Title

Informing epidemic (research) responses in a timely fashion by knowledge management - a Zika virus use case

Authors

Angela Bauch[1*], Johann Pellet[2], Tina Schleicher[1], Xiao Yu[3], Andrea Gelemanović[4], Cosimo Cristella[3], Pieter L. Fraaij[5], Ozren Polasek[4], Charles Auffray[2], Dieter Maier[1], Marion Koopmans[5¶], Menno D. de Jong[3¶]

Affiliations

[1] Biomax Informatics AG, Planegg, Germany

[2] European Institute of Systems Biology and Medicine, Lyon, France

[3] Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

[4] Department of Public Health, University of Split School of Medicine, Split, Croatia

[5] Department of Viroscience and Department of Paediatrics, Erasmus Medical Centre, Rotterdam, the Netherlands

* Corresponding author

E-mail: angela.bauch@biomax.com

[¶] These authors contributed equally

Abstract

The response of pathophysiological research to emerging epidemics often occurs after the epidemic and, as a consequence, has little to no impact on improving patient outcomes or on developing high-quality evidence to inform clinical management strategies during the epidemic. Rapid and informed guidance of epidemic (research) responses to severe infectious disease outbreaks requires quick compilation and integration of existing pathophysiological knowledge. As a case study we chose the Zika virus (ZIKV) outbreak that started in 2015 to develop a proof-of-concept knowledge repository. To extract data from available sources and build a computationally tractable and comprehensive molecular interaction map we applied generic knowledge management software for literature mining, expert knowledge curation, data integration, reporting and visualisation. A multi-disciplinary team of experts, including clinicians, virologists, bioinformaticians and knowledge management specialists, followed a pre-defined workflow for rapid integration and evaluation of available evidence. While conventional approaches usually require months to comb through the existing literature, the initial ZIKV KnowledgeBase (ZIKA KB) was completed within a few weeks. Recently we updated the ZIKA KB with additional curated data from the large amount of literature published since 2016 and made it publicly available through a web interface together with a step-by-step guide to ensure reproducibility of the described use case (S4). In addition, a detailed online user manual is provided to enable the ZIKV research community to generate hypotheses, share knowledge, identify knowledge gaps, and interactively explore and interpret data (S5). A workflow for rapid response during outbreaks was generated, validated and refined and is also made available. The process described here can be used for timely structuring of pathophysiological knowledge for future

threats. The resulting structured biological knowledge is a helpful tool for computational data analysis and generation of predictive models and opens new avenues for infectious disease research.

Availability: www.zikaknowledgebase.eu

Funding

European Commission's Seventh Framework Research Programme project PREPARE (FP7-Health n°602525) and ZIKALLIANCE (MK, H2020; No 734548).

Author summary

During the recent ZIKV outbreak there was little information about the interactions between Zika virus and the host, however, the massive research response lead to a steep increase in the number of relevant publications within a very short period of time. At the time, there was no structured and comprehensive database available for integrated molecular and physiological data and knowledge about ZIKV infection. Researchers had to manually review the literature (amounting to over 5000 articles on ZIKV during our last update of the ZIKA KB in September 2018) to extract information about host–pathogen interaction and affected molecular, cellular and organ pathways. We explored the use of automated literature analysis and a defined cooperative effort between experts from various scientific, biomedical and information-technology domains to rapidly compile existing pathophysiological knowledge as a potential tool to support investigations during an emergency. This tool is contrasted with conventional approaches that would take months to comb

through the massive amount of existing literature. In addition to providing background information for research, scientific publications can be processed to transform textual information into complex networks, which can be integrated with existing knowledge resources to suggest novel hypotheses that potentially contribute to innovative infectious disease research approaches. This study shows that the knowledge extraction and mapping process required to inform clinical and research responses to an emerging epidemic can be efficiently and effectively executed with a dedicated and trained group of experts, a validated process and the necessary tools. Our results further provide an overview of ZIKV biology, allow prediction of drug efficacy and indentify specific host factors and signalling pathways affected by ZIKV.

Introduction

The response to a (re-)emerging infectious disease (ID) epidemic requires a rapid compilation of existing pathophysiological knowledge to inform research priorities guiding basic and clinical research. Gaps in understanding of the underlying mechanisms make it difficult to design effective disease-modifying therapies. Hence, during an emerging ID outbreak, the available information at the time of its emergence and the subsequent rapid accumulation of scientific knowledge from various sources needs to be captured and analysed in a timely and comprehensive fashion. Responding to an ID outbreak therefore would benefit from the use of a knowledge repository that organizes the disease-related knowledge into pathway, molecular interaction and disease maps. Such maps are a relatively new concept that have been used in neurodegenerative and heart diseases (1,2), but which have

had limited application in the field of ID thus far (3–5).

Molecular interaction and disease maps are dynamic computer-based knowledge repositories developed to integrate data and information across information sources, in a manner that is customized to the research domain of interest. Data types include interactions between molecular components, such as genes, pathogens, compounds and diseases.

The Platform foR European Preparedness Against (Re-)emerging Epidemics (PREPARE) is an EU-funded research consortium and clinical research network with the aim to rapidly respond to severe ID outbreaks, generating real-time evidence to inform optimized clinical management of patients and public health response. The 2015 ZIKV outbreak was considered as a test case in the context of the PREPARE network, as the pathogenesis of neurologic or immune disease induced by ZIKV is not fully understood. ZIKV is a flavivirus belonging to the *Flaviviridae* family and had only marginally been researched prior to the 2015 epidemic was minimal (6–8). Outbreaks of ZIKV disease have been recorded in Africa, the Americas, Asia and the Pacific. Acute ZIKV infections are mostly asymptomatic or associated with mild and self-limiting symptoms of fever, rash, conjunctivitis, headache or joint pain (9,10). However, the unexpected association of ZIKV infection with pregnancy and the subsequent severe neurodevelopmental problems in offspring and with the occurrence of neurological illnesses such as Guillain-Barre syndrome (GBS) or meningoencephalitis in acutely infected patients, led to widespread global concerns and a Public Health Emergency of International Concern (PHEIC) declaration by World Health Organisation (WHO) in 2016 (7).

We used the ZIKV virus outbreak as a case study to develop and test the steps, tasks, protocols and tools necessary to rapidly gather and integrate existing and

emerging knowledge and to inform research priorities (Fig 1). Based on the available data and information we aimed to obtain a general overview of pathophysiological knowledge on ZIKV infection and its associated clinical manifestations described in the public domain. Other neurotropic flaviviruses, such as Dengue virus (DENV), West Nile virus (WNV), Japanese Encephalitis virus (JEV) and Tick-borne Encephalitis virus (TBEV) also cause nervous system infections, in particular encephalitis, but no association with neurodevelopmental disorders or GBS have been reported (11). To see whether including these viruses would shed additional light on ZIKV pathogenesis we compared available ZIKV information to other neurotropic flaviviruses in terms of neurovirulence and disease severity.

Fig 1. ZIKV KnowledgeBase generation process — Overview

Based on the research objectives and knowledge provided by clinical/virology domain experts a six-step process was applied. In the first step a multidisciplinary expert team is assembled, in step 2, a semantic representation ("data model") was designed by the knowledge management experts. This model includes details about the data sources for integration, how to transfer data into the system and how to report, visualize and export results, as well as the definition of the semantic context for objects, such as "gene", "cell type" and "strain". In a third step, a natural language processing algorithm was applied to the integrated PubMed literature source. In step 4 the relevant data, including literature mining results, was imported into the system and semantically mapped to the data model. In step 5, queries, views and reports were formulated. In the last step a web-browser based user interface was implemented to enable clinical/virology experts to review, validate and refine the

integrated information.

Methods

Rapid response protocol

Many procedures have been published to collect knowledge from literature and experts, including systematic literature reviews (12), clinical guideline consensus building (13) and literature mining (14). Based on these approaches we developed a dedicated six step protocol with a focus on rapid assembly of existing knowledge (see Fig 1 and S1 Fig):

1. Team organisation and process management

A multidisciplinary team of clinicians, virologists, bioinformaticians and knowledge management specialists was formed to collaboratively extract existing ZIKV related knowledge from the literature and from public databases, integrate the available information into a consistent summary and further connect integrated data to molecular and pharmaceutical information. To enable an efficient and consolidated initial result, tasks were distributed between individuals and results were discussed and integrated in weekly online conferences. The detailed protocol for knowledge base generation was developed in this initial exercise and is presented in the results section.

2. Knowledge Management

For data organization, integration and development of molecular interaction and disease maps, a dedicated knowledge management tool is required. In this case, one of the PREPARE partners contributed the BioXM™ Knowledge Management

Environment, a generic platform for dynamic modelling, visualization and analysis of biological and biomedical networks (15). For knowledge representation we applied a semantic network approach as described previously (16,17). Briefly, the abstract concepts necessary to capture and represent essential ideas and physical objects relevant to the domain of knowledge were defined. Based on input from clinical and virology experts, the concepts required to represent existing pathophysiological knowledge of infectious diseases were modelled with objects, such as "genes", "strains", "is expressed in" or "interacts with". For the ZIKV KB, we focused on concepts required to represent text-mining results and information from structured databases of protein–protein (PPI) and drug–protein interactions, namely genes, diseases, pathogens and drugs. Relationships between pathogens, genes, drugs and compounds extracted by text-mining were represented by three types of relations: up-regulation, down-regulation and regulation (for further details see S2 Fig). Where possible, each concept was referenced to unique entries from reference databases or ontologies such as ChEBI (18) for chemicals and ICD10 (19) for diseases. The defined semantic concepts become directly available in a natural-language-like query and reporting language. This language can be used to address specific questions and to summarise and visualise available knowledge. For example, the query "is a *drug* which *interacts with* is a *protein* from *organism human* which *is expressed by* a *gene* which is *associated with* a *organism Zika*" retrieves the number of drugs that interact with a protein of interest and generates a visualisation which applies a color coding to genes that indicates to the number of associated drugs.

3. Text mining

The integrated text mining tool uses syntactic text parsing and dictionary-based

named-entity recognition to extract semantically typed associations (such as "inhibits", "activates") between the defined semantic concepts (such as "gene", "strain") (20). The initial task creates a defined text corpus, which includes uploaded relevant full text articles if applicable. In principle the textual materials for mining can be derived from PubMed abstracts, text from the WHO or other news feeds or any document in the portable document format (PDF), Mircorsoft Word or American Standard Code for Information Interchange (ASCII) formats. For the case study described here we used all ZIKV PubMed abstracts and publicly openly available full text articles. From these sources, relationships between genes, diseases, pathogens and drugs were extracted. The extracted associations consist of a subject, an object and the linking predicate and are enriched by their supportive evidence and additional metadata. For example, one such relationship is "Zika virus (subject) causes (predicate) microcephaly (object)" (Fig 2). Genes, diseases, pathogens and drugs, can be used as subjects and as objects and the sum of all extracted associations form an initial knowledge network. Genes, diseases, pathogens and drugs were defined by dictionaries that we curated from public sources as described below. Each dictionary consists of a well-defined set of ontologies (including synonyms) or reference databases tailored to the research question of interest. For instance, the disease dictionary consists of Disease ontology entries (21) and relevant branches of the NCI Thesaurus (22), the organism dictionary of NCBI taxonomy entries (23), the compound dictionary of ChEBI entries (18), as well as of KEGG (24) and NCI Thesaurus compounds. The gene dictionary is based on genes derived from human and flavivirus genomes. Predicates are derived from a set of verbs, which can be modified. These predicates describe mainly molecular interactions but can also indicate causal associations between proteins or

compounds and diseases (for instance “activates”, “restricts”, “targets”). To optimise recall and specificity of the mining, we extended the dictionaries for viral names, acronyms and interaction predicates as well as defined a black-list of acronyms causing mostly false positives.

Fig 2. Predicated text mining relationship

A text mining relationship consists of a subject (ZIKV), an object (microcephaly) and a linking predicate (causes). Subject and object are defined by dictionaries consisting of ontologies or reference databases, whereas predicates are derived from a fixed set of verbs with assertions from integrated sources, such as Medline. The term "microcephaly" is a reusable scientific concept that participates not just in one "Subject-Predicate-Object" construct detected, but in all such constructs detected that mention " microcephaly". Supplementary information is associated with the "microcephaly" object, including, for example, information from the Disease ontology, and other integrated resources, such as Gene-Disease-Association data (DisGeNET). This expandable set of relationships forms a large network of knowledge that enables new knowledge to be inferred by "reasoning" based on the logic encoded in those relationships.

Finally, the extracted relationships can be curated to manually optimise quality and information content. A curation user interface was implemented to enable the expert team to support or refute the automatically generated relationships. At least two independent researchers (the “4-eye review mode”) manually evaluated the evidence for every extracted relationship. In the case that the evaluations from the

two researchers conflicted, the conflicts were either resolved during the weekly online conferences or were excluded, as our goal was to maximise specificity (correctness) rather than sensitivity (completeness) of the integrated information.

In addition, experts could expand the network with any relevant supporting evidence from other integrated sources, such as public or proprietary databases and experimental data.

4. Semantic mapping of experimental results, public data sources and ontologies

Semantic mapping describes the process of identifying and linking concepts that are shared between two information sources. We integrated the databases listed in Table 1 using existing concepts such as genes, pathogens or diseases which were identified by ontological descriptors. Semantically identical objects are mapped to descriptive data from literature and databases to allow informed and efficient querying of the overall collected information (e.g. “Dengue disease” is mapped to the following synonyms: “Breakbone fever”, “Dengue disorder”, “Dengue fever” and “Dengue”) (25). To this end, mapping scripts are created to resolve a given input data format and match the provided entity identifiers or ontology terms. Experimental data from key publications is mapped by the same approach. While these data are henceforth available for search and reporting they are not yet displayed as part of any specific molecular interaction and disease map.

Source database	Information type	Current statistics	Level of curation	Update frequency and version

ATC	Anatomical Therapeutic Classification System	6064		2016
BioGRID	Protein-protein interaction	293022	Manually curated from literature Different evidence codes	Updated monthly vVersion 3.4.137
BioGRID	Protein-drug interaction	10722	Manually curated from literature Different evidence codes	Updated monthly Version 3.4.137
ChEBI	Compound information	161090	Curated from different data sources	Updated weekly
DisGeNET	Gene-disease associations	429036	Integrated from several public data sources and literature Score for ranking asscoiations	Permanently updated Version 4.0.0.0
Disease ontology	Standardized ontology for human disease	15043	Manually curated	Updated weekly
DrugBank	Drug and Drug Target database	8203	Manually curated from literature	Updated weekly
EntrezGene	Gene functional information	>24 million	Curated information integrated from different databases, based on RefSeq genomes	Updated weekly

Human Phenotype Ontology	standardized vocabulary of phenotypic abnormalities in human disease	11592		2016
KEGG	Pathways and reactions	273	Manually curated from literature	2008
NCI Thesaurus	Controlled vocabulary of the National Cancer institute	118502	Manually curated from literature	Updated weekly
OMIM	Gene - disease relations	21395	Curated form literature	Updated weekly
PubMed	Literature	>24 million	Automatic collection with manual curation	Updated weekly
Reactome	Pathways and reactions	5334	Manually curated from literature	Updated quartely Aug 2016
UniProtKB	Protein sequences	>65 million	Manually curated	updated bi-weekly
VirHostNet	Virus/Host molecular interactions	44310	Manually curated from literature	2.0 (March 2016)

Table 1. External data sources integrated into ZIKV KnowledgeBase

5. Querying and visualization of integrated information in tables, networks and disease maps

To help the expert team establish a specific molecular interaction and disease map

we defined a number of queries to explore the collective knowledge. These queries were used, for example, to find diseases and genes associated with a virus of interest to find diseases associated with genes prioritized according to experimental evidence.

Based on these queries we developed a streamlined, wizard-based user interface to create disease maps by selecting the relevant relationships from the curated text mining from query results (S3 Fig). This basic network was further extended with interaction data (e.g. PPI & protein-drug interaction) by applying a network search algorithm based on genes extracted from text mining relationships. Finally, we defined queries to overlay additional information, such as literature evidence, experimental data, drug targets or host factors to obtain different perspectives of the same underlying molecular interaction or disease map.

6. Deployment of an open access, web-based user interface

To make the results of our internal test case generally available and to support ZIKV research, we provide and maintain a regularly updated ZIKA KB at the following URL www.zikaknowledgebase.eu. As we continue to extend this resource user registration for access will be implemented to ensure the knowledge base is used for research only.

Results

Semantic representation of ZIKV infection

The data model implemented to provide a semantic representation of ZIKV infection

is described in detail in supplemental Figure S2. Briefly, the model focuses on genes, diseases, pathogens and drugs, and distinguishes between associations derived from literature mining and those provided by experimental data such as PPIs.

Text mining results

We searched PubMed with the terms “Zika virus”, “Dengue”, “West Nile virus”, “Japanese encephalitis virus”, “Tick-borne encephalitis virus”, “Microcephaly” and “Guillain Barre Syndrome” initially in December 2016 and most recently in September 2018. The recent search resulted in 4927 hits for “Zika virus” and 19974, 7700, 5918, 5213, 14248 and 8615 hits for the other search terms, respectively. During the analysed time frame, literature on ZIKV increased substantially from 1414 in 2016 to the current 4927 hits (250%), whereas for all other terms, the increase in publications was closer to 10%. Accordingly, the recent search identified additional disease phenotypes, including carditis and skin diseases, that were reported to be associated with ZIKV that were not present in the previous search. An additional set of 236 open access full text articles about ZIKV were included. A natural language processing algorithm was applied to these sets of documents to efficiently extract the fast growing information in the biomedical literature. The text mining extracted a total of 11916 relationships, which were manually evaluated to 2982 verified relationships (Table 2). The distribution of the curated relationships is depicted in Fig 3, indicating that the largest overlap was for ZIKV and DENV and for DENV and WNV. The curated set of relationships was used for further analyses, including generation of molecular interaction and disease maps and querying for virus-associated genes or diseases.

Text corpus	Key word	Number of documents	Number of relationships	Curated relationships
ZIKV virus FT	Zika virus	4,927; 236 FT[a]	1,192	298
Medline	Dengue	19,974	5,868	1,277
Medline	West Nile virus	7,700	1,586	387
Medline	Japanese encephalitis virus	5,918	1,594	354
Medline	Tick-borne encephalitis virus	5,213	508	134
Medline	Microcephaly	6,896	749	314
Medline	Guillain Barre Syndrome	4,133	419	218
Total			11,916	2,982

Table 2. Text mining analyses. [a]FT: full text

Fig 3. Distribution of text mining relationships. Numbers represent the sum of different types of relationships, such as gene-pathogen, gene-disease, gene-gene, gene-compound and pathogen-disease relations, found for each virus in total as well as in overlap with other viruses (A), or ZIKV in overlap with text mining analyses of Microcephaly and Guillain-Barre Syndrome (B).

Integrated data

Overall, the ZIKA KB contains a network of 337332 human and host-pathogen PPI integrated from BioGRID and VirHostNet, as well as 18905 protein-drug interactions

integrated from BioGRID and DrugBank, and 450431 gene-disease associations from DisGeNET (Table 1). Recently, a variety of ZIKV- and other flavivirus-related large-scale data sets, including microarray gene expression (26,27), RNAseq (28) as well as CRISPR/Cas data (29), have become publicly available and were integrated to identify host factors that are affected during viral infection.

Molecular interaction and disease maps

Curated text mining results were used to populate the initial ZIKV molecular interaction and disease map. In a second step the map was extended with interaction data (PPI & protein-drug interaction by applying, a network search to implement the breadth-first algorithm (30) which connected genes extracted from text mining relationships based on the overall network. This set of interaction data can be filtered and explored interactively. In a systems medicine approach, a multidisciplinary expert team systematically analysed literature, public databases and experimental resources to create a formal, structured model of molecular and cellular ZIKV–host interactions (“molecular interaction and disease map”)

Publicly available ZIKA KB

After an assessment period of internal use, a web-browser based user interface was implemented to make the PREPARE ZIKA KB available to all ZIKV researchers. By openly sharing the collected data and information, the ZIKA KB allows researchers to generate hypotheses, identify knowledge gaps and interactively explore and interpret data. All data are currently in the public domain. Upon request, data submission can be modified to allow registered users to specify that submitted data should not be publicly available

.

Use of the ZIKA KB

In the following we provide several example use cases. For instance, publicly available interaction data, such as the PPI and protein–drug interaction data can be used to visualize drug targets and host factors involved in ZIKV pathogenesis. Alternatively, users can filter for PPIs whose source or target is a drug or refine search results to include only proteins localized to a specific cellular compartment, such as the endoplasmic reticulum. The returned networks can be interrogated subsequently to identify host factors that are targeted by the virus and to search for drugs that interact with these host factors. and thus might contribute to drug repositioning for future treatment options for ZIKV infection. The maps can also be explored further by using integrated expression and knockout data.

To explore the integrated literature knowledge for relevance or obtain an overview of drug targets or identify critical genes within the network consisting of gene-disease-pathogen relationships, predefined perspectives were overlaid onto the default map. The association of ZIKV with microcephaly was reported most frequently across all ZIKV literature and this association is visualized by the thickness of the edges (Fig 4A). Known drug targets interacting directly with ZIKV or microcephaly were highlighted in green for potential intervention evaluation (Fig 4B). Genes playing a role in ZIKV infected human neural progenitor cells (hNPCs) were also highlighted for comparative analyses of complementary experimental analyses (Fig 4C).

Fig 4. Different ZIKV molecular interaction and disease map perspectives. (A) The amount of literature evidence is depicted by relation strength. (B) Genes known to be drug targets are color coded: the gene is a target for one (green) or five or more (blue) drugs or no drugs (orange). (C) Genes (hNPCs challenged with ZIKV, Tang et al, 2016) are color coded to indicate up- (red) and down- (blue) regulation .

Diseases associated with flaviviruses

To further explore and validate the knowledge derived from the curated text mining analyses, diseases associated with a virus of interest were queried. The results are displayed in Fig 5. It was assumed that the disorders that were most frequently associated with a virus were the primary disorder for infection with the virus. Microcephaly and GBS, for instance, are the most frequently mentioned disorders associated with ZIKV infection. Dengue fever and hematopoietic system disorders (e.g. thrombocytopenia) are frequently listed for DENV, whereas encephalitis is the most frequently mentioned disease for WNV, JEV and TBEV. Each disease thus represents the corresponding primary manifestation of these viral infections. Encephalitis is equally frequently associated with ZIKV and DENV confirming that ZIKV is also an aetiological agent in encephalitis.

Fig 5. Diseases associated with flaviviruses. The amount of literature evidence (y-axis) for each of the diseases (x-axis) is grouped for the five flaviviruses. Symbols highlight neurological diseases associated with more than one virus. Triangle: GBS, moon: Encephalitis, pie-chart: Peripheral nervous system disease, star: Neurologic manifestations

396

397

398 *Potential inhibitors of ZIKV infection*

399

400 Currently there is no approved therapy to treat ZIKV infection. Barrows *et al* recently
401 performed a screen of 774 FDA-approved drugs to identify agents that could
402 potentially be repositioned as treatment options for ZIKV infection (31). Of these, 24
403 potential inhibitors of ZIKV infection were identified and validated in human neural
404 stem cells and primary amnion cells. In addition to their potential use for treatment,
405 these compounds provide a resource to study ZIKV pathogenesis and can contribute
406 to insights into the biology of ZIKV. To this end, the ZIKV molecular interaction and
407 disease map described in Figure 4 was extended and filtered to include these
408 potential "*ZIKV effective drugs*" which were connected to genes associated to ZIKV
409 through PPIs (Fig 6). After this extension ten of the identified *ZIKV effective drugs*
410 were part of the new map which we then used to gain insight into potential drug
411 mechanisms and ZIKV biology. One of the drugs, Bortezomib, is a known antiviral
412 compound that inhibits replication of flaviviruses (32). Bortezomib is a proteasome
413 inhibitor, suggesting that proteasome action is essential for ZIKV replication. This
414 conclusion is in agreement with published CRISPR screen data (29) identifying
415 genes associated with protein degradation required for ZIKV infectivity. Interestingly,
416 four of the predicted *ZIKV effective drugs* (Mefloquine, Mebendazole, Sorafenib and
417 Dactinomycin) are associated with genes which, through PPIs, are involved in ErbB
418 signalling. *ErbB* is associated with the development of neurodegenerative diseases
419 when inactivated (33). Four of these genes (*MYC*, *GSK3B*, *BRAF* and *MAP2K2*) are
420 reported to be up-regulated in a published RNAseq analysis (28) performed in

human embryonic cortical neural progenitor cells (Fig 7). These genes can serve as an entry point to be tested in specific assays designed to unravel molecular mechanisms between of ZIKV involvement in microcephaly.

Another predicted *ZIKV effective drug* is Sorafenib, a multi-target tyrosine kinase inhibitor. The ZIKV map was used to identify the effective target of Sorafenib:

1. Sorafenib interacts with 4 target genes, *FLT3*, *BRAF*, *VEGFR* (also known as *KDR*) and *PDGFR*. The latter two genes are known to interact with additional drugs, such as Sunitinib, Pazopanib, Dasatinib and Imatinib. These additional drugs were among those which had no effect in the ZIKV infection assay.

2. In ZIKV infection expression data, none of the genes producing protein products that interact with VEGFR and PDGFR through known PPI (orange edges) with ZIKV are differentially expressed (Fig 8). In contrast, *BRAF* and *SOCS2*, a FLT3 interactor, were unregulated upon ZIKV infection.

Based on the observations above drawn from the ZIKV map, we hypothesise that FLT3 or BRAF are the effective targets of Sorafenib in ZIKV infection, rather than VEGFR and PDGFR. This exemplifies how molecular interaction and disease maps can be used to provide further insight into ZIKV biology.

Fig 6. ZIKV molecular interaction and disease map extended for ZIKV effective drugs.

The following are the symbols used in the map. Orange circles: genes; green stars: ZIKV effective drugs; yellow rectangles: flaviviruses; pink rectangles: diseases; violet rectangles: GO processes or KEGG signalling pathways. The following edge colors are used in the map. Black edges: relationships derived from text mining; orange: protein-drug or PPIs. The latter were obtained by applying a network search

algorithm selecting the drug target of a ZIKV effective drug as start and STAT2 (contained in a direct relation with ZIKV) as the end point.

Fig 7. ZIKV molecular interaction and disease map extended for ZIKV effective drugs (enlarged perspective).

Up- and down-regulated genes (hNPCs challenged with ZIKV, Tang et al, 2016) are highlighted by a color code ranging from red to blue from enlarged perspectives surrounded by dotted boxes in Figure 6.

Fig 8. ZIKV molecular interaction and disease map extended for ZIKV effective drugs (enlarged perspective).

Up- and down-regulated genes (hNPCs challenged with ZIKV, Tang et al, 2016) are highlighted by a color code ranging from red to blue from enlarged perspectives surrounded by dotted boxes in Figure 7.

The network analysis described above could also be used to rank drugs according to their distance to known ZIKV associated genes, such as STAT2, to suggest a metric for priorisation in screening assays (34). The discussed extended ZIKV map contains 429 target gene products that interact with FDA approved drugs. Evaluation of the distance between STAT2 and drug targets via experimentally proven PPIs revealed that targets of ZIKV effective drugs were on average more proximal to STAT2 compared to other targets.

Combining the proximity measure with additional knowledge, for example the “FDA pregnancy” label, reduces the number of drugs to be screened from 774 to 64.

Discussion

In case of an emerging epidemic, public health as well as clinical and preclinical research responses are typically hampered by a lack of structured, curated and actionable knowledge. The results of this study describe an approach to knowledge extraction and mapping that can quickly provide an overview of existing and missing information if done by a dedicated and trained group of experts. The developed workflow does not follow formal expert consensus seeking processes, such as Delphi (13), systematic literature review processes such as Cochrane (35) and PRISMA (12), or medical guideline related processes (36,) as these processes are not compatible with the need for speed during emerging epidemics. Nevertheless, the workflow adopts several important aspects of good practice: it is systematic, independent and transparent, provides evidence for all integrated information and uses appropriate quality criteria. Combined with the software tools employed in the process, this pragmatic approach enabled much faster knowledge generation than more traditional methods.

The tools employed in the process need to be able to semantically integrate disparate structured resources of heterogeneous data with ease. However, much of the knowledge that represents scientific research advancements is locked within the unstructured text of classical publications, such as journal articles, newsfeeds or free-form web publications (e.g. Zika-related clinical information at http://www.ovid.com/site/zika/resources.html). The sheer volume of this published

information grows constantly and exponentially and, for the most active areas of research, far exceeds the capacity of individual scientists and medical doctors to identify and read all relevant articles. Literature mining, a well-established technology to extract meaningful information from text, provides valuable assistance in structuring the massive amounts of text data and, therefore, is an indispensable tool in the process of guidance generation. Dynamic integration of objects and the relationships they participate in that are present in the literature through the use of structured resources and experimental data is a pre-requisite for analysis and distinguishes the ZIKA KB from text mining-only solutions, such as ContentMine (http://contentmine.org/) or databases dedicated to specific questions such as SncRNAs linking to disease symptoms ((http://zikadb.cpqrr.fiocruz.br/zika/).

Beyond our initial analysis presented here, users can explore the ZIKA KB within a web-based user interface with the use of the online manual (S5) and the step-by-step guide (S4) to reproduce the presented results. We will collect and highly appreciate any user feedback to optimise user experience for broad adoptation. In contrast to alternative useful resources such as the Virus Pathogen Resource (viprbrc.org), which focus on gene and protein sequences, the ZIKA KB integrates genetic, phenotypic and drug knowledge about ZIKV to facilitate the generation of hypotheses, define research priorities and enable better understanding of viral pathogenesis. In addition to interactive exploration, a corresponding ranking of connections in a network based on integration of multiple pieces of biological evidence can also be performed systematically and on a large scale, for example, by applying the ChainRank method (37), which we plan to integrate in the future.

Based on the text mining analyses performed here, disease profiles for the set of five neurotropic flaviviruses were confirmed. Common knowledge was retrieved along with underlying literature evidence and rare manifestations, such as encephalitis, associating with ZIKV and DENV.

Using a molecular interaction and disease map based on ZIKV, Microcephaly, and GBS text mining analyses results, we showed that further exploration of the described map can provide insight into, for example, ZIKV biology, propose conclusions for research decisions, predict drug efficacy, as exemplified in the results section, as well as propose hypotheses on specific host factors and signalling pathways affected by ZIKV. The map can help to distinguish between multiple potential targets of a ZIKV effective drug. The integration of information about effectiveness of other drugs as well as their target genes and the information about genes whose expression is affected during ZIKV infection indicated that Sorafenib likely acts via its target genes, *FLT3* and/or *BRAF,* but not via its alternative target genes *VEGFR* or *PDGFR*. In addition, the number of drugs to be screened was reduced from 774 to 64 by filtering potential drug candidates based on their network distance to ZIKV infection associated genes and additional phenotype relevant additional knowledge, such as contained in the “FDA pregnancy” label.

The conclusions that can be drawn are limited by the initially low number of available publications and limited experimental data, a situation which is inherent to most emerging epidemics. Nevertheless, the work presented shows that the use of a knowledge integrating system can provide guidance for clinical and research responses, such as follow-up studies regarding the association between ZIKV,

546 microcephaly and epilepsy, the validation of candidate drugs for ZIKV treatment, and
547 the validation of candidate genes in specific functional assays to better understand
548 molecular ZIKV infection mechanisms. or to complement existing functional genomic
549 approaches with proteomics studies, such as the integrated proteomics approach
550 identifying cellular targets of ZIKV proteins (38,39). These studies allow additional
551 comparative analyses between ZIKV and other flavivirus family members in terms of
552 virulence and pathogenic traits.
553 Another limitation of the system is the restricted types of information which can be
554 retrieved by text mining. While qualitative associations between genes/proteins,
555 drugs, diseases and organisms are readily amendable to automatic approaches, it is
556 currently almost impossible to extract, for example, clinical study designs, detailed
557 quantitative information or complex treatment plans.
558 Finally, the ZIKA KB in its current stage enables exploration of the integrated
559 information, as well as generation and curation of text-mining analysis but is not a
560 public tool for molecular interaction and disease map generation. The functions
561 required for these tasks will need further refinement before they can be made
562 available in a general way.
563
564 In summary, this approach in our opinion, provides a feasible way to collect and
565 integrate existing knowledge to better understand the molecular mechanisms of an
566 emerging pathogen. In addition our approach helps to identify gaps in knowledge
567 and, together with the other features, guides rapid and effective responses to future
568 epidemics. We have made the specific outcome of our approach, the Zika
569 KnowledgeBase, publicly available as a hopefully valuable resource to the ZIKV
570 research community.

In the light of the current COVID-19 pandemic we now apply the described workflow to SARS-CoV-2 and other coronaviruses and will make the developed resource available as described.

Acknowledgements

The authors wish to thank their colleagues for discussion and feedback, and Sheridon Sauer and Shannon Frances for critical review of the article.

References

1. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. Mol Neurobiol. 2014 Feb;49(1):88–102.

2. Nim HT, Furtado MB, Costa MW, Kitano H, Rosenthal NA, Boyd SE. CARFMAP: A Curated Pathway Map of Cardiac Fibroblasts. PloS One. 2015;10(12):e0143274.

3. Guo X, Xu Y, Bian G, Pike AD, Xie Y, Xi Z. Response of the mosquito protein interaction network to dengue infection. BMC Genomics. 2010 Jun 16;11:380.

4. Le Breton M, Meyniel-Schicklin L, Deloire A, Coutard B, Canard B, de Lamballerie X, et al. Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast two-hybrid screen. BMC Microbiol. 2011 Oct 20;11:234.

5. Matsuoka Y, Matsumae H, Katoh M, Eisfeld AJ, Neumann G, Hase T, et al. A comprehensive map of the influenza A virus replication cycle. BMC Syst Biol. 2013 Oct 2;7:97.

6. Anderson KB, Thomas SJ, Endy TP. The Emergence of Zika Virus: A Narrative Review. Ann Intern Med. 2016 Aug 2;165(3):175–83.

7. Pierson TC, Diamond MS. The emergence of Zika virus and its new clinical syndromes. Nature. 2018;560(7720):573–81.

8. Pardy RD, Richer MJ. Zika Virus Pathogenesis: From Early Case Reports to Epidemics. Viruses. 2019 Sep 21;11(10).

9. Sharma V, Sharma M, Dhull D, Sharma Y, Kaushik S, Kaushik S. Zika virus: an emerging challenge to public health worldwide. Can J Microbiol. 2019 Nov 4;1–12.

10. Murray JS. Understanding Zika virus. J Spec Pediatr Nurs JSPN. 2017;22(1).

11. Carod-Artal FJ. Epidemiology and neurological complications of infection by the Zika virus: a new emerging neurotropic virus. Rev Neurol. 2016 Apr 1;62(7):317–28.

12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med [Internet]. 2009 Jul 21 [cited 2011 Feb 16];6(7):e1000100. Available from: http://dx.doi.org/10.1371/journal.pmed.1000100

615 13. Dalkey N, Helmer O. An Experimental Application of the Delphi Method to the
616 use of experts. Manag Sci. 1963;9(3):458–67.

617 14. Suh KS, Park SW, Castro A, Patel H, Blake P, Liang M, et al. Ovarian cancer
618 biomarkers for molecular biosensors and translational medicine. Expert Rev Mol
619 Diagn [Internet]. 2010 Nov [cited 2010 Dec 21];10(8):1069–83. Available from:
620 http://www.ncbi.nlm.nih.gov/pubmed/21080822

621 15. Losko S, Heumann K. Semantic data integration and knowledge management
622 to represent biological network associations. Methods Mol Biol Clifton NJ
623 [Internet]. 2009 [cited 2009 Jul 15];563:241–58. Available from:
624 http://www.ncbi.nlm.nih.gov/pubmed/19597789

625 16. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, et al. Knowledge
626 management for systems biology a general and visually driven framework
627 applied to translational medicine. BMC Syst Biol [Internet]. 2011 [cited 2011 Apr
628 27];5:38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21375767

629 17. Cano I, Tényi Á, Schueller C, Wolff M, Migueláñez MMH, Gomez-Cabrero D, et
630 al. The COPD Knowledge Base: enabling data analysis and computational
631 simulation in translational COPD research. J Transl Med [Internet]. 2014 Nov 28
632 [cited 2014 Nov 27];12(Suppl 2):S6. Available from: http://www.translational-
633 medicine.com/content/12/S2/S6/abstract

634 18. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et
635 al. ChEBI: a database and ontology for chemical entities of biological interest.
636 Nucleic Acids Res [Internet]. 2008 Jan [cited 2011 Jul 28];36(Database

637 issue):D344-350. Available from:
638 http://www.ncbi.nlm.nih.gov/pubmed/17932057

639 19. The international conference for the tenth revision of the International
640 Classification of Diseases. Strengthening of Epidemiological and Statistical
641 Services Unit. World Health Organization, Geneva. World Health Stat Q Rapp
642 Trimest Stat Sanit Mond. 1990;43(4):204–45.

643 20. Losko S, Wenger K, Kalus W, Ramge A, Wiehler J, Heumann K. Knowledge
644 Networks of Biological and Medical Data: An Exhaustive and Flexible Solution
645 to Model Life Science Domains. In: Data Integration in the Life Sciences.
646 Springer Berlin / Heidelberg; 2006. p. 232–9. (Lecture Notes in Computer
647 Science; vol. Volume 4075/2006).

648 21. Schriml LM, Arze C, Nadendla S, Chang Y-WW, Mazaitis M, Felix V, et al.
649 Disease Ontology: a backbone for disease semantic integration. Nucleic Acids
650 Res [Internet]. 2012 Jan [cited 2012 Jan 24];40(Database issue):D940-946.
651 Available from: http://www.ncbi.nlm.nih.gov/pubmed/22080554

652 22. de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus:
653 using science-based terminology to integrate cancer research results. Stud
654 Health Technol Inform [Internet]. 2004 [cited 2010 Nov 27];107(Pt 1):33–7.
655 Available from: http://www.ncbi.nlm.nih.gov/pubmed/15360769

656 23. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, et al.
657 Database resources of the National Center for Biotechnology Information.
658 Nucleic Acids Res. 2000 Jan 1;28(1):10–4.

24. Goto S, Nishioka T, Kanehisa M. LIGAND: chemical database for enzyme reactions. Bioinforma Oxf Engl [Internet]. 1998 [cited 2011 Jul 28];14(7):591–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9730924

25. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, et al. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst Biol. 2011 Mar 5;5:38.

26. Kumari B, Jain P, Das S, Ghosal S, Hazra B, Trivedi AC, et al. Dynamic changes in global microRNAome and transcriptome reveal complex miRNA-mRNA regulated host response to Japanese Encephalitis Virus in microglial cells. Sci Rep. 2016 Feb 3;6:20263.

27. Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR. Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell. 2016 May 5;18(5):591–6.

28. Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, et al. Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. Cell Stem Cell. 2016 May 5;18(5):587–90.

29. Zhang R, Miner JJ, Gorman MJ, Rausch K, Ramage H, White JP, et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature. 2016 07;535(7610):164–8.

30. Moore EF. The shortest path through a maze. In: Proceedings of the International Symposium on the Theory of Switching. Harvard University Press; 1959. p. 285–292.

681 31. Barrows NJ, Campos RK, Powell ST, Prasanth KR, Schott-Lerner G, Soto-
682 Acosta R, et al. A Screen of FDA-Approved Drugs for Inhibitors of Zika Virus
683 Infection. Cell Host Microbe. 2016 Aug 10;20(2):259–70.

684 32. Choy MM, Zhang SL, Costa VV, Tan HC, Horrevorts S, Ooi EE. Proteasome
685 Inhibition Suppresses Dengue Virus Egress in Antibody Dependent Infection.
686 PLoS Negl Trop Dis. 2015 Nov;9(11):e0004058.

687 33. Bublil EM, Yarden Y. The EGF receptor family: spearheading a merger of
688 signaling and therapeutics. Curr Opin Cell Biol. 2007 Apr;19(2):124–34.

689 34. Grant A, Ponia SS, Tripathi S, Balasubramaniam V, Miorin L, Sourisseau M, et
690 al. Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. Cell
691 Host Microbe. 2016 Jun 8;19(6):882–90.

692 35. Higgins J (editor), Green S (editor). Cochrane Handbook for Systematic
693 Reviews of Interventions [Internet]. 5.1.0. 2011. (The Cochrane Collaboration).
694 Available from: http://handbook.cochrane.org

695 36. Turner T, Misso M, Harris C, Green S. Development of evidence-based clinical
696 practice guidelines (CPGs): comparing approaches. Implement Sci [Internet].
697 2008 [cited 2016 Dec 15];3:45. Available from: http://dx.doi.org/10.1186/1748-
698 5908-3-45

699 37. Tényi Á, de Atauri P, Gomez-Cabrero D, Cano I, Clarke K, Falciani F, et al.
700 ChainRank, a chain prioritisation method for contextualisation of biological
701 networks. BMC Bioinformatics. 2016;17(1):17.

38. Scaturro P, Stukalov A, Haas DA, Cortese M, Draganova K, Płaszczyca A, et al. An orthogonal proteomic survey uncovers novel Zika virus host factors. Nature. 2018;561(7722):253–7.

39. Scaturro P, Kastner AL, Pichlmair A. Chasing Intracellular Zika Virus Using Proteomics. Viruses. 2019 Sep 19;11(9).

Supporting Information

S1 Fig. Molecular interaction and disease map generation protocol

S2 Fig. ZIKV KnowledgeBase Data Model

Based on input from clinical and virology experts the concepts required to represent existing pathophysiological knowledge of infectious diseases were modelled using the BioXM Knowledge management environment. To this end objects (as nodes) and relations (as edges) are defined on an abstract level such as “gene”, “disease” or “interacts with”. For the ZIKV KB we focused on concepts required to represent text mining results and information from structured databases of protein-protein and drug – protein interaction, namely genes, diseases, pathogens and drugs. Pathogens are displayed as yellow nodes and were uniquely identified by using the

NCBI Taxonomy ontology. Compounds (white nodes) were referenced to entries of the ChEBI ontology. Diseases (violet nodes) and genes (orange nodes) can reference one or multiple sources such as ICD10, Entrez or Ensembl. Relationships between pathogens, genes, drugs and compounds were defined by three types of text mining relations, upregulation (green edges), downregulation (red edges) and regulation (black edges). Associated information, not represented as object or relation, is being stored as so called “annotation” basically a note that can be assigned to any semantic concept. The annotation form “Textmining information evidence”, for example, stores information, such as the sentence containing the extracted statement (field “Reference”) or the predicate (field “Interaction type”) and links directly to the PubMed entry (represented as BioRS entry: MEDLINE). To enable experts to support or contradict extracted text mining relationships the annotation form “Textmining Validation” was introduced to generate a configurable curation workflow. Molecular interaction and disease maps are modelled using the “context” object type to group specific sub-networks (interactions between genes, diseases, pathogens and drugs) within the global semantic network. PPI and drug-protein interactions are represented by the “Interaction” relation (orange edges) with supportive evidence stored in two annotation forms (“BioGRID Interaction Evidence” and “Interaction Information” storing information from the VirHostNet database), and can be added as additional sub-networks to a disease map.

Based on the data model dynamic visualisation can be generated, for example drug targets or differentially expressed genes based on configurable queries and views. The data model is automatically transformed into a natural language like query and reporting language. This language can be used to e.g. define a query retrieving the number of drugs that interact with a protein of interest, and generating a view which

749 applies a colour code to a gene according to the number of interacting drugs.

750

751 **S3 Fig. Molecular interaction and disease map wizard.** (i) A text mining analysis
752 result (or multiple ones) can be chosen from to start populating the map. (ii) Based
753 on the genes extracted from text mining relationships (highlighted in orange, top left)
754 a network search algorithm is applied to extend the map with PPI & protein-drug
755 interactions. (iii) Upon selection of wanted interaction data the final map is
756 automatically rendered in different perspectives, displaying literature evidence,
757 experimental data or drug targets (shown from left to right, bottom).

758

759 **S4 Step-by-step guide to reproduce the results in the web interface**

760

761 **S5 Manual for web interface**

762

Elucidation of ZIKV pathogenesis/ biology
Identification of candidate drugs for repositioning

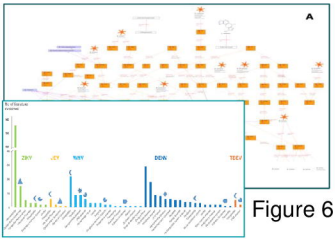


Figure 6

Figure 5

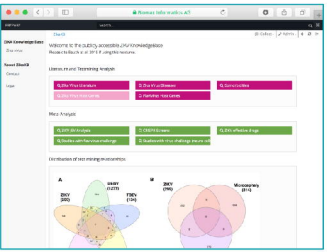
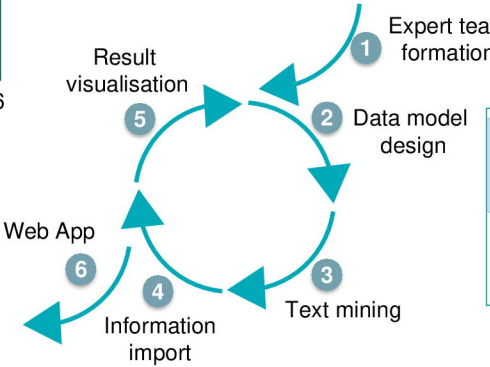
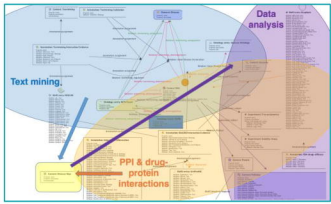


Figure 1 S5 Manual

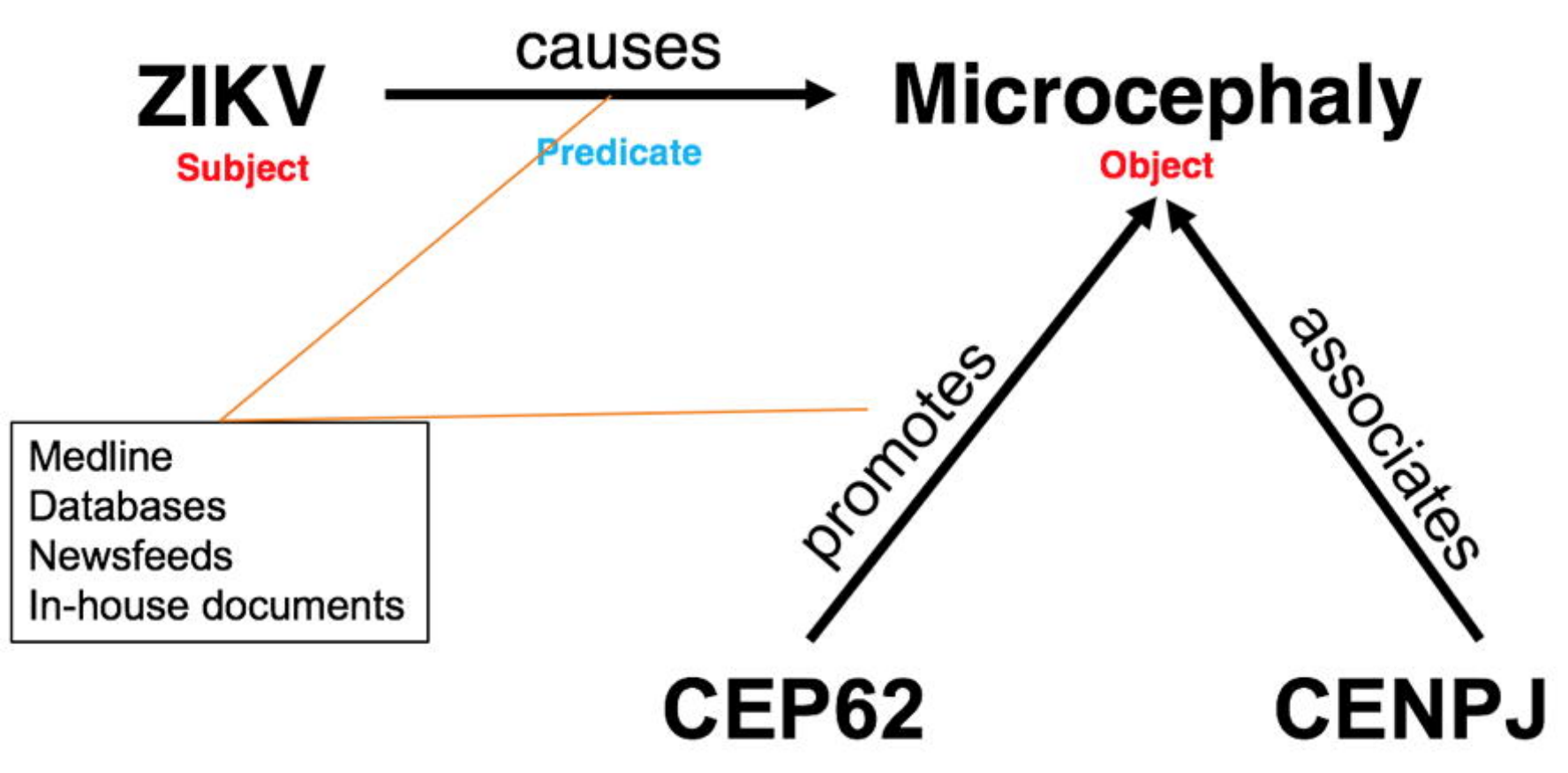


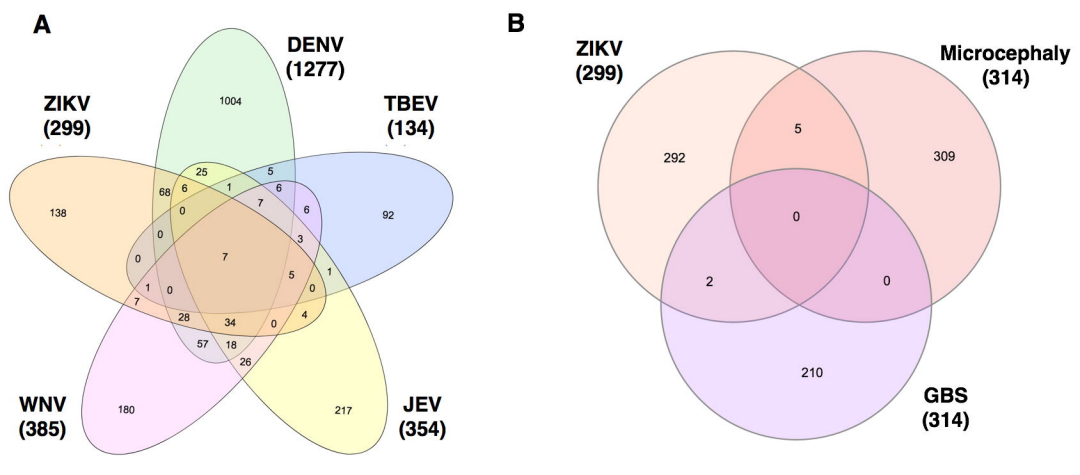


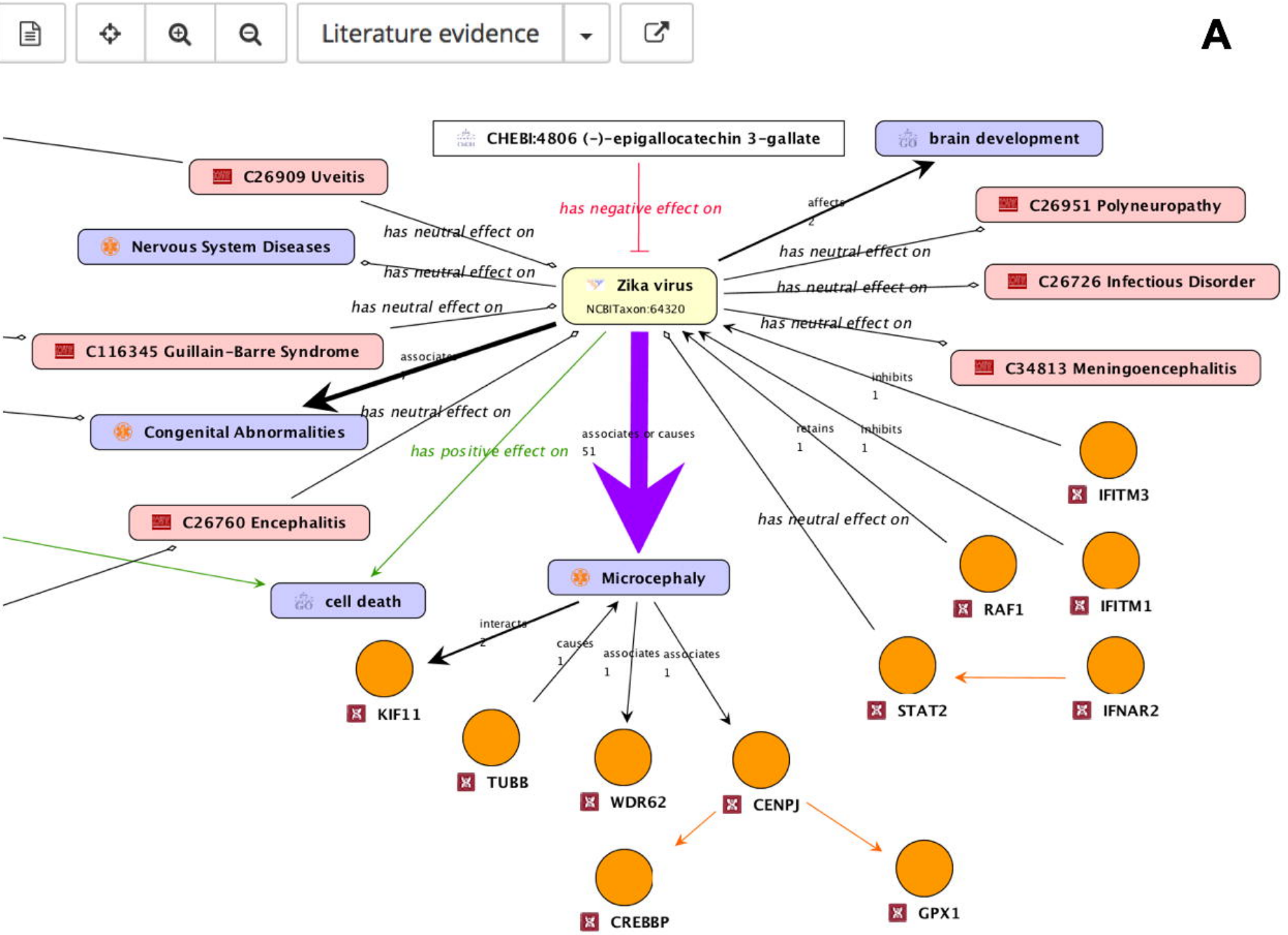
S2 Figure

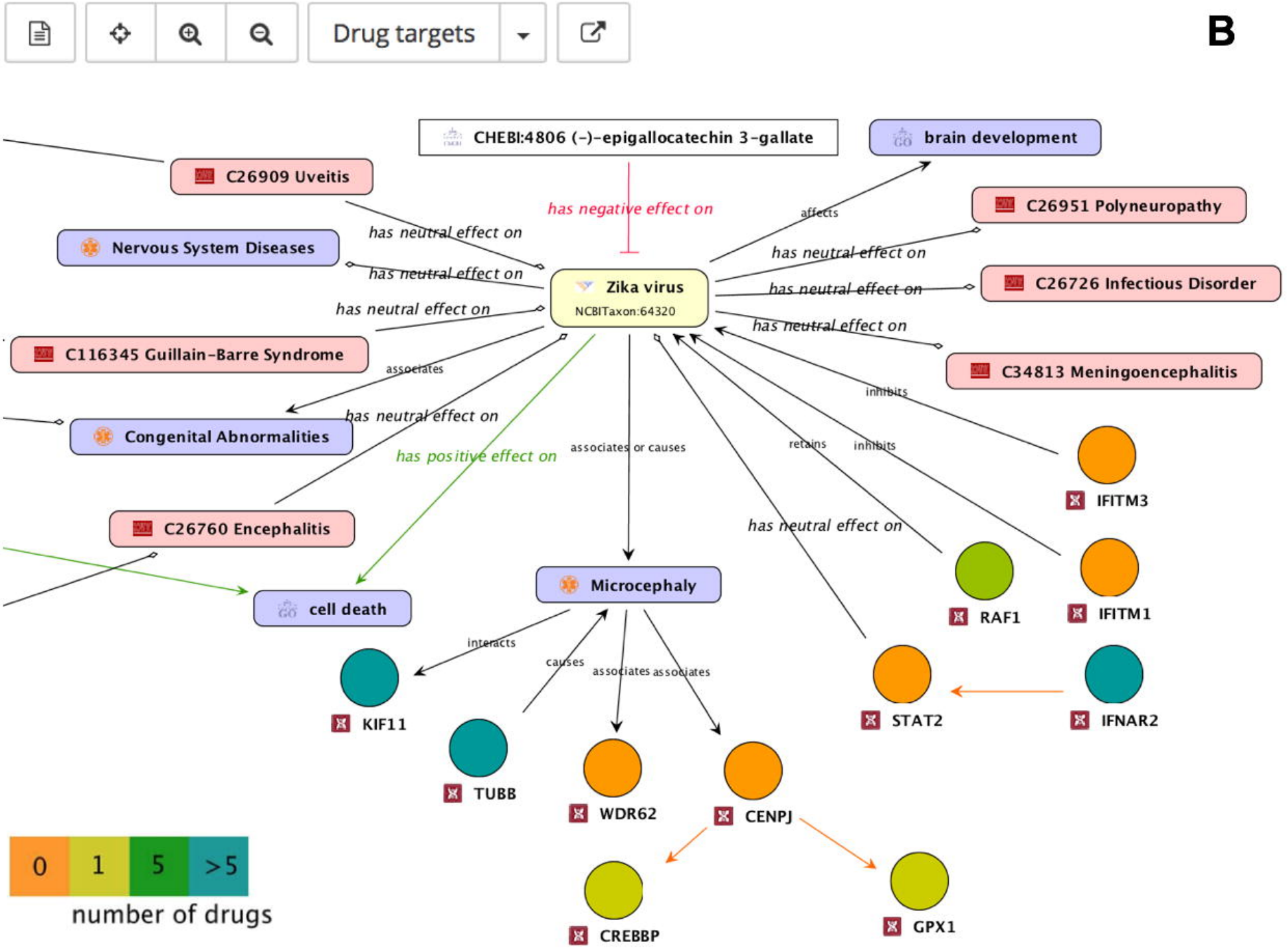
Data types:
Transcriptomics, Interactomics

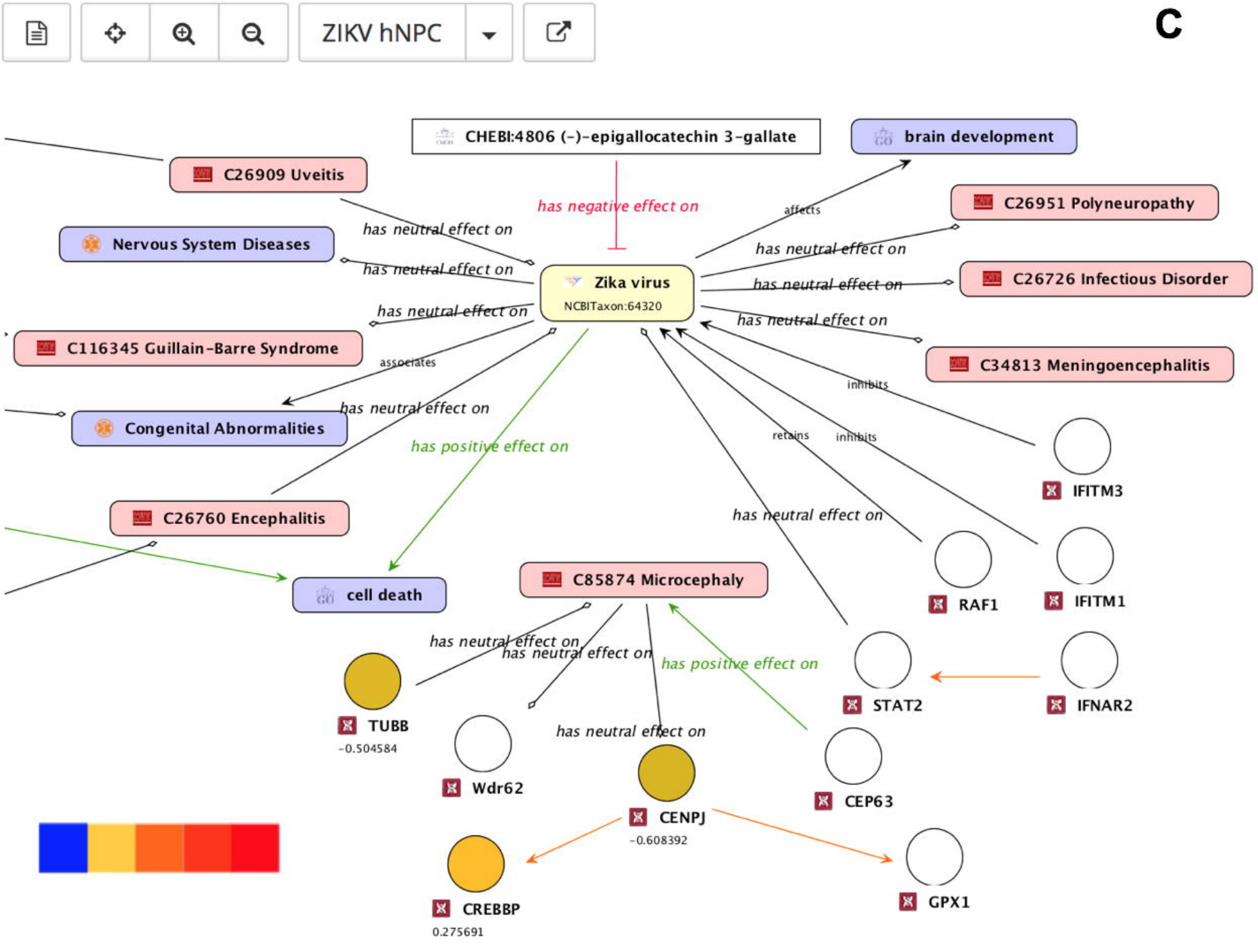
→ Molecular mechanisms between ZIKV and neurological disease
→ Novel research hypotheses

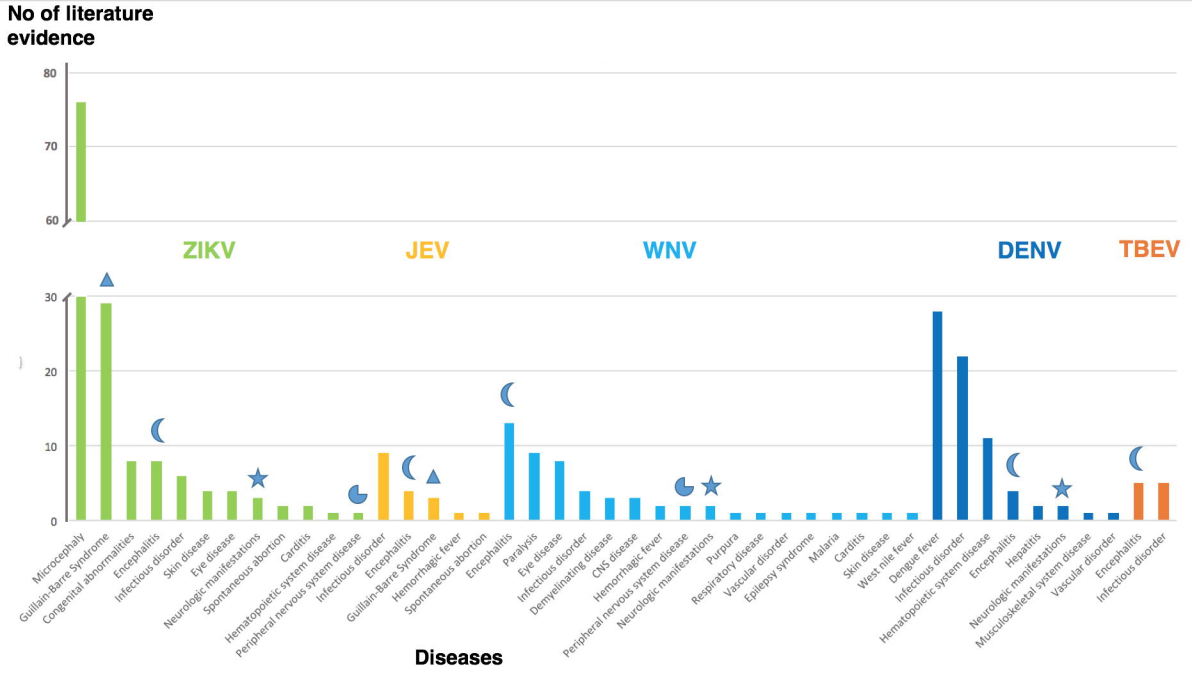

ZIKV
Subject
causes
Predicate
Microcephaly
Object
Medline
Databases
Newsfeeds
In-house documents
promotes
associates
CEP62
CENPJ

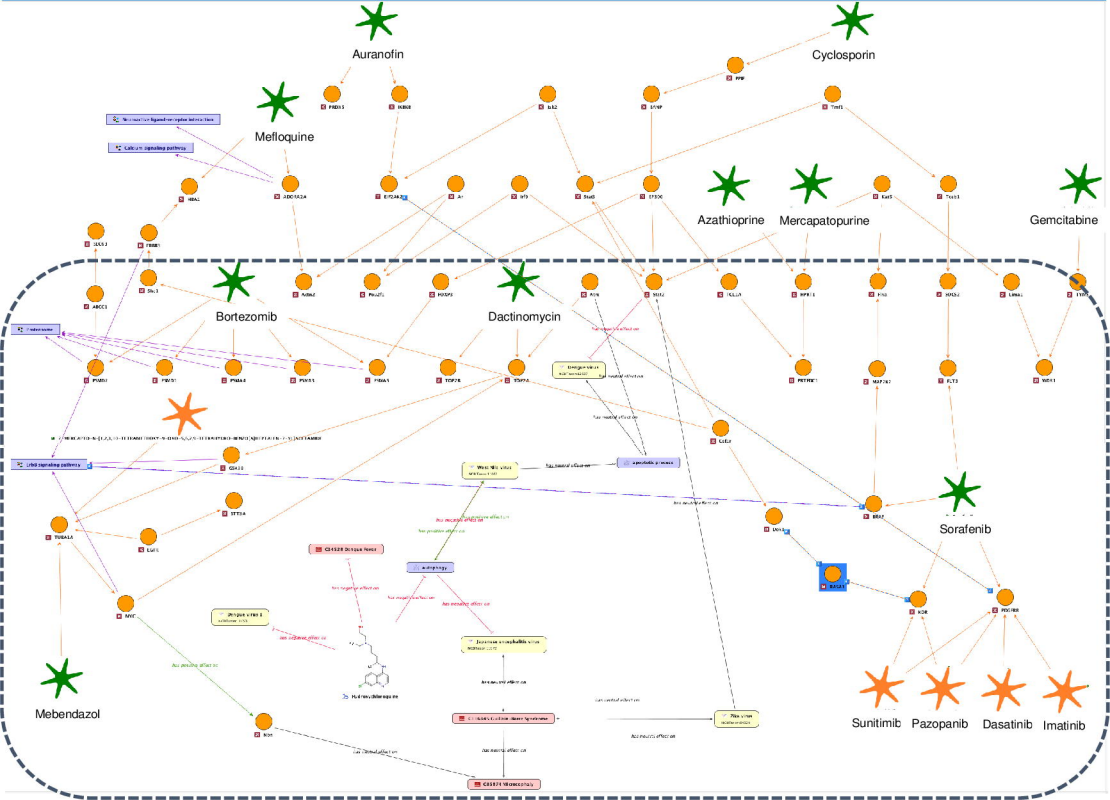

A
DENV
(1277)
ZIKV
(299)
TBEV
(134)
WNV
(385)
JEV
(354)
1004
138
92
180
217
25
5
68
6
1
6
0
7
6
0
3
0
7
5
1
0
1
0
7
28
34
0
4
57
18
26
B
ZIKV
(299)
Microcephaly
(314)
GBS
(314)
292
5
309
0
2
0
210

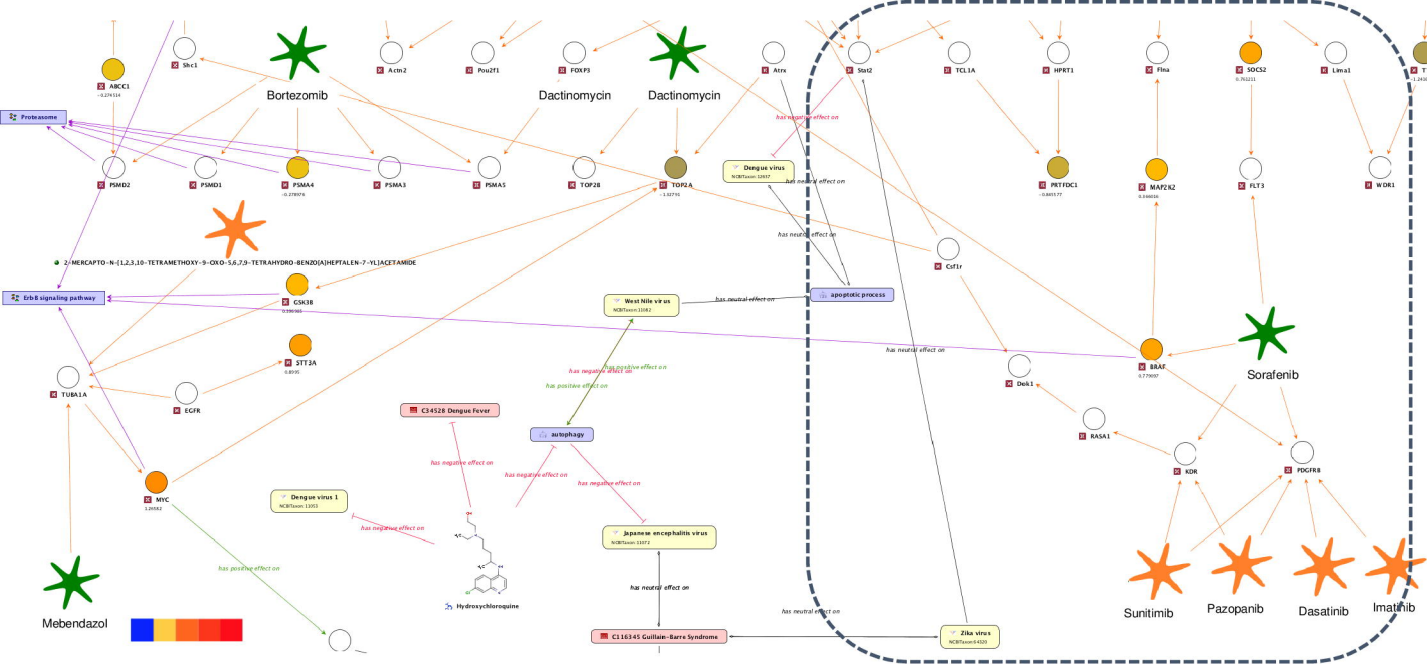

Literature evidence
A
CHEBI:4806 (-)-epigallocatechin 3-gallate
brain development
C26909 Uveitis
has negative effect on
affects
2
C26951 Polyneuropathy
has neutral effect on
Nervous System Diseases
has neutral effect on
has neutral effect on
Zika virus
NCBITaxon:64320
has neutral effect on
C26726 Infectious Disorder
has neutral effect on
C116345 Guillain-Barre Syndrome
associates
1
has neutral effect on
C34813 Meningoencephalitis
inhibits
1
Congenital Abnormalities
has neutral effect on
has positive effect on
associates or causes
51
retains
1
inhibits
1
IFITM3
C26760 Encephalitis
has neutral effect on
Microcephaly
RAF1
IFITM1
cell death
interacts
2
causes
1
associates
1
associates
1
KIF11
STAT2
IFNAR2
TUBB
WDR62
CENPJ
CREBBP
GPX1

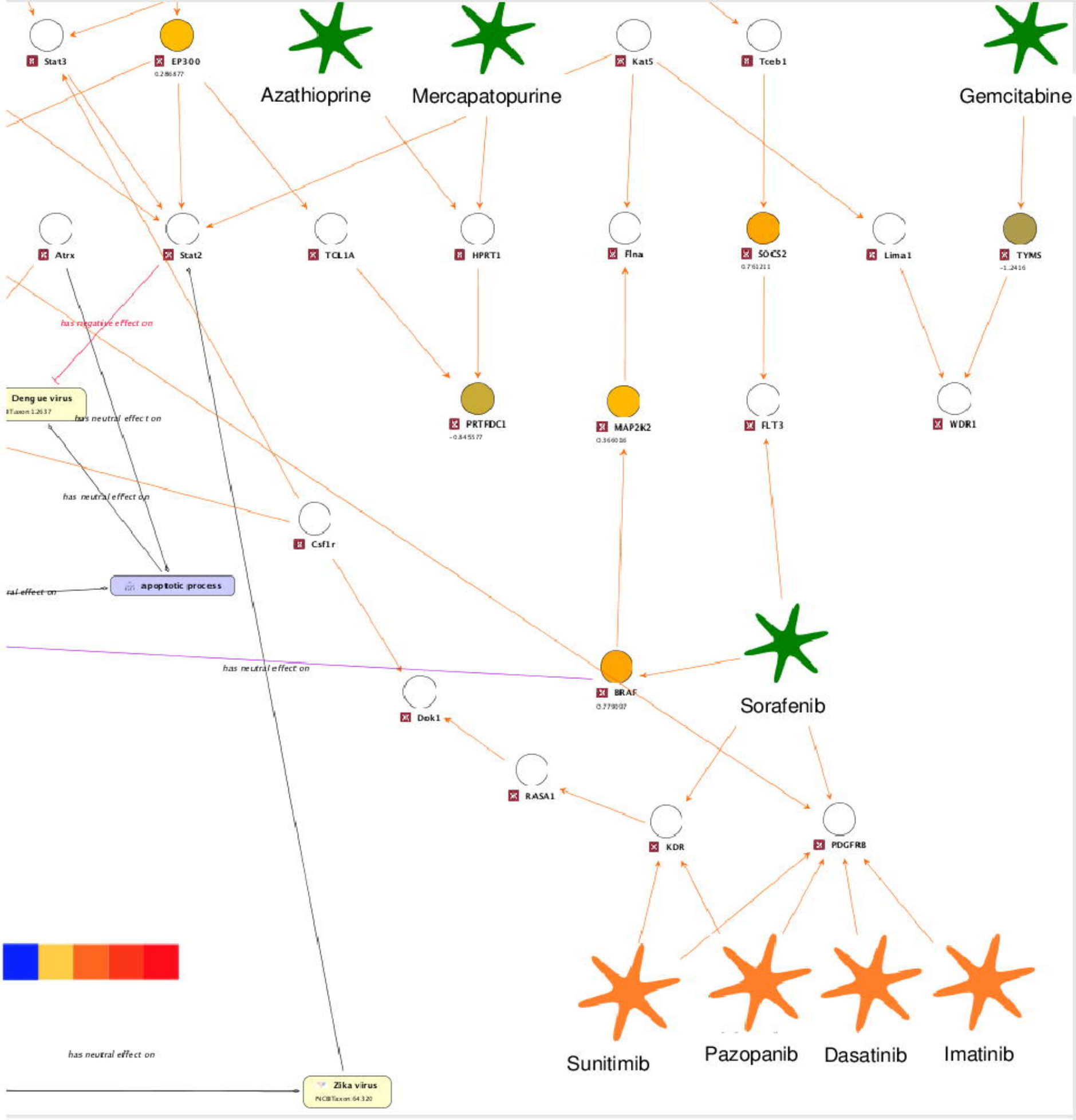

Drug targets
B
CHEBI:4806 (-)-epigallocatechin 3-gallate
brain development
C26909 Uveitis
C26951 Polyneuropathy
Nervous System Diseases
Zika virus
NCBITaxon:64320
C26726 Infectious Disorder
C116345 Guillain-Barre Syndrome
C34813 Meningoencephalitis
Congenital Abnormalities
C26760 Encephalitis
cell death
Microcephaly
IFITM3
IFITM1
RAF1
STAT2
IFNAR2
KIF11
TUBB
WDR62
CENPJ
CREBBP
GPX1
has negative effect on
has neutral effect on
has positive effect on
affects
associates
associates or causes
inhibits
retains
interacts
causes
associates
0
1
5
>5
number of drugs


ZIKV hNPC
C
CHEBI:4806 (-)-epigallocatechin 3-gallate
brain development
C26909 Uveitis
C26951 Polyneuropathy
Nervous System Diseases
Zika virus
NCBITaxon:64320
C26726 Infectious Disorder
C116345 Guillain-Barre Syndrome
C34813 Meningoencephalitis
Congenital Abnormalities
IFITM3
C26760 Encephalitis
cell death
C85874 Microcephaly
RAF1
IFITM1
STAT2
IFNAR2
TUBB
-0.504584
Wdr62
CENPJ
-0.608392
CEP63
CREBBP
0.275691
GPX1
has negative effect on
has neutral effect on
affects
has positive effect on
associates
inhibits
retains


No of literature evidence
80
70
60
30
20
10
0
ZIKV
JEV
WNV
DENV
TBEV
Microcephaly
Guillain-Barre Syndrome
Congenital abnormalities
Encephalitis
Infectious disorder
Skin disease
Eye disease
Neurologic manifestations
Spontaneous abortion
Carditis
Hematopoietic system disease
Peripheral nervous system disease
Infectious disorder
Encephalitis
Guillain-Barre Syndrome
Hemorrhagic fever
Spontaneous abortion
Encephalitis
Paralysis
Eye disease
Infectious disorder
Demyelinating disease
CNS disease
Hemorrhagic fever
Peripheral nervous system disease
Neurologic manifestations
Purpura
Respiratory disease
Vascular disorder
Epilepsy syndrome
Malaria
Carditis
Skin disease
West nile fever
Dengue fever
Infectious disorder
Hematopoietic system disease
Encephalitis
Hepatitis
Neurologic manifestations
Musculoskeletal system disease
Vascular disorder
Encephalitis
Infectious disorder
Diseases


Auranofin
Cyclosporin
Mefloquine
Azathioprine
Mercapatopurine
Gemcitabine
Bortezomib
Dactinomycin
Sorafenib
Mebendazole
Sunitimib
Pazopanib
Dasatinib
Imatinib


ABCC1
-0.229114
Shc1
Bortezomib
Actn2
Pou2f1
FOXP3
Dactinomycin
Dactinomycin
Atrx
Stat2
TCL1A
HPRT1
Flna
SOCS2
0.763211
Lima1
Proteasome
PSMD2
PSMD1
PSMA4
-0.278974
PSMA3
PSMA5
TOP2B
TOP2A
-1.12731
Dengue virus
NCBITaxon:12637
has neutral effect on
has neutral effect on
has neutral effect on
PKTFDC1
-0.840177
MAF2K2
0.144034
FLT3
WDR1
2-MERCAPTO-N-[1,2,3,10-TETRAMETHOXY-9-OXO-5,6,7,9-TETRAHYDRO-BENZO[A]HEPTALEN-7-YL]ACETAMIDE
ErbB signaling pathway
GSK3B
0.396365
STAT3A
0.8195
Csf1r
West Nile virus
NCBITaxon:11082
has neutral effect on
apoptotic process
has neutral effect on
has negative effect on
has positive effect on
BRAF
0.77697
Sorafenib
Dek1
TUBA1A
EGFR
C34528 Dengue Fever
autophagy
RASA1
KDR
PDGFRB
MYC
1.29582
Dengue virus 1
NCBITaxon:11053
has negative effect on
has negative effect on
has negative effect on
has negative effect on
Japanese encephalitis virus
NCBITaxon:11072
Hydroxychloroquine
has positive effect on
has neutral effect on
Mebendazol
C116345 Guillain-Barre Syndrome
has neutral effect on
Zika virus
NCBITaxon:64320
Sunitimib
Pazopanib
Dasatinib
Imatinib


Stat3
EP300
0.286877
Azathioprine
Mercapatopurine
Kat5
Tceb1
Gemcitabine
Atrx
Stat2
TCL1A
HPRT1
Flna
SOCS2
0.761211
Lima1
TYMS
-1.2416
has negative effect on
Dengue virus
has neutral effect on
PRTFDC1
-0.845577
MAP2K2
0.366036
FLT3
WDR1
has neutral effect on
Csf1r
apoptotic process
has neutral effect on
BRAF
0.779037
Sorafenib
has neutral effect on
Dok1
RASA1
KDR
PDGFRB
Sunitimib
Pazopanib
Dasatinib
Imatinib
has neutral effect on
Zika virus