Proceedings of the inaugural Dark Genome Symposium: November 2022

Jef D. Boeke[1,2,3], Kathleen H. Burns[4], Katherine B. Chiappinelli[5], Marie Classon[6], John M. Coffin[7], Daniel D. DeCarvalho[8,9], Joseph D. Dukes[10], Benjamin Greenbaum[11], George Kassiotis[12,13], Sarah K. Knutson[14], Arnold J. Levine[15], Avindra Nath[16], Sophie Papa[10,17]*, Daniel Rios[14], John Sedivy[18,19] and David T. Ting[20]

Abstract

In November 2022 the first Dark Genome Symposium was held in Boston, USA. The meeting was hosted by Rome Therapeutics and Enara Bio, two biotechnology companies working on translating our growing understanding of this vast genetic landscape into therapies for human disease. The spirit and ambition of the meeting was one of shared knowledge, looking to strengthen the network of researchers engaged in the field. The meeting opened with a welcome from Rosana Kapeller and Kevin Pojasek followed by a first session of field defining talks from key academics in the space. A series of panels, bringing together academia and industry views, were then convened covering a wide range of pertinent topics. Finally, Richard Young and David Ting gave their views on the future direction and promise for patient impact inherent in the growing understanding of the Dark Genome.

*Correspondence:
Sophie Papa
sophie.papa@enarabio.com

[1] Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA

[2] Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201, USA

[3] Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

[4] Department of Pathology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

[5] Department of Microbiology, Immunology and Tropical Medicine, The George Washington University, Washington, DC, USA

[6] Pfizer Centre for Therapeutic Innovation, San Diego, USA

[7] Department of Molecular Biology and Microbiology, Tufts University, Boston, MA 02111, USA

[8] Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

[9] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

[10] Enara Bio Limited, Magdalen Centre, 1 Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA, UK

[11] Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[12] Retroviral Immunology Laboratory, The Francis Crick Institute, London, UK

[13] Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK

[14] Rome Therapeutics, 201 Brookline Avenue, Suite 1001, Boston, MA, USA

[15] Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ, USA

[16] Section for Infections of the Nervous System, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

[17] School of Cancer and Pharmaceutical Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK

[18] Center on the Biology of Aging, Brown University, Providence, RI, USA

[19] Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA

[20] Department of Medical Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

In the following proceedings paper the contributing speakers, panel introduction speakers and panel moderators have summarized the topics covered.

Jef Boeke: *The Dark Matter Project: writing mammalian DNA*

The dark genome is littered with the dead bodies of retrotransposons – and although they may have lost their ability to jump around the genome, those insertions might have consequences for phenotype. However, retrotransposon insertions in mammalian introns are widely assumed to be of no consequence, since introns simply get spliced out. Mammalian retrotransposons include Long interspersed nuclear elements (LINEs) which include the machinery for retrotransposition and encode the machinery for their movement to new sites – they are thus called "Autonomous". Nonautonomous human elements that exploit the LINE-encoded machinery include the abundant Arthrobacter luteus (*Alu*) restriction endonuclease sequence, and the less abundant (ca. 1000 copies) SINE-VNTR-*Alu*s (SVA) sequence. However, we have made two sets of findings that challenge this widely held assumption. One of these relates to the human *Alu* Sequence, an element found in over a million copies in the human genome; many of those copies are intronic. Could any of these intronic copies have specific effects on gene expression? Bo Xia in the Yanai and Boeke groups explored the loss of the tail in hominids, one of the main anatomical evolutionary changes to have occurred along the lineage leading to humans. Morphological reprogramming in ancestral hominoids has been long considered to have accommodated a characteristic style of locomotion and contributed to the evolution of bipedalism in humans. However, the precise genetic mechanism that facilitated tail-loss evolution in hominoids remains largely unknown. Primate genome sequencing projects have made possible the identification of causal links between genotypic and phenotypic changes, and could enable the search for hominoid-specific genetic elements controlling tail development. Bo found evidence that tail-loss evolution was mediated by the insertion of an individual *Alu* element into the genome of the hominoid ancestor. He demonstrated that this *Alu* element – inserted deep in the middle of an intron of the T-Box Transcription Factor T *(TBXT)* gene (also called *T* or *Brachyury*) – pairs with a neighboring *Alu* element encoded in the reverse genomic orientation in the downstream intron and leads to a hominoid-specific alternative splicing event. Follow up studies in the mouse suggest that alternative splicing variation such as those caused by the *Alu* insertion can influence the length of the tail. This provides strong presumptive evidence that this particular *AluY* insertion was a pivotal molecular development of loss of the tail in hominoids,

as it is only observed in the hominoid lineage and not in other (tailed) primates [1].

A second case relates to the rare human disease X-linked dystonia parkinsonism (XDP), which affects males from the island of Panay in the Philippines in their fourth decade of life. This devastating disorder, which is uniformly fatal and has no cure or even treatment, is caused by an ancestral SVA retrotransposon insertion deep in the middle of a large intron (intron 32) of the essential TATA-box binding protein associated factor 1 (*TAF1)* gene. This SVA insertion presumably occurred once in a single female from Panay (a so called "founder effect") within relatively recent history, and the subsequent inheritance of the insertion by male offspring has revealed untold suffering in those unlucky enough to inherit it. SVA is a human-specific nonautonomous retrotransposon which is not very well studied. In order to better understand the disorder and its molecular etiology, and potentially to test therapeutics for treating it, a practical animal model would be useful. We used "Big DNA" technology, developed in our lab, to solve a thorny problem – how do you engineer the gene to contain a potentially toxic retrotransposon insertion in an essential gene on the X chromosome in mouse embryonic stem cells which are male. We developed a strategy for doing this using a conditional insertion carrying a "convertible allele" controlled by a recombinase. Mice derived from this "convertible allele" by breeding them with driver mice that convert the allele to partially human, in a tissue specific manner and observed phenotypes in the appropriate male mice (only) that resemble certain human XDP symptoms. Thus, a system now exists to better dissect the impact of a disease causing SVA insertion, in the appropriate tissues of a living animal.

Kathleen H. Burns: *Transposable elements in cancer*

This seminar began with an overview of the contributions of transposable elements and interspersed repeat sequences to human genome composition. The ongoing activity of LINE-1 sequences was highlighted, particularly with respect to its contributions to genetic diversity in human populations via retrotransposition of LINE-1, *Alu* and SVA sequences. Rarely, but recurrently, mobile element insertions generate loss-of-function alleles causing genetic disease, such as the hemophilia allele described by Kazazian and colleagues in 1988 [2]. More recent work, by our lab and others, has shown that commonly occurring insertion polymorphisms can also impact heritable disease risk by altering mRNA splicing or gene expression [3–5].

The primary focus of our discussion then turned to LINE-1 dysregulation in human cancers. Across a wide variety of commonly occurring, p53-mutated epithelial cancers, LINE-1 open reading frame 1 protein (ORF1p)

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 3 of 13

can be detected in tissue biopsies [6]. Moreover, we shared very recent evidence that this candidate 'binary' cancer biomarker can be appreciated in the peripheral blood of patients with advanced cancers using ultrasensitive protein detection assays [7]. These findings raise the possibility that LINE-1 expression may be leveraged for early detection or disease monitoring. We next reviewed data that LINE-1 expression is associated with somatic retrotransposition of LINE-1 sequences in cancers: several independent groups have now converged on the recognition that LINE-1 ORF2p is a mutator of cancer genomes [8]. We described published and unpublished data from our group that LINE-1 expression may incite chromosomal instability (CIN) and shape selective pressures on cells within cancers and premalignant lesions. For example, promoting tumor protein p53 (p53) loss or imposing requirements for DNA repair [9]. The paradigm that epigenetic dysregulation of LINE-1 is a hallmark of cancers where its expression promotes a 'mutator phenotype' and imposes molecular dependencies raises therapeutic implications, presenting important future directions for our field.

Arnold J. Levine: *Review of the role of Tp53 gene and protein in regulating epigenetic changes and LINE-1 gene expression*

The first demonstration that the p53 protein plays a role in the regulation of epigenetic changes came from the Jaenisch laboratory [10]. They created mice with a conditional cyclization recombinase (CRE) mediated deletion of the DNA methyltransferase 1 (DNMT1) gene (the copy DNA methy-cytosine transferase). This enzyme adds a methyl group (CH3) onto cytosines in CpG dinucleotides opposite GpC-methyl cytosine on the paired DNA strand after DNA replication, duplicating the epigenetic marks. In response to eliminating both DNMT1 genes with Cre cells were shown to replicate twice producing cells with no methylation of the C-residues in CpG dinucleotides and those cells died of apoptosis. When Tp53 genes were deleted from these DNMT1 deleted genes the cells failed to undergo apoptosis divided and ultimately became transformed and tumorigenic. These experiments suggested that alterations in epigenetic marks on the DNA were detected by p53 and these cells were killed by a p53 mediated apoptosis. This interpretation is confirmed by the results of the Yamanaka experiment [11]. In this case four different transcription factors are added to a fibroblastic cell in culture and through cell division the epigenetic marks are erased, and the cells dedifferentiate into a stem cell that can be reprogramed to produce many different cell and tissue types. When this occurs in cell culture the efficiency of stem cell production is very low (0.1-1.0%) and the time taken to produce these stem cells in culture is commonly one week to one month. However,

in the absence of the wild type p53 protein (using a temperature sensitive p53 protein) the efficiency increases up to 80% and the time to produce stem cells decreases to days. With a temperature sensitive p53 protein stem cells are produced at high frequencies in short times only at the non-permissive temperature. Temperature shifts map the times of cell death by apoptosis [12]. A third type of experiment also supports these ideas. The drugs azacytidine or decitabine (the deoxyribose form of the drug) are faulty incorporators into RNA or DNA at cytidine resides and prevent methylation of cytosines at CpG dinucleotides. These drugs will kill cells more efficiently that have lost the p53 wild type functions sparing cells that have wild type p53 functions [13]. A similar observation has been made in tumor xenografts in mice and in humans [14]. The reciprocal relationship between mutant p53 and epigenetic changes seems to be operative as well [15]. When p53 mutations occur in both alleles of the Tp53 gene epigenetic changes occur in selected regions of the genome and this results in Line-1 ORF-1 and 2 transcriptional gene expression. Presumably when LINE-1 transcription occurs in a cell with wild type p53 protein the cell will be eliminated by apoptosis or other methods of cell death mediated at least in part by the p53 pathway. As this happens over a lifetime, antibodies are produced by the adaptive immune system, (or perhaps the innate immune system) that give many individuals a background low level of antibodies directed against ORF-1 whose levels can increase as a cancer develops (talk by the Gudkov group). It is not uncommon to detect increasing levels of Tp53 mutations in people as they age [16].

The exact mechanism(s) by which the p53 protein or pathway senses changes in epigenetic marks is not yet clear.

John Coffin: *Endogenous Retroviruses are not Retrotransposons: The Case of HERV-K (HML-2)*

The genomes of all vertebrates – and many invertebrates – are laced with endogenous retroviruses (ERVs), segments of DNA proviruses derived from retroviral infection of the germline of a distant ancestor. In humans, such endogenous retroviruses (HERVs), numbering about 80,000 distinct elements, are all defective, likely the consequence of evolutionary selection against individuals with intact proviruses, those capable of giving rise to pathogenic infectious virus during development. While it is commonplace to consider ERVs as "retrotransposons," implying intracellular propagation within the germ line, I consider this to be a highly misleading classification, because of the large differences in biology and evolution between ERVs and true retrotransposons (such as LINEs). ERVs have several features that render them incompatible with an intracellular replication cycle, such as delaying proteolytic processing of virion proteins

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 4 of 13

until after budding and release of virions. Thus, while retrotransposons make a home in the genome, for ERVs, the genome is better considered as a graveyard where the losers are buried, providing us with a fossil record of long-extinct pandemic infection of our distant ancestors. Nevertheless, A few of the germline proviruses have been coopted for functions useful to the host, such as syncytins, ERV envelope (*env)* genes that play an important role in early fetal development. Perhaps the most common (and most underappreciated) ERV function is as a restriction factor to prevent infection with a related exogenous virus, for example by receptor blockade due to expression of an Env protein.

Our laboratory has been studying the most recently active HERV, human endogenous retrovirus-K (HERV-K/HML-2), which appears to have gone extinct in the human lineage within the last million years (although much more recently in the gorilla lineage), leaving about a thousand proviruses in our genome. 900 of these proviruses have been reduced to solo LTRs, and the remaining 100 or so are all defective for replication, although some are still capable of expressing one or a few viral gene products, and even non-infectious virus-like particles (VLPs). Much of the recent literature on HML-2 expression is based on two implicit assumptions: (1) That any detected HML-2 expression implies the presence of all functional gene products; (2) That expression is confined to disease states, particularly cancer. Neither of these is correct. Of all the proviruses with 2 long terminal repeats (LTRs), no two are alike in the pattern of inactivating mutations, and the potential for functional protein expression varies considerably from one to the next. Recently, we have used the Genotype-Tissue Expression (GTEx) RNA-seq database of 54 non-diseased tissues from 900 donors to assess the expression of all known HML-2 proviruses [17]. We found expression of one or more proviruses in one or more tissues, no two of which showed the same pattern. Of particular interest is the 8 + million-year-old provirus on chromosome 3 at 3q12.3, which was expressed in all 53 solid tissues tested due to a post-integration mutation in the 5'LTR. This mutation created a binding site for a common transcription factor, allowing widespread expression of a variant gag protein. We are now pursuing the hypothesis that this protein was coopted as a factor to restrict replication of the now extinct HML-2 virus.

George Kassiotis: *Antibodies to endogenous retroviruses underpin lung cancer immunotherapy*

The effectiveness of immunotherapies in a sizeable proportion of patients with several cancer types has not only transformed patients' lives, it has also validated a long-held belief that the immune system can be stimulated to protect against tumors. While most of this protective effect has been traditionally attributed to T cells, a growing body of evidence points to a possible contribution of B cells and, by extension, antibodies to immune protection against tumors and to the success of immunotherapies. However, the mechanism by which B cells may mediate their effects is still incompletely understood. We hypothesized that this effect is attributable to the production of anti-tumor antibodies and set out to identify their putative targets.

The team used a mouse model for lung adenocarcinoma, newly generated by Julian Downward's group at the Francis Crick Institute, that, unlike earlier models, is highly immunogenic, and presented evidence for the formation of tertiary lymphoid structures around lung tumors and for local germinal center reactions [18]. These were accompanied by the induction of highly protective anti-tumor antibodies. Perhaps unsurprisingly for a mouse cancer model, the team demonstrated that anti-tumor antibodies spontaneously induced during the course of tumor challenge targeted the envelope glycoproteins of infectious murine leukaemia viruses (MLV) "resurrected" from defective endogenous retrovirus (ERV) precursors in the tumor cells. Expression of MLV envelope glycoproteins was both necessary and sufficient for the immunogenicity of lung cancer cells and the induction of anti-tumor antibodies. Moreover, these antibodies were significantly boosted following checkpoint blockade, which further enhanced their protective capacity, suggesting that immunotherapies work at least in part through enhancement of the anti-tumor antibody response.

Given the evolutionary divergence of murine and human ERVs, the team then explored if phylogenetically unrelated ERVs in humans could serve the same purpose as the murine ones. By probing ERVs with an intact or near intact envelope gene, they identified three distinct proviruses with high expression in lung cancer samples, one of which, a member of the HERV-K(HML-2) family, additionally showed specific upregulation in lung adenocarcinoma (LUAD). Working with Charles Swanton's group at the Francis Crick Institute and using bespoke assays for antibodies against these envelope glycoproteins, the team reported spontaneous antibody responses specifically in LUAD and specifically to HERV-K(HML-2). The team further reported that antibodies to HERV-K(HML-2) envelope glycoproteins were induced or further boosted, if they pre-existed, in the weeks following immunotherapy of LUAD patients. Finally, in collaboration with colleagues at the Samsung Medical Centre in Seoul, Republic of Korea, they found that pre-treatment expression of the particular HERV-K(HML-2) provirus on Chromosome 1q22 predicted the outcome of immunotherapy in a larger cohort of LUAD patients.

Boeke *et al. Mobile DNA* (2023) 14:18

Page 5 of 13

The reported findings [18] raised the possibility that the reactivation of normally repressed ERVs provides non-mutated cancer-associated targets of viral origin for immune recognition of transformed cells, with implications for improvement of cancer immunotherapies.

Avindra Nath: *Role of retroviruses in Amyotrophic Lateral Sclerosis*

The possibility that retroviruses may be implicated in the pathophysiology of motor neuron disease or amyotrophic lateral sclerosis (ALS) dates to the 1960's, before the term "retrovirus" had been coined. PG Stansly described a "paralytic disease involving upper and lower motor neurons associated with virus-induced neoplasms of the mouse" [19]. He elaborated the history of the discovery, the clinical features, the pathology, the long latency, and the effect of the age of the host on the development of the syndrome by a filterable agent that was also associated with lymphomas. Subsequently, Murray Gardner and colleagues isolated the infectious agent and identified it as MLV. The pathology and clinical course resembles that of ALS with a long latency period and absence of inflammation [20]. Extensive work followed describing the molecular pathophysiology of MLV associated motor neuron disease [21–23]. This also initiated the hunt for retroviruses in patients with ALS.

Several research groups have consistently demonstrated the presence of reverse transcriptase activity in patients with ALS [24–27]. Further it has been shown that HIV and HTLV-I infection can also cause an ALS-like syndrome in some individuals [28, 29]. However, attempts to isolate a retrovirus from ALS patients have consistently failed. We and others have shown that an endogenous retrovirus HERV-K (subtype HML-2) gets activated in the brain and spinal cord of a subgroup of patients with ALS [30, 31]. Further, the loci activated in ALS have open reading frames for the viral proteins which include the envelope, reverse transcriptase and gag proteins [32]. Some investigators have not been able to demonstrate any difference between ALS and controls for HERV-K activation. We found that the envelope protein is toxic to neurons and transgenic animals that express the protein in neurons develop an ALS like syndrome [30]. The mechanism of neurotoxicity involves the cleavage of the signal peptide with nucleolar localization and disruption of protein synthesis (unpublished). Envelope protein can also be identified in the cerebrospinal fluid (CSF) of ALS patients [33] who develop an immune response against the virus, and analysis of the antibody binding epitopes shows that over the course of the illness there is evidence for epitope spreading with unique antibodies directed against the transmembrane domain of the protein in ALS patients only [34]. Through the process of drug screening, others and we have identified several antiretroviral drugs with variable efficacy against HML-2 [35, 36]. Of these a reverse transcriptase inhibitor, Abacavir has the best efficacy [37]. An open label study using Triumeq (abacavir, dolutegravir, and lamuvidine) in ALS patients showed that over a period of 24 weeks, the HML-2 viral load decreased in a subset of patients who also showed slower progression of the disease [38] (Garcia-Montojo et al., 2021). Based on these results, a double blind, placebo-controlled study has been initiated in Europe, Australia and New Zealand. However, there is a great need for better and more specific antiretrovirals against this virus.

John Sedivy: *Retrotransposon Activation and Consequences in Senescent Cells and Aging Tissue*

Aging is associated with extensive remodeling of the epigenome [39]. The age-related changes are complex but show a consistent trend for loss of constitutive heterochromatin [40–42]. Cellular senescence is an arrest of proliferation, studied primarily in mammals, which can be elicited by a variety of stresses including DNA damage [43]. While it has beneficial functions (such as tumor suppression), senescent cells accumulate in most tissues with age, and are an important component of the overall aging process [44]. In particular, due to their profound proinflammatory phenotype [45], senescent cells have been causally linked with many age-related pathologies and diseases [46]. Senescence is accompanied by widespread opening of heterochromatic regions, some of which were mapped to retrotransposable elements and associated with their derepression [47].

Retrotransposition in the soma appears to be very low. Line-1 mRNA is found in the adult brain [48–50] and neural stem cells can support retrotransposition [51], but insertions in adult neurons are only in the range of 0.5-1.0 events per cell [52–55]. In other tissues retrotransposition appears to be significantly lower [55]. Cancer is associated with increased LINE-1 expression [56] and many insertions, in some cases hundreds in an individual tumor, have been reported [8, 57–59]. Most appear to be passenger events in non-coding regions [60, 61].

In eukaryotic cells nucleic acid species such as DNA or double stranded RNA (dsRNA) in the cytoplasm are perceived as invading pathogens. In vertebrates this triggers the Type I Interferon (IFN-I) response: expression of interferons alpha and beta to alarm neighboring cells, a variety of cytokines and chemokines to communicate with the immune system, a large number of factors that interfere with viral replication, and in some cases cause cell death [62]. Cytoplasmic LINE-1 DNA is present in senescent cells and in tissues of mice in association with normal aging [63, 64]. Importantly, LINE-1 DNA was identified in association with the DNA sensor cyclic GMP-AMP Synthase (cGAS) [64].

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 6 of 13

Treatment with nucleoside reverse transcriptase inhibitor (NRTI) drugs, or a knockdown of LINE-1 mRNA with shRNAs against active LINE-1s, decreased cytoplasmic LINE-1 DNAs and ameliorated cGAS and IFN-I activation [63–65], indicating that cytoplasmic LINE-1 cDNA can be a potent trigger of an IFN-I response. The mechanisms by which cytoplasmic LINE-1 cytoplasmic DNAs are produced are not yet understood. Normally LINE-1 reverse transcription occurs in the nucleus by target-primed reverse transcription (TPRT) [9, 66]. One possibility is that cytoplasmic DNA species are abortive TPRT products that are somehow exported from the nucleus. Alternatively, reverse transcription could occur directly in the cytoplasm. Both purified ORF2 and LINE-1 RNPs purified from cells can reverse transcribe in vitro if provided with a primer [67, 68].

One consistent feature has been the efficacy of NRTIs, in multiple models and situations, to alleviate the effects of retrotransposable element activation. This finding is intriguing given that some aspects of the retrotransposable elements lifecycle, such as transcription of the elements, or stimulation of RNA sensors, should not be directly affected by NRTIs. One explanation might be downstream cross-talk between RNA and DNA sensing pathways and the existence of overall self-tolerance thresholds. Much more research needs to be done in this area, and it will thus be important to determine, for any given disease, which family of retrotransposable elements is being activated and whether DNA or RNA sensing pathways are engaged.

Panels.

Moderator: Joe Dukes

Mapping the dark genome

Opening speaker and panellist: Ben Greenbaum

Panellists: Rachel O'Neill

Menachem Fromer

Alice Lee

Samuel Lukowski

The first panel discussion explored the history and challenges of mapping traditionally thought to be noncoding regions of the human genome, where successes in this space emerged from and some of the remaining challenges especially in moving towards applications.

Benjamin Greenbaum provided an introductory overview talk on the topic. The main questions raised were what is the structure of information in the dark genome, what aspects continue to evolve and how do we assign function to repeats in healthy tissue, in disease and under therapy? In the pre-next generation sequencing era repeats were assessed by the rates of DNA reannealing. However next generation sequencing has opened an era of precise genome mapping and comparative genomics [69]. It has become clear that long-read sequencing has

uncovered missing dark genome events, and likewise that biased transcriptomics have missed repeats in cancer with immune consequences [70, 71].

Cancer was highlighted as an area where the underassaying of the repeatome has led to a loss of detection of consequential activity. Repeats which are expressed in cancer are being missed [70], and total RNAseq is needed to quantify them [71]. Recent work has shown that LINE-1 continues to reverse transcribe in human cancers [8, 72], particularly those associated with p53 mutation loss of function [73], indicating events that are missed due to lack of whole genome sequencing. Likewise, HERV-K expression has been associated with immunotherapy response in renal cell [74] and bladder cancer [71]. Immunostimulatory repeats which engage in viral mimicry by displaying viral associated molecular patterns are also being under quantified [71]. Emerging methods from statistical physics are likely able to predict their immune function when sequenced [75].

Rachel O'Neill kicked off the discussion by reminding the audience that despite what was dubbed the publication of the entire human genome in 2001 [76, 77], it was missing around 8% of the sequence due to limitations in the sequencing methods and technologies. Only in 2022, through the impressive collaborative efforts of the Telomer-to-Telomer (T2T) consortium along with technological advances, were this team able to finally publish hg38, a fully mapped human genome [78]. Some of the key challenges to overcome were the repetitive elements as well as sequencing through centromeres. A reoccurring theme of the panel discussion was how limited short read sequencing is for mapping and assembling genetic sequences. The advent of long read sequencing, especially through Oxford Nanopore and PacBio's platforms, have been key technological breakthroughs that have helped to unlock the ability to map such sequences. In addition, improvements in computational tools have been critical to success. The T2T group, consisting of collaborative efforts from around 100 different labs - in the absence of direct funding for this work, was touted as the poster child for how the field needs to continue to operate openly to drive efficient success.

Some of the key challenges that remain in the field were noted to be around the annotation of sequences and associated pipelines, composite repeats and variation in human repeat elements among individuals. The next frontier must also interrogate the evolutionary landscape to understand, especially in genomic hotspots of the repeateome, our evolutionary history to better elucidate function and ultimately application.

Alice Lee brought these points towards a discussion on population scale data, reiterating that long read sequencing of repeats is a necessary means to adequately

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 7 of 13

generate this data. In particular, rare but important transposons, such as those only present in a small proportion of cells, need to be mapped and further characterized in order to better understand human health and disease. The challenge of annotation was again highlighted, here in the context of transposon events. The other challenge in looking at rarer events and especially solving this with single cell analyses, is the ability to confidently distinguish between true events rather than artefacts of working with limited depth. Samuel Lukowski picked up on this point highlighting that more whole transcriptome sequencing needs to be performed and, with approaches such as long read RNA-seq at the bulk RNA and/or single cell level, would greatly enhance our view of the tissue-, cell type-, and disease-relevant events.

Menachem Fromer turned the discussion to the challenge of mapping sequence data using LINE1 as an example. Here a major challenge is how to accurately map where in the genome a sequence is coming from and what the mechanisms of control are, given the repetitive nature of these sequences, with traditional pipelines assigning location to thousands of potential genomic regions. This mapping is critical for applicability to the drug development setting to meaningfully treat human disease, and once again picked up on the common theme discussed among the panel of the value and necessity of long read sequencing platforms and approaches.

The conversation precipitated general agreement that ultimately seeking to determine and ascribe functionality and role in human health and disease is critical for application. Efforts should focus on parsing out areas of the dark genome where this can be ascertained, whilst recognising there still may be some of the novel sequence space which may well have little-to-no function.

Moderator: Daniel Rios

Viral Mimicry – Impact of the Dark Genome on the Innate Immune Response in Autoimmune Diseases, neurodegeneration and cancer

Opening speaker and panelist: Daniel De Carvalho

Panelists: Andy Satlin

Tomas Mustelin

Dennis Zaller

Andrei Gudkov

Daniel De Carvalho summarized recent findings and highlighted the observation from his groups and many others in the field that disruptions to several cellular processes including DNA methylation, histone modifications, splicing, A-to-I RNA editing, RNA degradation, RNA modifications and RNA-binding proteins can induce an innate immune response named viral mimicry [79, 80]. This response is characterized by a buildup of immunogenic transcripts from endogenous retroelements, which can trigger activation of pattern recognition receptors leading to loss of cell fitness, cell death and a type I/III interferon response that links this innate immune response to a downstream adaptive immune response. Among those endogenous retroelements, Dr. De Carvalho highlighted the role of inverted-repeats alus (IR-alu) pairs. Cryptic transcription of IR-alus due to loss of epigenetic repression or retention of intronic IR-alu pairs due to inhibition of spliceosome leads to formation of dsRNA by intramolecular pairing of the two inverted alus and downstream activation of Melanoma Differentiation- Associated protein 5 (MDA5) - mitochondrial antiviral signaling protein (MAVS) and protein kinase-R (PKR) [81].

Altogether, Dr. De Carvalho proposed the "Fire Alarm Hypothesis", whereas some endogenous retroelements, such as IR-alus, can be considered triggers for a fire alarm that is activated upon disruptions to core cellular processes. This hypothesis suggests a previously unappreciated tumor suppressor function for endogenous retroelements and suggest that cancer cells with disruptions to these core cellular processes may become dependent in mechanisms to suppress viral mimicry [82, 83]. Dr. De Carvalho showed presented adenosine deaminase RNA specific-1 (ADAR1) as one example of such dependency [81] and suggested that other druggable dependencies may also exist. Finally, he raised the possibility that a range of autoimmune diseases, characterized by overproduction of type 1 interferons, could be attributed to non-malignant defects in many of the pathways required for maintenance and sensing of transposable elements leading to misfiring of the viral mimicry response.

A spirited conversation during the panel discussion ensued that touched on two key topics – the relevance of the "fire alarm hypothesis" to cancer, autoimmunity, aging and neurodegeneration, and how these observations can be translated to the clinic.

Several salient points were brought up regarding the "fire alarm hypothesis". Dr. Mustelin speculated that the at baseline these transposable elements may promote a type 1 interferon "tone" which would enable swift responses in the face of a novel challenge. Dr. De Carvalho noted that several interferon stimulated genes had ALU elements in their 3'UTR which may trigger a feed forward loop for a type 1 interferon response. While Dr. Zaller noted that in the case of autoimmunity we should consider these triggers as "false alarms", which would suggest that it may be possible to return the system to baseline with little to no unwanted effects.

The topic of clinical translation of Dark Genome research extended from the moderated session into the open question forum. Dr. Satlin (CSO, Transposon Therapeutics) suggested that in the absence of a complete understanding of the mechanistic underpinnings of

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 8 of 13

disease in a preclinical setting, orchestration of precision phase-2 (Ph2) clinical trials would be key to our understanding of disease and the role of transposable elements in disease pathogenesis. Dr. Zaller highlighted the data science component of ROME therapeutics which will enable precise characterization, and quantification of transposable elements as a key aspect of successful clinical translation. This data science driven approach could also allow for "backtranslation" of pre-existing clinical trial data. Drs. Gudkov and Mustlin observed that the theoretical mechanisms of activation of putative therapies might be intrinsically challenging to demonstrate given their indirect effects.

Overall, the panel agreed on the potential for regulation of the Dark Genome to play a role in multiple diseases including cancer, autoimmunity, neurodegeneration and aging. However, it was also apparent that significant hurdles remain to successful translation of this exciting area of research. Events like the Dark Genome Symposium will be critical to increase the profile of a former scientific backwater and facilitate the development of novel therapies across an incredibly wide range of diseases.

Moderator: Sophie Papa

Dark Antigens in cancer

Opening speaker and panellist: Kate Chiappinelli

Panellists: Ralf Leonhardt

Robert Manguso

Nina Bhardwaj

Lelia Delamarre

Following on from immunogenicity in disease driven by viral mimicry, Kate Chiappinelli introduced evidence for cancer antigen discovery and modulation for therapeutic development across TEs, specifically focusing on ERVs and cryptic open reading frames (ORFs). She introduced the evidence that treatment of some cancers with epigenetic modulating drugs DNA methyltransferases inhibitors (DNMTi) and histone deacetylase inhibitors (HDACi) leads to type 1 interferon responses, upregulation of antigen expression machinery, and increased expression of some cancer antigens [84, 85]. Furthermore, treatment of T cells with DNMTi upregulates effector genes, reversing T cell exhaustion and increasing their activity against virally-infected cells and cancer cells [86–88]. Can we take advantage of this increased immunogenicity to target cancer?

We know that some, but not all, cancers demonstrate upregulation of TEs at the RNA level, and level of expression of TEs broadly correlates with antigenicity in samples from The Cancer Genome Atlas (TCGA) [89]. Kate Chiappinelli used examples from ovarian and endometrial cancers to illustrate cancer specific L1 ORF1P and ERV expression [90–92]. Specific examples of these non-canonical antigens include ERV-3

expression driving checkpoint inhibitor efficacy in kidney cancer [74] and ERV-K-Env pulsed dendritic cells (DCs) stimulating ovarian cancer patient T cell activation and cytotoxicity [93].

Indeed, chimeric antigen receptor T cell therapy directed against HERV-K shows activity in breast cancer models, which may translate to direct therapeutic potential [94]. To serve as T cell antigens, peptides derived from translates dark genome ORFs must be processed and presented on human leukocyte antigen (HLA) molecules for recognition by T cells. Evidence in support of this occurring in cancer includes loss of SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1), an epigenetic regulator of TE expression, resulting in TE upregulation in murine models of lung cancer and melanoma. Resultant HLA class I presentation of peptides and antigenicity was mitigated by knockout of beta-2 microglobulin [95]. Peptide-HLA prediction algorithms demonstrated upregulation of potential HLA-I epitopes in response to DNMTi and HDACi [96]; and CD8 T cell populations in acute myeloid leukemia (AML) patients recognized ERV-derived peptides both pre- and post DNMTi treatment [97]. Lastly, extension of the vast pool of potential antigens beyond TEs has recently been demonstrated through discovery of non-canonical ORFs in cancer [98]. These non-canonical peptides derive from unstable proteins, with evidence that presentation on HLA is up to 5 times more efficient, likely due to instability of the proteins [99].

Finally, the need to build out our limited evidence for differential impact of DNMTi inhibitors on non-canonical epitope presentation in cancer and normal tissue was touched on. The underlying epigenetic instability of cancers as well as compensatory epigenetic repression occurring in repeat regions that show DNA hypomethylation such as vulnerability derived from second level control loss from the SETDB1 pathway makes the likelihood of a wide therapeutic window between cancer and normal tissue high, but further work is needed to bring combination strategies into the clinic.

The panel discussion picked up on many of these themes. For antigens derived from the dark genome on cancer to compete with neoantigens as targets for vaccines or T cell therapies the panel highlighted the need for tumor specificity of peptide expression, homogeneous expression across cancer tissue, and a robust therapeutic window. Particularly for vaccine development, triggering of the innate immune response was considered key.

Next the panel discussed evidence for immunogenicity of TEs. The evidence is robust in murine models with dark genome derived antigenicity of MLV env a key exemplar. Lelia Delamarre stressed that even in murine models these epitopes are variable in their

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 9 of 13

immunogenicity. Evidence in humans is more patchy but exists and is expanding.

The topic of risk and therapeutic windows was then raised, especially in the context of combination therapies, including potential triplets for vaccines, checkpoint inhibitors and DNMTi. Robert Manguso highlighted that the clinical acceptance of off target inflammatory toxicity is relatively high with the adverse events routinely seen with checkpoint inhibitors. Highly cancer specific targets are key to derisking. Ralf Leonhardt emphasised that it is peptide expression that will define normal tissue risk rather than RNA levels. We have limited tools to quantify translation rather than transcription in tissues but newer approaches, such as Ribo-seq are helping enhance our confidence. Nina Bhardwaj made the point that risk is higher with T cell therapies than with vaccines but that we need to broaden and deepen our understanding of normal tissue and rare cell expression of targets of interest to build confidence.

Moderator: Sarah Knutson

The Dark Genome and gene regulation

Opening Speaker and panellist: Marie Classon

Panellist: Brad Bernstein

Josh Mandel-Brehm

Danuta Jeziorska

The Dark Genome makes up 98% of our genetic material, and misregulation of these non-coding and repetitive regions Has been implicated in a variety of diseases. We are just scratching the surface in our understanding of Dark Genome regulation: changes in DNA methylation and other epigenetic regulatory mechanisms, the role of transcription factors in tuning the expression of regulatory elements, genomic interactions in 3D and disease-linked variants. How do we illuminate more of the Dark Genome, and apply what we already know to development, disease, and drug discovery? The speakers in this session brought multiple perspectives to key challenges facing the field of Dark Genome regulation.

A key to addressing any disease is to understand if the target of interest is a cause or consequence of disease development. The vastness of the non-coding part of the genome provides an even grander challenge.

Expression of transposable elements and endogenous retroviruses are correlated with cancer and inflammatory diseases. The challenge is determining if and in which context there may be a vulnerability to repetitive element expression, by using genetic tools and small molecule inhibitors (as demonstrated by previous work of the academic, pharma, and biotech panel members). It should also be noted that the diseased cells may have harnessed adaptive mechanisms such as the downregulation of the Stimulator of Interferon Genes (STING) pathway during the evolution of some cancers as an example. Instead

of therapeutic targets, we could also consider repetitive element profiles as biomarkers, which would allow for a precision approach for patient selection. As an example, cancer cells that carry an abundance of double stranded Alu RNAs, or other cytoplasmic nucleic acid species (reminiscent of a viral infection), may be more sensitive to the loss of ADAR.

We are also moving toward understanding which mutations in the Dark Genome affect genome function, and the connection to which genes are directly affected in disease. Using this concept, complex disease systems (such as cancer) may not be the first place you start with for drug discovery; instead, starting with monogenic diseases could provide a more streamlined model for proof-of-concept?

Dark Genome elements and peptides are structurally and functionally different between mouse and human, and their regulation is not always conserved between species. This creates a challenge both to our understanding of basic biological mechanisms involving the dark genome and to therapeutic development opportunities that rely on the use of model organisms in this space. Promising efforts have been made to generate mice with large inserts of humanized genomic sequences to mimic the Dark Genome regions in monogenic diseases [100]. While the expression patterns are correct for those regions, the goal would be to achieve disease-level expression through improved cis-elements regulation and even larger genomic regions to correlate to the human disease.

A mouse model of human cancer specifically engineered to upregulate Dark Genome elements through loss of an epigenetic modifier (ref. SETDB1/HERV example) suggests an improved response to checkpoint inhibitors [95]. While the functions of epigenetic regulators may be similar between mouse and human, there is divergent repetitive element structure and function between species, and interpretation of the data would benefit from the compliment of Dark Genome analysis in related human tumor samples.

To delve into the specific regulation of the Dark Genome, the challenge is not necessarily around the existing tools. We have seen an evolution of capabilities to solve these challenges, such as long-read sequencing, single cell gene expression profiling, chromatin structural analysis, and 3D genomic interactions. The technology is available for integration of these "-omics" data sets using predictive modeling and machine learning. It really comes down to sample procurement and data generation. Generating and understanding the data sets required is a significant next step for academia and industry. The ultimate ask is for a Dark Genome database akin to the Encode project. Ideally it would contain bulk and single

cell data sets from relevant primary disease models and normal cell systems. And due to the variation of genomes per sample, -seqs from the same sample would allow for an apples-to-apples comparison. By collating databases with standard sets of tools, common annotations, and data sets, we can transform our understanding of Dark Genome regulation, the way these databases transformed the coding genome.

David Ting: *What is the future of the Dark Genome?*

Although we have known about the Dark Genome since the early days of molecular biology, we are just now developing the tools to accurately map these sequences that make up the majority of our genome. For the most part biomedical research has focused on the exome, which is highly conserved between individuals. However, as we develop the next generation of precision biomedicines, the Dark Genome encodes significant genetic heterogeneity that underlies the broad phenotypic variation between individuals. Illuminating the genetic and molecular rules of the Dark Genome will improve patient selection in our use of current drugs and open new therapeutic avenues with refined precision.

Application of current drugs that we know target Dark Genome biology include epigenetic therapies like DNA hypomethylating agents used for cancers and reverse transcriptase inhibitors commonly used for viral therapies. These initial footholds into the druggable landscape of the Dark Genome provide a path to expand refined compounds that can be precisely tuned to target particular parts of repeat element biology. Moreover, the computational tools to accurately quantify the genetic, epigenetic, and transcriptomic patterns of the Dark Genome provide opportunities to better understand the effects of existing therapies and potentially unlock new indications for those assets. Finally, emerging scientific work has revealed the ability of the Dark Genome to infect other cells through extracellular vesicles [101], or resurrection of endogenous retroviruses [102], that could trigger antibody responses to the Dark Genome [18]. The building of these molecular maps and the deeper mechanistic understanding of repeat element biology will open a new dimension to explore the next generation of biomedicines to target the Dark Genome.

The final talk of the meeting was given by Rick Young where he described transcriptional regulatory elements, their protein and RNA components and the multiple points at which they could be leveraged for therapeutic development.

Conclusion

This inaugural symposium provided an overview of the breadth of human biology impacted by the Dark Genome. It highlighted the importance of technological and platform development to open the door to enhanced understanding. Finally, it touched upon the potential impact on human disease enabled through Dark Genome focused drug development.

Abbreviations

ADAR1	adenosine deaminase RNA specific-1
ALS	amyotrophic lateral sclerosis
Alu	Arthrobacter luteus restriction endonuclease sequence
cGAS	cyclic GMP-AMP Synthase
CH3	methyl group
CRE	cyclization recombinase
CSF	cerebrospinal fluid
DCs	Dendritic cells
DNMT1	DNA methyltransferase 1
DNMTi	DNA methyltransferases inhibitors
dsRNA	double stranded ribonucleic acid
Env	envelope
ERVs	endogenous retroviruses
GTEx	Genotype- SETDB1Tissue Expression
HDACi	histone deacetylase inhibitors
HERV-K/HML-2	human endogenous retrovirus-K
HERVs	Human endogenous retroviruses
hg38	Homo sapiens (human) genome assembly GRCh38
HLA	human leukocyte antigen
IFN-I	Type I Interferon
LINE (-1)	Long Interspersed nuclear element (-1)
LTRs	long terminal repeats
LUAD	lung adenocarcinoma
MAVS	mitochondrial antiviral signaling protein
MDA5	Melanoma Differentiation- Associated protein 5
MLV	murine leukaemia viruses
NRTI	nucleoside reverse transcriptase inhibitor
ORF	open reading frame
ORF1p	LINE-1 open reading frame 1 protein
p53/*Tp53*	Tumor Protein p53
Ph2	phase-2
PKR	Protein Kinase-R
RTE	reterotransposon
SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
shRNAs	short hairpin RNAs
STING	Stimulator of Interferon Genes
SVA	SINE-VNTR-*Alu*s
TAF1	TATA-box binding protein associated factor 1
TBXT	T-Box Transcription Factor T
TCGA	Cancer Genome Atlas
TPRT	target-primed reverse transcription
VLP	virus-like particles
XDP	X-linked dystonia parkinsonism

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 11 of 13

Josh Mandel-Brehm, CEO, Camp4 Therapeutics.
Robert Manguso, Assistant Professor of Medicine, Harvard Medical School.
Tomas Mustelin, University of Washington.
Rachel O'Neill, University of Connecticut.
Andy Satlin, CMO Transposon Therapeutics.
Dennis Zaller, CSO Rome Therapeutics.

Authors' contributions

J.B., K.B., K.C., M.C., J.M., D.D., J.D., B.G., G.K., S.K., A.L., A.N., S.P., D.R., J.M.S., D.T., wrote the sections pertaining to the talks they gave or panels they moderated (as attributed in the text). All authors reviewed the manuscript.

Authors' information

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have consented to publication

Competing interests

J.B., is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to ReOpen Diagnostics, LLC and serves or served on the Scientific Advisory Board of the following: Logomix, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Sangamo, Inc., Tessera Therapeutics, Inc. and the Wyss Institute. K.B., interests are managed by Dana-Farber Cancer Institute K.C., Consultant, Rome Therapeutics. M.C., Employee at Pfizer Center for Therapeutic Innovation and may own Pfizer stock. J.C., Member of the scientific advisory board, Rome Therapeutics D.D.D.C., received research funds from Pfizer and Nektar Therapeutics. He is a co-founder and shareholder of DNAMx, Inc J.D., Employee and owner of share options in Enara Bio B.G., has received honoraria for speaking engagements from Merck, Bristol Meyers Squibb, and Chugai Pharmaceuticals; has received research funding from Bristol Meyers Squibb and Merck; and has been a compensated consultant for Darwin Health, Merck, PMV Pharma and Rome Therapeutics of which he is a co-founder. G.K., is a scientific co-founder of Enara Bio and a member of its scientific advisory board. He has consulted for EnaraBio and Repertoire Immune Medicines S.K., Employee and owner of share options in Rome TherapeuticsA.L., PMV Pharma and Chugai Pharmaceuticals A.N., has a joint patent with GeNeuro on a monoclonal antibody to HML-2 S.P., Employee and owner of share options in Enara BioD.R., Employee of Paragon Therapeutics J.M.S., is a cofounder and SAB chair of Transposon Therapeutics, Inc. and consults for Atropos Therapeutics, Inc., Gilead Sciences, Inc. and Oncolinea Inc. D.T.T., is a founder, consultant, and has equity in ROME Therapeutics. He has received consulting fees from Sonata Therapeutics and Tekla Capital, is a founder and has equity in PanTher Therapeutics and TellBio, Inc., is on the advisory board for Improve Bio, Inc., has received honorariums from Moderna and Ikena Oncology, receives research support from ACD-Biotechne, AVA LifeScience GmbH, and Incyte Pharmaceuticals, which are not related to this work. His interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies.

References

1. Xia B, Zhang W, Wudzinska A, Huang E, Brosh R, Pour M et al. The genetic basis of tail-loss evolution in humans and apes. bioRxiv. 2021:2021.09.14.460388.
2. Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature. 1988;332(6160):164–6.
3. Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, Ardeljan D, et al. Structural variants caused by Alu insertions are associated with risks for many human diseases. Proc Natl Acad Sci U S A. 2017;114(20):E3984–92.
4. Payer LM, Steranka JP, Ardeljan D, Walker J, Fitzgerald KC, Calabresi PA, et al. Alu insertion variants alter mRNA splicing. Nucleic Acids Res. 2019;47(1):421–31.
5. Payer LM, Steranka JP, Kryatova MS, Grillo G, Lupien M, Rocha PP, et al. Alu insertion variants alter gene transcript levels. Genome Res. 2021;31(12):2236–48.
6. Rodic N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol. 2014;184(5):1280–6.
7. Taylor MS, Wu C, Fridy PC, Zhang SJ, Senussi Y, Wolters JC, et al. Ultrasensitive detection of circulating LINE-1 ORF1p as a specific multi-cancer biomarker. Cancer Discov. Epub 2023 Sep 12.
8. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020;52(3):306–19.
9. Ardeljan D, Steranka JP, Liu C, Li Z, Taylor MS, Payer LM, et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. Nat Struct Mol Biol. 2020;27(2):168–78.
10. Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet. 2001;27(1):31–9.
11. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76.
12. Yi L, Lu C, Hu W, Sun Y, Levine AJ. Multiple roles of p53-related pathways in somatic cell reprogramming and stem cell differentiation. Cancer Res. 2012;72(21):5635–45.
13. Nieto M, Samper E, Fraga MF, Gonzalez de Buitrago G, Esteller M, Serrano M. The absence of p53 is critical for the induction of apoptosis by 5-aza-2'-deoxycytidine. Oncogene. 2004;23(3):735–43.
14. Yi L, Sun Y, Levine A. Selected drugs that inhibit DNA methylation can preferentially kill p53 deficient cells. Oncotarget. 2014;5(19):8924–36.
15. Levine AJ. The p53 protein plays a central role in the mechanism of action of epigenic drugs that alter the methylation of cytosine residues in DNA. Oncotarget. 2017;8(5):7228–30.
16. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature. 2019;565(7739):312–7.
17. Burn A, Roy F, Freeman M, Coffin JM. Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues. PLoS Biol. 2022;20(10):e3001826.
18. Ng KW, Boumelha J, Enfield KSS, Almagro J, Cha H, Pich O, et al. Antibodies against endogenous retroviruses promote Lung cancer immunotherapy. Nature. 2023;616(7957):563–73.
19. Stansly PG. Non-oncogenic infectious agents associated with experimental tumors. Prog Exp Tumor Res. 1965;7:224–58.
20. Gardner MB. Retroviruses and wild mice: an historical and personal perspective. Adv Cancer Res. 1994;65:169–201.
21. Paquette Y, Hanna Z, Savard P, Brousseau R, Robitaille Y, Jolicoeur P. Retrovirus-induced murine motor neuron disease: mapping the determinant of spongiform degeneration within the envelope gene. Proc Natl Acad Sci U S A. 1989;86(10):3896–900.
22. DesGroseillers L, Rassart E, Robitaille Y, Jolicoeur P. Retrovirus-induced spongiform encephalopathy: the 3'-end long terminal repeat-containing viral sequences influence the incidence of the disease and the specificity of the neurological syndrome. Proc Natl Acad Sci U S A. 1985;82(24):8818–22.

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 12 of 13

23. Cardona SM, Dunphy JM, Das AS, Lynch CR, Lynch WP. Astrocyte infection is required for retrovirus-induced spongiform neurodegeneration despite suppressed viral protein expression. Front Neurosci. 2019;13: 1166.

24. Viola MV, Frazier M, White L, Brody J, Spiegelman S. RNA-instructed DNA polymerase activity in a cytoplasmic particulate fraction in brains from Guamanian patients. J Exp Med. 1975;142(2):483–94.

25. McCormick AL, Brown RH Jr, Cudkowicz ME, Al-Chalabi A, Garson JA. Quantification of reverse transcriptase in ALS and elimination of a novel retroviral candidate. Neurology. 2008;70(4):278–83.

26. Andrews WD, Tuke PW, Al-Chalabi A, Gaudin P, Ijaz S, Parton MJ, et al. Detection of reverse transcriptase activity in the serum of patients with motor neurone disease. J Med Virol. 2000;61(4):527–32.

27. Steele AJ, Al-Chalabi A, Ferrante K, Cudkowicz ME, Brown RH Jr, Garson JA. Detection of serum reverse transcriptase activity in patients with ALS and unaffected blood relatives. Neurology. 2005;64(3):454–8.

28. Bowen LN, Tyagi R, Li W, Alfahad T, Smith B, Wright M, et al. HIV-associated motor neuron disease: HERV-K activation and response to antiretroviral therapy. Neurology. 2016;87(17):1756–62.

29. Alfahad T, Nath A. Retroviruses and amyotrophic lateral sclerosis. Antiviral Res. 2013;99(2):180–7.

30. Li W, Lee MH, Henderson L, Tyagi R, Bachani M, Steiner J, et al. Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med. 2015;7(307):307ra153.

31. Phan K, He Y, Fu Y, Dzamko N, Bhatia S, Gold J, et al. Pathological manifestation of human endogenous retrovirus K in frontotemporal dementia. Commun Med (Lond). 2021;1:60.

32. Douville R, Liu J, Rothstein J, Nath A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. Ann Neurol. 2011;69(1):141–51.

33. Steiner JP, Bachani M, Malik N, DeMarino C, Li W, Sampson K, et al. Human endogenous retrovirus K envelope in spinal fluid of amyotrophic lateral sclerosis is toxic. Ann Neurol. 2022;92(4):545–61.

34. Arru G, Mameli G, Deiana GA, Rassu AL, Piredda R, Sechi E, et al. Humoral immunity response to human endogenous retroviruses K/W differentiates between amyotrophic lateral sclerosis and other neurological diseases. Eur J Neurol. 2018;25(8):1076-e1084.

35. Contreras-Galindo R, Dube D, Fujinaga K, Kaplan MH, Markovitz DM. Susceptibility of human endogenous retrovirus type K to reverse transcriptase inhibitors. J Virol. 2017;91(23):e01309.

36. Tyagi R, Li W, Parades D, Bianchet MA, Nath A. Inhibition of human endogenous retrovirus-K by antiretroviral drugs. Retrovirology. 2017;14(1):21.

37. Baldwin ET, Gotte M, Tchesnokov EP, Arnold E, Hagel M, Nichols C, et al. Human endogenous retrovirus-K (HERV-K) reverse transcriptase (RT) structure and biochemistry reveals remarkable similarities to HIV-1 RT and opportunities for HERV-K-specific inhibition. Proc Natl Acad Sci U S A. 2022;119(27):e2200260119.

38. Garcia-Montojo M, Fathi S, Norato G, Smith BR, Rowe DB, Kiernan MC, et al. Inhibition of HERV-K (HML-2) in amyotrophic lateral sclerosis patients on antiretroviral therapy. J Neurol Sci. 2021;423:117358.

39. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2023;186(2):243–78.

40. Booth LN, Brunet A. The aging epigenome. Mol Cell. 2016;62(5):728–44.

41. Pal S, Tyler JK. Epigenetics and aging. Sci Adv. 2016;2(7):e1600584.

42. Sen P, Shah PP, Nativio R, Berger SL. Epigenetic mechanisms of longevity and aging. Cell. 2016;166(4):822–39.

43. Gorgoulis V, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, et al. Cellular senescence: defining a path forward. Cell. 2019;179(4):813–27.

44. Childs BG, Gluscevic M, Baker DJ, Laberge RM, Marquess D, Dananberg J, et al. Senescent cells: an emerging target for diseases of ageing. Nat Rev Drug Discov. 2017;16(10):718–35.

45. Coppe JP, Patil CK, Rodier F, Sun Y, Munoz DP, Goldstein J, et al. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS Biol. 2008;6(12):2853–68.

46. Pignolo RJ, Passos JF, Khosla S, Tchkonia T, Kirkland JL. Reducing senescent cell burden in aging and disease. Trends Mol Med. 2020;26(7):630–8.

47. De Cecco M, Criscione SW, Peckham EJ, Hillenmeyer S, Hamm EA, Manivannan J, et al. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell. 2013;12(2):247–56.

48. Guo C, Jeong HH, Hsieh YC, Klein HU, Bennett DA, De Jager PL, et al. Tau activates transposable elements in Alzheimer's Disease. Cell Rep. 2018;23(10):2874–80.

49. Sun W, Samimi H, Gamez M, Zare H, Frost B. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. Nat Neurosci. 2018;21(8):1038–48.

50. Tam OH, Rozhkov NV, Shaw R, Kim D, Hubbard I, Fennessey S, et al. Postmortem cortex samples identify distinct molecular subtypes of ALS: retrotransposon activation, oxidative stress, and activated glia. Cell Rep. 2019;29(5):1164–77.

51. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460(7259):1127–31.

52. Erwin JA, Paquola AC, Singer T, Gallina I, Novotny M, Quayle C, et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. Nat Neurosci. 2016;19(12):1583–91.

53. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell. 2012;151(3):483–96.

54. Evrony GD, Lee E, Park PJ, Walsh CA. Resolving rates of mutation in the brain using single-neuron genomics. Elife. 2016;5:5.

55. Zhao B, Wu Q, Ye AY, Guo J, Zheng X, Yang X, et al. Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. PLoS Genet. 2019;15(4):e1008043.

56. Ardeljan D, Taylor MS, Ting DT, Burns KH. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. Clin Chem. 2017;63(4):816–22.

57. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al. Landscape of somatic retrotransposition in human cancers. Science. 2012;337(6097):967–71.

58. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. 2014;24(7):1053–63.

59. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, et al. Mobile DNA in cancer: extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science. 2014;345(6196):1251343.

60. Burns KH. Our conflict with transposable elements and its implications for human disease. Annu Rev Pathol. 2020;15:51–70.

61. Mir AA, Philippe C, Cristofari G. euL1db: the European database of L1HS retrotransposon insertions in humans. Nucleic Acids Res. 2015;43(Database issue):D43-47.

62. Hartmann G. Nucleic acid immunity. Adv Immunol. 2017;133:121–69.

63. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature. 2019;566(7742):73–8.

64. Simon M, Van Meter M, Ablaeva J, Ke Z, Gonzalez RS, Taguchi T, et al. LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation. Cell Metab. 2019;29(4):871-85 e5.

65. Thomas CA, Tejwani L, Trujillo CA, Negraes PD, Herai RH, Mesci P, et al. Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. Cell Stem Cell. 2017;21(3):319-331.e8.

66. Mita P, Sun X, Fenyo D, Kahler DJ, Li D, Agmon N, et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. Nat Struct Mol Biol. 2020;27(2):179–91.

67. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription *in vitro*. EMBO J. 2002;21(21):5899–910.

68. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol. 2006;13(7):655–60.

69. Britten RJ, Kohne DE. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science. 1968;161(3841):529–40.

70. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science. 2011;331(6017):593–6.

Boeke *et al. Mobile DNA*   (2023) 14:18

Page 13 of 13

71. Solovyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, et al. Global cancer transcriptome quantifies repeat element polarization between immunotherapy responsive and T Cell suppressive classes. Cell Rep. 2018;23(2):512–21.

72. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Author Correction: pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2023;55(6):1080.

73. Solovyov A, Behr JM, Hoyos D, Banks E, Drong AW, Zhong JZ, et al. Mechanism-guided quantification of LINE-1 reveals p53 regulation of both retrotransposition and transcription. bioRxiv. 2023:2023.05.11.539471.

74. Panda A, de Cubas AA, Stein M, Riedlinger G, Kra J, Mayer T, et al. Endogenous retrovirus expression is associated with response to immune checkpoint blockade in clear cell renal cell carcinoma. JCI Insight. 2018;3(16):e121522.

75. Šulc P, Gioacchino AD, Solovyov A, Marhon SA, Sun S, Lindholm HT, et al. Repeats mimic pathogen-associated patterns across a vast evolutionary landscape. bioRxiv. 2023:2021.11.04.467016.

76. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.

77. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304–51.

78. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44–53.

79. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an Interferon response in cancer via dsRNA including endogenous retroviruses. Cell. 2015;162(5):974–86.

80. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-Demethylating agents Target Colorectal Cancer cells by inducing viral mimicry by endogenous transcripts. Cell. 2015;162(5):961–73.

81. Mehdipour P, Marhon SA, Ettayebi I, Chakravarthy A, Hosseini A, Wang Y, et al. Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. Nature. 2020;588(7836):169–73.

82. Chen R, Ishak CA, De Carvalho DD. Endogenous retroelements and the viral mimicry response in cancer therapy and cellular homeostasis. Cancer Discov. 2021;11(11):2707–25.

83. Lindholm HT, Chen R, De Carvalho DD. Endogenous retroelements as alarms for disruptions to cellular homeostasis. Trends Cancer. 2023;9(1):55–68.

84. Li H, Chiappinelli KB, Guzzetta AA, Easwaran H, Yen RW, Vatapalli R, et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. Oncotarget. 2014;5(3):587–98.

85. Moufarrij S, Srivastava A, Gomez S, Hadley M, Palmer E, Austin PT, et al. Combining DNMT and HDAC6 inhibitors increases anti-tumor immune signaling and decreases tumor burden in ovarian cancer. Sci Rep. 2020;10(1):3470.

86. Gomez S, Cox OL, Walker RR 3rd, Rentia U, Hadley M, Arthofer E, et al. Inhibiting DNA methylation and RNA editing upregulates immunogenic RNA to transform the tumor microenvironment and prolong survival in ovarian cancer. J Immunother Cancer. 2022;10(11):e004974.

87. Ghoneim HE, Fan Y, Moustaki A, Abdelsamed HA, Dash P, Dogra P, et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. Cell. 2017;170(1):142-157.e19.

88. Loo Yau H, Bell E, Ettayebi I, de Almeida FC, Boukhaled GM, Shen SY, et al. DNA hypomethylating agents increase activation and cytolytic activity of CD8(+) T cells. Mol Cell. 2021;81(7):1469-83e8.

89. Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nat Commun. 2019;10(1):5228.

90. Xia Z, Cochrane DR, Tessier-Cloutier B, Leung S, Karnezis AN, Cheng AS, et al. Expression of L1 retrotransposon open reading frame protein 1 in gynecologic cancers. Hum Pathol. 2019;92:39–47.

91. Pisanic TR 2nd, Asaka S, Lin SF, Yen TT, Sun H, Bahadirli-Talbott A, et al. Long Interspersed Nuclear element 1 retrotransposons become deregulated during the development of ovarian cancer precursor lesions. Am J Pathol. 2019;189(3):513–20.

92. Strissel PL, Ruebner M, Thiel F, Wachter D, Ekici AB, Wolf F, et al. Reactivation of codogenic endogenous retroviral (ERV) envelope genes in human endometrial carcinoma and prestages: emergence of new molecular targets. Oncotarget. 2012;3(10):1204–19.

93. Rycaj K, Plummer JB, Yin B, Li M, Garza J, Radvanyi L, et al. Cytotoxicity of human endogenous retrovirus K-specific T cells toward autologous Ovarian cancer cells. Clin Cancer Res. 2015;21(2):471–83.

94. Zhou F, Krishnamurthy J, Wei Y, Li M, Hunt K, Johanning GL, et al. Chimeric antigen receptor T cells targeting HERV-K inhibit breast cancer and its metastasis through downregulation of Ras. Oncoimmunology. 2015;4(11): e1047582.

95. Griffin GK, Wu J, Iracheta-Vellve A, Patti JC, Hsu J, Davis T, et al. Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. Nature. 2021;595(7866):309–14.

96. Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet. 2017;49(7):1052–60.

97. Saini SK, Orskov AD, Bjerregaard AM, Unnikrishnan A, Holmberg-Thyden S, Borch A, et al. Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. Nat Commun. 2020;11(1):5660.

98. Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet F, et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat Biotechnol. 2022;40(2):209–17.

99. Ruiz Cuevas MV, Hardy MP, Holly J, Bonneil E, Durette C, Courcelles M, et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. Cell Rep. 2021;34(10):108815.

100. Zhang W, Golynker I, Brosh R, Wudzinska AM, Zhu Y, Carrau L, et al. Mouse genomic rewriting and tailoring: synthetic *Trp53* and humanized *ACE2*. bioRxiv. 2022:2022.06.22.495814.

101. Porter RL, Sun S, Flores MN, Berzolla E, You E, Phillips IE, et al. Satellite repeat RNA expression in epithelial ovarian cancer associates with a tumor-immunosuppressive phenotype. J Clin Invest. 2022;132(16):e155931.

102. Liu X, Liu Z, Wu Z, Ren J, Fan Y, Sun L, et al. Resurrection of endogenous retroviruses during aging reinforces senescence. Cell. 2023;186(2):287-304e26.

Publisher's Note