Past and present: How has infection shaped the human genome?

In a society experiencing the true capacity of disease, the importance of studying infection has never been more apparent. As humans, we demonise infectious disease, though it is a biological inevitability that has perpetually shaped the evolution of our species. In this regard, I believe the modern cultural impact of infectious disease to be overruled by its genetic impact. To assess this impact of infection on the human genome, I shall discuss two leading factors: the natural selection of disease-resistant genes, and the integration of human endogenous retroviruses (HERVs).

Natural selection is the process by which genetic variants that benefit survival and reproduction are inherited by offspring, while adverse characteristics are selectively removed by nature's tendency to evolve. Throughout history, natural selection has shaped the human genome in response to environmental changes. Infectious pathogens are considered amongst the most defined selective forces for natural selection, increasing survival likelihood in those with disease-defensive genes, and the subsequent frequency of these favoured genes in the succeeding generation.[1]

From a historical perspective, the development of agriculture is a leading event by which infectious disease led natural selection in shaping the human genome. The agricultural revolution in 10,000 B.C involved the clearing of land for cultivation, increased sedentism, and domestication of animals.[2] The environmental shift provided ideal conditions for pathogen survival, and an accelerated spread of disease. In response, the human genome adapted through the natural selection of beneficial genes. For example, a mutation of the protein *connexin-26* increased in frequency – a mutation that thickened the skin, thus improving protection from skin infections. Furthermore, selection occurred on genes involved in Toll-like receptors (TLRs) that aid the immune system in recognising pathogens; another mechanism of defence.[3]

These examples of natural selection, alongside many others, have been proven through various methods of genome comparison. For instance, single nucleotide polymorphism (SNP) genotyping measures genetic variation by detecting the substitution of a single nucleotide.[4] As SNPs are conserved during evolution, this method is effectual in the direct comparison of human genomes over time. However, as argued by E. Karlsson, SNP genotyping has several disadvantages: the detection of non-selective deviations (irrelevant to study), and the inability to compare multiple population units in a single measure.[1] A more successful manner of research is large-scale genome-wide association studies. This approach targets multiple genomes simultaneously, scanning with markers that identify variation associated with particular infectious diseases. More recent examples of genome-wide association studies include that of type 2 diabetes, heart disorders and malaria. Research of the latter has determined malaria to be another model of how infection can alter the human genome through natural selection.

Considered as potentially *"the most important infectious disease on the planet"*[6], malaria is an endemic disease with significant geographical distribution; areas of risk cover 27% of the Earth's land surface[7]. Malaria is protozoan and relies on transmission via species of *Anopheles* mosquitos. In this sense, malaria's complexity is fascinating; the disease adjusts its biochemistry to suit the environment of both host and vector. Infection with malaria causes anaemia; malarian sporozoites destroy red blood cells during replication. Other consequences include fever and jaundice.

In areas most at risk to malaria, natural selection has favoured genetic mutations that provide disease-resistance. For example, the presence of one sickle cell allele copy (HbS allele) alongside one "normal" haemoglobin allele (HBB allele). Sickle red blood cells essentially act as a hostile environment for malaria sporozoites: they leak oxygen from their porous, abnormally shaped membranes, reducing the sufficient release of energy required for the malaria parasite to function.[8] Similarly, natural selection has increased the frequency of glucose-6-phosphate dehydrogenase (G6PD) deficiencies. Functioning G6PD protects cells from oxidative damage; deficiencies can incite haemolysis.[6] As of the sickle cell allele mutation, malaria species cannot complete their reproductive cycle in misfunctioning red blood cells. Therefore, the deficiency provides resistance. In all, natural selection has heightened the frequency of malaria-resistant alleles, and thus the structure of the human genome has been altered.

Alike to malaria, human immunodeficiency virus (HIV) acts as ideal evidence for the selective shaping of the genome as caused by infection. For HIV to infect a host cell, the cell must possess receptors that are compatible to a glycoprotein on the virus's surface, known as gp120. These receptors are labelled CD4 (primary receptor) and CCR5 (co-receptor). The "CCR5-Δ32" mutation, as discovered in 1996, lacks the co-receptor, henceforth resisting infection.[6] This genetic deviation rose in frequency due to, as described by J. Brinkworth, a *"selective sweep."*[9] Although, M. Wayne scrutinises the contribution of natural selection in CCR5-Δ32 mutation frequency, depicting it as a *"happy accident"* of *"genetic drift."*[6]

Genetic drift is a process that, alike to natural selection, incites changes in the human genome. The processes differ in that genetic drift is based on chance; there are no selective ideas involved. Though importantly and somewhat unexpectedly, there is a link between HIV and genetic drift: mutations of HIV experience ten times more genetic drift than the expected factor for its population size, as studied in a replicated population by Voronin in 2009.[10] There is no absolute explanation for the cause of this phenomenon aside from the unpredictable nature of viral replication.

Despite the unknown, it remains undisputable that both genetic drift and natural selection are catalysed by infectious disease; infection gives rise to fluctuations in the frequency of resistant gene expression. As pathogens evolve and new mutations arise,

natural selection and genetic drift will continue to shape the genome for an ongoing future.

In contrast, the process of human endogenous retrovirus integration is one of absolution. HERVs are defined as retroviruses that have integrated into the human genome, thus resulting in heritable viral sequences. 8% of human DNA is composed of HERVs that, unlike naturally selected genes, were permanently incorporated into the genome up to 30 million years ago.[11] An initial explanation of retroviral DNA integration was devised by T. Fujiwara and K. Mizuuchi in 1988, through the method of simulating integration and analysing the reaction intermediates. Their investigation revolved around the joining of the 3' ends of retroviral long terminal repeat (LTR) sequences to samples of target DNA. LTRs are a transposable element of retroviruses. Studying of the intermediate confirmed the transformation of a viral RNA genome, through the protein *reverse transcriptase,* into a DNA genome that could feasibly integrate. The structure of the intermediate also suggested endonuclease activity; enzymes that cleave the phosphodiester bond within a polynucleotide DNA backbone. This cleavage has been theorised to provide energy for the action of integrase, an enzyme produced by the retrovirus that can form covalent bonds between the retroviral genetic information and the host cell.[12] As of this model, human endogenous retroviruses have played a substantial role in shaping the human genome. Not only have they physically altered the genome but held structural relevance throughout millennia. To fully understand the genetic influence of HERVs, I shall further explain the accelerating field of study surrounding the subject that illustrates their progressive relevance in genomics.

Despite the deep-rooted inhabitancy of HERVs in the genome, their existence was first established in the late 1960s, through the combination of virological and immunological methods with Mendelian genetic theory. Their existence was later confirmed by a DNA identification method known as nucleic acid hybridisation [11]. The identification methods of HERVs have expanded exponentially since this initial discovery: in 2004, R. Belshaw utilised RepeatMaster software and phylogeny estimation to study HERVs alongside other techniques. The former is screening program for DNA sequences, used in this context to search for elements homologous to HERV-K10 and other known retroviral components. The latter estimated the evolution of HERVs by reconstructing a typical structure: a sequence of three retroviral genes (*gag, pol* and *env*), flanked by two LTRs.[13] Contrarily in 2016, L. Vargiu led the research of HERV sequences through a different identification program known as RetroTector. An identification of 3137 HERV sequences was accomplished. To further contrast the work of Belshaw, Vargiu references an alternate internal structure, in which four retroviral genes are present; the addition of a *pro* gene.[14] The continuous development of research methods indicates the complexity and elusive nature of the subject, and consequently the importance of studying human endogenous retroviruses.

In assessing the impact of HERVS, a clear division can be established between beneficial and detrimental effects. One fundamental advantage of HERV integration is the

physiological role of envelope proteins in placenta function. Two envelope glycoproteins, encoded separately by members of the retroviral variant families HERV-W and HERV-FRD, possess anti-receptors that bind to corresponding receptors on multiple host cell membranes. This mechanism mediates membrane fusions between cells, inducing the formation of a vital placental layer - the syncytiotrophoblast. This multinucleated layer enables for embryo implantation, protection against pathogens, and hormonal secretions, alongside acting as an exchange surface.[15]

Furthermore, HERVs can be advantageous in terms of viral infection resistance. For example, a defective endogenous retrovirus related to HERV-L known as *Fv1* can interact with the core of a murine leukaemia virus particle and prevent infection.[16] Another example is the contribution to vaginal HSV-2 disease protection. An investigation led by R. Jayewickreme found that mice with particular infectious endogenous retroviruses can be detected through the presence of a Toll-like receptor 7 (TLR7) deficiency. The study compared two groups of mice, one group with the deficiency (and the consequent ERV) and one without, both exposed to HSV-2. Conclusions observed that, after 3 days, the TLR7 deficient group experienced minimal local inflammation of the vagina, while no members of the second group survived the infection. Observation of vaginal tissue revealed an evident correlation between endogenous retrovirus expression and HSV-2 resistance.[17] Although, this evidence derives from the study of a mouse genome, thus reducing its validity in terms of the human genome; mice and human infection restrictions occur at differing viral entry stages. As of this, it can be considered that beneficial shaping of the genome via HERVs lacks sufficient evidence for confirmation.

To further question the overall consequence of human endogenous retroviruses on the genome, I shall explore the potential threat they pose on human health. One argument for the disadvantages of HERVs is the correlation between abnormal retrovirus expression and many physiological disorders. For example, in multiple sclerosis patients, overexpression of HERV-H, -K and -W occurs. HERV envelope proteins, secreted by HERV-W, block the activation of neuroreceptors, hence exacerbating the axonal disruption that causes MS. [18] Furthermore, the same upregulated HERV has been associated with schizophrenia, due to its presence in both the brain tissue and cerebrospinal fluid of schizophrenia patients.[16] Overexpression of HERV-W in blood cells is common in chronic inflammatory demyelinating polyradiculoneuropathy patients and that of HERV-K has been linked with sporadic amyotrophic lateral sclerosis patients; the frequent correlations are undisputable examples of HERVs detrimentally shaping the human genome.[18]

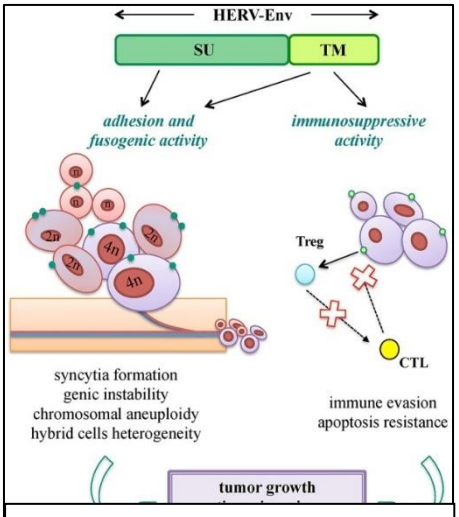

Figure 1[11]

Moreover, subsequent to their discovery, continual studies have sought to prove the link between HERVs and tumour induction. Though, as prior mentioned, cell-cell binding due to HERV integration can be considered an advantage in its placental development role, it's additionally observed in relation to cancerous tumours. The fusion of cells

underpins the growth of tumours [Fig. 1], the migration of cancer cells around an organism (metastasis), and resistance to therapeutic treatment.[19] Another connection between HERVs and tumour formation refers once again to envelope proteins produced by members of the HERV-K family. The proteins bind the factor that transcribes DNA into RNA for spermatogenesis, thus impairing the process. Disrupting spermatogenesis is associated with increased germ-cell tumour frequency.[16] In this regard, one can argue that the genome has been detrimentally shaped by the consequences of HERV integration.

However, the overall conclusion is still disputed amongst experts in the field; Vargiu states *"there is so far no proof of HERV-induced disease."* [14] In addition, though HERV-K envelope proteins threaten human health, their frequent correlation to tumorigenesis subjects them to use as diagnostic markers.[20] By opening the door to new screening techniques that allow for early cancer detection, I consider the presence of HERVs in shaping the human genome to be one of excitement and unascertained potential.

When comparing the role of human endogenous retroviruses to that of natural disease susceptibility in shaping the genome, it is difficult to assess which factor has had a greater impact. In one perspective, HERVs have summated greater effects. Not only are they a predicted component of the genome (unlike mutations), but they reside in every member of the human population. This is dissimilar to disease-resistant genes; changes in mutation frequency only tend to occur in smaller populations that are most at-risk to infection.[6] However, recent studies prove a recent decline in the rate of HERV integration. In 2015, Margiorkinis correlated this to an increase in body mass over time, and the gradual extinction of a HERV family.[21] As of this, perhaps the relevance of HERVs shaping the genome will disappear; yet another unknown that evolution is yet to determine.

To conclude, infection has domineered genetic alteration for the entirety of human evolution. As of pathogenic complexity, it is assumable that disease will continue to shape the human genome in ways we cannot yet predict. Importantly, studying infection in the context of the genome may help us clarify the mechanisms of disease progression, and the identification of therapeutic treatments.[9] In a topical manner, the impact of COVID-19 has accelerated this field of study. Though the infectious disease has put countless lives at detriment, I believe we can use this factor to perceive COVID-19 with a optimistic outlook: we are gaining the ability to better understand infection, human genetics, and the medical progress that adjoins scientific discovery.

Holly Powell

Bibliography

1. Karlsson, E., Kwiatkowski, D., Sabeti, P. (2014), *Natural selection and infectious disease in human populations.* [online] Last accessed 7 March 2022: https://www.nature.com/articles/nrg3734
2. Inhorn, M., Brown, P. (1990) 'Annual Review of Anthropology', *The Anthropology of Infectious Disease,* vol. 19, pp. 94
3. Tibayrenc, M. (2017), *Genetics and Evolution of Infectious Diseases,* Amsterdam, Elsevier
4. Kalendar, R. et al., (2022) *Modified "Allele-Specific qPCR" Method for SNP Genotyping Based on FRET.* [online] Last accessed 7 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8784781/
5. Katzourakis, A., Rambaut, A., Pybus, O. (2005), *The evolutionary dynamics of endogenous retroviruses.* [online] Last accessed 6 March 2022: https://doi.org/10.1016/j.tim.2005.08.004
6. Wayne, M., Bolker, B. (2015), *Infectious Disease A Very Short Introduction*, Oxford, Oxford University Press
7. Hay, S. et al., (2011) *The global distribution and population at risk of malaria: past, present, and future*. [online] Last accessed 7 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3145123/
8. Tankeshwar, A. (2021) *Relationship between sickle cell anemia and malaria.* [online] Last accessed 8 March 2022: https://microbeonline.com/relationship-sickle-cell-anemia-malaria/
9. Brinkworth, J. (2017) 'Human Biology', *Infectious Disease and the Diversification of the Human Genome,* vol. 89, no 1, pp. 50
10. Voronin, Y., Holte, S., Overbaugh, J., Emerman, M. (2009) *Genetic Drift of HIV Populations in Culture.* [online] Last accessed 8 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652835/
11. Weiss, R.A. (2006), *The discovery of endogenous retroviruses*. [online] Last accessed 6 March 2022: https://doi.org/10.1186/1742-4690-3-67
12. Fujiwara, T., Mizuuchi, K. (1988) *Retroviral DNA integration: structure of an integration intermediate*. [online] Last accessed 6 March 2022: https://pubmed.ncbi.nlm.nih.gov/3401925/
13. Belshaw, R. et al., (2004), *Long-term reinfection of the human genome by endogenous retroviruses*. [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC387345/
14. Vargiu, L. et al., (2016), *Classification and characterization of human endogenous retroviruses; mosaic forms are common.* [online] Last accessed 6 March 2022: https://rdcu.be/cIy0S
15. Lavialle, C. et al., (2013), *Paleovirology of 'syncytins', retroviral* env *genes exapted for a role in placentation*. [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758191/
16. Griffiths, D. (2001), *Endogenous retroviruses in the human genome sequence*. [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC138943/
17. Jayewickreme, R. et al., (2022), *Endogenous Retroviruses Provide Protection Against Vaginal HSV-2 Disease.* [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8764156/

18. Katsura, Y., Asai, S. (2019), *Evolutionary Medicine of Retroviruses in the Human Genome*. [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7093845/
19. Grandi, N., Tramontano, E. (2018), *HERV Envelope Proteins: Physiological Role and Pathogenic Potential in Cancer and Autoimmunity.* [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861771/
20. Bannert, N., Hofmann, H., Block, A., Hohn, O. (2018), *HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors.* [online] Last accessed 6 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5816757/
21. Magiorkinis, G., Blanco-Melo, D., Belshaw, R. (2015) *The decline of human endogenous retroviruses: extinction and survival.* [online] Last accessed 8 March 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4335370/