HHS Public Access

Author manuscript

Cancer. Author manuscript; available in PMC 2018 October 15.

Published in final edited form as:

Cancer. 2017 October 15; 123(20): 3916–3924. doi:10.1002/cncr.30837.

RNA sequencing of esophageal adenocarcinomas identifies novel fusion transcripts, including NPC1-MELK, arising from a complex chromosomal rearrangement

Zhixiong Wang, PhD[1,2,*], **Yulan Cheng, PhD**[2,3,*], **John M. Abraham, PhD**[2,3,*], **Rong Yan, PhD**[2], **Xi Liu, PhD**[2], **Wei Chen, PhD**[1], **Sariat Ibrahim, PhD**[2,3], **Gary P. Schroth, PhD**[5], **Xiquan Ke, PhD**[2], **Yulong He, PhD**[1], and **Stephen J. Meltzer, MD**[2,3,4]

[1]Department of Gastrointestinal Surgery, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[2]Division of Gastroenterology, Johns Hopkins University School of Medicine, Baltimore, USA

[3]Department of Medicine, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, USA

[4]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, USA

[5]Core Applications Group, Illumina, Inc., San Diego, CA, USA

Abstract

Background—Studies of chromosomal rearrangements and fusion transcripts have elucidated mechanisms of tumorigenesis and led to targeted cancer therapies. In this study, we aimed to identify novel fusion transcripts in esophageal adenocarcinomas.

Methods—To identify new fusion transcripts associated with esophageal adenocarcinoma, we performed targeted RNA sequencing and PCR verification in 40 esophageal adenocarcinomas (EACs) and matched non-malignant specimens from the same patients. Genomic PCR and Sanger sequencing were performed to find the breakpoint of fusion genes.

Corresponding author: Stephen J. Meltzer, MD, Division of Gastroenterology, Department of Medicine, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, 1503 E. Jefferson St., Rm. 112. Baltimore, MD 21287; Fax: (410) 502-1329; smeltzer@jhmi.edu.

*The first 3 authors contributed equally to this article.

Conflict of interest disclosures

No any conflicts of interest.

Author Contributions

Conceived and designed the experiments: SJM, GPS, and ZW; Wrote the paper: ZW and SJM; Performed the experiments: ZW, YC, JA, XL, LCW, SMG, and XK; Analyzed the data: ZW, RY, YH, and WC.

Data access

The raw and processed sequencing data generated in this study have been submitted to the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94869).

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Johns Hopkins University School of Medicine. All patients provided written informed consent.

Results—Five novel in-frame fusion transcripts were identified and verified in 40 EACs as well as in a validation cohort of 15 additional EACs (in total, 55 patients): FGFR2-GAB2 (2/55, or 3.6%), NPC1-MELK (2/55, or 3.6%), USP54-CAMK2G (2/55, or 3.6%), MKL1-FBLN1 (1/55, or 1.8%), and CNOT2-C12orf49 (1/55, or 1.8%). Genomic analysis indicated that NPC1-MELK arose from a complex inter-chromosomal translocation event involving chromosomes 18, 3, and 9 with three rearrangement points, consistent with chromoplexy.

Conclusions—These data indicate that fusion transcripts occur at a stable frequency in EAC. Furthermore, our results indicate that chromoplexy is an underlying mechanism that generates fusion transcripts in EAC. These and other fusion transcripts merit further study as diagnostic markers and potential therapeutic targets in EAC.

Keywords

FGFR2 fusion; NPC fusion; fusion transcript; esophageal adenocarcinomas; chromoplexy

Background

Esophageal cancer is the eighth most common cancer and the sixth leading cause of cancer-related death in the world [1]. Esophageal cancer is composed of two main histological subtypes: squamous cell carcinoma and adenocarcinoma [2]. Over the past 30 years, the incidence of EAC has risen approximately7-fold in the United States [3]. Esophageal adenocarcinoma (EAC) is now the predominant subtype of esophageal cancer in the United States [1]. While the five-year survival of EAC has remained below 25%, largely due to late diagnosis [4], Barrett's esophagus (BE), a pre-malignant lesion, is the most significant risk factor for developing EAC[5]. It has been reported that 5% of BE cases will ultimately progress into EAC[6]. Mechanisms underlying malignant transformation in BE are poorly understood, but improved understanding of these mechanisms could lead to improved survival by providing biomarkers for early detection as well as therapeutic target strategies.

Molecular genetic studies of EAC have revealed that mutations in key cancer-causing genes, such as TP53, CDKN2A, and SMAD4, are involved in the genesis and advancement of EAC[7–9]. In addition, chromosomal rearrangements may also play a critical role in EAC development. Recently, whole-genome sequencing (WGS) of EAC have confirmed that chromosomal rearrangements occur in EAC[9,10]. Chromosomal rearrangements can exert their effects through one of two alternative mechanisms: 1) dysregulation, usually resulting in the overexpression of a normal gene at one of the breakpoints, or 2) the creation of a fusion protein[11]. There is extensive evidence that fusion proteins participate in carcinogenesis. One classic example is the fusion gene BCR-ABL, which occurs in patients with CML, ALL and AML[11]. Notably, fusion genes can also serve as diagnostic markers and therapeutic targets, as shown for FGFR3-TACC3 fusions in glioblastoma and EML4-ALK fusions in lung cancer[12,13]. Although chromosomal rearrangements have been reported in EAC, no driver events caused by fusion genes have yet been described in this malignancy. Recently, Blum and colleagues identified RPS6KB1-VMP1 as a fusion transcript that modulates autophagy in EAC[14]. With the exception of this single study, however, there have been no comprehensive reports of fusion transcripts in EAC. Identification of fusion events is essential for enhancing our understanding of EAC and for promoting the development of

early detection biomarkers as well as targeted therapeutic strategies to improve outcome in this disease, as discussed above.

In the current report, we identified and characterized five novel fusion transcripts using next generation RNA sequencing in 40 paired EAC specimens and non-malignant specimens. In addition, genomic analysis revealed that one fusion transcript, NPC1-MELK, resulted from a complex structural rearrangement event involving three chromosomes, typical of the recently proposed phenomenon known as chromoplexy[15].

Methods

Patients and samples

EAC specimens and adjacent non-malignant tissues from 40 patients were obtained from the Johns Hopkins University and University of Maryland Hospitals. An additional independent cohort of 15 patients from the Johns Hopkins University and University of Maryland Hospitals was also studied. All patients provided written informed consent under protocols approved by the institutional review boards at the Johns Hopkins University or University of Maryland Schools of Medicine. All tissues were pathologically confirmed as EAC.

Cell lines

Three human EAC-derived cell lines (OE33, SKGT4, and Flo-1), 5 human gastric cancer-derived cell lines (MKN28, KATOIII, AGS, NCIN87, and SNU1), immortalized normal esophageal epithelial cells (Het1A), and immortalized normal gastric epithelial cells (HFE145) from were also analyzed in this study. Cell lines, KATOIII, AGS, NCIN87, SNU1 and Het1A, were purchased from ATCC (American Type Culture Collection, Manassas, Virginia, USA). MKN28 was gifted by Dr. Gen Tamura (Department of Pathology, Yamagata University School of Medicine, Japan), and HFE145 cells was gifted by Dr. Duane Smoot (Division of Gastroenterology, Howard University College of Medicine, USA). OE33, SKGT4, and Flo-1, were purchased from the European Collection of Cell Culture (Porton Down, UK), Sigma Chemical (St Louis, Missouri, USA), and ScienCell Research Laboratories (Carlsbad, California, USA), respectively.

cDNA library preparation and sequencing

Total RNA from 40 paired EAC specimens was extracted using an RNeasy Mini Kit (QIAGEN). Total RNA was converted into cDNA library, then selectively enriched for 1385 cancer-associated genes using the TruSight RNA Pan-Cancer Panel kit according to the manufacturer's instructions (Illumina). The TruSight RNA Pan-Cancer panel is a targeted sequencing strategy that enables measurement of gene expression, variant calling, and fusion detection of cancer-related genes with limited sample quantity. This approach has the ability to detect fusions between any of the 1385 genes in the panel, but also between any of these genes and other novel fusion partners, for instance FGFR1-X fusions, where X is any possible gene. All RNA sequencing was performed on MiSeq instruments with V3 reagents according to the manufacturer's instructions (Illumina).

Bioinformatics analysis

We used The BaseSpace® RNA-Seq Alignment v1.0 App to analyze sequence data (Illumina). The RNA-Seq Alignment App uses the following methods to analyze sequencing data: STAR, Manta, Cufflinks, and Tuxedo Suite (including Bowtie, Bowtie2, and TopHat). The Homo sapiens UCSC hg19 reference genome was used for alignment. Input reads were filtered out as per the following conditions: 1) Reads that failed base calling quality checks; 2) Reads marked as PCR duplicates; 2) Paired-end reads not marked as a proper pair; and 3) Reads with a mapping quality < 20. Spliced Transcripts Alignment to a Reference (STAR) is a fast RNA-seq read mapper, with support for splice-junction and fusion read detection. STAR aligns reads by finding the Maximal Mappable Prefix (MMP) hits between reads (or read pairs) and the reference genome. Manta was used in combination with STAR to identify fusion transcripts. Fusion transcripts were filtered based on the following criteria: 1) Read counts across fusion and alignment qualities; 2) Genome-wide realignment of fusion contigs to filter candidates that could be explained by a local alignment elsewhere in the genome; and 3) length of coverage around breakpoints, indicating the presence of stable fusion transcripts. Only high-confidence candidate fusion transcripts passing these filters were selected for downstream analysis.

Reverse transcription-PCR (RT-PCR)

Total RNA from tissues or cell lines was extracted using an RNeasy Mini kit (QIAGEN, Aarhus, Denmark). 0.5 μg of total RNA were reverse-transcribed into cDNA using a ThermoScript RT-PCR System kit (Invitrogen, Grand Island, NY), following the manufacturer's instructions. Resulting cDNA was used as a template for PCR amplification using the following primers spanning fusion transcript junctions:

FGFR2-sense: 5'-ACAAAAAGACCACCAATGGGC-3';

GAB2-anti-sense: 5'-TGCTCTGGACCCACTTATTCA-3';

NPC1-sense: 5'-AACAATGATTCCCTGGTGCA-3';

MELK-anti-sense: 5'-TATCACACCCACACTCATCCG-3';

MKL1-sense: 5'-TGAGCGGAAGAATGAAAAATG-3';

FBLN-anti-sense: 5'-TACAGACCACAGTCACTGGCA;

USP54-sense: 5'-AATTTGGTGGACCAAGCCC-3';

CAMK2G-anti-sense: 5'TCTGGCCACAATGTCTTCAA-3';

CNOT2-sense: 5'-TGCAATAAAACTTGGCCGA-3';

C12orf19-anti-sense: 5'-ACAACACGAGAAGGGGTCAAA-3';

MEGF9-sense: 5'-AGCAGCAACAGCAGCGTCCT-3';

GSN-anti-sense: 5'-ACCATGCTGTTGGGCTAACAA-3'.

PCR amplification was performed using Taq DNA polymerase (Invitrogen). PCR products were analyzed by agarose gel electrophoresis.

Genomic DNA PCR

In order to precisely locate the DNA breakpoints resulting in the fusion transcript NPC1-MELK, genomic DNA PCR was performed. DNA was extracted from frozen tissues containing the NPC1-MELK transcript using a DNeasy Blood and Tissue Kit (QIAGEN). 0.1 μg of DNA were used as a template for PCR amplification using Platinum Taq DNA polymerase (Invitrogen). The primers used were: Forward primer: 5'-ATTCCTGCCCATGTTCCTTT-3'; Reverse primer: 5'-CCCAAAATCTGACTGTGTTTG-3'. PCR products were analyzed by agarose gel electrophoresis.

TOPO cloning and Sanger sequencing

Amplified products of RT-PCR and genome PCR were cloned into vector pCR™4-TOPO TA Vector and transformed into E.coli according to the manufacturer's instructions (Thermo-Fisher Scientific). Then, the vector was extracted using a Plasmid Mini Kit (QIAGEN) and sequenced via Sanger sequencing according to the manufacturer's instructions.

Results

RNA sequencing of 40 EACs and matched normal specimens

Forty freshly frozen EAC specimens and their matched non-malignant tissues were collected for targeted RNA sequencing, as described above. The demography of the patients is shown in Supplemental Table S1. The mean read count was 3,543,321 and the mean ratio of aligned reads was 96%. To identify fusion transcripts, we analyzed paired-end sequencing reads using Manta according to the strategy mentioned above.

Identification and validation of fusion transcripts

After Manta workflow, we identified 6 novel high-confidence fusion transcripts in the initial 40-patient cohort. These 6 fusion transcripts comprised FGFR2-GAB2 (3/40), NPC1-MELK (2/40), USP54-CAMK2G (2/40), MKL1-FBLN1 (2/40), CNOT2-C12orf49 (1/40), and MEGF9-GSN (1/40). In order to validate fusion transcripts found by RNA-seq, PCR was performed using primers spanning the junctions of each fusion transcript. Except for MEGF9-GSN, 5 of these 6 identified fusion transcripts were validated by PCR (Figures 1C–G). In addition, one of three samples showing FGFR2-GAB2 and one of two samples showing MKL1-FBLN1 were not validated by RT-PCR, suggesting that these findings were due to false-positive RNA-seq reads. Supplemental Table S2 shows read counts that support fusion transcripts and the RNA sequences in those conjunctions. These read maps show that FGFR2-GAB2 and NPC1-MELK were detected at high sequencing depth (Figures 1A and 1B). Next, we employed Sanger sequencing to sequence the junction regions amplified by PCR, which yielded 100% concordance between RNA-seq and Sanger sequencing (Figures 1C–G). Sequencing of these junction regions demonstrated that exon 19 of FGFR2 was fused to exon 1 of GAB2, exon 20 of NPC1 to exon 18 of MELK, exon 2 of MKL1 to exon 17 of FBLN1, exon 14 of USP54 to exon 4 of CAMK2G, and exon 15 of CNOT2 to exon 4 of C12orf19.

To confirm the recurrent nature and frequencies of these 5 fusions, we screened 15 additional independent paired EAC samples, 3 EAC cell lines (OE33, SKGT4, Flo-1), one immortalized normal esophageal epithelial cell line (Het1a), 5 gastric cancer cell lines (MKN28, KATOIII, AGS, NCIN87, SNU1) and one immortalized normal gastric epithelial cell line (HFE145). However, none of these five fusion transcripts were observed in this additional cohort or in any of the studied cancer lines (Figure 1C, data from cell lines not shown).

FGFR2-GAB2 and NPC1-MELK arose from fusion of sequences located on different chromosomes. USP54-CAMK2G, MKL1-FBLN1 and CNOT2-C12orf49 arose from fusion of sequences located on the same chromosome. Distances between the two partners of fusion transcripts located in the same chromosome were 331,209–46,415,586 base pairs (331,209 in USP54-CAMK2G, 5,136,964 in MKL1-FBLN1, and 46,415,596 in CNOT2-C12orf49) (Table 1). Given the large distances between the respective partners of these fusion genes, we speculate that these fusion genes more likely resulted from chromosomal rearrangement, than from transcriptional read-through. Given that the two partners of NPC1-MELK and FGFR2-GAB2 were located on different chromosomes, we prioritized NPC1-MELK and FGFR2-GAB2 for further studies.

The FGFR2-GAB2 fusion transcript in human EACs

Fusion events involving FGFR2 are frequent in human solid tumors, including intrahepatic cholangiocarcinoma (iCCA), glioblastoma, breast, lung and prostate cancers - *e.g.*, FGFR2-PPHLN1, FGFR2-BICC1, and FGFR2-CCDC6 [16,17]. In the current study, two of 55 EAC patients, EAC3869 and EAC5061, were validated by PCR to be harboring the FGFR2-GAB2 fusion transcript (Figures 1C and D). RNA sequencing of the junction region showed that exon 19 of FGFR2 was fused to exon 2 of GAB2 in-frame (Figure 2A). In the normal genome, FGFR2 and GAB2 map to chromosomes 10q26 and 11q14.1, respectively. Given the high frequency of fusion events involving FGFR2 in other cancer types, we conclude that the FGFR2-GAB2 fusion transcript reported here likely resulted from a genomic DNA rearrangement. Based on sequencing of this FGFR2-GAB2 fusion transcript, the putative structure of the FGFR2-GAB2 fusion protein is displayed in Figure 2B. Most of FGFR2 and a portion of GAB2 were retained in this putative fusion protein.

Complex chromosome translocation underlying the NPC1-MELK fusion transcript

In order to characterize the genomic translocation giving rise to the NPC1-MELK fusion transcript, we extracted DNA from the two EAC specimens (EAC3253 and EAC3274) containing this fusion, then PCR-amplified the junction containing the breakpoint using genomic DNA PCR and sequenced this product using Sanger sequencing (Figure 3B). Interestingly, sequencing results revealed that two additional DNA fragments were inserted into the junction region. One fragment consisted of partial intron 3 and partial exon 4 of MDM2, located on chromosome 3, while the other fragment comprised a partial antisense isoform of MELK (Figure 3C). Both of these intronic fragments were apparently removed, and the end of exon 20 of NPC1 was apparently spliced to the beginning of exon 18 of MELK, during the process of RNA splicing (Figure 3A). Therefore, we concluded that the NPC1-MELK fusion transcript was derived from complex inter- and intra-chromosomal

rearrangements consisting of three fusion points involving chromosomes 18, 3 and 9, consistent with the recently described phenomenon of chromoplexy [15]. Additionally, the two samples harboring NPC1-MELK (EAC3253-T and EAC3274-T) also contained other fusion transcripts, as shown in Figure 1C, suggesting that EAC3253-T and EAC3274-T were subject to complex chromosomal rearrangements having at least three fusion points involving more than a single chromosome[18]. In addition, both patients had advanced tumors (T4N2M0 and T3aN1M0, respectively). Finally, both patients had received preoperative adjuvant chemoradiotherapy and had obtained a partial response (PR) to therapy.

Discussion

Chromosomal rearrangements are a hallmark of carcinogenesis in many human tumor types. Recently, several large cohort studies based on next-generation sequencing have reported high-frequency chromosomal rearrangements in esophageal adenocarcinoma (EAC) [9,10]. However, these studies did not explore the effects of these rearrangements and consequent fusion transcripts on progression of EAC. In the current study, we identified 5 novel fusion transcripts in EAC, comprising FGFR2-GAB2, NPC1-MELK, USP54-CAMK2G, MKL1-FBLN1, and CNOT2-C12orf49. In addition, our data reveal that complex genomic rearrangements, typical of the recently described concept of chromoplexy [15], gave rise to the NPC1-MELK fusion transcript.

FGFR2 fusion events have been recently described in multiple cancer types, including glioblastoma, iCCA, bladder, lung, breast and prostate cancers. Multiple partner genes were identified in these fusions, including PPHLN1, BICC1, AFF3, CASP7, KIAA1967, OFD1, CCDC6, TXLN1, KCTD1, and ACSL5[16,17,19,20]. FGFR2 fusions are, in fact, the most recurrent chromosomal rearrangements in human tumors. It was reported that FGFR2-PPHLN1 occurred in 45% of iCCA, and FGFR2-BICC1 occurred in 38% of iCCA[16]. However, the FGFR2-GAB2 fusion found in this study was a low-frequency event, being detected in only 2 of 55 (3.6%) patients with EAC, suggesting that different cancer types may be unique in this regard. In addition, we also tested these 5 fusion genes in gastric cancer cell lines, where however no fusion gene was found. Whether these fusion genes occur in other tumor types is still unknown and worthy of further study. Given that FGFR2 is a frequent target of genomic rearrangement in many human tumors, it will be quite worthwhile to search for FGFR2-GAB2 in other tumor types. In addition, exploring these fusion genes in other datasets is a good strategy to discover whether these fusion transcripts are recurrent in esophageal adenocarcinomas or other tumor types in other clinical centers. We hope to collaborate with bioinformatics specialists to tackle this undertaking in future studies.

Interestingly, nine of ten FGFR2 fusions reported to date involve the same breakpoint (exon 19) and contain the principal functional domain of WT FGFR2[16,17]. It was reported that most of the FGFR fusion partners contribute oligomerization domains to induce activation of FGFR kinase [17]. For example, Sia *et al.* verified that the phosphorylation level of FGFR2-PPHLN1 fusion protein and activation of downstream MAP kinase ERK1/2 increased compared with wild-type FGFR2. In the current study, we also demonstrated that exon1~19 of FGFR2 were present in the 5’ end of our fusion transcript. Thus, FGFR2-GAB2 has a

functional structure analogous to FGFR2 kinase fusions described in other cancer types, suggesting a shared mechanism involving the kinase. While the precise function of FGFR2-GAB2 in EAC is still unknown, this hypothesis merits further study.

According to the findings of FGFR2 fusion genes in previous studies as well as our own, the breakpoint location of FGFR2 rearrangements are stable in intron 19. However, this finding doesn't necessarily mean that other genes involved in fusions also have stable breakpoints. Actually, even the same fusion gene may have many isoforms and different breakpoints. For example, FGFR3-TACC3 is a recurrent fusion gene found in nasopharyngeal carcinoma. It was reported that different isoforms of FGFR3-TACC3 occur, such as a gene fusion between exon 18 of FGFR3 and exon 6 or exon 14 of TACC3, or between exon 19 of FGFR3 and exon 11 of TACC3[21].

Additionally, it was reported that loss of the 3' UTR, which contains targets of miRNAs, results in FGFR3 fusion transcript overexpression[22]. Consistent with these observations, the 5' FGFR2 portion of FGFR2-GAB2 in our study also lost the 3'-UTR. This finding suggests that up-regulated expression of FGFR2-GAB2 in EAC may be due, at least in part, to escape from miRNA regulation. In addition, drug targeting of FGFR2 fusions has proven to be an effective therapeutic approach in certain cancers. For example, Daniela *et al.* proved that HUCCT1 cells overexpressing the FGFR2-PPHLN1 fusion protein showed enhanced sensitivity to the chemotherapeutic agent BGJ398 compared with the parental cell line transfected with empty vector[16]. Thus, FGFR2-GAB2 may represent a potential target of FGFR2 inhibitors in EAC patients; this possibility merits further study.

With the increasing application of whole-genome and whole-transcriptome sequencing approaches, it has become evident that chromosomal rearrangements are more frequent and complex than previously recognized[23,24]. Complex chromosomal rearrangements which contain at least three genomic breakpoints involving more than a single chromosome, have been detected in a broad spectrum of tumors, including prostate cancer, bone cancer, lung cancer, and medulloblastoma[25,26]. For example, Baca *et al.* reported that 88% of prostate tumors contained chains of five or more re-arrangements[15]. It is difficult to explain this phenomenon with the classic view that chromosomal rearrangements are caused by a gradual accumulation of DNA damage. Recently, several new mechanisms, such as chromothripsis and chromoplexy, have been proposed to account for complex chromosomal rearrangements. According to these and other new mechanisms, complex rearrangements can be derived from a single catastrophic event, rather than gradually [27]. The distinction between these two mechanisms is that genomic breakpoints in chromothripsis, usually numbering in the hundreds, are clustered within specific regions of one or two chromosomes; whereas those in chromoplexy are unclustered, usually numbering in the tens and involving multiple chromosomes[23].

Complex chromosomal rearrangements similar to chromothripsis were found in 44 of 123 EACs [10]. In the current study, we detected multiple fusion transcripts in 2 of 55 EAC patients (NPC1-MELK, USP54-CAMK2G, MKL1-FBLN1, and CNOT2-C12orf49 in EAC3253 and NPC1-MELK, and USP54-CAMK2G in EAC3274). In addition, three fusion points were detected in the junction region of fusion transcript NPC1-MELK, involving

complex inter- and intra- chromosomal rearrangements. The relatively low frequency of complex chromosomal rearrangements detected in our study may have been due to the targeted RNA sequencing method we used, which covers only 1385 tumor-related genes rather than the whole transcriptome. The NPC1-MELK fusion transcript, involving 3 breakpoints and three chromosomes in total, is consistent with chromoplexy. We speculate that NPC1-MELK, USP54-CAMK2G, MKL1-FBLN1, and CNOT2-C12orf49 may comprise components of this chromoplexy. Chromoplexy accounts for several known fusion genes. For example, the TMPRSS2-ERG fusion, firstly found by Tomlins in 2005, is a high-frequency recurrent fusion gene in prostate cancer [28]. Baca *et al.* demonstrated that TMPRSS2-ERG was often found as part of chromoplexy in prostate cancer [15,23]. Taken together, our and others' finding suggest that chromoplexy may comprise an underlying mechanism causing complex fusion events in EAC. Next-generation DNA sequencing will be an efficient method to identify chromoplexy in additional EAC and other types of tumors.

In addition, both chromoplexic tumors were of advanced stages. Interestingly, both patients with chromoplexy also had favorable responses to chemoradiotherapy. While this sample size is too small to statistically analyze, these findings imply that the involvement of chromoplexy in cancer progression and chemoradiosensitivity, as well as the functional effects of fusion transcripts, merit further research.

Conclusions

In conclusion, we discovered and validated five novel fusion transcripts in esophageal adenocarcinoma. Chromoplexy may underlie complex chromosomal rearrangements that produce fusion genes in EAC. These fusion transcripts and this mechanism merit further study as potential diagnostic markers and therapeutic targets in EAC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Stephen M. Gross and Lisa C. Watson (Scientist II, Core Applications, Illumina) for their help with data analysis. Zhixing Wang was supported by a Scholarship from the China Scholarship Council (CSC). Dr. Meltzer is an American Cancer Society Clinical Research Professor and the Harry and Betty Myerberg/Thomas R. Hendrix Professor of Gastroenterology at Johns Hopkins.

Funding support

This work was supported by the National Institutes of Health (Grants: CA190040, CA211457).

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65(2):87–108. 2015-03-01. [PubMed: 25651787]
2. DeMeester SR. Adenocarcinoma of the esophagus and cardia: a review of the disease and its treatment. Ann Surg Oncol. 2006; 13(1):12–30. 2006-01-01. [PubMed: 16378161]
3. Pohl H, Sirovich B, Welch HG. Esophageal adenocarcinoma incidence: are we reaching the peak? Cancer Epidemiol Biomarkers Prev. 2010; 19(6):1468–1470. 2010-06-01. [PubMed: 20501776]

4. Koppert LB, Wijnhoven BP, van Dekken H, Tilanus HW, Dinjens WN. The molecular biology of esophageal adenocarcinoma. J Surg Oncol. 2005; 92(3):169–190. 2005-12-01. [PubMed: 16299787]
5. Maley CC, Rustgi AK. Barrett's esophagus and its progression to adenocarcinoma. J Natl Compr Canc Netw. 2006; 4(4):367–374. 2006-04-01. [PubMed: 16569389]
6. Li X, Galipeau PC, Paulson TG, et al. Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. Cancer Prev Res (Phila). 2014; 7(1):114–127. 2014-01-01. [PubMed: 24253313]
7. Dolan K, Walker SJ, Gosney J, Field JK, Sutton R. TP53 mutations in malignant and premalignant Barrett's esophagus. Dis Esophagus. 2003; 16(2):83–89. 2003-01-20. [PubMed: 12823203]
8. Frankel A, Armour N, Nancarrow D, et al. Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. Genes Chromosomes Cancer. 2014; 53(4):324–338. 2014-04-01. [PubMed: 24446147]
9. Dulak AM, Stojanov P, Peng S, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat Genet. 2013; 45(5):478–486. 2013-05-01. [PubMed: 23525077]
10. Nones K, Waddell N, Wayte N, et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. Nat Commun. 2014; 5:5224. 2014-10-29. [PubMed: 25351503]
11. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer. 2007; 7(4):233–245. 2007-04-01. [PubMed: 17361217]
12. Laszlo A, Thotala D, Hallahan DE. Membrane phospholipids, EML4-ALK, and Hsp90 as novel targets in lung cancer treatment. Cancer J. 2013; 19(3):238–246. 2013-05-01. [PubMed: 23708071]
13. Singh D, Chan JM, Zoppoli P, et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. Science. 2012; 337(6099):1231–1235. 2012-09-07. [PubMed: 22837387]
14. Blum AE, Venkitachalam S, Guo Y, et al. RNA Sequencing Identifies Transcriptionally Viable Gene Fusions in Esophageal Adenocarcinomas. Cancer Res. 2016; 76(19):5628–5633. 2016-10-01. [PubMed: 27503924]
15. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013; 153(3):666–677. 2013-04-25. [PubMed: 23622249]
16. Sia D, Losic B, Moeini A, et al. Massive parallel sequencing uncovers actionable FGFR2-PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. Nat Commun. 2015; 6:6087. 2015-01-22. [PubMed: 25608663]
17. Wu YM, Su F, Kalyana-Sundaram S, et al. Identification of targetable FGFR gene fusions in diverse cancers. Cancer Discov. 2013; 3(6):636–647. 2013-06-01. [PubMed: 23558953]
18. Kloosterman WP, Guryev V, van Roosmalen M, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet. 2011; 20(10):1916–1924. 2011-05-15. [PubMed: 21349919]
19. Nakamura H, Arai Y, Totoki Y, et al. Genomic spectra of biliary tract cancer. Nat Genet. 2015; 47(9):1003–1010. 2015-09-01. [PubMed: 26258846]
20. Kim SY, Ahn T, Bang H, et al. Acquired resistance to LY2874455 in FGFR2-amplified gastric cancer through an emergence of novel FGFR2-ACSL5 fusion. Oncotarget. 2017 2017-01-21.
21. Yuan L, Liu ZH, Lin ZR, Xu LH, Zhong Q, Zeng MS. Recurrent FGFR3-TACC3 fusion gene in nasopharyngeal carcinoma. Cancer Biol Ther. 2014; 15(12):1613–1621. 2014-01-20. [PubMed: 25535896]
22. Parker BC, Annala MJ, Cogdell DE, et al. The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. J Clin Invest. 2013; 123(2):855–865. 2013-02-01. [PubMed: 23298836]
23. Shen MM. Chromoplexy: a new category of complex rearrangements in the cancer genome. Cancer Cell. 2013; 23(5):567–569. 2013-05-13. [PubMed: 23680143]
24. Malhotra A, Lindberg M, Faust GG, et al. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. Genome Res. 2013; 23(5):762–776. 2013-05-01. [PubMed: 23410887]

25. Kloosterman WP, Guryev V, van Roosmalen M, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet. 2011; 20(10):1916–1924. 2011-05-15. [PubMed: 21349919]

26. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011; 144(1):27–40. 2011-01-07. [PubMed: 21215367]

27. Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. Genes Dev. 2013; 27(23):2513–2530. 2013-12-01. [PubMed: 24298051]

28. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310(5748):644–648. 2005-10-28. [PubMed: 16254181]

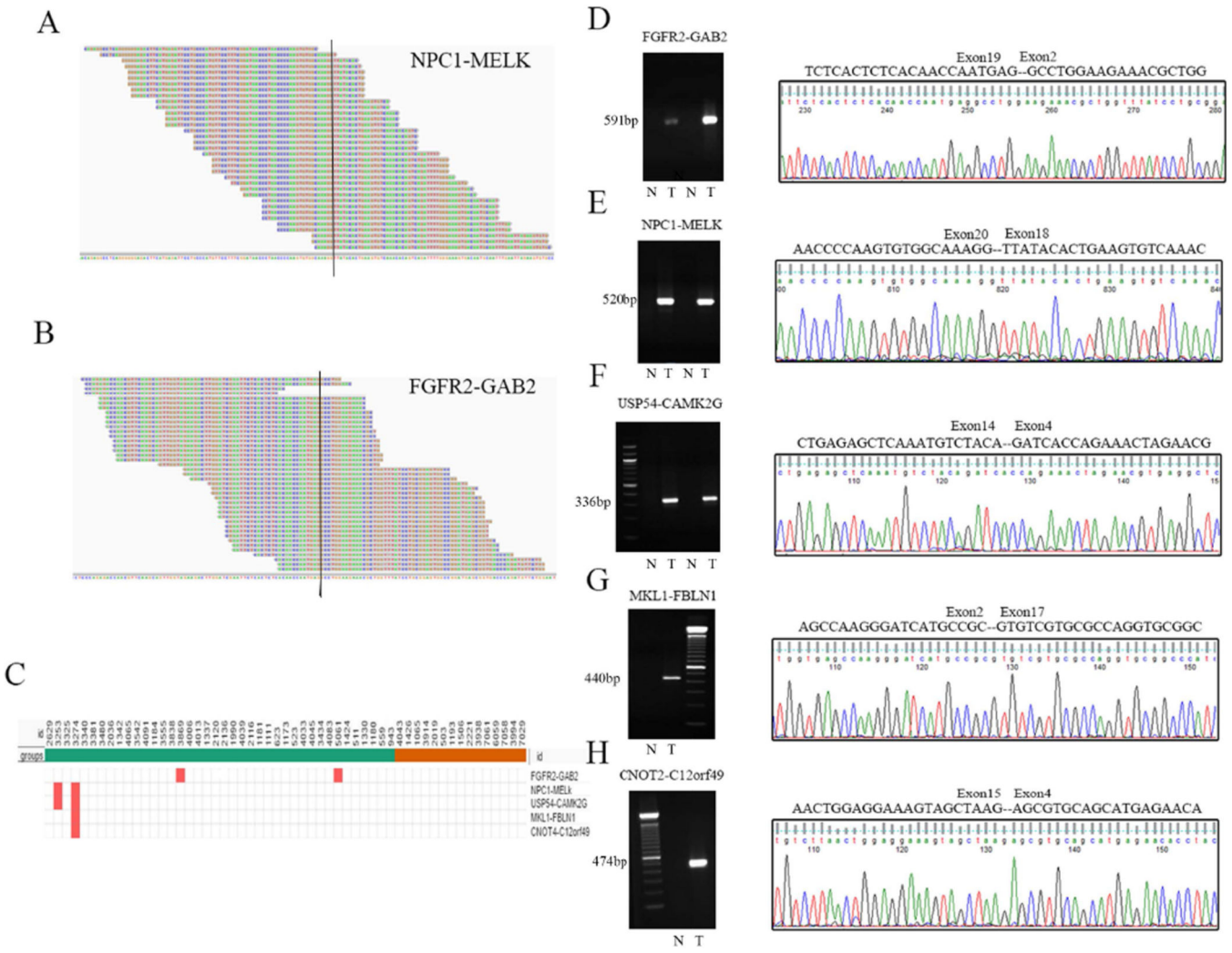

Figure 1. RNA-seq identifies 5 novel fusion transcripts in esophageal adenocarcinomas

(A) Read map of NPC1-MELK; (B) Read map of FGFR2-GAB2; (C) 55 total cases (40 cases in the green group belong to the initial RNA-seq cohort) were screened for the presence of fusion transcripts, yielding the following frequencies: FGFR2-GAB2 (2/55, or 3.6%), NPC1-MELK (2/55, or 3.6%), USP54-CAMK2G (2/55, or 3.6%), MKL1-FBLN1(1/55, or 1.8%), and CNOT2-C12orf49 (1/55, or 1.8%). (D-H) RT-PCRs of paired tumor/normal cases using primers spanning each fusion junction confirm that bands corresponding to the 5 fusion transcripts were present only in tumor, but not in normal tissue. Sanger sequencing confirms fusion junction sequences of all 5 fusion transcripts and revealed that exon 19 of FGFR2 was fused to exon 1 of GAB2, exon 20 of NPC1 to exon 18 of MELK, exon 2 of MKL1 to exon 17 of FBLN1, exon 14 of USP54 to exon 4 of CAMK2G, and exon 15 of CNOT2 to exon 4 of C12orf19.

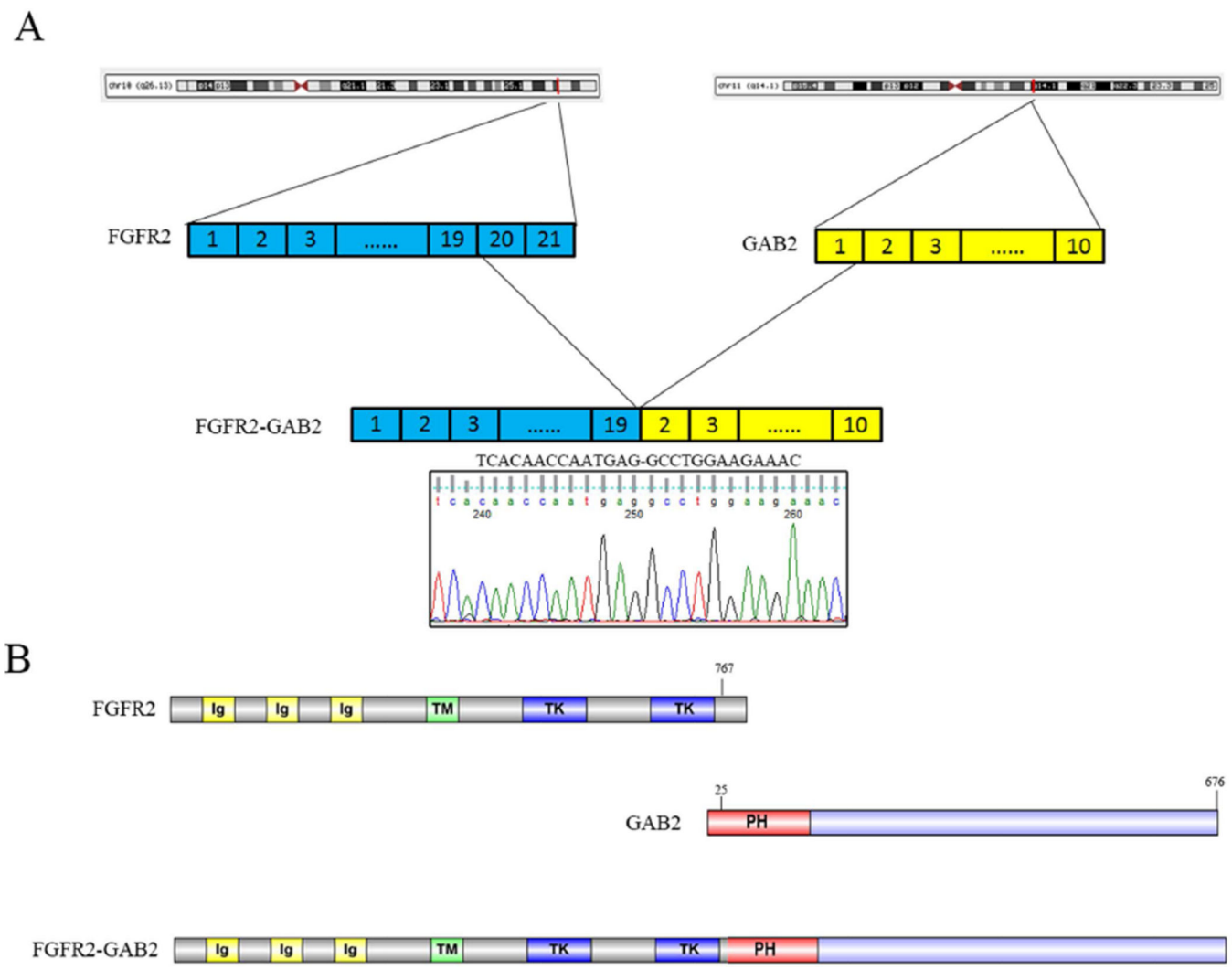

Figure 2. Schematic representations of the predicted FGFR2-GAB2 fusions identified by RNA sequencing

A: FGFR2-GAB2 fusion transcript consists of 5' exons 1~19 of FGFR2 and 3' exons 2~10.

B: FGFR2-GAB2 protein (1418aa) consists of N-terminal FGFR2 and C-terminal GAB2.

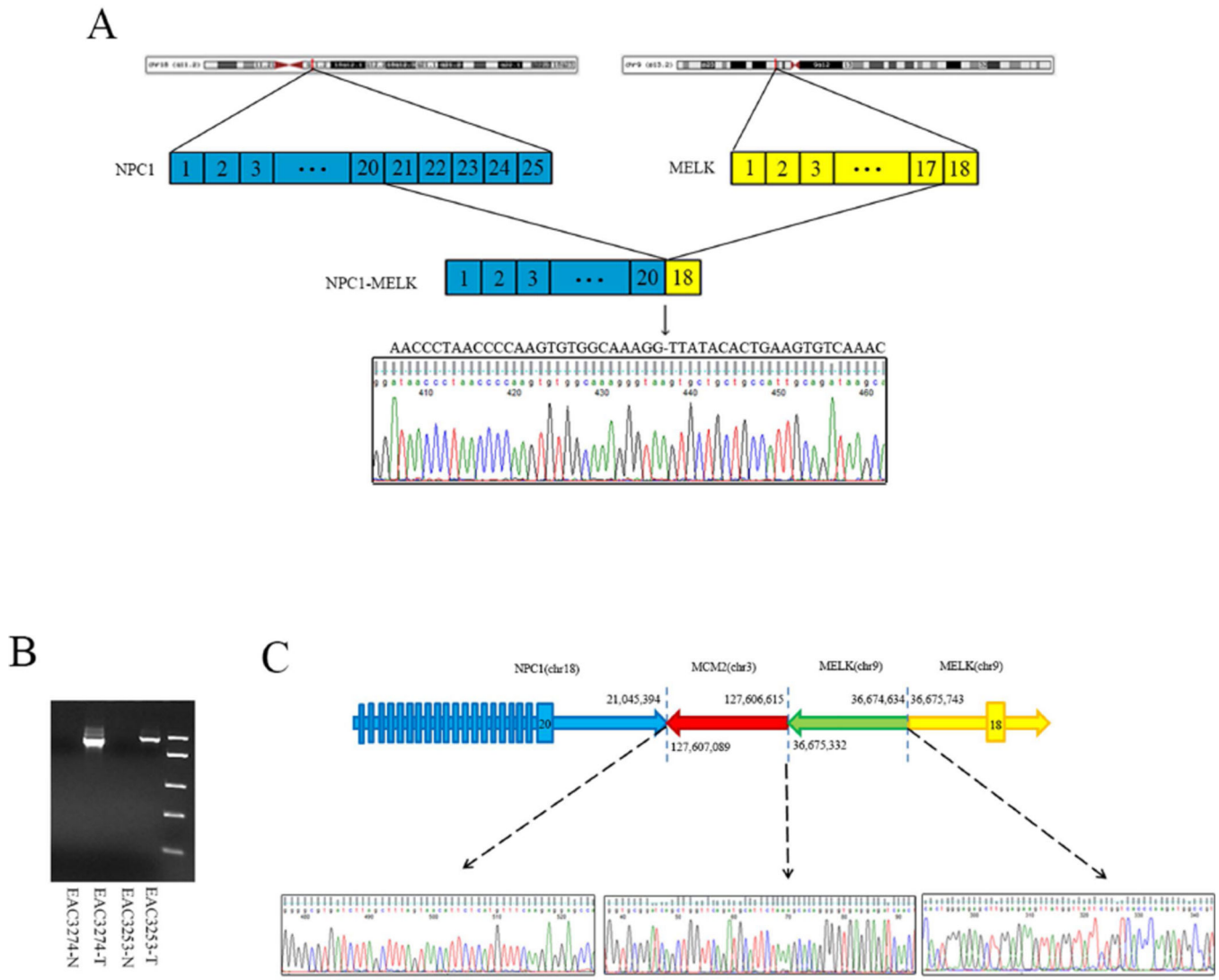

Figure 3. Genomic analysis identifies complex fusion events responsible for the generation of fusion genes in EAC

(A) Schematic representation of the NPC1-MELK transcript. NPC1 and MELK genes map to 18q11 and 9p13, respectively. Sanger sequencing confirms the chimeric junction between NPC1 and MELK. (B) Genomic DNA PCRs were performed on DNA extracted from EAC3253 and EAC3274 using a primer set whose forward primer was located in exon 20 of NPC1 and whose reverse primer was located in exon 18 of MELK. A band was produced only in tumor, but not in normal tissue. (C) Schematic representation of complex rearrangement events. Sanger sequencing confirms that the NPC1-MELK fusion gene is derived from complex inter- and intra-chromosomal rearrangements consisting of three fusion points involving chromosomes 18, 3 and 9, consistent with chromoplexy (the entire DNA sequence in the junction region is shown in Supplementary Fig. S1).

Table 1

Fusion transcripts identified by RNA-seq and validated by PCR

Fusion transcripts	5’ gene (chr)	3’ gene (chr)	5’ position	3’ position	Type	Validated
FGFR2-GAB2	FGFR2(chr10)	GAB2(chr11)	123,243,211	77,991,946	Inter-chromosomal	Yes
NPC1-MELK	NPC1(chr18)	MELK(chr9)	21,118,506	36,677,156	Inter-chromosomal	Yes
MKL1-FBLN1	MKL1(chr22)	FBLN1(chr22)	40,859,223	45,996,187	intra-chromosomal	Yes
USP54-CAMK2G	USP54(chr10)	CAMK2G(chr10)	75,289,438	75,620,647	intra-chromosomal	Yes
CNOT2-C12orf49	CNOT2(chr12)	C12orf49(chr12)	70,740,102	117,155,698	intra-chromosomal	Yes