BMC Genomics

RESEARCH Open Access

Landscape of antisense genes in the human genome and identification of new human hepatic antisense RNAs by long-read sequencing

Juan Jose Rojo-Carrillo[1†], Pedro Garrido-Rodríguez[1†], Maria Llamas-López[1], Rosa Cifuentes-Riquelme[1], Jose Padilla[1], Bruno Ramos-Molina[2], Maria Luisa Lozano[1], Belen de la Morena-Barrio[1*], Maria Eugenia de la Morena-Barrio[1*†] and Javier Corral[1†]

Abstract

Background Protein-coding genes have been considered the functional part of the genome, although they represent only 2% of the genome. In contrast, more than 90% of the genome produces non-coding RNA (ncRNA), including antisense (AS) genes, a type of long non-coding genes (encoding transcripts > 200 nucleotides) located on the opposite strand of coding genes. Therefore, antisense RNA (asRNA) can be complementary to the counterpart sense RNA, supporting a regulatory role with potential pathogenic consequences, as their deregulation has been associated with cardiovascular disease, cancer, and diabetes.

Results We performed an in-depth review of AS genes in *Ensembl* and evaluated the expression of AS genes in human liver by third-generation RNA sequencing methods. Currently, 1656 AS genes overlapping with 1556 sense genes have been identified in the human genome. Coding genes with antisense counterparts were significantly larger and had higher transcriptional activity than genes without antisense counterparts. RNA nanopore sequencing of 15 human livers identified 185 transcripts (55 novel) from 150 known AS genes, and 1316 transcripts from 807 sense genes, with 111 sense-antisense pairs. Sense transcripts showed higher expression than antisense transcripts. Remarkably, RNA nanopore sequencing of human livers identified 146 transcripts (67 novel) corresponding to 118 possible novel AS genes.

[†]Juan Jose Rojo-Carrillo, Pedro Garrido-Rodríguez, Maria Eugenia de la Morena-Barrio ME and Javier Corral contributed equally to this work.

*Correspondence:
Belen de la Morena-Barrio
belendelamorenabarrio@gmail.com
Maria Eugenia de la Morena-Barrio
uge2985@hotmail.com

Full list of author information is available at the end of the article

Conclusion This study reveals the landscape of AS genes in the human genome and demonstrates the power of long-read RNA sequencing to identify novel transcripts and even novel AS genes, exploring the relationship between sense and AS gene expression as well.

Keywords Antisense gene, Antisense RNA, Long-read sequencing, Genome, Transcriptome, Long non-coding RNAs

Background

Protein-coding genes have always been considered the functional part of the genome, even though they represent only about 2%. In contrast, more than 90% of the genome gives rise to non-coding RNAs (ncRNA) [1]. Recent studies have shown that regions previously considered "junk DNA" are involved in important regulatory functions [2]. In fact, there is a positive correlation between the complexity of an organism and the abundance of ncRNA [3].

Antisense RNAs (asRNA) are a type of ncRNA that belong to the group of long non-coding RNAs (lncRNA) because they produce transcripts of more than 200 nucleotides [4]. asRNAs are produced by antisense (AS) genes located on the opposite strand of a protein-coding gene (sense gene) [5]. AS genes are designated by the name of the coding gene followed by the suffix "-AS" (Hugo Gene Nomenclature Committee, HGNC, www.genenames.org/) [6]. There is indeed great ambiguity regarding AS genes naming, as there is no convention about how to refer to those elements. While on expression-level authors tend to use terms such as "antisense RNA", the same element is logged in genomic databases as gene instances. Regarding their position with respect to their target gene, AS genes have been observed in introns, exons, promoters, enhancers, UTRs, and gene flanking sequences, and they can be oriented face-to-face, tail-to-tail, or internally and are [7]. This high conservation and complementarity to the coding gene makes AS genes particularly interesting [8].

The production of asRNAs requires RNA polymerase II and typical RNA processing, so it is believed that AS genes could have expression levels close to the sense gene [9]. They show high degree of tissue specificity [7]. asRNAs have the highest conservation and stability among lncRNAs [8] and have been found in both the nucleus and cytoplasm, with higher abundance in the latter [10].

asRNAs can assemble other macromolecules in three dimensions to participate in a coordinated manner in transcriptional, post-transcriptional and epigenetic processes [10–12], such as DNA methylation and histone modification. Thus, asRNAs are thought to be involved in the regulation of gene expression and genome integrity through transcriptional interference. asRNA regulation can be in *cis*, acting on the originating gene through interaction with its promoter or complementary region, or in *trans*, acting on other genes through imperfect complementarity [10]. Although the physiological role of asRNAs has not been fully characterized, dysregulation of asRNAs has been observed in several diseases such as cancer, diabetes, cardiovascular conditions and neurological disorders [2, 13].

The current method used for lncRNA detection is RT-PCR [10]. However, the low expression of most asRNAs and their high heterogeneity make their study difficult [13]. To the best of our knowledge, long-read-based RNA sequencing methods (which allow the detection of unfragmented, long-length, unmanipulated RNA) [14], have never been applied to evaluate asRNA expression in human liver. Therefore, in this study we dissected the landscape of AS genes in the human genome and described asRNAs detected by long-read RNA sequencing in 15 human liver samples.

Results

Identification of 1685 AS genes described in the human genome

AS genes already described in the human genome were identified according to the procedure described in Fig. 1.

This approach allowed us to identify 18,806 genes classified as lncRNA in *Ensembl*. 1819 genes showed the suffix "-AS"; 134 of them were mapped in alternative contigs. Thus, 1685 AS genes and 1556 sense genes were reported across the human genome (Supplementary Tables 1 and 2); according to their nomenclature, 1656 AS genes were identified with defined sense gene (Fig. 1 and supplementary Table 3).

Most AS genes (N=1503) showed a number 1 in their suffix (-AS1), meaning they were the only AS gene of their corresponding sense gene. However, 157 AS genes had numbers ranging from 2 to 6; i.e. –AS4 means that at least 4 different AS genes have been identified to the same sense gene. *NAV2* was the gene with more associated AS genes (*NAV2-AS1* to *NAV2-AS6*). Finally, 25 AS genes showed no number in their suffix (Supplementary Table 4).

In addition, some AS genes did not follow this convention: (1) *LARGE-AS1* should be the AS gene for *LARGE1*, although there is also a *LARGE1-AS1* gene, and both AS genes share close chromosomal regions. (2) *MYL12-AS1* includes two genes: *MYL12A* and *MYL12B*. Similarly, *KRTAP5-AS1* overlaps with *KRTAP5-1* and *KRTAP5-2*. In fact, there are 38 AS genes overlapping with two or more sense genes (Supplementary Table 5). (3) *TMEM246-AS1* did not have associated gene with this name ("*TMEM246*"), but according to its chromosomal

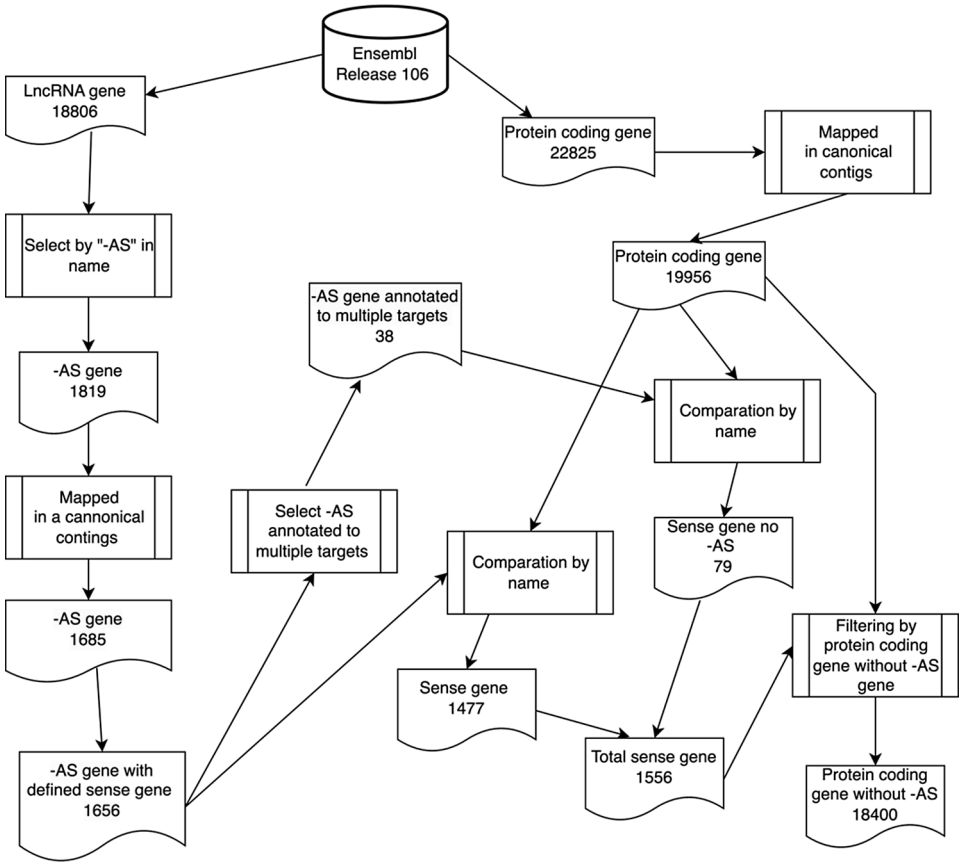

Fig. 1 Workflow for antisense gene identification

coordinates, it could be the AS gene of *PGAP4* and thus should be named *PGAP4-AS1*.

The AS genes of the HOX gene clusters deserved a detailed description. Twelve AS genes were identified for the HOX gene clusters (4 clusters and 39 genes distributed on chromosomes 2, 7, 12, and 17). However, only 3 AS genes, *HOXA10-AS*, *HOXA11-AS* and *HOXAC13-AS*, could be confidently related to the sense gene. The remaining HOX-AS genes overlapped with more than one gene. Thus, it could not be assigned to a single gene. In addition, two AS genes named *HOTAIR* and *HOTAIRM1* also appear in this cluster but were not considered for this study because they are related to the entire *HOXC* and *HOXA* gene clusters, respectively, without specifying the sense gene.

In summary, our study identified 1656 AS genes that could be linked to 1556 sense genes.

Genes with AS genes are longer and enclose more alternative transcripts

Our study showed only 1/10 of protein coding genes in the human genome have annotated AS genes. Our first goal was to study the differences between both groups of genes to identify some features that could explain why some genes have AS genes while others do not.

First, the distribution of (sense) genes with and without AS genes was evaluated across chromosomes in absolute values (Fig. 2a), and in percentages related to the total number of genes per chromosome (Fig. 2b). Chromosome 13 was highly enriched on genes with AS (18%). In contrast, chromosomes 16 and Y showed the lowest percentage (4%) as shown in Fig. 2b.

Although gene size among those with and without AS showed a large heterogeneity, median length of genes with AS was significantly higher than that of those without AS (median: 63394 bp; IQR: 22324–171915 vs. 25956 bp; IQR 9232–66092; $p<2.2e\text{-}16$, Cohen's D 0.732; Fig. 3a). Transcriptional heterogeneity between groups was also evaluated. Genes with AS showed a higher median number of alternative transcripts than those without AS (8; IQR: 4–14 vs. 6; IQR 3–11; $p<2\text{-}2e\text{-}16$, Cohen's D 0.340; Fig. 3b). We observed a higher exon count in genes with AS than those without AS (median: 11; IQR 6–18 vs. 8; IQR 13−4; $p<2.2e\text{-}16$, Cohen's D 0.373; Supplementary Fig. 1a). Standardization by gene size was also performed to ascertain the impact of this factor. The results indicated genes without antisense had more exons per base pair ($p<2.2e\text{-}16$, Cohen's D 0.342; Supplementary Fig. 1b).

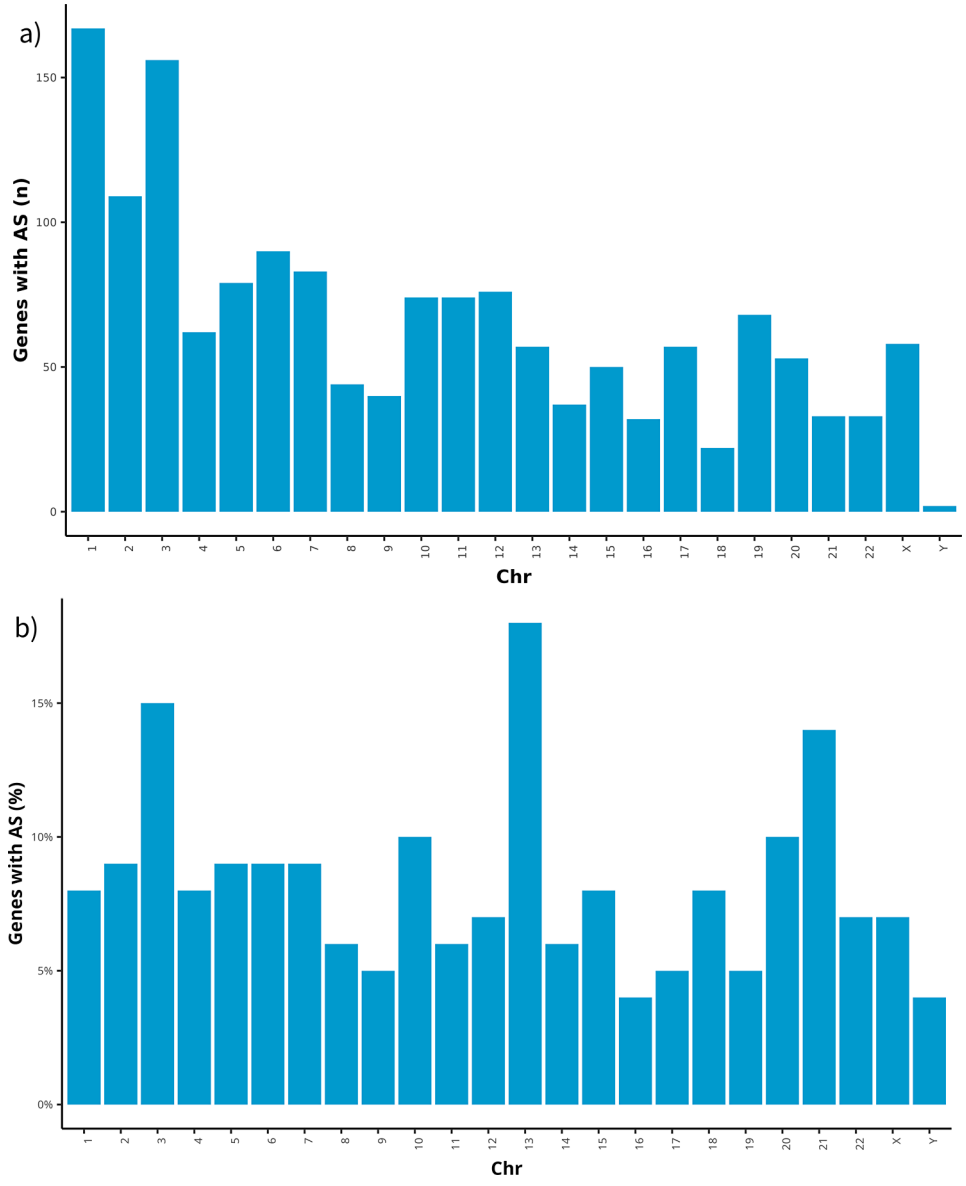

Fig. 2 Distribution of genes with AS in each chromosome. (**a**) Absolute values. (**b**) Density

Finally, genes with or without AS showed differences in repeated sequences per gene (107 & IQR 33–275 vs. 50 & IQR 16–123; respectively). However, this difference might be attributed to their longer length. Indeed, after standardization by gene size, genes without AS showed a higher median of repeated sequences per gene, but with a negligible Cohen's D ($p=0.0289$, Cohen's D 0.031Supplementary Fig. 2a). Additionally, repetitive sequences have been examined in terms of family and subtype in Supplementary Fig. 2b y 2c.

AS shows variable overlapping degree with their sense genes

We compared the relative position of the AS with respect to its sense gene. Near half of the AS genes (762, 46%) completely overlapped with their sense counterparts. Subsequently, we found a variable degree of overlap between the AS and its respective sense gene, even at low levels, such as 0.02%, in the case of *POT1-AS1*. Indeed, there were 58 genes with AS annotation (3.5%) that showed no overlap with their sense gene (Supplementary Table6).

We also compared the size of sense and AS genes. Overall, the median size of sense genes was larger than that of AS (63394 bp; IQR: 22324–171915 vs. 10380 bp; IQR: 3410–34290; $p<2.2e-16$, Cohen's D 0.640; Fig. 4a),

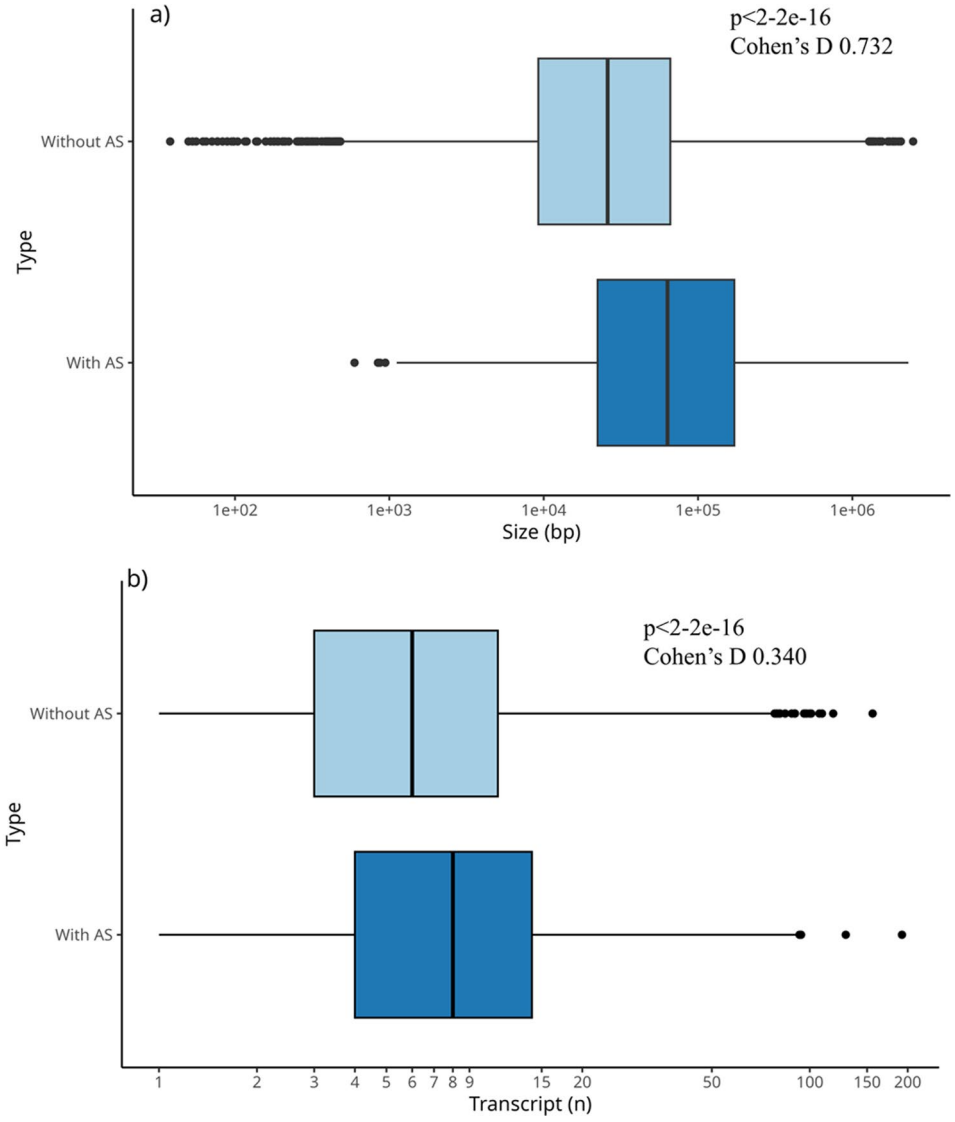

Fig. 3 Comparison of genes with and without AS. (**a**) Gene Size; (**b**) Transcripts count

perhaps because there were 114 sense genes that contained multiple AS genes. Nevertheless, we observed a large variability in size for both sense and AS genes (330 bp to 1064 Kb for AS genes, 595 bp to 2305 Kb for sense genes). Regarding transcriptional heterogeneity, we found sense genes had more alternative transcripts than AS genes (8; IQR: 4–14 vs. 2; IQR: 1–5; $p<2.2e-16$, Cohen's D 0.589; Fig. 4b). However, counter examples can be found, with *PCBP1-AS1* and *PCBP1* being extreme examples, as these genes have 283 and 1 transcripts, respectively.

No correlation was observed between gene size (Kendall τ 0.012; Supplementary Fig. 3a) nor transcripts count of the sense gene and the presence of AS counterparts (Kendall τ -0.049; Supplementary Fig. 3b).

Finally, the analysis of repetitive elements in AS genes revealed a grand total of 96,340 elements across all AS genes. The most frequently detected family was SINE ($N=37611$, 39.0%), followed by LINE ($N=32995$, 34.3%), LTR sequences ($N=13870$, 14.4%), and DNA repeat sequences ($N=11281$, 11.7%). On the other hand, 348,276 repeat sequences were detected across sense genes. The most frequent family was SINE ($N=146197$, 42.0%), followed by LINE ($N=117986$, 33.9%), DNA repeat sequences ($N=44483$, 12.8%) and LTR ($N=37483$, 10.8%), as shown in Supplementary Fig. 4. The median number of repetitive elements per gene was higher in sense genes than in AS (107; IQR: 33–275 vs. 21; IQR: 6–67) with no significant differences (normalized data: $p=0.1512$).

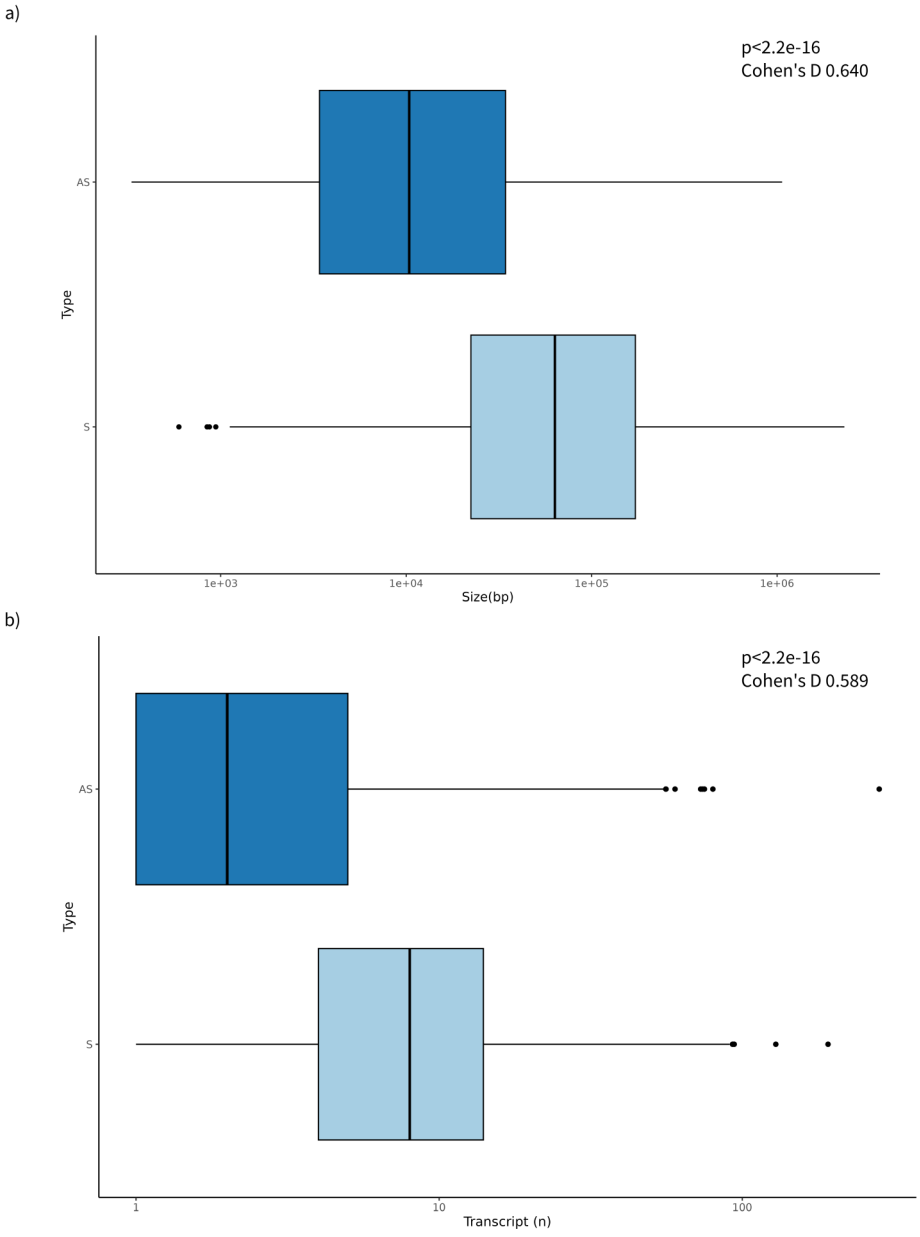

Fig. 4 Comparison of Sense and AS genes: (**a**) By gene size; (**b**) By transcripts count

Identification of 15 AS genes of non-coding genes

We identified 15 AS genes had a non-protein coding sense counterpart (8 long non-coding and 7 pseudogenes). The characteristics of these sense and AS genes are summarized in Table 1.

Expression profile of AS by tissue reveals different expression patterns

To assess the expression patterns of AS genes across different human tissues, we studied the expression levels of AS genes based on Adult GTEx data. As can be seen on Fig. 5, AS show different expression levels across tissues, with some tissues having similar AS patterns as well, such as the left ventricle and atrial appendage tissues from the heart. On the other hand, it is interesting to highlight the high expression levels on testis tissue in comparison with others. Thus, it would be of interest to get an in-depth picture of such differences in further studies.

Identification by nanopore RNA sequencing of 118 potential new AS and comparison with Sense expression in human liver.

Direct RNA nanopore sequencing of 15 human liver samples identified 185 transcripts, corresponding to 150 previously described AS genes. More than a quarter of

Table 1 Characteristics of AS genes corresponding to non-protein-coding genes

Chr	AS gene	Size (bp)	Transcripts	Type	Sense gene	Size (bp)	Transcripts	Type
5	*C5orf64-AS1*	2941	3	lncRNA	*C5orf64*	182,977	34	lncRNA
12	*CLLU1-AS1*	8054	1	lncRNA	*CLLU1*	9471	5	lncRNA
2	*DIRC3-AS1*	641,716	18	lncRNA	*DIRC3*	472,574	11	lncRNA
1	*GAS5-AS1*	1468	1	lncRNA	*GAS5*	10,323	78	lncRNA
1	*HHLA3-AS1*	875	1	lncRNA	*HHLA3*	30,553	8	lncRNA
20	*MIR1-1HG-AS1*	7361	6	lncRNA	*MIR1-1*	70	1	lncRNA
15	*SPATA8-AS1*	11,307	2	lncRNA	*SPATA8*	2226	5	lncRNA
22	*PRR34-AS1*	3505	3	lncRNA	*PRR34*	4666	1	lncRNA
1	*CYMP-AS1*	2578	1	lncRNA	*CYMP*	10,525	3	pseudogene
17	*CCDC144NL-AS1*	133,843	22	lncRNA	*CCDC144NL*	59,710	5	pseudogene
9	*PGM5P3-AS1*	18,164	7	lncRNA	*PGM5P3*	657	2	pseudogene
2	*PGM5P4-AS1*	15,902	8	lncRNA	*PGM5P4*	21,361	1	pseudogene
8	*TDH-AS1*	1754	1	lncRNA	*TDH*	28,815	5	pseudogene
8	*ZNF252P-AS1*	3235	1	lncRNA	*ZFN252P*	29,323	11	pseudogene
7	*TRG-AS1*	47,826	20	lncRNA	*TRGV*	650	1	pseudogene

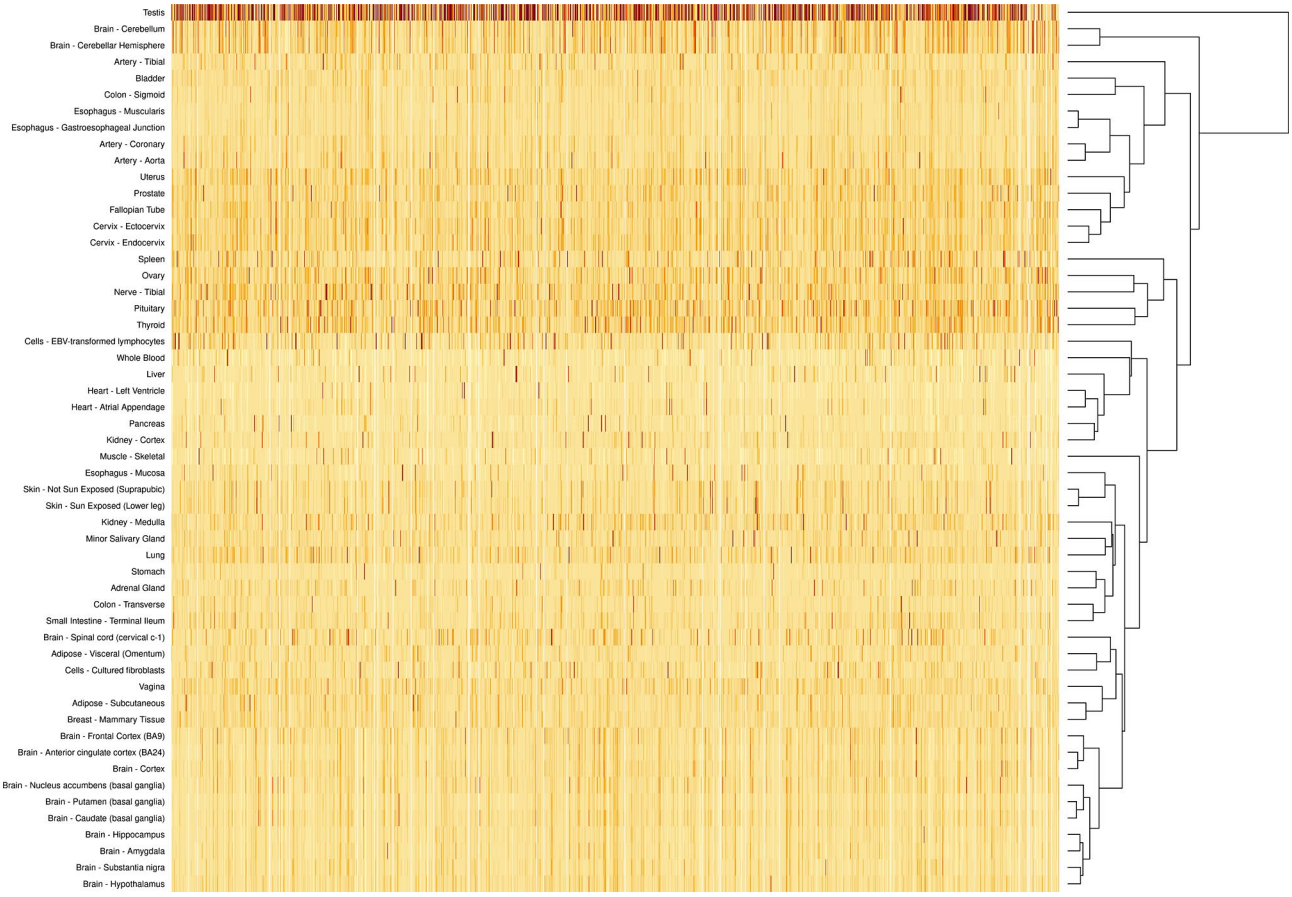

Fig. 5 Heatmap of expression profile of AS by tissue

the 185 transcripts identified in our study ($N=55$) were novel forms of previously described AS genes. On average, 75 asRNAs were detected per sample. 22 asRNAs were detected in more than 10 subjects, while 2 were reported in only one sample. Most asRNAs in human liver showed low expression, with a median of 7.7 TPM (transcripts per million); IQR: 6.4–10.7, (Fig. 6). A novel transcript of *FGD5-AS1* was the asRNA with highest expression (median: 43.1 TPM; IQR: 32.43–56.22). Expression of asRNAs in each sample is shown in Supplementary Fig. 5.

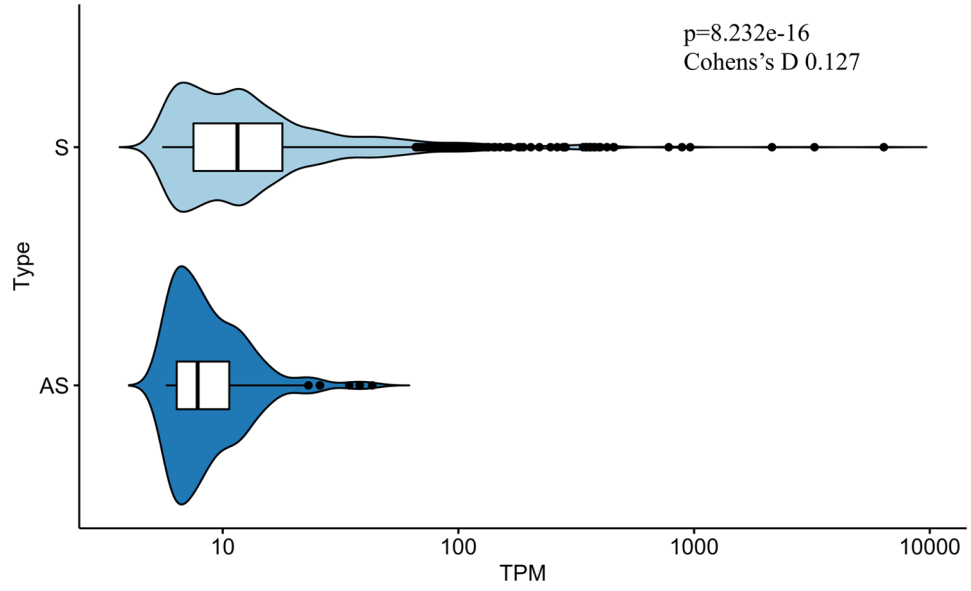

Fig. 6 Transcripts identified in human liver by nanopore sequencing according to their expression levels (TPM) in antisense (AS) and sense (S) genes

Nanopore RNA-Seq of human liver identified 1316 transcripts for 807 sense genes, all with previously reported AS genes. Most of the sense gene transcripts also showed low expression (median 11.5 TPM; IQR: 7.5–18.0), but this was statistically higher than the expression of the AS genes (p=8.232e-16 and Cohens's D 0.127) (Fig. 6).

Supplementary Fig. 6 shows the expression levels of sense gene transcripts according to their expression in each sample.

Our approach detected transcripts for 111 pairs of sense-AS genes in human liver samples (Supplementary Table7). For these pairs, sense genes showed more transcripts (median 2; IQR: 1–6 vs. 1; IQR: 1–1; p=0.002425, Cohen's D 0.438) and higher expression than their AS counterparts (median 11.6 TPM; IQR: 8-19.1 TPM vs. 8.0 TPM; 6.6–11.6 TPM; p=1.424e-08, Cohen's D 0.420) (Fig. 7).

Next, we conducted an analysis to verify if there was any correlation between the expression of AS and sense gene pairs (Fig. 8). We cannot assert the existence of a correlation (Kendall τ -0.0540). Nevertheless, qualitatively, certain outlier values can observe could be of interest.

In addition, we consider of great interest the relative expression of sense and AS gene pairs, which is clearly unbalanced in favor of the sense gene in general. Nonetheless, in 28.7% of the sense-antisense pairs, transcripts of the AS gene had a higher expression than those from the sense counterpart (Fig. 9). The individual variability of these relationships and their potential impact on different disorders needs to be addressed in further studies.

Our study found expression of 696 sense genes without co-expression of their AS (Supplementary Table8). Aiming to find an explanation to this finding, we compared the characteristics of sense genes with or without AS expression, but no differences were observed in transcript number (median 1; IQR: 1–2 vs. 1; IQR: 1–2; p=0.43) nor in expression levels (median 11.6 TPM; IQR: 8.0—19.1 vs. 11.5; TPM IQR: 7.5–17.9; p=0.37) (Supplementary Fig. 7).

On the other hand, 51 transcripts from 40 AS genes were expressed in human liver, while their sense counterparts were not, even though they tended to be expressed in human liver according to public databases (Supplementary Table9). These orphan asRNAs showed low expression (median 7.5 TPM; IQR: 6.2–9.5) and few alternative transcripts (median 1; IQR: 1–1). Interestingly, 19 of these 51 forms corresponded to novel transcripts detected by long reads. We compared the characteristics of AS genes whose sense genes were expressed in the liver with those of AS genes whose sense genes were not expressed. No differences were observed either in the number of transcripts (median 1; IQR 1–1 vs. 1; IQR: 1–1; p=0.25) or in the expression levels (median 7.5 TPM; IQR 6.2–9.5 vs. 8.0 TPM; IQR: 6.6–11.6; p=0.17) (Supplementary Fig. 8).

In addition, we explored if we could identify novel AS genes using long reads. Among all the sequences, we searched for lncRNAs (>200 nucleotides) that overlapped with protein-coding genes (≥1 nucleotide) in antisense orientation. This strategy revealed 118 potential new AS genes with 144 different transcripts, 67 of which were novel (Table 2). In the description of 58 genes identified in our study already included the definition of

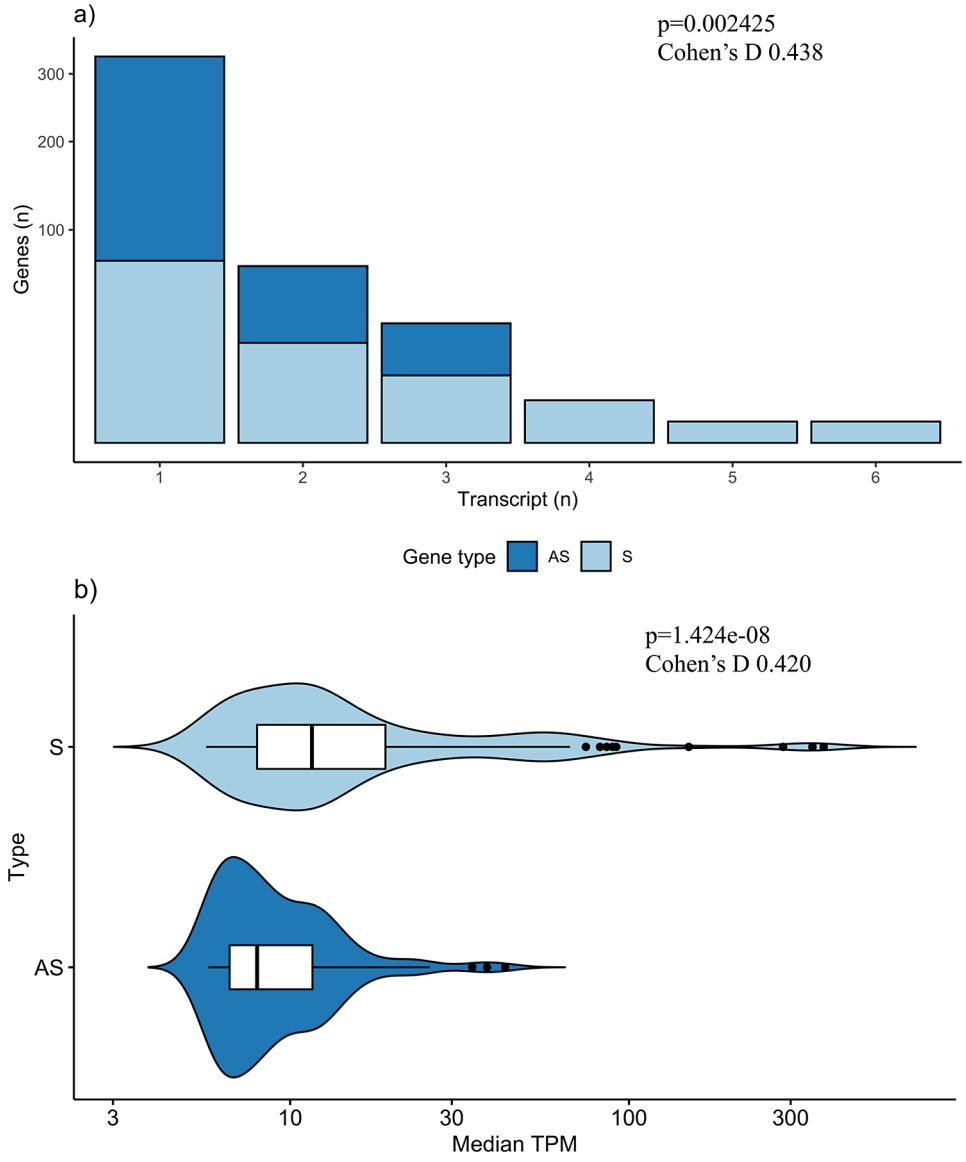

Fig. 7 Comparison of sense transcripts (S) ($N = 111$) and antisense (AS) gene pairs expressed in human liver samples: (**a**) number of transcripts; (**b**) Expression levels in TPM median values

antisense; 7 were described as antisense RNA and 2 as divergent. In fact, 10 of the potential new AS genes identified in this study already had an alias including 'AS'. One was considered an artifact, twenty had a previous name unrelated to AS, and 30 were completely new transcripts. For 14 potential new AS genes, at least 2 different transcripts were identified in this study. We also measured the overlap of AS transcripts with their proposed target gene(s). This information is summarized in Table 2.

In order to validate these results, we evaluated the expression of selected transcripts by qRT-PCR. We selected three previously described transcripts that that were not described as AS, as well as three novel AS transcripts. All selected transcripts amplified by qRT-PCR in the subject in which these transcripts were identified by nanopore RNAseq. Further validation of these novel AS transcripts involved electrophoresis on 5% and 6.25% polyacrylamide gels of specific qRT-PCR and PCR amplification products and stained by using silver. Additionally, the PCR amplicons were sequenced using nanopore sequencing, confirming the sequences identified by RNA-seq.

The name we proposed for the new AS genes identified in this study followed these rules: unnamed AS genes were labeled with the name of the overlapping sense gene, followed by the suffix 'AS' and a correlative numbering if previous AS were associated with such target. If the AS gene overlapped two sense genes, it took the name of the gene with higher mean overlap followed by 'AS' suffix, unless this gene had no name. In case the overlapping

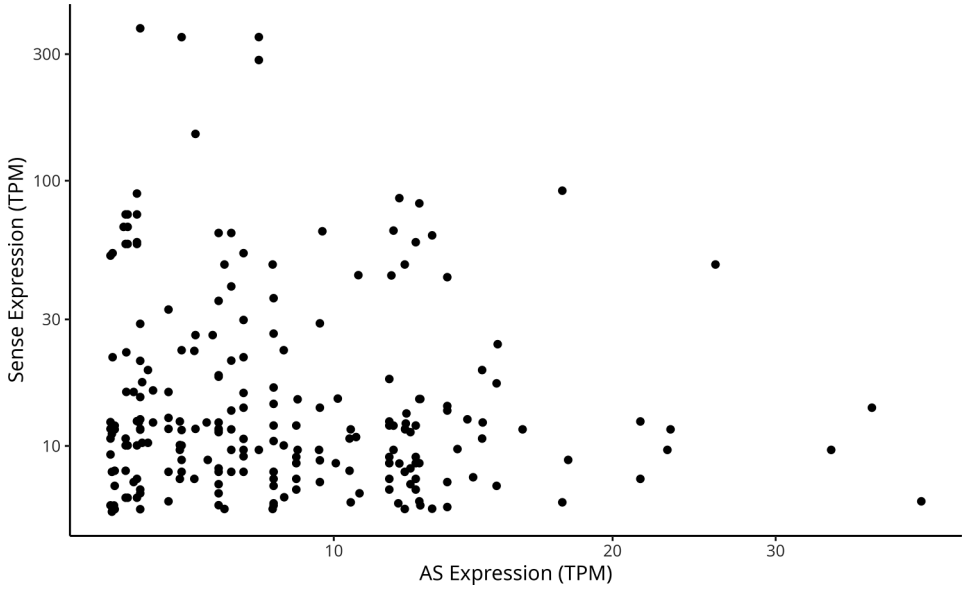

Fig. 8 Correlation of expression levels for sense and antisense (AS) gene pairs in TPM values

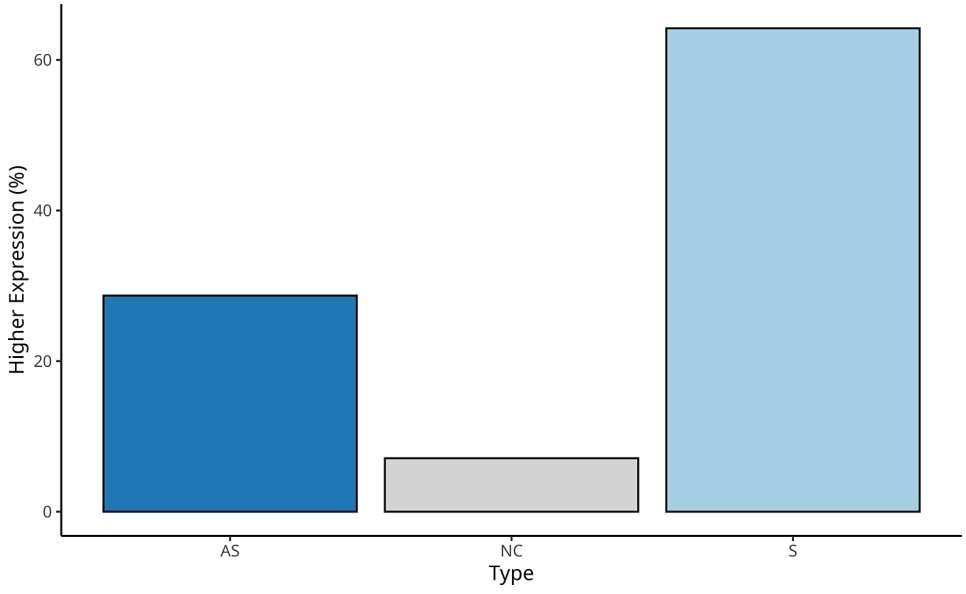

Fig. 9 Percentage of sense-AS pair transcripts with higher expression for one element of the pair. NC: No expression change

gene had no name, we propose to keep the code followed by 'AS' suffix. Finally, if the AS gene had already had a name (not recognized as AS), we add an AS alias created according to the aforementioned indications.

In supplementary Table 10, we present expression (TPM) on a matrix format for all expression of all the transcripts analyzed in this study.

Discussion

The gap between a huge genome and a few coding genes is being filled by the elucidation of new functions for what was previously considered junk DNA [1]. There is increasing evidence supporting the importance of non-coding genes, which have not only structural but also regulatory functions [2]. In addition, the highly optimized genome of prokaryotic organisms is apparently wasted in eukaryotes, as we mainly use only one strand of a gene, the template strand for transcription [15]. However, this perspective is changing with the increased interest in AS genes, as they are located in the complementary strand of their corresponding sense gene. We have reviewed the current knowledge of AS genes in the human genome, and by using nanopore long-read-based sequencing technology for RNA we provide further evidence that this field needs to be explored in depth.

Table 2 Potential novel AS genes identified in human liver samples by long-read RNA sequencing

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
0be7179e-8f08-48b4-ae0c-6b1a6d09fe3b	822	***PAXIP1*** 822		Novel transcript, antisense to *PAXIP1* ENSG00000272760	*PAXIP-AS3*
22f2f8fe-dc27-41d2-af21-6da301239ecb 14f5d420-9c14-4d48-84c2-1ab5f24d8c03 91fa9b64-f6ce-4a19-b3c5-59a39103249e 29c12e98-ecc9-40dc-a312-55f20b71502c	859 555 376 899	***CD36*** 3279 2971 2971 4985		Novel transcript, antisense to *CD36* ENSG00000232667	*CD36-AS1*
15b13279-951c-40d6-ade0-27b2377cc467	995	***TBC1D14*** 995		Novel transcript, antisense to *TBC1D14* ENSG00000287104	*TBC1D14-AS1*
163ab1f5-89ac-41c0-b8c2-d7acc807d51f	669	***SS18L1*** 217		Novel transcript, antisense to *SS18L1* ENSG00000289537	*SS18L1-AS1*
17367111-8f95-47 cd-b896-c88eebd34248	605	***CUL5*** 385		Novel transcript, antisense to *CUL5* ENSG00000288012	*CUL5-AS1*
17fa8bbc-6c7a-4394-803a-d7bcb428bd2a	766	***CTDSP1*** 766		Novel transcript, antisense to *CTDSP1* ENSG00000273361	*CTDSP1-AS1*
1958df77-4661-4fbf-a864-40ed3a41499e	897	***TUT1*** 53	**ENSG00000255508** 897	Novel transcript, antisense to *TUT1* ENSG00000289562	*TUT1-AS1*
19e743a5-bbb7-4a2a-816e-b595171186e7	1180	***MARCHF6*** 186		Novel transcript, antisense to *MARCH6* ENSG00000259802	*MARCHF6-AS1*
ENST00000572046.2 ENST00000417897.2 ENST00000689184.1	844 344 817	***CD68*** 137 137 137	***MPDU1*** 2 544 2	Novel transcript, antisense to *MPDU1* and *CD68* ENSG00000233223	*MPDU1-AS1*
c1068eb5-edb5-4a6b-afe3-6049b9be0640	1267	***RBM33*** 755		*RBM33* divergent transcript ENSG00000216895 *(RBM33-DT)*	*RBM33-AS1*
ENST00000684981.1 489112f4-1c98-4baa-83f7-f81e76757fa8	1455 2289	***TNS2*** 95		*TNS2* antisense RNA 1 ENSG00000257337 *(TNS2-AS1)*	*TNS2-AS1*
ENST00000690863.1	911	***DSP*** 1465		*DSP* antisense RNA 1 ENSG00000261189 *(DSP-AS1)*	*DSP-AS1*
ENST00000688747.1 ENST00000602674.2	370 373	***FARS2*** 5433 5433		*FARS2* antisense RNA 1 ENSG00000269985 *(FARS2-AS1)*	*FARS2-AS1*
ENST00000613438.2 ENST00000621378.2	520 803	***RIPOR1*** 10,266 10,266		Novel transcript, antisense to *FAM65A* ENSG00000276075	*RIPOR1-AS1*
27c32315-9bc9-43b0-9ff1-4a075d443bc4 ENST00000669440.2	403 316	***ERVK3-1*** 2025 2025	**ENSG00000283515** 396 396	Novel transcript, antisense to *ERVK3-1* ENSG00000283103	*ERVK3-1-AS1*
ENST00000597230.3	529	**ZNF544** 18,047	**ENSG00000283515** 19,360	ZNF8 divergent transcript (ZNF8-DT) ENSG00000268516	ZNF544-AS1
ENST00000330539.1	2216	***ERLIN2*** 956		Novel transcript, antisense to *ERLIN2* ENSG00000183154	*ERLIN2-AS1*

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
ENST00000409905.1	642	***NGEF*** 643		Novel transcript, antisense to *NGEF* ENSG00000222001	*NGEF-AS1*
ENST00000593917.2	630	***SMIM29*** 528		Novel transcript, antisense to *C6orf1* ENSG00000225339	*SMIM29-AS1*
ENST00000657821.1	741	***EPHB1*** 7981		Novel transcript, antisense to *EPHB1* ENSG00000240086	*EPHB1-AS1*
8f1508a6-0820-423f-9f59-57402cea0793	2792	***CACNB2*** 28,325		Novel transcript, antisense to *CACNB2* ENSG00000240291	*CACNB2-AS1*
ENST00000511361.1	3000	***PDK2*** 3787		Novel transcript, antisense to *PDK2* ENSG00000250282	*PDK2-AS1*
ENST00000534620.1	408	***KIRREL3*** 26,019		Novel transcript, antisense to *KIRREL3* ENSG00000254607	*KIRREL3-AS1*
ENST00000526623.2	1539	***CAPN1*** 999		*CAPN1* antisense RNA 1 ENSG00000254614 *(CAPN-AS1)*	*CAPN1-AS1*
ENST00000649789.1	585	***SLC29A2*** 1240		*B4GAT1* divergent transcript ENSG00000255468 *(B4GAT1-DT)*	*B4GAT1-DT (SCL29A2-AS1)*
ENST00000684881.1	940	***LTA4H*** 10,952		Novel transcript, antisense to *LTA4H* ENSG00000257878	*LTA4H-AS1*
ENST00000555011.3	589	***ACOT6*** 3205		Novel transcript, antisense to *ACOT6* ENSG00000258603	*ACOT6-AS1*
ENST00000556301.2	492	***RAD51B*** 1493		Novel transcript, antisense to *RAD51L1* ENSG00000259038	*RAD51L1-AS1*
33c630b6-0c5f-4fc1-866f-54f3a4d9526b	1669	***SIX5*** 1115	***DMPK*** 98	Novel transcript, antisense to *SIX5* ENSG00000259605	*SIX5-AS1*
a942c6e6-bee4-405c-b2da-6253ef6d739e	1521	***CBFA2T3*** 2418		Novel transcript, antisense to *CBFA2T3* ENSG00000259881	*CBFA2T3-AS1*
b492a6a5-65a7-494e-b20b-c67864714d77	695	***TGFBR3L*** 453		Novel transcript, antisense to CTD-3193O13.2 ENSG00000260500	*TGFBR3L-AS1*
ENST00000570952.1	381	***TBC1D16*** 2177		Novel transcript, antisense to *TBC1D16* ENSG00000261978	*TBC1D16-AS1*
ENST00000580714.1	493	***SLC46A1*** 1557		Novel transcript, antisense to *SLC46A1* ENSG00000265254	*SLC46A1-AS1*
c8c060f3-d855-4d49-8dfd-929919f78ceb	729	***RBBP8*** 135,504		Novel transcript, antisense to *RBBP8* ENSG00000266850	*RBBP8-AS1*
8b0cd22f-e27a-44d0-8957-fd5575f37737	1354	***APC2*** 1354		Novel transcript, antisense to *APC2* ENSG00000267317	*APC2-AS1*

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
f0e53bf1-f4a3-4add-a193-9ef06b71411b	3082	***ABHD16B*** 1491		Novel transcript, antisense to *ABHD16B* ENSG00000268858	*ABHD16B-AS1*
e53f9e95-03ca-4e10-bdbb-f79ddb0ecacf	1320	***TAF6L*** 1320		Novel transcript, antisense to *TAF6L* ENSG00000269176	*TAF6L-AS1*
23c63d00-ecb8-4450-af9e-72a3a7c7a774	1372	***TARBP2*** 613		Novel transcript, antisense to *TARBP2* ENSG00000270175	*TARBP2-AS1*
829320ca-8d68-4d09-94d4-ed50d1e53b89	1173	***SHC1*** 447	***PYGO2*** 1173	Novel transcript, antisense to *SHC1* and *PYGO2* ENSG00000271380	*PYGO2-AS1*
ENST00000605945.1	401	***ZNF165*** 278		Novel transcript, antisense to *ZNF165* ENSG00000272009	*ZNF165-AS1*
235fdc67-3a23-421e-a4cd-c7c6089b5e21	561	***PDHB*** 561		Novel transcript, antisense to *PDHB* ENSG00000272182	*PDHB-AS1*
9da7467c-4b25-426b-8500-5cf044d017d0	884	***RIPK1*** 884		Novel transcript, antisense to *RIPK1* ENSG00000272277	*RIPK1-AS1*
ENST00000548468.2	641	***CERS5*** 36,330		Novel transcript, antisense to *CERS5* ENSG00000272368	*CERS5-AS1*
2c7c8e20-f03e-4081-9312-1883287c1a8e	679	***RABGEF1*** 679	***ENSG00000284461*** 679	Novel transcript, antisense to *RABGEF1* ENSG00000272831	*RABGEF1-AS1*
ENST00000608016.1	3516	***SELENOO*** 3958		*SELENOO* antisense RNA 1 ENSG00000273137 *(SELENOO-AS1)*	*SELENOO-AS1*
50f099da-7043-4c71-99eb-cf7d837896a3	429	***EXTL2*** 252		Novel transcript, antisense to *EXTL2* ENSG00000273204	*EXTL2-AS1*
8e80aacc-95cb-4316-8d99-12fcbb111f21	3806	***PROX1*** 3806		Novel transcript, antisense to *PROX1* ENSG00000274895	*PROX1-AS2*
e213331a-8c98-4011-bc4e-5072d455a2a1	700	***DENND1C*** 700		Novel transcript, antisense to *DENND1C* ENSG00000275234	*DENND1C-AS1*
200c56b3-583a-4fb8-91d8-fd7f81067a62	467	***SGF29*** 261		Novel transcript, antisense to *CCDC101* ENSG00000275441	*SGF29-AS1*
c9390c69-74f3-433f-a27a-2aeebdf2111d	974	***XRN2*** 374		Novel transcript, antisense to *XRN2* ENSG00000275457	*XRN2-AS1*
8ebbdba1-65f9-4f50-ba8a-bd5c68d6140e	770	***CNOT1*** 770		Novel transcript, antisense to *CNOT1* ENSG00000276259	*CNOT1-AS1*
c6a08ce8-039c-4fb5-bcbf-ef45adb769e0	2304	***SLCO4A1*** 2304		*SLCO4A1* antisense RNA 2 ENSG00000277496 *(SLCO4A1-AS2)*	*SLCO4A1-AS2*
9b0ae63c-7980-43eb-a876-f75d8b865c11	1626	***SMARCE1*** 562	**ENSG00000264058** 1626	Novel transcript, antisense to *SMARCE1* ENSG00000278834	*SMARCE1-AS1*

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
ENST00000633182.1	1052	***LAD1*** 412		Novel transcript, antisense to *LAD1* ENSG00000282221	*LAD1-AS1*
ENST00000650724.2	599	***BCAT2*** 3412		Novel transcript, antisense to *BCAT2* ENSG00000286024	*BCAT2-AS1*
ENST00000661159.1	1605	***MAP3K5*** 2745		*MAP3K5* antisense RNA 2 ENSG00000286646 *(MAP3K5-AS2)*	*MAP3K5-AS2*
ENST00000667155.2	1307	***CNPY3*** 625		Novel transcript, antisense to *CNPY3* ENSG00000287825	*CNPY3-AS1*
ENST00000652867.1	627	***GRIN2B*** 38,047		Novel transcript, antisense to *GRIN2B* ENSG00000287928	*GRIN2B-AS1*
ENST00000669726.2	355	***RABGGTA*** 411	***DHRS1*** 8261	Novel transcript, antisense to *RABGGTA* and DHRS1 ENSG00000288044	*DHRS1-AS1*
5d650a5d-f957-4c36-bc9a-b5aaf9eef867	576	***HSPA8*** 576		Novel transcript, antisense to *HSPA8* and CLMP ENSG00000288061	*HSPA8-AS1*
fa4deeac-521c-44ed-9fbb-ed04e33714ba	878	***SLC19A1*** 878		Novel transcript, antisense to *SLC19A1* ENSG00000288885	*SLC19A1-AS1*
2e156f29-7402-4cc6-8dff-f17418ac17c8	718	***STT3A*** *53*	***CHEK1*** 718	Novel transcript, antisense to *CHEK1* ENSG00000288907	*CHEK1-AS1*
adcdede2-ec3e-418c-9899-5edefe5ceb66	1441	***PARP8*** 1441		Novel transcript, antisense to *PARP8* ENSG00000289056	*PARP8-AS1*
ab8bc225-49fc-4442-832a-5439bf86407e	622	***ACYP2*** 622		Novel transcript, antisense to *ACYP2* ENSG00000289065	*ACYP2-AS1*
80d83732-9c04-48f2-88bd-9e3f189b7293	1768	***GSDMD*** 1768		Novel transcript, antisense to *GSDMD* ENSG00000289161	*GSDMD-AS1*
3d476ea2-1c18-4b2d-a794-a698302c1232	431	***GPR107*** 431		Novel transcript, antisense to *GPR107* ENSG00000289226	*GPR107-AS1*
ENST00000686111.1	615	***SLC29A3*** 7400		Novel transcript, antisense to *SLC29A3* ENSG00000289607	*SLC29A3-AS1*
ENST00000475947.7 ENST00000660462.1 ENST00000481027.5 c6e73a37-8ff5-4187-a4ca-621492844a7d ENST00000475953.6 ENST00000470491.7	906 988 1041 931 909 808	***LRRC75A*** *442* 442 442 442 442		*SNHG29* ENSG00000175061 *C17orf76-AS1* *FAM211A-AS1* ***LRRC75A-AS1***	*SNHG29* *(LRRC75A-AS1)*
ENST00000648127.1	749	***TRNAU1AP*** 5		*SNHG12* ENSG00000197989	*SNHG12* *(TRNAU1AP-AS1)*
ENST00000565342.5	2438	***BOLA2B*** 415		*SLX1A-SULT1A3* ENSG00000213599	*SLX1A* *(BOLA2B-AS1)*

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
ENST00000351325.9	498	***LGI4***		Novel transcript	*LGI4-AS1*
ENST00000592818.5	972	3611		ENSG00000221857	
1fb06730-b21e-4c37-8faa-b3b47aee30b2	494	3611			
		3011			
ENST00000669154.1	554	**ENSG00000284686**		*LINC01767*	*LINC01767*
ENST00000451914.2	454	1002		ENSG00000223956	*(ENSG00000284686-*
ENST00000690755.1	644	913			*AS1)*
		913			
129b45cb-e1bd-44ab-b41d-102f755f9cd0	720	***CROCC2***		*UICLM*	*UICLM*
		1774		ENSG00000233392	*(CROCC2-AS1)*
				(CROCC2-AS1)	
ENST00000425110.1	532	***CROCC2***		Novel transcript	*CROCC2-AS2*
		2745		ENSG00000223991	
1852842a-c5b8-4cd7-9fea-f60ceb6e96ef	774	***KCNE2***		Novel transcript	*KCNE2-AS1*
		867		ENSG00000225555	
60f67cbf-7bf6-43a6-9733-f01eab9489ea	1492	***PHF10***		Novel transcript	*PHF10-AS1*
		963		ENSG00000227704	
ENST00000454219.2	726	***SLC35A3***		Novel transcript	*SLC35A3-AS1*
		585		ENSG00000228084	
2398dfc1-cca5-4b89-9b06-d1dfc6f50e9c	2739	***RGL4***		*GUSBP11*	*GUSBP11*
		3807		ENSG00000228315	*(RGL4-AS1)*
faf9036f-4689-489d-9393-d8d0c2e94ece	1793	**ENSG00000285444**	***DHRS12***	Novel transcript	*DHRS12-AS1*
		1793	*326*	ENSG00000231856	
ENST00000416970.5	783	***DIPK1B***		*SNHG7*	*SNHG7*
ENST00000653276.1	898	85		ENSG00000233016	*(DIPK1B-AS1)*
		85			
ENST00000654602.1	420	***TSNAX***		Novel transcript	*TSNAX-AS1*
		44		ENSG00000233461	
1a81bb5f-e4f2-4c47-93c1-6a3e8ca7a1f2	611	***AHR***		Novel transcript	*AHR-AS1*
ENST00000433005.1	469	13,012		ENSG00000237773	
		13,012			
ENST00000416869.2	332	***CROCC***		Novel transcript	*CROCC-AS1*
ENST00000422124.1	474	1097		ENSG00000238142	
ENST00000457075.1	330	904			
		1092			
ENST00000520944.6	415	***MCMDC2***		*SNHG6*	*SNHG6*
ENST00000521399.6	303	63		ENSG00000245910	*(MCMDC2-AS1)*
ENST00000686474.1	346	63			
		63			
ENST00000539259.1	433	***ERC1***		Novel transcript	*ERC1-AS1*
		614		ENSG00000250132	
ENST00000655818.1	558	***SNCAIP***		*MGC32805*	*MGC32805*
		9972		ENSG00000250328	*(SNCAIP-AS1)*
ENST00000508564.1	1369	***SMUG1***		LINC02381	LINC02381
		6746		ENSG00000250742	*(SMUG1-AS1)*
ENST00000515205.1	424	**ENSG00000170846**		*LINC02482*	*LINC02482*
ENST00000513179.1	413	3108		ENSG00000251580	(ENSG00000170846
		3103			*-AS1)*
ENST00000523657.2	749	***LYNX1-SLURP2***	***SLURP2***	Novel transcript	*SLURP2-AS1*
		2097	2093	ENSG00000253196	
ENST00000527564.2	794	***PPP3CA***		*FLJ20021*	*FLJ20021*
ENST00000529296.2	620	503		ENSG00000254531	*(PPP3CA-AS1)*
		503			
ENST00000529369.2	688	***DHCR7***		Novel transcript	*DHCR7-AS1*
		3492		ENSG00000254682	

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)		AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
ENST00000527726.1	372	***GALNT18*** 3699		Novel transcript ENSG00000255462	*GALNT18-AS1*
ENST00000659090.1	575	***ATP23*** 4592		*GIHCG* ENSG00000257698	*GIHCG* *(ATP23-AS1)*
26,641,821-b3a5-449c-b834-98207e980e6b	277	***CCDC196*** 1682		Novel transcript ENSG00000258561	*CCDC196-AS1*
ENST00000532213.2 2f486c7e-f133-4926-ae5a-bccbf4ff5548	443 1727	***NDRG2*** 4950 2133		Novel transcript ENSG00000258604	*NDRG2-AS1*
ENST00000565747.5	804	***COG7*** 408		Novel transcript ENSG00000260136	*COG7-AS1*
c2e04f15-9c7c-4471-970d-b8a8b5d20496	1430	***MAN2C1*** 458		Novel transcript ENSG00000260274	*MAN2C1-AS1*
eadb6a6d-92a4-461f-aae7-7feb3f3592b9	576	**ENSG00000284512** 576		Novel transcript ENSG00000261061	*ENSG00000284512-AS1*
a2030f1e-c1c6-4330-8de1-753fa97911dc	830	***TMEM238L*** 9192	***PIRT*** 15,618	Novel transcript ENSG00000263508	*PIRT-AS1*
95887790-14c3-4cab-a2ea-d9b552722b09	986	***SPDYE17*** 1456		*DTX2P1-UPK3BP1-PMS2P11* ENSG00000265479	*DTX2P1* *(SPDYE17-AS1)*
ENST00000443997.1	800	***RNASEK-C17orf49*** 1704	***C17orf49*** 1709	*MIR497HG* ENSG00000267532	*MIR497HG* *(RNASEK-AS1)*
656bbbc0-12b3-4de1-8f18-246f2685e7aa	561	***U2SURP*** 561		Novel transcript ENSG00000268129	*U2SURP-AS1*
ba023482-f1b6-4cf4-87f5-d000b3b02757	4107	***PDXDC1*** 22,975		*PKD1P6-NPIPP1* ENSG00000270580	*PKD1P6* *(PDXDC1-AS1)*
861975c5-4aac-4f25-b3b6-fa156f40e840	658	***VDAC1*** 71		Novel transcript ENSG00000271737	*VDAC1-AS1*
fef6ac5d-cbd0-4e9a-8fd2-ea5c1c8c9c0c	1854	***MIB2*** 342		Artifact ENSG00000272106	*MIB2-AS1*
d477fddc-3ef4-41eb-92b4-b68c1170bdbe	2043	***SNTB1*** 1957		Novel transcript ENSG00000272502	*SNTB1-AS1*
31e5077f-cc28-48af-b2db-cfe10e79deda	1231	***AMMECR1L*** 13		Novel transcript ENSG00000272667	*AMMECR1L-AS1*
ENST00000608612.1	420	***RAB11FIP5*** 613		Novel transcript ENSG00000272702	*RAB11FIP5-AS1*
e22a70a2-ddb2-482f-8d50-270d3558d2b1	1010	***CCNYL1*** 1010		Novel transcript ENSG00000272851	*CCNYL1-AS1*
ENST00000623943.2	1185	***G3BP1*** 464		Novel transcript ENSG00000275765	*G3BP1-AS1*
03095496-f5a1-44b7-a666-61a821e0e186	1598	***TM9SF2*** 2395		*LINC01232* ENSG00000280734	*LINC01232* *(TM9SF2-AS1)*
ENST00000629441.2	742	***DCPS*** 8287		*GSEC* ENSG00000280832 *(ST3GAL4-AS1)* *(DCPS-AS1)*	*GSEC* *(DCPS-AS1)*
13973899-1028-4379-a408-21f2d851788d	1710	***ZSWIM5*** 1710		*LINC01144* ENSG00000281912	*LINC01144* *(ZSWIM5-AS1)*
ENST00000479646.1	383	***KIAA1755*** 567		Novel transcript ENSG00000285347	*KIAA1755-AS1*
ENST00000535372.1	890	***ACAT2*** 288		*SOD2-OT1* ENSG00000285427	*SOD2-OT1* *(ACAT2-AS1)*
ENST00000507305.5	765	***FBXL5*** 11		Novel transcript ENSG00000288606	*FBXL5-AS1*
a88cfd62-b06e-4c24-a517-866daa7db5c1	659	**ENSG00000254692** 659		Novel transcript ENSG00000288820	*ENSG00000254692-AS1*

Table 2 (continued)

Transcript	Length (exons, bp)	Sense gene (genomic overlap, bp)	AS description/ ID/ Gene Synonyms or actual name	Proposed name (Alias)
9065a2c1-1743-4868-a4c7-59400f92428f	673	***CDK2AP2*** 181	Novel transcript ENSG00000289343	*CDK2AP2-AS1*

An exhaustive review of AS genes in *Ensembl* shows that the number of AS genes is not negligible. Considering only those named as AS (N=1685), such genes represent almost 7% of the total number of genes. Although the identification of AS and their nomenclature has improved in recent years, there are AS genes that need to be renamed.

More interesting is the biological relevance that AS may have. Our analysis started by evaluating the distribution of AS genes in the human genome. We highlight the significant enrichment of AS genes in chromosome 13. Although further studies are needed to explain this concentration of AS genes in said chromosome, it is probably related to the presence of large genes, with a higher number of alternative transcripts, and the enrichment in repetitive elements, being the difference between genes with AS from genes without AS genes. In addition, it is worth mentioning that AS genes may not be restricted to sense-coding protein genes. Fifteen AS genes corresponded to non-protein-coding genes, suggesting that AS may also modulate indirect regulators.

Our study also explored asRNAs using disruptive RNA sequencing methods. Nanopore sequencing allows the sequencing of long reads covering entire RNA molecules [16]. This is particularly interesting for the correct identification of the complete set of transcripts for a Sense or AS gene, allowing for more accurate transcript-level quantification (instead of gene-level, as provided by traditional short-reads NGS systems) and alternative transcripts discovery [17], both relevant features for long non-coding RNAs such as asRNAs. The study, performed in 15 human liver samples, revealed a relatively low amount of asRNAs (185 transcripts corresponding to 150 AS genes), less than reported in previous studies, based on short-read sequencing methods. This was also observed for the expression of sense genes in human liver: we detected 807 genes and the reported number was 1034 [18]. Thus, studies using this technology and other RNA-Seq methods need to be performed in different tissues to unravel the expression pattern of AS genes and to identify the strengths and limitations of this technology.

A clear strength of nanopore sequencing is its ability to identify novel transcripts. Our study reported 55 novel transcripts of AS genes (more than a quarter of the total identified forms). If alternative transcripts of coding genes may be involved in increasing protein heterogeneity, the biological relevance of alternative transcripts for AS genes remains to be further investigated.

Our study also confirmed the low expression of AS genes. However, there are some asRNAs with high expression. As mentioned, although there not inverse correlation between the expression of the AS gene and that of the sense gene, there are still some pairs that are worth mentioning. The most notable case is that of the gene *FGD5-AS1*, which not only exhibits higher expression but also corresponds with lower expression of its sense counterpart. The individual variability of these relationships and their potential impact on different disorders needs to be addressed in further studies. Furthermore, although sense genes exhibit higher expression compared to antisense genes, we identified 28.7% antisense genes that are expressing more than their sense counterparts. Despite the absence of a negative correlation in this co-expression, we continued to investigate the potential regulation of a transcript produced by an antisense gene relative to the transcript produced by a sense gene.

In this context, it is interesting to note the identification of 51 transcripts for 40 AS genes whose sense counterpart was not expressed. This finding could be explained by a regulatory role of the expression of the sense gene by the AS, although further studies are needed to validate this hypothesis.

Probably the most relevant result of this study was the identification of new potential AS genes. In this work, 146 different transcripts of 118 potential new AS genes were found, only in the liver and sequencing a relatively small number of samples. Interestingly, the number of potential new AS expressed in the human liver detected by nanopore sequencing was similar to the number of known AS genes whose transcripts were detected in the same samples (185 transcripts corresponding to 150 previously described AS). These results, the ability to detect whole transcripts using long reads, and the fact that for 75 of these new AS exist additional evidence supporting that they are transcripts of new AS genes, strongly suggest that the number of AS genes in the human genome could be higher. Thus, similar approaches can be performed in other tissues to complete the landscape of AS in the human genome and their physiopathological impact.

Conclusion

This study reveals the landscape of AS genes in the human genome and demonstrates the power of nanopore sequencing to provide insights into the relative

expression of sense and AS gene pairs and to identify novel transcripts and even novel AS genes. These results suggest that the human genome's antisense landscape may be more extensive than previously known, warranting further exploration.

Methods

Systematic search of AS genes in the human genome

Coding and AS genes were obtained from *Ensembl* Release 106 (GRCh38.p13), retaining lncRNA genes. Downstream analysis was performed with bash v3.2.57, Python v3.8.13, and R v4.2.2. Analysis was based on whether genes names contained "-AS" or not. Sense genes were compared with the antisense ones based on their names. BEDtools v2.28.0 (https://bedtools.readthedocs.io/en/latest/) was used to check for AS genes overlap with sense genes. Manual curation was required due to inconsistencies in AS gene nomenclature. IGV v.2.13.1 [19] was used to visualize chromosomal coordinates. Analyzed genes were screened for repetitive elements using RepeatMasker v4.1.2-p1 and UCSC nomenclature for family and classes [20]. GENCODE v39 [21] was used as a reference. The expression profiles of antisense genes by tissue were performed using normalized bulk tissue expression v1.1.9 from Adult GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2, https://gtexportal.org/home/).

Sequencing of human liver RNA using nanopore technology

Liver biopsy samples were obtained from 15 patients aged 32–68, recruited from bariatric surgery at Virgen de la Arrixaca University Hospital (Murcia, Spain). Fresh samples were snap frozen in less than 30 min after surgery and stored at -80ºC until further analysis.

RNA was extracted using TRIzol (Thermo Fisher Scientific, Madrid, Spain) according to a protocol for total RNA extraction. Subsequently, mRNA was purified using Poly-A magnetic beads (Thermo Fisher Scientific, Madrid, Spain).

Human liver mRNA samples were sequenced using MinION (Oxford Nanopore Technologies, Oxford, United Kingdom). Complete transcript reads were obtained using the Direct RNA Sequencing Kit SQK-RNA002 and the MinION Flowcell R9.4 for sequencing.

Basecalling was performed using MinKNOW v21.02.01 and Guppy (versions 3.0.4+e7dbc23 to 3.2.8+bd67289). Alignment was performed with minimap2 v2.17-r941 [22] against GRCh38. FLAIR v1.5.1 [23] was used for de novo assembly and quantification. SQANTI v4.1.3 [24] was used for quality control and transcript classification. Expression was reported as transcripts per million (TPM). Included transcripts had an expression level greater than 0.5 TPM, following Expression Atlas cutoffs [25].

Discovery of candidate new AS genes using long-reads RNA-Seq was based on BEDtools v2.28.0 (https://bedtools.readthedocs.io/en/latest/) and GENCODE v43 to find target protein-coding genes. Identification of novel AS genes was based on Python v3.9.12, R 4.1.2 and BEDTools v2.3.0. A lncRNA was considered antisense provided their transcript length was over 200 nucleotides, and they overlapped with a protein-coding gene described on GENCODE (≥1 nucleotide) in antisense orientation.

All mentioned tools have been executed with default parameters, following the corresponding developers' manuals.

To validate the new AS detected by RNA seq using nanopore technology, we performed qRT-PCR of selected transcripts, three previously described transcripts that were not annotated as AS, and another three novel transcripts reported in our study as new AS. These transcripts had to meet the following criteria: they needed to be polyexonic, with no overlap with other sequences, and demonstrate expression in at least three samples. qRT-PCR was done on the Applied Biosystems™ QuantStudio 5 device (Applied Biosystems™, Madrid, Spain) using custom primers specifically designed to amplify the novel transcripts (identification name AS1-AS6) (Supplementary Table11). Supplementary Figs. 9 and 11a shows the PCR design for AS6 amplification, along with the amplicon products analyzed on a 5% acrylamide gel, where they were compared to GAPDH PCR products from the qRT-PCR. The AS products from qRT-PCR and from the re-amplification of each AS from the qRT-PCR products are presented in Supplementary Fig. 10a, 10b and 11, respectively. Additionally, all AS PCR products were sequenced using nanopore sequencing, confirming the sequences.

Naming conventions

To avoid use of conflictive labeling conventions used across the literature, in this study we use "antisense gene" (AS gene) when referring to an antisense element at genomic level, whilst the transcripts of AS gene are referred as "asRNA".

Statistical analysis

Kolmogorov-Smirnov test with Lilliefors correction was used to check for normality. Comparisons between groups were made using nonparametric tests as the distributions were not normal. Wilcoxon signed-rank test was used to compare continuous variables. Statistical analysis was performed with R v4.2.2.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11017-3.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Supplementary Material 9

Supplementary Material 10

Supplementary Material 11

Supplementary Material 12

Supplementary Material 13

Supplementary Material 14

Supplementary Material 15

Supplementary Material 16

Supplementary Material 17

Supplementary Material 18

Supplementary Material 19

Supplementary Material 20

Supplementary Material 21

Supplementary Material 22

Author contributions
ME.M-B; B.M-B and J.C: designed the research, conducted the research, analyzed data, and wrote the article. JJ.R-C and P.G-R: conducted the research, analyzed data and wrote the article. M.L-L, R.C-R, and J.P: conducted the research, and analyzed data. B.R-M, and ML.L: analyzed data, and wrote the article.

Funding
The study was supported by Instituto de Salud Carlos III (ISCIII) through projects "PI21/00174", "PMP21/00052" and "PI21/00137", which were co-funded by the European Union and by the European Union – Next Generation EU, respectively; and by Fundación Séneca (21886/PI/22). JJ R-C holds a predoctoral contract from Sociedad Española de Trombosis y Hemostasia (SETH-FETH). P.G-R and B.M-B hold contracts from CIBERER - ISCIII (CB15/00055); ME M-B has a Ramon y Cajal contract at the University of Murcia (Ministerio de Ciencia e Innovacion RYC2021-031000-I); R C-R has a PFIS predoctoral contract (ISCIII co-funded by the Fondo Europeo de Desarrollo Regional-FEDER) and M L-L has a predoctoral contract from Instituto de Salud Carlos III ("PMP21/00052"). J P holds contract from Servicio Murciano de Salud.

Data availability
All data created in this study are included as Supplementary Material in this article. Raw se-quencing data for the human liver samples are available on the European Nucleotide Archive under project ID PRJEB72070. Description of transcripts found on such samples have been uploaded as GTF and FASTA with this article. Further data are available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Clinical Research Ethics Committee at Virgen de la Arrixaca University Hospital and conducted in accordance with the 1964 Declaration of Helsinki and their later amendments. All included subjects gave their written informed consent to enter the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Servicio de Hematología, Centro Regional de Hemodonación, Hospital General Universitario Morales Messeguer, Universidad de Murcia, IMIB-Pascual Parrilla, CIBERER-ISCIII, Ronda de Garay s/n, Murcia 30003, Spain
[2]Grupo de Obesidad, Diabetes y Metabolismo, Instituto Murciano de Investigación Biosanitaria (IMIB-Pascual Parrilla), Murcia, Spain

Received: 9 January 2024 / Accepted: 8 November 2024
Published online: 27 November 2024

References

1. Xu Jzhong, Zhang J lan, Zhang Wguo. Antisense RNA: the new favorite in genetic research. J Zhejiang Univ-Sci B. 2018;19(10):739–49.
2. Khorkova O, Myers AJ, Hsiao J, Wahlestedt C. Natural antisense transcripts. Hum Mol Genet. 2014;23(R1):R54–63.
3. Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete S, Raya FJ. P, Implicaciones Del ARN no codificante en biología y evolución: desde Los primeros homínidos hasta Los Humanos Modernos - Revisión. Archaeobios. 2020;14(1):120–31.
4. Wright MW. A short guide to long non-coding RNA gene nomenclature. Hum Genomics. 2014;8(1):7.
5. Shen T, Li H, Song Y, Yao J, Han M, Yu M, et al. Antisense transcription regulates the expression of sense gene via alternative polyadenylation. Protein Cell. 2018;9(6):540–52.
6. Seal RL, Tweedie S, Bruford EA. A standardised nomenclature for long non-coding RNAs. IUBMB Life. 2023;75(5):380–9.
7. Albrecht AS, Ørom UA. Bidirectional expression of long ncRNA/protein-coding gene pairs in cancer. Brief Funct Genomics. 2016;15(3):167–73.
8. Li B, Hu Y, Li X, Jin G, Chen X, Chen G, et al. Sirt1 antisense long noncoding RNA promotes Cardiomyocyte Proliferation by enhancing the Stability of Sirt1. J Am Heart Assoc. 2018;7(21):e009700.
9. Deng Y, Wei Z, Huang M, Xu G, Wei W, Peng B, et al. Long non-coding RNA F11-AS1 inhibits HBV-related hepatocellular carcinoma progression by regulating NR1I3 via binding to microRNA-211-5p. J Cell Mol Med. 2020;24(2):1848–65.
10. Liu B, Xiang W, Liu J, Tang J, Wang J, Liu B, et al. The regulatory role of antisense lncRNAs in cancer. Cancer Cell Int. 2021;21(1):459.
11. Zhao H, Xu Q. Long non-coding RNA DLX6-AS1 mediates proliferation, invasion and apoptosis of endometrial cancer cells by recruiting p300/E2F1 in DLX6 promoter region. J Cell Mol Med. 2020;24(21):12572–84.
12. Barman P, Reddy D, Bhaumik SR. Mechanisms of Antisense Transcription Initiation with implications in Gene expression, genomic Integrity and Disease Pathogenesis. Non-Coding RNA. 2019;5(1):11.
13. Nishizawa M, Ikeya Y, Okumura T, Kimura T. Post-transcriptional inducible gene regulation by natural antisense RNA. Front Biosci-Landmark. 2015;20(1):1–36.
14. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–65.
15. Girbig M, Misiaszek AD, Müller CW. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. Nat Rev Mol Cell Biol. 2022;23(9):603–22.
16. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-generation sequencing: the spearhead towards the Radical Transformation of Modern Genomics. Life Basel Switz. 2021;12(1):30.
17. Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, et al. Transcriptome variation in human tissues revealed by long-read sequencing. Nature. 2022;608(7922):353–9.
18. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, et al. Comparison of the transcriptional landscapes between human and mouse tissues. Proc Natl Acad Sci. 2014;111(48):17224–9.
19. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the Integrative Genomics Viewer. Cancer Res. 2017;77(21):e31–4.
20. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet TIG. 2000;16(9):418–20.

21. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. Nucleic Acids Res. 2021;49(D1):D916–23.
22. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
23. Tang AD, Soulete CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat Commun. 2020;11(1):1438.
24. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, del Risco H, et al., et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28(3):396–411.
25. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, et al. Expression Atlas update: from tissues to single cells. Nucleic Acids Res. 2020;48(D1):D77–83.

Publisher's note