Molecular fusion events in carcinogenic organisms: a bioinformatics study for the detection of fused proteins between viruses, bacteria and eukaryotes.

Eleni Papakonstantinou, Kalliopi Io Diakou, Thanasis Mitsis, Konstantina Dragoumani, Flora Bacopoulou, Vasilis Megalooikonomou, Sophia Kossida, George P Chrousos, Dimitrios Vlachakis

Abstract

Molecular fusion events have a prominent role in the initial steps of carcinogenesis. In this study, a bioinformatics analysis was performed between four organisms that are known to induce cancer development in humans: two viruses, *Human Herpesvirus 4*, and *Human T-cell leukaemia* virus, one bacterium, *Helicobacter Pylori,* and one trematode, *Schistosoma mansoni*. The annotated proteomes from these organisms were analysed using the SAFE software to identify protein fusion events, which may provide insight into protein function similarities and possible merging events during the course of evolution. Based on the results, five fused proteins with very similar functions were detected, whereas proteins with different functions that might act in the same molecular complex or biochemical pathway were not found. Thus, this study analysed the above four well-known cancer-related organisms with *de novo* bioinformatics programs and provided useful information on protein fusion events, hopefully leading to deeper understanding of carcinogenenesis.

Key Points

- Molecular fusion events have a prominent role in the initial steps of carcinogenesis.
- Identification of fusion events provide a useful basis for the discovery of novel potential therapeutic targets against cancer and other diseases.
- A bioinformatics analysis was performed between four organisms that are known to induce cancer development in humans.
- Four fusion events and one linked protein were identified in the present study from the comparison analysis of four prominent cancer-related organisms.

Introduction

Cancer is one of the leading causes of death in the world with 9.6 million cancer related deaths in 2018 and a projection of more than 16 million by the year 2040 (World Health Organisation). Carcinogenic effects appear with the transformation of a normal cell into a tumour cell and its unrestrained proliferation, with potential to invade beyond normal tissue boundaries and metastasize to distant organs (Tomasetti *et al.*, 2017). Genetic factors are involved in cancer, but external agents can also play a key role in the disruptive cellular changes. These external factors are separated into three main categories, the physical carcinogens, the chemical carcinogens and the biological carcinogens (Vineis *et al.*, 2010). This study focuses on the biological type of carcinogens, with the goal to provide new perspectives on cancer-generating events. The term biological carcinogens can refer to infections from various organisms, such as certain viruses, bacteria or parasites. Herein, possible molecular fusion events and probably linked proteins are studied between organisms including bacteria (*Helicobacter pylori I)*, eukaryotes (*Schistosoma* flatworms) and viruses *(Human Herpesvirus 4* and *Human T-cell leukaemia virus*).

Helicobacter Pylori is a non-invasive, gram-negative bacterium which colonizes the stomach and is present in almost half of the human population. It has been classified as a member of the group I carcinogenic agents by the International Agency for Research on Cancer (Thorell *et al.*, 2017). The transmission of these bacteria is interhuman and infection often occurs during childhood. In most cases, the bacteria can live in the human stomach for years without causing any problem (Burucoa and Axon, 2017; Lopes *et al.*, 2014). *H. pylori* is considered to be the most important risk factor for gastric cancer (Polk and Peek, 2010). Infection by this bacterium induces chronic gastritis that can progress to intestinal dysplasia and, through complex mechanisms, can ultimately lead to gastric cancer (Parkin *et al.*, 2005). Studies *in vivo* showed that *H. pylori* infection induced double-stranded breaks, a type of DNA damage (Xie *et al.*, 2014). Endogenous DNA damage is inherently linked to genomic instability, which in turn is one of the most prevalent onsets of tumorigenesis. Meta-analysis of randomized control trials showed that eradication of the bacterium may reduce the risk of gastric cancer (Liou *et al.*, 2020), while the safety and efficacy of the eradication was reviewed in a recent study (Liou *et al.*, 2019).

Schistosoma (blood flukes) is a genus of parasitic flatworms that are responsible for schistosomiasis (Utzinger *et al.*, 2009). This disease affects millions of people, especially in developing countries (Africa and South America) and is considered the second most socioeconomically devastating parasitic disease after malaria (World Health Organization). Out of the five species that infect humans, *Schistosoma mansoni* is the most common and generally the one used in laboratory studies. Humans are their primary host, but the larvae need to pass through an intermediate freshwater snail host to infect another mammalian host (Aguiar *et al.*, 2017; Protasio *et al.*, 2012). Clinical manifestations of the disease pass through various acute, sub-acute and chronic stages. Schistosomiasis has been associated with development of malignancy in the rectum, bladder and lymphoid tissue (Palumbo, 2007; Barsoum *et al.*, 2013). Recently, glutathione transferase from *Schistosoma japonicum* was studied as a potential drug design target towards this parasite (Platis *et al.*, 2020).

Human Herpesvirus 4 (HHV-4), also called *Epstein–Barr virus* (EBV), is one of the most common viruses in humans, affecting nearly 90% of the adult population in the world. The virus is transmitted by saliva and, in the majority of cases, it causes infectious mononucleosis (Smatti, Al-Sadeq *et al.*, 2018). Additionally, the virus is aetiologically linked to two pre-malignant lymphoproliferative diseases (LPDs) and up to nine distinct human tumours. The LPDs include B-cell origin diseases, such as Burkitt lymphoma, Hodgkin lymphoma and immune impairment associated tumour pathologies, such as Plasmablastic lymphoma (Kanda *et al.*, 2019). Furthermore, the virus is involved in various tumour pathologies like nasopharyngeal cancer, gastric carcinoma and leiomyosarcoma (Abe, Kaneda *et al.*, 2015) (Hall *et al.*, 2015).

Human T-cell leukaemia virus is the first pathogenic human retrovirus to have been discovered (Coffin, 2015). *Human T-cell leukaemia virus type I* (HTLV-I) is the most common one, having infected an estimated 5-20 million individuals. Among other disorders, HTLV-1 causes a form of T-cell lymphoproliferation, characterized as leukemia or lymphoma, and termed adult T-cell leukemia (ATL) or adult T-cell leukemia-lymphoma (ATLL) (Cook *et al.*, 2017; Meissner *et al.*, 2017). Therapies that block HTLV-1 replication include integrase inhibitors and nucleoside reverse transcriptase inhibitors, while a combination of interferon alpha and zidovudine seems to have significant effects on the chronic/acute forms of ATLL and not on the lymphoma sub-type of ATLL (Nasr, El-Hajj *et al.*, 2011). *Human T-cell leukaemia virus type II* (HTLV-II) has been related to T-cell variant of hairy cell leukaemia, and like HTLV-I, it is capable of transforming normal human peripheral blood into lymphocytes *in vitro,* however its mechanism remains elusive (Murphy 2016; Shima *et al.*, 1986).

Proteins control almost all biological systems in a cell, and while some proteins work independently, the vast majority perform their actions in collaboration with other proteins. Protein-protein interactions are established in all cellular processes, such as in the cell cycle and intracellular signalling, and are also prognostic factors of diseases such as cancer (Chen, Sam *et al.*, 2010). As the same proteins may be involved in different cellular processes, the study of their interactions allows finding correlations between different diseases or cellular conditions, and promotes research in the field of drug design (Tsaniras SC *et al.*, 2015). The analysis of protein-protein interaction is, therefore, critical for a more profound understanding of biological processes (Papageorgiou *et al.*, 2014; Steinhauf *et al.*, 2014). Although there are experimental methods towards determining protein-protein interactions, the techniques are labour-intensive, time consuming and not necessarily easily performed in high-throughput analyses (Berggard *et al.*, 2007). A useful alternative in the face of those hindrances is a bioinformatic approach, which can take multiple, equally interesting forms, as evidenced by various studies (Li, Wang *et al.*, 2018;Chen, Wang *et al.*, 2019)

Functional links between proteins can be examined through various ways. The occurrence of a fusion event between two proteins can suggest the likelihood of a functional connection between them, for example by either being part of the same protein complex, or by acting in the same pathway or biological network (Tsagrasoulis *et al.*, 2012). A bioinformatic analysis of protein fusion events between two species can, therefore, enable the detection of functional links between proteins (Enright *et al.*, 1999).

Molecular fusion events have a prominent role in the initial steps of carcinogenesis, where chromosome translocations and gene fusions result in the deregulation of physiological molecular mechanisms (Yu *et al.*, 2019). Gene fusions have been detected in all types of human neoplasias, with a varying proportion among different types (Mitelman *et al.*, 2007). The analysis of fusion events employs genomic structure and sequence analysis of two or more genomes to detect possibly connected protein pairs without any prior knowledge, which would not have been necessarily detected by experimental analysis (Dimitriadis *et al.*, 2011). Nowadays, the development of techniques, such as next generation sequencing (NGS) or transcriptome analysis, and of various bioinformatics algorithms, has led to the creation of several databases containing information on gene fusion events and their functions (Panigrahi *et al.*, 2018). A recent example of such an accomplishment is the establishment of the ChimberDB, an extensive database of fusion genes (Jang, Jang *et al.*, 2020). The SAFE software (Tsagrasoulis *et al.*, 2012) was used in the past towards the successful detection of protein interactions in several studies (Dimitriadis *et al.*, 2011; Trimpalis *et al.*, 2013; Vlachakis *et al.*, 2013). SAFE is an application used to identify, filter and visualize fusion events. It enables the analysis and representation of fusion proteins by performing pairwise alignments of protein sets, permitting an independent research on fusion proteins and their subsequent imaging in this specific application, thus simplifying the analysis and providing optimum results (Tsagrasoulis *et al.*, 2012).

Methods

Database sequence search

The proteome is the entire set of proteins expressed by a specific organism at a certain time (Jensen, 2006). In contrast to the genome, the proteome is not static and continually changes in response to external and internal events.

In order to run the comparison with SAFE software and detect fusion events, the proteome sequences of the four organisms were used. The FASTA files of the proteomes were downloaded by UniProtKB (http://www.uniprot.org), a freely accessible protein database containing protein sequences and biological information. More specifically, its "Proteomes" subsection allows access to the whole proteome of numerous organisms. It includes both manually reviewed (UniProtKB/Swiss-Prot) and unreviewed (UniProtKB/TrEMBL) entries. The carcinogenic organisms included in their study and their respective UniProt Proteome ID are listed in Table 1.

Fusion events analyses

Fusion events analyses were performed with SAFE software (Software for the Analysis of Fusion Events 3). SAFE is designed to search for fusion events between a set of organisms given as input in a FASTA file format, one file for each organism to analyse.

Its algorithm is implemented as follows: The first task conducted is the removal of potentially duplicated proteins from each sequence file. This is done according to a user-defined parameter, which represents the percentage of minimal identities to consider two proteins as duplicated ones. Therefore, for each FASTA proteome file, each protein is blasted against its respective organism proteome, considering the user parameter "Max Blast identities" to identify duplicates. Within this step, each shorter duplicated protein over the max blast identity threshold is removed from the results. The proteins that are found above threshold are then saved in a new FASTA file called [initial_file_name] _reduced.txt.

The second task to perform is the actual analysis of fusion events from this reduced proteomes, using the following parameter for the fusion detection processes: Min. Domain Length: 70; Min. Blast Identities: 27; Min. Fused Protein coverage: 70; Max. overlaps Region in Domains: 0; Multiple Proteins cut-off: 5; E-value: 9 .10−3. The main SAFE algorithm is described in Figure 1.

Results

The proteomes of the two viruses, *Human Herpesvirus 4* and *Human T-cell leukaemia virus 2*, that can cause cancer in humans, as well as the proteomes of *Helicobater pylori*, a gram-negative microaerophilic bacterium of the human stomach that is correlated to gastric cancer, and *Schistosoma mansoni*, a human parasite that is responsible for intestinal schistosomiasis, were all analysed for potential protein fusion events. The following analysis pairs were carried out; the proteome of *Helicobacter pylori* against the proteome *of Schistosoma mansoni*, the proteome of *Helicobacter pylori* against the proteome of *Human T-cell leukaemia virus 2*, the proteome of *Human T-cell leukaemia virus 2* against *Schistosoma mansoni* and the proteome of *Human T-cell leukaemia virus 2* against *Human Herpesvirus 4,* as well as the backward BLAST analysis for each of these sets.

No putative protein fusion events in any comparison within the two viruses were identified. 20 fusion events were detected in the comparison of *Helicobacter pylori* and *Schistosoma mansoni* proteomes. Out of those 20 fusion events, only 5 of them are proteins that satisfy the unique protein threshold. 19 were detected in the proteome of *Helicobater pylori* (4 of them above the unique protein cut-off) and 1 was detected in the proteome of *Schistosoma mansoni*, which also satisfies the unique protein cut-off (Table 2).

The fusion event that took place in *Schistosoma mansoni* represents the fusion of two ATPase domains, one from a ATP-dependent zink metalloprotease FtsH and one from a cell division protein FtsH from the AAA family (ATPases Associated with diverse cellular Activities) from *Helicobater pylori* into a cell division control protein 48 from the AAA family in *Schistosoma mansoni*. In this case, only parts of the proteins from *Schistosoma mansoni* were used to build domains in the fused protein in *Helicobater pylori* (Figure 2A). The fused protein has the same predicted enzymatic function as the two original proteins.

In all four fusion events detected in *Helicobater pylori* it was visible that almost the full length of the two proteins from *Schistosoma mansoni* fused to form the majority of the protein in *Helicobater pylori* (Figure 2). In three of these cases, the proteins from *Schistosoma mansoni* had the same function, leading to a protein with similar function in *Helicobater pylori*. In one case, the two original proteins seem to have slightly different functions, which lead to a fused protein that incorporates both functions.

In the first fusion event, two putative Radical SAM proteins containing one $[4Fe\text{-}4S]^{+}$ cluster each, involved in RNA modifications, fused almost entirely to form tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase (miaB) in *Helicobacter pylori*, containing two $[4Fe\text{-}4S]^{+}$ clusters (Figure 2). In the second fusion event, two Phosphoglycerate kinases from *Schistosoma mansoni* fused entirely to form a Phosphoglycerate kinase in *Helicobater pylori* (Figure 2). In the third fusion event, the C-terminal 3/4 part of a threonyl-tRNA synthetase, including the threonyl-tRNA synthetase-domain, was fused with the whole of an uncharacterized protein that carries an anticodon-binding domain (prediction carried out by InterPro (http://www.ebi.ac.uk/interpro). This fusion built a threonyl-tRNA synthetase in *Helicobater pylori* (Figure 2). In the last fusion event, two proteins with slightly different function fused to form an enzyme with both functions. An Aspartyl-tRNA synthetase with a putative asparagine-tRNA ligase activity was almost entirely fused with the C-terminal 3/4 of an Aspartyl-tRNA synthetase with a putative aspartate-tRNA ligase activity, to form the majority of an Aspartate-tRNA (Asp/Asn) ligase in *Helicobacter pylori*. The *H. pylori* ligase presents a relaxed tRNA specificity since it can aspartylate not only its cognate aspertat-tRNA, but also asparagine-tRNA (Figure 2).

Discussion

No fusion events were discovered when the two viruses, *Human Herpesvirus 4* and *Human T-cell leukaemia virus 2*, were compared. This may be an expected result when one considers that viruses do not share a lot with other evolution branches. Moreover, the chosen viruses have a very small genome (Vlachakis *et al.*, 2013), which inherently limits the chances of finding fusion events amongst them and against other organisms. On the other hand, the fusion events detected through the comparison of *Helicobater pylori* and *Schistosoma mansoni* were between proteins of the same or similar function. The performed analysis was not able to detect potential interactions of proteins with different function that may act in the same complex, or the same pathway. This does not mean that the fusion events that were detected could not point to a potential interaction of proteins with similar function, rather that the fusion events may have been stabilised during the course of evolution. For example, a Phosphoglycerate kinase with two phosphoglyceratekinase domains may be more efficient than a Phosphoglycerate kinase with only one.

The fusion events that were found by comparing the proteomes of *Schistosoma mansoni* and *Helicobater pylori* were almost all found in *Helicobater pylori* (19 out of 20) and only one was found in *Schistosoma mansoni*. *Schistosoma mansoni* is a worm and thereby higher up in the "tree of life" than *Helicobater pylori*, which is a prokaryote. These fusion events seem to be in opposite direction of evolution and could also represent protein fission events.

The detected fusion proteins were a cell division control protein 48 from the AAA family in *Schistosoma mansoni*, containing ATPase domains from the proteins in *Helicobater pylori*. A tRNAsynthase in *Helicobacter pylori* containing two $[4Fe\text{-}4S]^{+}$ cluster created through the fusion of two proteins with one $[4Fe\text{-}4S]^{+}$ cluster each from *Schistosoma mansoni*. A Phosphoglycerate kinase in *Helicobater pylori* created through the fusion of two Phosphoglycerate kinase from *Schistosoma mansoni*. A threonyl-tRNA synthetase in *Helicobater pylori* created through the fusion a threonyl-tRNA synthetase, including the threonyl-tRNA synthetase-domain fused with an uncharacterized protein that carries an anticodon-binding domain from *Schistosoma mansoni*. An Aspartate-tRNA (Asp/Asn) ligase in *Helicobacter pylori* that it is able to aspartylate aspertat-tRNA and asparagine-tRNA created through the fusion of two threonyl-tRNA synthetase, one with a putative asparagine-tRNA ligase activity and one with a putative aspartate-tRNA ligase activity.

The present study conducted amongst these four organisms revealed a limited number of fusion events, thus it is difficult to make any statement on whether the proteins that were detected could be in any relation with a possible cancer induction stemming from the presence of the organism in humans. Replicated in a larger scale and for an increased number of carcinogenic organisms, the study of putative fusion events could reveal potentially interacting proteins and connection to cancer by analysing GO terms and involved pathways.

Conclusions

Chromosome translocations, chromosomal interstitial deletion and inversion, and the resultant fusion events frequently underlie cancer development, through deregulation of molecular mechanisms and the generation of fused gene products, which often possess oncogenic properties. Four fusion events and one linked protein were identified in the present study from the comparison analysis of four prominent cancer-related organisms using the SAFE software. Identification of such fusion events may provide a useful basis for the discovery of novel potential therapeutic targets against cancer and other diseases.

Acknowledgements

DV would like to acknowledge funding from: i) AdjustEBOVGP-Dx (RIA2018EF-2081): Biochemical Adjustments of native EBOV Glycoprotein in Patient Sample to Unmask target Epitopes for Rapid Diagnostic Testing. A European and Developing Countries Clinical Trials Partnership (EDCTP2) under the Horizon 2020 'Research and Innovation Actions' DESCA, and ii) "MilkSafe: A novel pipeline to enrich formula milk using omics technologies", a research co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T2EDK-02222).

Figure legends

Figure 1. SAFE software, the main algorithm steps.

Figure 2. Graphical representation of the discovered fusion events, that passed the unique cutoff filter, recuperated from the SAFE software. (A) The fusion event detected leading to a cell division control protein 48 from the AAA family in *Schistosoma mansoni*, containing ATPase domains from the proteins in *Helicobater pylori*. (B-E) The fusion events detected in *Helicobater pylori.* (B) The tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase (miaB) in *Helicobacter pylori* containing two $[4Fe\text{-}4S]^{+}$ cluster created through the fusion of two putative radical SAM proteins with each one $[4Fe\text{-}4S]^{+}$cluster from *Schistosoma mansoni*. (C) The Phosphoglycerate kinase in *Helicobater pylori* created through the fusion of two Phosphoglycerate kinase from *Schistosoma mansoni*. (D) A threonyl-tRNA synthetase in *Helicobater pylori* created through the fusion of the C-terminal 3/4 part of a threonyl-tRNA synthetase, including the threonyl-tRNA synthetase-domain are fused with an uncharacterized protein that carries an anticodon-binding domain from *Schistosoma mansoni*. (E) An Aspartate-tRNA(Asp/Asn) ligase in *Helicobacter pylori* that it is able to aspartylate aspertat-tRNA and asparagine-tRNA created through the fusion of two threonyl-tRNA synthetase, one with a putative asparagine-tRNA ligase activity and one with a putative aspartate-tRNA ligase activity.

Tables

Table 1: List of selected organisms.

Organism common name	Proteome Name	UniProt Proteome ID
Human herpesvirus 4	Epstein-Barr virus (strain B95-8)	UP000007640
Helicobacter Pylori	Helicobacter pylori (strain ATCC 700392 / 26695) (Campylobacter pylori)	UP000000429
Human T-cell leukaemia virus	Human T-cell leukaemia virus 2	UP000009254
***Schistosoma mansoni* (Blood fluke)**	Schistosoma mansoni Puerto Rican	UP000008854

Table 2: Detected fusion events between the analysed organisms. The number in brackets indicates the fusion events without considering the unique protein cut-off.

	Human Herpesvirus 4	***Human T-cell leukaemia virus 2***	***Helicobater pylori***	***Schistosoma mansoni***
Human Herpesvirus 4	-	0	0	0
Human T-cell leukaemia virus 2	0	-	0	0
Helicobater pylori	0	0	-	1 (1)
Schistosoma mansoni	0	0	4 (19)	-

References

Abe H, Kaneda A and Fukayama M (2015) Epstein-Barr Virus-Associated Gastric Carcinoma: Use of Host Cell Machineries and Somatic Gene Mutations. *Pathobiology* **82**(5), 212-223. http://dx.doi.org/10.1159/000434683

Aguiar PHN, Fernandes NMGS, Zani CL and Mourão MM (2017) A high-throughput colorimetric assay for detection of Schistosoma mansoni viability based on the tetrazolium salt XTT. *Parasites & vectors* **10**(1), 300-300. http://dx.doi.org/10.1186/s13071-017-2240-3

Barsoum RS, Esmat G and El-Baz T (2013) Human Schistosomiasis: Clinical Perspective: Review. *Journal of Advanced Research* **4**(5), 433-444. http://dx.doi.org/https://doi.org/10.1016/j.jare.2013.01.005

Berggård T, Linse S and James P (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7**(16), 2833-2842. http://dx.doi.org/10.1002/pmic.200700131

Burucoa C and Axon A (2017) Epidemiology of Helicobacter pylori infection. *Helicobacter* **22 Suppl 1**. http://dx.doi.org/10.1111/hel.12403

Chen J, Sam L, Huang Y, Lee Y, Li J *et al.* (2010) Protein interaction network underpins concordant prognosis among heterogeneous breast cancer signatures. *Journal of Biomedical Informatics* **43**(3), 385-396. http://dx.doi.org/https://doi.org/10.1016/j.jbi.2010.03.009

Chen K-H, Wang T-F and Hu Y-J (2019) Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC Bioinformatics* **20**(1), 308. http://dx.doi.org/10.1186/s12859-019-2907-1

Coffin JM (2015) The discovery of HTLV-1, the first pathogenic human retrovirus. *Proc Natl Acad Sci U S A* **112**(51), 15525-15529. http://dx.doi.org/10.1073/pnas.1521629112

Cook L, Melamed A, Yaguchi H and Bangham CR (2017) The impact of HTLV-1 on the cellular genome. *Curr Opin Virol* **26**, 125-131. http://dx.doi.org/10.1016/j.coviro.2017.07.013

Dimitriadis D, Koumandou VL, Trimpalis P and Kossida S (2011) Protein functional links in Trypanosoma brucei, identified by gene fusion analysis. *BMC Evolutionary Biology* **11**(1), 193. http://dx.doi.org/10.1186/1471-2148-11-193

Enright AJ, Iliopoulos I, Kyrpides NC and Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**(6757), 86-90. http://dx.doi.org/10.1038/47056

Hall LD, Eminger LA, Hesterman KS and Heymann WR (2015) Epstein-Barr virus: dermatologic associations and implications: part I. Mucocutaneous manifestations of Epstein-Barr virus and nonmalignant disorders. *J Am Acad Dermatol* **72**(1), 1-19; quiz 19-20. http://dx.doi.org/10.1016/j.jaad.2014.07.034

Jang YE, Jang I, Kim S, Cho S, Kim D *et al.* (2020) ChimerDB 4.0: an updated and expanded database of fusion genes. *Nucleic Acids Res* **48**(D1), D817-d824. http://dx.doi.org/10.1093/nar/gkz1013

Jensen ON (2006) Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology* **7**(6), 391-403. http://dx.doi.org/10.1038/nrm1939

Kanda T, Yajima M and Ikuta K (2019) Epstein-Barr virus strain variation and cancer. *Cancer Sci* **110**(4), 1132-1139. http://dx.doi.org/10.1111/cas.13954

Li LP, Wang YB, You ZH, Li Y and An JY (2018) PCLPred: A Bioinformatics Method for Predicting Protein-Protein Interactions by Combining Relevance Vector Machine Model with Low-Rank Matrix Approximation. *Int J Mol Sci* **19**(4). http://dx.doi.org/10.3390/ijms19041029

Liou JM, Lee YC, El-Omar EM and Wu MS (2019) Efficacy and Long-Term Safety of H. pylori Eradication for Gastric Cancer Prevention. *Cancers (Basel)* **11**(5). http://dx.doi.org/10.3390/cancers11050593

Liou JM, Malfertheiner P, Lee YC, Sheu BS, Sugano K *et al.* (2020) Screening and eradication of Helicobacter pylori for gastric cancer prevention: the Taipei global consensus. *Gut* **69**(12), 2093-2112. http://dx.doi.org/10.1136/gutjnl-2020-322368

Lopes D, Nunes C, Martins MC, Sarmento B and Reis S (2014) Eradication of Helicobacter pylori: Past, present and future. *J Control Release* **189**, 169-186. http://dx.doi.org/10.1016/j.jconrel.2014.06.020

Meissner ME, Mendonça LM, Zhang W and Mansky LM (2017) Polymorphic Nature of Human T-Cell Leukemia Virus Type 1 Particle Cores as Revealed through Characterization of a Chronically Infected Cell Line. *J Virol* **91**(16). http://dx.doi.org/10.1128/jvi.00369-17

Mitelman F, Johansson B and Mertens F (2007) The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**(4), 233-245. http://dx.doi.org/10.1038/nrc2091

Murphy EL (2016) Infection with human T-lymphotropic virus types-1 and -2 (HTLV-1 and -2): Implications for blood transfusion safety. *Transfus Clin Biol* **23**(1), 13-19. http://dx.doi.org/10.1016/j.tracli.2015.12.001

Nasr R, El Hajj H, Kfoury Y, de Thé H, Hermine O *et al.* (2011) Controversies in targeted therapy of adult T cell leukemia/lymphoma: ON target or OFF target effects? *Viruses* **3**(6), 750-769. http://dx.doi.org/10.3390/v3060750

Palumbo EM (2007) Association Between Schistosomiasis and Cancer: A Review. *Infectious Diseases in Clinical Practice* **15**, 145-148.

Panigrahi P, Jere A and Anamika K (2018) FusionHub: A unified web platform for annotation and visualization of gene fusion events in human cancer. *PLoS ONE* **13**(5), e0196588. http://dx.doi.org/https://doi.org/10.1371/journal.pone.0196588

Papageorgiou L, Loukatou S, Koumandou VL, Makałowski W, Megalooikonomou V *et al.* (2014) Structural models for the design of novel antiviral agents against Greek Goat Encephalitis. *PeerJ* **2**, e664. http://dx.doi.org/10.7717/peerj.664

Parkin DM, Bray F, Ferlay J and Pisani P (2005) Global cancer statistics, 2002. *CA Cancer J Clin* **55**(2), 74-108. http://dx.doi.org/10.3322/canjclin.55.2.74

Platis M, Vlachakis D, Foudah AI, Muharram MM, Alqarni MH *et al.* (2021) The Interaction of Schistosoma Japonicum Glutathione Transferase with Cibacron Blue 3GA and its Fragments. *Med Chem* **17**(4), 332-343. http://dx.doi.org/10.2174/1573406416666200403074742

Polk DB and Peek RM (2010) Helicobacter pylori: gastric cancer and beyond. *Nature Reviews Cancer* **10**(6), 403-414. http://dx.doi.org/10.1038/nrc2857

Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M *et al.* (2012) A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. *PLoS neglected tropical diseases* **6**(1), e1455-e1455. http://dx.doi.org/10.1371/journal.pntd.0001455

Shima H, Takano M, Shimotohno K and Miwa M (1986) Identification of p26Xb and p24Xb of human T-cell leukemia virus type II. *FEBS Lett* **209**(2), 289-294. http://dx.doi.org/10.1016/0014-5793(86)81129-2

Smatti MK, Al-Sadeq DW, Ali NH, Pintus G, Abou-Saleh H *et al.* (2018) Epstein-Barr Virus Epidemiology, Serology, and Genetic Variability of LMP-1 Oncogene Among Healthy Population: An Update. *Front Oncol* **8**, 211. http://dx.doi.org/10.3389/fonc.2018.00211

Steinhauf D, Rodriguez A, Vlachakis D, Virgo G, Maksimov V *et al.* (2014) Silencing motifs in the Clr2 protein from fission yeast, Schizosaccharomyces pombe. *PLoS ONE* **9**(1), e86948. http://dx.doi.org/10.1371/journal.pone.0086948

Thorell K, Lehours P and Vale FF (2017) Genomics of Helicobacter pylori. *Helicobacter* **22**(S1), e12409. http://dx.doi.org/https://doi.org/10.1111/hel.12409

Tomasetti C, Li L and Vogelstein B (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**(6331), 1330-1334. http://dx.doi.org/10.1126/science.aaf9011

Trimpalis P, Koumandou VL, Pliakou E, Anagnou NP and Kossida S (2013) Gene fusion analysis in the battle against the African endemic sleeping sickness. *PLoS ONE* **8**(7), e68854. http://dx.doi.org/10.1371/journal.pone.0068854

Tsagrasoulis D, Danos V, Kissa M, Trimpalis P, Koumandou VL *et al.* (2012) SAFE Software and FED Database to Uncover Protein-Protein Interactions using Gene Fusion Analysis. *Evol Bioinform Online* **8**, 47-60. http://dx.doi.org/10.4137/ebo.s8018

Tsaniras S, Vlachakis D and Taraviras S (2015) The Nucleophosmin-Pin1 interaction links the cell cycle, cancer and pluripotency. *J Mol Biochem* **4**(3), 50-51.

Utzinger J, Raso G, Brooker S, De Savigny D, Tanner M *et al.* (2009) Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology* **136**(13), 1859-1874. http://dx.doi.org/10.1017/s0031182009991600

Vineis P, Schatzkin A and Potter JD (2010) Models of carcinogenesis: an overview. *Carcinogenesis* **31**(10), 1703-1709. http://dx.doi.org/10.1093/carcin/bgq087

Vlachakis D, Champeris Tsaniras S, Karozou A and Kossida S (2013a) An update on virology and emerging viral epidemics. *J Mol Biochem* **2**(2), 80-84.

Vlachakis D, Pavlopoulou A, Kazazi D and Kossida S (2013b) Unraveling microalgal molecular interactions using evolutionary and structural bioinformatics. *Gene* **528**(2), 109-119. http://dx.doi.org/10.1016/j.gene.2013.07.039

Xie C, Xu LY, Yang Z, Cao XM, Li W *et al.* (2014) Expression of γH2AX in various gastric pathologies and its association with Helicobacter pylori infection. *Oncol Lett* **7**(1), 159-163. http://dx.doi.org/10.3892/ol.2013.1693

Yu Y-P, Liu P, Nelson J, Hamilton RL, Bhargava R *et al.* (2019) Identification of recurrent fusion genes across multiple cancer types. *Scientific Reports* **9**(1), 1074. http://dx.doi.org/10.1038/s41598-019-38550-6