Of mice and human-specific long noncoding RNAs

Amr R. Ghanam · William B. Bryant · Joseph M. Miano

Received: 12 October 2021 / Accepted: 10 January 2022 / Published online: 1 February 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022, corrected publication 2022

Abstract

The number of human LncRNAs has now exceeded all known protein-coding genes. Most studies of human LncRNAs have been conducted in cell culture systems where various mechanisms of action have been worked out. On the other hand, efforts to elucidate the function of human LncRNAs in an in vivo setting have been limited. In this brief review, we highlight some strengths and weaknesses of studying human LncRNAs in the mouse. Special consideration is given to bacterial artificial chromosome transgenesis and genome editing. The integration of these technical innovations offers an unprecedented opportunity to complement and extend the expansive literature of cell culture models for the study of human LncRNAs. Two different examples of how BAC transgenesis and genome editing can be leveraged to gain insight into human LncRNA regulation and function in mice are presented: the random integration of a vascular cell-enriched LncRNA and a targeted approach for a new LncRNA immediately upstream of the ACE2 gene, which encodes the receptor for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiologic agent underlying the coronavirus disease-19 (COVID-19) pandemic.

Approximately ninety-eight percent of our genome is non-coding. Contrary to initial descriptions of this vast sea of sequence comprising “junk DNA” (Ohno 1972), comparative genomics and various next-generation sequencing studies have revealed millions of transcription factor binding sites (TFBS) (Vierstra et al. 2020) and tens of thousands of noncoding genes, most notably the class of long non-coding RNAs (LncRNAs), defined currently as processed transcripts of length > 200 base pairs with no protein-coding capacity (Rinn and Chang 2020; Statello et al. 2021). The widespread transcription of LncRNAs and abundance of regulatory sequences such as enhancers support the concept of a genome that is largely functional (ENCODE Project Consortium 2012). Such a dynamic genome should not be surprising given the complex nature of gene expression and gene function necessary for embryonic and postnatal development as well as disease processes.

Unlike coding genes, which are ultimately translated into proteins with conserved domains predictive of function, most LncRNAs lack conserved sequence motifs that foretell biological utility. Consequently, the study of LncRNA genes has been challenging, with few examples of well-defined functions in an in vivo setting (Rinn and Chang 2020; Statello et al. 2021). At a minimum, mechanistic insight into the biological role of an LncRNA requires an understanding of (a) where the processed LncRNA accumulates in a cell (Kopp and Mendell 2018), (b) the molecular docking sites of an LncRNA for nucleic acid or protein association (McDonel and Guttman 2019), and (c) phenotypes (e.g., developmental, metabolic, transcriptomic) manifested following LncRNA loss-of-function in vivo (Sauvageau et al. 2013). It should be noted that in some cases, the mere act of transcribing the LncRNA confers functionality on the expression of an adjacent transcription unit, with the processed LncRNA perhaps having an independent role (Ali and Grote 2020; Anderson et al. 2016; Paralkar et al. 2016). Mature LncRNAs, or regulatory elements embedded within the LncRNA locus, may activate or repress local gene transcription (Gil and Ulitsky 2020). Further, a number of LncRNA loci are host genes for other genic units such as microRNAs that provide another level of finely-tuned gene expression (Sun et al. 2021).

Most wet-lab studies of human-specific LncRNAs are confined to cells in a dish. For example, a frequently reported role of human LncRNAs in vitro relates to their competition with mRNAs for microRNA binding. These so-called
competing endogenous RNAs fine-tune gene expression by “sponging” microRNAs that otherwise bind the 3’ untranslated region of an mRNA, targeting the mRNA for degradation. However, interpretation of most data ascribing a competing RNA function to LncRNAs is difficult in the absence of careful stoichiometric measures of the LncRNA, target mRNA, and associated microRNA (Denzler et al. 2014). Gene editing of a microRNA binding site (MREs) within an LncRNA represents a rigorous approach to invoke a competing endogenous RNA mechanism of action. Surprisingly, there are very few studies that target an endogenous MRE via editing tools such as CRISPR and none have yet to do so in the mouse (Bassett et al. 2014; Broughton et al. 2016; Ohtsuki et al. 2021). Given the expansive number of human-specific LncRNAs reported to function as competing endogenous RNAs, largely through standard luciferase assays that interrogate an MRE out of normal sequence context, there should be increased efforts to formally demonstrate the importance of an MRE in vivo through genome editing approaches (Wu et al. 2017). This is of particular interest since mammalian MREs may carry functionally relevant single-nucleotide polymorphisms (Miller et al. 2014).

Growth, migration, differentiation, and MRE functionality should be assayed in cell culture or organoid model systems to gain some foundational insight into the biology of human-specific LncRNAs. However, illuminating the function of human-specific LncRNAs in the complex milieu of a multisystem organism requires a combination of evolving technologies in mouse genetics and genome editing. Herein, some strengths and weaknesses of mouse transgenesis and genome editing are briefly summarized in the context of elucidating expression and regulation of LncRNAs. Two examples are then presented as to how specialized transgenesis, combined with genome editing, may afford important insight into the biological role of human-specific LncRNAs in the mouse.

Transgenic Human LncRNAs in Mice

Traditional approaches to study gene regulation and function in the mouse involve pronuclear injection of a cDNA encoding a protein or a reporter gene such as beta galactosidase under the control of a strong heterologous or cell-restricted promoter (Brinster et al. 1989). Transgenic mice carrying the human hepatitis C virus regulated 1 LncRNA exhibited deleterious expression of the mouse sterol regulatory element binding protein and reduced lipid metabolism (Li et al. 2017). In a similar fashion, overexpression of the human colon cancer associated transcript 2 LncRNA caused chromosomal instability with resultant myeloid malignancies (Shah et al. 2018). Although these examples offer some insight into the in vivo function of human LncRNAs, they are limited by the heterologous nature of the promoter driving widespread expression of the LncRNA. Moreover, even if the endogenous promoter were to have been utilized, distal regulatory regions may be absent from the transgene precluding full recapitulation of the LncRNA’s expression profile. To circumvent these constraints, artificial chromosome vectors have evolved to better capture all regulatory elements and avoid the ambiguity of a strong heterologous promoter which often directs supraphysiological levels of an LncRNA that otherwise exhibits low-level, cell compartment-specific expression.

The development of yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) vectors represented a significant advance in mouse transgenesis (Giraldo and Montoliu 2001; Heaney and Bronson 2006). Artificial chromosome vectors can harbor large (> 100 kilobases) sequences, thus enabling the integration of human transgenes that exceed the cloning capacity of conventional vectors into the mouse genome. In addition, the transgene within an artificial chromosome will contain most, if not all, regulatory sequences, including enhancers and insulators in their correct sequence context, ensuring proper spatiotemporal expression of the transgene (Long and Miano 2007). Relatively few human LncRNAs have been incorporated into the mouse genome through artificial chromosome transgenesis. The human X inactivation specific transcript, XIST, is 32 kilobases in length and the processed 19 kilobase transcript drives X chromosome dosage compensation in females through propagated hypoacetylation. The human XIST LncRNA was packaged in a 480 kilobase YAC for transfer into the mouse genome, and results revealed expression and X chromosome inactivation in the mouse, demonstrating the conservation of XIST function between human and mouse (Migeon et al. 1999). The imprinted human H19 LncRNA is a host gene for microRNA-675 (Cai and Cullen 2007). This LncRNA was studied in the context of a 100 kilobase artificial chromosome and found to be correctly expressed in the mouse, but incorrectly imprinted suggesting species-specific mechanisms for methylation-dependent repression of H19 (Jones et al. 2002). Using a BAC scanning reporter assay in mice, the human moesin pseudogene 1 antisense (MSNPS1AS) LncRNA was found to be expressed in cortex, striatum, and cerebellum, and expression was ascribed to enhancer regions that overlap a series of single-nucleotide polymorphisms implicated in autism spectrum disorder (ASD) (Inoue and Inoue 2016). These findings suggest that elevated levels of MSNPS1AS, shown recently to provoke neuronal phenotypes considered important in ASD (Luo et al. 2020), may occur through altered enhancer activities. Of note, the BAC transgenes under study contained the variants associated with ASD; however, expression levels of MSNPS1AS were not assessed in the context of a wild-type allele (Inoue and Inoue 2016).
While YAC/BAC integration of human lncRNAs has the advantage of native promoter and enhancer sequences for proper expression levels, pronuclear transgenes insert randomly in the genome, often as concatenators and sometimes in more than one locus, complicating the genotyping of mice homozygous for the transgene (Nakanishi et al. 2002). The emergence of PacBio and Oxford Nanopore Technologies sequencing platforms (Amarasinghe et al. 2020) allows for the determination of the site of transgene integration as well as transgene copy number, thus permitting facile breeding strategies to distinguish heterozygous from homozygous mice (Nicholls et al. 2019). These third-generation sequencing platforms will be of great utility in pinpointing the integration site of many of the 95% of reported transgenes that remain unmapped in mouse models (Nicholls et al. 2019). Another challenge to overcome with random integration of a BAC/YAC carrying an lncRNA is the possible disruption of coding or noncoding genic units or regulatory sequences such as enhancers or individual transcription factor binding sites (TFBS). The disruption of regulatory cassettes is of particular concern given widespread transcription of the genome and the presence of millions of predicted TFBS (Jensen et al. 2013; Vierstra et al. 2020). Beyond the obvious perturbation in local sequence topology, random insertion of a transgene can result in loss of host genome sequence with unpredictable consequences (Suzuki et al. 2020). Finally, phenotyping of mice could be confounded by disruption of a genic unit exhibiting haploinsufficiency. To circumvent these limitations, it should be possible to target a human lncRNA and associated coding gene/regulatory regions to the corresponding mouse region using a recombinase-mediated strategy wherein an entire mouse genomic region is swapped out for the orthologous human sequence (Devoy et al. 2011). This method of orthologous gene replacement has yet to be done in the context of a BAC-containing human lncRNA, though we shall introduce a potentially important candidate below. However, before introducing this idea, the power of genome editing of lncRNAs is summarized.

### Genome Editing of lncRNAs in Mice

The clustered regularly interspaced short palindromic repeat (CRISPR) platform of gene editing (Jinek et al. 2012) has forever transformed the development of genetically modified mouse models (Harms et al. 2014; Miano et al. 2016; Singh et al. 2015). Whereas germline transmission of a genetic modification in mice, using traditional embryonic stem cell targeting, can take a year or more (or never), a CRISPR edit enables germline transmission in a matter of just a few months (Miano et al. 2019). Since the initial reporting of CRISPR editing in mice (Shen et al. 2013), additional gene editing systems have been developed, including base editing and the very recent prime editing (Anzalone et al. 2020).

The absence of well-annotated functional motifs in most lncRNAs renders CRISPR targeting of this class of genes in the mouse challenging, though not insurmountable (Miano et al. 2019). Indeed, several lncRNAs have been targeted with CRISPR in rodents through large deletions of multiple exons or the entire lncRNA locus (Han et al. 2014; Zhou et al. 2021b; Zhuang et al. 2021). The approach of removing such large sequences runs the risk of deleting regulatory elements or small intronic RNAs that may compromise accurate interpretation of phenotypes. Alternatively, smaller deletions such as in the promoter region or a single exon of an lncRNA have been reported that minimize the risk of removing other functionally important sequences (Allou et al. 2021; Li et al. 2021; Saba et al. 2021). In addition, CRISPR-mediated insertion of a polyadenylation signal that arrests transcription of an lncRNA can be used to address the role of active transcription in lncRNA function (Allou et al. 2021; Anderson et al. 2016; Ballarino et al. 2018). An alternative approach to permanently silence transcription of an lncRNA is through strategic nucleotide substitutions across a key TFBS (Choi et al. 2020). Using the prime editing platform (Anzalone et al. 2019), a recent study showed that a single-nucleotide substitution in a TFBS nearly extinguished expression of an lncRNA. Interestingly, this single base change also nullified the expression of a divergently transcribed protein-coding gene (Gao et al. 2021). The latter finding highlights the need for careful deliberation over the specific strategy implemented in gene editing of an lncRNA in mice (Miano et al. 2019). For example, there could be a TFBS embedded inside the lncRNA locus that controls the expression of another locus independent of the transcribed lncRNA (Ali and Grote 2020). As of this writing, there has been no report of the editing of a human-specific lncRNA in mice. Below, we introduce two examples of human-specific lncRNA integration in the mouse and how genome editing may unveil important regulatory and functional features of each lncRNA.

### A Humanized Mouse Model for SENCr

The smooth muscle and endothelial cell-enriched migration/differentiation-associated long Non-Coding RNA (SENCr, pronounced sen-sor) was first reported in early 2014 from an RNA-seq study of human coronary artery smooth muscle cells (Bell et al. 2014). This 3-exon lncRNA overlaps the 5’ end of Friend Leukemia Integration 1 (FLI1), a member of the E26 transformation specific family of DNA-binding transcription factors. SENCr and FLI1 display similar patterns of tissue-specific RNA expression (Fig. 1). However, data thus far suggest that the RNA expression of one is
independent of the other (Bell et al. 2014). Further, whereas FLI1 is a nuclear transcription factor, most SENCR transcripts are cytoplasmic suggesting each gene product exerts distinct functions (Bell et al. 2014). Knockdown studies combined with RNA-seq revealed functions of SENCR related to the maintenance of a non-motile, differentiated smooth muscle cell phenotype (Bell et al. 2014). A subsequent study demonstrated SENCR to promote the commitment of human embryonic stem cells to an endothelial cell lineage (Boulberdaa et al. 2016). SENCR also facilitated endothelial cell proliferation and migration, key processes in angiogenesis (Boulberdaa et al. 2016). In this context, patients with critical limb ischemia or premature coronary artery disease showed reduced levels of SENCR in ischemic tissue or in endothelial cells derived from blood vessels, respectively (Boulberdaa et al. 2016). The latter report provided some intriguing insight into the in vivo function of SENCR. However, these proposed functions and others require validation and further study of SENCR in an animal model.

To date, there has been no compelling evidence for a mouse ortholog of human SENCR. CRISPR-directed SENCR deletion studies in an immortalized human endothelial cell line (EA.hy926 cells) were thwarted by the presence of four copies of the host chromosome 11 (unpublished). However, the in vivo function of SENCR could be revealed by its introduction into the mouse genome, with the assumption that spatial expression and function of SENCR in the mouse would mirror SENCR expression and function in the human body. To begin to address these important points, a recent study reported the integration of a 217 kilobase BAC harboring the entire human FLI1 and SENCR genes into the mouse using the piggyBAC transposase system of transgene integration (Lyu et al. 2019). Studies in cultured human endothelial cells revealed an increase in SENCR expression under laminar flow conditions, which approximated the biophysical forces endothelial cells encounter with blood flow in vivo (Lyu et al. 2019). Notably, immuno-RNA fluorescence in situ hybridization experiments disclosed expected increases in SENCR expression where laminar flow conditions exist across the aortic arch of the humanized mouse model (Lyu et al. 2019). These results demonstrated the utility of studying proposed functions of SENCR as a mediator of smooth muscle and endothelial cell homeostasis in vivo.

Fig. 1 Tissue RNA profile of SENCR and FLI1. Data obtained from the GTEx portal website (https://www.gtexportal.org/home/)
In addition, the opportunity now exists to uncouple FLI1 and SENCR through BAC editing in the background of a FLI1 null mouse. Since genetic loss of FLI1 is embryonic lethal (Spyropoulos et al. 2000), the expectation is human FLI1 will rescue the lethal phenotype. One important caveat to the BAC editing of the FLI1/SENCR human transgene is the need for a single-copy BAC transgene. The piggyBAC system for in vivo BAC integration supports a single-copy integration event (Jung et al. 2016). However, transgene copy number and the site of integration will require third-generation sequencing platforms (Amarasinghe et al. 2020) to determine the suitability for BAC editing and the breeding of heterozygous mice to homozygosity for gene dosage effects. As discussed next, targeting a single-copy human LncRNA-mRNA gene pair to a defined locus obviates the need for such mapping studies.

**An ACE2-LncRNA Gene Pair and Development of a New Mouse Model for COVID-19**

Over the last two years, Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2), the etiologic agent underlying the COronaVIrus Disease-2019 (COVID-19) pandemic, has ravaged the world, precipitating economical, sociological, and political upheaval as well as an unprecedented ‘infodemic’ that has hampered efforts to disseminate scientific facts regarding SARS-CoV-2 infection, COVID-19, and the vaccination campaign (Tentolouris et al. 2021). Moreover, health care systems around the world have been overstrained, making prioritization of health care delivery ever-challenging. Cumulatively, as of January 27, 2022, the COVID-19 pandemic has resulted in 363,582,071 positive cases of SARS-CoV-2 infection and 5,629,317 deaths, 15% of which have occurred in the United States (https://corona virus.jhu.edu/map.html).

The receptor mediating SARS-CoV-2 entry into human cells is angiotensin-converting enzyme 2 (ACE2) (Zhou et al. 2020). There are at least three isoforms of human ACE2, spanning ~41 kilobases of DNA on the X chromosome, each of which appears to be under control of distinct promoters (Fig. 2). The longest isoform of ACE2 comprises 19 exons, encoding an 805 amino acid protein. A slightly shorter isoform of ACE2 exists, encoding the same number of amino acids (Fig. 2). High-level expression of ACE2 protein is seen in human small intestine and kidney (Fig. 3A). Several human cell lines also express ACE2 protein, though levels of ACE2 are undetectable in vascular smooth muscle cells and endothelial cells (Fig. 3A). The latter cell type has been the focus of numerous studies given the mounting evidence for SARS-CoV-2-induced endotheliopathy, considered an important contributor to the pathogenesis of COVID-19 (Goshua et al. 2020). The undetectable levels of ACE2 protein in human endothelial cells shown here is consistent with a recent report that failed to detect ACE2 mRNA in several human endothelial cell types (McCracken et al. 2021), but inconsistent with other reports (Hamming et al. 2004; Targosz-Korecka et al. 2021; Wagner et al. 2021). These disparate findings highlight the ongoing controversy over whether endothelial cells are prone to SARS-CoV-2 infection (Goldsmith et al. 2020; McCracken et al. 2021; Targosz-Korecka et al. 2021; Varga et al. 2020; Wagner et al. 2021).

In addition to the two long isoforms of ACE2, there is at least one shorter isoform (Fig. 2). This shorter deltaACE2 (dACE2) isoform is elevated following interferon stimulation of several human cell lines, including nasal epithelial cells (Onabajo et al. 2020). Similar induction of the dACE2 isoform is observed upon stimulation of Caco-2 cells (immortal colorectal adenocarcinoma cell line) with interferon alpha, interferon beta, or interferon gamma (unpublished). The dACE2 isoform lacks ~350 N-terminal amino acids and does not bind SARS-CoV-2 (Onabajo et al. 2020).

Interestingly, a non-overlapping antisense LncRNA, designated GS1-594A7.3, resides just upstream of the human ACE2 locus (Fig. 2). This LncRNA, which is poorly conserved across vertebrate species (Fig. 2), is only 722 base pairs upstream of the longest ACE2 isoform, suggesting the ACE2-GS1-594A7.3 mRNA-LncRNA gene pair may share a common promoter. Evidence in support of such a bifunctional promoter exists with the partial overlap in RNA expression of ACE2 and GS1-594A7.3 across human tissues (Fig. 4). Rapid amplification of cDNA ends and

---

**Fig. 2** ACE2 locus. Modified UCSC Genome Browser (http://genome.ucsc.edu/) screenshot showing the three isoforms of ACE2 and the upstream GS1-594A7.3 LncRNA. Vertebrate conservation (bottom green track) reveals conservation of ACE2 coding exons, but a notable lack of conservation in the two exons of GS1-594A7.3 (cream-colored rectangles overlapping exons)
long range qRT-PCR validated the annotation of GS1-594A7.3 as an independently transcribed LncRNA (unpublished). Of intriguing importance is the finding that the GS1-594A7.3 LncRNA is confined largely to the nucleus of several human cell lines (Fig. 3B–C). This observation suggests that GS1-594A7.3 possesses the potential to regulate ACE2 levels in cis. However, repeated attempts to CRISPR edit this LncRNA in cultured cells have been unsuccessful, likely because of the known difficulties in establishing stable cell lines in Caco-2 and Calu-3 cells and their state of aneuploidy.

X-ray crystallographic analysis of the receptor binding domain of SARS-CoV-2 bound to human ACE2 (Lan et al. 2020) revealed critical contact residues that are not conserved in the mouse ACE2 protein, rendering mice resistant to SARS-CoV-2 infection and disease (Lan et al. 2020). Accordingly, several humanized ACE2 mouse models for SARS-CoV-2 infection and COVID-19 exist (Lutz et al. 2020). Most of these mouse models were generated through pronuclear transgenesis (Table 1). As discussed earlier, limitations of mouse transgenesis include the unknown site of integration and copy number of transgene. Moreover, the majority of humanized ACE2 mouse models utilize chimeric or cell-specific promoters that likely do not fully recapitulate the human ACE2 pattern of expression in humans (Table 1), though at least one of these models has proved useful for testing vaccines and therapeutics (Chen et al. 2021; Hoffmann et al. 2021). To control the inherent limitations of transgenesis and more closely approximate the endogenous expression profile of human ACE2, two models targeted exon 2 of the endogenous mouse Ace2 locus with a human ACE2 cDNA (Table 1). These knockin models not only safeguard against multiple copies of the transgene, but also take advantage of the mouse Ace2 regulome, thus better modeling the true spatiotemporal pattern of ACE2 protein expression. However, there may be differences between promoter/enhancer sequences in the mouse Ace2 regulome versus the human ACE2 regulome. Moreover, the GS1-594A7.3
LncRNA appears to be a human-specific LncRNA as there is no similarly arranged LncRNA in the mouse, and analysis of sequencing data around the 5’ region of mouse Ace2 has failed to reveal transcription of an LncRNA. Since there presently is no evidence for a mouse Ace2-associated LncRNA, humanized BAC transgenic studies, as described above for the SENCR LncRNA, offer a unique opportunity to assess the expression and function of GS1-594A7.3 in the mouse.

Beginning in the summer of 2020, this lab set out to develop a new humanized ACE2 mouse model in order to capture the entire human ACE2 locus (Fig. 2) as well as the upstream GS1-594A7.3 LncRNA. However, rather than risk random integration of the BAC harboring the ACE2-GS1-594A7.3 mRNA-LncRNA gene pair (BAC clone CTD-2522M16), a different strategy was used. The basic approach involves the swapping in of the entire human ACE2 locus for the mouse Ace2 locus. A CRISPR-mediated method has been used for targeting large sequences, such as BACs, to define gene loci in the rat genome (Yoshimi et al. 2016). An alternative method, and the one we adopted, uses recombinase-mediated genomic replacement (RMGR) in mouse embryonic stem cells, which are then implanted into the blastocyst for generation of chimeric mice (Wallace et al. 2007). In this model, all human ACE2 coding exons and non-coding introns are present in their proper sequence context, allowing for all isoforms, including the interferon-induced dACE2 (Fig. 2), to be expressed. Important validations are required, including correct spatiotemporal expression of ACE2 mRNA and ACE2 protein using molecular probes and scRNA-seq; susceptibility of mice to SARS-CoV-2 infection and attending pathology in the lung, blood, intestinal tract, and brain; phenotyping homozygous ACE2 mice for evidence of developmental defects, altered blood pressure regulation or behavioral deficits due to loss of the mouse Ace2 locus and unannotated critical noncoding genes that are unable to be rescued by the human ACE2 locus; and, most importantly, the expression and localization of the GS1-594A7.3 LncRNA.

Beyond the targeting of a single-copy transgene to a defined locus, there are several advantages to this more fully humanized ACE2 mouse model. First, the definitive
role of the upstream GS1-594A7.3 LncRNA can be studied with genome editing, either through deletion of the entire LncRNA locus, insertion of a polyadenylation signal sequence in the first exon, or more subtle editing of a TFBS as reported in other mouse models of LncRNA regulation (Choi et al. 2020; Gao et al. 2021). The hypothesis would be that loss of GS1-594A7.3 LncRNA will alter normal expression of human ACE2, rendering mice either more or less susceptible to SARS-CoV-2 infection and COVID-like symptoms. Second, a more representative human ACE2 expression profile would likely reflect the nuanced expression of this receptor, especially under conditions that model human comorbidities (e.g., type 2 diabetes, hypertension, and obesity), where the risk for severe COVID-associated pathology and death is high. Third, mechanisms underlying so-called long COVID (Nalbandian et al. 2021) may be illuminated with the correct spatiotemporal expression of human ACE2 and multi-system infection and pathology; the increasingly problematic ‘long COVID’ has been barely touched upon in in vivo setting as well as neuronal- and circulatory-derived inputs. To circumvent these limitations, whole organ or human embryonic stem cell-derived organoid model systems have been developed to interrogate human LncRNAs. For example, the human-specific LncRNA, SMILR, was investigated in organ cultures of human saphenous vein grafts to define its role in mediating smooth muscle cell proliferation (Mahmoud et al. 2019). Meanwhile, the PAUPAR LncRNA was studied in human organoids and shown to regulate cortical differentiation (Xu et al. 2021). These ex vivo model systems represent a higher order level of investigation over simple, two-dimensional cell culture models. In order to realize whether what is observed in vitro or in organ culture models applies to a complex living animal, humanized BAC rodent models offer another level of exploration.

Challenges, Limitations, and Alternative Approaches to Humanized BAC Mice

The study of human-specific LncRNAs has been confined mainly to cell culture models. However, most cell culture systems are either transformed or phenotypically altered with poor reproduction of their natural in vivo state. Further, cells in a dish lack correct integration with neighboring cell types encountered in an in vivo setting as well as neuronal- and circulatory-derived inputs. To circumvent these limitations, whole organ or human embryonic stem cell-derived organoid model systems have been developed to interrogate human LncRNAs. For example, the human-specific LncRNA, SMILR, was investigated in organ cultures of human saphenous vein grafts to define its role in mediating smooth muscle cell proliferation (Mahmoud et al. 2019). Meanwhile, the PAUPAR LncRNA was studied in human organoids and shown to regulate cortical differentiation (Xu et al. 2021). These ex vivo model systems represent a higher order level of investigation over simple, two-dimensional cell culture models. In order to realize whether what is observed in vitro or in organ culture models applies to a complex living animal, humanized BAC rodent models offer another level of exploration.

To be sure, there are several limitations and challenges with humanized BAC transgenic mouse experiments. First, BAC transgenesis, whether via pronuclear injection or RMGR, requires highly skilled methods of handling and delivery into the mouse genome with no guarantee of targeting or germline transmission. Beyond academic cores, commercial vendors can perform these genetic manipulations, typically at a cost >$30,000. Second, some LncRNAs (e.g., the 363 kilobase STXBP5-AS1) exceed the cloning capacity of BAC vectors, thus requiring larger

| Transgenic Reference | PMID |
|----------------------|------|
| Tseng et al. (2007)   | 17108019 |
| Menachery et al. (2016) | 26976607 |
| Asaka et al. (2021)  | 34463644 |

| Knock In Reference | PMID |
|-------------------|------|
| Sun et al. (2020) | 32485164 |
| Zhou et al. (2021a) | 33636719 |
cloning capacity vectors such as YACs (see above). The latter limitation serves as a reminder that annotation of many LncRNAs may be incomplete with rapid amplification of cDNA ends needed to fully extend the LncRNA transcript at both the 5' and 3' ends (Freedman and Miano 2017). Third, phenotypic analysis of a mouse carrying a human LncRNA can be challenging if insertion of the BAC disrupts a critical regulatory or coding sequence or if human-specific sequences such as enhancers or other genic units within the BAC create an unrelated phenotype to that of the LncRNA. Fourth, BAC models of human LncRNAs may confer phenotypes not easily discerned in the mouse (e.g., cognitive functions). Fifth, human LncRNAs may not fully recapitulate their spatiotemporal pattern expression profile in the mouse, due to the absence of human-specific regulatory cassettes or cofactors. Finally, the random, multicopy integration of BAC transgenes in the mouse requires mapping analysis using, for example, third-generation sequencing platforms, in order to optimize breeding schedules and learn of any potential genetic confounders such as disruption of a protein-coding gene or regulatory sequence. Where conserved LncRNAs exist, we suggest replacement of a mouse locus with the orthologous human sequence using RMGR, as described for the ACE2-GS1-594A7.3 mRNA-LncRNA gene pair, as an alternative approach to pronuclear injection of a BAC for the study of human LncRNA expression regulation and function in the mouse. In addition to a single integration event at a known genomic location, thus facilitating genotyping of heterozygous intercrosses for the generation of homozygous animals, RMGR renders the mouse more amenable to genome editing strategies (Fig. 5). The development of new mouse models, coupled with genome editing, holds promise for advancing our understanding of the expression and function of human LncRNAs under normal and pathological conditions.

**Acknowledgements**  LncRNA and mouse editing work is supported by NIH grants HL138987, HL136224, and HL147476 as well as institutional support. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

**Declarations**

**Conflict of interest**  On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Data transparency and distribution**  This manuscript alludes to unpublished data that we can make available upon written request.

---

**Fig. 5** Schematic of two methods of generating humanized BAC transgenic mice. Hypothetical LncRNA-mRNA gene pair within a human BAC (top). At least two methods exist for incorporating a BAC-containing LncRNA transgene into the mouse genome (arrows). One involves standard pronuclear injection with attending random integration, often as multiple copies (right arrow). An alternative method is recombinase-mediated genomic replacement (RMGR) (left arrow). Features of each method are indicated at bottom. See text for details.
References

Ali T, Grote P (2020) Beyond the RNA-dependent function of LncRNA genes. Elife 9

Allou L, Balzano S, Magg A, Quinodoz M, Royer-Bertrand B, Schopfli R et al (2021) Non-coding deletions identify Maenli IncRNA as a limb-specific Enl regulator. Nature 592:93–98

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouli Q (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21:30

Anderson KM, Anderson DM, McNally JR, Shelton JM, Bassel-Duby R, Olson EN (2016) Transcription of the non-coding RNA upper-hand controls Hand2 expression and heart development. Nature 539:433–436

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM et al (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576:149–157

Anzalone AV, Koblan LW, Liu DR (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol 38:824–844

Asaka MN, Utsumi D, Kamada H, Nagata S, Nakachi Y, Yamaguchi T et al (2021) Highly susceptible SARS-CoV-2 model in CAG promoter-driven hACE2 transgenic mice. JCI Insight 6:e152529

Bakhshandeh B, Sorboni SG, Javanmard AR, Mottaghi SS, Mehrabi MR, Sororui F et al (2021) Variants in ACE2; potential influences on virus infection and COVID-19 severity. Infect Genet Evol 90:104773

Ballarino M, Cipriano A, Tita R, Santini F, Desideri F, Morlando et al (2018) Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. EMBO J 37:e99697

Bassett AR, Azzam G, Wheatley L, Titbit C, Rajakumar T, Ballarino M, Cipriano A, Tita R, Santini T, Desideri F, Morlando et al (2018) Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. EMBO J 37:e99697

Bassett AR, Azzam G, Wheatley L, Titbit C, Rajakumar T, McGowan S et al (2014) Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. Nat Commun 5:4640

Bell RD, Long X, Lin M, Bergmann JH, Newby GA, Zhang W et al (2021) Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression. Genome Biol 22:83

Gil N, Ulitsky I (2020) Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev Genet 21:102–117

Giraldo P, Montoliu L (2001) Size matters: use of YACs, BACs, and PACs in transgenic animals. Transgenic Res 10:83–103

Goldsmith CS, Miller SE, Martines RB, Bullock HA, Zaki SR (2020) Electron microscopy of SARS-CoV-2: a challenging task. Lancet 395:e99

Gosuha G, Pine AB, Meizlish ML, Chang CH, Zhang H, Bahel P et al (2020) Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study. Lancet Haematol 7:e575–e582

Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H (2004) Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol 203:631–637

Han J, Zhang J, Chen L, Shen B, Zhou J, Hu B et al (2014) Efficient in vivo detection of a large imprinted IncRNA by CRISPR/Cas9. RNA Biol 11:829–835

Harms DW, Quadros RM, Seruggia D, Ohtsuka M, Takahashi G, Montoliu L et al (2014) Mouse genome editing using the CRISPR/Cas system. Curr Protoc Hum Genet 83:15.17.11-27

Heaney JD, Bronson SK (2006) Artificial chromosome-based transgenes in the study of genome function. Mamm Genome 17:799–807

Hoffmann D, Corleis B, Rauch S, Roth N, Muhe J, Halwe NJ et al (2021) CVαCoV and CV2CoV protect human ACE2 transgenic mice from ancestral B.1.529 and emerging B.1.351 SARS-CoV-2. Nat Commun 12:4048

Inoue YU, Inoue T (2016) Brain enhancer activities at the gene-poor TALEN mediated BAC transgenesis in the zygote for the generation of humanized SIRPA rats. Sci Rep 6:31455

Inoue YU, Inoue T (2016) Brain enhancer activities at the gene-poor TALEN mediated BAC transgenesis in the zygote for the generation of humanized SIRPA rats. Sci Rep 6:31455

Jensen TH, Jacquier A, Libri D (2013) Dealing with pervasive transcription. Mol Cell 52:473–484

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

Jones BK, Levorse J, Tilghman SM (2002) A human H19 transgene exhibits impaired paternal-specific imprint acquisition and maintenance in mice. Hum Mol Genet 11:411–418

Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. Cell 172:393–407

Krajcovicova J, Miskova A, Cermakova J, Balcerova B, Sotakova I, Mihalikova et al (2016) Comparative analysis of piggyBac, CRISPR/Cas9 and TALEN mediated BAC transgenesis in the zygote for the generation of humanized SRIPa rats. Sci Rep 6:31455

Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. Cell 172:393–407

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S et al (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215–220

Li D, Cheng M, Niu Y, Chi X, Liu X, Fan J et al (2017) Identification of a novel human long non-coding RNA that regulates hepatic lipid metabolism by inhibiting SREBP-1c. Int J Biol Sci 13:349–357

Devoy A, Bunton-Stasyszyn RK, Tybulewicz VL, Smith AJ, Fisher EM (2011) Genomically humanized mice: technologies and promises. Nat Rev Genet 13:14–20

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74

Feng Y, Xia H, Cai Y, Halabi CM, Becker LK, Santos RA et al (2010) Brain-selective overexpression of human Angiotensin-converting enzyme type 2 attenuates neurogenic hypertension. Circ Res 106:373–382

Freedman JE, Miano JM (2017) Challenges and opportunities in linking long noncoding RNAs to cardiovascular, lung, and blood diseases. Arterioscler Thromb Vasc Biol 37:21–25

Gao P, Lyu Q, Ghanam AR, Lazzarotto CR, Newby GA, Zhang W et al (2021) Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression. Genome Biol 22:83

290 A. R. Ghanam et al.
Li PY, Li SQ, Gao SG, Dong DY (2021) CRISPR/Cas9-mediated gene editing on Sox2ot promoter leads to its truncated expression and does not influence neural tube closure and embryonic development in mice. Biochem Biophys Res Commun 573:107–111

Long X, Miano JM (2007) Remote control of gene expression. J Biol Chem 282:15941–15945

Luo T, Ou JN, Cao LF, Peng QX, Li YM, Tian YQ (2020) The autism-related lncRNA MSNPIA1 regulates Moesin protein to influence the RhoA, Rac1, and PI3K/Akt pathways and regulate the structure and survival of neurons. Autism Res 13:2073–2082

Lutz C, Maher L, Lee C, Kang W (2020) COVID-19 preclinical models: human angiotensin-converting enzyme 2 transgenic mice. Hum Genomics 14:20

Lyu Q, Xu S, Lyu Y, Choi M, Christie CK, Slivano OI et al (2019) SENCr stabilizes vascular endothelial cell adhesions junctions through interaction with CKAP4. Proc Nat Acad Sci USA 116:546–555

Mahmoud AD, Ballantyne MD, Miscianinov V, Pinel K, Hung J, Scanlon JP et al (2019) The human-specific and smooth muscle cell-enriched LncRNA SMILR promotes proliferation by regulating mitotic CENPF mRNA and drives cell-cycle progression which can be targeted to limit vascular remodeling. Circ Res 125:535–551

McCracken IR, Saginc G, He L, Huseynov A, Daniels A, Fletcher S et al (2021) Lack of evidence of angiotensin-converting enzyme 2 expression and replicative infection by SARS-CoV-2 in human endothelial cells. Circulation 143:865–868

McCray PB Jr, Pewe L, Wohlford-Lenane C, Hickey M, Manzel L, Shi L et al (2007) Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. J Virol 81:813–821

McDonel P, Guttmann M (2019) Approaches for understanding the mechanisms of long noncoding RNA regulation of gene expression. Cold Spring Harb Perspect Biol 11

Menachery VD, Yount BL Jr, Sims AC, Debbink K, Agnihothram SS, Gralinski LE et al (2016) SARS-like WIV1-CoV poised for human emergence. Proc Natl Acad Sci U S A 113:3048–3053

Miano JM, Zhu QM, Lowenstein CJ (2016) A CRISPR path to engineering new genetic mouse models for cardiovascular research. Arterioscler Thromb Vasc Biol 36:1058–1075

Miano JM, Long X, Lyu Q (2019) CRISPR links to long noncoding RNA function in mice: a practical approach. Vascul Pharmacol 114:1–12

Migeon BR, Kazi E, Haisley-Royster C, Hu J, Reeves R, Call L et al (1999) Human X inactivation center induces random X chromosome inactivation in male transgenic mice. Genomics 59:113–121

Miller CL, Haas U, Diaz R, Leeper NJ, Kundra RK, Patlolla B et al (2014) Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. PLoS Genet 10:e1004263

Nadarajah R, Milagres R, Dilauro M, Gutsol A, Xiao F, Zimpelmann J et al (2012) Podocyte-specific overexpression of human angiotensin-converting enzyme 2 attenuates diabetic nephropathy in mice. Kidney Int 82:292–303

Nakanishi T, Kuroiwa A, Yamada S, Isotani A, Yamashita A, Taira K et al (2002) FISH analysis of 142 EGFP transgene integration sites in mouse embryos. Genomics 80:564–574

Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS et al (2021) Post-acute COVID-19 syndrome. Nat Med 27:601–615

Nicholls PK, Bellott DW, Cho TJ, Pintikova T, Page DC (2019) Locating and characterizing a transgene integration site by nanopore sequencing. G3 (Bethesda) 9, 1481–1486

Ohno S (1972) So much “junk” DNA in our genome. Brookhaven Symp Biol 23:366–370

Ohtsuki N, Kizawa K, Mori A, Nishizawa-Yokoi A, Komatsu T, Yoshida H et al (2021) Precise genome editing in mRNA target site via gene targeting and subsequent single-strand-annaeling-mediated excision of the marker gene in plants. Front Genome Ed 2:617713

Onabajo OO, Banday AR, Staniifer ML, Yan W, Obajemu A, Santer DM et al (2020) Interferons and viruses induce a novel truncated ACE2 isoform and not the full-length SARS-CoV-2 receptor. Nat Genet 52:1283–1293

Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkow AV, Prasad R et al (2016) Unlinking an lncRNA from its associated cis element. Mol Cell 62:104–110

Renzsch B, Todiras M, Illescu R, Popova E, Campos LA, Oliveira ML et al (2008) Transgenic angiotensin-converting enzyme 2 overexpression in vessels of SHRSP rats reduces blood pressure and improves endothelial function. Hypertension 52:967–973

Rinn JL, Chang HY (2020) Long noncoding RNAs: molecular modalities to organizational functions. Annu Rev Biochem 89:283–308

Saba LM, Hoffman PL, Homanics GE, Mahaffey S, Daulatalev SB, Janga SC et al (2021) A long non-coding RNA (Lrap) modulates brain gene expression and levels of alcohol consumption in rats. Genes Brain Behav 20:e12698

Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C et al (2013) Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife 2:e01749

Shay MH, Ferracini M, Filezzi V, Chen B, Redis R, Fabris L et al (2018) Cancer-associated rs6983267 SNP and its accompanying long noncoding RNA CCAT2 induce myeloid malignancies via unique SNP-specific RNA mutations. Genome Res 28:432–447

Shen B, Zhang J, Wu H, Wang J, Ma K, Li Z et al (2013) Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell 23:720–723

Singh P, Schimenti JC, Bolcun-Filas E (2015) A mouse geneticist’s practical guide to CRISPR applications. Genetics 199:1–15

Spyropoulos DD, Pharr PN, Lavenburg KR, Jackers P, Papas TS, Ogawa M et al (2000) Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Flt1 transcription factor. Mol Cell Biol 20:5643–5652

Statello L, Guo CJ, Chen LL, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22:96–118

Sun SH, Chen Q, Gui JJ, Yang G, Wang YX, Huang XY et al (2020) A mouse model of SARS-CoV-2 infection and pathogenesis. Cell Host Microbe 28(124–133):e124

Sun Q, Song YJ, Prasanth KV (2021) One locus with two roles: microRNA-independent functions of microRNA-host-gene locus-encoded long noncoding RNAs. Wiley Interdiscip Rev RNA 12:e1625

Suzuki O, Koura M, Uchio-Yamada K, Sasaki M (2020) Analysis of the transgenic insertion pattern in a transgenic mouse strain using long-read sequencing. Exp Anim 69:279–286

Targosz-Korecka M, Kubisak A, Kloska D, Kopacz A, Grochot-Przezefczak A, Szymonski M (2021) Endothelial glycoalyx shields the interaction of SARS-CoV-2 spike protein with ACE2 receptors. Sci Rep 11:12157

Tentolouris A, Ntanasis-Stathopoulos I, Vlachakos PK, Tsilimigras DI, Gavriatopoulou M, Dimopoulos MA (2021) COVID-19: time to flatten the infodemic curve. Clin Exp Med 21:161–165

Tseng CT, Huang C, Newman P, Wang N, Narayanan K, Watts DM et al (2007) Severe acute respiratory syndrome coronavirus infection of mice transgenic for the human Angiotensin-converting enzyme 2 virus receptor. J Virol 81:1162–1173

Varga Z, Flammer AJ, Steiger P, Haberecker M, Andermatt R, Zink-ernagel AS et al (2020) Endothelial cell infection and endotheliitis in COVID-19. Lancet 395:1417–1418
Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D et al (2020) Global reference mapping of human transcription factor footprints. Nature 583:729–736

Wagner JUG, Bojkova D, Shumliakivska M, Luxan G, Nicin L, Aslan GS et al (2021) Increased susceptibility of human endothelial cells to infections by SARS-CoV-2 variants. Basic Res Cardiol 116:42

Wallace HA, Marques-Kranc F, Richardson M, Luna-Crespo F, Sharpe JA, Hughes J et al (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. Cell 128:197–209

Wang X, Raghavan A, Peters DT, Pashos EE, Rader DJ, Musunuru K (2018) Interrogation of the atherosclerosis-associated SORT1 (Sortilin 1) locus with primary human hepatocytes, induced pluripotent stem cell-hepatocytes, and locus-humanized mice. Arterioscler Thromb Vasc Biol 38:76–82

Wu Q, Ferry QRV, Baeumler TA, Michaels YS, Vitsios DM, Habib O et al (2017) In situ functional dissection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. Nat Commun 8:2109

Xu Y, Xi J, Wang G, Guo Z, Sun Q, Lu C et al (2021) PAUPAR and PAX6 sequentially regulate human embryonic stem cell cortical differentiation. Nucleic Acids Res 49:1935–1950

Yang XH, Deng W, Tong Z, Liu YX, Zhang LF, Zhu H et al (2007) Mice transgenic for human angiotensin-converting enzyme 2 provide a model for SARS coronavirus infection. Comp Med 57:450–459

Yoshimi K, Kunihiro Y, Kaneko T, Nagahora H, Voigt B, Mashimo T (2016) ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat Commun 7:10431

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273

Zhou B, Thao TTN, Hoffmann D, Taddeo A, Ebert N, Labroussaa F et al (2021a) SARS-CoV-2 spike D614G change enhances replication and transmission. Nature 592:122–127

Zhou L, Li J, Liu J, Wang A, Liu Y, Yu H, et al. (2021b) Investigation of the IncRNA THOR in mice highlights the importance of non-coding RNAs in mammalian male reproduction. Biomedicines 9

Zhuang A, Calkin AC, Lau S, Kiriazis H, Donner DG, Liu Y et al (2021) Loss of the long non-coding RNA OIP5-AS1 exacerbates heart failure in a sex-specific manner. iScience 24:102537

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.