BMC Bioinformatics

SOFTWARE

Open Access

AStruct: detection of allele-specific RNA secondary structure in structuromic probing data

Qingru Xu[1,2†], Xiaoqiong Bao[1†], Zhuobin Lin[3†], Lin Tang[1], Li-na He[1], Jian Ren[1], Zhixiang Zuo[1*] and Kunhua Hu[3*]

†Qingru Xu, Xiaoqiong Bao and Zhuobin Lin contributed equally to this work.

*Correspondence: zuozhx@sysucc.org.cn; hukunh@mail.sysu.edu.cn

[1] State Key Laboratory of Oncology in South China, Cancer Center, Collaborative Innovation Center for Cancer Medicine, School of Life Sciences, Sun Yat-sen University, Guangzhou 510060, China
[2] Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[3] Guangdong Key Laboratory of Liver Disease Research, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Abstract

Background: Uncovering functional genetic variants from an allele-specific perspective is of paramount importance in advancing our understanding of gene regulation and genetic diseases. Recently, various allele-specific events, such as allele-specific gene expression, allele-specific methylation, and allele-specific binding, have been explored on a genome-wide scale due to the development of high-throughput sequencing methods. RNA secondary structure, which plays a crucial role in multiple RNA-associated processes like RNA modification, translation and splicing, has emerged as an essential focus of relevant research. However, tools to identify genetic variants associated with allele-specific RNA secondary structures are still lacking.

Results: Here, we develop a computational tool called 'AStruct' that enables us to detect allele-specific RNA secondary structure (ASRS) from RT-stop based structuromic probing data. AStruct shows robust performance in both simulated datasets and public icSHAPE datasets. We reveal that single nucleotide polymorphisms (SNPs) with higher AStruct scores are enriched in coding regions and tend to be functional. These SNPs are highly conservative, have the potential to disrupt sites involved in m6A modification or protein binding, and are frequently associated with disease.

Conclusions: AStruct is a tool dedicated to invoke allele-specific RNA secondary structure events at heterozygous SNPs in RT-stop based structuromic probing data. It utilizes allelic variants, base pairing and RT-stop information under different cell conditions to detect dynamic and functional ASRS. Compared to sequence-based tools, AStruct considers dynamic cell conditions and outperforms in detecting functional variants. AStruct is implemented in JAVA and is freely accessible at: https://github.com/canceromics/AStruct.

Keywords: Allele-specific events, RNA secondary structure, Structuromic probing data, Functional variants

Introduction

Numerous genetic variants have been reported as risk or causative factors for diseases at the population level, as they disrupt DNA structures or regulatory elements and thereby impact transcription [1, 2]. There is increasing evidence that genetic variants can also

affect RNA splicing [3], RNA secondary structure [4], RNA *N6*-methyladenosine (m6A) modification [5], and protein binding [6–8], etc. In individuals, these variants can act in an allele-specific manner, which are termed allele-specific events. Investigating the potential variants involved in allele-specific events is an important and effective approach to discovering and annotating functional variants.

RNA secondary structure is an essential feature that exerts a significant influence on several stages of the RNA life cycle, including RNA transcription [9], splicing [10] and translational control [11]. Facilitated by next-generation sequencing, a number of innovative techniques have been developed that combine traditional chemical probing methods with high-throughput sequencing to capture genome-wide RNA secondary structure. Foremost among them are reverse transcription stop (RT-stop) based structural probing techniques, including SHAPE-seq [12], icSHAPE [13] and smartSHAPE [14], etc.

Some genetic variants have been found to cause local or global RNA secondary structure changes [4] and have been proved to be associated with many genetic diseases [15, 16]. In the case of heterozygous variants, two alleles might have distinct effects on RNA secondary structure, namely allele-specific RNA secondary structure (ASRS). Several computational methods have been developed to predict ASRS without considering allele dosage [17–20]. However, analyzing ASRS remains a great challenge. Since the existing tools rely primarily on static sequence information, they are unable to capture the complex and dynamic cell conditions from the experimentally-derived data. Moreover, the RT-stop feature of sequencing technologies poses a challenge in effectively segregating reads by alleles and retaining sufficient structural information for analyzing the structure of each allele within a single sample.

To address the limitations mentioned above, we developed a computational tool called 'AStruct'. To the best of our knowledge, AStruct is the first software capable of identifying allele-specific RNA secondary structure events at heterozygous single nucleotide polymorphisms (SNPs) within one sample using RT-stop based structuromic probing data.

Methods

Calculating allelic structure score from icSHAPE sequencing data

Six icSHAPE datasets of six human cell lines were downloaded from GEO [21] (Gene Expression Omnibus). Specifically, the K562, HepG2, HEK293, HeLa, and H9 datasets were obtained from GSE145805 [22], while the HEK293T dataset was obtained from GSE74353 [23]. smatSHAPE dataset of HEK293T was accessed through number GSE155961 [14]. Each dataset comprised two replicates of NAI-N3 treated samples and two replicates of control (DMSO) samples. The raw reads of each dataset were preprocessed using Trimmomatic [24] to remove barcodes and adapters. The preprocessed reads were then mapped to the human genome (GRCh38.p12) using STAR [25]. PCR duplicates were discarded by collapsing the reads with identical sequences.

673,668,919 human SNPs were download from NCBI SNP Database (dbSNP) (GCF000001405.38) [26]. Read coverage was calculated for each SNPs in each aligned sample. Only the heterozygous SNPs with reads covering both the reference (Ref) allele and the alternate (Alt) allele were retained. Moreover, the SNPs with total reads less than

10 or Ref reads less than 2 or Alt reads less than 2 were filtered out. For each retained SNP, a SNP window was determined by the start base of the most upstream SNP-spanning read and the end base of the most downstream SNP-spanning read. The structure score for each base in the window was calculated for Ref allele and Alt allele separately according to the following method modified from icSHAPE-pipe [27]:

Firstly, we calculated the scores for the RT-stop events and background in a SNP window using the reads overlapping the SNP window. Reads overlapping the SNP window but not spanning the SNP would be utilized for both Ref and Alt alleles. The first upstream base of a read mapping start represented an RT-stop event (R) and all the bases in a read represented the background (B). Accordingly, R score and B score were calculated for each base in the window using all reads.

Secondly, to smooth the R scores and B scores, all the values were divided by a normalization factor which was the 95th percentile of all scores in the window (see formulas 1 and 2).

$$R_i = R_i / \boldsymbol{R}[0.95 \times N] \tag{1}$$

$$B_i = B_i / \boldsymbol{B}[0.95 \times N] \tag{2}$$

where $\boldsymbol{R}$ were arranged in ascending order as $R_1 \leq R_2 \leq \ldots \leq R_N$ representing the RT-stop event occurring in all bases, $\boldsymbol{B}$ were arranged in ascending order as $B_1 \leq B_2 \leq \ldots \leq B_N$ representing the background in all bases, i was the *ith* base in the window, and N was the window length.

Thirdly, an enrichment score (ES) was calculated for each base in the window using the smoothed R scores and B scores in NAI-N_3 treated samples and DMSO control samples (see formula 3).

$$ES_i = \left(R_i^T - s \times R_i^C\right) \Big/ B_i^C \tag{3}$$

where ES_i was the enrichment score for the *ith* base in the window, R_i^T was the R score for the *ith* base in the window in treated sample, R_i^C was the R score for the *ith* base in the window in control sample, B_i^C was the B score for the *ith* base in the window in control sample, and s was the predefined factor.

Fourthly, the enrichment score was converted between 0 and 1 using formula 4 to represent the structure score.

$$ES_i = \max\left(0, \min\left(1, \frac{ES_i - \boldsymbol{ES}[0.05 \times N]}{\boldsymbol{ES}[0.95 \times N] - \boldsymbol{ES}[0.05 \times N]}\right)\right) \tag{4}$$

where $\boldsymbol{ES}$ were arranged in ascending order as $ES_1 \leq ES_2 \leq \ldots \leq ES_N$.

Finally, we obtained a set of structure scores for *Ref* allele and *Alt* allele, separately.

Statistical test for the structure difference between Ref allele and Alt allele

To test the structure difference between *Ref* allele and *Alt* allele, we firstly calculated *Pearson Correlation* between the structure scores of two alleles. Then the *Pearson Correlation Coefficient* was converted to the experimental structural disruption coefficient (*eSDC*) as described below [28] (see formula 5).

$$eSDC = \left(1 - {}^{p}CC\right) \times \sqrt{N} \tag{5}$$

where ${}^{p}CC$ was the *Pearson Correlation Coefficient* of the structure scores between two alleles, and N was the length of the SNP window. The larger *eSDC* meant more difference in structure scores between the two alleles.

To evaluate the statistical significance of the difference, 1000 permutations were performed by assigning reads to *Ref* and *Alt* alleles following 1:1 ratio based on *Poisson distribution*. The P value of the difference was calculated by comparing the true *eSDCs* and permuted *eSDCs*. Some studies have shown the robustness of measuring the gene expression difference by combining the magnitude and significance [29]. Thus, borrowing from this idea, we also combined *eSDC* (magnitude) and P (statistical significance) to measure the structure difference. Because the reads that overlapping with the SNP window but not spanning SNP smoothed the structure difference between two alleles, we used the ratio between the number of reads spanning SNP and total reads in the window to adjust the difference score. Finally, the AStruct score was defined as formula 6:

$$AStruct\ score = -\log_{10} P \times eSDC \times \left(1 - \log_{10}\left(\frac{R_d}{R}\right)\right) \tag{6}$$

where $\boldsymbol{R_d}$ was the number of reads spanning SNP, $\boldsymbol{R}$ was the number of total reads in the SNP window.

Calculating the structure difference for each base in a SNP window

To provide more details of the structure difference between the two alleles for each base, we applied the method from a previous study [30] as described below (see formula 7):

$$StrucDiff_i = \sum_{k=i-2}^{k=i+2} abs\left(ES_{k,Alt} - ES_{k,\mathrm{Ref}}\right)/_5 \tag{7}$$

A sequence of P values named *StrucDiff* for each base were obtained by permutation test.

Performance evaluation using simulated datasets

To evaluate the reliability of our method, we simulated the icSHAPE reads under different read depths (10M, 20M,…,100M) based on *Poisson distribution*. The pair or unpair structure information were generated from RNAsubopt [31]. In order to simulate the RT-stop event, the artificial stop intervention was evenly added on the unpaired bases. For the sequences without allelic structure difference, *Ref* and *Alt* alleles were designed to have the same set of unpaired bases (RT-stop bases). For the sequences with allelic structure difference, *Ref* and *Alt* alleles were supposed to have different sets of unpaired bases.

Functional annotation of allele specific RNA secondary structure

All the retained SNP sites were divided into three groups according to the magnitude of AStruct score, which are 'Low' (0), 'Medium' (0, 1], and 'High' $(1, +\infty)$. For RNAsnp, the 'High-RNAsnp' and 'Low-RNAsnp' groups were defined by smallest 10% r and largest

10% *r* values of the same SNP sites [20]. The *r* value represented the *Pearson Correlation Coefficient* for the structural comparison between the reference sequence and the alternative sequence (±100bps of SNP sites). SNPs in linkage disequilibrium (LD) were retrieved from HaploReg ($r^2 > 0.8$) [32].

To explore conservation, we gathered PhastCon100 scores from UCSC genome browser [33]. In examining the relationship between ASRS and other AS events, genetic variants affecting RNA modifications (mainly m6Asnp) were obtained from RMVar [34]. Allele-specific RBP binding (ASRBP) were collected from ADASTRA [6] overlapped with two RBP datasets [35, 36]. To investigate ASRS association with diseases, FATHMM-XF scores were annotated to all SNPs [37]. Disease-related SNPs were obtained from ClinVar [38].

Results

AStruct pipeline for evaluating allele-specific structure events

AStruct is implemented in Java (JDK 8), and simply takes sorted BAM files from structure sequencing data as input. The rationale behind AStruct is that the two alleles involved in an allele-specific structure event are expected to have dissimilar structure scores using structure sequencing technologies such as icSHAPE and smartSHAPE [13, 14]. The workflow of AStruct was illustrated in Fig. 1a. First, the preprocessed sequencing reads were aligned to the reference genome using STAR [25]. Second, all the SNPs obtained from dbSNP [26] were used to search against the alignment data, and only the heterozygous SNPs were retained for further analysis. Third, the aligned reads around each heterozygous SNP were separated into two groups according to the SNP, with each group representing one allele. Finally, the AStruct score for each SNP was calculated based on the similarity test between the structure scores of two alleles. A higher AStruct score indicated higher allele specificity of RNA secondary structure.

AStruct is robust in predicting allele-specific structure events

To evaluate the performance of AStruct, we simulated 10 icSHAPE datasets with sequencing depths ranging from 10 to 100M. Each dataset consisted of two control samples and two treated samples. The AUC values of the simulated datasets ranged from 0.712 to 0.892, indicating a robust performance of AStruct. As expected, the performance of AStruct improved with the increase of sequencing depth (Fig. 1b).

We next applied AStruct to public icSHAPE datasets of six different cell lines. As a result, we obtained 1616, 3046, 930, 4415, 2591, and 8640 heterozygous SNPs with enough reads coverage and enough extra reads stop of structure information from K562, HepG2, HEK293, HEK293T, HeLa, and H9, respectively (Fig. 1c). According to the AStruct score, the heterozygous SNPs above were categorized into three groups: the 'Low' group with a score of 0, the 'Medium' group with a score ranging from 0 to 1, and the 'High' group with a score greater than 1. We applied a published method [30] to measure the structure difference at the single base level. As shown in Fig. 1d, the 'High' group had much more bases with significantly different structure scores between two alleles, compared to the 'Low' group and 'Medium' groups. We found a high correlation between AStruct scores of different cell lines (Fig. 1e), further demonstrating the robustness of AStruct. Additionally, we found that SNPs with higher AStruct scores were more

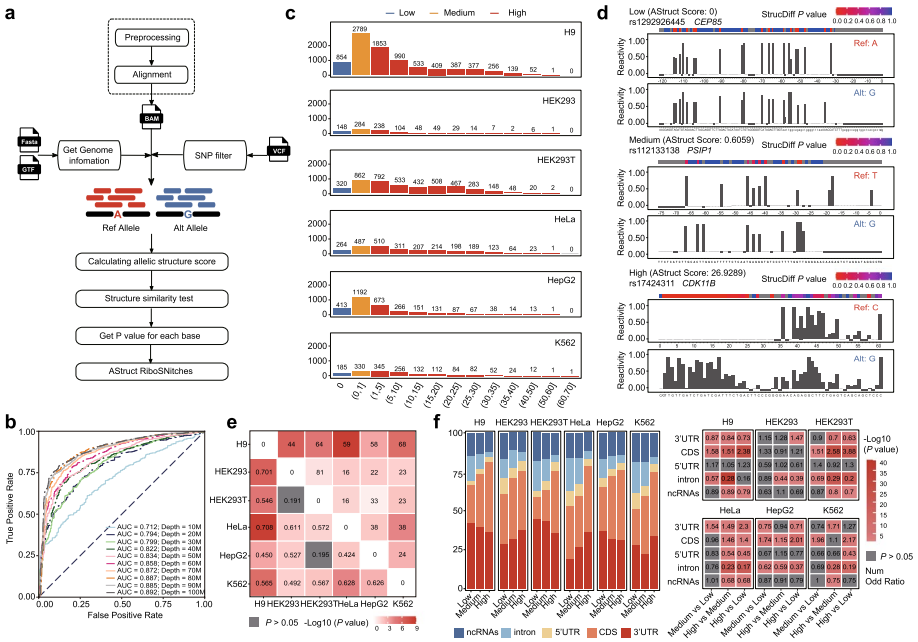

Fig. 1 Robustness of AStruct in predicting allele-specific structure events. **a** Flowchart of AStruct pipeline. **b** ROC curves of simulated datasets under different sequencing depths. **c** The AStruct score distribution of all heterozygous SNPs in six cell lines (H9, HEK293, HEK293T, HeLa, HepG2, and K562). **d** The allele level reactivity score derived from the icSHAPE data of K562 cell line for three SNPs representing the three AStruct group (Low: rs1292926445; Medium: rs12133138; High: rs17424311). The *P* values indicating the difference in reactivity score between two alleles for each base were shown in the top. **e** The correlation of ASRS SNPs between different cell lines. The value in top-right triangle presents the number of common ASRS SNPs; the value in bottom-left triangle presents the *Pearson correlation coefficient* of the AStruct scores. Background color presents the *P* Value, where gray indicates $P > 0.05$. **f** Distribution of ASRS SNPs in different genomic regions (3′UTR, CDS, 5′UTR, intron, and ncRNAs). Left: Stacked bar plot for three AStruct groups. Right: Comparison the sites number of different regions between different groups (*Fisher's exact test*). The number represents the Odds Ratio. Background color represents the significant *P* Value

likely to be located in the CDS region rather than the intron region (Fig. 1f). Moreover, AStruct could also be applied to smartSHAPE, another RT-stop based sequencing technology, implying broader applicability (Additional file 1: Fig. 1).

AStruct can help identify functional variants

Numerous causal SNPs associated with traits or diseases were defined in previous GWAS studies, but most of them were usually not genotyped but are in linkage disequilibrium (LD) with the genotyped SNPs [39]. Recent studies on SNP pairs in high LD indicated their interplay on m^6A modification [40] and RNA structure [41]. Consequently, in order to take the effects of LD-SNPs into account, we annotated the LD-SNPs ($r^2 > 0.8$) for ASRS SNPs from HaploReg [8] and grouped ASRS SNPs with their LD-SNPs into new 'High', 'Medium' and 'Low' groups for the following functionality analyses. We presumed that SNPs in the 'High' group are prone to include allele-specific structure features than SNPs in the 'Medium' or 'Low' groups.

As shown in Fig. 2a, SNPs with higher AStruct scores illustrated a higher degree of conservation, highlighting their important function. As for the same test on RNAsnp, which predicted ASRS based on sequence, only three cell lines showed significant

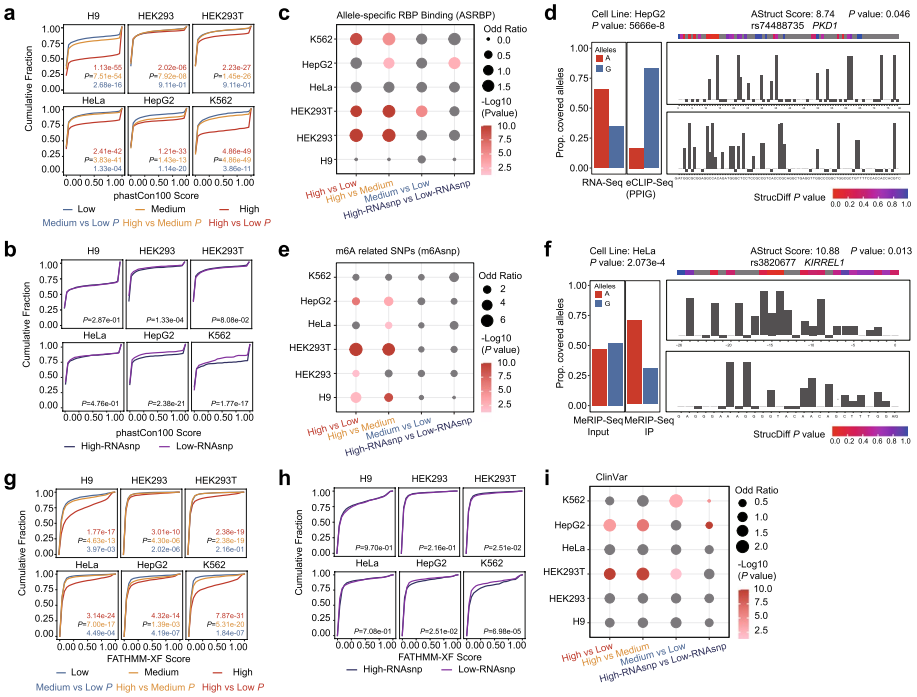

Fig. 2 AStruct is valuable in assisting with the identification of functional variants. **a**, **b** The cumulative distribution curve of the phastCon100 score for different groups. a for AStrcut, b for RNAsnp. **c** The bubble plot shows the comparison of the number of ASRS SNPs annotated as allele-specific RBP binding events between different groups. **d** A G-to-A SNP (rs74488735) in the 'High' group shows allele-specific pattern of a PPIG binding site in PXD1 of HepG2 cell line. It is illustrated by the different proportion of reads count covering each allele in the eCLIP-Seq data and the control RNA-Seq data (left), and the significant structure difference with each allele (right). **e** The bubble plot shows the comparison of the number of ASRS SNPs annotated as m6A related variants between different groups. **f** A A-to-G SNP (rs3820677) in the 'High' group shows allele-specific pattern of a m6A modification site in KIRREL1 of HeLa cell line. It is illustrated by the different proportion of reads count covering each allele in the MeRIP-Seq IP data and the control Input data (left), and the significant structure difference with each allele (right). **g**, **h** The cumulative distribution curve of the FATHMM-XF score for different groups. **g** For AStrcut, h for RNAsnp. **i** The bubble plot shows the comparison of the number of disease-related ASRS SNPs annotated by ClinVar between different groups. The *P* value of cumulative distribution curve was calculated using *Kolmogorov–Smirnov (KS) test*. The *P* value of bubble plot was calculated using *Fisher's exact test*

conservation differences (Fig. 2b). This observation led us to concentrate on investigating the biological features associated with ASRS.

ASRS SNPs with high AStruct scores demonstrated an enrichment of allele-specific RBP binding sites (ASRBP) and m6A related SNPs (m6Asnp) compared to SNPs with medium or low scores. Again, we did not detect any significant differences between 'High-RNAsnp' and 'Low-RNAsnp' groups (Fig. 2c and 2e). Besides, we found 39 original ASRS SNPs with high AStruct scores in HepG2 cell line that can induce allele-specific binding pattern of the PPIG RBP (Additional file 1: Table 1). For example, the SNP rs74488735 located in *PKD1* showed both allele-specific RNA secondary structure (AStruct score = 8.74, $P = 0.046$) and allele-specific PPIG RNA binding pattern ($P = 5.666e{-}8$) (Fig. 2d). Moreover, we found 34 original ASRS SNPs with high AStruct scores in HeLa cell line that can influence the m6A modification (Additional file 1: Table 2). For instance, the SNP rs3820677 located in *KIRREL1* showed both

allele-specific RNA secondary structure (AStruct score = 10.88, $P = 0.013$) and allele-specific m6A modification ($P = 2.073e{-}4$) (Fig. 2f).

Furthermore, to investigate the association between ASRS SNPs and disease, the FATHMM-XF score was used to evaluate the pathogenicity [37]. A higher FATHMM-XF score means mutations are more likely to be pathogenic, while a lower score means more benign. We found that SNPs with higher AStruct scores were more pathogenic (Fig. 2g) while RNAsnp did not reveal any significant differences [20] between different groups (Fig. 2h). We also annotated the ASRS SNPs and their LD-SNPs using ClinVar database [38], and found SNPs with higher AStruct scores had a higher proportion of variants associated with disease, such as Lynch syndrome, Argininosuccinate lyase deficiency, and Farber disease (Fig. 2i, Additional file 2: Table 3).

Taken together, AStruct demonstrated superior capability in detecting functional ASRS SNPs, in contrast to tools that only focused on static sequence information. ASRS SNPs with high AStruct scores tended to be more conserved and were more likely to influence other allele-specific events such as RNA–protein interactions and RNA modifications. These ASRS SNPs are also potentially linked with specific traits- or diseases-related variants.

Discussion

Despite the fact that the impact of genetic variants on RNA structure and transcriptional regulation were widely acknowledged, specialized tools to systematically profile ASRS have been lacking. Identification of ASRS events is expected to contribute to revealing functional variants and illuminating the molecular mechanism of associated diseases at the allele level. Here, we provided AStruct as a robust algorithm for detecting ASRS from RT-stop based structuromic probing data.

Theoretically, RNAsnp and other sequence-based tools are designed to evaluate the impact of SNPs on local RNA secondary structure. They typically utilize thermodynamic models on the input sequence. However, the structures and functions are mostly varied because of cell specificity, cell status and their microenvironment that must be considered. Taking full advantage of experimental structure sequencing technologies, AStruct utilized allelic variants, base pairing, and RT-stop information under in vivo conditions to detect dynamic and functional ASRS. In practice, we analyzed the association of ASRS ANPs detected from AStruct and RNAsnp with other AS events as well as diseases. The results further demonstrated the superiority of AStruct. Additional tests on smartSHAPE further indicated the broad applicability of AStruct.

However, it is worth noting that read coverage is important for recovering RNA structure and providing enough SNPs coverage for statistical analyses. In this study, we set coverage threshold for each SNPs. Because of the sparsity of alleles and the truncation in RT-stop data, only a part of SNPs satisfied the requirements and were used as candidates. Therefore, a complete mapping of ASRS have not yet been established. These limitations can be overcome by using high sequencing depth data and high resolution technologies. Moreover, we are working to update AStruct to adapt it to RT-Mut structure sequencing data. While not mentioned in this article, we had successfully tested AStruct on SHAPE-MaP datasets and obtained similar results. It is expected that AStruct can be applied

to more types of structure sequencing data and provide insights to allele-specific RNA secondary structure.

Conclusions

In conclusion, AStruct can be an excellent candidate for detecting allele-specific RNA secondary structure. The advantages of AStruct are obvious. Using high-throughput RT-stop based experimental data, AStruct has the capability to capture the allele-specific RNA secondary structure in the real cellular environment. In addition, AStruct allows allele-specific comparisons within a single sample by accurately separating alleles based on heterozygous SNPs, resulting in more accurate and reliable results.

Availability and requirements

Project name: AStruct. Project homepage: https://github.com/canceromics/AStruct. Operating system: Platform independent. Programming language: Java. Other requirements: JDK 8. License: GNU GPL v3. Any restrictions to use by non-academics: none.

Abbreviations

AS	Allele-specific
ASE	Allele-specific gene expression
ASM	Allele-specific methylation
ASB	Allele-specific binding
ASRS	Allele-specific RNA secondary structure
ASRBP	Allele-specific RNA-binding protein
RBP	RNA-binding protein
RT	Reverse transcription
PCR	Polymerase chain reaction
SHAPE-seq	Selective 2′-hydroxyl acylation analyzed by primer extension and sequencing
icSHAPE	In vivo click selective 2-hydroxyl acylation and profiling experiment
smartSHAPE	Small amount random RT icSHAPE
NAI-N3	2-Methylnicotinic acid imidazolide, by the addition of an azide to the nicotinic acid ring
DMSO	Dimethylsulfoxide
SNP	Single nucleotide polymorphism
m^6A	N6-methyladenine modification
Ref	Reference
Alt	Alternate
ES	Enrichment score
eSDC	Experimental structural disruption coefficient
LD	Linkage disequilibrium

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-024-05704-x.

Additional file 1: Fig. 1. AStruct in smartSHAPE. **a** Venn diagram of icSHAPE and smartSHAPE results. Pearson Correlation was calculated using the AStruct scores of the intersection set. **b** The cumulative distribution curve of the FATHMM-XF score for three AStruct groups identified using smartSHAPE. **Table 1.** ASRS SNPs in 'High' group result in allele-specific RNA-protein-interaction of PPIG protein in HepG2 cell line. **Table 2.** ASRS SNPs in 'High' group result in allele-specific m6A modification in HeLa cell line.

Additional file 2: Table 3. Disease annotation for ASRS SNPs and LD-SNPs.

Acknowledgements
Not applicable.

Author contributions
HKH, ZZX designed the work. QXR, BXQ, LZB contributed to the algorithm development and conducted the programming. TL, HLN tested the software. RJ provided conceptual guidance. QXR, BXQ wrote the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by grants from the Program for Guangdong Basic and Applied Basic Research Foundation [2021B1515020108]; Guangdong Introducing Innovative and Entrepreneurial Teams [2017ZT07S096]; the Guangzhou Municipal Institute Enterprise Joint Funding Project [SL2024A03J00312]; the Guangdong Provincial Natural Science Foundation General Project [2021A1515010780].

Availability of data and materials
Public datasets used in this article are available under GEO accession number GSE145805, GSE74353 (icSHAPE), GSE155961 (smartSHAPE), GSE198145 (RNA Seq), GSE91478 (CLIP Seq), and GSE102493 (MeRIP Seq).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 23 October 2023 Accepted: 14 February 2024
Published online: 01 March 2024

References

1. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. Nature. 2020;578:112–21.
2. Genetic associations of protein-coding variants in human disease—PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8891017/. Accessed 28 Jan 2024.
3. Blakes AJM, Wai HA, Davies I, Moledina HE, Ruiz A, Thomas T, et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 genomes project. Genome Med. 2022;14:79.
4. Lokody I. RiboSNitches reveal heredity in RNA secondary structure. Nat Rev Genet. 2014;15:219–219.
5. Cao S, Zhu H, Cui J, Liu S, Li Y, Shi J, et al. Allele-specific RNA N6-methyladenosine modifications reveal functional genetic variants in human tissues. Genome Res. 2023;33:1369–80.
6. Abramov S, Boytsov A, Bykova D, Penzar DD, Yevshin I, Kolmykov SK, et al. Landscape of allele-specific transcription factor binding in the human genome. Nat Commun. 2021;12:2751.
7. Yang E-W, Bahn JH, Hsiao EY-H, Tan BX, Sun Y, Fu T, et al. Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. Nat Commun. 2019;10:1338.
8. Bahrami-Samani E, Xing Y. Discovery of allele-specific protein-RNA interactions in human transcriptomes. Am J Hum Genet. 2019;104:492–502.
9. Wanrooij PH, Uhler JP, Simonsson T, Falkenberg M, Gustafsson CM. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc Natl Acad Sci. 2010;107:16072–7.
10. Buratti E, Baralle FE. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol. 2004;24:10505–14.
11. Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene. 2005;361:13–37.
12. Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB. SHAPE–Seq: high-throughput RNA structure analysis. Curr Protoc Chem Biol. 2012;4:275–97.
13. Flynn RA, Zhang QC, Spitale RC, Lee B, Mumbach MR, Chang HY. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat Protoc. 2016;11:273–90.
14. Piao M, Li P, Zeng X, Wang X-W, Kang L, Zhang J, et al. An ultra low-input method for global RNA structure probing uncovers Regnase-1-mediated regulation in macrophages. Fundam Res. 2022;2:2–13.
15. Aouacheria A, Navratil V, López-Pérez R, Gutiérrez NC, Churkin A, Barash D, et al. In silico whole-genome screening for cancer-related single-nucleotide polymorphisms located in human mRNA untranslated regions. BMC Genom. 2007;8:2.
16. Ben-Hamo R, Zilberberg A, Cohen H, Bahar-Shany K, Wachtel C, Korach J, et al. Resistance to paclitaxel is associated with a variant of the gene BCL2 in multiple tumor types. Npj Precis Oncol. 2019;3:1–11.
17. Salari R, Kimchi-Sarfaty C, Gottesman MM, Przytycka TM. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. Nucleic Acids Res. 2013;41:44–53.
18. Lin J, Chen Y, Zhang Y, Ouyang Z. Identification and analysis of RNA structural disruptions induced by single nucleotide variants using Riprap and RiboSNitchDB. NAR Genom Bioinform. 2020;2:lqaa057.
19. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet. 2010;6:e1001074.
20. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. Hum Mutat. 2013;34:546–56.
21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41:D991–5.
22. Sun L, Xu K, Huang W, Yang YT, Li P, Tang L, et al. Predicting dynamic cellular protein–RNA interactions by deep learning using in vivo RNA structures. Cell Res. 2021;31:495–516.

23. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA duplex map in living cells reveals higher-order transcriptome structure. Cell. 2016;165:1267–79.
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.
25. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
26. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
27. Li P, Shi R, Zhang QC. icSHAPE-pipe: a comprehensive toolkit for icSHAPE data analysis and evaluation. Methods. 2020;178:96–103.
28. Ritz J, Martin JS, Laederach A. Evaluating our ability to predict the structural disruption of RNA by SNPs. BMC Genom. 2012;13:S6.
29. Xiao Y, Hsiao T-H, Suresh U, Chen H-IH, Wu X, Wolf SE, et al. A novel significance score for gene selection and ranking. Bioinformatics. 2014;30:801–7.
30. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature. 2014;505:706–9.
31. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.
32. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease|Nucleic Acids Research|Oxford Academic. https://academic.oup.com/nar/article/44/D1/D877/2503117. Accessed 28 Jan 2024.
33. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al. The UCSC genome browser database: 2021 update. Nucleic Acids Res. 2021;49:D1046–57.
34. Luo X, Li H, Liang J, Zhao Q, Xie Y, Ren J, et al. RMVar: an updated database of functional variants involved in RNA modifications. Nucleic Acids Res. 2021;49:D1405–12.
35. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. Nucleic Acids Res. 2011;39(suppl_1):D301–8.
36. Caudron-Herger M, Jansen RE, Wassmer E, Diederichs S. RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. Nucleic Acids Res. 2021;49:D425–36.
37. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics. 2018;34:511–3.
38. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44:D862–8.
39. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primer. 2021;1:1–21.
40. Li Y, Zhou D, Liu Q, Zhu W, Ye Z, He C. Gene POLYMORPHISMS of m6A erasers FTO and ALKBH1 associated with susceptibility to gastric cancer. Pharmacogenom Pers Med. 2022;15:547–59.
41. Martin JS, Halvorsen M, Davis-Neulander L, Ritz J, Gopinath C, Beauregard A, et al. Structural effects of linkage disequilibrium on the transcriptome. RNA. 2012;18:77–87.

Publisher's Note