
TLR3
IFNAR2
MAPK9
ELK1
TLR2
TLR7
CASP8
TRAF2
RIPK1
TRAF3
MAPK1
MAPK3
FOS
ATF2
NFKB1
IKBKG
JUN
STAT1
JAK1
AZI2
FADD
ISGF3G
IRF3
IRF1
TYK2
III
II
III
V
VI
IV
Papillomaviridae
Retroviridae
Herpesviridae
Flaviviridae

SFigure 6


a
All Data – By Family (Figure 3a)
Modulation index gradient colour
0
0.5
1
1.5
2
IRF5
IFNAR1
TANK
NFKBIA
NFKBIE
NFKBIB
MAP3K8
CASP10
IFNAR2
MAPK9
ELK1
STAT2
LGP2
MAP3K7IP2
NFKB2
RELB
MAP2K1
TIRAP
MAP3K7
IRAK1
MAP3K7IP1
MAP2K6
TICAM2
MAP3K7IP3
TLR4
IRAK2
IRAK4
DDX58
TLR8
IFIH1
TLR9
ZBP1
TMEM173
MAPK8
MAP2K3
MAPK14
MAP2K2
TLR3
TLR2
TLR7
IKBKB
CHUK
IKBKE
TRADD
JAK1
MYD88
TICAM1
VISA
AZI2
RIPK1
TRAF3
TRAF6
IRF7
RELA
TBK1
TRAF2
CASP8
TYK2
IRF3
ISGF3G
FADD
IRF1
NFKB1
IKBKG
MAPK1
MAPK3
ATF2
JUN
STAT1
FOS
I
II
III
IV
V
VI
Papillomaviridae
Retroviridae
Herpesviridae
Flaviviridae
b
All Data – By Virus
Modulation index gradient colour
0
0.5
1
1.5
2
JAK1
CASP8
TBK1
IKBKE
IKBKB
CHUK
TRADD
MYD88
VISA
TICAM1
MAP3K7IP2
NFKBIB
RELB
NFKBIA
NFKBIE
MAP3K7IP1
IFNAR2
MAP3K8
MAP2K3
IKBKG
MAP3K7
IRAK1
NFKB2
TRAF6
MAPK14
MAP2K2
MAPK8
CASP10
MAPK9
MAP2K6
MAP2K1
TLR4
TLR2
IFIH1
DDX58
IRAK4
TICAM2
TLR7
TLR8
TMEM173
ZBP1
TLR9
IRAK2
MAP3K7IP3
IFNAR1
TANK
ATF2
TIRAP
STAT2
ELK1
TLR3
IRF5
FADD
AZI2
JUN
IRF1
FOS
IRF3
RIPK1
RELA
TRAF3
TYK2
IRF7
ISGF3G
TRAF2
NFKB1
LGP2
MAPK1
MAPK3
STAT1
II
IV
III
V
HPV-16
HIV-1
HHV-4
HCV

STable 3: Number of virus-host protein interactions on VIN and its neighbourhood.

	in VIN	in neighborhood	Total
Virus-host protein interaction	109	838	947
Targeted host protein	37	372	409
Distinct viral protein targeting host protein	62	263	292
Distinct viral species targeting host protein	34	80	89
Distinct viral families targeting host protein	13	25	26

description	*class*	*total*	*modulated*	*odds*
cell cycle and apoptosis				
cell cycle arrest	GO:0007050	81	13	6,26974607
interphase	GO:0051325	86	12	5,4509778
apoptosis	GO:0006915	715	81	4,42558408
regulation of caspase activity	GO:0043281	45	8	6,9449495
Pathways				
MAPKKK cascade	GO:0000165	159	21	5,15957333
JAK-STAT cascade	GO:0007259	40	8	7,81306818
epidermal growth factor receptor signaling pathway	GO:0007173	27	7	10,1280514
ER-nuclear signaling pathway	GO:0006984	23	7	11,8894516
steroid hormone receptor signaling pathway	GO:0030518	53	11	8,10790094
DNA structure and modification				
DNA packaging	GO:0006323	341	25	2,86402793
chromatin remodeling	GO:0006338	54	9	6,51089015
chromatin modification	GO:0016568	184	22	4,67085598
Protein modification				
ubiquitin cycle	GO:0006512	484	41	3,30925409
phosphorylation	GO:0016310	714	47	2,57152804
Transcription and translation				
regulation of transcription	GO:0045449	2295	119	2,02561027
regulation of translation	GO:0006417	171	16	3,65523658
Intracellular trafficking				
nucleocytoplasmic transport	GO:0006913	145	18	4,8494906
NLS-bearing substrate import into nucleus	GO:0006607	13	7	21,0351836
Cellular activation, proliferation, migration and differentiation				
lymphocyte activation	GO:0046649	188	22	4,57147606
mononuclear cell proliferation	GO:0032943	67	12	6,99677748
regulation of mononuclear cell proliferation	GO:0032944	45	10	8,68118687
Molecular metabolism				
regulation of lipid metabolic process	GO:0019216	24	6	9,76633523
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	GO:0019219	2350	126	2,09456722
regulation of DNA metabolic process	GO:0051052	44	11	9,76633523
regulation of protein metabolic process	GO:0051246	292	31	4,14734784
regulation of phosphate metabolic process	GO:0019220	62	9	5,67077529
regulation of amine metabolic process	GO:0033238	51	9	6,89388369
Innate and adaptative response				
regulation of immune response	GO:0050776	218	20	3,58397623
inflammatory response	GO:0006954	323	22	2,66079721
JNK				
regulation of JNK activity	GO:0043506	23	6	10,1909585
Virus				
viral reproduction	GO:0016032	50	10	7,81306818

STable 5: **Gene Ontology functional enrichment analysis of VIN neighbours.**