Comprehensive Characterization of RNA Editing in Primary Gastric Adenocarcinoma through RNA-seq Data Analysis

Javad Behroozi
Tarbiat Modares University Faculty of Medical Sciences https://orcid.org/0000-0001-6429-0295
Shirin Shahbazi (sh.shahbazi@modares.ac.ir)
https://orcid.org/0000-0002-7634-5350
Mohammad Reza Bakhtiarizadeh
University of Tehran
Habibollah Mahmoodzadeh
Tehran University of Medical Sciences

Research article

Keywords: Primary Gastric Adenocarcinoma, RNA-seq Data Analysis, RNA Editing

Posted Date: April 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23058/v1

Abstract

RNA editing is a post-transcriptional nucleotide modification in humans. Of the various types of RNA editing, the adenosine to inosine substitution is the most widespread in higher eukaryotes, which is mediated by ADAR family enzyme. Inosine is recognized by the biological machineries as guanosine, therefore, editing can potentially rendering substantial functional effects throughout the genome, depending on where it located. RNA editing could contribute to cancer by either exclusive editing of tumor suppressor/promoting genes or by introducing transcriptomic diversity to promote cancer progression. Here, we provided a comprehensive overview of the RNA editing sites in gastric adenocarcinoma and highlighted some of their possible contributions to gastric cancer. RNA-seq data corresponding to 8 gastric adenocarcinoma and their paired non-tumor counterparts were retrieved from GEO database. After pre-possessing and variant calling steps, a stringent filtering pipeline was employed to distinguish potential RNA editing sites from SNPs. The identified potential editing sites were annotated and compared with those in DARNED database. Totally, 12362 high-confidence adenosine to inosine RNA editing sites were detected across all samples. Of these, 12105 and 257 were known and novel editing events, respectively. These editing sites were unevenly distributed across genomic regions, nearly half of them were located in 3´UTR. Indeed, 4868, 3985 and 3509 editing sites were found to be common in both tissue, normal specific and cancer specific, respectively. Further analysis revealed significant number of differentially edited events among these sites, which were located in protein coding genes and microRNAs. Given the distinct pattern of RNA editing in gastric adenocarcinoma and adjacent normal tissue, edited sites have the potential to serve as biomarkers and therapeutic targets in gastric cancer diagnose, management and treatment.

Introduction

RNA editing is a common and essential post-transcriptional alteration of RNA sequences, affecting millions of bases, expanding the transcriptome diversity and the functions of RNA transcripts [1]. Although several types of RNA editing have been characterized, conversion of adenosine residues to inosine (A to I) is the most frequent type of editing in humans, which is catalyzed by the double stranded RNA (dsRNA) specific adenosine deaminase that act on RNA (ADAR) family [2, 3]. Resulting inosine is recognized by most of the biological machineries as a guanosine (G), consequently, editing could have protein recoding outcome, generating proteomic and phenotypic diversity [4].

RNA editing diversifies the transcriptome when editing located in coding mRNA sequences, also, editing in the non-coding sequence could have a fundamental consequence. 3´UTRs usually comprise key elements and it has been found to be involved in numerous regulatory processes, editing in these elements can modulate the regulation of mRNA expression [5]. microRNAs identify their target genes primarily by sequence complementarity between the microRNA seed region and a target site, hence, editing in the seed sequence could affect target recognition [6]. Indeed, editing in the non-seed sequence may alter microRNA structure or stability, leading to biological consequences. It is also believed that editing of precursor microRNA may inhibits its processing to mature microRNA [7].

RNA editing are critical for growth and development in mice and humans. Hence, knockout mice for either of ADAR1 or ADARB1 genes die early in development [8, 9]. In addition, ADARs mutation are associated with several human diseases, mutations in the ADAR1 gene, mainly in its deaminase domain, are associated with the pathogenesis of Aicardi-Goutieres syndrome (AGS; OMIM #225750) [10] and dyschromatosis symmetrica hereditaria (DSH; OMIM #127400) [11]. Transcriptome of nearly all normal cell types are actively edited, particularly, in the immune system and the central nervous system, which exhibit fundamental flexibility of function. On the other hand, editing seems to be decreased in static cells, such as muscle cells, where there is no need for novel adaptations [12]. Moreover, it has been reported that RNA editing events are a positive contribution to cancer development and progression [5]. RNA editing dysregulation has been linked to cancer by either editing in coding [13, 14] or by editing in noncoding [15, 16] sequences. Also, there is a strong consensus on the effect of global editing levels in cancer, increased genome-wide editing rates has been reported in some cancers including; breast tumors, head/neck squamous cell, thyroid, lung adenocarcinoma and kidney renal cell carcinomas. Conversely, decreased whole editing rates were seen in kidney chromophobe and renal papillary carcinoma [17].

ADAR proteins bind a specific dsRNA structure formed either intramolecularly or intermolecularly, thus, ADAR edits A to I only on RNAs that adopt this the double strand structure [18]. There are also some modifying features including; RNA sequence preference associated with neighbor editing sites [19], editing inducer elements distant from editing position [20] and base opposing the edited adenosine [21]. Despite the identification of these regulation elements, the main controlling feature of ADAR target recognition and how the ADAR nominates an adenosine for edition, remains to be further studied. Since, these elements do not allow the prediction of editing sites, identification of editing events is therefore dependent on sequencing data [22].

The advent of next-generation sequencing (NGS) has greatly improved the genome-wide identification of RNA editing sites through RNA sequencing (RNA-Seq) technologies and so far several million high confidence editing sites have been recognized in the human genome [23]. Identification of editing sites from RNA-seq data seems to be straightforward. Simply, aligning RNA-seq reads to the reference genome and searching for A to G mismatches, leads to detection of editing sites [24]. However, there are several sources of disagreement between RNA sequence and the reference genome, making the identification of actual editing sites challenging. The major challenge in identifying RNA editing events using RNA-seq data is the discrimination of genuine editing sites from somatic mutations, SNPs and sequencing errors, therefore, robust bioinformatical approaches need to overcome this challenge [25]. However, dozens of outstanding studies have successfully employed RNA-seq data alone to identify editing events [25–34].

To the best of our knowledge, there has been no comprehensive study investigating the editome in gastric adenocarcinoma and many outstanding questions on the extent and consequences of RNA editing in gastric cancer remain concealed. In this study we leveraged publicly available sequencing datasets to characterize RNA editing in gastric cancer.

Materials And Methods

RNA-seq datasets

Raw paired-end RNA-seq samples related to eight primary gastric adenocarcinoma and their paired non-tumor counterparts were retrieved from publicly available GEO database (Gene Expression Omnibus database, accession number GSE85465). Non-tumor counterparts refers to samples harvested from the stomach, from sites distant from the tumor and exhibiting no visible evidence of tumor or intestinal metaplasia/dysplasia upon surgical assessment. The original data and sample details are described by Ooi et al. [35]. RNA-seq libraries of these samples were constructed using Illumina Stranded Total RNA Sample Prep Kit v2 and the dataset was generated using the Illumina HiSeq 2000 platform and the paired-end 101 bp read option.

Quality control and read mapping

First, FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was employed to control raw reads quality [27]. Furthermore, sequencer adapter removal and quality trimming was performed with Trimmomatic v0.32 (parameters: trailing 20 Maxinfo 60:0.95 and minimum length 60) [36]. Then, clean reads were aligned to the human reference genome (GRCh38) using Hisat2 v2.0.5, as it is more efficient at providing editing prediction from RNA-Seq data than other programs [37]. To reduce the potential bias caused by short read alignment, only uniquely and concordantly aligned reads were kept. PCR-induced duplicated reads that mapped to the same location were marked and excluded from analysis using the MarkDuplicates tool from the Picard package (http://picard.sourceforge.net/), except those with the highest mapping quality score [26]. To promote the aligning in the flanking of the indel regions and to improve the quality of reads, the remaining reads were locally realigned around putative indels and the base quality values were recalibrated by GATK tool v3.5 (https://www.broadinstitute.org/gatk).

Variant calling and identification of RNA editing sites

To perform variant calling, single nucleotide variants (SNVs) were first called using the HaplotypeCaller from the GATK tool with a stand_call_conf and stand_emit_conf value of 30 and mbq of 25 [38]. Next, the SNVs were removed from further analysis if they were corresponded to known SNPs found in Ensembl human SNP database version 151. Then, the remaining variants were filtered using the GATK standard filters including; 1) total depth of coverage < 10, to remove variants with less than 10 reads that passed the caller's internal quality control metrics. 2) HomopolymerRun > 5, to eliminate the variants with a homopolymer run larger than 5 bp on either side. 3) RMSMappingQuality < 40, to exclude variants with root mean square mapping quality less than 40 over all the reads at the site. 4) MappingQualityRankSum <-12.5, this parameter compares the mapping qualities of the reads supporting the reference allele and the alternate allele and employed to avoid mapping quality bias. A negative value indicates the mapping qualities of the reference allele are higher than those supporting the alternate allele. 5) QualitybyDepth < 2, this annotation is intended to normalize the variant confidence in order to avoid inflation caused when there is deep coverage. 6) ReadPosRankSum <-8, this annotation compares whether the positions of the reference and alternate alleles are different within the reads and eliminates variant distance bias [39].

Additionally, several quality-aware filtering steps employed to increase the accuracy of identifying true RNA editing sites. First, the sites with more than one non-reference type and homozygous sites for the alternative allele were filtered. Second, we discarded the sites with fewer than three reads supporting the SNV and only those sites, which at least 10 reads cover that site were kept for further analysis. Further, the SNV sites with an extreme or a rare degree of variation (threshold for the editing ratio was between 10% and 90%) were removed under the assumption that 100% editing efficiency is unrealistic. Third, SNVs located in regions with bidirectional transcription (transcription that occurs on both the positive and negative strands) were filtered. Fourth, GMATo software used for detection of simple sequence repeats (SSR) patterns and SNVs located in SSR regions were considered as biased with an offset of ± 3 bases [40]. Fifth, SNVs occurred within 5 bp intronic flanking region were removed. Finally, to reduce false-positive SNVs because of misalignment of sequencing reads to other parts of the genome, we filtered out SNVs in paralogs or repetitive regions by retrieving and aligning 100 bp of flanking sequence (50 upstream and 50 downstream of the SNV) using BLAT [41]. Only the SNVs that were located in uniquely mapped sequences considered as RNA editing site. A to G and editing sites were kept for further analysis and other non-canonical editing sites were excluded. Ultimately, we compare identified RNA editing sites with those in DARNED [42] database and categorized them as “known RNA editing site”, if they were in the database, and as “novel editing site” if they were not. An overview of our computational analysis pipeline for identifying the RNA editing sites is shown in Fig. 1.

Neighborhood profile of editing

In order to predict the conservation of the editing sites neighborhood nucleotides, 10 bp upstream and 10 bp downstream of the edited sites were extracted. Then, WebLogo software was employed to generate a consensus sequence logo and investigate the sequence context flanking the identified potential editing sites [43].

Annotation of RNA editing sites

The functional annotation and genomic location of the RNA editing sites were performed using SnpEff v4.3 [44]. The gene set used for annotation was Ensembl version GRCh38.92. In order to identify the biological functions associated with edited genes in cancer and normal tissue, we used Enrichr web-application to conduct a functional enrichment analysis based on Gene Ontology (GO) biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms (Adjusted P-value ≤ 0.05) [45].

Validation of detected editing sites

To validate detected RNA editing sites, we used the publicly available human expressed sequence tags (ESTs) (ftp://ftp.ncbi.nih.gov/repository/UniGene/) to investigate whether the editing events identified by our pipeline were also present in these sequences. First, 50 bp upstream and downstream flanking regions of editing sites were extracted and queried against the human EST sequences using BLAST. Then, alignments with e-values < 10 − 5 were considered as significant and counted. On the other hand,

since most of A to I RNA editing occurs in Alu repeats [46], we evaluate intersection of Alu repeats with identified editing sites. To do this, genomic positions of Alu repeats were downloaded from UCSC database (http://genome.ucsc.edu/) and their distribution pattern across the genome were compared with pattern of identified A to I editing sites.

Impaired microRNAs targeting

In order to predict microRNAs whose binding is affected by RNA editing, we downloaded the predicted microRNA binding data of highly conserved miRNA families from miRcode database [47]. Then, we applied intersect feature of BedTools to find RNA editing sites that overlap with target site of microRNAs [48].

Statistical Analyses

Statistical significance for differences between cancer and normal tissue editing ratio were assessed by paired Student's t-test. Spearman's correlation coefficient was used to determine the relationship between chromosome length, number of Alu elements, number of protein coding genes and number of editing sites. Differences were considered significant when the P-value or adjusted P-value was < 0.05

Results And Discussion

Identification of RNA editing sites

High-throughput RNA-seq technology have facilitated the discovery of transcriptome-wide RNA editing events across individuals and tissues at unprecedented throughput and resolution. However, the main obstacle in identifying bona fid RNA editing sites using RNA-seq data is the distinction of RNA editing sites from rare SNPs and technical artifacts caused by sequencing or read-mapping error. To accurately detect the RNA editing sites at the transcriptome-wide level in gastric cancer, we developed a computational approach by using a precise strategy (see Fig. 1). This strategy enabled us to identify the potential RNA editing sites using RNA sequencing data alone, without the need for available matched DNA sequence from the same sample. We obtained 1725 million reads from RNA-seq data of eight gastric adenocarcinoma and their paired normal tissues. After quality trimming, a total of 1492.1 million reads were generated from all samples (on average, 93.3 million reads per sample). The clean reads were aligned to the reference genome with an average mapping rate of 91.67%. Also, the average rate of uniquely and concordantly mapped reads was 74.34% (range 59–84%). Initial analysis led to the identification of 1370502 variants and after excluding SNPs and INDELs 141347 SNVs remained. Finally, after applying multiple stringent filters to exclude false-positives, a total of 12362 unique A to G RNA editing sites were identified across all samples, which 12105 sites were previously reported in DARNED database and 257 variants were novel editing sites. These editing sites were distributed in 2406 unique gene. Based on our filtering criteria, all of these editing sites were located in unique genomic positions and were not close to any splice junction, bidirectional transcription or low complexity regions (such as

SSRs). A summary of the statistics of raw and clean reads and mapping information as well as the number of identified SNVs and editing sites in different samples is provided in S1 File.

Sequence preferences analysis

ADAR enzyme targets dsRNA of any sequence, but it has a sequence preference in the vicinity of the editing sites. Consistent with the known attributes of ADAR substrates, our results showed that the nucleotide immediately upstream (relative − 1 location) of edited site had a strong preference for G depletion and T enrichment. While, nucleotide immediately downstream (relative + 1 location) of the editing site showed significantly depleted T and favored G (see Fig. 2).

Validation of identified editing

The location of identified editing sites were compared with the position of Alu elements across genome. Interestingly, distribution of A to G editing sites and Alu elements were very similar across the genome. This is more obvious when we look closely at chromosomes 1, 9, 16 and 19, where, ends of these chromosomes are reach in Alu repeats but the middle of chromosomes are relatively vacant (see Fig. 3). Next, to validate whether the identified RNA editing sites were true positive, we searched for evidences of the identified RNA editing sites in expressed sequence tags (ESTs) based on NCBI database. Of 12105 known and 257 novel editing sites, 10944 (90.4%) and 218 (84.8%) sites were found in EST clones, respectively. Moreover, further investigation revealed that 7643 (68.5%) of the identified editing events were validated in more than five EST clones, which reinforce the accuracy of our method.

Distribution of the editing sites across genomic regions

First, the location of editing sites was annotated according to Ensembl database. As shown in Table 1 the most biotype of edited transcripts were “Protein coding” and the least were “snoRNA”. Also, 42 editing sites were located in miRNAs, which were belonged to 17 unique microRNAs. Of these, MIR34A included 10 cancer-specific editing sites (Table 2). Investigation of genomic distribution of editing sites showed that the number of RNA editing sites greatly are varied across genomic regions. Overall, the 3´UTR was the most edited region, with 5870 editing sites (45.5% of all detected editing sites), followed by the upstream (23.5%), and the 5´UTR had the least number of editing sites (less than 1%). Indeed, 192 (1.6%) of the editing sites were located in exons, including 81 sites (42% of exonic editing sites) with non-synonymous effect and 43 sites (22%) with synonymous effect. Exonic RNA editing leads to at least one premature termination codon and two stop loss mutations. Also, one editing site with start-gain mutation effect was detected (see Fig. 4).

Table 1
Number of different edited biotypes

Biotype	No.
Protein coding	9000
snRNA	32
Processed transcript	893
Retained intron	996
lincRNA	347
antisense	453
miRNA	42
Sense intronic	116
Sense overlapping	24
Pseudogene	212
TEC	63
snoRNA	12
MiscRNA	21
Intergenic	122
NA	29

Table 2
List of edited microRNAs in gastric cancer and normal tissue

Symbol	Chr.	Specification	No. of editing site(s)	Role in cancer
miR-1205	Chr8	Cancer	2	[49, 50]
miR-143	Chr5	Normal	6	[51–53]
miR-24-1	Chr9	Normal	1	[54, 55]
miR-3176	Chr16	Cancer	3	[56, 57]
miR-34A	Chr1	Cancer	10	[58–60]
miR-4315-2	Chr17	Cancer	1	[61]
miR-4522	Chr17	Cancer	1	[62]
miR-4539	Chr14	Common	1	[63]
miR-4728	Chr17	Cancer	2	[64, 65]
miR-559	Chr2	Normal	1	[66, 67]
miR-5692C2	Chr7	Normal	2	[68]
miR-612	Chr11	Cancer	1	[69, 70]
miR-621	Chr13	Normal	3	[71, 72]
miR-635	Chr17	Common	3	[73, 74]
miR-642B	Chr19	Normal	3	[75, 76]
miR-650	Chr22	Common	1	[77, 78]
miR-8071-1	Chr14	Common	1	[79]

Gene editing rate and RNA editing level

RNA editing sites often appear in clusters, due to simultaneous editing of multiple adenosines by ADAR proteins. Therefore, we investigated whether the identified editing sites were in clusters or not. We found that 34% of genes were edited in more than five sites. Furthermore, gene editing rate (number of edited sites located in gene) were calculated to evaluate clustering of editing sites. Overall, each gene in our study showed editing rate equal 5.1, which means on average each gene had five editing sites. Interestingly, editing rates were different when genomic regions were considered separately. Editing rate in 3´UTRs was predominant, 7.1 editing site per gene, and exons showed the least editing rate, 1.4 editing site per gene. Editing rate in upstream and downstream regions, which included a large number of editing sites, was 3.5 and 3.8 site per gene, respectively. The frequency distribution of gene editing rates across genomic regions is shown in Fig. 5A. RNA editing level was also calculated for all edited sites, using the following formula [80]:

RNA editing level= (number of reads supporting edited allele × 100) / (total number of reads at a site)

Average RNA editing level across all sites was 30.72, which means, approximately 31% of each gene transcripts were edited in a given site. Editing level for most of the identified editing sites in the present study ranged from 15 to 25. The frequency distribution of RNA editing levels is shown in Fig. 5B.

Association between chromosome length, Alu elements, protein coding genes and number of editing sites

Pearson's correlation coefficient was used to investigate the association between the number of editing sites and length of chromosomes. As expected, the number of RNA editing sites tended to be associated with chromosome length, but the association was weak when all chromosomes were included ($r = 0.47$, $P = 0.02$). As show in Fig. 6A, chromosome 19 has the highest editing frequency according to its size. Excluding the chromosome 19 from the analysis showed a significant correlation between number of RNA editing sites and length of chromosomes ($r = 0.6$ $P < 0.002$). In addition, correlation of editing with both number of Alu elements and number of protein coding genes were calculated. Surprisingly, we found that correlation of editing with number of protein coding genes was stronger than number of Alu elements, where Spearman's correlation coefficient was 0.91 and 0.85, respectively (Fig. 6B and 6C). To further investigate we calculate editing rate for each chromosome as number of editing sites in one kilobase (kb). Our results showed that chromosome 19 has the most rate of editing with one editing sites per 40 kb, followed by chromosome 17 with one editing sites per 112 kb. On average one editing site was identified in 250 kb of human genome and gene-poor chromosomes (18, 4, 21, 13 and Y) have the least rate of editing (S2 File).

Cancer and normal specific editing sites

Among the 12362 editing sites, 4868 sites were found within both normal and cancer samples. On the other hand, 3985 and 3509 editing sites were specific to normal and cancer tissues, respectively. Statistical analysis revealed 285 differentially edited events among common editing sites. Notably, 129 cancer-specific and 173 normal-specific editing sites were found to be differentially edited (see Fig. 7). Functional enrichment analysis of the cancer and normal-specific edited genes showed a larger number of significant terms in cancer-specific edited genes. Nine GO term were significantly enriched in cancer-specific edited genes, on the other hand only one term was significantly enriched in normal-specific edited genes. GO and KEGG pathways categories of the top five cancer-and normal-specific edited genes are shown in Table 3. These significantly enriched terms could help us a lot to further understand the role of edited genes in gastric cancer.

Table 3
GO and KEGG pathway enrichment analysis of edited gene between cancer and normal tissue

Term (Gene Ontology)	overlap	Term (KEGG)	overlap
Cancer-specific edited genes			
chromatin remodeling at centromere (GO:0031055)	12/32*	Herpes simplex virus 1 infection	69/492*
centromere complex assembly (GO:0034508)	12/36*	Homologous recombination	9/41
DNA replication-independent nucleosome assembly (GO:0006336)	12/39*	Endocytosis	30/244
CENP-A containing nucleosome assembly (GO:0034080)	10/30*	Non-homologous end-joining	4/13
CENP-A containing chromatin organization (GO:0061641)	10/30*	Hepatitis C	19/155
Normal-specific edited genes			
rRNA processing (GO:0006364)	31/202	Herpes simplex virus 1 infection	65/492*
rRNA metabolic process (GO:0016072)	30/200	Ribosome	21/153
snRNA transcription from RNA polymerase II promoter (GO:0042795)	15/70	RNA polymerase	6/31
protein targeting to ER (GO:0045047)	19/97	Measles	17/138
CENP-A containing chromatin organization (GO:0061641)	9/30	Folate biosynthesis	5/26
*indicates adjusted P-value is significant.			

Functional impacts of RNA editing sites

The functional impact of RNA editing could induce by vast range of molecular mechanisms. For instance, it can lead to amino acid recoding, causing changes in seed sequences of microRNAs or affect microRNA targeting sites. In search of amino acid recoding mutations, 81 editing sites were found across 63 genes that could lead to non-synonymous change (S3 File), including 12 novel editing sites. Interestingly, MUC4, an epithelial glycoprotein coding gene, was edited in two positions (3:195780295 and 3:195780902), which caused p.L3762P and p.S2560P, respectively (Table 4). Also, microRNAs targeting could affect by editing. In this regard, 44 editing sites were detected that affect microRNA target recognition in normal and cancerous tissue of gastric (Table 5). In addition, 294 editing sites with nonsense-mediated decay impact were found that affect 92 protein coding genes. Of these, 80 and 111 sites were identified only in cancer and normal samples, respectively. Also, 103 nonsense-mediated decay editing sites were found in both cancer and normal tissues (S4 File).

Table 4
List of novel editing sites with non-synonymous change

Position	Gene ID	Editing effect
1: 246885532	AHCTF1	p.N883S
7: 142529491	TRBV7-9	p.N26D
17: 2333110	TSR1	p.S386G
11: 130914721	SNX19	p.S407G
17: 31856838	COPRS	p.S43G
11: 1018295	MUC6	p.I1502M
X: 315276	GTPBP6	p.I171V
7: 100958135	MUC3A	p.M2119T
3: 58156064	FLNB	p.M2324V
3: 195780295	MUC4	p.L3762P
3: 195783902	MUC4	p.S2560P
22: 22376266	IGLV5-45	p.C44R

Table 5
List of editing sites that affect microRNA target recognition

Chr.	Position	Gene	specify	Affected microRNA(s)
chr1	9100841	GPR157	Common	miR-490-3p
chr1	10459831	DFFA	Common	miR-150/5127
chr1	10460010	DFFA	Common	miR-208ab/208ab-3p
chr1	10460010	DFFA	Common	miR-499-5p
chr1	179073347	FAM20B	Common	miR-125a-5p/125b-5p/351/670/4319
chr1	179073347	FAM20B	Common	let-7/98/4458/4500
chr1	179075081	FAM20B	Normal	miR-22/22-3p
chr1	179075107	FAM20B	Normal	miR-146ac/146b-5p
chr1	179075144	FAM20B	Normal	miR-143/1721/4770
chr4	2839669	SH3BP2	Cancer	miR-199ab-5p
chr4	2840078	SH3BP2	Common	miR-217
chr4	2840078	SH3BP2	Common	miR-200bc/429/548a
chr4	2938644	NOP14	Common	miR-24/24ab/24-3p
chr4	17626928	MED28	Normal	miR-455-5p
chr4	17632277	FAM184B	Cancer	miR-144
chr5	34906645	RAD1	Common	miR-143/1721/4770
chr5	37290314	NUP155	Common	miR-24/24ab/24-3p
chr5	43377383	CCL28	Normal	miR-383
chr5	43380635	CCL28	Common	miR-24/24ab/24-3p
chr5	75378054	COL4A3BP	Cancer	miR-103a/107/107ab
chr6	53100576	FBXO9	Normal	miR-103a/107/107ab
chr7	44802500	PPIA	Common	miR-22/22-3p
chr7	100212870	CASTOR3	Common	miR-128/128ab
chr7	100212870	CASTOR3	Common	miR-27abc/27a-3p
chr8	41542121	GINS4	Common	miR-26ab/1297/4465
chr8	43029279	HOOK3	Common	miR-26ab/1297/4465
chr8	43029280	HOOK3	Cancer	miR-26ab/1297/4465

Chr.	Position	Gene	specify	Affected microRNA(s)
chr9	128305442	TRUB2	Common	miR-15abc/16/16abc/195/322/424/497/1907
chr9	128305442	TRUB2	Common	miR-103a/107/107ab
chr11	768850	GATD1	Common	miR-141/200a
chr11	769393	GATD1	Cancer	miR-24/24ab/24-3p
chr11	111728428	SIK2	Normal	miR-142-3p
chr14	21460342	RAB2B	Common	miR-196abc
chr16	66887684	PDP2	Common	miR-7/7ab
chr19	1032835	CNN2	Common	miR-25/32/92abc/363/363-3p/367
chr19	1777958	ONECUT3	Common	miR-142-3p
chr19	1778139	ONECUT3	Common	miR-194
chr19	1778241	ONECUT3	Common	miR-218/218a
chr19	1778303	ONECUT3	Common	miR-103a/107/107ab
chr19	2835224	ZNF554	Normal	miR-17/20ab/20b-5p/93/106ab/427/518a-3p/519d
chr19	4653661	TNFAIP8L1	Common	miR-150/5127
chr19	16631220	SMIM7	Cancer	miR-192/215
chr19	18368012	PGPEP1	Common	miR-455-5p
chr20	3869487	MAVS	Normal	miR-338/338-3p

Discussion

The identification of RNA editing sites deeply depends on sequencing technology and bioinformatics approaches. We developed a pipeline for identifying RNA editing events in primary gastric cancer and normal tissues by screening RNA differences from reference genome followed by successive and rigorous filtering criteria. Most of previous studies have used coupled RNA and DNA sequences to identify editing events [28, 81], by the contrary, we identified RNA editing sites using RNA sequencing data alone. Our analyses found significant number of editing sites, vast majority of them harbored in 3´UTR regions, which has been reported in previous studies [80, 82]. Also a few novel editing sites were found, which were reported for the first time in the current study. Although the number of identified RNA editing sites was huge, most of the sites exhibited low editing levels and approximately half of the identified sites were edited in less than 27% of their related transcripts.

Our analyses found that the RNA editing sites were highly associated with both number of protein coding genes and Alu elements distribution in the genome. Also, frequency of editing sites were correlated with size of chromosomes. These results are in a good agreement with Chigaev et al. study, who reported that correlation of editing frequency with protein coding genes is stronger than lincRNA density [80]. However, these correlation could result from the bias of the library preparation step of RNA sequencing projects. Since oligo-dT primers apply to capture the RNA through the poly-A tail, most of the reads will be related to protein coding genes.

To date, no specific sequence has been found that characterize editing sites of any of the ADAR enzymes. However, in the neighborhood of edited adenosine, there are preferred and opposed preferences. Consistent with previous studies, there was an over-representation of guanosine in the neighboring position downstream, while guanosine was depleted in the upstream neighboring position [26, 82]. Since some of adenine bases in the right context do not edit, other features proposed to be involved in determination of editing. Daniel et al. described editing inducer elements distance from the edited adenine, which increase the editing efficiency and specificity of a highly edited site [20]. Wong et al. reported that editing efficiency is strongly influenced by the base opposing the edited adenosine. They found that when there is an A:C mismatch at the editing site, editing by ADAR enzyme was enhanced compared to when A:A or A:G mismatches or A:U base pairs occurred at the same site [21]. Due to the contradictory results, it is difficult to make definitive conclusions about potential editing sites.

We wonder whether RNA editing could function as an additional mechanism contributing to tumorigenesis by generating specific RNA editing sites that are unique to cancer samples. In the search of the answer to this question we found that 28.4% and 32.2% of the identified editing sites were specific to cancer and normal tissues, respectively. These tissue specific editing sites could contribute to cancer initiation and progression, if they located in important gene. Some of cancer-specific editing sites and their role in pathogenesis of cancer have been identified in previous studies. RNA editing of transcription factor PROX1, a candidate tumor suppressor, leads to several missense substitutions including E328G, R334G, and H536R and loses tumor suppressive functions. These editing events have been seen in a number of esophageal, pancreatic, and colon cancer samples, but no such editing is seen in a number of cDNA libraries of many normal tissues [17].

We also found a remarkable number of common editing events between cancer and normal tissues, which their editing levels were significantly different in cancer and normal tissue. Deregulated editing level in cancer and normal common editing sites could be an important contributor in tumorigenesis. Chen et al. reported that RNA editing level of AZIN1 increases by at least 10% in hepatocellular carcinoma compared to adjacent normal liver. The edited isoform compared with wild-type AZIN1 has increased affinity to antizyme, which leads to neutralization of antizyme-mediated degradation of ornithine decarboxylase and cyclin D1 and promotes cell proliferation [83]. In this regard, Han et al. reported a higher level of editing on RHOQ in tumor compared with normal tissue in colorectal cancer, which results in N136S amino acid substitution. This RNA mutation increases RHOQ protein activity, actin cytoskeletal reorganization and invasion potential [84]. On the contrary, hypo-editing of several genes are associated

with cancer phenotypes. The pre-mRNA transcript encoding the GluR-B has two functionally important editing sites (Q/R and R/G sites) and the Q/R site almost entirely edited, which is necessity for normal function of receptor. It has been proved, in malignant tissue of human brain tumors, this editing site of GluR-B considerably under-edited compared with control tissues [85]. Our results corroborate that the RNA editing frequency can be regulated in a tissue specific manner, which is consistent with observations reported previously.

Our results showed that the vast majority of editing sites in gastric cancer were located in 3´UTR and up/down stream regions as well as a large number of editing sites were observed in coding regions. According to their genomic location, these RNA editing events could lead to various functional impacts and apply their effects through several dominant mechanisms. First and most important, RNA editing events in exonic region can cause amino acid change and imitate cancer-associated missense mutations. Our pipeline identified 81 editing events with non-synonymous effect, including 12 novel editing events. Notably, we found four missense RNA mutations in mucin family (MUC3A, MUC4 and MUC6). Normal gastric epithelial cells transcribe MUCs, which have several functions including; protection against mechanical and infectious lesions, lubrication and acid resistance [86]. Several studies have been reported that transcription profile of mucins are changed in gastrointestinal cancers, which overall suggests an important role for MUCs in gastric cancer [87–89]. Our results reinforced the hypothesis that inappropriate RNA editing can be involved in gastric cancer development.

Second, RNA editing could affect microRNAs target recognition and subsequently affect the expression profile of the genes. Previous computational analyses suggested that RNA editing tends to avoid microRNA target sites in general, even though RNA editing events have a potential to block the microRNA target recognition. Dysregulation of microRNA target recognition has been linked to cancers [90, 91]. In this context, 44 editing events were found in the present study, where at least one microRNA binding was disrupted. In consistent with our research, Soundararajan et al. identified 652 editing events in lung cancer, which were located in the 3´UTR of 205 target genes and mapped to 932 potential microRNA target binding sites [92]. All together these findings are inconsistent with Liang and Landweber previous computational analyses, where they suggested that RNA editing tends to avoid microRNA target sites in general, even though RNA editing events have a potential to block the microRNA target recognition [93]. It is worth to remind, RNA editing events in addition to disrupting existing microRNA binding sites, could generate novel microRNA regulatory networks. In a completely separate mechanism from what has been mentioned, RNA editing could affect microRNA biosynthesis. miR-142 is highly expressed in hematopoietic tissues, conversely it is not expressed in non-hematopoietic tissues. Also, its expression in patients with acute myeloid leukemia is significantly lower than that in controls. Yang et al. showed that editing of pri-miR-142, leads to suppression of its processing by Drosha and subsequently it degradation [94].

Third, editing of microRNA sequences could alter their binding affinity or target recognition properties. Since microRNAs play a role in nearly all cellular pathways and pathological processes, including cancer initiation and progression, fluctuations of their targeting are an important contributor to cancer [95]. Our

analysis revealed 42 editing sites in 17 cancer-associated microRNAs, some of them exclusively edited in cancerous tissue. Consistent with our results, Nigita et al. identified 40 and 18 potential editing sites in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma, respectively [96]. Indeed, our results showed miR-34a, a cancer-specific edited microRNA, was edited in 10 position. Previous studies have been identified this microRNA as a tumor suppressor in gastric cancer cell lines [58]. On the other hand, it was shown miR-34a epigenetically down-regulated or silenced in gastric cancer tissues and cell lines [97]. We therefore speculate that editing in some positions could terminate the function of miR-34a, but further studies are required to confirm this possibility.

To our knowledge, this is the first time to comprehensively characterize editome of normal and cancerous tissue of gastric. Findings of the current study uncovered relatively large number of RNA editing sites, which were unevenly distributed across genome. Editing level of these sites and editing rate of different genes had diverse distribution. We also found a significant number of exclusively edited genes in cancer and normal tissue, which are likely to contribute to cancer initiation and progression.

Conclusions

Gastric cancer initiation and progression is driven by the cumulative effects of genetic and epigenetic alterations, RNA editing a widespread post-transcriptional mechanism could be part of these alterations. Depending on genomic location and level of editing, this phenomenon could leads to missense mutations, affecting microRNA biosynthesis and targeting, changing splicing patterns and modifying microRNA target sites. Editome of gastric cancer vastly differ from adjacent tissue in terms of both type and number of editing sites. Given the distinct pattern of RNA editing between gastric cancer and normal tissue, edited sites have the potential to serve as biomarkers and therapeutic targets in gastric cancer diagnosis, management and treatment.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in the GEO repository [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85465]. All data generated during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Tarbiat Modares University and Iran National Science Foundation (grant number: 97014362).

Authors' contributions

The study concept and design involved SS, MRB, HM and JB. SS and JB were responsible for the recruitment and data collection. Data analysis was completed by MRB and JB. JB drafted the original manuscript. The article was revised by SS, MRB and HM. All authors read and approved the final version.

Acknowledgements

Not applicable.

References

1. Roth SH, Levanon EY, Eisenberg E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. Nature methods. 2019:1-8.
2. Huang W-H, Tseng C-N, Tang J-Y, Yang C-H, Liang S-S, Chang H-W. RNA editing and drug discovery for cancer therapy. The Scientific World Journal. 2013;2013.
3. Jayan GC. RNA editing in hepatitis delta virus: unsolved puzzles. The Scientific World Journal. 2004;4:628-37.
4. Danan-Gotthold M, Levanon EY. Promoting RNA editing by ADAR attraction. Genome biology. 2017;18(1):196.
5. Huang WH, Tseng CN, Tang JY, Yang CH, Liang SS, Chang HW. RNA editing and drug discovery for cancer therapy. The Scientific World Journal. 2013;2013.
6. Mingardi J, Musazzi L, De Petro G, Barbon A. miRNA editing: New insights into the fast control of gene expression in health and disease. Molecular neurobiology. 2018;55(10):7717-27.
7. Nishikura K, Sakurai M, Ariyoshi K, Ota H. Antagonistic and stimulative roles of ADAR1 in RNA silencing. RNA biology. 2013;10(8):1240-7.
8. Higuchi M, Maas S, Single FN, Hartner J, Rozov A, Burnashev N, et al. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature. 2000;406(6791):78-81.
9. Wang Q, Khillan J, Gadue P, Nishikura K. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. Science. 2000;290(5497):1765-8.
10. Rice GI, Kasher PR, Forte GM, Mannion NM, Greenwood SM, Szynkiewicz M, et al. Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nature genetics. 2012;44(11):1243.

11. Ganem NS, Ben-Asher N, Lamm AT. In cancer, A-to-I RNA editing can be the driver, the passenger, or the mechanic. Drug Resistance Updates. 2017;32:16-22.
12. Gagnidze K, Rayon-Estrada V, Harroch S, Bulloch K, Papavasiliou FN. A new chapter in genetic medicine: RNA editing and its role in disease pathogenesis. Trends in molecular medicine. 2018;24(3):294-303.
13. Anadón C, Guil S, Simó-Riudalbas L, Moutinho C, Setien F, Martínez-Cardús A, et al. Gene amplification-associated overexpression of the RNA editing enzyme ADAR1 enhances human lung tumorigenesis. Oncogene. 2016;35(33):4407.
14. Qin Y-R, Qiao J-J, Chan THM, Zhu Y-H, Li F-F, Liu H, et al. Adenosine-to-inosine RNA editing mediated by ADARs in esophageal squamous cell carcinoma. Cancer research. 2014;74(3):840-51.
15. Zipeto MA, Court AC, Sadarangani A, Santos NPD, Balaian L, Chun H-J, et al. ADAR1 activation drives leukemia stem cell self-renewal by impairing Let-7 biogenesis. Cell Stem Cell. 2016;19(2):177-91.
16. Beghini A, Ripamonti CB, Peterlongo P, Roversi G, Cairoli R, Morra E, et al. RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. Human molecular genetics. 2000;9(15):2297-304.
17. Baysal BE, Sharma S, Hashemikhabir S, Janga SC. RNA editing in pathogenesis of cancer. Cancer research. 2017;77(14):3733-9.
18. Song C, Sakurai M, Shiromoto Y, Nishikura K. Functions of the RNA editing enzyme ADAR1 and their relevance to human diseases. Genes. 2016;7(12):129.
19. Kuttan A, Bass BL. Mechanistic insights into editing-site specificity of ADARs. Proceedings of the National Academy of Sciences. 2012;109(48):E3295-E304.
20. Daniel C, Venø MT, Ekdahl Y, Kjems J, Öhman M. A distant cis acting intronic element induces site-selective RNA editing. Nucleic acids research. 2012;40(19):9876-86.
21. Wong SK, Sato S, Lazinski DW. Substrate recognition by ADAR1 and ADAR2. Rna. 2001;7(6):846-58.
22. Eisenberg E, Levanon EY. A-to-I RNA editing—immune protector and transcriptome diversifier. Nature Reviews Genetics. 2018;19(8):473-90.
23. Diroma MA, Ciaccia L, Pesole G, Picardi E. Elucidating the editome: bioinformatics approaches for RNA editing detection. Briefings in bioinformatics. 2019;20(2):436-47.
24. Porath HT, Carmi S, Levanon EY. A genome-wide map of hyper-edited RNA reveals numerous new sites. Nature communications. 2014;5(1):1-10.
25. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'connell MA, et al. Identifying RNA editing sites using RNA sequencing data alone. Nature methods. 2013;10(2):128.
26. Bakhtiarizadeh MR, Salehi A, Rivera RM. Genome-wide identification and analysis of A-to-I RNA editing events in bovine by transcriptome sequencing. PloS one. 2018;13(2):e0193316.
27. Shafiei H, Bakhtiarizadeh M, Salehi A. Large-scale potential RNA editing profiling in different adult chicken tissues. Animal genetics. 2019;50(5):460-74.

28. Peng Z, Cheng Y, Tan BC-M, Kang L, Tian Z, Zhu Y, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nature biotechnology. 2012;30(3):253.
29. Hwang T, Park C-K, Leung AK, Gao Y, Hyde TM, Kleinman JE, et al. Dynamic regulation of RNA editing in human brain development and disease. Nature neuroscience. 2016;19(8):1093.
30. Han L, Diao L, Yu S, Xu X, Li J, Zhang R, et al. The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. Cancer cell. 2015;28(4):515-28.
31. Moran B, Butler ST, Creevey CJ. Comparison and characterisation of mutation calling from whole exome and RNA sequencing data for liver and muscle tissue in lactating holstein cows divergent for fertility. BioRxiv. 2017:101733.
32. Zhu S, Xiang J-F, Chen T, Chen L-L, Yang L. Prediction of constitutive A-to-I editing sites from human transcriptomes in the absence of genomic sequences. BMC genomics. 2013;14(1):206.
33. Blanc V, Park E, Schaefer S, Miller M, Lin Y, Kennedy S, et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. Genome biology. 2014;15(6):R79.
34. Gu T, Buaas FW, Simons AK, Ackert-Bicknell CL, Braun RE, Hibbs MA. Canonical A-to-I and C-to-U RNA editing is enriched at 3' UTRs and microRNA target sites in multiple mouse tissues. PloS one. 2012;7(3).
35. Ooi WF, Xing M, Xu C, Yao X, Ramlee MK, Lim MC, et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. Nature communications. 2016;7:12983.
36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
37. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nature methods. 2015;12(4):357.
38. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491.
39. De Summa S, Malerba G, Pinto R, Mori A, Mijatovic V, Tommasi S. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. BMC bioinformatics. 2017;18(5):119.
40. Wang X, Lu P, Luo Z. GMATo: a novel tool for the identification and analysis of microsatellites in large genomes. Bioinformation. 2013;9(10):541.
41. Kent WJ. BLAT—the BLAST-like alignment tool. Genome research. 2002;12(4):656-64.
42. Kiran A, Baranov PV. DARNED: a DAtabase of RNa EDiting in humans. Bioinformatics. 2010;26(14):1772-6.
43. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome research. 2004;14(6):1188-90.

44. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92.
45. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics. 2013;14(1):128.
46. Daniel C, Silberberg G, Behm M, Öhman M. Alu elements shape the primate transcriptome by cis-regulation of RNA editing. Genome biology. 2014;15(2):R28.
47. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics. 2012;28(15):2062-3.
48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
49. Anauate AC, Leal MF, Wisnieski F, Santos LC, Gigek CO, Chen ES, et al. Analysis of 8q24. 21 miRNA cluster expression and copy number variation in gastric cancer. Future medicinal chemistry. 2019(0).
50. Dai B, Kong D, Tian J, Liu T, Zhou H, Wang Z. microRNA-1205 promotes cell growth by targeting APC2 in lung adenocarcinoma. European review for medical and pharmacological sciences. 2019;23(3):1125-33.
51. Wu X-L, Cheng B, Li P-Y, Huang H-J, Zhao Q, Dan Z-L, et al. MicroRNA-143 suppresses gastric cancer cell growth and induces apoptosis by targeting COX-2. World journal of gastroenterology: WJG. 2013;19(43):7758.
52. Lei C, Du F, Sun L, Li T, Li T, Min Y, et al. miR-143 and miR-145 inhibit gastric cancer cell migration and metastasis by suppressing MYO6. Cell death & disease. 2017;8(10):e3101.
53. Obermannova R, Redova-Lojova M, Vychytilova-Faltejskova P, Grell P, Cho WC, Sachlova M, et al. Tumor expression of miR-10b, miR-21, miR-143 and miR-145 is related to clinicopathological features of gastric cancer in a central European population. Anticancer research. 2018;38(6):3719-24.
54. Duan Y, Hu L, Liu B, Yu B, Li J, Yan M, et al. Tumor suppressor miR-24 restrains gastric cancer progression by downregulating RegIV. Molecular cancer. 2014;13(1):127.
55. Xie L, Wang T, Yu S, Chen X, Wang L, Qian X, et al. Cell-free miR-24 and miR-30d, potential diagnostic biomarkers in malignant effusions. Clinical biochemistry. 2011;44(2-3):216-20.
56. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer research. 2011;71(1):78-86.
57. Li J, Yang X, Guan H, Mizokami A, Keller ET, Xu X, et al. Exosome-derived microRNAs contribute to prostate cancer chemoresistance. International journal of oncology. 2016;49(2):838-46.
58. Hui W-T, Ma X-B, Zan Y, Wang X-J, Dong L. Prognostic significance of MiR-34a expression in patients with gastric cancer after radical gastrectomy. Chinese medical journal. 2015;128(19):2632.

59. Deng X, Zheng H, Li D, Xue Y, Wang Q, Yan S, et al. MicroRNA-34a regulates proliferation and apoptosis of gastric cancer cells by targeting silent information regulator 1. Experimental and therapeutic medicine. 2018;15(4):3705-14.
60. Peng Y, Fan J-Y, Xiong J, Lou Y, Zhu Y. miR-34a Enhances the Susceptibility of Gastric Cancer to Platycodin D by Targeting Survivin. Pathobiology. 2019;86(5-6):296-305.
61. Golan T, Stossel C, Schvimer M, Atias D, Halperin S, Buzhor E, et al. Pancreatic cancer ascites xenograft–an expeditious model mirroring advanced therapeutic resistant disease. Oncotarget. 2017;8(25):40778.
62. Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WW, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood. 2010;116(23):e118-e27.
63. Pellatt DF, Stevens JR, Wolff RK, Mullany LE, Herrick JS, Samowitz W, et al. Expression profiles of miRNA subsets distinguish human colorectal carcinoma and normal colonic mucosa. Clinical and translational gastroenterology. 2016;7(3):e152.
64. Floros KV, Lochmann TL, Hu B, Monterrubio C, Hughes MT, Wells JD, et al. Coamplification of miR-4728 protects HER2-amplified breast cancers from targeted therapy. Proceedings of the National Academy of Sciences. 2018;115(11):E2594-E603.
65. Liu Z, Zhang J, Gao J, Li Y. MicroRNA-4728 mediated regulation of MAPK oncogenic signaling in papillary thyroid carcinoma. Saudi journal of biological sciences. 2018;25(5):986-90.
66. Bahreini F, Ramezani S, Shahangian S, Salehi Z, Mashayekhi F. miR-559 polymorphism rs58450758 is linked to breast cancer. British journal of biomedical science. 2019:1-6.
67. Qin W, Tang Y, Yang N, Wei X, Wu J. Potential role of circulating microRNAs as a biomarker for unexplained recurrent spontaneous abortion. Fertility and sterility. 2016;105(5):1247-54. e3.
68. Watahiki A, Wang Y, Morris J, Dennis K, O'Dwyer HM, Gleave M, et al. MicroRNAs associated with metastatic prostate cancer. PloS one. 2011;6(9):e24950.
69. Sheng L, He P, Yang X, Zhou M, Feng Q. miR-612 negatively regulates colorectal cancer growth and metastasis by targeting AKT2. Cell death & disease. 2015;6(7):e1808.
70. Tao Z-H, Wan J-L, Zeng L-Y, Xie L, Sun H-C, Qin L-X, et al. miR-612 suppresses the invasive-metastatic cascade in hepatocellular carcinoma. Journal of Experimental Medicine. 2013;210(4):789-803.
71. Xue J, Chi Y, Chen Y, Huang S, Ye X, Niu J, et al. MiRNA-621 sensitizes breast cancer to chemotherapy by suppressing FBXO11 and enhancing p53 activity. Oncogene. 2016;35(4):448.
72. Zhang Y, You W, Zhou H, Chen Z, Han G, Zuo X, et al. Downregulated miR-621 promotes cell proliferation via targeting CAPRIN1 in hepatocellular carcinoma. American journal of cancer research. 2018;8(10):2116.
73. Zhang Y, Sun Z, Zhang Y, Fu T, Liu C, Liu Y, et al. The microRNA-635 suppresses tumorigenesis in non-small cell lung cancer. Biomedicine & Pharmacotherapy. 2016;84:1274-81.

74. Vila-Casadesús M, Vila-Navarro E, Raimondi G, Fillat C, Castells A, Lozano JJ, et al. Deciphering microRNA targets in pancreatic cancer using miRComb R package. Oncotarget. 2018;9(5):6499.
75. Hussein NAEM, El Kholy ZA, Anwar MM, Ahmad MA, Ahmad SM. Plasma miR-22-3p, miR-642b-3p and miR-885-5p as diagnostic biomarkers for pancreatic cancer. Journal of cancer research and clinical oncology. 2017;143(1):83-93.
76. Zhang Y, Luo J, Wang X, Wang HL, Zhang XL, Gan TQ, et al. A comprehensive analysis of the predicted targets of miR-642b-3p associated with the long non-coding RNA HOXA11-AS in NSCLC cells. Oncology letters. 2018;15(5):6147-60.
77. Ningning S, Libo S, Chuanbin W, Haijiang S, Qing Z. MiR-650 regulates the proliferation, migration and invasion of human oral cancer by targeting growth factor independent 1 (Gfi1). Biochimie. 2019;156:69-78.
78. Tang X, Ding Y, Wang X, Wang X, Zhao L, Bi H. miR-650 promotes non-small cell lung cancer cell proliferation and invasion by targeting ING4 through Wnt-1/β-catenin pathway. Oncology letters. 2019;18(5):4621-8.
79. Gong M, Yan C, Jiang Y, Meng H, Feng M, Cheng W. Genome wide bioinformatics analysis reveals CTCFL is upregulated in high-grade epithelial ovarian cancer. Oncology letters. 2019;18(4):4030-9.
80. Chigaev M, Yu H, Samuels D, Sheng Q, Oyebamiji O, Ness S, et al. Genomic positional dissection of RNA Editomes in tumor and normal samples. Frontiers in genetics. 2019;10:211.
81. Wang LY, Guo J, Cao W, Zhang M, He J, Li Z. Integrated sequencing of exome and mRNA of large-sized single cells. Scientific reports. 2018;8(1):1-11.
82. Daniel C, Lagergren J, Öhman M. RNA editing of non-coding RNA and its role in gene regulation. Biochimie. 2015;117:22-7.
83. Chen L, Li Y, Lin CH, Chan THM, Chow RKK, Song Y, et al. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nature medicine. 2013;19(2):209.
84. Han S-W, Kim H-P, Shin J-Y, Jeong E-G, Lee W-C, Kim KY, et al. RNA editing in RHOQ promotes invasion potential in colorectal cancer. Journal of Experimental Medicine. 2014;211(4):613-21.
85. Maas S, Patt S, Schrey M, Rich A. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proceedings of the National Academy of Sciences. 2001;98(25):14687-92.
86. Senapati S, Chaturvedi P, Sharma P, Venkatraman G, Meza JL, El-Rifai W, et al. Deregulation of MUC4 in gastric adenocarcinoma: potential pathobiological implication in poorly differentiated non-signet ring cell type gastric cancer. British journal of cancer. 2008;99(6):949-56.
87. King RJ, Yu F, Singh PK. Genomic alterations in mucins across cancers. Oncotarget. 2017;8(40):67152.
88. Boltin D, Niv Y. Mucins in gastric cancer–an update. Journal of gastrointestinal & digestive system. 2013;3(123):15519.
89. Gupta B, Sikander M, Jaggi M. Overview of mucin (MUC13) in gastrointestinal cancers. J Gastroenterol Dig Dis 2016; 1 (2): 19-18 J Gastroenterol Dig Dis 2016 Volume 1 Issue.2.

90. Si W, Shen J, Zheng H, Fan W. The role and mechanisms of action of microRNAs in cancer drug resistance. Clinical epigenetics. 2019;11(1):25.
91. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. Trends in molecular medicine. 2014;20(8):460-9.
92. Soundararajan R, Stearns TM, Griswold AJ, Mehta A, Czachor A, Fukumoto J, et al. Detection of canonical A-to-G editing events at 3' UTRs and microRNA target sites in human lungs using next-generation sequencing. Oncotarget. 2015;6(34):35726.
93. Liang H, Landweber LF. Hypothesis: RNA editing of microRNA target sites in humans? Rna. 2007;13(4):463-7.
94. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, et al. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nature structural & molecular biology. 2006;13(1):13-21.
95. Wang Y, Liang H. When MicroRNAs meet RNA editing in cancer: a nucleotide change can make a difference. Bioessays. 2018;40(2):1700188.
96. Nigita G, Distefano R, Veneziano D, Romano G, Rahman M, Wang K, et al. Tissue and exosomal miRNA editing in non-small cell lung cancer. Scientific reports. 2018;8(1):1-8.
97. Xiong S, Hu M, Li C, Zhou X, Chen H. Role of miR-34 in gastric cancer: From bench to bedside. Oncology reports. 2019;42(5):1635-46.

Supplementary Materials

S1 File: Summary of the statistics of raw and clean reads and mapping information.

S2 File: Editing rate for each chromosome.

S3 File: List of editing events with recoding impact.

S4 File: Editing events with nonsense-mediated decay impact.

Figures

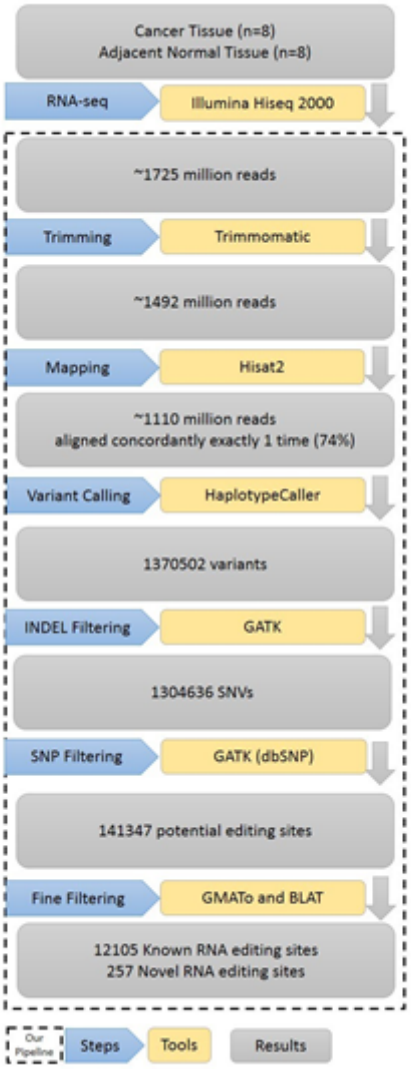

Figure 1

Bioinformatical approach used for the identification of RNA editing sites in normal and cancerous gastric tissue from RNA-seq datasets.

Figure 2

Neighborhood sequence preferences of nucleotides for RNA editing sites.

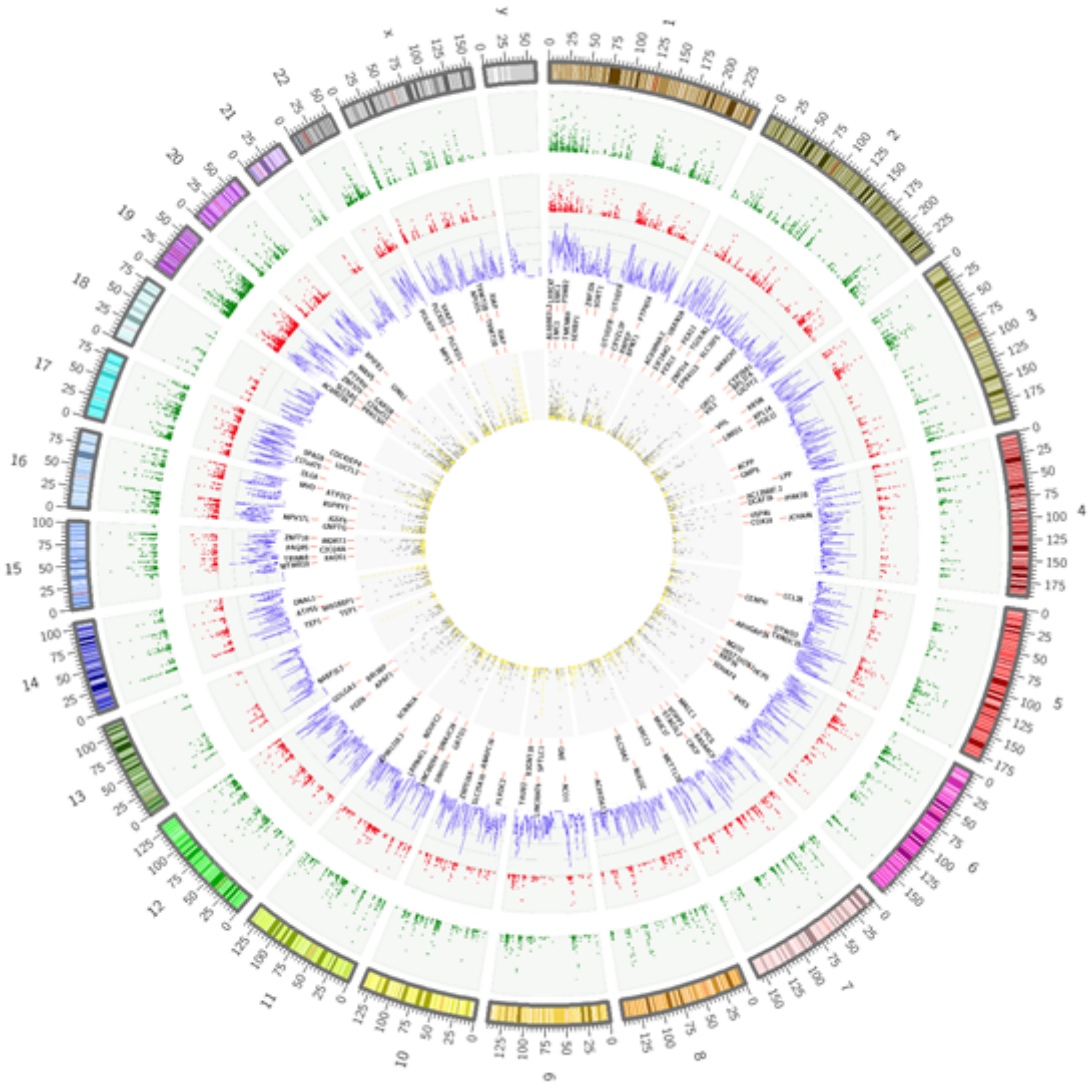

Figure 3

Profiling of RNA editing sites in normal and cancer tissues of gastric cancer patients. Human genome represented as the outermost ring. Each of normal and cancer tissues editing sites is shown by green and red dots. The purple line plot indicates Alu repeat distribution across genome. Also, outer and inner text circles indicate normal-specific differentially edited genes and cancer-specific differentially edited genes, respectively. Yellow bars represent microRNA targeting sites in the genome and grey scatter dots indicate editing sites in these regions.

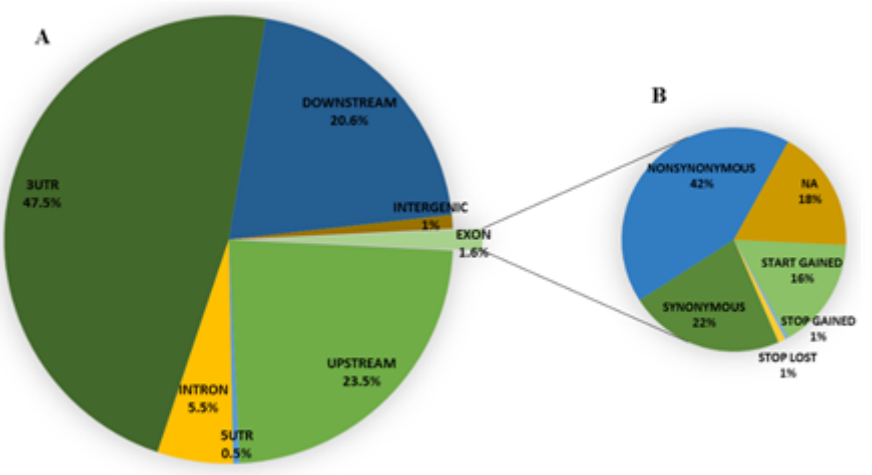

Figure 4

Distribution of RNA editing sites in different genomic regions.

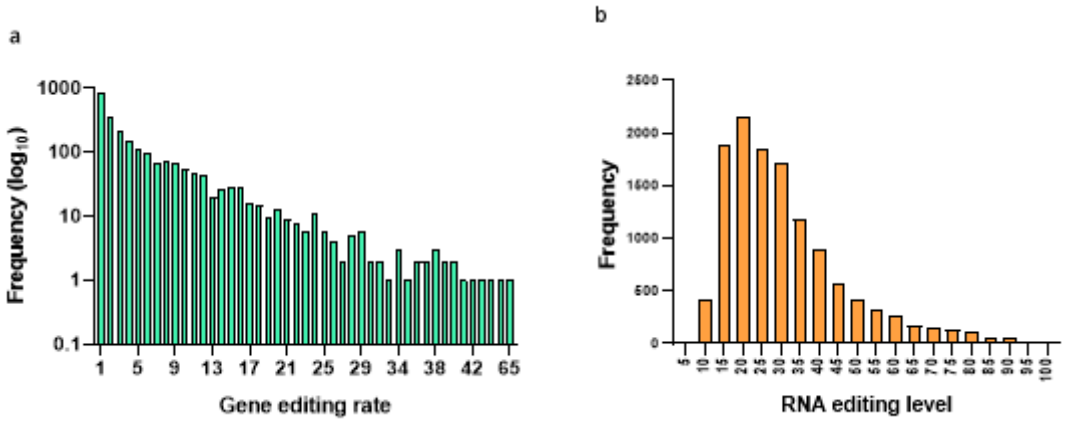

Figure 5

Frequency histogram of gene editing rate (a) and frequency distribution plots of RNA editing levels

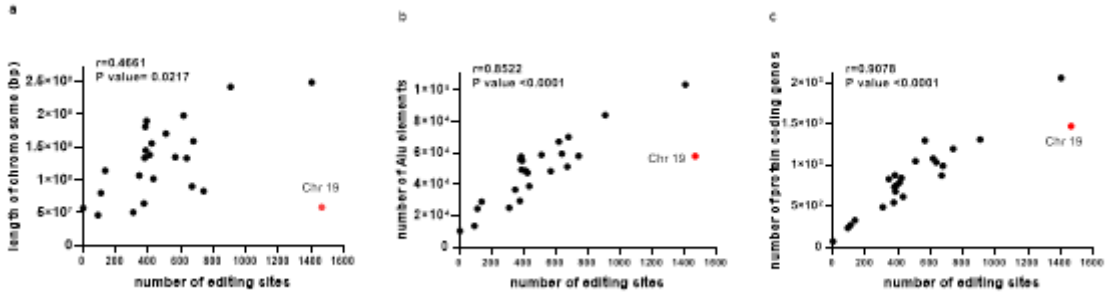

Figure 6

Association between number of editing sites and (a) length of chromosome, (b) number of Alu elements and (c) number of protein coding genes.

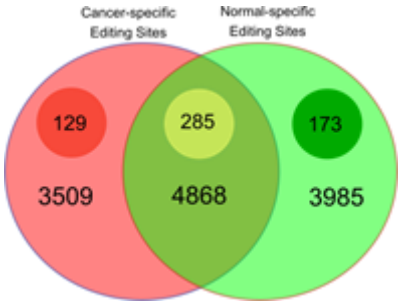

Figure 7

Number of editing sites in cancer and normal tissue. Inner circles indicate number of differentially edited sites between two groups (P value < 0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S1File.xlsx
- S3File.xlsx
- S4File.xlsx
- S2File.xlsx