Genome-wide Detection of Chimeric Transcripts in Early-stage Non-small Cell Lung Cancer

YAROSLAV ILNYTSKYY[1], LARS PETERSEN[2], JOHN B. MCINTYRE[2], MIE KONNO[3],
ADRIJANA D'SILVA[2], MICHELLE DEAN[2], ANIFAT ELEGBEDE[2], ANDREY GOLUBOV[1],
OLGA KOVALCHUK[1], IGOR KOVALCHUK[1] and GWYN BEBB[2]

[1]University of Lethbridge, Lethbridge, Alberta, Canada;
[2]University of Calgary, Calgary, Alberta, Canada;
[3]Alberta Health Services, Calgary, Alberta, Canada

Abstract. *Background/Aim: Lung cancer remains the main culprit in cancer-related mortality worldwide. Transcript fusions play a critical role in the initiation and progression of multiple cancers. Treatment approaches based on specific targeting of discovered driver events, such as mutations in EGFR, and fusions in NTRK, ROS1, and ALK genes led to profound improvements in clinical outcomes. The formation of chimeric proteins due to genomic rearrangements or at the post-transcriptional level is widespread and plays a critical role in tumor initiation and progression. Yet, the fusion landscape of lung cancer remains underexplored. Materials and Methods: We used the JAFFA pipeline to discover transcript fusions in early-stage non-small cell lung cancer (NSCLC). The set of detected fusions was further analyzed to identify recurrent events, genes with multiple partners and fusions with high predicted oncogenic potential. Finally, we used a generalized linear model (GLM) to establish statistical associations between fusion occurrences and clinicopathological variables. RNA sequencing was used to discover and characterize transcript fusions in 270 NSCLC samples selected from the Glans-Look specimen repository. The samples were obtained during the early stages of disease prior to the initiation of chemo- or radiotherapy. Results: We identified a set of 792 fusions where 751 were novel, and 33 were recurrent. Four of the 33 recurrent fusions were significantly associated with clinicopathological variables. Several of the fusion partners were represented by well-established oncogenes ERBB4, BRAF, FGFR2, and MET. Conclusion: The data presented in this study allow researchers to identify, select, and validate promising candidates for targeted clinical interventions.*

Correspondence to: Yaroslav Ilnytskyy, Ph.D., Department of Biological Sciences, University of Lethbridge, Alberta T1K3M4, Canada. Tel: +1 4033821526, e-mail: slava.ilyntskyy@alumni.uleth.ca

In the United States (USA) the risk of developing lung cancer (LC) over a lifetime stands at 6.7% for men and 6.0% for women. Despite the plateauing incidence trend, lung and bronchus cancer remain the leading cause of cancer-associated mortality with 135,720 estimated deaths in 2020 in the USA alone, exceeding the mortality of prostate, breast, brain, and colorectal cancers combined (1). Non-small cell lung cancer (NSCLC) comprises 85% of all LC cases further sub-typed into 3 major classes: adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC). Currently AC, is the most prevalent type comprising over 64% of NSCLC cases, originates from more distal airways and peripheral lung, whereas SCC making-up 25-30% arises from cells in the more central airway epithelium. The remaining cases belong to poorly differentiated and aggressive tumors assigned to the LCC type (2).

Clinical outcomes of NSCLC remain poor, for example, 5-year relative survival of patients with advanced disease stands at 6.1%. The survival is significantly improved reaching 61% in cases of localized NSCLC indicating the importance of early detection, relapse prognosis, and recurrence prevention (3). The aggressive nature of NSCLC and poor outcomes of non-targeted radio- and chemotherapies highlight the need for further research to uncover personalized targeted approaches that rely on the knowledge of molecular alterations behind the initiation and progression of NSCLC. Several driver mutations such as *EGFR* and translocations such as *ALK, ROS1, NTRK* have been identified and successfully targeted using small-molecule inhibitors (4, 5). Although not curative, this approach has led to profound improvements in outcomes to

the minority of patients whose tumors harbor these changes. Identifying new translocations that can be successfully targeted clinically remains a pivotal strategy in improving lung cancer patients experience (6).

Genome-wide studies of NSCLC have revealed a complex mutational landscape involving point-mutations, insertion, deletions, gene amplification, and larger structural rearrangements (7-9). The formation of chimeric proteins is a common way of oncogene activation across epithelial cancers (10).

Fusion transcript is a hybrid transcript formed by exons that belong to different genes. Mechanistically fusions may form both as a result of genomic rearrangements and at the post-transcriptional level *via* trans-splicing (11).

Fusion-mediated oncogenic activation is an established mechanism behind the initiation and progression of many NSCLC cases with a handful of tyrosine kinases (TK) identified as culprits. Initially, transforming *EML4:ALK* fusion was detected 6.7% of studied cases (12), later *RET* and *ROS1* fusions were added to the list of clinically relevant alterations (13, 14). More recent efforts in the analysis of lung cancer genomes and transcriptomes identified targetable fusions in *NTRK1/2/3* (15), *NRG1* (16), and *FGFR1/2/3* (17). Wider application of high-throughput methods revealed that fusion transcripts in epithelial cancer are common and are not limited to tyrosine kinases but involve a wide array of genes including transcription factors, metabolic genes, signal transducers, and chromatin modifiers (10).

For the most part, the detection of fusion transcripts in lung tumors relies on single-gene tests and targeted multiplexed panels (18). Both approaches require pre-existing knowledge of markers that need to be tested. The discovery of new targets requires unbiased genome-wide studies based on whole-genome or transcriptome sequencing. The majority of reports on the large-scale fusion detection that involves lung cancer is based on re-analysis of TCGA samples (19-21). Only a few studies were done using dedicated NSCLC fusion collections (22-24). The majority of fusions detected in these studies were not reported previously, showing the potential for novel findings.

Herein, we used RNA sequencing to conduct genome-wide discovery of transcript fusion events in a panel of 270 archived NSCLC tumor specimens. Our study focused on the characterization of transcript fusions in early-stage NSCLC prior to introduction of radio- or chemo-therapy. We also show that RNA sequencing can be used to detect transcript fusion in formalin-fixed, paraffin-embedded samples (FFPE) samples, despite the challenges with RNA quality. Furthermore, we analyzed recurrent fusions, promiscuous fusion partners, and predicted fusion oncogenic potential using computational methods. Finally, we characterized the associations between recurrent fusions, recurrent partner genes, and key clinicopathological covariates. We believe that this study may serve as a valuable resource for the selection of candidates with oncogenic and precision medicine potential.

Materials and Methods

Patient inclusion criteria and tumor samples. This was a retrospective study aimed at the characterization of transcript fusions in early stages non-small lung cancer (NSCLC). The study was approved by the Health Research Ethics Board of Alberta. All patients were diagnosed with NSCLC at the Tom Baker Cancer Centre in Calgary, Canada, between 2003 and 2010 (25). Clinical data were obtained retrospectively by chart review and abstracted into the Glans-Look Lung Cancer Database (26). The Glans-Look Lung Cancer Research (GLR) database is an Institutional database which captures patient-level demographics, clinical, treatment, response, and outcome data *via* chart reviews of electronic medical records for all patients with a diagnosis of lung cancer who present for diagnosis and treatment within the Canadian province of Alberta. Study data within the GLR Database are collected and managed using the Research Electronic Data Capture (REDCap) data capture tools hosted at the University of Calgary (27, 28).

The tumor samples were surgically collected taking utmost care to avoid non-cancerous tissue. The samples were gathered over the period of 10 years, they were fixed in formalin and preserved in paraffin blocks as a part of Glans-Look tumor repository. The series represents NSCLC at the early (I, II and III) stages of oncogenesis prior to the start of radio- and chemotherapy.

RNA isolation and sequencing library construction. Formalin-fixed paraffin-embedded (FFPE) NSCLC samples were excised from paraffin blocks and subjected to an RNA extraction process. RNA was isolated using the FFPE isolation protocol following the manufacturer's instructions. Briefly, the RNA samples were treated with DNAse I to remove the traces of genomic DNA contamination and quality tested using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The sequencing libraries were constructed from 100 ng of RNA using NEB Next Ultra II directional RNA library reagents kit with ribosomal depletion step (New England Biolabs, Ipswitch, MA, USA). The library preparation was performed following the manufacturer's instructions omitting the fragmentation step as recommended in the manual for the samples with a low RNA integrity index (RIN). The samples were multiplexed with NEBNext Multiplex oligonucleotides (New England Biolabs). The fragment libraries were sequenced using NextSeq500 (Illumina, San Diego, CA, USA) with a single-end, 75 cycle protocol. To minimize potential batch effects, all library preparation and sequencing steps were conducted by the same personnel and using the same set of equipment.

Quality control and fusion detection. Base-calling and demultiplexing steps were done with bcl2fastq v2.17.1.14 script from Illumina CASAVA 1.9 pipeline (Illumina). Various quality parameters of the sample sequencing libraries were assessed using FastQCv.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Base qualities were excellent across all samples, and no adapter contamination was noted. The reads were mapped to the human genome (Ensembl GRCh37) using HISAT v.2.2.1 (29) and raw read counts were obtained with featureCounts v.2.0.1 software (30). The reference genome was downloaded from Illumina iGenomes website (31). This step was conducted as a part of quality control to estimate alignment rates and gene expressions.

Fusion detection was conducted using JAFFA v. 1.09 pipeline in 'hybrid' mode that uses transcript assembly combined while incorporating the reads that do not map to either reference

transcriptome or the assembly. JAFFA developers recommended the 'hybrid' mode the single-end reads in the range of 60-99 bp (32). JAFFA was used with default GENCODE hg38 (UCSC version) transcript reference. The workings of the JAFFA pipeline are described in detail in the original publication and the software documentation https://github.com/Oshlack/JAFFA. Briefly, JAFFA pre-filters the reads to remove those mapping to intergenic, intronic, and mitochondrial regions. The remaining reads are assembled into contigs with Oases (33). The resulting contigs and reads are mapped to reference transcripts; those spanning multiple genes are selected as candidate fusions. Next, the pipeline counts the reads spanning breakpoints and aligns the contigs to the genome to determine break-point coordinates. The fusion candidates are then characterized and ranked based on confidence.

Fusions with genomic gap below 200 bp and no evidence of genomic re-arrangement are flagged as "potential regular transcripts" and excluded from default reporting. Fusions with breakpoints spanning exact exon-exon boundaries with overlapping reads and read-pairs are classified as "high confidence". The fusion receives "medium confidence" if only reads and no read pairs are present at the break-point. Since the present sequencing was single-end, the highest achievable rank was "medium confidence". Fusions with no breakpoints aligning to exon-exon boundaries were classified as "low confidence" and excluded from the analysis.

Filtering of fusion candidates. To reduce the number of potential false-positives we applied a set of filtering criteria to the dataset. We retained all of the fusions ranked as "medium confidence" with at least 2 reads spanning the break-point. Fusions involving adjacent (less than 50 kb gap) and overlapping genes were excluded as potential read-through events. We also removed fusions involving pseudogenes, immunoglobulins, and fusions formed between paralogues due to potential misalignments. Pseudogene and paralog data was obtained using biomaRt v2.48.1 (34). Finally, we excluded fusions that were observed in non-cancerous cells. The list of non-cancer fusion was obtained from FusionHub (20) and ChimerDB 4.0 databases (19).

Prediction of oncodriver potential. The oncogenic potential of discovered fusions was predicted using Oncofuse v.1.1.1 with epithelial training dataset and hg38 genome as a reference (21). The rest of the software options were kept at default. Oncofuse is a naïve Bayesian classifier that predicts the oncogenic potential of the experimental fusions based on genomic hallmarks of known fusions. Oncodriver score generated by the software is easily interpreted as *p*-value, further referred to as "driver probability". Bioinformatics workflow involved in the detection, filtering and prediction of transcript fusion is shown in Figure 1.

Statistical analysis. All of the statistical analyses were conducted in R v4.1.0. The associations between clinicopathologic covariates and recurrent fusions or partner genes were done using a multivariate logistic regression model implemented in glm() function with the "family" option set as "binomial". The model formula of constructed as follows: *Fusion/Gene ~ Sex + Smoking + Histology + Stage + Age + Recurrence*. Associations with the *p*-values below 0.05 were considered significant. The associations between samples fusion burden and clinicopathologic variables were conducted in the same way except the model family was "gaussian".

The associations with survival were analyzed with Cox proportional hazards model implemented as coxph() function from
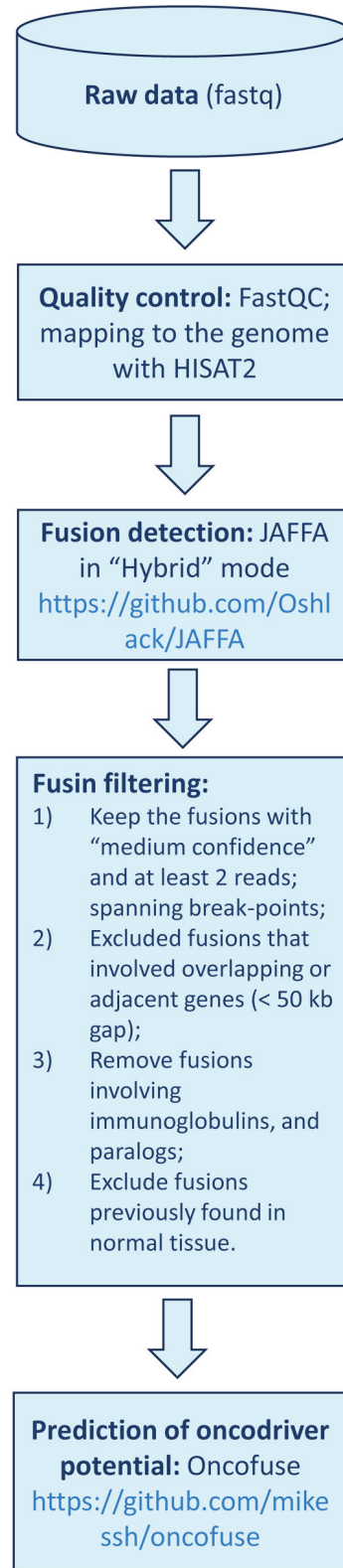

Figure 1. *The schematics of bioinformatics pipeline used to detect transcript fusions in early-stage non-small lung cancer. A full description of the workflow is available in the Materials and Methods section.*

Table I. *Clinical and demographic characteristics of the patient population.*

Clinical or demographic covariate	Summary statistics
Sex	
Female	143
Male	115
Age (years)	
Range	34.53-87.75
Median	65.42
Mean	65.01
Age (over/under 65)	
>65	135
<65	123
Smoking history	
Ever	196
Never	25
Missing	37
Histology (squamous/non-squamous)	
Non-squamous	187
Squamous	71
Histology (detailed)	
Adenocarcinoma	141
Adenosquamous carcinoma	13
Large cell carcinoma	9
Large cell neuroendocrine	5
Mixed NSCLC (synchronous)	3
Mixed NSCLC (within tumor)	1
Other NSCLC	15
Squamous cell carcinoma	71
Stage	
I	161
II	71
III	26
Recurrence	
Recurred	155
No recurrence	103

Figure 2. *Recurrent fusions observed in at least 3 samples in NSCLC RNA sequencing dataset. The numbers to the right of the bars indicate the number of samples where the fusion was observed.*

survival v3.2-11 R package. The model incorporated the same clinical and demographic variables as in GLM analysis. In addition, Kaplan-Meyers curves were drawn for fusions and gene partners with Cox model *p*-values below 0.05, and the survival of fusion-positive and negative cases was compared with the log-rank test. This analysis was conducted using survminer v0.4.9 package.

Results

Clinical and demographic characteristics. The lung tumor samples were selected from the Glans-Look archive of clinical specimens. The non-small lung cancer (NSCLC) samples used in the study were surgically resected at early stages (I-III) of the disease and prior to the start of chemo- or radiotherapy. The sample selection process was designed to generate a data set with a balanced representation of clinical and demographic covariates. The demographic characteristics considered in the analysis were age at diagnosis, sex, smoking history, and
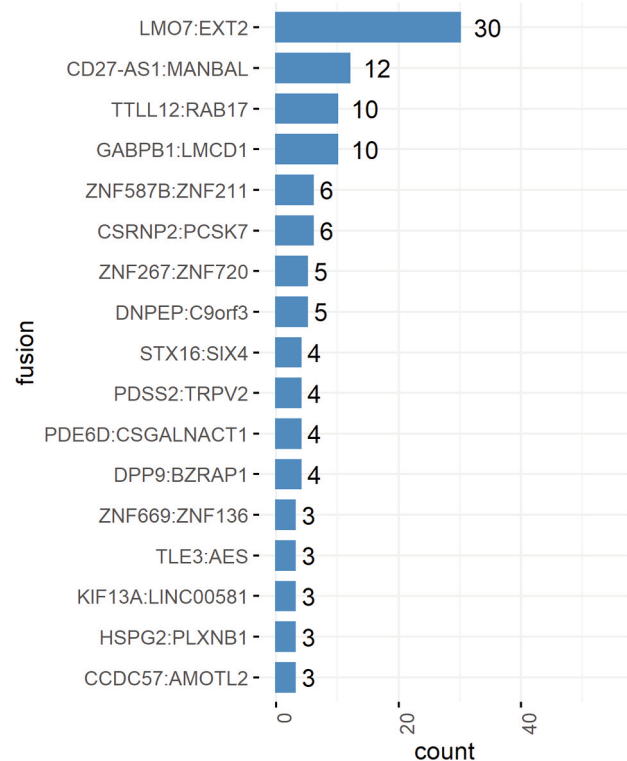
survival from the date of diagnosis until the event of death. The "age" variable was divided into 2 intervals - under and over 65 years. Clinical variables were the stage of tumor development (I, II, III) and the histological type – either squamous or non-squamous. Stage and the histological type were determined using histopathological methods by qualified pathologists. The summary statistics for demographic and clinical variables are available in Table I.

Summary of sequencing. We used high-throughput RNA-sequencing to gain insight into the transcript fusion landscape of early-stage NSCLC. In total, we sequenced 270 FFPE tumor specimens that represented individual patients. Initial quality control showed that the base qualities were good across all samples, minor adapter contamination was successfully removed. The sizes of the sequencing libraries ranged from 7 to 123 million reads with a median of 33. The reads were mapped to GRCh37 (Ensembl) human genome assembly resulting in the alignment rate in the range of 24-95% with a median of 88%. The sequencing detected 25,533 expressed features with at least 10 reads in 5 samples. The majority of 64.66% of detected features were protein-coding genes, 17.2% were long non-coding RNAs, 13.62% represented pseudogenes

Table II. *The list of 31 fusions that overlap between the current study and the ChimerDB 4.0 fusion data base.*

Fusion	Number of patients	Cancer type	5' gene role	3' gene role	Highly reliable sequence
ATP2A2:IFT81	2	LUAD	Unknown	Unknown	Seq+
NOC4L:FBRSL1	2	OV	Unknown	Unknown	
AMBRA1:CKAP5	1	BLCA	Unknown	Unknown	Seq+
ARMC7:HN1	1	LUSC	Unknown	Unknown	Seq+
ASPH:TACC1	1	BLCA	Unknown	Unknown	Seq+
ATG5:PDSS2	1	GBM	Unknown	TSG	Seq+
ATP11A:CUL4A	1	LUSC	Unknown	Unknown	Seq+
ATP1B1:NRG1	1	PAAD	Unknown	Unknown	Seq+
B4GALT1:BAG1	1	CESC	Unknown	Unknown	Seq+
BCAS3:VMP1	1	BRCA	Unknown	Unknown	Seq+
CD74:ROS1	1	LUAD	Unknown	Oncogene	Seq+
EIF3B:MAD1L1	1	BRCA	Unknown	TSG	
GCN1L1:PLA2G1B	1	LIHC	Unknown	Unknown	
GSK3B:LSAMP	1	PRAD	Kinase; TSG	TSG	Seq+
HIF1A:PRKCH	1	LUAD	Oncogene; TSG; TF	Unknown	Seq+
KHDRBS3:ANXA13	1	SKCM	Unknown	Unknown	Seq+
KMT2C:XRCC2	1	SKCM	Oncogene; TSG	Oncogene	
LINC00393:KLF12	1	HNSC	Unknown	TF	
LSR:GRAMD1A	1	LIHC	Unknown	Unknown	Seq+
MAP3K1:GPBP1	1	OV	Kinase; Oncogene	TF	
MET:CAV1	1	BRCA	Kinase; Oncogene	TSG	Seq+
MIRLET7BHG:ATXN10	1	LUAD	Unknown	Unknown	
PMS1:HIBCH	1	HNSC	Oncogene	Unknown	
PRDM2:PEX14	1	THCA	TSG; TF	Unknown	
PRIM2:ZNF451	1	HNSC	Unknown	TF	
RAPGEF5:STEAP1B	1	BRCA	Unknown	Unknown	Seq+
SNX8:CHST12	1	UCS	Unknown	Unknown	Seq+
SUV420H1:C11orf24	1	UCEC	Unknown	Unknown	Seq+
TBC1D22A:GRAMD4	1	BRCA	Unknown	Unknown	
UVRAG:C11orf30	1	BRCA	TSG	Unknown	Seq+
ZMPSTE24:CAP1	1	OV	Unknown	Unknown	Seq+

The annotations (cancer type, role in cancer and sequence reliability) were obtained from the ChimerDB. Cancer type corresponds to TCGA study abbreviations available at this link: https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations. The potential role of protein product in cancer is described as kinase, oncogene, tumor suppressor gene (TSG), or transcription factor (TF). Seq+ descriptor identifies a subset of highly reliable fusions (ChimerSeq-Plus) predicted by multiple programs in ChimerDB 4.0 study.

and other RNAs, and the biotype of the remaining 4.52% was not assigned (Supplementary Figure 1).

Overview of transcript fusion landscape in NSCLC. The results of the initial JAFFA analysis were filtered to exclude fusions with less than 2 reads mapping to the break-point, fusions between mitochondrial and autosomal genes, those involving pseudogenes, fusions originating from adjacent or overlapping genes, and those where both partner genes were paralogues. Following filtering, the fusion detection resulted in 896 transcript fusion events across 270 NSCLC tumor samples. This data-set could be reduced to 782 unique fusions after removing redundant events with the same fusion partners. The number of fusions per sample ranged from 0 to 21, with mean and median values of 3.0 and 3.32 respectively. The full list of fusions retained after filtering is available in supplementary material (Supplementary Table I).

A fraction of transcript fusions equal to 4.22% or 33 out of 782 occurred more than once in any of the samples and were deemed recurrent. The rate of the highest recurrence was 30 or 11.1% of samples for *LMO7:EXT2* fusion, other highly recurrent fusions were *CD27-AS1:MANBAL* (4.4%), and *TTLL12:RAB17* (4.4%). The top 17 fusions that occurred in over 3 samples are shown in Figure 2. The full table with fusion occurrence data can be found in Supplementary Table II.

The majority of fusions, 480 (61.3%) originated from inter-chromosomal rearrangements while in 302 (38.6%) cases both partners were located on the same chromosome. Less than half of 37.6% of these fusions were in-frame and expected to produce viable protein while 62.4% were predicted to generate out-of-frame transcripts (Supplementary Figure 2).
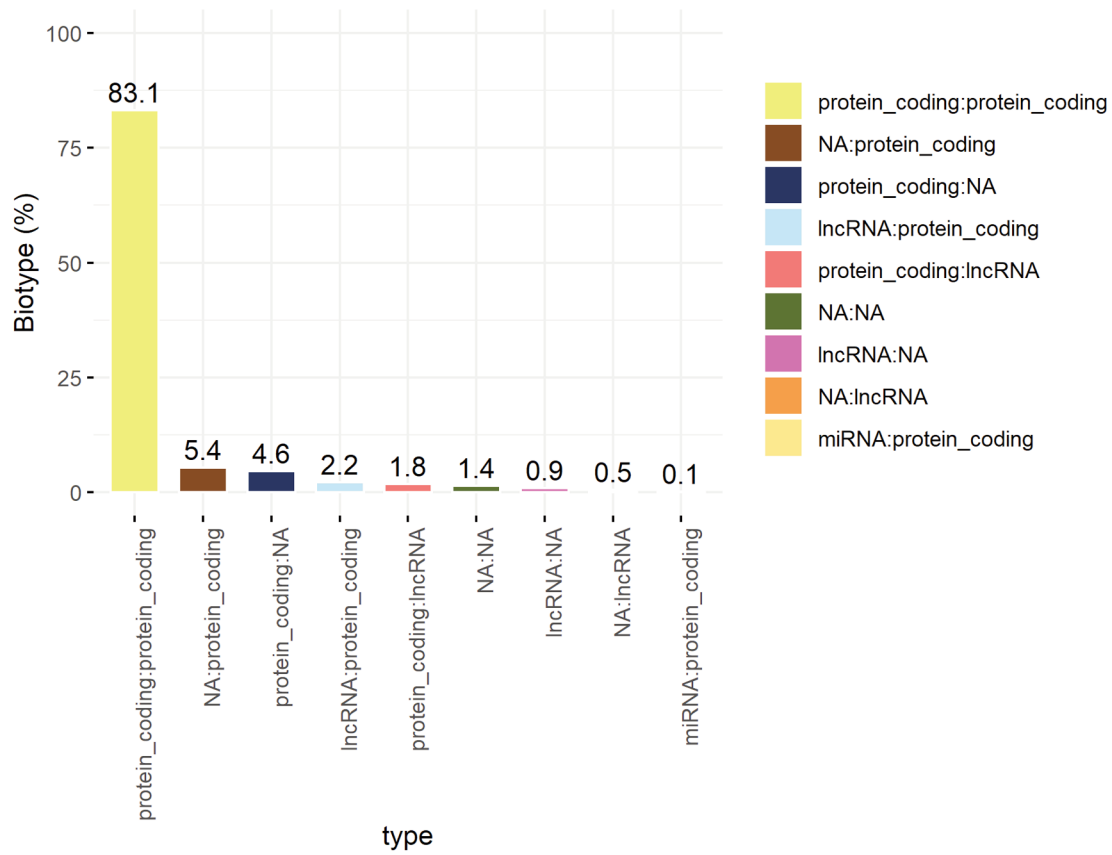
Figure 3. *Classification of NSCLC fusions by biological type retrieved from the biomart database using biomaRt Biocondustor package. "NAs" indicate that the biotype could not be identified.*

Analysis of the distribution of fusions across chromosomes showed significantly higher fusion frequencies in chromosome 19. This increase was observed both for inter- and intra-chromosomal fusions (Supplementary Figure 3A and B). This data suggests that chromosome 19 serves as a genomic instability hotspot concerning both translocations and deletions resulting in inter- and intra-chromosomal chimeric genes.

BiomaRt Bioconductor package was used to annotate fusion partners with biotype: protein-coding, long non-coding RNA (lncRNA), microRNA, constant and variable T-cell receptor genes (*TR C* and *TR V*), and unknown. The absolute majority or 83.1% of all fusions contained protein-coding genes as both partners, a much smaller fraction (10%) had protein-coding partner fused with a gene with "unknown" biotype, and 4% consisted of protein-coding gene joined to lncRNA, other categories were less numerous (Figure 3).

Fusions detected in this study were searched against ChimerDB v.4.0, a large fusion database that incorporates 67,610 fusions sourced from the prediction based on deep sequencing, text mining of PubMed articles, and derived from other databases (19). Only 31 fusions found in our study were also present in ChimerDB 4.0, while 751 or 96% were absent and considered novel. Known fusions were observed across 15 cancer types, with the exception of glioblastoma multiforme and melanoma, all were carcinomas. Four fusions (*ATP2A2:IFT81*, *CD74:ROS1*, *HIF1A:PRKCH*, *MIRLET7BHG:ATXN10*) were previously observed in lung adenocarcinomas and 2 (*ARMC7:HN1*, *ATP11A:CUL4A*) in lung squamous cell carcinomas. Genes classified as oncogenes in the ChimerDB database were observed in 6 out of 31 known fusions and include *ROS1*, *HIF1A*, *PMS1*, *MET*, *MAP3K1*, *XRCC2*, and *KMT2C* (Table II). The prevalence of novel fusions (751 or 96%) indicates high diversity and widespread presence of these aberrations combined with the relative lack of relevant data mining efforts.

Recurrent transcript fusions in NSCLC. Research suggests that gene fusions observed in multiple cases either within or across multiple cancer types are functional regarding the induction or progression of the malignant process. To identify such candidates, we analyzed fusions that occurred in at least 2 tumor samples in our dataset. In total, we found that 33 of 782
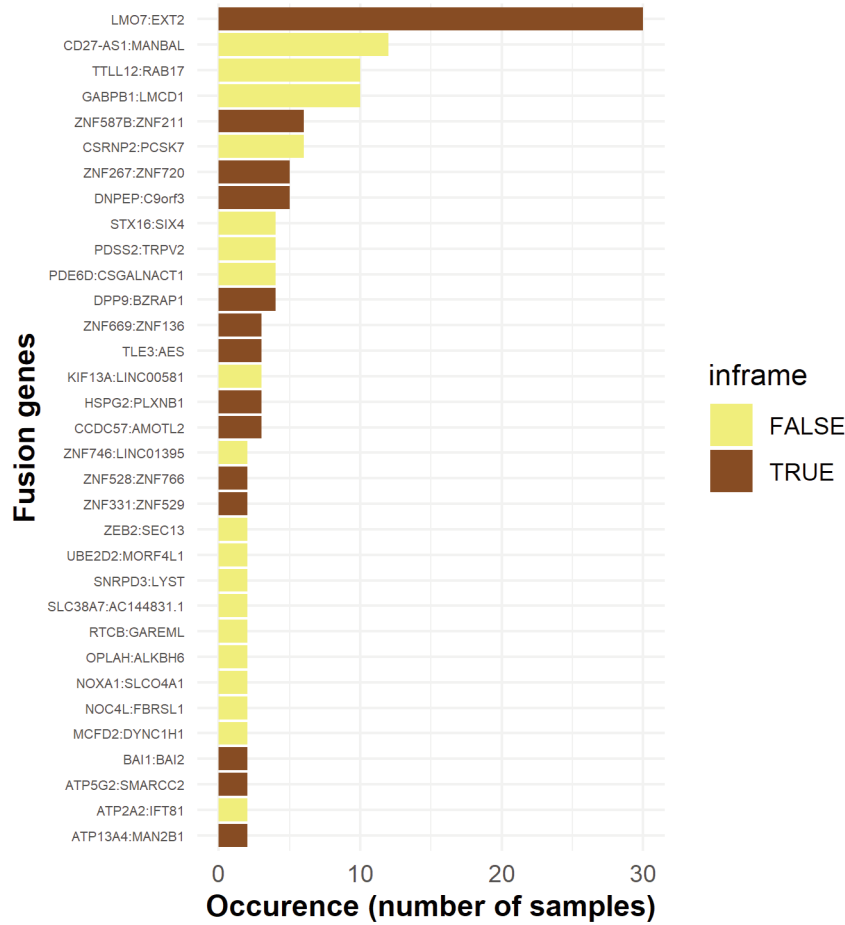
Figure 4. *The graph shows all of the recurrent fusions observed in the sequencing data set. The bar heights correspond to the number of separate NSCLC tumor samples where a gene fusion was observed. The total number of samples in the dataset was 270. In-frame fusions are shown in dark brown and out-of-frame in yellow colors.*

(4.22%) fusions were recurrent. Four recurrent fusions were observed in at least 10 samples (Figure 4). Out of 33 recurrent fusions 19 or 57.6% were out-of-frame, and 14 were in-frame (Supplementary Table III). Certain fusions had unusually high frequency, for example, *LMO7:EXT2* had the highest recurrence and was found in 30 (11.11%) samples, *CD27-AS1:MANBAL* was detected in 12 (4.44%) samples, while *GABPB1:LMCD1* and *TTLL12:RAB17* each appeared in 10 (3.7%) samples (Supplementary Table III). These fusions were absent from ChimerDB and the existing literature. It is unclear whether these are false-positives originating from misalignments or library preparation artifacts, however, all of the fusions included in the analysis had breakpoints at the exact exon-exon boundaries, which we find unlikely in the case of artifacts. It is possible that highly recurring fusions originate from trans-splicing events observed in normal non-cancerous tissue (35).

Only 2 of 33 recurrent fusions from our data set were also found in ChimerDB: *ATP2A2:IFT81* and *NOC4L:FBRSL1*.

Both occurred in 2 patients and were included in ChimerSeq-Plus subset of highly reliable fusions. *NOC4L:FBRSL1* was associated with ovarian serous cystadenocarcinoma (OV), and also detected in B-cell acute lymphoblastic leukemia (36). In the ChimerSeq data set, *ATP2A2:IFT81* was found in lung adenocarcinoma, uterine corpus endometrial, and liver hepatocellular carcinomas. The presence of this fusion in lung adenocarcinoma was also evidenced in a recent survey of gene fusions across multiple cancer types (37).

The majority of recurrent fusions in our data set were novel. In order to gain further insight into their potential role, we cross-referenced gene-fusion partners with Cancer Gene Census (CGC) and the list of oncodrivers obtained from the literature (38). Also, literature searches were conducted to gather additional evidence and insight into the role of fusion partners in lung cancer and in malignant processes in general (Supplementary Table IV). Two genes, *EXT2* in *LMO7:EXT2* fusion and *ZNF669:ZNF136* were previously characterized as

Table III. *Recurrent fusions with the highest oncogenic potential (Oncofuse driver probability over 0.8).*

Fusion	Driver probability (Oncofuse)	Number of samples (out 270)	Comment
ZEB2:SEC13	1	2	Zeb2 is epithelial-to-mesenchymal transition factor up-regulated in many cancers including lung. Zeb2 drives invasive and microbiota-dependent carcinoma in mouse model (84).
ZNF587B:ZNF211	0.97	6	ZNF587B suppresses cell proliferation and metastasis in ovarian cancer (49).
ZNF528:ZNF766	0.91	2	No evidence of cancer involvement.
ZNF669:ZNF136	0.86	3	No evidence of cancer involvement.
HSPG2:PLXNB1	0.85	3	CD100-PlexinB1 axis induces epithelial-to-mesechmal transition and promotes metastasis in head and neck squamous cell carcinoma (48).
GABPB1:LMCD1	0.84	10	LMCD1 mutations promote cell migration and metastasis in hepatocellular carcinoma (51). Increased expression of GABPB1 is associated with poor survival outcome in renal carcinoma (50).
STX16:SIX4	0.81	4	SIX4 homeobox gene is highly expressed in NSCLC tissue and its exogenous expression greatly enhances proliferation, migration, and invasion of these cells.

TSGs according to CGC. Twenty-five of the recurrent fusions had gene partners with clear involvement in a malignant process described in the literature. Only 6 fusions had partners with no known involvement in cancer.

Recurrence of fusion partners and gene promiscuity. Genes participating in fusions are frequently promiscuous in terms of partners and form fusions across a range of cancer histologies (39). Such genes play important roles in cancer initiation and progression as either oncogenes or tumor suppressor genes (TSGs). Their disruption by genomic rearrangement results in a formation of a chimeric protein that benefits cancerous cells and survives selective pressure regardless of the second partner. Participation of a certain gene in multiple fusions indicates its importance in the malignant process and offers potential for clinical intervention. To identify promiscuous fusion genes in our data set, we analyzed the recurrence of individual fusion partners and cross-referenced them with the lists of known oncogenes and TSGs, while focusing on fusion genes with multiple partners. In total, we identified 1,363 unique genes as fusion participants with 229 observed in more than 1 sample (recurrent). Matching the list of fusion partners to the collection of known cancer-related genes revealed 51 oncogenes and 48 TSGs. Of these 23 had dual roles and could serve as both drivers and suppressors. Out of recurrent fusion partners, 13 were oncogenes while 7 played the role of tumor suppressors (Supplementary Table V). The highest degree of recurrence (31/270 samples) was observed in the case of *EXT2* tumor suppressor gene (40) that, with one exception, had *LMO7* as an exclusive partner. Other known cancer genes had much lower recurrence; *TBL1XR1* and *ZBTB20* were observed

in 3 patients, each forming chimeric transcripts with different partners. *TBL1XR1* is known for both oncogenic and tumor suppressor roles. In NSCLC, *TBLXR1* mediates cell proliferation, survival, and metastasis by regulating *MEK* and *AKT* pathway through c-Met (41). *ZBTB20* was shown to promote cell proliferation in NSCLC through transcriptional repression of *FOXO1* tumor suppressor (42). Other notable oncogenes included serine/threonine kinase *BRAF*, fibroblast growth factor receptor family member *FGFR2*, and tyrosine kinase receptor *MET*. These genes were observed in fusions across 2 samples each.

In total, 179 genes formed transcript fusions with multiple partners, 24 genes took part in at least 3 different fusion configurations (Supplementary Table IV). PubMed searches revealed cancer involvement for 18 of these genes with 8 participating in the development of lung cancer. The highest degree of promiscuity was found in the case of *ZBTB7A* and *RNF213* genes that each formed fusions with 4 different partners. None of these genes was present in the reference list of oncodrivers or TSGs, however, a literature search revealed their involvement in the development of NSCLC. *ZBTB7A* is a transcriptional suppressor of glycolysis down-regulated in many cancer types including LC. Tumors deficient for this gene progress fast and display heightened sensitivity to glycolysis inhibition (43). *RNF213* mutations discovered in circulating DNA discriminate between early-stage lung cancer and benign disease pulmonary nodules (44).

Prediction of oncogenic potential with Oncofuse. The oncogenic potential of detected fusions was predicted using a naive Bayesian classifier implemented in Oncofuse software (21). In a total, 657 fusions were available for the prediction

of oncogenic potential expressed as driver probability. The full list of oncodriver prediction results is available in Supplementary Table VI. There were 88 fusions with driver probability over 0.95, eight of them were formed with the participation of known oncogenes *ERBB4*, *SMAD2*, *BRAF*, *PRKAR1A*, *FGFR2*, *KMT2C*, *ZNF236*, and *MET*.

Six fusions with high oncodriver probability (>0.8) were recurrent (Table III). *ZEB2:SEC13* was the recurrent fusion with the highest driver probability of 1.0. This fusion was detected in 2 out 270 (0.74%) of samples, SCC and AC. *ZEB2* is an E-box homeobox binding transcription factor that regulates epithelial-to-mesenchymal transition (EMT) and is subject to recurrent translocations in T-cell acute lymphoblastic leukemia (45). MicroRNA mediated inhibition of *ZEB2* suppresses migration and invasion in NSCLC (46, 47). Another EMT-inducing factor PlexinB1 was observed as a part of *HSPG2:PLXNB1* transcript fusion present in 3 SCC samples. Activation of PlexinB1 by *CD100* promotes EMT and metastasis in head and neck SCC (48). Other recurrent fusions with high driver probability involved zink finger proteins with putative DNA binding and gene regulation functions. For example, *ZNF587B:ZNF211* had a driver probability 0.97 and was found in 4 AC samples, 1 LCC and 1 SCC. *ZNF587B* is a *C2H2*-type zinc finger protein (ZFP) family protein that was shown to inhibit proliferation, colony formation, migration, and invasion in ovarian cancer (49).

GABPB1:LMCD1 fusion with driver probability 0.84 was observed in 10 or 3.70% of the samples, both squamous and non-squamous. Down-regulation of long non-coding RNA *GABPB1-IT1* expressed from the intronic region of *GABPB1* is associated with poor prognosis in NSCLC (50). Somatic mutations in a 3' partner *LMCD1* promoted migration and metastasis in hepatocellular carcinoma (HCC) and systemic lung metastasis in murine model (51).

Associations of transcript fusions with clinicopathologic characteristics and survival analysis. Generalized linear models (GLM) method was used to investigate the association of recurrent fusions and promiscuous fusion partners with known clinicopathologic characteristics: sex (M/F), smoking status (never/ever), age (over/under 65), histology, recurrence, stage (I, II and III). The survival analysis included the same set of clinical and demographic covariates. Considering that transcript fusions may arise from genomic rearrangements, the number of fusions within the tumor serves as a surrogate of genomic stability. Our GLM analysis failed to detect any significant association between a number of fusions per sample and clinicopathological characteristics; similarly, no such association was found with patient survival (data not shown).

Four of the recurrent fusions were significantly associated with clinicopathologic covariates (GLM *p*-value <0.05) (Supplementary Table VII). *CD27-AS1:MANBAL* was associated
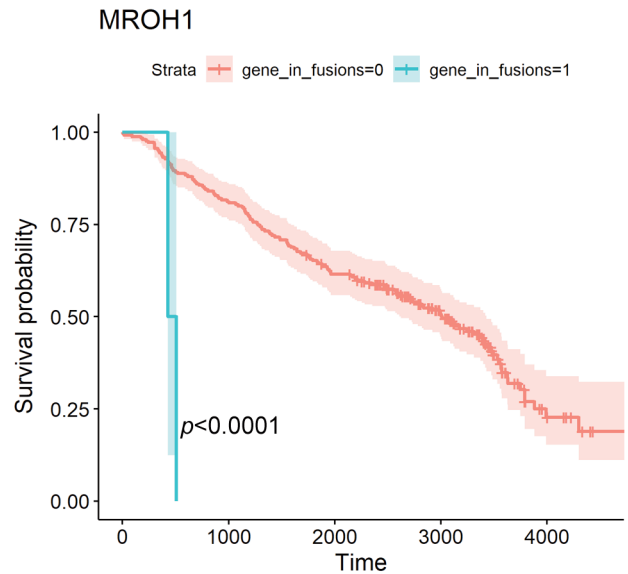

MROH1

Figure 5. *Kaplan-Meier survival analysis of patients that were either positive or negative for MROH1 transcript fusions. The sample size was 2 MROH1 fusion positive (blue) and 250 negative cases (red). The survival time (X-axis) was expressed as days. Kaplan-Meier survival analysis showed decreased life expectancy of MROH1 fusion positive cases (HR=14.04; p=0.0004; 95% CI=3.22-60.94).*

with a female gender (*p*=0.03); *DNPEP:C9ORF3* was predominantly found in cases with no recurrence (*p*=0.04). The remaining fusions were associated with tumor histological types. *GABPB1:LMCD1* was found in 1 out 5 (20%) of large cell neuroendocrine type while only 8 out 253 (3.16%) were positive in the remaining cases (*p*=0.03). *LMO7:EXT2* fusion was prevalent in adenosquamous carcinoma with 4 out 13 (30.76%) cases as opposed to 8.62% in the rest of the samples (*p*=0.03). The same method was applied to analyze potential relationships between recurrent promiscuous genes and clinical characteristics did not reveal any significant results.

Integrative multivariate analysis failed to detect any recurrent transcript fusions associated with survival when the model included relevant clinical and demographic characteristics. Among clinical covariates, the later stage at diagnosis (II) and recurrence strongly predicted poor survival (*p*<0.001). Improved survival was expected in never smokers (*p*=0.01), while large cell histological type was associated with poor outcome (*p=0.03*) (Supplementary Table VIII). Three recurrent genes participating in multiple fusions *MROH1*, *N4BP2L2* and *TRAPPC9* were significantly associated with survival. However only in the case *MROH1* the association was highly significant in both Cox regression (HR=14.04, *p*<0.001) and log-rank test (Figure 5; Supplementary Table VIII). Recent research suggests the involvement of *MROH1* in the late actin-dependent stage of

exocytosis (52); also, a recent study detected copy number variation of *MROH1* in drug-resistant ovarian cancer (OC), while duplication of MROH1 gene was associated with decreased survival in these patients (53). In addition, MROH1 was identified as one of the top genes that undergo CNV in TCGA dataset (53).

Discussion

In this study, we demonstrated the utility of RNA-sequencing approaches applied to archived tumor specimens in the identification of transcript fusions with a potential contribution to the development of NSCLC. Our work involved a large collection (270 samples) of early-stage FFPE tumor samples that were surgically resected prior to the start of radio- or chemo-therapy. This allowed us to observe tumors in their native state without the selective pressure and added genomic instability induced by therapeutic intervention. We found a diverse fusion landscape in early-stage NSCLC with chimeric transcripts present in the majority of samples. Fusion candidates were filtered to reduced false-positives and compared to existing fusion datasets. The absolute majority (751 or 96%) of the fusion candidates were novel showing that there is a plethora of chimeric transcripts yet to be fully described. We provide the research community with plenty of supporting data including full list fusion candidates, the sequences of chimeric transcripts, prediction of oncodriver potential, data regarding recurrent fusions, genes with multiple partners, evidence of their roles in cancer, and associations with clinicopathologic variables. This data is available in supplementary materials and is meant to generate research leads resulting in the identification and validation of clinically relevant fusion events.

Currently, the majority of NGS applications are focused on the detection of fusions in lung cancer specimens using either amplicon-based or hybrid-capture enrichment approaches. These methods rely on custom and commercial panels with known targets (18). NGS is also an excellent method for the discovery of novel fusions with clinical significance. Normally researchers mine the collection of TCGA datasets to discover and characterize fusions across various cancers (54-56). Considering a limited number of lung cancer cases in TCGA and generally low frequency of fusions, there is a need to expand fusion detection to other case studies. A recent study of 54 Taiwanese patients with NSCLC was conducted on cryo-preserved specimens and resulted in the discovery 218 fusions where only 24 were previously reported (22). Earlier, Rudin *et al*. applied NGS to characterize genomic characteristics including gene fusions in small cell lung cancer (SCLC) (24). This study used RNA-seq to detect transcript fusions in 55 samples composed of primary tumors, adjacent normal tissue, and cancer cell lines. In total, 44 transcript fusions were detected of which 4 were recurrent. Another large study applied RNA-seq to detect 45 fusions across 87 AC cases, 8 of them involved in transformation of tyrosine kinases (57). Other similar studies involved a much smaller number of cases (23, 58).

Compared to earlier reports, our project involved a larger panel of 270 cases representing diverse histological NSCLC types. Our samples were selected to represent early stages of malignant development, mostly stages I and II with the lesser representation of stage III. The tumors were resected before chemo- and radiation therapy to take the snapshot of the cancer genome in its native state. Importantly, we successfully conducted fusion detection using RNA extracted from FFPE samples collected and stored over a long time period, showing the feasibility of large retrospective fusion studies with archived tumor material. In total, we detected 782 unique chimeric transcripts with an average occurrence rate of 3.32 per sample in line with the fusion frequency described in a recent Taiwanese study (22). Also, similar to Chang *et al*., the absolute majority of the fusions, or 96% were not reported previously. Despite the similar nature of the samples, there was no overlap between fusions detected in Chang *et al*. (2021) and the current study. Comparing our results with 48 adenocarcinoma fusions from Seo *et al*. (2012) report found a single overlap - *CD74:ROS1* fusion.

We observed a significantly higher frequency of transcript fusion events where either one or both partners were located on chromosome 19. This chromosome is known to have some unique properties such us increased gene density, unusual richness in clustered gene families, CpG islands, and repetitive elements (59). The repeats located on chromosome 19 are mainly comprised of Alu and LINE elements (59). A 32-38 bp long MSR1 minisatellite tandem repeat concentrated in 19q13 can contribute to genomic instability, influence gene expression by changes in copy number variation and increase a risk of cancer (60). Interestingly, aberrations in chromosome 19 are associated with lung asbestos-induced lung cancer (61).

Chromosome-Centric Human Proteome Project (C-HPP) determined that chromosome 19 aberrations are present in various lung cancer subtypes and lead to loss of tumor suppressor genes, changes in DNA repair genes, *TGFB* pathway genes, and formation of fusion oncogene *BRD4-NUT* (62). At least 2 chromosome 19 genes detected in *C-HPP* project (*LTBP4* and *PPP1R13L*) were identified as fusion partners in our study. It appears that chromosome 19 is prone to forming aberrations in lung cancer, however published evidence is still very limited.

A handful of genes contain mutations with a well-established oncodriver role in NSCLC (63). Of those, *ROS1*, *FGFR2* and *BRAF* formed fusions in 2 samples, while *NRG1* and *MET* fusions were single occurrences. Only one of these fusions *CD74:ROS1* was previously described (64), all others recombined with novel fusion partners. Erythroblastic oncogene 2, also known as *HER2*, is up-regulated in 1-4%

cases of adenocarcinoma and is a well-established biomarker in a variety of other cancers (65). Although we did not find any fusions harboring *HER2/ERBB2* oncogene, 2 fusions involving related genes *ERBB2IP* (*ERBB2IP:FAM156B*) and *ERBB4* (*PSME4:ERBB4*) were observed. *ERBB2IP* is a scaffold protein that binds to *ERBB2* with high specificity and stabilizes it at the plasma membrane (66). The IntOGen web platform (intogen.org) identified *ERBB2IP* as a mutational cancer driver in lung squamous cell carcinoma and pan-cancer analysis (67). *ERBB4* is a receptor tyrosine kinase (*RTK*) known to carry activating mutations in NSCLC (68) and forming fusions with multiple partners; a novel *KIF5B:ERBB4* fusion was recently identified in AC patients (69). Other notable *RTK* oncogenes observed as fusion partners included *MET*, *IGF1R*, *FYN*, and *SRC*. *MET* fusions were found in 2 samples while the rest were non-recurrent. Of these only 2 fusions were described previously, specifically, *FYN:REV3L* was found in breast cancer (70) and *CAV1:MET* was detected in SCLC case with the positive response to Crizotinib and Osimertinib (71). The rest of the *RTK* fusions were novel.

We hypothesized that genes that recombine with multiple partners play a role in carcinogenesis. This notion is based on the assumption that a disruption of a single oncogene or a TSG is required to facilitate the malignant process while the second fusion partner may vary due to the randomness of genomic rearrangement. To detect such genes, we analyzed promiscuity in fusion formation defined as a number of different partners that fuse with a gene of interest. Literature searches showed that most of the promiscuous genes were involved in either lung or other types of cancer. Two genes with the highest promiscuity, *RNF213*, and *ZBTB7A*, formed fusions with 4 fusion partners. All fusions that involved *RNF213* and *ZBTB7A* had very high predicted driver probability (detected by Oncofuse) with a mean of 0.92 and a range of 0.78-0.99. *RNF213* fusion was recently reported and validated in chronic myeloid leukemia (72); furthermore, knockdown of *RNF213* reverses hypoxia-induced death and promotes tumorigenicity (73). The mutational status of *RNF213* was shown to differentiate early lung cancer from benign pulmonary modules (44). *ZBTB7A* is a *POZ/BTB* and Kruppel (*POK*) family transcription factor that acts as an oncogene or a TSG depending on cellular context. Originally described as an oncogene in hematological cancers, *ZBTB7A* role was expanded to solid cancers including NSCLC where this gene is frequently amplified and over-expressed (74, 75).

Naive Bayes classifier implemented in Oncofuse software was used to predict oncodriver potential of detected fusion candidates. Seven recurrent fusions had high oncogenic potential (Table III). Two of them, *ZEB2:SEC13* and *HSPG2:PLXNB1* involved key regulators of epithelial-to-mesenchymal transition. *ZEB2:SEC13* had a driver probability of 1 and was observed in 2/270 or 0.74% of samples, *HSPG2:PLXNB1* had the same occurrence. Also, *ZEB2* formed a third *ZEB2:SRGAP* rearrangement with low predicted driver potential. Novel *ZEB2* fusions were recently identified by RNA sequencing in AML (76) and lipoblastoma (77). MicroRNA-mediated targeting of *ZEB2* blocks *EMT* and *WNT*/Beta-catenin pathway (78); furthermore, *ZEB2* cooperates with *PAX6* to promote EMT, metastasis, and cisplatin resistance in NSCLC (79). Little is known about the role of *PLXNB1* in lung cancer, however, it is an established cancer gene in other malignancies. *PLXNB1* serves as a receptor of semaphorins in a pathway that regulates cellular shape and motility (80). Mutations of *PLXNB1* drive invasion and metastasis in prostate cancer (81); in breast cancer, *PLXNB1* directly interacts with *ERBB2* to promote metastasis (82). Other recurrent novel fusion candidate *GABPB1:LMCD1* had driver probability 0.84 and was found in 2.7% of samples. Recently, *LMCD1* was identified as a transcriptional factor regulating EMT in NSCLC and breast cancer cell lines (83); somatic mutations of this gene promoted migration and metastasis in hepatocellular carcinoma (51). Overall, most of the recurrent fusions with driver potential have known involvement in cancer by either promoting EMT and metastasis (*ZEB2*, *PLXNB1*, *LMCD1*) or acting as tumor suppressors (*ZNF587B:ZNF211*). To our knowledge, none of these fusions were detected previously.

We used linear models to find the relationships between fusion candidates and their partner genes with clinicopathological variables and survival. The rate of genomic rearrangements can be used as a proxy for genomic stability. In our analysis, the number of fusion candidates per sample was not significantly associated with any of the clinical covariates or patient survival. There were significant associations (raw *p*-value <0.05) between individual fusions and certain demographic and clinical variables. For example, *CD27-AS1:MANBAL* had a higher occurrence in female patients, *DNPEP:C9ORF3* showed association with recurrence, *GABPB1:LMCD1* had a higher frequency in large cell neuroendocrine histology, and *LMO7:EXT2* in adenosquamous carcinoma. All of these fusions were novel with no literature data available. Survival analysis with Cox proportional hazards model identified detection at stage II and recurrence as strong predictors (*p*<0.001) of poor outcomes when other variables were accounted for (Supplementary Table VIII). The association of survival with stage III did not reach significance, likely due to the smaller samples size (n=36). Weaker associations (*p*<0.05) were detected with smoking, where "never" smokers had improved survival, while the poorer outcome was associated with LCC histology. No significant associations were found between survival and the occurrence of individual fusion candidates. The presence of *MROH1* as a variable fusion partner was significantly associated with decreased survival (HR=14.04; *p*<0.001), however, these results are based on 2 positive samples only. Despite being a large and highly conserved gene, *MROH1* is not well researched, a single functional study suggests its involvement in vesicle transport

(52). There is no research on the involvement of this gene in cancer, however, *MROH1:BOP1* fusion was detected and validated in CCLE cell lines (56); the data available in the COSMIC portal shows *MROH1* over-expression in multiple cancers, including intestinal, stomach, lung, and ovary.

In conclusion, we applied the JAFFA fusion detection pipeline to identify potential transcript fusion candidates in 270 RNA-seq libraries. All the datasets obtained from archived NSCLC tumor samples representing early stages of cancer development. In total we detected 782 transcript fusions, only 33 of them were found in fusion databases and the rest appear novel. Many fusions included well-known and emerging oncogenes with the evidence of lung cancer involvement revealed by literature searches. Considering the association between fusion recurrence and their role in malignancy, we compiled the lists of recurrent fusions and promiscuous fusion partners annotated with the literature evidence of cancer role. In addition to the analysis of published data, fusions with high oncodriver potential were predicted with computation methods. Considering the availability of clinicopathological data we analyzed the associations between recurrent fusions and fusion partners with select clinical covariates including survival. The major drawbacks and limitations of this study lie in the lack of molecular validation of the promising fusion candidates. The nature of the study that used RNA sequencing reads only makes it impossible to differentiate transcript fusions originating from genomic re-arrangements and those caused at the post-transcriptional level (trans-splicing). Identification of a wide range of transcript fusion candidates enables us and other interested researchers to select, validate, and characterize clinically actionable targets to improve patient outcomes.

Supplementary Material

Available at: https://github.com/slavailn/fusion_manuscript_suppl_materials

Conflicts of Interest

The Authors disclose no conflicts of interest.

Authors' Contributions

YI: Bioinformatics data analysis, investigation, writing the original manuscript. LP: Conceptualization, data curation, review. JBM, MK, AE, and AG: experimental work. OK and IK conceptualization, review, editing. GB: conceptualization, review, editing, supervision.

Acknowledgements

References

1 Siegel RL, Miller KD and Jemal A: Cancer statistics, 2020. CA Cancer J Clin 70(1): 7-30, 2020. PMID: 31912902. DOI: 10.3322/caac.21590

2 Duma N, Santana-Davila R and Molina JR: Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc 94(8): 1623-1640, 2019. PMID: 31378236. DOI: 10.1016/j.mayocp.2019.01.013

3 Schegoleva AA, Khozyainova AA, Fedorov AA, Gerashchenko TS, Rodionov EO, Topolnitsky EB, Shefer NA, Pankova OV, Durova AA, Zavyalova MV, Perelmuter VM and Denisov EV: Prognosis of different types of non-small cell lung cancer progression: current state and perspectives. Cell Physiol Biochem 55(S2): 29-48, 2021. PMID: 33687819. DOI: 10.33594/000000340

4 Gibson AJW, Box A, Dean ML, Elegbede AA, Hao D, Sangha R and Bebb DG: Retrospective real-world outcomes for patients with ALK-rearranged lung cancer receiving ALK receptor tyrosine kinase inhibitors. JTO Clin Res Rep 2(4): 100157, 2021. PMID: 34590010. DOI: 10.1016/j.jtocrr.2021.100157

5 Yuan M, Huang LL, Chen JH, Wu J and Xu Q: The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. Signal Transduct Target Ther 4: 61, 2019. PMID: 31871778. DOI: 10.1038/s41392-019-0099-9

6 Huang L, Jiang S and Shi Y: Tyrosine kinase inhibitors for solid tumors in the past 20 years (2001-2020). J Hematol Oncol 13(1): 143, 2020. PMID: 33109256. DOI: 10.1186/s13045-020-00977-0

7 Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, Bose R, Ornitz D, Xiong D, You M, Dooling DJ, Watson M, Mardis ER and Wilson RK: Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 150(6): 1121-1134, 2012. PMID: 22980976. DOI: 10.1016/j.cell.2012.08.024

8 Mao G, Xuxing P, Hao W, Junbing W, Suyue L, Qiang W, Ying L, Ren Q, Liu J and Melançon CE: Mutational landscape of early-stage non-small cell lung cancer patients using a 451 gene panel. JCO 37(15_suppl): e20048-e20048, 2019. DOI: 10.1200/JCO.2019.37.15_suppl.e20048

9 Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES and Meyerson M: Characterizing the cancer genome in lung adenocarcinoma. Nature 450(7171): 893-898, 2007. PMID: 17982442. DOI: 10.1038/nature06358

10 Tuna M, Amos CI and Mills GB: Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. Oncotarget 10(21): 2095-2111, 2019. PMID: 31007851. DOI: 10.18632/oncotarget.26777

11 Jia Y, Xie Z and Li H: Intergenically spliced chimeric RNAs in cancer. Trends Cancer 2(9): 475-484, 2016. PMID: 28210711. DOI: 10.1016/j.trecan.2016.07.006

12 Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y and Mano H: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448(7153): 561-566, 2007. PMID: 17625570. DOI: 10.1038/nature05945

13 Davies KD, Le AT, Theodoro MF, Skokan MC, Aisner DL, Berge EM, Terracciano LM, Cappuzzo F, Incarbone M, Roncalli M, Alloisio M, Santoro A, Camidge DR, Varella-Garcia M and Doebele RC: Identifying and targeting ROS1 gene fusions in non-small cell lung cancer. Clin Cancer Res 18(17): 4570-4579, 2012. PMID: 22919003. DOI: 10.1158/1078-0432.CCR-12-0550

14 Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH and Seo JS: A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res 22(3): 436-445, 2012. PMID: 22194472. DOI: 10.1101/gr.133645.111

15 Cocco E, Scaltriti M and Drilon A: NTRK fusion-positive cancers and TRK inhibitor therapy. Nat Rev Clin Oncol 15(12): 731-747, 2018. PMID: 30333516. DOI: 10.1038/s41571-018-0113-0

16 Laskin J, Liu SV, Tolba K, Heining C, Schlenk RF, Cheema P, Cadranel J, Jones MR, Drilon A, Cseh A, Gyorffy S, Solca F and Duruisseaux M: NRG1 fusion-driven tumors: biology, detection, and the therapeutic role of afatinib and other ErbB-targeting agents. Ann Oncol 31(12): 1693-1703, 2020. PMID: 32916265. DOI: 10.1016/j.annonc.2020.08.2335

17 Wang L, Hao K, Yang T and Wang C: Role of the lung microbiome in the pathogenesis of chronic obstructive pulmonary disease. Chin Med J (Engl) 130(17): 2107-2111, 2017. PMID: 28741603. DOI: 10.4103/0366-6999.211452

18 Bruno R and Fontanini G: Next generation sequencing for gene fusion analysis in lung cancer: a literature review. Diagnostics (Basel) 10(8): 521, 2020. PMID: 32726941. DOI: 10.3390/diagnostics10080521

19 Jang YE, Jang I, Kim S, Cho S, Kim D, Kim K, Kim J, Hwang J, Kim S, Kim J, Kang J, Lee B and Lee S: ChimerDB 4.0: an updated and expanded database of fusion genes. Nucleic Acids Res 48(D1): D817-D824, 2020. PMID: 31680157. DOI: 10.1093/nar/gkz1013

20 Panigrahi P, Jere A and Anamika K: FusionHub: A unified web platform for annotation and visualization of gene fusion events in human cancer. PLoS One 13(5): e0196588, 2018. PMID: 29715310. DOI: 10.1371/journal.pone.0196588

21 Shugay M, Ortiz de Mendíbil I, Vizmanos JL and Novo FJ: Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. Bioinformatics 29(20): 2539-2546, 2013. PMID: 23956304. DOI: 10.1093/bioinformatics/btt445

22 Chang YS, Tu SJ, Yen JC, Lee YT, Fang HY and Chang JG: The fusion gene landscape in Taiwanese patients with non-small cell lung cancer. Cancers (Basel) 13(6): 1343, 2021. PMID: 33809651. DOI: 10.3390/cancers13061343

23 Kim SC, Jung Y, Park J, Cho S, Seo C, Kim J, Kim P, Park J, Seo J, Kim J, Park S, Jang I, Kim N, Yang JO, Lee B, Rho K, Jung Y, Keum J, Lee J, Han J, Kang S, Bae S, Choi SJ, Kim S, Lee JE, Kim W, Kim J and Lee S: A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. PLoS One 8(2): e55596, 2013. PMID: 23405175. DOI: 10.1371/journal.pone.0055596

24 Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, Bergbower EA, Guan Y, Shin J, Guillory J, Rivers CS, Foo CK, Bhatt D, Stinson J, Gnad F, Haverty PM, Gentleman R, Chaudhuri S, Janakiraman V, Jaiswal BS, Parikh C, Yuan W, Zhang Z, Koeppen H, Wu TD, Stern HM, Yauch RL, Huffman KE, Paskulin DD, Illei PB, Varella-Garcia M, Gazdar AF, de Sauvage FJ, Bourgon R, Minna JD, Brock MV and Seshagiri S: Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet 44(10): 1111-1116, 2012. PMID: 22941189. DOI: 10.1038/ng.2405

25 Petersen LF, Klimowicz AC, Otsuka S, Elegbede AA, Petrillo SK, Williamson T, Williamson CT, Konno M, Lees-Miller SP, Hao D, Morris D, Magliocco AM and Bebb DG: Loss of tumour-specific ATM protein expression is an independent prognostic factor in early resected NSCLC. Oncotarget 8(24): 38326-38336, 2017. PMID: 28418844. DOI: 10.18632/oncotarget.16215

26 Otsuka S, Klimowicz AC, Kopciuk K, Petrillo SK, Konno M, Hao D, Muzik H, Stolte E, Boland W, Morris D, Magliocco AM and Bebb DG: CXCR4 overexpression is associated with poor outcome in females diagnosed with stage IV non-small cell lung cancer. J Thorac Oncol 6(7): 1169-1178, 2011. PMID: 21623238. DOI: 10.1097/JTO.0b013e3182199a99

27 Harris PA: Research Electronic Data Capture (REDCap) - planning, collecting and managing data for clinical and translational research. BMC Bioinformatics 13(12): A15, 2012. DOI: 10.1186/1471-2105-13-S12-A15

28 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N and Conde JG: Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 42(2): 377-381, 2009. PMID: 18929686. DOI: 10.1016/j.jbi.2008.08.010

29 Kim D, Langmead B and Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12(4): 357-360, 2015. PMID: 25751142. DOI: 10.1038/nmeth.3317

30 Liao Y, Smyth GK and Shi W: featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7): 923-930, 2014. PMID: 24227677. DOI: 10.1093/bioinformatics/btt656

31 Illumina. iGenomes. Ready-To-Use Reference Sequences and Annotations. Available at: https://support.illumina.com/sequencing/sequencing_software/igenome.html [Last accessed on June 29, 2023]

32 Davidson NM, Majewski IJ and Oshlack A: JAFFA: High sensitivity transcriptome-focused fusion gene detection. Genome Med 7(1): 43, 2015. PMID: 26019724. DOI: 10.1186/s13073-015-0167-x

33 Schulz MH, Zerbino DR, Vingron M and Birney E: Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28(8): 1086-1092, 2012. PMID: 22368243. DOI: 10.1093/bioinformatics/bts094

34 Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A and Huber W: BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21(16): 3439-3440, 2005. PMID: 16082012. DOI: 10.1093/bioinformatics/bti525

35 Babiceanu M, Qin F, Xie Z, Jia Y, Lopez K, Janus N, Facemire L, Kumar S, Pang Y, Qi Y, Lazar IM and Li H: Recurrent chimeric fusion RNAs in non-cancer tissues and cells. Nucleic Acids Res 44(6): 2859-2872, 2016. PMID: 26837576. DOI: 10.1093/nar/gkw032

36 Liu YF, Wang BY, Zhang WN, Huang JY, Li BS, Zhang M, Jiang L, Li JF, Wang MJ, Dai YJ, Zhang ZG, Wang Q, Kong J, Chen B, Zhu YM, Weng XQ, Shen ZX, Li JM, Wang J, Yan XJ, Li Y, Liang YM, Liu L, Chen XQ, Zhang WG, Yan JS, Hu JD, Shen SH, Chen J, Gu LJ, Pei D, Li Y, Wu G, Zhou X, Ren RB, Cheng C, Yang JJ, Wang KK, Wang SY, Zhang J, Mi JQ, Pui CH, Tang JY, Chen Z and Chen SJ: Genomic profiling of adult and pediatric B-cell acute lymphoblastic leukemia. EBioMedicine 8: 173-183, 2016. PMID: 27428428. DOI: 10.1016/j.ebiom.2016.04.038

37 Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H and Verhaak RG: The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene 34(37): 4845-4854, 2015. PMID: 25500544. DOI: 10.1038/onc.2014.406

38 Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, Ng PK, Jeong KJ, Cao S, Wang Z, Gao J, Gao Q, Wang F, Liu EM, Mularoni L, Rubio-Perez C, Nagarajan N, Cortés-Ciriano I, Zhou DC, Liang WW, Hess JM, Yellapantula VD, Tamborero D, Gonzalez-Perez A, Suphavilai C, Ko JY, Khurana E, Park PJ, Van Allen EM, Liang H, MC3 Working Group, Cancer Genome Atlas Research Network, Lawrence MS, Godzik A, Lopez-Bigas N, Stuart J, Wheeler D, Getz G, Chen K, Lazar AJ, Mills GB, Karchin R and Ding L: Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 173(2): 371-385.e18, 2018. PMID: 29625053. DOI: 10.1016/j.cell.2018.02.060

39 Stangl C, de Blank S, Renkens I, Westera L, Verbeek T, Valle-Inclan JE, González RC, Henssen AG, van Roosmalen MJ, Stam RW, Voest EE, Kloosterman WP, van Haaften G and Monroe GR: Partner independent fusion gene detection by multiplexed CRISPR-Cas9 enrichment and long read nanopore sequencing. Nat Commun 11(1): 2861, 2020. PMID: 32504042. DOI: 10.1038/s41467-020-16641-7

40 Stickens D, Clines G, Burbee D, Ramos P, Thomas S, Hogue D, Hecht JT, Lovett M and Evans GA: The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. Nat Genet 14(1): 25-32, 1996. PMID: 8782816. DOI: 10.1038/ng0996-25

41 Zhang T, Liu C, Yu Y, Geng J, Meng Q, Xu S, Zhou F, Chen Y, Jin S, Shen J, Pan B, Meng F and Liu F: TBL1XR1 is involved in c-Met-mediated tumorigenesis of human nonsmall cell lung cancer. Cancer Gene Ther 27(3-4): 136-146, 2020. PMID: 31243347. DOI: 10.1038/s41417-019-0111-0

42 Zhao JG, Ren KM and Tang J: Zinc finger protein ZBTB20 promotes cell proliferation in non-small cell lung cancer through repression of FoxO1. FEBS Lett 588(24): 4536-4542, 2014. PMID: 25311537. DOI: 10.1016/j.febslet.2014.10.005

43 Liu XS, Haines JE, Mehanna EK, Genet MD, Ben-Sahra I, Asara JM, Manning BD and Yuan ZM: ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. Genes Dev 28(17): 1917-1928, 2014. PMID: 25184678. DOI: 10.1101/gad.245910.114

44 Jiang N, Zhou J, Zhang W, Li P, Liu Y, Shi H, Zhang C, Wang Y, Zhou C, Peng C, Zhang W, Hao Y, Sun Q, Li Y and Zhao X: RNF213 gene mutation in circulating tumor DNA detected by targeted next-generation sequencing in the assisted discrimination of early-stage lung cancer from pulmonary nodules. Thorac Cancer 12(2): 181-193, 2021. PMID: 33200540. DOI: 10.1111/1759-7714.13741

45 De Coninck S, Berx G, Taghon T, Van Vlierberghe P and Goossens S: ZEB2 in T-cells and T-ALL. Adv Biol Regul 74: 100639, 2019. PMID: 31383581. DOI: 10.1016/j.jbior.2019.100639

46 Hou Y, Zhen J, Xu X, Zhen K, Zhu B, Pan R and Zhao C: miR-215 functions as a tumor suppressor and directly targets ZEB2 in human non-small cell lung cancer. Oncol Lett 22(2): 600, 2021. PMID: 34188702. DOI: 10.3892/ol.2021.12861

47 Lin X, Yang Z, Zhang P, Liu Y and Shao G: miR-154 inhibits migration and invasion of human non-small cell lung cancer by targeting ZEB2. Oncol Lett 22(2): 612, 2021. PMID: 34188714. DOI: 10.3892/ol.2021.12873

48 Zhang C, Qiao H, Guo W, Liu Y, Yang L, Liu Y, Jin B, Fu M, Wang G and Li W: CD100-plexin-B1 induces epithelial-mesenchymal transition of head and neck squamous cell carcinoma and promotes metastasis. Cancer Lett 455: 1-13, 2019. PMID: 30981760. DOI: 10.1016/j.canlet.2019.04.013

49 Liu Y, Ouyang Q, Sun Z, Tan J, Huang W, Liu J, Liu Z, Zhou H, Zeng F and Liu Y: The novel zinc finger protein 587B gene, ZNF587B, regulates cell proliferation and metastasis in ovarian cancer cells in vivo and *in vitro*. Cancer Manag Res 12: 5119-5130, 2020. PMID: 32617020. DOI: 10.2147/CMAR.S252347

50 Chen SC, Yen MC, Chen FW, Wu LY, Yang SJ, Kuo PL and Hsu YL: Knockdown of GA-binding protein subunit β1 inhibits cell proliferation *via* p21 induction in renal cell carcinoma. Int J Oncol 53(2): 886-894, 2018. PMID: 29845229. DOI: 10.3892/ijo.2018.4411

51 Chang CY, Lin SC, Su WH, Ho CM and Jou YS: Somatic LMCD1 mutations promoted cell migration and tumor metastasis in hepatocellular carcinoma. Oncogene 31(21): 2640-2652, 2012. PMID: 21996735. DOI: 10.1038/onc.2011.440

52 Thomason PA, King JS and Insall RH: Mroh1, a lysosomal regulator localized by WASH-generated actin. J Cell Sci 130(10): 1785-1795, 2017. PMID: 28424231. DOI: 10.1242/jcs.197210

53 Sharbatoghli M, Fattahi F, Aboulkheyr Es H, Akbari A, Akhavan S, Ebrahimi M, Asadi-Lari M, Totonchi M and Madjd Z: Copy number variation of circulating tumor DNA (ctDNA) detected using NIPT in neoadjuvant chemotherapy-treated ovarian cancer patients. Front Genet 13: 938985, 2022. PMID: 35938032. DOI: 10.3389/fgene.2022.938985

54 Dehghannasiri R, Freeman DE, Jordanski M, Hsieh GL, Damljanovic A, Lehnert E and Salzman J: Improved detection of gene fusions by applying statistical methods reveals oncogenic RNA cancer drivers. Proc Natl Acad Sci USA 116(31): 15524-15533, 2019. PMID: 31308241. DOI: 10.1073/pnas.1900391116

55 Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, Liao WW, Reynolds SM, Wyczalkowski MA, Yao L, Yu L, Sun SQ, Fusion Analysis Working Group, Cancer Genome Atlas Research Network, Chen K, Lazar AJ, Fields RC, Wendl MC, Van Tine BA, Vij R, Chen F, Nykter M, Shmulevich I and Ding L: Driver fusions and their implications in the development and treatment of human cancers. Cell Rep 23(1): 227-238.e3, 2018. PMID: 29617662. DOI: 10.1016/j.celrep.2018.03.050

56 Vellichirammal NN, Albahrani A, Banwait JK, Mishra NK, Li Y, Roychoudhury S, Kling MJ, Mirza S, Bhakat KK, Band V, Joshi SS and Guda C: Pan-Cancer analysis reveals the diverse landscape of novel sense and antisense fusion transcripts. Mol Ther Nucleic Acids 19: 1379-1398, 2020. PMID: 32160708. DOI: 10.1016/j.omtn.2020.01.023

57 Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, Kim J, Lee ER, Kang CH, Park IK, Rhee H, Lee SH, Kim JI, Kang JH and Kim YT: The transcriptional landscape and mutational profile of lung

adenocarcinoma. Genome Res 22(11): 2109-2119, 2012. PMID: 22975805. DOI: 10.1101/gr.145144.112

58 Miyanaga A, Masuda M, Motoi N, Tsuta K, Nakamura Y, Nishijima N, Watanabe SI, Asamura H, Tsuchida A, Seike M, Gemma A and Yamada T: Whole-exome and RNA sequencing of pulmonary carcinoid reveals chromosomal rearrangements associated with recurrence. Lung Cancer 145: 85-94, 2020. PMID: 32417679. DOI: 10.1016/j.lungcan.2020.03.027

59 Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, Aerts A, Altherr M, Ashworth L, Bajorek E, Black S, Branscomb E, Caenepeel S, Carrano A, Caoile C, Chan YM, Christensen M, Cleland CA, Copeland A, Dalin E, Dehal P, Denys M, Detter JC, Escobar J, Flowers D, Fotopulos D, Garcia C, Georgescu AM, Glavina T, Gomez M, Gonzales E, Groza M, Hammon N, Hawkins T, Haydu L, Ho I, Huang W, Israni S, Jett J, Kadner K, Kimball H, Kobayashi A, Larionov V, Leem SH, Lopez F, Lou Y, Lowry S, Malfatti S, Martinez D, McCready P, Medina C, Morgan J, Nelson K, Nolan M, Ovcharenko I, Pitluck S, Pollard M, Popkie AP, Predki P, Quan G, Ramirez L, Rash S, Retterer J, Rodriguez A, Rogers S, Salamov A, Salazar A, She X, Smith D, Slezak T, Solovyev V, Thayer N, Tice H, Tsai M, Ustaszewska A, Vo N, Wagner M, Wheeler J, Wu K, Xie G, Yang J, Dubchak I, Furey TS, DeJong P, Dickson M, Gordon D, Eichler EE, Pennacchio LA, Richardson P, Stubbs L, Rokhsar DS, Myers RM, Rubin EM and Lucas SM: The DNA sequence and biology of human chromosome 19. Nature 428(6982): 529-535, 2004. PMID: 15057824. DOI: 10.1038/nature02399

60 Rose AM, Krishan A, Chakarova CF, Moya L, Chambers SK, Hollands M, Illingworth JC, Williams SMG, McCabe HE, Shah AZ, Palmer CNA, Chakravarti A, Berg JN, Batra J and Bhattacharya SS: MSR1 repeats modulate gene expression and affect risk of breast and prostate cancer. Ann Oncol 29(5): 1292-1303, 2018. PMID: 29509840. DOI: 10.1093/annonc/mdy082

61 Ruosaari ST, Nymark PE, Aavikko MM, Kettunen E, Knuutila S, Hollmén J, Norppa H and Anttila SL: Aberrations of chromosome 19 in asbestos-associated lung cancer and in asbestos-induced micronuclei of bronchial epithelial cells *in vitro*. Carcinogenesis 29(5): 913-917, 2008. PMID: 18339684. DOI: 10.1093/carcin/bgn068

62 Wang X, Zhang Y, Nilsson CL, Berven FS, Andrén PE, Carlsohn E, Horvatovich P, Malm J, Fuentes M, Végvári Á, Welinder C, Fehniger TE, Rezeli M, Edula G, Hober S, Nishimura T and Marko-Varga G: Association of chromosome 19 to lung cancer genotypes and phenotypes. Cancer Metastasis Rev 34(2): 217-226, 2015. PMID: 25982285. DOI: 10.1007/s10555-015-9556-2

63 Chevallier M, Borgeaud M, Addeo A and Friedlaender A: Oncogenic driver mutations in non-small cell lung cancer: Past, present and future. World J Clin Oncol 12(4): 217-237, 2021. PMID: 33959476. DOI: 10.5306/wjco.v12.i4.217

64 Matsuura S, Shinmura K, Kamo T, Igarashi H, Maruyama K, Tajima M, Ogawa H, Tanahashi M, Niwa H, Funai K, Kohno T, Suda T and Sugimura H: CD74-ROS1 fusion transcripts in resected non-small cell lung carcinoma. Oncol Rep 30(4): 1675-1680, 2013. PMID: 23877438. DOI: 10.3892/or.2013.2630

65 Batra U, Sharma M, Dewan A, Diwan H and Nathany S: A narrative review of ERBB2 in non-small cell lung carcinoma. Cancer Res Stat Treat 5(1): 97, 2022. DOI: 10.4103/crst.crst_323_21

66 Mei L and Borg JP: ERBB2 oncogenicity: ERBIN helps to perform the job. Mol Cell Oncol 2(3): e995033, 2015. PMID: 27308480. DOI: 10.4161/23723556.2014.995033

67 Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A and Lopez-Bigas N: IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods 10(11): 1081-1082, 2013. PMID: 24037244. DOI: 10.1038/nmeth.2642

68 Kurppa KJ, Denessiouk K, Johnson MS and Elenius K: Activating ERBB4 mutations in non-small cell lung cancer. Oncogene 35(10): 1283-1291, 2016. PMID: 26050618. DOI: 10.1038/onc.2015.185

69 Guenzi E, Pluvy J, Guyard A, Nguenang M, Rebah K, Benrahmoune Z, Lamoril J, Cazes A, Gounant V, Brosseau S, Boileau C, Zalcman G and Théou-Anton N: A new KIF5B-ERBB4 gene fusion in a lung adenocarcinoma patient. ERJ Open Res 7(1): 00582, 2021. PMID: 33816604. DOI: 10.1183/23120541.00582-2020

70 Alaei-Mahabadi B, Bhadury J, Karlsson JW, Nilsson JA and Larsson E: Global analysis of somatic structural genomic alterations and their impact on gene expression in diverse human cancers. Proc Natl Acad Sci U S A 113(48): 13768-13773, 2016. PMID: 27856756. DOI: 10.1073/pnas.1606220113

71 Liu J, Li X and Peng J: A novel CAV1-MET fusion in SCLC transformation responds to crizotinib and osimertinib treatment. J Thorac Oncol 14(6): e126-e128, 2019. PMID: 31122565. DOI: 10.1016/j.jtho.2019.01.025

72 Zhou JB, Zhang T, Wang BF, Gao HZ and Xu X: Identification of a novel gene fusion RNF213 SLC26A11 in chronic myeloid leukemia by RNA-Seq. Mol Med Rep 7(2): 591-597, 2013. PMID: 23151810. DOI: 10.3892/mmr.2012.1183

73 Banh RS, Iorio C, Marcotte R, Xu Y, Cojocari D, Rahman AA, Pawling J, Zhang W, Sinha A, Rose CM, Isasa M, Zhang S, Wu R, Virtanen C, Hitomi T, Habu T, Sidhu SS, Koizumi A, Wilkins SE, Kislinger T, Gygi SP, Schofield CJ, Dennis JW, Wouters BG and Neel BG: PTP1B controls non-mitochondrial oxygen consumption by regulating RNF213 to promote tumour survival during hypoxia. Nat Cell Biol 18(7): 803-813, 2016. PMID: 27323329. DOI: 10.1038/ncb3376

74 Apostolopoulou K, Pateras IS, Evangelou K, Tsantoulis PK, Liontos M, Kittas C, Tiniakos DG, Kotsinas A, Cordon-Cardo C and Gorgoulis VG: Gene amplification is a relatively frequent event leading to ZBTB7A (Pokemon) overexpression in non-small cell lung cancer. J Pathol 213(3): 294-302, 2007. PMID: 17907153. DOI: 10.1002/path.2222

75 Zhao ZH, Wang SF, Yu L, Wang J, Chang H, Yan WL, Zhang J and Fu K: Overexpression of Pokemon in non-small cell lung cancer and foreshowing tumor biological behavior as well as clinical results. Lung Cancer 62(1): 113-119, 2008. PMID: 18550205. DOI: 10.1016/j.lungcan.2008.02.014

76 Torkildsen S, Gorunova L, Beiske K, Tjønnfjord GE, Heim S and Panagopoulos I: Novel ZEB2-BCL11B fusion gene identified by RNA-sequencing in acute myeloid leukemia with t(2;14)(q22;q32). PLoS One 10(7): e0132736, 2015. PMID: 26186352. DOI: 10.1371/journal.pone.0132736

77 Krsková L, Němečková T, Balko J, Brož P and Vícha A: Novel ZEB2-PLAG1 fusion gene identified by RNA sequencing in a case of lipoblastoma. Pediatr Blood Cancer 68(3): e28691, 2021. PMID: 32918527. DOI: 10.1002/pbc.28691

78 Cui J, Pan G, He Q, Yin L, Guo R and Bi H: MicroRNA-545 targets ZEB2 to inhibit the development of non-small cell lung cancer by inactivating Wnt/β-catenin pathway. Oncol Lett 18(3): 2931-2938, 2019. PMID: 31452774. DOI: 10.3892/ol.2019.10619

79 Wu DM, Zhang T, Liu YB, Deng SH, Han R, Liu T, Li J and Xu Y: The PAX6-ZEB2 axis promotes metastasis and cisplatin resistance in non-small cell lung cancer through PI3K/AKT signaling. Cell Death Dis 10(5): 349, 2019. PMID: 31024010. DOI: 10.1038/s41419-019-1591-4

80 Alto LT and Terman JR: Semaphorins and their signaling mechanisms. Methods Mol Biol 1493: 1-25, 2017. PMID: 27787839. DOI: 10.1007/978-1-4939-6448-2_1

81 Wong OG, Nitkunan T, Oinuma I, Zhou C, Blanc V, Brown RS, Bott SR, Nariculam J, Box G, Munson P, Constantinou J, Feneley MR, Klocker H, Eccles SA, Negishi M, Freeman A, Masters JR and Williamson M: Plexin-B1 mutations in prostate cancer. Proc Natl Acad Sci USA 104(48): 19040-19045, 2007. PMID: 18024597. DOI: 10.1073/pnas.0702544104

82 Worzfeld T, Swiercz JM, Looso M, Straub BK, Sivaraj KK and Offermanns S: ErbB-2 signals through Plexin-B1 to promote breast cancer metastasis. J Clin Invest 122(4): 1296-1305, 2012. PMID: 22378040. DOI: 10.1172/JCI60568

83 Du L, Yamamoto S, Burnette BL, Huang D, Gao K, Jamshidi N and Kuo MD: Transcriptome profiling reveals novel gene expression signatures and regulating transcription factors of TGFβ-induced epithelial-to-mesenchymal transition. Cancer Med 5(8): 1962-1972, 2016. PMID: 27318801. DOI: 10.1002/cam4.719

84 Slowicka K, Petta I, Blancke G, Hoste E, Dumas E, Sze M, Vikkula H, Radaelli E, Haigh JJ, Jonckheere S, Taminau J, Vandamme N, Wullaert A, Tulchinsky E, Nittner D, Van Vlierberghe P, De Hertogh G, Baldin P, Etlioglu E, Wirapati P, Boon L, Lambrecht BN, Callewaert C, Tejpar S, Goossens S, Berx G, Vereecke L and van Loo G: Zeb2 drives invasive and microbiota-dependent colon carcinoma. Nat Cancer 1(6): 620-634, 2020. PMID: 35121975. DOI: 10.1038/s43018-020-0070-2