

Challenges in LncRNA Biology: Views and Opinions

Donald A. Adjeroh ^{1,*}, Xiaobo Zhou ², Alexandre Rossi Paschoal ^{3,4} , Nadya Dimitrova ⁵, Ekaterina G. Derevyanchuk ⁶, Tatiana P. Shkurat ⁶, Jeffrey A. Loeb ⁷, Ivan Martinez ⁸ and Leonard Lipovich ^{9,10,11,*}

- ¹ Lane Department of Computer Science and Electrical Engineering, West Virginia University (WVU), Morgantown, WV 26506, USA
- ² Department of Bioinformatics and Systems Medicine, University of Texas Health Science Center, Houston, TX 77030, USA; xiaobo.zhou@uth.tmc.edu
- ³ Department of Computer Science, Bioinformatics and Pattern Recognition Group, Federal University of Technology—Paraná—UTFPR, Curitiba 86300-000, Brazil; paschoal@utfpr.edu.br
- ⁴ Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot OX11 0FA, UK
- ⁵ Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA; nadya.dimitrova@yale.edu
- ⁶ Department of Genetics, Southern Federal University, Rostov-on-Don 344090, Russia; derevyanchuk@sfedu.ru (E.G.D.); tshkurat@sfedu.ru (T.P.S.)
- ⁷ Department of Neurology and Rehabilitation, The Center for Clinical and Translational Science, The University of Illinois NeuroRepository, University of Illinois, Chicago, IL 60607, USA; jalob@uic.edu
- ⁸ Department of Microbiology, Immunology & Cell Biology, WVU Cancer Institute, West Virginia University (WVU) School of Medicine, Morgantown, WV 26505, USA; ivmartinez@hsc.wvu.edu
- ⁹ Shenzhen Huayuan Biological Science Research Institute, Shenzhen Huayuan Biotechnology Co., Ltd., Shenzhen 518000, China
- ¹⁰ Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, Detroit, MI 48201, USA
- ¹¹ College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou 325060, China
- * Correspondence: donald.adjeroh@mail.wvu.edu (D.A.A.); llipovich@med.wayne.edu (L.L.)



Citation: Adjeroh, D.A.; Zhou, X.; Paschoal, A.R.; Dimitrova, N.; Derevyanchuk, E.G.; Shkurat, T.P.; Loeb, J.A.; Martinez, I.; Lipovich, L. Challenges in LncRNA Biology: Views and Opinions. *Non-Coding RNA* **2024**, *10*, 43. <https://doi.org/10.3390/ncrna10040043>

Academic Editors: Leandro Castellano, Eleonora Leucci and Suresh K Alahari

Received: 2 March 2024
Revised: 26 June 2024
Accepted: 4 July 2024
Published: 1 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: This is a mini-review capturing the views and opinions of selected participants at the 2021 IEEE BIBM 3rd Annual LncRNA Workshop, held in Dubai, UAE. The views and opinions are expressed on five broad themes related to problems in lncRNA, namely, challenges in the computational analysis of lncRNAs, lncRNAs and cancer, lncRNAs in sports, lncRNAs and COVID-19, and lncRNAs in human brain activity.

Keywords: lncRNAs; cancer; COVID-19; lncRNA datasets; deep learning; large language models

1. Introduction

This mini-review collates selected views and opinions on certain topics in the area of long non-coding RNA (lncRNA) biology, bioinformatics, genomics, and therapeutics, from the perspectives of several keynote, plenary, and invited speakers, panelists, and organizers of the IEEE BIBM 3rd Annual LncRNA Workshop, which was held in Dubai, UAE, in December 2021, with both remote and in-person participation. As a mini-review, this article is not intended to be either comprehensive or complete, nor does it cover all the important issues surrounding today's lncRNA research. Contributors were asked to write briefly on a topic of their choice as it related to their presentations or the discussions at the Workshop. This mini-review touches on five broad themes chosen for their relevance and appeal to the lncRNA community: challenges in the computational analysis of lncRNAs, lncRNAs and cancer, lncRNAs in sports, lncRNAs and COVID-19, and lncRNAs in human brain activity.

Given the focus on discussion and presentations at the 2021 Workshop, we acknowledge the limited scope of this mini-review, as it could not cover various equally important topics related to lncRNA—for instance, the accurate prediction of lncRNA targets and

functions; tissue, cell type, and context specificity of lncRNAs and their roles in development, differentiation, and tissue homeostasis; roles of lncRNAs in cell communication and immune response; lncRNA function in the brain beyond epilepsy; the relationship of the lncRNAome to the conundrum of bifunctional RNAs; and other themes.

The views and opinions expressed are those of the authors and do not necessarily represent those of the IEEE or BIBM. Not all invited speakers contributed to this mini-review.

2. Current Challenges in the Computational Analysis of lncRNAs

Alexandre Rossi Paschoal, Federal University of Technology—Paraná—UTFPR, Brazil; Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, UK, Xiaobo Zhou, University of Texas Health Science Center, Houston, TX, USA, Donald A. Adjeroh, West Virginia University (WVU), Morgantown, WV, USA.

Long non-coding RNA (lncRNA) is arbitrarily defined as a transcript greater than 200 nucleotides (nt) with low or no potential to encode a protein. These lncRNA genes may reside in regions of the genome between protein-coding genes, but they may also overlap protein-coding regions, which results in several lncRNA biotypes, including intergenic, antisense, and intronic lncRNAs. According to GENCODE version 44 (v44) website statistics (www.genencodegenes.org/human/stats.html (accessed on 28 September 2023)), 31.8% (19,922/62,700) of the total genes annotated are lncRNA genes, while 23.0% (58,246/252,835) of the annotated transcripts are lncRNA loci transcripts. In contrast, just a tiny fraction of these is experimentally or manually curated and available and, in general, this is typically obtained from model species. These observations can tremendously impact lncRNA research. For that, we provide some discussion that may help researchers reflect on the current challenges in long non-coding RNA research. We point out open challenges crucial for the mathematical and computational modeling and analysis of lncRNAs and their roles in the cell. Some of these challenges require immediate attention, while others may require longer-term research to address. All the same, the goal is to raise the curiosity of the lncRNA community about these problems and to call attention to potential directions toward their resolution. We organized our discussion on the computational challenges into three broad groups: RNA biology, data challenges, and machine learning in lncRNA analysis.

2.1. RNA Biology: lncRNAs and Their Functions, ORFs, and Micropeptides

Long non-coding RNAs are essential players in many cellular processes, from regular development to oncogenic transformation. Figure 1 provides a schematic view of the functions of lncRNAs. The rapid growth of genome-wide translation profiling and ribosome profiling has revealed that a number of small open reading frames (sORFs) within ncRNAs actually have peptide- or protein-coding potential [1,2]. The peptides or short proteins encoded by transcripts annotated as lncRNAs have been shown to be critical players in cancer development and progression. For example, a conserved 53 aa peptide encoded by the putative lncRNA HOXB-AS3 suppresses colon cancer growth by regulating the alternative splicing of pyruvate kinase M (PKM) and tumor metabolic reprogramming [1,3]. FBXW7-185aa is encoded by the circRNA FBXW7 and inhibits glioma growth [4]. Moreover, miPEP-200a and miPEP-200b are encoded by miR-200a and miR-200b, and respectively, inhibit the migration of prostate cancer cells via the suppression of the process of epithelial-to-mesenchymal transition [5].

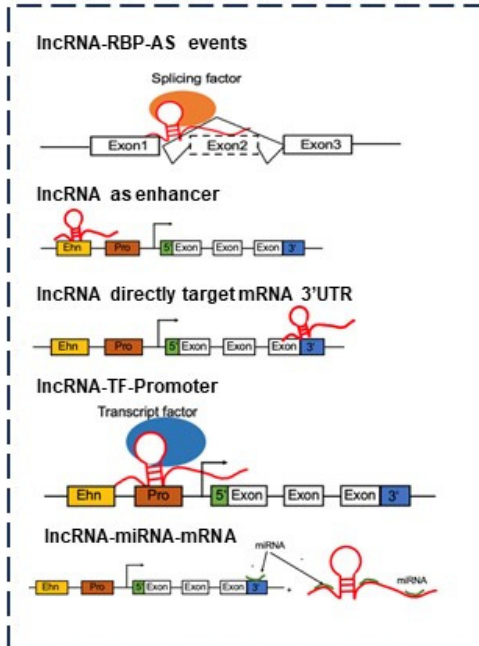
The addiction to ORF annotation in “non-coding” biology. A recent study discussed the feature extraction approaches for biological sequences using different mathematical formulations [6]. In particular, the paper provided detailed mapping of the timeline on various tools for lncRNA identification (see Figure 1 in [6]). Of these tools, only 3 out of 21 (14.2%) did not use open reading frame (ORF) predictions or annotations in their methods. The conclusion was that most current approaches to lncRNA analysis are, paradoxically, still highly dependent on ORF information. How would such tools that highly depend on ORF information handle lncRNAs without ORF information? With the rise in awareness of the frequent incidence of small ORFs (sORFs or smORFs) in lncRNAs, what will the

behavior of these tools be? How do they and how should they handle sORFs inside lncRNA genes? These are some open questions that the field will face imminently. One essential suggestion could be to consider alignment-free methods, using tools such as BASiNET [7], an alignment-free tool that uses network-based features to analyze biological sequences.

A. Multifunctional polymers



B. As RNA molecules



C. Encoding peptides

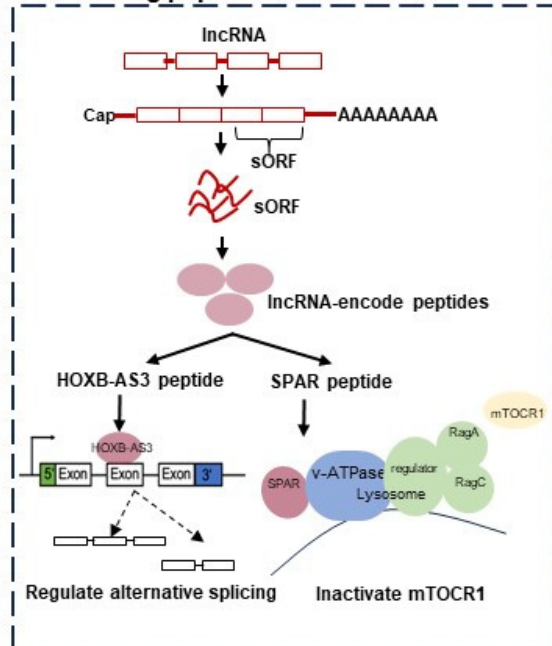


Figure 1. The functions of lncRNAs can be considered from two perspectives, namely, (i) from the viewpoint of RNA function alone and (ii) in terms of encoding peptides or proteins. (A) The multifunctional molecular complexes containing lncRNAs, including lncRNA–protein, lncRNA–DNA, lncRNA–mRNA, and lncRNA–miRNA. (B) The functions of lncRNAs as RNA molecules: lncRNAs regulate target genes by functioning as enhancers, lncRNAs regulate target genes by recruiting transcription factors, lncRNAs regulate alternative splicing events by interacting with splicing factors, and lncRNAs regulate mRNA by targeting the 3′-UTR region. (C) The functions of lncRNAs via encoding peptides or proteins able to regulate alternative splicing (e.g., HOXB-AS3 peptide) and gene expression (e.g., SPAR peptide).

Coding or non-coding RNA: What is a gene? The definition of a gene has been substantially updated since the Central Dogma of Molecular Biology proposed by Watson and Crick. Researchers of the first post-genomic century have been deliberating over the increasingly complex question of what are coding and non-coding genes. This scenario brings one particular challenge, namely, when is the gene actually coding for a protein? For example, if a gene has 12 isoforms, of which some are coding and some are non-coding transcripts, should it be considered a coding or a non-coding gene? This is the case for the gene KCN4-TEX40 (ID: ENSG00000257069), which has six transcripts annotated as protein-coding (including one with its CDS not defined), four annotated as “RNA with retained introns”, one as a product of nonsense-mediated decay, and a last one as an lncRNA. The question is the following: Do the isoforms not matter? This issue is not trivial, and also relates to the problem of bifunctional RNAs [8–10] where certain RNAs

biologically exhibit both a protein-coding capacity and non-coding functional properties, as well as to complex transcriptional units such as SRA1, where a combination of alternative splicing and promoter choice lead to the production of distinct coding and non-coding transcripts, with different functions, from a single gene locus.

Small in size, but a bigger challenge. On a related challenge, there have been discussions about small, short, or micropeptides, or small ORFs. These peptides are derived from a myriad of genomic locations including coding, non-coding, intergenic, and untranslated regions (UTRs) with upstream or downstream coding gene regions, unveiling the remarkable complexity of cells [11]. Principally, the identification of lncRNA-encoded small peptides confronts the basic molecular biological concepts and brings up the question of whether or not these micropeptides are just noise. There are indications that some of these may not be just noise. A small peptide of 53 aa encoded by the lncRNA HOXB-AS3 binds to an mRNA and inhibits the formation of PKM2 to induce a tumor-suppressive effect [12] (see also Figure 1).

2.2. The Data Challenge in LncRNA Research: Datasets, Availability, Beyond Human Data

Datasets. In the era of data-hungry machine learning and other artificial intelligence (AI) techniques, data have become a premium in successful applications of these approaches in a given domain. Not surprisingly, the availability of data and datasets has also become an urgent challenge in the computational analysis of lncRNAs and their functions. Several datasets have become available on lncRNAs. For instance, RNALocate v2.0 [13] contains RNA subcellular localization entries validated by experimental evidence. The database contains information on lncRNAs located in the nucleus, cytoplasm, ribosome, exosome, nucleoplasm, chromatin, cytosol, endoplasmic reticulum, and plasma membrane. LncAtlas [14], a key database for the subcellular localization of lncRNAs, provides a cytoplasmic/nuclear relative concentration index (CN-RCI), derived from GENCODE (Ensembl) RNA-Seq measurements for lncRNAs and mRNAs across 15 cell lines.

From the emerging importance of ncRNAs in cancer, studying regulations related to ncRNAs and their coding potential is essential to fully understanding cancer progression and identifying therapeutic targets. Several databases are now available on cancer-related studies providing annotations for interaction and ncRNA–disease associations. Examples of these include RNACentral [15], NONCODE [16], NoncoRNA [17], NPInter [18], ViR-Base [19], MNDR [20], EVAAtlas [21], ncRNA-eQTL [22], and NSDNA [23]. Some datasets such as ncEP [24], FuncPEP [25], and SPENCER [26] only contain experimentally validated and functionally characterized ncRNA-encoded peptides. lncRNAfunc [27] is a newly developed knowledgebase focusing on the annotation of lncRNA function in human cancer. TransLnc [28] provides the potential peptides encoded by lncRNAs, but this database only includes lncRNAs and lacks circRNAs and miRNAs, which are also required to fully understand ncRNAs' coding potential of action in cancer. Recently, a manually curated dataset of bifunctional RNAs was introduced [10].

To date, systematic annotations of coding potentials for ncRNAs have been unavailable. There is also an urgent need to develop a comprehensive resource for translatable ncRNAs that extends neoantigens for investigating the translation capacity of ncRNAs and for expanding investigations on the cancer immunopeptidome. Similar comprehensive resources are also required for applications of lncRNAs and other ncRNAs outside of cancer. But other challenges still abound.

Benchmarking, standardized nomenclature, and discrepancies. As noted by the creators of the Extensive de novo TE Annotator (EDTA) in [29], there is a significant lack of benchmarking and standardized studies to help to organize the diversity of studies and datasets in lncRNA research. We have the HUGO Gene Nomenclature Committee [30] and similar reports [31,32], for example. All these initiatives are crucial for biological research; however, they have focused mainly on humans. Science is much more than the human genome. In plants, there are disparate lncRNA datasets, with differences in their annotations. A case in point is the difference in lncRNA annotations between two popular

datasets CANTATAdb [33,34] and GreeNC [35,36]. What is the overlap among these available databases? The lack of experimental data raises the need for a careful analysis of potential discrepancies and to possibly arrive at a consensus across different datasets.

Big data age does not always mean good data availability. By 2019, there were over 230 databases dedicated only to ncRNAs [37,38]. For miRNA, the most studied ncRNA, only about 13% of the databases made their FASTA sequences or similar files available, and for lncRNAs, it was even a tinier fraction [37,38]. The scientific community has to wake up to these unbelievable statistics because information is needed to build novel and cutting-edge methods. Having access to data is fundamental to improving in silico research. This seems obvious, but it is not always the case. This challenge brings one more reflection that less is actually more. From a supervised machine learning point of view, a higher quality but smaller number of experimentally validated data samples is perhaps better than big data without validated data or curated labels. We also do concede that, when one considers certain machine learning models, such as self-supervised learning, large unannotated datasets can still be valuable for learning important representations from the data. A related challenge is about how the scientific community makes data available. This is perhaps a more general problem, beyond lncRNAs and genomics research. There are cases where a dataset was published but not updated; even worse, the dataset may have been discontinued and made no longer available, without providing access to the old versions of the dataset. This has been observed even for datasets published as part of some papers in relatively high-quality (Q1) journals. Another aspect is that, oftentimes, data are available as supplementary material and are often incomplete, for instance, without complete sequence information or other relevant attributes. These data need to be easier to obtain. Making data available in several formats (e.g., FASTA, GFF, BED, TSV), using an uncomplicated download protocol is the way to go. Some journals and conferences are already moving toward this requirement.

Beyond lncRNA data from humans and other model organisms. The pandemic created an awareness and significantly increased our consciousness with respect to the importance of using data from non-model species in important biomedical investigations. Whether such data are coming from a bat or from a pangolin does not matter. What really matters is that only focusing on humans or mammals limits our analysis and perspective to just a tiny fraction of biodiversity. Only by deeply exploring other species can we prepare ourselves for future pandemics and novel drug development. A well-explored model plant is *Arabidopsis thaliana*, whose genome, for example, is quite comparatively different from the coffee genome. For lncRNAs, and for ncRNAs or other types of genomic research for that matter, there is an urgent need to consider information and genomic data from other species, beyond the human genome, or the standard model species.

2.3. Modern Machine Learning in lncRNA Research: Deep Learning and LLMs

We consider there to be two problems in the application of artificial intelligence (AI) techniques in lncRNA analysis, specifically the use of deep learning models in lncRNA localization and the use of large language models (LLMs) in studying lncRNAs.

lncRNA subcellular localization: More challenging than it appears. There has been vigorous research on computational approaches to the problem of lncRNA identification [39–41] and on the issue of mRNA localization [42–45]. lncRNA subcellular localization has also attracted some recent attention [46], and various machine learning methods have been proposed [47–49]. Some of these methods reported relatively high performance, for instance, relatively high accuracy, and in some cases, above 90%. However, a closer analysis showed that most of these methods had problems in their data, for instance, working with significantly imbalanced data, where the machine learning model was basically memorizing the majority class. In some others, there was observed significant leakage in the data, whereby there was leakage of data from the training to the validation sets or to the test sets. In some other cases, some methods tended to directly or indirectly ignore or eliminate the difficult cases (for instance, based on the thresholds used to declare the

localization, say nuclear versus cytoplasmic localization, for the case of binary localization). The result was that the true performance of these methods was typically in the 50% to 65% range (for binary nuclear versus cytoplasmic localization), generally much lower than some of the reported performance levels. However, the results for these methods demonstrated that the computational analysis of lncRNA localization, be it for binary class or multi-class localization, is still a critical challenge, especially using only sequence information. This thus calls for more careful benchmarking of existing machine learning methods for this problem. For fairness, such benchmarking will have to be performed on the same dataset across different methods and on the same or comparable hardware configurations. (This obviously relates to the challenge of the datasets introduced previously). Beyond benchmarking, there is also need for more specialized machine learning and deep learning techniques that can exploit the special nature and characteristics of lncRNAs, including information on their sequences, structures, conformations, and types. Clearly, this need for more sophisticated deep learning algorithms also applies to the general area of lncRNA and ncRNA analysis, and not just on lncRNA localization.

Large Language Models (LLMs) for lncRNAs. Current lncRNA detection methods are mainly designed based on transcriptome sequencing data. Language models such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-training (GPT) have received an extensive attention recently because of the good performance of the transformer involved. They have been applied in various directions and fields. Recently, scientists found that some sORFs were also derived from lncRNAs. sORF-encoded peptides (SEPs) with a length of less than 100 amino acids were found to be involved in cancer suppression and cellular metabolism. Feng et. al. developed LncCat [50], an ORF attention model for identifying lncRNAs using a natural language model, which showed an impressive performance. This model could be improved by pre-training and fine-tuning the BERT and GPT models. Other models such as the BART model [51] could be applied to this problem too. We could potentially identify a lot of more lncRNAs with these latest approaches. Along the lines of the improved techniques mentioned earlier, a similar case can also be made beyond methods for improved lncRNA identification, but for the case of developing specialized LLMs for general analysis and understanding lncRNAs and their functions.

2.4. Looking to the Future: Toward Improved Research and Applications

Two aspects are helping to provide better lncRNA annotation and, consequently, discover better lncRNA functional information. The first is developments in long-read sequencing technology which will help to sequence full-length transcripts and lncRNAs. The second is the deep sequencing of genomes (for example, the Genome 10K Project [52]; 100,000 Genomes— in Africa, for Africa [53]; 100,000 Genome Project in the UK [54]; and 10KP: 10,000 Plant Genomes Project [55]). These are examples of how the large-scale availability of samples in the same species could help to comprehend the diversity, conservation, and variability among lncRNAs. In the same way, single-cell sequencing will help to look at these data as individual cells, enabling cell-to-cell communication. Further, the increasing availability of information about chromatin packaging and accessibility, for instance, from assays for transposase-accessible chromatin with sequencing (ATAC-Seq), could also open up another vista for the improved computational analysis of lncRNAs. Combining insights from ATAC-Seq with information from RNA-Seq could lead to new breakthroughs in the machine understanding of the functions and activities of ncRNAs in general, and lncRNAs in particular. These technologies coupled with improvements and new developments in computational methods (both hardware (e.g., processor architectures) and software (e.g., AI)) will improve future research on lncRNAs and create new applications and possibilities like never before seen.

3. Challenges and Rewards of Modelling Cancer-Associated Long Non-Coding RNAs In Vivo

Nadya Dimitrova, Yale University, New Haven, CT, USA

Cancer initiation and progression to a deadly and metastatic disease is a multi-step process that involves the gradual acquisition of numerous genetic and epigenetic aberrations of cellular effectors, involved in proliferation, survival, differentiation, and cell-to-cell interactions [56]. The discovery that mammalian genomes encode thousands of long non-coding RNAs (lncRNAs) has spurred investigation into whether the non-coding genome contains an untapped reservoir of novel modulators or perhaps even novel effectors of cancer processes [57,58]. To date, the in vivo evidence that lncRNAs are themselves effectors in cancer is sparse. On the other hand, the potential for lncRNAs to act as modulators of key nodes in cancer pathways is steadily gathering experimental support, including in organismal models [59]. Here, we focus on how the dysregulation of lncRNAs can rewrite cellular pathways, creating points of vulnerability exploited during cancer evolution. We then describe how genetically engineered mouse models (GEMMs) of cancer can be carefully harnessed to define the scope of lncRNAs' contribution in cancer.

While lncRNAs are rarely mutated in cancer, comparative analyses of normal and tumoral tissues have revealed extensive cancer-linked alterations in lncRNA expression patterns [60]. Global gene expression analyses have highlighted numerous differentially expressed lncRNAs and, in many cases, robustly linked differential lncRNA expression patterns with cancer types and clinical outcomes. For example, the overexpression of the lncRNAs metastasis-associated lung adenocarcinoma transcript 1 (*Malat1*) and Hox transcript antisense RNA (*HOTAIR*) has been strongly associated with metastatic lung and breast adenocarcinomas [61,62]. In some cases, genomic analyses of non-coding regions have provided a genetic mechanism for altered lncRNA expression, such as the somatic copy number variations (SCNVs) in plasmacytoma variant translocation 1 (*PVT1*) [63] and focally amplified lncRNA 1 (*FAL1*) [64] loci and the single nucleotide polymorphisms (SNPs) detected in enhancer and promoter regions of prostate cancer-associated transcript 1 (*PCAT1*) [65] and neuroblastoma-associated transcript 1 (*NBAT1*) [66], respectively. The contribution of dysregulated lncRNA expression to cancer has been an important ongoing question to pursue.

Current models propose that lncRNAs act primarily through two mechanisms: *cis*-regulatory, acting as local activators or repressors of the expression of neighboring genes, and *trans*-regulatory, acting as global modulators of cellular processes throughout the nucleus and in the cytoplasm [67]. The subcellular localization of an lncRNA detected by single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) can provide clues about its mechanism. Many *cis*-regulatory lncRNAs, including *PVT1*, accumulate at their site of transcription [68], while *trans*-regulatory lncRNAs, such as *Malat1*, tend to have a wider cellular distribution [69]. When designing gain-of-function and loss-of-function models to elucidate lncRNA function, it has been essential to consider the mechanism involved [70]. For example, the functions of *cis*-regulatory lncRNAs cannot be reliably studied using RNAi- or ASO-based knockdown approaches and their activities cannot be recapitulated with exogenously expressed constructs. In addition, genetic modifications of *cis*-regulatory lncRNA loci have to be assessed for their RNA-independent effects through perturbations of transcriptional and DNA regulatory elements. On the other hand, lncRNA/target accessibility and relative abundance have to be incorporated in models for *trans*-regulatory lncRNAs.

When one considers these caveats, it becomes apparent that designing experiments to study the contribution of dysregulated lncRNA expression in cancer, especially at the organismal level, can be very challenging. Efforts to investigate two well-studied cancer-associated lncRNAs, *PVT1* and *Malat1*, described below, illustrate the discordant results from alternative approaches and highlight the importance of carefully designed studies to dissect the contributions of lncRNAs in cancer initiation and progression.

Frequently amplified in cancer, *PVT1* offers a striking example of the challenges encountered when studying cancer lncRNAs. Initial work supported the proposed oncogenic function of *PVT1* in a murine model of breast cancer where genetically engineered *Myc-Pvt1* co-amplification was observed to be more tumorigenic than *Myc* amplification alone [71]. However, subsequent studies discovered tumor-suppressive elements in the *Pvt1* locus, including a DNA boundary element [72] and a p53-regulated isoform, *Pvt1b*, which acted *in cis* to downregulate *Myc* expression in response to genotoxic and oncogenic stress and during cellular senescence [68]. Inhibition of the p53-activated isoform *Pvt1b* by CRISPR mutagenesis of the p53 binding element *in vivo* led to larger tumors in an autochthonous mouse model of lung adenocarcinoma [68]. This finding placed *Pvt1b* downstream of the p53 tumor-suppressive pathway and indicated a specific role for *Pvt1b* in restraining growth early in tumor development through *Myc* repression. Thus, there is exciting potential for using the *Pvt1* locus as a therapeutic handle to modulating *Myc* levels in cancer. However, additional carefully designed GEMMs are needed to deconvolve the oncogenic and tumor-suppressive elements in the *Pvt1* locus and to dissect the interplay between regulatory DNA and RNA elements.

Another example of the biological and technical complexities associated with studying the roles of lncRNAs in cancer is provided by *Malat1*, whose overexpression is strongly associated with metastatic disease and poor patient prognosis across a remarkably wide range of cancer types [73]. Despite its ubiquitous expression, strong conservation, and high abundance in all mammalian cells, both the normal and the cancer functions of *Malat1* have remained a mystery. One of the main caveats of previous studies has been the use of three independent *Malat1* knockout mouse models, which have led to conflicting results likely because they did not properly recapitulate the potential gain-of-function or dominant phenotypes that *Malat1* overexpression may have imparted in the context of cancer [74,75]. Indeed, in unpublished work from our lab, CRISPR activation-guided epigenetic overexpression of endogenous *Malat1* in an autochthonous mouse model of lung adenocarcinoma confirmed that increased *Malat1* expression is a driver of cancer progression. We found that while *Malat1* does not modulate cellular proliferation, it cooperates with p53 loss to reprogram the tumor microenvironment and promote metastatic disease. The data indicated that *Malat1* drives tumor progression, rather than tumor initiation, consistent with its association with advanced disease and poor prognosis in patients. One of the outstanding questions is the link or lack thereof between *Malat1*'s normal and oncogenic functions and mechanisms. Another important area for further investigation is how targeting *Malat1* or its downstream targets may be harnessed to limit metastatic disease.

Attempting to model the dysregulation of lncRNAs in cancer models has not been without its challenges. However, growing efforts that take into account the evolving functions and mechanisms of lncRNAs also reflect the important insights that can be gained from well-designed studies.

4. The Human Y Chromosome, Long Non-Coding RNAs, and Cancer: Challenges and Opportunities

Ivan Martinez, West Virginia University, Morgantown, WV, USA

From amongst the 23 pairs of chromosomes comprising the human genome, the human Y chromosome is the smallest, containing only a little over 62 million base pairs. (Chromosome 1, the largest, spans about 249 million base pairs.) [76] For that reason, it is not surprising that the Y chromosome harbors, and expresses, fewer protein-coding genes than any other chromosome in the human genome [77]. Originally, the X and Y chromosomes contained the same amount of genetic information, but during evolution, the Y chromosome lost most of its genes, while maintaining the sex-determining SRY gene. Although the reference assembly of the human genome, building upon the Human Genome Project's original sequence completed in 2003, was updated in 2013, and again with supplementary annotations in 2019, named GRCh38 [78], the complete sequencing

of the human Y chromosome was not achieved until 2023 [76]. The main difficulty in sequencing the Y chromosome was engendered by its high amount of complex repetitive DNA structures, such as long palindromes, tandem repeats, and segmental duplications. This property is unique to the Y chromosome. There are other difficulties in the study of the human Y chromosome: in particular, during the process of aging, men's cells start to lose their Y chromosome [79], a condition that can be exacerbated by smoking [80]. The occasional loss of the Y chromosome in different male cell types is known as mosaic loss of Y chromosome (mLoY). This uneven mLoY has been linked to several diseases, such as a higher incidence of Alzheimer's disease [81], cardiac fibrosis [82], and deficient immune response [83], and a higher risk of cancer development [84]. A recent study using a pan-cancer approach revealed that different types of male cancers have different amounts of mLoY [85]. The study suggested that mLoY is a reflection of genetic instability in cancer cells and could be associated with a higher incidence of mutation rates [85]. All these findings pose an interesting question: How could the loss of a small chromosome, such as the Y chromosome, which only contains a few protein-coding genes, be so important in cancer development and aggressiveness? The answer may be found in the different types of non-coding RNA (ncRNA) transcripts expressed from the Y chromosome.

In comparison to other ncRNAs, such as microRNAs, small nuclear RNAs, and other "classical" ncRNA types, long non-coding RNAs (lncRNAs) have a higher genetic and functional diversity. In general, lncRNAs demonstrate a low rate of gene expression in normal cells, relative to the typical expression levels of protein-coding mRNAs, but lncRNA expression is altered in a wide variety of human cancers [86]. Interestingly, most lncRNAs, instead of having a high extent of evolutionary sequence conservation comparable to the majority of protein-coding genes, rely more on secondary structure conservation, which affords them a higher level of functional versatility [87]. Previously, it was accepted that lncRNAs, by definition, were unable to be translated into proteins, as they were assumed to lack bona fide open reading frames (ORFs). However, recent studies, starting with the findings from the Lipovich laboratory within the framework of Phase II of the ENCODE Consortium, experimentally showed, using direct methods such as mass spectrometry and indirect methods such as ribosome profiling, that certain lncRNAs have short ORFs and are able to be translated into "micropeptides" with a median size of 44 amino acids [88,89], even though these ORFs are distinct—in terms of their micro size, (recent) evolution, and (lack of) any database homologies or predicted functions—from conventional mRNA ORFs, indicating that despite containing short ORFs, these lncRNAs should not be reclassified as protein-coding. In general, lncRNAs have been classified into four major groups: antisense to protein-coding genes, sense-overlapping, intergenic, and intronic [90]. Antisense lncRNAs are transcribed from the antisense DNA strand of a protein-coding "sense" gene, and in contemporary nomenclature, they are commonly named with the protein-coding gene name with the addition of an -AS (antisense) at the end [91], for instance, FAM83H-AS1, AFAP1-AS1, and ASMTL-AS1. Sense lncRNAs are encoded on the same DNA strand as the gene they have been named after even though their sequence includes some of the gene's exonic regions which may not be translated (because they reside in untranslated regions or because of isoform dependence of the reading frame), as is the case with the lncRNA DAPLAR [92]. Intergenic lncRNAs, also known as long intergenic non-coding RNAs (lincRNAs), are the most common lncRNA type. These lncRNAs do not overlap any known coding genes [93], and include the lincSPRY3 family [94]. Intronic lncRNAs, as the name implies, are expressed within the intronic regions of protein-coding genes [95]. All these types of lncRNAs are transcribed from the human Y chromosome, but intergenic lncRNAs are the most commonly found there—perhaps due to the paucity of annotated protein-coding genes on this chromosome. Until now, the expression of more than 50 lncRNAs from the human Y chromosome has been documented, and evidence is emerging from recent studies that some of these ncRNAs are contributing to the etiology and pathogenesis of several human diseases.

A global gene expression analysis of male human brain embryos has unraveled six new lncRNAs (KDM5D, TTTY14, UTY, TTTY15, ARSEP1, and TXLNG2P) expressed from the Y chromosome that could be potentially involved in central nervous system development [96]. Dysregulation of these lncRNAs and others plays a role in many diseases including cancer. For example, lnc-KDM5D-4 has a role in fatty liver formation and cellular inflammation associated with atherosclerosis and coronary artery disease in men [97]. Also, the expression of the Y chromosome lncRNA TTTY15 was found to be frequently elevated in prostate cancer tissue, in comparison to normal tissue, proving that this lncRNA is not a mere biomarker but a functional contributor to cell proliferation; repression of this lncRNA showed a tumor-suppressive effect [98]. In multiple myeloma, the antisense lncRNA ZFY-AS1 was found to be a protective prognostic biomarker through bioinformatics analysis [99]. Another example is LINC00278, an lncRNA able to be translated into a small peptide (micropeptide), known as Yin Yang 1, that is important in the development of esophageal carcinomas [100]. Interestingly, LINC00278 was found to inhibit the growth of laryngeal squamous cell carcinoma cells in *in vitro* and *in vivo* models, by negatively regulating the AKT/mTOR signaling pathway via the downregulation of COL4A1/COL4A2 [101].

Our group discovered a family of lncRNAs expressed from the Y chromosome that directly contribute to radiation therapy susceptibility in male non-small-cell lung carcinomas (NSCLCs) [94]. We found three lncRNAs named lnc-SPRY3 RNAs (also known as lnc-BPY2C), which exhibited dose-dependent expression after radiation in radiosensitive male NSCLC cell lines, but not in radioresistant male NSCLC cell lines. Cytogenetic analysis revealed that the radioresistant male NSCLC cell lines lost their Y chromosome (mLoY), and consequently, lnc-SPRY3 RNAs did not exist in those cell lines. Gain- and loss-of-function experiments confirmed that the lnc-SPRY3 family functionally contributes to the radiation susceptibility of male NSCLC cells. More recent unpublished findings from our group have highlighted that the exogenous expression of the lnc-SPRY3 RNAs in female NSCLC cells, as well as mouse NSCLC cells, which also induced radiation susceptibility, suggesting that not only are the pathways affected by these lncRNAs present in female cells and other mammalian cells, but also that a potential use of these lncRNAs as therapeutic molecules may be feasible. The advantage of using mouse NSCLC cells expressing these human lncRNAs is enormous, because we can ask questions about the tumor microenvironment and the immune response, as well as address the functional impact of the differences between the mouse and human Y chromosomes, in well-established *in vivo* mouse xenograft models of cancer. One of the greatest obstacles of using mouse models to study lncRNAs expressed from the Y chromosome is the vast difference between the human and the mouse Y chromosomes. For example, the mouse Y chromosome is 99.9% euchromatin, expressing almost 700 protein-coding genes [102]. In contrast, the human Y chromosome is largely heterochromatic and has less than 40% euchromatin [76], serving as an example of the extensive, and often underappreciated, human–mouse differences in the non-coding sequence space of the genome. In this context, it is important to remember that, unlike protein-coding genes, most lncRNAs are not conserved between mammalian lineages, and hence, most human lncRNAs appear to be primate-specific in evolution, based on conventional nucleotide sequence homology analyses and multispecies alignments. Reflecting this trend, the linc-SPRY3 lncRNA genes are only present in humans and some nonhuman primates and are absent in mice.

An orthogonal opportunity to understand the importance of Y chromosome-encoded lncRNAs in cancer is afforded by the increasing adoption of artificial intelligence (AI) in bioinformatics. There is emerging evidence that AI can predict chromosomal aneuploidies in glioblastomas, using an imaging-based fully automated method [103]. Furthermore, AI was used to review histopathological images of colorectal cancer tumors to predict DNA yields [104]. Our group is interested in using AI to identify and characterize specific histopathological features found in male NSCLC tumors with the loss of the Y chromosome, in order to correlate these images with tumor staging, overall survival, and smoking history. By using DNA-FISH and hematoxylin and eosin (H&E) stained images from male NSCLC

tissue arrays, we are planning to test different segmentation and convolutional neuronal network algorithms to identify differences in histopathological features (including cell size, nuclear cluster, and nuclear texture) that predict the loss of the Y chromosome.

Summarily, here we highlight that the study of lncRNAs expressed from the human Y chromosome is an uncharted new field with exceptionally high potential to generate new cancer-related findings that could explain in part the sex-specific differences in cancer treatment efficacy as well as differences in overall survival between male and female cancer patients, simultaneously providing the mechanistic insights that could guide the development of lncRNA-driven therapeutics in NSCLC and beyond.

5. Non-Coding RNAs in the Genetics of Sports

Ekaterina G. Derevyanchuk, Southern Federal University, Rostov-on-Don, Russian Federation. Leonard Lipovich, Shenzhen Huayuan Biological Science Research Institute, Shenzhen, China, and School of Medicine, Wayne State University, Detroit, Michigan, USA.

To date, over 200 genes have been associated with the development and manifestation of human traits related to physical activity, exercise, and endurance [105,106]. The majority of these genes were annotated by Bouchard and colleagues in a human genetic map of physical activity [107,108]. A detailed study of these genes is required in order to understand how they convey the underlying phenotypes, and hence, how to properly design a personalized sports training regimen that accounts for genetics, accurately predicts individual athletes' capabilities, proactively identifies exercise-related health problems, and facilitates the detection of the new threat of gene-based doping. In the past two decades, the progressive emergence of high-throughput technologies to monitor transcriptome profiling has revealed that, while the exons of protein-coding genes account for a mere 1.5% of the 3,300,000,000-base human genome, more than 90% of the human genome is transcribed into non-coding RNA (ncRNA) transcripts. The most abundant class of these long non-coding RNAs (lncRNAs), encoded by 40,000 of the estimated 60,000 human genes, has been shown to be essential for a wide range of fundamental cellular and biological processes in health and disease [109–111].

Motivated by these revelations, our studies have been focused on the identification and localization of miRNA binding sites in the 3'UTRs of, and lncRNA transcripts adjoining and overlapping (including antisense and bidirectionally promoted transcripts of), key genes previously shown to be among the determinants of endurance and speed strength qualities (*PPARD*, *NRF2*, *PGC-1*, *HIF-1*, *HIF-2*, *GYS1*, *HBA1*, *HBA2*, *HBB*, *ADRB2*, *NOS3*, *CHRM2*, *UCP2*, *UCP3*, *VEGF*, and *EPO*), genes for muscle work efficiency (*ACE*, *CK-MM*, *ACTN3*, *MLCK*, *AMPD1*, and *IGF-1*), and genes responsible for psychological characteristics relevant to exercise (*5HTT*, *BDNF*—already known to be regulated by the antisense lncRNA *BDNF-AS1*—as discussed in the Human Brain Activity section of this mini-review, *HTR2A* (*SR*), *DRD2*, and *MTHFR*).

We used the TargetScan Human 8.0 resource to identify microRNA binding sites inside the studied genes (http://www.targetscan.org/vert_80/ (accessed on 5 October 2021)). TargetScan predicts the biological targets of miRNAs at protein-coding mRNA 3'UTRs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA. Sites are identified with mismatches in the seed region that are compensated by conserved 3' pairing. We used a Pct cutoff = 0.8 to select only significant miRNA binding sites. We identified binding sites for 26 microRNAs inside the studied genes.

Furthermore, we performed a search for disease-related SNPs in miRNA binding sites. We used miRdSNP, a database of disease-associated SNPs in miRNA target sites on 3'UTRs of human genes (<http://mirdsnp.ccr.buffalo.edu/>, (accessed on 28 September 2023)). We identified 21 miRNAs that were literature-supported (PubMed) disease-associated SNPs at their target sites in our selected set of genes. This study needs to be complemented in the future by competing endogenous RNA (ceRNA) prediction tools that may help identify, for example, miRNA–lncRNA interactions that sponge miRNAs away from important mRNA targets in the endurance and physical ability contexts.

Although, in addition to miRNAs, lncRNAs play critical roles in various biological functions and disease processes, human physical qualities related to lncRNAs have rarely been investigated to date. We carried out bioinformatics analysis using the UCSC Genome Browser and its underlying UCSC Genome Database (<https://genome.ucsc.edu/>, (accessed on 23 May 2023)) to find lncRNAs transcribed near, or overlapping, selected genes. We analyzed the length and tissue expression distribution of the candidate lncRNA identified, which included several TCONS transcripts (transcripts of uncertain coding potential, contributed to databases by high-throughput transcriptome projects but lacking detailed laboratory-based annotations), and we found that these characteristics vary widely. In particular, overall, we observed a lack of their expression in muscle tissue, except the HELLPAR lncRNA (a cause of HELLP Syndrome, a complication of pregnancy), supported by the TCONS_12_00006386, TCONS_12_00006387, and TCONS_12_00006388 transcripts located downstream of the *IGF1* gene and, intriguingly, expressed mainly in skeletal muscle, according to the “LincRNA RNA-seq” UCSC Genome Browser dataset.

It was already evident from the literature and from our pilot analysis that both short and long non-coding RNAs had great potential for further studies in the nascent field of sports genomics. The obtained findings should promote the understanding of possible lncRNA functions in the regulation of human physical qualities, with implications that range well beyond extreme sports such as deep-sea diving and high-altitude alpinism, in particular for the transcriptome-based control of endurance and extreme physiology, in situations such as recovery from natural and manmade disasters.

6. Long Non-Coding RNAs in COVID-19

Tatiana P. Shkurat, Southern Federal University, Rostov-on-Don, Russian Federation. Leonard Lipovich, Shenzhen Huayuan Biological Science Research Institute, Shenzhen, China, and School of Medicine, Wayne State University, Detroit, Michigan, USA.

During the opening decades of the post-genomic era, it has become increasingly clear that the majority of significant disease-associated variants discovered in Genome-Wide Association Studies (GWAS) across all human diseases reside in non-coding regions of the genome [112], and at the same time, long non-coding RNAs (lncRNAs) have been emerging as drivers of different diseases [113]. Tens of thousands of lncRNAs genes regulate the activity of protein-coding genes in various tissues and their response to environmental influences. Nevertheless, most lncRNAs' disease contributions and druggability potential remain poorly understood. To date (23 May 2023), more than 6.9 million people on our planet have died from the SARS-CoV-2 virus and more than 765.9 million have been infected (<https://covid19.who.int>, (accessed on 23 May 2023)). Extraordinary worldwide research and clinical efforts have been undertaken to understand the complex mechanisms of SARS-CoV-2 infection.

Coronaviruses that cause disease in humans have been transmitted to humans from animals and to domestic animals from bats. Viral mutations occur in animal hosts. Current consensus converges on the conclusion that, during passage through animal hosts, the virus accumulated mutations in the spike protein gene. This spike is important for the docking of SARS-CoV-2 to human cells. Comparative analysis of SARS-CoV-2 with other coronaviruses showed that this virus has a stretch of six new amino acids in its spike protein, Tyr–Leu–Thr–Pro–Gly–Asp [114]. New mutations increase dramatically in the SARS-CoV-2 spike protein when the virus is transmitted person to person, as the virus mutates mainly when it replicates in host cells. Analysis of 48635 SARS-CoV-2 genomes highlighted a total of 353341 mutation events, relative to the NC_045512.2 Wuhan reference genome, with an average of 7.2 mutations per sample [115]. Frequently occurring variants have been found in both the non-coding and the protein-coding regions of the viral genome. But the severity of the disease does not depend solely on the characteristics of the virus. The clinical manifestations of COVID-19 are extraordinarily variable and range from a complete absence of symptoms to severe respiratory failure and death. This extreme and unusual clinical variability suggests that host genetics (i.e., genomic variants) play a strong

role in the susceptibility to symptomatic disease and in and impact of the wide spectrum of manifestations of COVID-19.

We meta-analyzed data from GWAS and other host-side genetic studies of COVID, noticing the high diversity of the datasets. Nine causal SNPs and eleven pathogenic genes in COVID-19 have been identified in populations in Spain and Italy [116]. In the Chinese population, SNP variants common with other populations (including in the *ABO* blood group gene, previously implicated in susceptibility to diabetes and metabolic disease and harboring an antisense lncRNA that overlaps its first exon (L. Lipovich and E.L. Kleinrbink, pers. comm., unpublished), and in the antisense lncRNA *FOXP4-AS1*) were found to be significantly associated with COVID-19 severity. The rare risk variant in *MEF2B*, on the other hand, is specific to East Asian populations and confers an approximately eightfold increase in the risk of severe COVID-19 among carriers [117]. A meta-analysis of GWAS from more than 125,000 cases in 60 studies from 25 countries showed that genes at several loci, including *SFTPD*, *MUC5B*, and *ACE2*, contained variants that could be confidently used to predict the susceptibility to, and the severity of, symptomatic disease [118].

Intriguingly, several studies have highlighted that the expression of specific antisense lncRNAs was deregulated in patients with COVID-19 and correlated with the severity of COVID-19, and that this may play a role in the pathogenesis of this disease. Reduced expression of the lncRNAs *A2M-AS1* and *FLVCR1*, and increased expression of the lncRNAs *DBH-AS1*, *FLVCR1-DT*, and *NCBP2AS2*, was observed in patients with COVID-19. Both *FLVCR1-DT* and *NCBP2AS2* showed a positive correlation with interleukin-6 (IL-6). *DBH-AS1* and *FLVCR1-DT* had a significant association with mortality, complications, and mechanical ventilation. A significant negative correlation was found between *A2M-AS1* and *NCBP2AS2-1* as well as between *FLVCR1* and *FLVCR1-DT* [119], the latter consistent with many findings of antisense lncRNAs negatively regulating their cognate-sense mRNAs transcribed from the same locus, mentioned in other sections of this mini-review. In another recent study [120], it was shown that, in the severe form of COVID-19, 898 differentially expressed lncRNAs were detected in the peripheral blood of patients: 414 upregulated and 484 downregulated. In that study, through GO and KEGG gene ontology and functional category enrichment analysis, the authors reported the detected lncRNAs to have previously known functions in neurons, lung cancer, and organ injury.

We conducted a bioinformatics analysis of the localization of human long non-coding RNAs genes located within 50 kb of each significant phenotype-associated genetic variant SNP mapped in COVID GWAS and virus–host interactions. It is interesting to note that the majority of these SNPs are located on chromosome 2. In comparison, on chromosomes 6 (heavily involved in immunity thanks to hosting the MHC complex genes) and 19, only very small numbers of significant SNPs are located on similarly sized genomic fragments. We used data from the COVID GWAS v4 (genome-euro.ucsc.edu) and COVID19-hg GWAS round 5 meta-analyses (www.covid19hg.org). We examined all SNPs reported by these efforts for possible localization within lncRNA genes. We then constructed maps of interactions for the identified COVID severity, SNP-containing lncRNAs (*IL10RB-AS1*, *MGC57346*, *CCR5AS*) with miRNA and protein-coding genes. Using the LncRRlsearch web server (<http://rtools.cbrc.jp/LncRRlsearch/>, (accessed on 23 May 2023)), we predicted interactions between the miRNAs known to target the highlighted genes and various lncRNAs. Then, we analyzed the significant SNPs in the protein-coding genes targeted by miRNAs in this context and identified all lncRNAs that had the potential to interact with them. For example, we identified three lncRNAs that showed the highest probability of interaction with the tyrosine kinase (*TYK2*) gene: RP11-573D15.8-01 (antisense to the *FETUB*/fetuin-B protease inhibitor gene), AP006621.9-001 (antisense to the *EPS8L2* gene involved in actin cytoskeleton organization), and RP11-95O2.5-001 (antisense to *CELF4*). Each of these lncRNAs also interacts with more than 15 different miRNAs. This prototype points to the possibility of a massive and unbiased analysis of how, specifically, lncRNAs regulate coding gene activity, and of their relationship with miRNAs and with each other. These studies will empower quantitative assessments of the viral infection phenotypic

response range of any given host genotype, hence heralding the implementation of genome-driven, RNA-empowered, personalized medicine for more effective responses to current and future pandemics.

7. Deciphering Roles for lncRNAs in Human Brain Activity, Disease, and Death

Jeffrey A. Loeb, Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, USA

The human brain has evolved to generate extraordinary capabilities enabling complex behaviors not present in other species. Due to their often recent evolutionary origins relative to older conserved protein-coding genes, as well as to their more rapid rate of evolution as a consequence of being unencumbered by protein-coding open reading frames, lncRNAs have a significant potential to contribute to recent-origin and species-specific traits, including the phenotypic uniqueness of the human brain, through a wide variety of functions that are still poorly understood. At the same time, and perhaps because of these capabilities, the human brain suffers from a number of diseases related to both too much brain activity (epilepsy) and too little (neurodegeneration). Given these strong evolutionary differences, a major challenge in deciphering the roles of lncRNAs in human brain function and disease comes from a critical need to study them in the human brain rather than in other species.

For most other human brain diseases including Alzheimer's disease, we have to wait until death to collect tissues for research. When comparing the genomic landscape between fresh human brain samples and postmortem samples, we discovered marked differences in coding and non-coding gene expression suggesting that a dead brain is not entirely dead [121]. While neuronal genes involved in brain activity are rapidly degraded, other cells and "zombie genes" come to life, making the interpretation of postmortem human brain genomic studies challenging. In fact, we found that neuronal genes are rapidly lost during the postmortem interval, yet genes involved in brain repair from glial cells including astrocytes and microglia actually increase substantially. Histological inspection at these same time points confirmed these cellular changes, as predicted, from clustering the transcriptional changes. Of note, we found that non-coding genes including microRNAs and lncRNAs were lost at a much faster rate than protein-coding genes [122]. This finding demonstrates a critical importance to studying lncRNAs in freshly isolated tissues rather than postmortem tissues, especially from the human brain.

Patients who suffer from epileptic seizures can obtain significant improvements by removing portions of their brains. Epileptic electrical networks in fact are often quite large and complex, often requiring very large portions of the brain to be removed surgically in both children and adults. In order to make critical decisions about which portions to remove and which to leave intact, extensive long-term electrical recordings are performed in vivo. When precisely localized to the electrical recordings, the brain tissues removed from these surgical procedures offer a rare opportunity to explore differences in human brain regions with abnormal and normal brain activity in freshly isolated human brain that is not affected by a postmortem interval [122].

We took a multi-systems approach to study these priceless tissues linking genomic, proteomic, and molecular differences to quantitative electrical activities, brain imaging studies, and clinical data through a relational database [123]. Previously, we found many lncRNAs that were differentially expressed as a function of human brain epileptic spiking activity from human brain biopsies at regions of low and high epileptic activities (4). One of these, BDNFOS (BDNF-AS1), was an antisense lncRNA overlapping coding portions of the Brain-Derived Neurotrophic Factor (BDNF) gene that has important roles in synaptic remodeling and neuronal survival. We found that downregulation of BDNFOS led to the upregulation of BDNF which suggested that increasing BDNFOS expression could be a way of reducing pathological BDNF signaling in epileptic brain regions. More recently, we further explored lncRNA expression and function discovering that many lncRNAs are coregulated in the human brain with known MAPK signaling pathway protein coding

genes. Similar to BDNFOS, we found that some of these, particularly those antisense to nearby protein-coding genes, also regulate these coding genes, suggesting a large network of lncRNAs that can regulate important protein-coding genes within this important pathway that could lead to pathologic epileptic activities [124]. Consistently, we found that an MAPK inhibitor could prevent the development of epileptic spiking in an animal model [8].

8. Conclusions

This mini-review collated selected views and opinions on lncRNA biology, bioinformatics, genomics, and therapeutics, from the perspectives of several keynote, plenary and invited speakers, panelists, and organizers of the IEEE BIBM 3rd Annual lncRNA Workshop, which was held in Dubai, UAE, in December 2021. Given the focus on the 2021 Workshop, we acknowledge the limited scope of this mini-review, as it could not cover various equally important topics that were not discussed at the Workshop. Examples here include the accurate prediction of lncRNA targets and functions, context specificity of lncRNAs, lncRNAs in cell communication and immune response, connection between lncRNAs and bifunctional RNAs, etc. These provide promising avenues for future investigations.

Author Contributions: Conceptualization, D.A.A., X.Z., I.M. and L.L.; writing—original draft preparation, all authors; writing—review and editing, all authors; funding acquisition, D.A.A., X.Z., N.D., T.P.S., J.A.L. and I.M. All authors have read and agreed to the published version of the manuscript.

Funding: The 2021 BIBM-lncRNA Workshop and this work were supported in part by grants from the US National Science Foundation (NSF), Award #: 1747788 (DAA and XZ), 1920920 (DAA and IM), and 2125872 (to DAA). XZ was partially supported by the US National Institutes of Health—NIH (R01GM123037, U01AR069395, and R01CA241930), and by the NSF (Award #s 2217515 and 2326879); ARP was supported by the Fundação Araucária in NAPI Bioinformática (#66.2021) and the CNPq (440412/2022-6). ND was supported in part by the US National Cancer Institute (NCI), part of the NIH (R37CA230590 and R01CA262286) and by a Developmental Research Program Grant from the Yale SPORE in Lung Cancer (P50CA196530). TPS and EGD were supported by funds from the Ministry of Science and Higher Education of the Russian Federation (No. FENW-2023-0018). JAL was supported by funds from the NIH (R01 NS109515 and UG3 TR004501). IM was supported in part by a Research Scholar Grant (RSG-24-1039619-01-RMC) from the American Cancer Society.

Institutional Review Board Statement: For human brain tissue studies, informed consent was obtained from each patient (or legal guardian/parent for subjects under 18) at Wayne State University (Detroit, MI, USA) (IRB #2015-0457, approved 5/19/13) and the University of Illinois (Chicago, IL) (IRB#2019-1082, approved 11/3/21) with IRB approvals at both institutions. All methods were performed in accordance with the relevant guidelines and regulations.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge the help of Weijun Yi (Computer Science PhD student at West Virginia University) and Douglas Silva Domingues (Department of Genetics, “Luiz de Queiroz” College of Agriculture, University of São Paulo, ESALQ/USP, Piracicaba, Brazil) for their assistance in preparing this mini-review. We thank Emily Rice for her assistance in Dr. Martinez’s review. Figure 1 was contributed by Dr. MY Yang and Dr. X. Zhou.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

About the Authors

Dr. Donald Adjero is a Professor, Associate Departmental Chair, and Graduate Coordinator of Computer Science at the Lane Department of Computer Science and Electrical Engineering, West Virginia University (WVU), Morgantown, WV, USA. He is the Founding Director of WVU AI+Health Engineering Center. His research interests include machine learning, artificial intelligence (AI), search data structures, bioinformatics, and digital health. His research results have appeared in over 190 peer reviewed publications, including papers in several reputable conferences and journals, such as the

IEEE Transactions series, top journals in theoretical computer science, and renowned journals in bioinformatics and biomedicine. Since 2019, he has served as the Founding Chair for the IEEE BIBM Annual Workshop on Long Non-Coding RNAs (BIBM LncRNA). He also served as the Program Co-Chair for the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) held in two locations, Las Vegas, USA, and Changsha, China, on Dec. 6–9, 2022. Dr. Adjeroh's research has been funded by various federal funding agencies in the USA, such as the NSF, NASA, DHS, DoD/ONR, and DoJ/NIJ, including a current US 4M NSF RII Track-2 award and another current US 3M NSF NRT project on Bridges in Digital Health, for both of which he serves as the PI. He received the Department of Energy (DOE) CAREER award in 2002, the best paper award at the 25th Workshop on Information Technology and Systems (WITS 2015), the WVU Statler Outstanding Researcher award in 2009 and 2012, the WVU Statler Excellence in Research award in 2018, 2021, and 2024, and the WVU Statler Researcher of the Year award in 2021 and 2024. Dr. Adjeroh is an Associate Editor of IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB).

Dr. Xiaobo Zhou is a Fellow of the American Institute for Medical and Biological Engineering (AIMBE), a distinguished and tenured professor, and holds the Dr. and Mrs. Carl V. Vartian Professorship at UTHealth[®] Houston, where he is the Director of the Center for Computational Systems Medicine. Dr. Zhou is a world-class scholar and expert in applying translational omics, bioimaging, medical imaging, and EMR data to precision medicine. He has published over 350 international journal articles. Per Google Scholar, as of Jan 2024, Dr. Zhou's research has been cited over 19,000 times, and his h-index is 68. Since 2005, he has been fully and continuously funded by the NIH, NSF, and other major funding entities. He is a well-established, well-funded investigator and center director with internationally recognized expertise in basic, translational, and clinical research. Among his noteworthy accomplishments, Dr. Zhou has pioneered high-content cellular imaging informatics, bioinformatics, systems biology, systems modeling-guided cancer and regenerative medicine, and imaging-aided surgical design/optimization.

Dr. Alexandre Paschoal is a Group Leader in Artificial Intelligence and Informatics at the Rosalind Franklin Institute, United Kingdom, and concurrently serves as an Associate Professor at the Federal University of Technology—Parana (UTFPR), Brazil. Having earned his Ph.D. in Bioinformatics from the University of São Paulo in 2012 in Brazil, Dr. Paschoal exemplifies a commitment to interdisciplinary research. Dr. Paschoal is broadly trained in computer science, with his primary research focus lying in build cutting edge AI techniques to translate biological data into actionable knowledge utilizing principles from the field of bioinformatics. His vision involves contributions to link the fields involving data sciences and life/biological sciences to extract insights to guide decision-making. His primary biological focus lies in non-coding RNAs and transposable elements.

Dr. Nadya Dimitrova is an Assistant Professor in the Departments of Molecular, Cellular and Developmental Biology and Genetics at Yale University and a member of the Yale Center for RNA Science and Medicine and the Yale Cancer Center. She has a long-standing interest in understanding the functions and mechanisms of action of long non-coding RNAs (lncRNAs) in the context of cancer. Her work is focused on developing genetic models to study the roles of lncRNAs in normal and disease states and gaining deeper insights into mechanisms of lncRNAs. Nadya Dimitrova is the recipient of the HHMI Predoctoral Fellowship, the Damon Runyon Postdoctoral Fellowship Award, the Lung Cancer Research Foundation Scientific Merit Award, the Pew-Stewart Award for Cancer Research, the V Scholar Award, and the NCI Merit Award.

Dr. Ekaterina G. Derevyanchuk is an Associate Professor at the Genetics Department at the Academy of Biology and Biotechnology named after D.I. Ivanovsky, Southern Federal University (Rostov-on-Don, Russian Federation). She is a member of the Vavilov Society of Geneticists and Breeders. Her work focuses on the study of the role of non-coding RNAs in various diseases and the development of diagnostic test systems based on the data obtained. Particularly, her studies are related to the search of localization of ncRNA binding sites around and inside genes associated with various diseases. Dr. Derevyanchuk is a winner of the All-Russian contest, held by the Foundation for Assistance to Small Innovative Enterprises (FASIE), a German Academic Exchange Service (DAAD) scholarship holder, and a winner of the young scientists contest held within the framework of the International Congress "Early Pregnancy" (RUDN, Moscow, Russia). She is also co-owner of a small gold medal for the development of bioinformatics technology for searching for the relationship between organization scenarios in animal and human genomes of non-coding DNA and protein-coding DNA (RosBioTech, Moscow, Russia). She is the author of the best poster presentation at the Interregional Scientific and Practical Conference "Cellular Technologies for Practical Healthcare" (Yekaterinburg, Russia) and winner of the competition for the best poster report in the

framework of the International Conference of Young Scientists “New directions in Life Sciences” (Yerevan, Armenia).

Dr. Tatiana Shkurat is Professor and Head of the Genetics Department at the Academy of Biology and Biotechnology named after D.I. Ivanovsky, Southern Federal University (Rostov-on-Don, Russian Federation). She is PI and group leader at her home institution. She has extensive experience in supervising research teams and training doctoral candidates and has led more than 30 grant-funded research projects in Russia as well as internationally sponsored projects. She has developed extensive intellectual property (IP), having been granted patents on methods of gene diagnostics of monogenic and multifactorial diseases, certificates of the computer program for automatic search of motifs, and visual analysis of images of nucleotide sequences. Recently, Prof. Shkurat has focused on the interface of non-coding DNA/RNA, evolutionary genomics, and personalized medicine. Her preliminary results represent a significant advancement toward the future deployment of approaches for the validation and utilization of novel therapeutic targets in patients with different diseases, based on lncRNA and genome editing.

Dr. Jeffrey Loeb is a practicing neurologist, epileptologist, and neuroscientist who currently holds the John S. Garvin Chair, Professor and Head of the Department of Neurology and Rehabilitation. He received his A.B. in Chemistry, M.D., and Ph.D. from the University of Chicago. After completing a residency in Neurology at the Massachusetts General Hospital in Boston, he joined the faculty of Harvard Medical School where he also had fellowship training in epilepsy at Harvard’s Beth Israel Deaconess Hospital. Dr. Loeb conducted postdoctoral work in the Department of Neurobiology at Harvard Medical School with Dr. Gerald Fischbach, where he became interested in how understanding early neural development can teach us what goes wrong in human disease and suggest new treatments. This and other work focusing on translating basic discoveries into new treatments has resulted in high-profile publications, federal and foundation grant support, and a number of patents, including one for a targeted drug that blocks neuregulin signaling and shows promise in degenerative disease by blocking neural inflammation, such as in ALS and Alzheimer’s. His work in treating epilepsy patients led to a one-of-a-kind, large-scale, systems biology project of human brain tissues that links the genes, molecular, and cellular signals that underlie the abnormal electrical activities that characterize this common neurological disorder. Dr. Loeb is director of the University of Illinois NeuroRepository, co-director of biomedical informatics in UIC’s Center for Clinical and Translational Science, and Chief Clinical Strategist for the Sturge-Weber Foundation. To build on his scientific efforts and translate discoveries back to patients through public-private partnerships, he recently formed “I-BRAIN” (Illinois Brain Analytics) in close collaboration and with support from the Discovery Partners Institute.

Dr. Ivan Martinez is an Associate Professor at the West Virginia University (WVU) Cancer Institute and the Department of Microbiology, Immunology and Cellular Biology at WVU School of Medicine. Originally from Mexico City, he graduated with honors from The National Autonomous University of Mexico (UNAM). He completed his Ph.D. in Molecular Genetics and Biochemistry at the University of Pittsburgh and continued his training as a postdoctoral fellow in the Department of Genetics at Yale University School of Medicine for over 5 years. After his training, he was recruited at WVU in 2013. The goal of Dr. Martinez’s research is to understand the importance of non-coding RNAs (ncRNAs), such as microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular non-coding RNAs (circRNAs), in the process of carcinogenesis and viral infections. Dr. Martinez has been invited to present his research at prestigious institutions such as Harvard University, the University of Massachusetts Chan Medical School, the National Cancer Institute (NCI) at the National Institutes of Health (NIH), Gottingen University-Max Planck Institutes in Germany, and the National Institute of Genomic Medicine (Instituto Nacional de Medicina Genómica) in Mexico City. Since 2019, he has served as the Co-Chair for the International Conference on Bioinformatics and Biomedicine (IEEE BIBM) Annual Workshop on Long Non-Coding RNAs (BIBM lncRNA). He was also the Co-Chair in the organization of the 2022 Keystone Symposium “Small Regulatory RNAs: From Bench to Bedside” in Santa Fe, New Mexico. Dr. Martinez was also the co-leader of a team of researchers involved in the development of the COVID-19 PCR test at WVU (from in-house reagents to a fully automated testing robot). The test developed by Dr. Martinez’s team was the main test used not only at WVU but in the entire State of West Virginia during the COVID-19 pandemic. Furthermore, Dr. Martinez has been a member of the World Health Organization (WHO) COVID-19 Animal Model Group since 2021.

Dr. Leonard Lipovich is an internationally recognized pioneer in the long non-coding RNA (lncRNA) biology of human disease. He led a major international project with the Japan-based

FANTOM Consortium (which built the first comprehensive catalog of mouse protein-coding and lncRNA genes), where he discovered that lncRNAs are often not conserved in evolution between primates and rodents and are more numerous than protein-coding genes. He built the first published catalog of human lncRNA genes supported by full-length experimental data. Leading a platform effort within the ENCODE (Encyclopedia of DNA Elements) Consortium—a direct successor to the Human Genome Project—Prof. Lipovich was the first to empirically reveal the unexpected ribosomal translation of short open reading frames from lncRNAs in human cells using mass spectrometry. His work has been fundamental to establishing that primate-specific lncRNAs in humans directly contribute to the etiology of cancer, diabetes, and epilepsy. His current focus is on identifying, and validating in the laboratory, lncRNAs from genome-wide association studies and personalized genome sequencing, as well as from in vivo transcriptome analysis and high-throughput screens, as novel and druggable causes of those diseases.

Leonard Lipovich earned his B.A. (*cum laude*) in Genetics and Development from Cornell University (1998) and his Ph.D. in Molecular Biotechnology from the Department of Genome Sciences at the University of Washington, Seattle (2003). He completed postdoctoral training at the Genome Institute of Singapore, where he characterized the first mammalian lncRNA functional in stem cell pluripotency and discovered the prevalence of primate-specific lncRNAs in sense–antisense gene pairs. Dr. Lipovich joined Wayne State University in Detroit, Michigan, USA in 2007, as an assistant professor, attaining an associate professorship and tenure in 2013, then a full professorship in 2019. In 2024, he transitioned to Wenzhou-Kean University in Wenzhou, China. He is active in lncRNA-based drug development efforts, through commercialization projects with startup RNA-biotech companies in China and elsewhere. Dr. Lipovich received the U.S. National Institutes of Health (NIH) Director’s New Innovator Award, for the 2014–2019 project cycle. In 2015, he chaired both the Keystone Symposium and the Royal Society international scientific meeting on lncRNA. To date, Dr. Lipovich (h-index: 45) has published 87 peer-reviewed papers (including 36 as a first, last, or principal author), delivered over 50 invited presentations, and chaired numerous sessions at key conferences. Dr. Lipovich is a Founding Co-Chair (2019–present) of the Annual lncRNA Workshop at the IEEE-BIBM. His goal is to improve human health through personalized, lncRNA-targeted, post-genomic therapeutics.

References

1. Wang, J.; Zhu, S.; Meng, N.; He, Y.; Lu, R.; Yan, G.-R. ncRNA-Encoded Peptides or Proteins and Cancer. *Mol. Ther.* **2019**, *27*, 1718–1725. [[CrossRef](#)]
2. Zhu, S.; Wang, J.; He, Y.; Meng, N.; Yan, G.-R. Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers. *Front. Pharmacol.* **2018**, *9*, 1295. [[CrossRef](#)] [[PubMed](#)]
3. Huang, J.-Z.; Chen, M.; Chen, D.; Gao, X.-C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.-R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171–184.e6. [[CrossRef](#)] [[PubMed](#)]
4. Yang, Y.; Gao, X.; Zhang, M.; Yan, S.; Sun, C.; Xiao, F.; Huang, N.; Yang, X.; Zhao, K.; Zhou, H.; et al. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* **2018**, *110*, 304–315. [[CrossRef](#)]
5. Fang, J.; Morsalin, S.; Rao, V.N.; Reddy, E.S.P. Decoding of Non-Coding DNA and Non-Coding RNA: Pri-Micro RNA-Encoded Novel Peptides Regulate Migration of Cancer Cells. *J. Pharm. Sci. Pharmacol.* **2017**, *3*, 23–27. [[CrossRef](#)]
6. Bonidia, R.P.; Sampaio, L.D.H.; Domingues, D.S.; Paschoal, A.R.; Lopes, F.M.; de Carvalho, A.C.P.L.F.; Sanches, D.S. Feature extraction approaches for biological sequences: A comparative study of mathematical features. *Brief. Bioinform.* **2021**, *22*, bbab011. [[CrossRef](#)] [[PubMed](#)]
7. Ito, E.A.; Katahira, I.; Vicente, F.F.d.R.; Pereira, L.F.P.; Lopes, F.M. BASiNET—BiologicAl Sequences NETwork: A case study on coding and non-coding RNAs identification. *Nucleic Acids Res.* **2018**, *46*, e96. [[CrossRef](#)]
8. Kumari, P.; Sampath, K. cncRNAs: Bi-functional RNAs with protein coding and non-coding functions. *Semin. Cell Dev. Biol.* **2015**, *47–48*, 40–51. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
9. Chen, G.; Yang, J.; Chen, J.; Song, Y.; Cao, R.; Shi, T.; Shi, L. Identifying and annotating human bifunctional RNAs reveals their versatile functions. *Sci. China Life Sci.* **2016**, *59*, 981–992. [[CrossRef](#)]
10. Huang, Y.; Wang, J.; Zhao, Y.; Wang, H.; Liu, T.; Li, Y.; Cui, T.; Li, W.; Feng, Y.; Luo, J.; et al. cncRNAdb: A manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. *Nucleic Acids Res.* **2021**, *49*, D65–D70. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
11. Couso, J.-P.; Patraquim, P. Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 575–589. [[CrossRef](#)]
12. Wang, S.; Mao, C.; Liu, S. Peptides encoded by noncoding genes: Challenges and perspectives. *Signal Transduct. Target. Ther.* **2019**, *4*, 57. [[CrossRef](#)] [[PubMed](#)]

13. Cui, T.; Dou, Y.; Tan, P.; Ni, Z.; Liu, T.; Wang, D.; Huang, Y.; Cai, K.; Zhao, X.; Xu, D.; et al. RNALocate v2.0: An updated resource for RNA subcellular localization with increased coverage and annotation. *Nucleic Acids Res.* **2022**, *50*, D333–D339. [[CrossRef](#)] [[PubMed](#)]
14. Mas-Ponte, D.; Carlevaro-Fita, J.; Palumbo, E.; Hermoso Pulido, T.; Guigo, R.; Johnson, R. LncAtlas database for subcellular localization of long noncoding RNAs. *RNA* **2017**, *23*, 1080–1087. [[CrossRef](#)]
15. The RNAcentral Consortium RNAcentral: A hub of information for non-coding RNA sequences. *Nucleic Acids Res.* **2019**, *47*, D221–D229. [[CrossRef](#)]
16. Zhao, Y.; Li, H.; Fang, S.; Kang, Y.; Wu, W.; Hao, Y.; Li, Z.; Bu, D.; Sun, N.; Zhang, M.Q.; et al. NONCODE 2016: An informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* **2016**, *44*, D203–D208. [[CrossRef](#)] [[PubMed](#)]
17. Li, L.; Wu, P.; Wang, Z.; Meng, X.; Zha, C.; Li, Z.; Qi, T.; Zhang, Y.; Han, B.; Li, S.; et al. NoncoRNA: A database of experimentally supported non-coding RNAs and drug targets in cancer. *J. Hematol. Oncol.* **2020**, *13*, 15. [[CrossRef](#)]
18. Teng, X.; Chen, X.; Xue, H.; Tang, Y.; Zhang, P.; Kang, Q.; Hao, Y.; Chen, R.; Zhao, Y.; He, S. NPInter v4.0: An integrated database of ncRNA interactions. *Nucleic Acids Res.* **2020**, *48*, D160–D165. [[CrossRef](#)]
19. ViRBase v3.0: A Virus and Host ncRNA-Associated Interaction Repository with Increased Coverage and Annotation—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/34723320/> (accessed on 7 February 2024).
20. Ning, L.; Cui, T.; Zheng, B.; Wang, N.; Luo, J.; Yang, B.; Du, M.; Cheng, J.; Dou, Y.; Wang, D. MNDR v3.0: Mammal ncRNA-disease repository with increased coverage and annotation. *Nucleic Acids Res.* **2021**, *49*, D160–D164. [[CrossRef](#)]
21. Liu, C.-J.; Xie, G.-Y.; Miao, Y.-R.; Xia, M.; Wang, Y.; Lei, Q.; Zhang, Q.; Guo, A.-Y. EVAtlas: A comprehensive database for ncRNA expression in human extracellular vesicles. *Nucleic Acids Res.* **2022**, *50*, D111–D117. [[CrossRef](#)]
22. Li, J.; Xue, Y.; Amin, M.T.; Yang, Y.; Yang, J.; Zhang, W.; Yang, W.; Niu, X.; Zhang, H.-Y.; Gong, J. ncRNA-eQTL: A database to systematically evaluate the effects of SNPs on non-coding RNA expression across cancer types. *Nucleic Acids Res.* **2020**, *48*, D956–D963. [[CrossRef](#)] [[PubMed](#)]
23. Wang, J.; Cao, Y.; Zhang, H.; Wang, T.; Tian, Q.; Lu, X.; Lu, X.; Kong, X.; Liu, Z.; Wang, N.; et al. NSDNA: A manually curated database of experimentally supported ncRNAs associated with nervous system diseases. *Nucleic Acids Res.* **2017**, *45*, D902–D907. [[CrossRef](#)] [[PubMed](#)]
24. Liu, H.; Zhou, X.; Yuan, M.; Zhou, S.; Huang, Y.-E.; Hou, F.; Song, X.; Wang, L.; Jiang, W. ncEP: A Manually Curated Database for Experimentally Validated ncRNA-encoded Proteins or Peptides. *J. Mol. Biol.* **2020**, *432*, 3364–3368. [[CrossRef](#)] [[PubMed](#)]
25. Dragomir, M.P.; Manyam, G.C.; Ott, L.F.; Berland, L.; Knutsen, E.; Ivan, C.; Lipovich, L.; Broom, B.M.; Calin, G.A. FuncPEP: A Database of Functional Peptides Encoded by Non-Coding RNAs. *Noncoding RNA* **2020**, *6*, 41. [[CrossRef](#)] [[PubMed](#)]
26. Luo, X.; Huang, Y.; Li, H.; Luo, Y.; Zuo, Z.; Ren, J.; Xie, Y. SPENCER: A comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. *Nucleic Acids Res.* **2022**, *50*, D1373–D1381. [[CrossRef](#)] [[PubMed](#)]
27. Yang, M.; Lu, H.; Liu, J.; Wu, S.; Kim, P.; Zhou, X. lncRNAfunc: A knowledgebase of lncRNA function in human cancer. *Nucleic Acids Res.* **2022**, *50*, D1295–D1306. [[CrossRef](#)] [[PubMed](#)]
28. Lv, D.; Chang, Z.; Cai, Y.; Li, J.; Wang, L.; Jiang, Q.; Xu, K.; Ding, N.; Li, X.; Xu, J.; et al. TransLnc: A comprehensive resource for translatable lncRNAs extends immunopeptidome. *Nucleic Acids Res.* **2022**, *50*, D413–D420. [[CrossRef](#)] [[PubMed](#)]
29. Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J.R.A.; Hellinga, A.J.; Lugo, C.S.B.; Elliott, T.A.; Ware, D.; Peterson, T.; et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **2019**, *20*, 275. [[CrossRef](#)]
30. Wright, M.W. A short guide to long non-coding RNA gene nomenclature. *Hum. Genom.* **2014**, *8*, 7. [[CrossRef](#)]
31. Seal, R.L.; Tweedie, S.; Bruford, E.A. A standardised nomenclature for long non-coding RNAs. *IUBMB Life* **2023**, *75*, 380–389. [[CrossRef](#)]
32. Seal, R.L.; Chen, L.-L.; Griffiths-Jones, S.; Lowe, T.M.; Mathews, M.B.; O'Reilly, D.; Pierce, A.J.; Stadler, P.F.; Ulitsky, I.; Wolin, S.L.; et al. A guide to naming human non-coding RNA genes. *EMBO J.* **2020**, *39*, e103777. [[CrossRef](#)]
33. Szcześniak, M.W.; Rosikiewicz, W.; Makołowska, I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol.* **2016**, *57*, e8. [[CrossRef](#)]
34. Szcześniak, M.W.; Bryzghalov, O.; Ciombarowska-Basheer, J.; Makołowska, I. CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs. *Methods Mol. Biol.* **2019**, *1933*, 415–429. [[CrossRef](#)]
35. Paytuví Gallart, A.; Hermoso Pulido, A.; Anzar Martínez de Lagrán, I.; Sanseverino, W.; Aiese Cigliano, R. GREENC: A Wiki-based database of plant lncRNAs. *Nucleic Acids Res.* **2016**, *44*, D1161–D1166. [[CrossRef](#)] [[PubMed](#)]
36. Di Marsico, M.; Paytuvi Gallart, A.; Sanseverino, W.; Aiese Cigliano, R. GreenC 2.0: A comprehensive database of plant long non-coding RNAs. *Nucleic Acids Res.* **2022**, *50*, D1442–D1447. [[CrossRef](#)] [[PubMed](#)]
37. Maracaja-Coutinho, V.; Paschoal, A.R.; Caris-Maldonado, J.C.; Borges, P.V.; Ferreira, A.J.; Durham, A.M. Noncoding RNAs Databases: Current Status and Trends. *Methods Mol. Biol.* **2019**, *1912*, 251–285. [[CrossRef](#)] [[PubMed](#)]
38. Paschoal, A.R.; Maracaja-Coutinho, V.; Setubal, J.C.; Simões, Z.L.P.; Verjovski-Almeida, S.; Durham, A.M. Non-coding transcription characterization and annotation: A guide and web resource for non-coding RNA databases. *RNA Biol.* **2012**, *9*, 274–282. [[CrossRef](#)]
39. Ernst, C.; Morton, C.C. Identification and function of long non-coding RNA. *Front. Cell Neurosci.* **2013**, *7*, 168. [[CrossRef](#)]

40. Miller, J.R.; Adjeroh, D.A. Exploring Neural Network Models for LncRNA Sequence Identification. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2920–2927.
41. Zheng, H.; Talukder, A.; Li, X.; Hu, H. A systematic evaluation of the computational tools for lncRNA identification. *Brief. Bioinform.* **2021**, *22*, bbab285. [[CrossRef](#)]
42. Dönnes, P.; Höglund, A. Predicting protein subcellular localization: Past, present, and future. *Genom. Proteom. Bioinform.* **2004**, *2*, 209–215. [[CrossRef](#)]
43. Garg, A.; Singhal, N.; Kumar, R.; Kumar, M. mRNAloc: A novel machine-learning based in-silico tool to predict mRNA subcellular localization. *Nucleic Acids Res.* **2020**, *48*, W239–W243. [[CrossRef](#)] [[PubMed](#)]
44. Babaiha, N.S.; Aghdam, R.; Ghiam, S.; Eslahchi, C. NN-RNALoc: Neural network-based model for prediction of mRNA sub-cellular localization using distance-based sub-sequence profiles. *PLoS ONE* **2023**, *18*, e0258793. [[CrossRef](#)] [[PubMed](#)]
45. Musleh, S.; Islam, M.T.; Qureshi, R.; Alajez, N.M.; Alam, T. MSLP: mRNA subcellular localization predictor based on machine learning techniques. *BMC Bioinform.* **2023**, *24*, 109. [[CrossRef](#)]
46. Cabili, M.N.; Dunagin, M.C.; McClanahan, P.D.; Biaesch, A.; Padovan-Merhar, O.; Regev, A.; Rinn, J.L.; Raj, A. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **2015**, *16*, 20. [[CrossRef](#)] [[PubMed](#)]
47. Gudenäs, B.L.; Wang, L. Prediction of LncRNA Subcellular Localization with Deep Learning from Sequence Features. *Sci. Rep.* **2018**, *8*, 16385. [[CrossRef](#)] [[PubMed](#)]
48. Yi, W.; Adjeroh, D.A. A Deep Learning Approach to LncRNA Subcellular Localization Using Inexact q-mers. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 2128–2133.
49. Yuan, G.-H.; Wang, Y.; Wang, G.-Z.; Yang, L. RNAlight: A machine learning model to identify nucleotide features determining RNA subcellular localization. *Brief. Bioinform.* **2023**, *24*, bbac509. [[CrossRef](#)] [[PubMed](#)]
50. Feng, H.; Wang, S.; Wang, Y.; Ni, X.; Yang, Z.; Hu, X.; Yang, S. LncCat: An ORF attention model to identify LncRNA based on ensemble learning strategy and fused sequence information. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1433–1447. [[CrossRef](#)] [[PubMed](#)]
51. Gong, H.; Wen, J.; Luo, R.; Feng, Y.; Guo, J.; Fu, H.; Zhou, X. Integrated mRNA sequence optimization using deep learning. *Brief. Bioinform.* **2023**, *24*, bbad001. [[CrossRef](#)]
52. Genome10k | Unveiling Animal Diversity. Available online: <https://genome10k.ucsc.edu/> (accessed on 1 March 2024).
53. 100,000 genomes—In Africa, for Africa. *Nat. Biotechnol.* **2022**, *40*, 988. [[CrossRef](#)]
54. 100,000 Genomes Project. Available online: <https://www.genomicsengland.co.uk/initiatives/100000-genomes-project> (accessed on 1 March 2024).
55. Home | 10KP | CNGB. Available online: <https://db.cngb.org/10kp/> (accessed on 1 March 2024).
56. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [[CrossRef](#)]
57. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of transcription in human cells. *Nature* **2012**, *489*, 101–108. [[CrossRef](#)] [[PubMed](#)]
58. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [[CrossRef](#)] [[PubMed](#)]
59. Olivero, C.E.; Dimitrova, N. Identification and characterization of functional long noncoding RNAs in cancer. *FASEB J.* **2020**, *34*, 15630–15646. [[CrossRef](#)]
60. Yan, X.; Hu, Z.; Feng, Y.; Hu, X.; Yuan, J.; Zhao, S.D.; Zhang, Y.; Yang, L.; Shan, W.; He, Q.; et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell* **2015**, *28*, 529–540. [[CrossRef](#)]
61. Ji, P.; Diederichs, S.; Wang, W.; Böing, S.; Metzger, R.; Schneider, P.M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E.; et al. MALAT-1, a novel noncoding RNA, and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **2003**, *22*, 8031–8041. [[CrossRef](#)] [[PubMed](#)]
62. Gupta, R.A.; Shah, N.; Wang, K.C.; Kim, J.; Horlings, H.M.; Wong, D.J.; Tsai, M.-C.; Hung, T.; Argani, P.; Rinn, J.L.; et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **2010**, *464*, 1071–1076. [[CrossRef](#)]
63. Tseng, Y.-Y.; Bagchi, A. The PVT1-MYC duet in cancer. *Mol. Cell. Oncol.* **2015**, *2*, e974467. [[CrossRef](#)] [[PubMed](#)]
64. Hu, X.; Feng, Y.; Zhang, D.; Zhao, S.D.; Hu, Z.; Greshock, J.; Zhang, Y.; Yang, L.; Zhong, X.; Wang, L.-P.; et al. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell* **2014**, *26*, 344–357. [[CrossRef](#)]
65. Guo, H.; Ahmed, M.; Zhang, F.; Yao, C.Q.; Li, S.; Liang, Y.; Hua, J.; Soares, F.; Sun, Y.; Langstein, J.; et al. Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat. Genet.* **2016**, *48*, 1142–1150. [[CrossRef](#)]
66. Pandey, G.K.; Mitra, S.; Subhash, S.; Hertwig, F.; Kanduri, M.; Mishra, K.; Fransson, S.; Ganeshram, A.; Mondal, T.; Bandaru, S.; et al. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* **2014**, *26*, 722–737. [[CrossRef](#)]
67. Winkler, L.; Dimitrova, N. A mechanistic view of long noncoding RNAs in cancer. *Wiley Interdiscip. Rev. RNA* **2022**, *13*, e1699. [[CrossRef](#)] [[PubMed](#)]

68. Olivero, C.E.; Martínez-Terroba, E.; Zimmer, J.; Liao, C.; Tesfaye, E.; Hooshdaran, N.; Schofield, J.A.; Bendor, J.; Fang, D.; Simon, M.D.; et al. p53 Activates the Long Noncoding RNA Pvt1b to Inhibit Myc and Suppress Tumorigenesis. *Mol. Cell* **2020**, *77*, 761–774.e8. [CrossRef] [PubMed]
69. Hutchinson, J.N.; Ensminger, A.W.; Clemson, C.M.; Lynch, C.R.; Lawrence, J.B.; Chess, A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genom.* **2007**, *8*, 39. [CrossRef] [PubMed]
70. Bassett, A.R.; Akhtar, A.; Barlow, D.P.; Bird, A.P.; Brockdorff, N.; Duboule, D.; Ephrussi, A.; Ferguson-Smith, A.C.; Gingeras, T.R.; Haerty, W.; et al. Considerations when investigating lncRNA function in vivo. *Elife* **2014**, *3*, e03058. [CrossRef] [PubMed]
71. Tseng, Y.-Y.; Moriarity, B.S.; Gong, W.; Akiyama, R.; Tiwari, A.; Kawakami, H.; Ronning, P.; Reuland, B.; Guenther, K.; Beadnell, T.C.; et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature* **2014**, *512*, 82–86. [CrossRef]
72. Cho, S.W.; Xu, J.; Sun, R.; Mumbach, M.R.; Carter, A.C.; Chen, Y.G.; Yost, K.E.; Kim, J.; He, J.; Nevins, S.A.; et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **2018**, *173*, 1398–1412.e22. [CrossRef]
73. Gutschner, T.; Hämmerle, M.; Diederichs, S. MALAT1—A paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* **2013**, *91*, 791–801. [CrossRef] [PubMed]
74. Sun, Y.; Ma, L. New Insights into Long Non-Coding RNA MALAT1 in Cancer and Metastasis. *Cancers* **2019**, *11*, 216. [CrossRef] [PubMed]
75. Arun, G.; Aggarwal, D.; Spector, D.L. MALAT1 Long Non-Coding RNA: Functional Implications. *Noncoding RNA* **2020**, *6*, 22. [CrossRef]
76. Rhie, A.; Nurk, S.; Cechova, M.; Hoyt, S.J.; Taylor, D.J.; Altemose, N.; Hook, P.W.; Koren, S.; Rautiainen, M.; Alexandrov, I.A.; et al. The complete sequence of a human Y chromosome. *Nature* **2023**, *621*, 344–354. [CrossRef]
77. National Center for Biotechnology Information. Chromosome Map. In *Genes and Disease [Internet]*; National Center for Biotechnology Information (US): Bethesda, MD, USA, 1998.
78. Schneider, V.A.; Graves-Lindsay, T.; Howe, K.; Bouk, N.; Chen, H.-C.; Kitts, P.A.; Murphy, T.D.; Pruitt, K.D.; Thibaud-Nissen, F.; Albracht, D.; et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **2017**, *27*, 849–864. [CrossRef] [PubMed]
79. Zhou, W.; Machiela, M.J.; Freedman, N.D.; Rothman, N.; Malats, N.; Dagnall, C.; Caporaso, N.; Teras, L.T.; Gaudet, M.M.; Gapstur, S.M.; et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat. Genet.* **2016**, *48*, 563–568. [CrossRef] [PubMed]
80. Dumanski, J.P.; Rasi, C.; Lönn, M.; Davies, H.; Ingelsson, M.; Giedraitis, V.; Lannfelt, L.; Magnusson, P.K.E.; Lindgren, C.M.; Morris, A.P.; et al. Smoking is associated with mosaic loss of chromosome Y. *Science* **2015**, *347*, 81–83. [CrossRef]
81. Dumanski, J.P.; Lambert, J.-C.; Rasi, C.; Giedraitis, V.; Davies, H.; Grenier-Boley, B.; Lindgren, C.M.; Campion, D.; Dufouil, C.; European Alzheimer’s Disease Initiative Investigators; et al. Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am. J. Hum. Genet.* **2016**, *98*, 1208–1219. [CrossRef]
82. Sano, S.; Horitani, K.; Ogawa, H.; Halvardson, J.; Chavkin, N.W.; Wang, Y.; Sano, M.; Mattisson, J.; Hata, A.; Danielsson, M.; et al. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science* **2022**, *377*, 292–297. [CrossRef] [PubMed]
83. Mattisson, J.; Danielsson, M.; Hammond, M.; Davies, H.; Gallant, C.J.; Nordlund, J.; Raine, A.; Edén, M.; Kilander, L.; Ingelsson, M.; et al. Leukocytes with chromosome Y loss have reduced abundance of the cell surface immunoprotein CD99. *Sci. Rep.* **2021**, *11*, 15160. [CrossRef] [PubMed]
84. Forsberg, L.A.; Rasi, C.; Malmqvist, N.; Davies, H.; Pasupulati, S.; Pakalapati, G.; Sandgren, J.; Diaz de Ståhl, T.; Zaghlool, A.; Giedraitis, V.; et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **2014**, *46*, 624–628. [CrossRef]
85. Why loss of Y? A pan-cancer genome analysis of tumors with loss of Y chromosome—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/36874157/> (accessed on 7 February 2024).
86. Wu, Z.; Liu, X.; Liu, L.; Deng, H.; Zhang, J.; Xu, Q.; Cen, B.; Ji, A. Regulation of lncRNA expression. *Cell. Mol. Biol. Lett.* **2014**, *19*, 561–575. [CrossRef]
87. Johnsson, P.; Lipovich, L.; Grandér, D.; Morris, K.V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta (BBA) Gen. Subj.* **2014**, *1840*, 1063–1071. [CrossRef]
88. Ruiz-Orera, J.; Villanueva-Cañas, J.L.; Albà, M.M. Evolution of new proteins from translated sORFs in long non-coding RNAs. *Exp. Cell Res.* **2020**, *391*, 111940. [CrossRef]
89. Bánfai, B.; Jia, H.; Khatun, J.; Wood, E.; Risk, B.; Gundling, W.E.; Kundaje, A.; Gunawardena, H.P.; Yu, Y.; Xie, L.; et al. Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **2012**, *22*, 1646–1657. [CrossRef]
90. Ma, L.; Bajic, V.B.; Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **2013**, *10*, 925–933. [CrossRef] [PubMed]
91. ASMTL-AS1 ASMTL Antisense RNA 1 [Homo Sapiens (Human)]—Gene—NCBI. Available online: <https://www.ncbi.nlm.nih.gov/gene/80161> (accessed on 7 February 2024).
92. Perez, C.A.G.; Adachi, S.; Nong, Q.D.; Adhitama, N.; Matsuura, T.; Natsume, T.; Wada, T.; Kato, Y.; Watanabe, H. Sense-overlapping lncRNA as a decoy of translational repressor protein for dimorphic gene expression. *PLoS Genet.* **2021**, *17*, e1009683. [CrossRef] [PubMed]
93. Ransohoff, J.D.; Wei, Y.; Khavari, P.A. The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 143–157. [CrossRef] [PubMed]

94. Brownmiller, T.; Juric, J.A.; Ivey, A.D.; Harvey, B.M.; Westemeier, E.S.; Winters, M.T.; Stevens, A.M.; Stanley, A.N.; Hayes, K.E.; Sprowls, S.A.; et al. Y Chromosome LncRNA Are Involved in Radiation Response of Male Non-Small Cell Lung Cancer Cells. *Cancer Res.* **2020**, *80*, 4046–4057. [[CrossRef](#)] [[PubMed](#)]
95. Tahira, A.C.; Kubrusly, M.S.; Faria, M.F.; Dazzani, B.; Fonseca, R.S.; Maracaja-Coutinho, V.; Verjovski-Almeida, S.; Machado, M.C.C.; Reis, E.M. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Mol. Cancer* **2011**, *10*, 141. [[CrossRef](#)] [[PubMed](#)]
96. Johansson, M.M.; Pottmeier, P.; Suci, P.; Ahmad, T.; Zaghlool, A.; Halvardson, J.; Darj, E.; Feuk, L.; Peuckert, C.; Jazin, E. Novel Y-Chromosome Long Non-Coding RNAs Expressed in Human Male CNS During Early Development. *Front. Genet.* **2019**, *10*, 891. [[CrossRef](#)] [[PubMed](#)]
97. Molina, E.; Chew, G.S.; Myers, S.A.; Clarence, E.M.; Eales, J.M.; Tomaszewski, M.; Charchar, F.J. A Novel Y-Specific Long Non-Coding RNA Associated with Cellular Lipid Accumulation in HepG2 cells and Atherosclerosis-related Genes. *Sci. Rep.* **2017**, *7*, 16710. [[CrossRef](#)] [[PubMed](#)]
98. Xiao, G.; Yao, J.; Kong, D.; Ye, C.; Chen, R.; Li, L.; Zeng, T.; Wang, L.; Zhang, W.; Shi, X.; et al. The Long Noncoding RNA TTTY15, Which Is Located on the Y Chromosome, Promotes Prostate Cancer Progression by Sponging let-7. *Eur. Urol.* **2019**, *76*, 315–326. [[CrossRef](#)]
99. Zhou, M.; Zhao, H.; Wang, Z.; Cheng, L.; Yang, L.; Shi, H.; Yang, H.; Sun, J. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. *J. Exp. Clin. Cancer Res.* **2015**, *34*, 102. [[CrossRef](#)]
100. Wu, S.; Zhang, L.; Deng, J.; Guo, B.; Li, F.; Wang, Y.; Wu, R.; Zhang, S.; Lu, J.; Zhou, Y. A Novel Micropeptide Encoded by Y-Linked LINC00278 Links Cigarette Smoking and AR Signaling in Male Esophageal Squamous Cell Carcinoma. *Cancer Res.* **2020**, *80*, 2790–2803. [[CrossRef](#)]
101. Yang, C.; Cao, H.; Yang, J.-W.; Wang, J.-T.; Yu, M.-M.; Wang, B.-S. The ETS1-LINC00278 negative feedback loop plays a role in COL4A1/COL4A2 regulation in laryngeal squamous cell carcinoma. *Neoplasia* **2022**, *69*, 841–858. [[CrossRef](#)] [[PubMed](#)]
102. Soh, Y.Q.S.; Alföldi, J.; Pyntikova, T.; Brown, L.G.; Graves, T.; Minx, P.J.; Fulton, R.S.; Kremitzki, C.; Koutseva, N.; Mueller, J.L.; et al. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **2014**, *159*, 800–813. [[CrossRef](#)]
103. Calabrese, E.; Villanueva-Meyer, J.E.; Cha, S. A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas. *Sci. Rep.* **2020**, *10*, 11852. [[CrossRef](#)] [[PubMed](#)]
104. Osinski, B.L.; BenTaieb, A.; Ho, I.; Jones, R.D.; Joshi, R.P.; Westley, A.; Carlson, M.; Willis, C.; Schleicher, L.; Mahon, B.M.; et al. Artificial intelligence-augmented histopathologic review using image analysis to optimize DNA yield from formalin-fixed paraffin-embedded slides. *Mod. Pathol.* **2022**, *35*, 1791–1803. [[CrossRef](#)]
105. Sellami, M.; Elrayess, M.A.; Puce, L.; Bragazzi, N.L. Molecular Big Data in Sports Sciences: State-of-Art and Future Prospects of OMICS-Based Sports Sciences. *Front. Mol. Biosci.* **2022**, *8*, 815410. [[CrossRef](#)] [[PubMed](#)]
106. Ahmetov, I.I.; Egorova, E.S.; Gabdrakhmanova, L.J.; Fedotovskaya, O.N. Genes and Athletic Performance: An Update. *Med. Sport. Sci.* **2016**, *61*, 41–54. [[CrossRef](#)] [[PubMed](#)]
107. Bray, M.S.; Hagberg, J.M.; Pérusse, L.; Rankinen, T.; Roth, S.M.; Wolfarth, B.; Bouchard, C. The human gene map for performance and health-related fitness phenotypes: The 2006–2007 update. *Med. Sci. Sports Exerc.* **2009**, *41*, 35–73. [[CrossRef](#)] [[PubMed](#)]
108. Bouchard, C.; Rankinen, T.; Timmons, J.A. Genomics and genetics in the biology of adaptation to exercise. *Compr. Physiol.* **2011**, *1*, 1603–1648. [[CrossRef](#)]
109. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)]
110. Ulitsky, I.; Bartel, D.P. lincRNAs: Genomics, evolution, and mechanisms. *Cell* **2013**, *154*, 26–46. [[CrossRef](#)] [[PubMed](#)]
111. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [[CrossRef](#)] [[PubMed](#)]
112. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **2012**, *337*, 1190–1195. [[CrossRef](#)] [[PubMed](#)]
113. Mattick, J.S.; Amaral, P.P.; Carninci, P.; Carpenter, S.; Chang, H.Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M.E.; Fitzgerald, K.A.; et al. Long non-coding RNAs: Definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **2023**, *24*, 430–447. [[CrossRef](#)] [[PubMed](#)]
114. Sironi, M.; Hasnain, S.E.; Rosenthal, B.; Phan, T.; Luciani, F.; Shaw, M.-A.; Sallum, M.A.; Mirhashemi, M.E.; Morand, S.; González-Candelas, F.; et al. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. *Infect. Genet. Evol.* **2020**, *84*, 104384. [[CrossRef](#)] [[PubMed](#)]
115. Mercatelli, D.; Giorgi, F.M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **2020**, *11*, 1800. [[CrossRef](#)] [[PubMed](#)]
116. Zong, Y.; Li, X. Identification of Causal Genes of COVID-19 Using the SMR Method. *Front. Genet.* **2021**, *12*, 690349. [[CrossRef](#)]
117. Wu, P.; Ding, L.; Li, X.; Liu, S.; Cheng, F.; He, Q.; Xiao, M.; Wu, P.; Hou, H.; Jiang, M.; et al. Trans-ethnic genome-wide association study of severe COVID-19. *Commun. Biol.* **2021**, *4*, 1034. [[CrossRef](#)] [[PubMed](#)]

118. Badr, E.A.; El Sayed, I.E.; Gabber, M.K.R.; Ghobashy, E.A.E.; Al-Sehemi, A.G.; Algarni, H.; Elghobashy, Y.A. Are Antisense Long Non-Coding RNA Related to COVID-19? *Biomedicines* **2022**, *10*, 2770. [[CrossRef](#)]
119. Wu, Y.; Zhao, T.; Deng, R.; Xia, X.; Li, B.; Wang, X. A study of differential circRNA and lncRNA expressions in COVID-19-infected peripheral blood. *Sci. Rep.* **2021**, *11*, 7991. [[CrossRef](#)]
120. Dacht, F.; Brown, J.B.; Valyi-Nagy, T.; Narayan, K.D.; Serafini, A.; Boley, N.; Gingeras, T.R.; Celniker, S.E.; Mohapatra, G.; Loeb, J.A. Selective time-dependent changes in activity and cell-specific gene expression in human postmortem brain. *Sci. Rep.* **2021**, *11*, 6078. [[CrossRef](#)] [[PubMed](#)]
121. Kirchner, A.; Dacht, F.; Loeb, J.A. Identifying targets for preventing epilepsy using systems biology of the human brain. *Neuropharmacology* **2020**, *168*, 107757. [[CrossRef](#)] [[PubMed](#)]
122. Lipovich, L.; Dacht, F.; Cai, J.; Bagla, S.; Balan, K.; Jia, H.; Loeb, J.A. Activity-dependent human brain coding/noncoding gene regulatory networks. *Genetics* **2012**, *192*, 1133–1148. [[CrossRef](#)] [[PubMed](#)]
123. Kirchner, A.; Dacht, F.; Lipovich, L.; Loeb, J.A. Activity-Dependent Non-Coding RNA MAPK Interactome of the Human Epileptic Brain. *Noncoding RNA* **2023**, *9*, 3. [[CrossRef](#)]
124. Barkmeier, D.T.; Senador, D.; Leclercq, K.; Pai, D.; Hua, J.; Boutros, N.N.; Kaminski, R.M.; Loeb, J.A. Electrical, molecular and behavioral effects of interictal spiking in the rat. *Neurobiol. Dis.* **2012**, *47*, 92–101. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.