

Evidence of prostate cancer-linked virus zoonoses from biophysical genomic variations

Daniah Alsufyani^{1,2} and James Lindesay³

¹ College of Sciences and Health Professions, King Saud Bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia

² King Abdullah International Medical Research Center, Jeddah, Saudi Arabia

³ Howard University, Washington, D.C., United States

ABSTRACT

An ongoing double-blind examination of (mathematically) smooth functional dependences of population-based genomic distributions of single nucleotide polymorphisms (SNPs) on quantified environmental parameters has flagged a SNP that has been associated with prostate cancer for dependence on zoonotic viruses. The SNP rs13091518 is an intergenic variant near the gene/pseudo-gene COX6CP6 on chromosome 3. The risk T allele, which is the major allele in all homeostatic populations considered, clearly demonstrates a negative adaptive force of about -0.1 universal genomic energy units/zoonotic virus unit. This biophysical perspective has thus provided evidence for a causative relationship between zoonotic viruses and prostate cancer. Our findings are consistent with other studies that have found an association between several zoonotic viruses and prostate cancer. This result demonstrates the significance of an intergenic variant in the adaptive response to a viral zoonotic pathogen.

Subjects Bioinformatics, Biophysics, Computational Biology, Genetics, Genomics

Keywords Adaptive force, Population health, SNPs, Genodynamics, Zoonotic viruses, Prostate cancer

INTRODUCTION

Since the inception of the HapMap project of population-based human genomes, access to the study of the subtleties and dependencies of the distribution of genomic variants has resulted in considerable insights into the dynamics of the factors that influence genomic variations. Since the data are specific to each population, this allows an examination of how quantifiable environmental influences can affect the populations subject to those influences. Due to migrations, human populations often smoothly adapt to environmental changes along their various migration routes. Some models have developed statistical free energy functions to describe the evolutionary dynamics of finite populations (*Sella & Hirsh, 2005*), where perturbations of a population's free energy evolve toward "fitness peaks" for the variants (*Agozzino et al., 2020*). However, many such tools utilized in bioinformatics lack a *universal* genomic dimensional unit that can quantify the pressures on the genomic variants of a given population. In particular, genome-wide association

Submitted 19 June 2024
Accepted 4 November 2024
Published 6 December 2024

Corresponding author
Daniah Alsufyani,
sufyanid@ksau-hs.edu.sa

Academic editor
Vladimir Uversky

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.18583

© Copyright
2024 Alsufyani and Lindesay

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

studies (GWAS) only search for associations of single nucleotide polymorphisms (SNPs) with a particular disease or trait.

In our approach, genodynamics ([Lindesay et al., 2013](#)) develops a universal dimensional biophysical unit (a genomic energy unit) associated with genomic variations. The existence of this unit then allows adaptive *forces* (i.e., differences in genomic variant distributions between homeostatic populations due to environmental differences) to be ascertained. Genodynamics searches for meaningful relationships between the distributions of population-based genomic variants and quantifiable environmental influences. It seeks to model the information dynamics of bi-allelic SNPs due to those influences. The approach presently being undertaken involves a double-blind examination of all SNP variants on chromosome 3. The establishment of smooth functional dependencies of stable population distributions upon any impinging influence implies that all adaptations optimize the overall health of each population. This approach is thus absent any focus on disease. In this article, the dynamics of a particular common genomic variant between homeostatic populations that *presently* reside in their ancestral geographic homelands as a function of zoonotic viruses will be presented. Population homeostasis implies that several generations of the population have maintained the same distribution of genomic variants, which is analogous to equilibrium in thermodynamics. Subsequent to flagging rs13091518 with richness of zoonotic viruses, a literature search summarized in the discussion section revealed several biological studies indicating a direct link of the risk allele to various viruses associated with prostate cancer.

METHODS

The dispersion of the human species throughout a wide variety of stable environments has generated a large diversity in both phenotypes and genotypes. The resulting genetic stability often exhibits Hardy-Weinberg equilibrium within each population, where the frequencies of allelic variants are maintained throughout several generations or other arbitrary subgroups ([Hardy, 1908](#)). These observations motivated the development of a biophysical approach that formulates the dynamics of the information codified within the populations' genomes, enabling the investigation of meaningful relationships between environmental stimuli and genomic SNPs. This approach is referred to as "genodynamics". For the present study (on-going for 4 years), we have chosen 10 populations genotyped within their ancestral geographic regions to search for any dependencies of maintained genomic order on a set of 15 quantifiable ancestral environmental parameters. Obtaining quantified environmental parameters remains one of the most challenging aspects of this study.

Genodynamics modeling approach

Geodynamics utilizes a well-established measure of the degree of maintained order within a dynamic system defined as the information content (IC) ([Shannon & Weaver, 1963](#); [Bergstrom & Lachmann, 2004](#); [Suskind & Lindesay, 2004](#)). A more portable measure of maintained order is the normalized information content (NIC). The NIC can be expressed

in terms of the entropy of the distribution $s = -\sum_a p_a \log_2 p_a$, which itself is defined in terms of the probabilities (frequencies) of variants a , here labeled p_a . For the variants in a population's genome, the NIC_{genome} is defined as

$$NIC_{genome} = \frac{s_{max} - s_{genome}}{s_{max}},$$

where s_{genome} is the entropy of the whole genome, and s_{max} is the entropy of maximal variation. The entropy of the complete genome of a population sums the maintained entropies of each SNP variant on all chromosomes. The maximum possible entropy of a human genome assigns maximum variation for each of those variants. When utilizing bi-allelic SNPs as the genomic variants, the maximum entropy is given by $s_{max} = N_{SNPs}$, where N_{SNPs} is the total number of (biallelic) SNPs on the genome. From the NIC, a degree of intrinsic disorder analogous to the temperature in thermodynamics can be derived by minimizing the overall genomic free energy (Alsufyani & Lindsay, 2022). This

dimensional environmental potential is defined by $T_E = \frac{\check{\mu}}{NIC_{genome}}$, where the unit $\check{\mu}$ is the maximum SNP potential of a non-linked bi-allelic SNP and is assigned a value of 1 (human) genomic energy unit (GEU). From this relation, it is clear that populations with a greater degree of disorder have a higher T_E .

Bi-allelic SNPs that are not in linkage disequilibrium have additive genomic potentials related to allelic frequencies $p_a^{(S)}$ as defined *via* the expression

$$\mu_a^{(S)} = \check{\mu} - T_E - T_E \log_2 p_a^{(S)}.$$

If only a single allele is shared amongst the whole population (*i.e.*, $p_a^{(S)} = 1$, $p_a^{(S)} = 0$), then SNP potential is fixed at the *fixing potential* $\check{\mu} - T_E$. As previously mentioned, these potentials take the value of 1 GEU for bi-allelic SNPs of maximum variation (Alsufyani & Lindsay, 2022). For a particular SNP (S), the SNP potential $\mu^{(S)}$ is just the population averaged value of its allelic potentials $\mu^{(S)} = \sum_a p_a^{(S)} \mu_a^{(S)}$. For the populations examined in what follows, the environmental potentials T_E are given in Table 1. The characterization of the environmental potential T_E requires evaluation of the whole genome information content of each population under consideration. A previous exploration of the (somewhat smaller) whole genome information content of HapMap populations convinced the authors that the information contents of chromosome 3 accurately represents that of the whole genomes within a percent (Dunston *et al.*, 2014).

Allelic variants that are in linkage disequilibrium maintain a collective order unique to a given population that decreases the overall entropy of those variants. The collection of such SNPs form haploblocks with the various haplotypes defining the variants within the haploblocks. In the same manner that each allelic and SNP potential can be defined for an individual genomic locus, the haplotype potentials and haploblock potential $\mu^{(H)}$ (which is the population average of the haplotype potentials) can directly be defined for a linked set of genomic loci (Alsufyani, 2024). One should note that the haploblock structure of any population is unique. However, most of the GWAS data that will be utilized to ascertain biological associations focus on individual SNPs. For this reason, we have developed a

Table 1 Environmental potentials of the examined populations.

Population	T_E
PEL	1.108
CLM	1.125
FIN	1.125
KHV	1.106
JPT	1.108
TSI	1.122
MSL	1.237
CHB	1.108
IBS	1.110
LWK	1.225

mechanism for meaningfully distributing the haploblock potentials amongst the SNPs (Lindesay et al., 2018):

- if a specific allele is fixed for *all* members of a population, its distributed SNP potential is defined to be the fixing potential $\mu_S^{(H)} = \check{\mu} - T_E$;
- the sum of the distributed SNP potentials within a haploblock is the same as the haploblock potential $\langle \mu^{(H)} \rangle = \sum_S \mu_S^{(H)}$;
- the overall haploblock potential is linearly distributed amongst its constituent SNPs proportionate with the occurrence of the alleles consistent with the previous requirements according to $\mu_S^{(H)} = \mu_{fixed} + \left[\langle \mu^{(H)} \rangle - n^{(H)} \mu_{fixed} \right] \left[\frac{\bar{P}_{s'}}{\sum_{s'} \bar{P}_{s'}} \right]$.

The increased degree of maintained order due to the linkage disequilibrium thus defines a ‘binding potential’ $\varepsilon_{binding}^{(S)} = \mu_S^{(H)} - \langle \mu^{(S)} \rangle$ relative to the averaged SNP potential assigned if the SNPs were not linked. The distributed SNP potentials can finally be used to define distributed allelic potentials $\mu_{a_s}^{(H)} = \mu_{a_s}^{(S)} + \varepsilon_{binding}^{(S)}$ which share the lowered genomic potential of the SNP.

We search for allelic potentials whose values vary smoothly between the populations in such a manner that simple mathematical forms assign those potentials as functions of quantified environmental parameters. Physical systems tend to slide down the slopes of potential energy, defining a force as the slope of the potential energy curve. In analogy, the adaptive force (or adaptive “pressure”) is defined as the slope down the gradient of an allelic potential curve as a function of the environmental parameter $f_a = -\frac{\partial \mu_a}{\partial \lambda}$.

Genomic analysis in environmental context

The populations examined in this study were genotyped within their ancestral geographic environments. These include Peruvian in Lima, Peru (PEL), Colombian in Medellín, Colombia (CLM), Finnish in Finland (FIN), Kinh in Ho Chi Minh City, Vietnam (KHV), Japanese in Tokyo, Japan (JPT), Toscani in Italy (TSI), Mende in Sierra Leone (MSL), Han

Chinese in Beijing (CHB), Iberian population in Spain (IBS), and Luhya in Webuye, Kenya (LWK). The populations were double-blind scanned for allelic dependencies on the 15 quantified environmental parameters including, altitude, humidity, pressure, temperature, rain, UVB, wind speed, and pathogens (virus, bacteria, helminth, protozoa) associated with zoonotic hosts (chiroptera, primates, rodentia, soricomorpha). Environmental data was population-averaged using a variety of cities distributed throughout the ancestral regions. Maps and non-tabular data were independently quantified by the authors, and averaged results are ultimately utilized in the searches. To this point, 95 thousand SNPs have been scanned, and only about 0.007% have flagged a meaningful dependency so far.

Criteria for flagging allelic potential dependencies

The formulation requires a smooth mathematical dependency on a specific environmental parameter for a result to be meaningful. For our purposes, only relatively simple dependencies of the allelic potentials on environmental parameters are being considered. During a given search procedure, the root-mean-squared (RMS) deviation of the distribution of the allelic potentials of the populations from optimized fitted curves is calculated. A smooth dependency of an allelic potential on an environmental parameter will only be flagged if the dimensionless ratio of the RMS deviation to the maximum change in that allelic potential is within 10%.

RESULTS

Allelic potentials that quantify the pressures induced on genomic variants due to environmental influences continue to provide a useful tool for discovering the natural response to stimuli and pathogens that affect the given population. For the present study, only simple functional forms monotonic in allelic potentials (but not necessarily in the SNP potential) have been considered (Alsufyani & Lindesay, 2022, 2023). The SNP rs13091518 only flagged the populations for the pathogen ‘zoonotic virus richness’, where ‘richness’ is defined in the reference as follows (Alsufyani & Lindesay, 2023):

Richness: the number of unique species within a particular geographic area; richness is a count-based metric for quantifying diversity, which contrasts with other metrics, such as functional trait diversity (the different types of traits represented within a geographic area) or genetic diversity.

It is noteworthy that for rs13091518, the potentials of both alleles as well as the SNP itself flagged dependencies upon the zoonotic viruses. The type of flag of the SNP emphasizes a *direct* dependence of the SNP *potential* due to the pathogen. The *risk* allele T exhibits a *negative* adaptive force of about -0.1 GEUs/zoonotic virus unit due to the pathogen, consistent with an increasing risk to exposed populations.

Figure 1 demonstrates the SNP potential of rs13091518 vs. exposure to zoonotic viruses, which notably flagged with relative RMS deviation of less than 0.03. The SNP potential reflects the favorability of maintaining an overall degree of variation of *both* alleles within the population. In the figure, blue points represent SNPs that are not in linkage disequilibrium and red points represent SNPs in linkage disequilibrium. The figure clearly demonstrates that *any* presence of this pathogen breaks the healthy maintained *correlated*

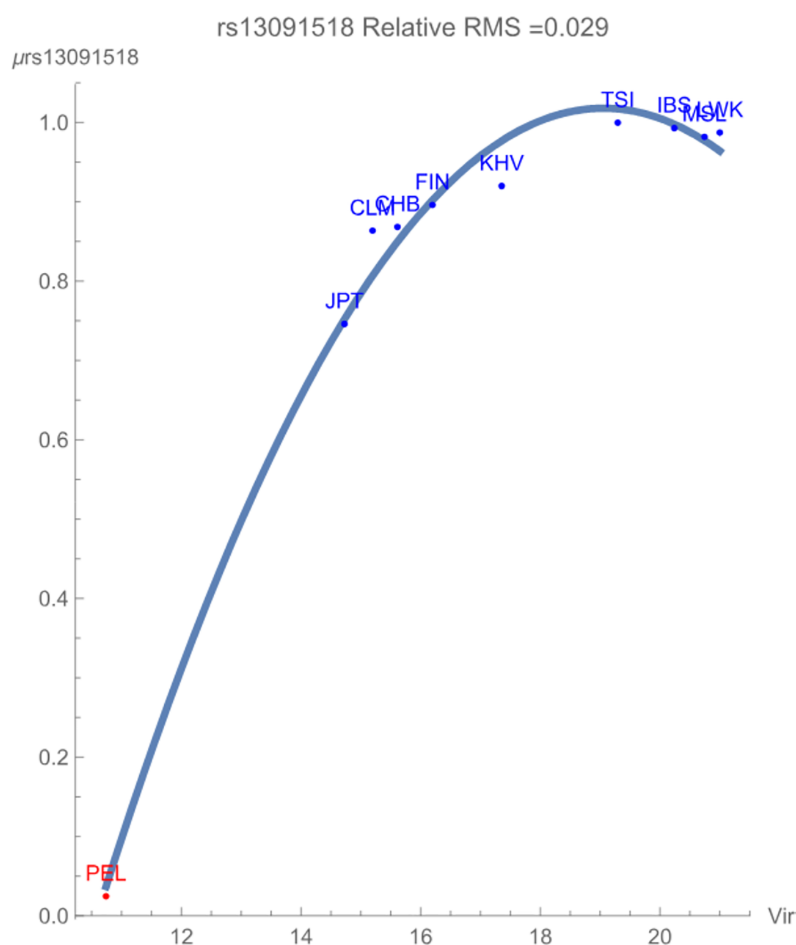


Figure 1 The correlation between the rs13091518 SNP potential values (in GEUs) and the richness (number of species) of zoonotic virus pathogens. [Full-size !\[\]\(1679558f37f6db0dd8360a2a7e913e90_img.jpg\) DOI: 10.7717/peerj.18583/fig-1](https://doi.org/10.7717/peerj.18583/fig-1)

genomic variation of the SNP haploblock within the population PEL. The increased conservation of the block of SNPs is reflected in its considerably lower SNP potential when compared to the more exposed populations. There is some indication that the non-risk allele C begins to have an overall more favorable contribution to population health for the highly exposed populations due to the downturn in the potential curve.

Figure 2 illustrates the allelic potential of the T allele, which is the risk allele for prostate cancer. This allele remains the *major* allele in all the populations chosen for consideration. This risk is clearly demonstrated through the negative adaptive force of -0.1 GEUs/zoonotic virus unit, with a relative RMS deviation of 0.055, well within our flagging criterion. The T allele is highly conserved with an allelic potential of -0.37 GEUs for the population PEL, which has the lowest exposure to zoonotic viruses. This is in contrast to the allelic potential of nearly 1 GEU for the population TSI, indicative of highest allelic variation.

Finally, Fig. 3 displays the allelic potential of the C allele, which (barely) remains the minor allele in all populations considered. It is clear from the plots that an increased

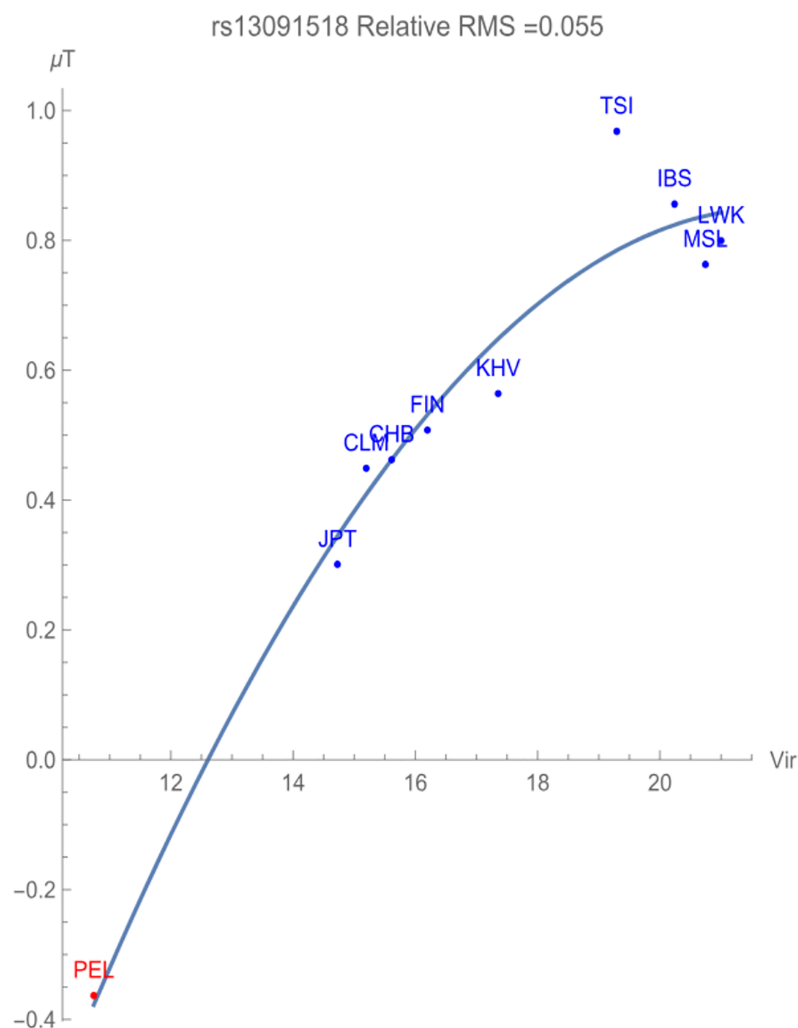


Figure 2 The correlation between T allelic potential (in GEUs) of rs13091518 and zoonotic viruses host richness. The adaptive force is about -0.1 GEUs/zoonotic virus unit.

Full-size DOI: [10.7717/peerj.18583/fig-2](https://doi.org/10.7717/peerj.18583/fig-2)

presence of viral zoonoses induces higher variation in the SNP, which is provided by the C allele. This mitigation of risk translates into an adaptive force of $+0.16$ GEUs/zoonotic virus unit, with a relative RMS deviation of 0.06. The type of flag emphasizes a direct dependence of the *frequencies* of alleles in the population as opposed to the allelic potential itself. This indicates that the increased conservation of the C allele acts to ‘switch on’ an enhancement to population health.

DISCUSSION

The SNP rs13091518 is an intergenic variant on chromosome 3 at location 70747545 (Conti *et al.*, 2021). Genome wide association studies have relatively recently (2021) associated the T allele of this variant (the *major* allele in all of the examined populations) with an increased risk of prostate cancer (Sollis *et al.*, 2023). The nearest *gene/pseudogene*, cytochrome c oxidase subunit 6C pseudogene 6 (COX6CP6), is located between 70751058

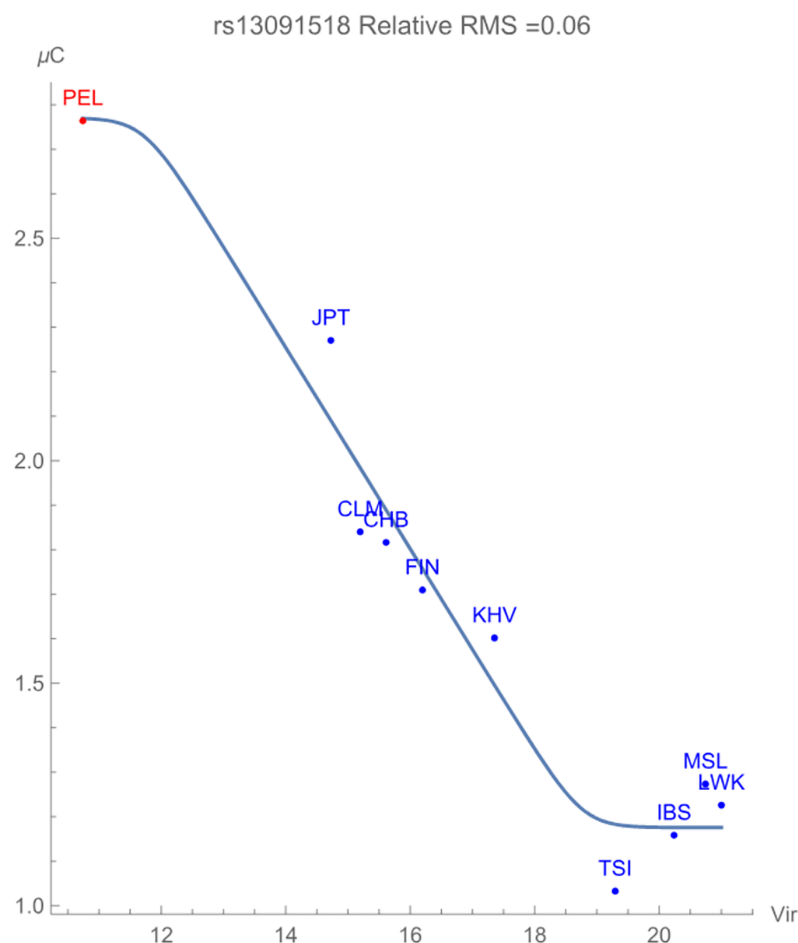


Figure 3 The correlation between C allelic potential (in GEUs) of rs13091518 and zoonotic virus richness. The adaptive force is about +0.16 GEUs/zoonotic virus unit.

Full-size  DOI: [10.7717/peerj.18583/fig-3](https://doi.org/10.7717/peerj.18583/fig-3)

and 70751523. Some pseudogenes have been suggested to play a role in regulating protein-coding transcripts (Chan & Chang, 2014). More recent developments in sequencing technology have discovered a large number of pseudogenes that can interact with DNA, RNA, and proteins that modulate target gene expression. This indicates their potentially strong involvement in the development and progression of certain diseases, especially cancer (Xu et al., 2020).

It is noteworthy in our results that the *only* population for which this SNP is in linkage disequilibrium is that population with the *lowest* viral load. This loss of coordinated genomic variance with increasing viral zoonoses amplifies the increased virally induced disorder amongst the populations. This is consistent with the observed *negative* adaptive force associated with the risk allele T. Evidently, this risk major allele maintains a beneficial biological function moderated through its viral risk association. Previous examinations of the MHC region on chromosome 6 (associated with adaptive immunity) have demonstrated that regulatory regions display a lower degree of maintained order as

quantified by the NIC when compared to genic regions within a given population (Lindesay et al., 2014).

Other studies support our hypothesis, suggesting a link between increased risk of prostate cancer and exposure to zoonotic viruses. A study in 2007 about prostate cancer pathogenesis involves both heritable and environmental factors. Environmental factors have been implicated in studies involving immigrant Asians living in Western countries as compared to their counterparts living in Asia (De Marzo et al., 2007). In a review article (Maia & Hansen, 2017), the authors have argued that viruses have been shown to play a significant role in the etiology of prostate cancer, as well as in promoting a pro-inflammatory microenvironment. Such microenvironments have been shown to be immunosuppressive and to support cancer progression.

Numerous studies have established a correlation between specific zoonotic viruses and an increased risk of developing prostate cancer, which will be elaborated upon in the following sections. Both human and non-human polyomaviruses consist of several types of non-enveloped DNA that encode oncogenes associated with human disease. The BK virus, which is a polyomavirus, has been found as an etiological cofactor of early-stage prostate cancer progression (Das, Wojno & Imperiale, 2008). This virus has been suggested to infect the urinary tract, encoding tumor antigens that inactivate the tumor suppressors *p53* and *Rb1*. However, this virus is not detected in those cells exhibiting more advanced cancer progression. Another study examining the JC (John Cunningham) polyomavirus (JCV) as a risk factor in prostate cancer detected the presence of the JCV Large T antigen (L Tag) in cases vs. controls to assess its risk (Gorish, 2022). The study found a significant relationship between JCV infection and the probability of developing prostate cancer, with the JCV LTag in the cases group higher than that in the controls with a *p*-value of 0.006. In addition, viral load was also significantly higher with a *p*-value of 0.002.

Polyomaviruses encode oncogenes and are associated with cancer (Scuda et al., 2013). They are known to cause infections in patients that have compromised immune responses. Recent investigations suggest that certain polyomaviruses from non-human primates circulate among humans (Kuchipudi et al., 2021). Furthermore, the possibility of zoonosis (Scuda et al., 2013) from bovine polyomaviruses due to close human encounters is suspected. Human JCV has recently been found to cause multifocal leukoencephalopathy (Dugan, Gasparovic & Atwood, 2008). In humans, sera reactivity against BKV and JCV has been found to be cross reactive with simian virus 40 (or old-world monkey virus) (Scuda et al., 2013; Kean et al., 2009). BKV has been found to induce tumors in newborn hamsters and mice (Roka, Prieto & Gershenwald, 2008). Furthermore, some of the evidence suggesting that BKV is oncogenic has resulted from studies involving tumorigenesis in rodents inoculated with the virus (Ambalathingal et al., 2017). However, due to the cell specificity of BKV and narrow range of hosts, the development of animal models to study this infection is difficult (Borriello et al., 2022).

Kaposi's sarcoma-associated herpes virus (KSHV), formerly known as humangammaherpesvirus 8 (HHV-8), is one of several known rhadinoviruses. KSHV has been found to be associated with prostate cancer amongst those who carry the IFNL4 gene.

IFNL4 is a polymorphic pseudogene on chromosome 19, sometimes encoding the interferon lambda 4 protein. The HHV-8 virus has been associated with both aggressive and non-aggressive prostate cancer amongst those carrying at least one G allele in IFNL4 (Jenkins *et al.*, 2023). Herpesviruses have been shown to cause several diseases in both humans and animals (Woźniakowski & Samorek-Salamonowicz, 2015). Of the human herpesviruses, KSHV is the most closely related to zoonotic rhadinoviruses (Dollery, 2019; McGeoch, Rixon & Davison, 2006). It is also the only rhadinovirus known to infect humans (Chang *et al.*, 1994). Several studies indicate that KSHV can infect most human cell types, and several animal cell types (Pica & Volpi, 2007; Schäfer, Blumenthal & Katz, 2015). The range of hosts of herpesviruses is narrow. However, due to specifics in the kind of encoded nucleic acid, it is possible that herpesviruses can be transmitted to other species. Even if the presence of a herpesvirus in its natural host is asymptomatic or only manifests mild symptoms, its transmission to another species might lead to severe disease or death (Woźniakowski & Samorek-Salamonowicz, 2015).

Finally, it should be noted that the Zika virus has been found to actually decrease the proliferation of a prostate cancer cell line (Delaflori *et al.*, 2019). This indicates the possibility that environmental exposure to this virus might be associated with a positive adaptive response, reflected by the enhanced conservation of the C allele.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Daniah Alsufyani conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- James Lindesay conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The codes for calculating SNPs and alleles potentials and to find the associations between environmental parameters and potentials are available in the [Supplemental Files](#).

The SNP rs13091518 frequencies can be generated at the Allele Frequency Calculator: https://grch37.ensembl.org/Homo_sapiens/Tools/AlleleFrequency?db=core.

All the maps utilized for the virus richness data are available at Global Patterns of Zoonotic Disease in Mammals, <https://doi.org/10.1016/j.pt.2016.04.007>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.18583#supplemental-information>.

REFERENCES

- Agozzino L, Balázs G, Wang J, Dill KA. 2020. How do cells adapt? Stories told in landscapes. *Annual Review of Chemical and Biomolecular Engineering* **11**(1):155–182.
- Alsufyani D. 2024. Evidence of zoonotic pathogens through biophysically induced genomic variance. *Quarterly Reviews of Biophysics* **57**:e2 DOI [10.1017/S0033583524000039](https://doi.org/10.1017/S0033583524000039).
- Alsufyani D, Lindesay J. 2022. Quantification of adaptive forces on SNP rs1010211 due to viral zoonotic pathogens. *Journal of Biological Physics* **48**(2):227–236 DOI [10.1007/s10867-022-09606-y](https://doi.org/10.1007/s10867-022-09606-y).
- Alsufyani D, Lindesay J. 2023. Evidence of cancer-linked rodent zoonoses from biophysical genomic variations. *Scientific Reports* **13**(1):9541 DOI [10.1038/s41598-023-41257-4](https://doi.org/10.1038/s41598-023-41257-4).
- Ambalathingal GR, Francis RS, Smyth MJ, Smith C, Khanna R. 2017. BK polyomavirus: clinical aspects, immune regulation, and emerging therapies. *Clinical Microbiology Reviews. American Society for Microbiology* **30**:503–528 DOI [10.1128/CMR.00074-16](https://doi.org/10.1128/CMR.00074-16).
- Bergstrom CT, Lachmann M. 2004. Shannon information and biological fitness. In: *Information Theory Workshop*. IEEE, 50–54.
- Borriello M, Ingrosso D, Perna AF, Lombardi A, Maggi P, Altucci L, Maggi P, Altucci L, Caraglia M. 2022. BK virus infection and BK-virus-associated nephropathy in renal transplant recipients. *Genes* **13**:1290 MDPI DOI [10.3390/genes13071290](https://doi.org/10.3390/genes13071290).
- Chan WL, Chang JG. 2014. Pseudogene-derived endogenous siRNAs and their function. *Methods in Molecular Biology* **1167**:227–239 DOI [10.1007/978-1-4939-0835-6](https://doi.org/10.1007/978-1-4939-0835-6).
- Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated kaposi's sarcoma. *Science* **266**(5192):1865–1869 DOI [10.1126/science.7997879](https://doi.org/10.1126/science.7997879).
- Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, Schumacher FR, Olama A, Benlloch S, Dadaev T, Brook MN, Sahimi A, Hoffmann TJ, Takahashi A, Matsuda K, Momozawa Y, Fujita M, Muir K, Lophatananon A, Wan P, Le Marchand L, Wilkens LR, Stevens VL, Gapstur SM, Carter BD, Schleutker J, Tammela TLJ, Sipeky C, Auvinen A, Giles GG, Southey MC, MacInnis RJ, Cybulski C, Wokołorczyk D, Lubiński J, Neal DE, Donovan JL, Hamdy FC, Martin RM, Nordestgaard BG, Nielsen SF, Weischer M, Bojesen SE, Røder MA, Iversen P, Batra J, Chambers S, Moya L, Horvath L, Clements JA, Tilley W, Risbridger GP, Gronberg H, Aly M, Szulkin R, Eklund M, Nordström T, Pashayan N, Dunning AM, Ghoussaini M, Travis RC, Key TJ, Riboli E, Park JY, Sellers TA, Lin H-Y, Albanes D, Weinstein SJ, Mucci LA, Giovannucci E, Lindstrom S, Kraft P, Hunter DJ, Penney KL, Turman C, Tangen CM, Goodman PJ, Thompson IM, Hamilton RJ, Fleshner NE, Finelli A, Parent ME, Stanford JL, Ostrander EA, Geybels MS, Koutros S, Beane Freeman FM, Wolk A, Håkansson N, Andriole GL, Hoover RN, Machiela MJ, Sørensen KD, Borre M, Blot WJ, Zheng W, Yeboah ED, Mensah JE, Lu YJ, Zhang H-W, Feng N, Mao X, Wu Y, Zhao SC, Sun Z, Thibodeau SN, McDonnell SK, Schaid DJ, West CML, Burnet N, Barnett G, Maier C, Schnoeller T, Luedeke M, Kibel AS, Drake BF, Cussenot O, Cancel-Tassin G, Menegaux F, Truong T, Koudou YA, John EM, Grindedal EM, Maehle L, Khaw K-T, Ingles SA, Stern MC, Vega A, Gómez-Caamaño A, Fachal L,

- Rosenstein BS, Kerns SL, Ostrer H, Teixeira MR, Paulo P, Brandão A, Watya S, Lubwama A, Bensen JT, Fonham ETH, Mohler J, Taylor JA, Kogevinas M, Llorca J, Castaño-Vinyals G, Cannon-Albright L, Teerlink CC, Huff CD, Strom SS, Multigner L, Blanchet P, Brureau L, Kaneva R, Slavov C, Mitev V, Leach RJ, Weaver B, Brenner H, Cuk K, Holleczech B, Saum KU, Klein EA, Hsing AW, Kittles RA, Murphy AB, Logothetis CJ, Kim J, Neuhausen SL, Steele L, Ding YC, Isaacs WB, Nemesure B, Hennis AJM, Carpten J, Pandha H, Michael A, De Ruyck K, De Meerleer G, Ost P, Xu J, Razack A, Lim J, Teo SH, Newcomb LF, Lin DW, Fowke JH, Neslund-Dudas C, Rybicki BA, Gamulin M, Lessel D, Kulis T, Usmani N, Singhal S, Parliament M, Claessens F, Joniau S, Van den Broeck T, Gago-Dominguez M, Castela JE, Martinez ME, et al. 2021. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature Genetics* 53:65–75 DOI 10.1038/s41588-020-00748-0.
- Das D, Wojno K, Imperiale MJ. 2008. BK virus as a cofactor in the etiology of prostate cancer in its early stages. *Journal of Virology* 82(6):2705–2714 DOI 10.1128/JVI.02461-07.
- De Marzo AM, Platz EA, Sutcliffe S, Xu J, Grönberg H, Drake CG, Nakai Y, Isaacs WB, Nelson WG. 2007. Inflammation in prostate carcinogenesis. *Nature Reviews Cancer* 7(4):256–269 DOI 10.1038/nrc2090.
- Delafiori J, de Lima EO, Dabaja MZ, Dias-Audibert FL, de Oliveira DN, Melo CFOR, Morishita KN, Sales GM, Ruiz ALTG, da Silva GG, Lancellotti M, Catharino RR. 2019. Molecular signatures associated with prostate cancer cell line (PC-3) exposure to inactivated Zika virus. *Scientific Reports* 9(1):587 DOI 10.1038/s41598-019-51954-8.
- Dollery SJ. 2019. Towards understanding KSHV fusion and entry. *Viruses* 11:1073 MDPI AG DOI 10.3390/v11111073.
- Dugan AS, Gasparovic ML, Atwood WJ. 2008. Direct correlation between sialic acid binding and infection of cells by two human polyomaviruses (JC Virus and BK Virus). *Journal of Virology* 82(5):2560–2564 DOI 10.1128/JVI.02123-07.
- Dunston GM, Mason TE, Hercules W, Lindesay J. 2014. Single nucleotide polymorphisms: a window into the informatics of the living genome. *Advances in Bioscience and Biotechnology* 5(7):623–626 DOI 10.4236/abb.2014.57073.
- Gorish BMT. 2022. JC polyoma virus as a possible risk factor for prostate cancer development—immunofluorescence and molecular based case control study. *Cancer Control: Journal of The Moffitt Cancer Center* 29(3):107327482211407 DOI 10.1177/10732748221140785.
- Hardy GH. 1908. Mendelian proportions in a mixed population. *New Series* 28:49–50 DOI 10.1126/science.28.706.49.
- Jenkins FJ, Minas TZ, Tang W, Dorsey TH, Ambs S. 2023. Human herpesvirus 8 infection is associated with prostate cancer among IFNL4-ΔG carriers. *Prostate Cancer and Prostatic Diseases* 26(2):338–346 DOI 10.1038/s41391-022-00546-1.
- Kean JM, Rao S, Wang M, Garcea RL. 2009. Seroepidemiology of human polyomaviruses. *PLOS Pathogens* 5(3):e1000363 DOI 10.1371/journal.ppat.1000363.
- Kuchipudi SV, Nelli RK, Gontu A, Satyakumar R, Nair MS, Subbiah M. 2021. Sialic acid receptors: the key to solving the enigma of zoonotic virus spillover. *Viruses* 13:262 MDPI AG DOI 10.3390/v13020262.
- Lindesay J, Mason TE, Hercules W, Dunston GM. 2014. Development of genodynamic metrics for exploring the biophysics of DNA polymorphisms. *Journal of Computational Biology and Bioinformatics Research* 6(1).

- Lindesay J, Mason TE, Hercules W, Dunston GM. 2013. The foundations of genodynamics: the development of metrics for genomic-environmental interactions. ArXiv preprint DOI 10.48550/arXiv.1312.3260.
- Lindesay J, Mason TE, Hercules W, Dunston GM. 2018. Mathematical modeling the biology of single nucleotide polymorphisms (SNPs) in whole genome adaptation. *Advances in Bioscience and Biotechnology* 9(10):520–533 DOI 10.4236/abb.2018.910036.
- Maia MC, Hansen AR. 2017. A comprehensive review of immunotherapies in prostate cancer. *Critical Reviews in Oncology/Hematology* 113(1):292–303 DOI 10.1016/j.critrevonc.2017.02.026.
- McGeoch DJ, Rixon FJ, Davison AJ. 2006. Topics in herpesvirus genomics and evolution. *Virus Research* 117(1):90–104 DOI 10.1016/j.virusres.2006.01.002.
- Pica F, Volpi A. 2007. Transmission of human herpesvirus 8: an update. *Current Opinion in Infectious Diseases* 20(2):152–156 DOI 10.1097/QCO.0b013e3280143919.
- Roka F, Prieto VG, Gershenwald JE. 2008. Cutaneous desmoplastic melanoma. In: *Encyclopedia of Cancer*. Berlin Heidelberg: Springer, 785–786.
- Schäfer G, Blumenthal MJ, Katz AA. 2015. Interaction of human tumor viruses with host cell surface receptors and cell entry. *Viruses* 7:2592–2617 MDPI AG DOI 10.3390/v7052592.
- Scuda N, Madinda NF, Akoua-Koffi C, Adjogoua EV, Wevers D, Hofmann J, Cameron KN, Leendertz SAJ, Couacy-Hymann E, Robbins M, Boesch C, Jarvis MA, Moens U, Mugisha L, Calvignac-Spencer S, Leendertz FH, Ehlers B, Wang D. 2013. Novel polyomaviruses of nonhuman primates: genetic and serological predictors for the existence of multiple unknown polyomaviruses within the human population. *PLOS Pathogens* 9(6):e1003429 DOI 10.1371/journal.ppat.1003429.
- Sella G, Hirsh AE. 2005. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences* 102(27):9541–9546.
- Shannon CE, Weaver W. 1963. The mathematical theory of communication. Champaign: University of Illinois Press.
- Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J, Ibrahim A, Ji Y, John S, Lewis E, MacArthur JAL, McMahon A, Osumi-Sutherland D, Panoutsopoulou K, Pendlington Z, Ramachandran S, Stefancsik R, Stewart J, Whetzel P, Wilson R, Hindorff L, Cunningham F, Lambert S A, Inouye M, Parkinson H, Harris LW. 2023. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Research* 51(D1):D977–D985 DOI 10.1093/nar/gkac1010.
- Susskind L, Lindesay J. 2004. *Introduction to black holes, information and the string theory revolution, an: the holographic universe*. Singapore: World Scientific.
- Woźniakowski G, Samorek-Salamonowicz E. 2015. Animal herpesviruses and their zoonotic potential for cross-species infection. *Annals of Agricultural and Environmental Medicine. Institute of Agricultural Medicine* 22(2):191–194 DOI 10.5604/12321966.1152063.
- Xu Z, Liu C, Zhao Q, Lü J, Ding X, Luo A, He J, Wang G, Li Y, Cai Z, Wang Z, Liu J, Liu S, Li W, Yu Z. 2020. Long non-coding RNA CCAT2 promotes oncogenesis in triple-negative breast cancer by regulating stemness of cancer cells. *Pharmacological Research* 152(5):104628 DOI 10.1016/j.phrs.2020.104628.