



---

Original article

## FARE-CAFE: a database of functional and regulatory elements of cancer-associated fusion events

Praveen Kumar Korla<sup>1</sup>, Jack Cheng<sup>2</sup>, Chien-Hung Huang<sup>3</sup>,  
Jeffrey J. P. Tsai<sup>1</sup>, Yu-Hsuan Liu<sup>3</sup>, Nilubon Kurubanjerdjit<sup>4</sup>,  
Wen-Tsong Hsieh<sup>5</sup>, Huey-Yi Chen<sup>6</sup> and Ka-Lok Ng<sup>1,7,\*</sup>

<sup>1</sup>Department of Bioinformatics and Medical Engineering, Asia University, Taichung 41354, Taiwan, <sup>2</sup>Graduate Institute of Integrated Medicine, College of Chinese Medicine, China Medical University, Taichung 40402, Taiwan, <sup>3</sup>Department of Computer Science and Information Engineering, National Formosa University, Yunlin 632, Taiwan, <sup>4</sup>School of Information Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand, <sup>5</sup>Department of Pharmacology, China Medical University, Taichung 40402, Taiwan, <sup>6</sup>Department of Obstetrics and Gynecology, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan, and <sup>7</sup>Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan

\*Corresponding author: Tel: +886 4 23394541; Fax: +886 4 23320718; Email: ppiddi@gmail.com

Citation details: Korla,P.K., Cheng,J., Huang,C.-H., *et al.* FARE-CAFE: a database of functional and regulatory elements of cancer-associated fusion events. *Database* (2015) Vol. 2015: article ID bav086; doi:10.1093/database/bav086

Received 29 April 2015; Revised 14 August 2015; Accepted 18 August 2015

### Abstract

Chromosomal translocation (CT) is of enormous clinical interest because this disorder is associated with various major solid tumors and leukemia. A tumor-specific fusion gene event may occur when a translocation joins two separate genes. Currently, various CT databases provide information about fusion genes and their genomic elements. However, no database of the roles of fusion genes, in terms of essential functional and regulatory elements in oncogenesis, is available. FARE-CAFE is a unique combination of CTs, fusion proteins, protein domains, domain–domain interactions, protein–protein interactions, transcription factors and microRNAs, with subsequent experimental information, which cannot be found in any other CT database. Genomic DNA information including, for example, manually collected exact locations of the first and second break points, sequences and karyotypes of fusion genes are included. FARE-CAFE will substantially facilitate the cancer biologist’s mission of elucidating the pathogenesis of various types of cancer. This database will ultimately help to develop ‘novel’ therapeutic approaches.

**Database URL:** <http://ppi.bioinfo.asia.edu.tw/FARE-CAFE>

---

## Introduction

Chromosomal rearrangements play a crucial role in the progression of cancer, and in particular, in chromosomal translocation (CT) events; their corresponding fusion genes (FGs) are essential in the initiation and/or development of cancer (1). FGs are frequently examined in clinical diagnosis, treatment and cancer prognosis (2). FG events account for at least 20% of all cancer cases (3).

In the past two decades, commonly used cytogenetic techniques (FISH, SKY, CGH and PCR) and recent advances in sequencing technologies have revealed a high number of CT events in human tumors (4). These experimental discoveries have improved our understanding of the pathogenetic importance of CT events in carcinogenesis (5, 6).

Mitelman's database (7), dbCRID (4), TICdb (8), HYBRIDdb (9) and ChimerDB (10) are databases of cancer-associated CTs and FGs that have been identified experimentally and in the literature. The four databases other than Mitelman's database provide fusion sequences with the positions of break points at nucleotide level. Only dbCRID provides the 'exact' positions of break points, and it does so for only a 'few' FGs; in contrast, this work provides 'exact' break point information for 'all' FGs including their 'isoforms'. Break point positions are mapped using the Human genome assembly hg19 (NCBI Build 37.1 Feb 2009).

During transcription, transcription factors (TFs) regulate the promoter region of the FG. These TFs enhance or inhibit the transcription of the tumorigenic FGs. At the post-transcriptional modification level, the expression of the fusion transcript that was regulated by microRNAs (miRNAs) was studied. These miRNAs may target the 3'-end of the FG, suppressing the expression of the FG, and thereby promote apoptosis and reduce cell proliferation (11, 12). In summary, these two regulatory elements, TFs and miRNAs, crucially affect cancer progression.

Based on the first and second break points of FG, the coding regions were extracted and the domain composition of the fusion proteins is annotated. 5' and 3' partner genes of FG have fewer domains than wild-type 5' and 3' genes due to translocation, so the FGs failed to exhibit certain domain-domain interactions (DDIs) after translation. The absence of these domains disrupted certain DDIs and DDI-mediated protein-protein interactions (PPIs) (Figure 1), further disrupting the regular biological processes that are associated with cancer-associated pathways.

In this work, a number of resources, including FGs, TFs, miRNAs, miRNA targets, domains, information

about DDIs and PPIs are integrated to construct a database of functional and regulatory elements of cancer-associated fusion events (FARE-CAFE). This database supports an examination of the role of cancer-associated FGs/proteins at various levels of regulation, which are transcription initiation (TF→FG), post-transcriptional regulation (miRNA→FG) and functional interaction corresponding to domains, DDIs and PPIs. To the best of our knowledge, this database is the first to address the relationships among FGs, TFs, miRNAs, DDIs and PPIs in a study of cancer (Figure 1).

## FARE-CAFE database

### Input data

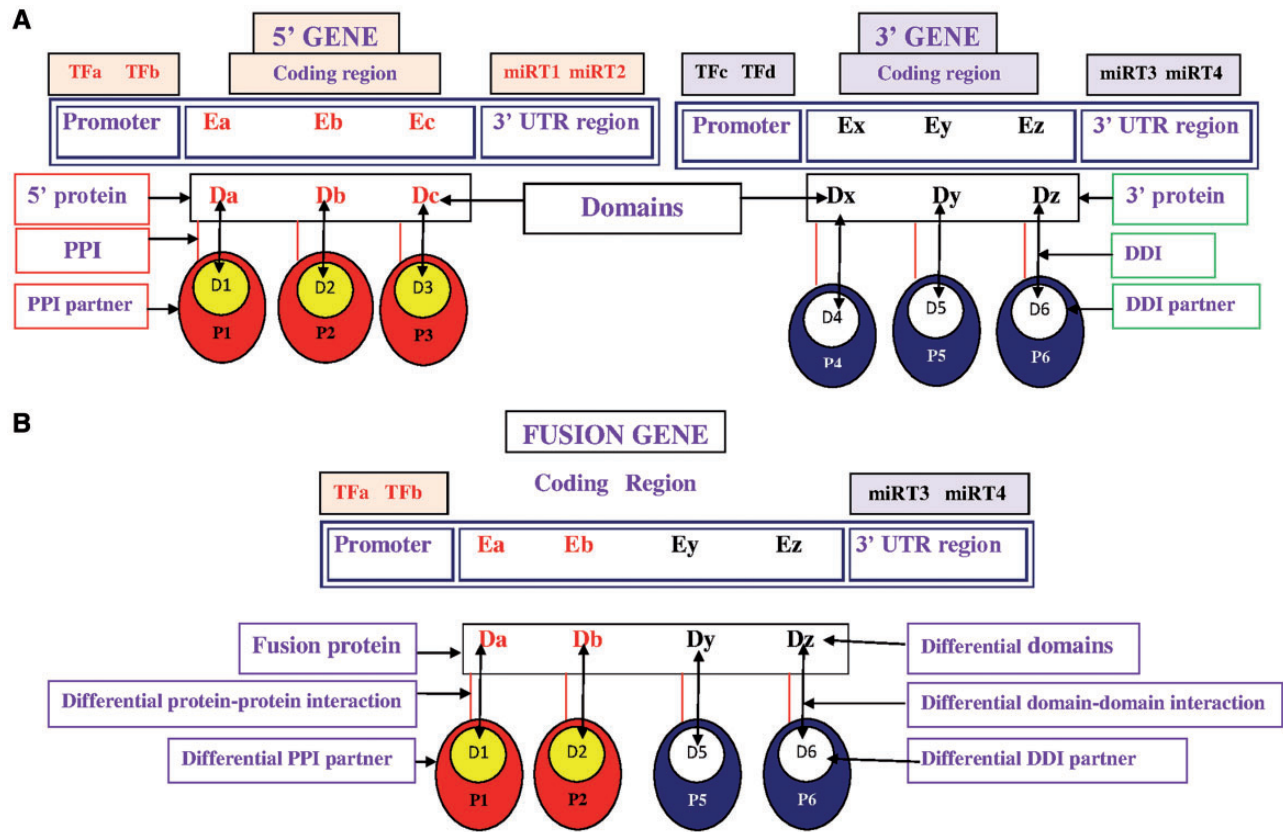
To construct FARE-CAFE, functional, regulatory and genomic information about FGs were collected from various data sources (Supplementary Table S1). Figure 2 presents the workflow of FARE-CAFE. Genomic information about the FGs was obtained from Mitelman's database, dbCRID and TICdb. Positions of break point at nucleotide level in fusion sequences were obtained from dbCRID and TICdb. Data concerning functional elements, such as domains from Pfam (13), DDIs from 3DID (14), PPIs from both of BioGrid (15) and Metacore, were also used. With respect to regulatory elements, TFs are obtained from PAZAR (16), miRNA targets from MirTarBase (17), karyotypes and types of cancer from OMIM (18), and mRNA and protein sequences from the Refseq database. The first break point (FBP) and second break point (SBP) in the exon/intron are retrieved from the UCSC genome browser (19) using the pairwise alignment tool, BLAT (20). Information about sub-cellular localization and the tissue specificity of wild-type proteins is obtained from Swiss-Prot (21). Table 1 presents statistics concerning the input data.

### Implementation

FARE-CAFE provides a user-friendly interface that can be utilized by researchers to search for cancer-specific fusion events and information about their functional and regulatory elements, recorded using MySQL. The graphical web interface was built using the PHP language.

### Query interface

FARE-CAFE provides a web-based interface with which users query and access information about cancer-specific FGs. Query keywords include cancer types, FGs, names of 5' proteins and names of 3' proteins, all of which can be



**Figure 1.** Pictorial representation of TFs, miRNA target, differential domains, DDIs and PPIs in FG. **(A)** Wild-type 5' gene and 3' gene with their TFs, miRNAs, domains, DDIs and PPIs. **(B)** FG associates with a different set of (i) TFs, i.e. TFa and TFb; (ii) miRNA targets (miRT) at 3'-UTR region, i.e. miRT3 and miRT4; (iii) domains denoted by Da, Db, Dy and Dz; (iv) DDIs denoted as Da-D1, Db-D2, Dy-D5 and Dz-D6; and (v) DDI-mediated PPI partners, i.e. P1, P2, P5 and P6.

selected using a pull down menu. For instance, if the user selects an FG as a query option, then the database will return cancer types, genomic, functional, regulatory and sequence features on the output page. The user can search on keywords that specify cancer type, or 5' and 3' gene names to access cancer-associated data and information about wild-type proteins functional (domains, DDIs and PPIs) and regulatory elements (TFs and miRNA targets). Apart from the search page, the other web pages provide essential information, such as database statistics, pictorial representations of the workflow, TFs, miRNAs, domains, disrupted domains, DDIs, PPIs, case studies and downloadable files that are related to the FGs in the FARE-CAFE database.

### Case studies

To demonstrate the usefulness of FARE-CAFE in identifying the role of regulatory and functional elements of FGs in cancer progression, three case studies are examined; these case studies will increase the impacts of our database in cancer biology research.

### Results and discussion

The FARE-CAFE database includes a comprehensive collection of FGs, CT events and their essential genetic elements which are necessary to understand the mechanism of cancer progression, and are not provided by most existing databases. FARE-CAFE includes 518 FGs, 1587 FG isoforms, 122 major types and subtypes of cancers, 270 five-prime proteins and 295 three-prime proteins annotated with key genetic elements that are involved in cancer. All existing CT-related databases provide only gene level or partial genomic level information. Each of these databases has distinct strengths and uses. A comparison of FARE-CAFE to other database resources is given in Table 2.

FARE-CAFE is the only database that provides information about TFs, miRNAs, domains, DDIs and PPIs for FGs. A snapshot of the search and result pages of FGs is depicted in Figure 3. The missing TF, miRNA, domain and DDI components of FGs are determined by comparison with wild-type 5' and 3' gene partners. The significance of the presented database is demonstrated using three case studies, which illustrate the role of the regulatory

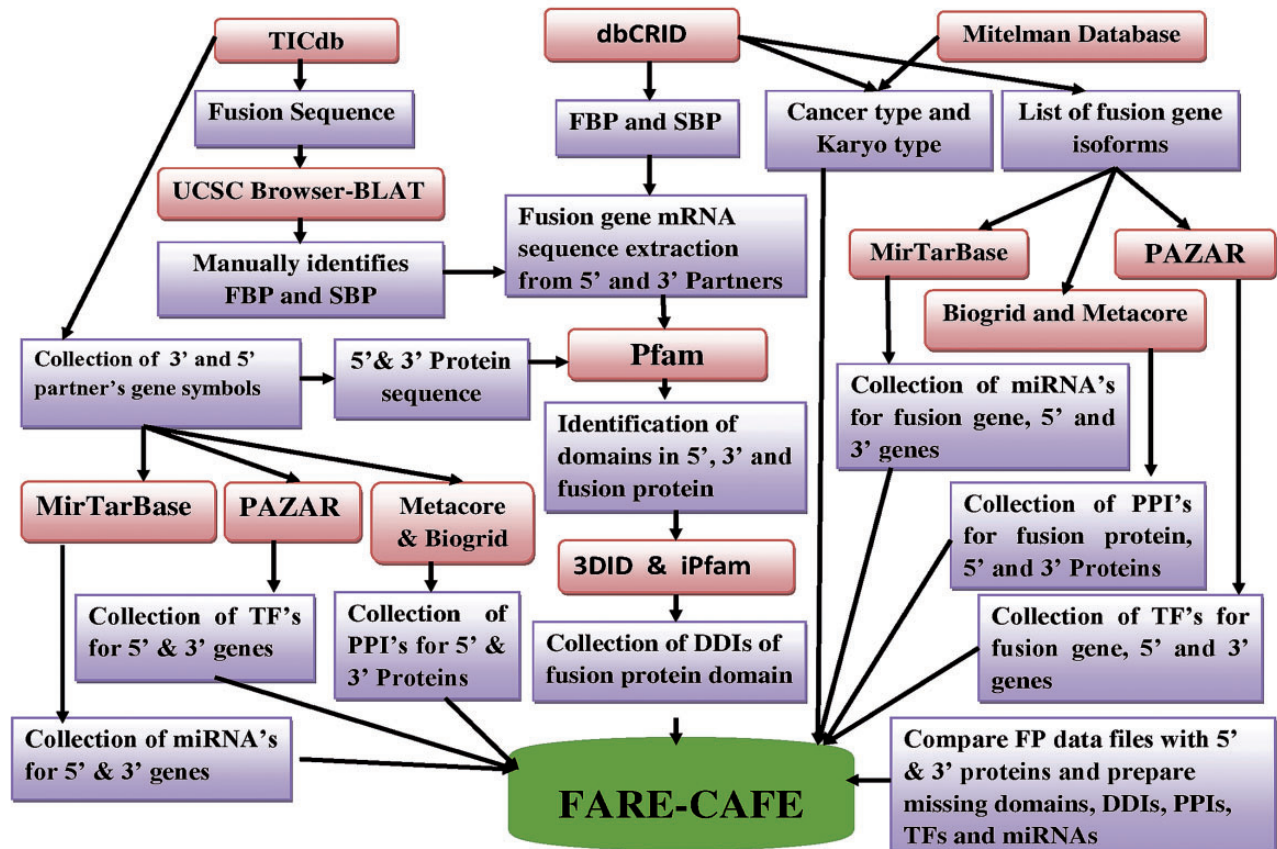


Figure 2. FARE-CAFE methodology workflow.

Table 1. FARE-CAFE database statistics

FARE-CAFE Database	Number of entities
Total number of FGs	518
Total number of FG isoforms	1587
Total number of 5' gene partners	270
Total number of 3' gene partners	295
Major types and sub-types of cancers	122

elements and functional elements in cancer formation by CTs.

### Case study 1: Impact of TFs on FG-induced cancer progression

FGs with the following features were collected: (i) the 5' gene was transcribed by TFs (using TF information from the PAZAR database and activity information from Metacore database) (Figure 4a), (ii) the 3' gene is oncogenic (from NCG 4.0) (22), (iii) the 3' gene expression is controlled by a transcription repressor (from the Metacore database) and (iv) the 3' gene is not targeted by miRNAs (Figure 4b). Figure 4c displays the FG structure. A total of 14 FGs satisfy the specified constraints (Supplementary Table S2).

### Observations and analysis

- The FG is transcribed by various TFs, which bind the 5' gene promoter region.
- After translocation, the 3' oncogene lost its TF binding sites for the transcription repressor and fused with the 5' gene.
- miRNA does not target the 3' partner gene of the FG (e.g. *BCL2-IGH*), so the 3' gene-specific transcription repressor or miRNA does not suppress this kind of FG (e.g. all FGs reported in this case study). The *ETS1* (from Metacore) acts as a transcription repressor for the wild-type 3' gene but a transcription enhancer for two FGs in which *ETS1* activates the 5' gene—*PAX5-JAK2* (23, 24) and *RUNX1-CBFA2T3* (25, 26). Another TF, EBF, is found to act on the FG, *PAX5-ZNF521* (23). Here, the EBF activates the wild-type 5' gene *PAX5*, but represses the wild-type 3' gene *ZNF521* (from Metacore).

In summary, owing to CT, the wild-type 3' oncogene lost its transcription inhibition site and fused with the 5' gene. This FG (all FGs reported in this case study) may exhibit oncogenic characteristics after translocation. The 3'

**Table 2.** A comparison of FARE-CAFE and other related databases

Features	TICdb	ChimerDB	dbCRID	FARE-CAFE
FGs with break points information	✓	✓	✓	✓
Fusion sequence/Junction sequence	✓	✓	✓	✓
FGs identified methods documented	—	—	✓	✓
Manually identified exact break point positions	—	—	Partially	✓
Cancer type information	—	✓	✓	✓
List of domains in fusion protein	—	—	—	✓
DDIs of fusion protein domains	—	—	—	✓
PPIs of fusion protein	—	—	—	✓
5' and 3' partner proteins domains, DDIs and PPIs information	—	—	—	✓
Differential miRNAs at FG—3'-UTR	—	—	—	✓
Differential TF at FG—5'-end	—	—	—	✓
FG break points located in exon or in intron region documented	—	—	—	✓
5' and 3' proteins sub-cellular localization and tissue specificity	—	—	—	✓

gene fragment does not lose its oncogenic nature because it utilizes the 5' gene promoter for transcription.

### Case study 2: Impact of miRNAs on FG-induced cancer progression

FGs with the following two features were obtained from miRTarbase and NCG 4.0: (i) the 5' gene is oncogenic with experimentally verified miRNAs that are targeted at its 3' UTR (Figure 5a) and (ii) the 3' gene is not targeted by any miRNA (Figure 5b). To address feature (ii), three miRNA target prediction algorithms—miRDB (27), MiRanda (28) and TargetScan (29)—were utilized to determine whether the 3' UTR of the 3' gene was targeted by any miRNA or not. Then, the predicted miRNAs were compared with the miRNAs of the 5' gene. If any one of the predictions made using the algorithm matched the 5' gene miRNA, then a score (S) of 1 was assigned to the prediction, so the maximum score was 3. To improve the reliability of the prediction, the predicted miRNAs were predicated against the miRCancer database (30) to determine whether those miRNAs had been recorded. MiRCancer is a database of cancer-related miRNAs. Table 3 summarizes the four possible outcomes of this process.

A total of 33 FGs that satisfied the above constraints (Figure 5c) were found, about which detailed information can be found in Supplementary Table S3. Notably, if a 3' gene is not targeted by any miRNA and its host FG is activated, then that FG is more likely to have an oncogenic role than wild-type 3' gene.

### Observations and analysis

After CT, all of the 33 collected FGs lost the 3'-UTR regions from their 5' oncogenes, and so were not regulated by any miRNA.

- The 5' oncogenes fused with 3' genes, which were not targeted by any miRNA.
- A majority (more than 95%) of the 3' genes predicted miRNAs were not shared by the 5' genes. The remaining 5% were overlapping miRNAs.
- The predictions are supported by the findings that over 90% of the 3' genes' predicted miRNAs were not associated with any record in miRCancer. The other 10% miRNAs were identified in mirCancer.

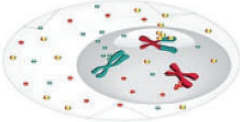
Owing to CT, the 5' oncogenes of the 33 FGs lost their miRNA sites and fused with 3' partner genes that were not targeted by any miRNA. As the FGs did not have any accessible site that was targeted by miRNAs, those FGs may have had a key role in enhancing cell proliferation and inhibited apoptosis, which leads to cancer progression.

Our case study is strengthened by the work of Shugay *et al.* (31). The previous work reported that 3' translocation partner genes (TPGs) contain significantly shorter 3'-UTRs than 5' TPGs. Hence, there were fewer regulatory elements and miRNA target sites in the 3'-UTRs of 3' TPGs, thus disrupting FGs from miRNA-mediated gene repression (32). Gomez-Benito *et al.* (33), shown that substituting the 3'-UTR of the *MLL* gene with the 3'-UTR of its fusion partner, resulted in significantly post-transcriptional inhibition of the expression of *MLL*, which leads to *MLL*-mediated CT events.

### Case study 3: Impact of functional elements on FG-induced cancer progression

The impacts of the functional elements—domains, DDIs and PPIs—of the fusion proteins in acute myeloid leukemia (AML), chronic myeloid leukemia (CML) and Ewing's sarcoma cancer progression are presented in the following. A total of 38 fusion proteins with 203 isoforms are

**A**



## FARE-CAFE

*A Database of Functional and Regulatory Elements of Cancer-Associated Fusion Events*

Home	About Database	Search	Case studies	Download	Help	Glossary	Contact
------	----------------	--------	--------------	----------	------	----------	---------

**Search cancer genomic information**

Select one of the cancer type from drop down menu and submit query to get cancer genomic information

Acute basophilic leukemia ▼ **by Cancer type**

**Search differential elements of fusion gene**

Select one of the fusion gene from drop down menu and submit query to get Cancer genomic, functional and regulatory information of fusion gene.

ABI1-MLL ▼ **by Fusion gene**

**Search genomic, functional and regulatory elements of wild type 5' Protein**

Select one of the wild type 5' protein name from drop down menu and submit query to get Genomic, functional and regulatory elements information of query protein

ABI1 ▼ **by 5' Protein**

**Search genomic, functional and regulatory elements of wild type 3' Protein**

Select one of the wild type 3' protein name from drop down menu and submit query to get Genomic, functional and regulatory elements information of query protein

ABI1 ▼ **by 3' Protein**

**B Query fusion gene:ABI1-MLL**

**Fusion gene cancer genomic information**

- Genomic Information**

Fusion_Protein	Fusion_Protein_Isoforms	5'_Protein	F.B.P	FBP_in_Exon/Intron	5'_Protein_on_Positive/Negative_Strands	3'_Protein	S.B.P	SBF
ABI1-MLL	ABI1-MLL	ABI1	27112096	1	0	MLL	118354896	0

**Note:**  
 FP=Fusion protein; FBP= First break point ; SBP= Second break point  
 0 = FBP in intron; 0 = SBP in intron; 0 = 5' Protein on negative strands; 0 = 3' Protein on negative strands  
 1 = FBP in exon; 1 = SBP in exon; 1 = 5' Protein on positive strands; 1 = 3' Protein on positive strands

- Cancer related Information**

Fusion_Protein	Cancer_Type	FP_Experimental_Method	FP_Karyotype	FP_Reference
ABI1-MLL	Acute myelomonocytic leukemia (FAB type M4)	PCR,DNA sequencing,RT-PCR	t(10;11)(p12;q23)	PubMed: <a href="#">11477655</a>

**Figure 3.** FARE-CAFE search and result page. (a) Search for ABI1-MLL fusion gene. (b) Result page with cancer genomic elements information of ABI1-MLL fusion gene. (c) Result page with functional elements information of ABI1-MLL fusion gene. (d) Result page with regulatory elements information of ABI1-MLL fusion gene.

associated with the three selected types of cancer. [Supplementary Table S4](#) presents the functional elements of those fusion proteins and their wild-type 5' and 3' partner proteins.

### Observations and analysis

- After translocation, fusion protein, which is composed of a smaller group of domains than the union of its 5' and 3' wild-type genes, is formed.

**C**

### Fusion gene functional elements information

- Domains and DDI Information**

Fusion_Protein	Fusion_Protein_Isoforms	Domains_in_5'Partner	DDIs_for_Domains_in_5'Partner
ABI1-MLL	ABI1-MLL	1) <a href="#">Abi_HHR</a> 2) <a href="#">SH3_1</a>	1) <a href="#">FragX_IP</a> , <a href="#">Nckap1</a> 2) <a href="#">SH3_1</a> , <a href="#">Ank_2</a> , <a href="#">P53</a> , <a href="#">ubiquitin</a> , <a href="#">Asp</a> , <a href="#">PB1</a> , <a href="#">SH3_2</a> , <a href="#">F-protein</a> , <a href="#">Pkinase_Tyr</a> , <a href="#">Trypsi</a>

- PPI Information**

Fusion_Protein	FP_PPI_partner	FP_PPI_Effect	FP_PPI_Mechanism	FP_PPI_Reference
ABI1-MLL	-	-	-	-

### Fusion gene regulatory elements information

**D**

- MiRNA Information**

Fusion_Protein	miRNAs_Targets_Fusion_protein	Missing_MiRNAs_Targets_Fusion_protein	FP_miRNA_Validation_method
ABI1-MLL	<a href="#">hsa-miR-128-3p</a>	-	-

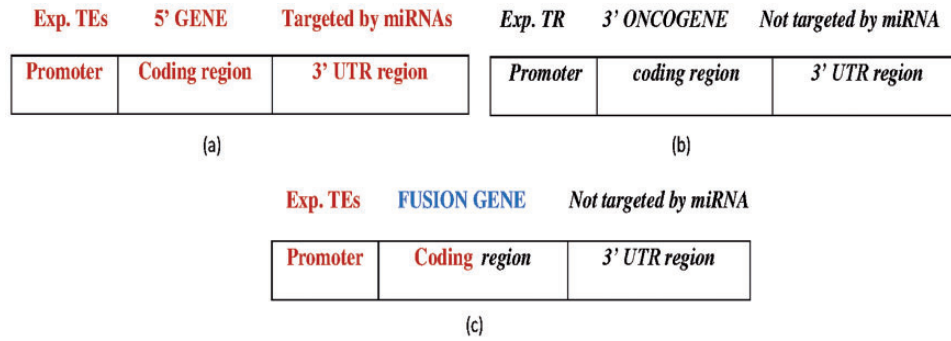
- Transcription Factors Information**

Fusion_Protein	5'P_Ref_mRNA_seqID	FP_Transcription_factors	FP_TF_Reference	FP_Missing_TFs
ABI1-MLL	<a href="#">NM_001012750.2</a>	1) <a href="#">TRIM28_HUMAN</a> 2) <a href="#">STAT1_HUMAN</a> 3) <a href="#">RBL_HUMAN</a> 4) <a href="#">RBL2_HUMAN</a> 5) <a href="#">ZNF263_HUMAN</a> 6) <a href="#">E2F4_HUMAN</a> 7) <a href="#">ETS1_HUMAN</a> 8) <a href="#">EGR1_HUMAN</a> 9) <a href="#">FOS_HUMAN</a> 10) <a href="#">CTCF_HUMAN</a> 11) <a href="#">BRG1_HUMAN</a> 12) <a href="#">TFAP2C_HUMAN</a> 13) <a href="#">FOXA1_HUMAN</a> 14) <a href="#">SPI1_HUMAN</a> 15) <a href="#">TRIM24_HUMAN</a>	1) <a href="#">17542650</a> 2) <a href="#">19122651</a> 3) <a href="#">20385362</a> 4) <a href="#">20385362</a> 5) <a href="#">19887448</a> 6) <a href="#">21247883</a> 7) <a href="#">20019798</a> 8) <a href="#">20690147</a> 9) <a href="#">20139302</a> 10) <a href="#">21795385</a> 11) <a href="#">19505939</a> 12) <a href="#">21572391</a> 13) <a href="#">21572391</a> 14) <a href="#">21241896</a> 15) <a href="#">21164480</a>	-

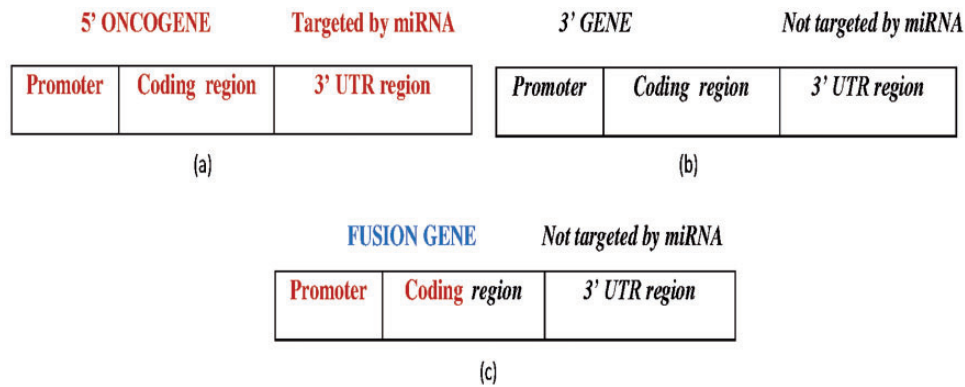
Figure 3. Continued.

- As the fusion protein is composed of a smaller set of domains, certain biologically significant DDIs and DDI-mediated PPIs are disrupted.
- For instance, the *BCR-ABL* fusion protein (which causes AML and CML) interacts with the *GAB2* (which contains a PH domain) and *H-Ras* (which contains a Ras

domain) proteins via DDI with the RhoGEF domain (which contains in *BCR*). These PPIs play a key role in the ‘CML pathway’ (KEGG ID: hsa05220). Also, *BCR-ABL* interacts with the *CRK* and *STAT5A* proteins (with the SH2 domain in *CRK* and *STAT5A*) via DDI with the C2 domain (which contains in *BCR*). These PPIs also



**Figure 4.** (a) The wild-type 5' gene is activated by experimentally reported transcription enhancers (Exp.TEs) at the promoter region and with experimentally reported targeted miRNAs at the 3' UTR region. (b) The wild-type 3' oncogene inhibited by experimentally reported transcription repressor (Exp. TR) at the promoter region and not targeted by miRNA at the 3'-UTR region. (c) The FG composed of 5' gene transcribed by TEs at the promoter region, and coding region forms from the fusion of the 5' gene and 3' oncogene which is not targeted by miRNA at the 3'-UTR region. Bold-faced and italic fonts denote the wild-type 5' gene and 3' gene, respectively.



**Figure 5.** (a) The wild-type 5' oncogene targeted by experimentally reported miRNAs at the 3'-UTR region. (b) The wild-type 3' gene not targeted by any miRNA. (c) The FG with a 5' oncogene promoter, coding region fragment forms from the fusion of the 5' oncogene and 3' gene, where the 3' gene is not targeted by any miRNA. Bold-faced and italic fonts denote the wild-type 5' gene and 3' gene, respectively.

**Table 3.** The four possible outcomes based on the results of score *S* and miRCancer

Score, <i>S</i>	miRNA found in miRCancer	Hypothesis
$S \geq 1$	No	miRNA has a low possibility of regulating the FG
$S \geq 1$	Yes	miRNA has high possibility of regulating the FG
$S = 0$	No	miRNA has a very low possibility of regulating the FG
$S = 0$	Yes	miRNA has a low possibility of regulating the FG

play a role in the ‘CML pathway’. However, certain *BCR-ABL* fusion isoforms (34, 35) lost the RhoGEF and C2 domains, disrupting the ‘CML pathway’.

- With respect to Ewing’s sarcoma cancer, the RRM\_6 domain in protein *EWSR1* interacts with the *SF3B4* and *SRSF5* proteins; this interaction is presented in the ‘Spliceosome pathway’ (hsa03040). The PPI is mediated by the DDI between domains RRM\_6 and RRM\_1 (which contains in *SF3B4* and *SRSF5* proteins). Also, the *zf-RanBP* domain in *EWSR1* interacts with the *RAD23A* and *HERPUD1* proteins. This interaction is recorded in

‘Protein processing in endoplasmic reticulum pathway’ (hsa04141).

The PPI is mediated by the DDI between domains *zf-RanBP* and ubiquitin (which contains in *RAD23A* and *HERPUD1* proteins). After translocation, *EWSR1* lost both RRM\_6 and *zf-RanBP* domains and fused with *FLI1* to form the oncogenic fusion protein *EWS-FLI1*, causing Ewing’s sarcoma. [For more details, see ‘Sarcomas’ under ‘Transcriptional Mis-regulation in Cancer’ (hsa05202).]

Relative to their 5' and 3' gene partners, the fusion proteins lost some essential functional elements, such as



domains (Supplementary Table S4), and subsequently disrupting a few DDIs or DDI-mediated PPIs, interrupting certain biological processes that may play crucial roles in cancer formation.

Case study 3 is in compatible with the work by Shugay *et al.* Both works have certain aspects in common, for example, their work made use of protein interaction interfaces information of a TPG to list its interaction partners, whereas our work employed DDI and PPI databases to achieve this goal.

It is interesting to note that Shugay *et al.* classified all protein domains into five functional classes and studied the co-occurrence of any class using Fisher exact test. It was also found that 5' fusion partner contains domain with non-oncogenic properties could contribute a strong promoter and 3' fusion partner associates with oncogenic domains might stabilize the fusion mRNA by contributing its 3'-UTR.

### Future works for the FARE-CARE database

In future version of FARE-CAFE, data obtained from four databases, i.e. COSMIC (36), Mitelman (7), TCGA Fusion Gene Data Portal database (37) and ChimerDB, will be incorporated.

COSMIC is the largest and the most comprehensive resource for somatic mutations in human cancers. The Mitelman database provides huge collection of FG events and their tumor characteristics information. The TCGA Fusion Gene Data Portal database provides a list of fusion events with clinical relevance that have not been previously recognized in various cancer types. This database group detects fusion transcripts based on integrated analysis of RNA sequencing. ChimerDB is one of the large scale fusion events database with huge number of fusion transcripts annotated with mRNA and EST and NGS fusion transcripts information.

We will utilize the above-mentioned resources and include additional information in FARE-CAFE. For instance, new/novel FGs (from Mitelman database and TCGA Fusion Gene Data Portal database) with exon annotations, CDS mutation sites, mutation types, mutation frequency and tissue-specific information for FGs (from COSMIC), mRNA sequence and EST sequence (from ChimerDB, but it is not update since 2010), which are crucial for cancer diagnosis and therapy. Furthermore, FARE-CAFE will provide RNA sequencing information of cancer samples by linking to the TCGA Fusion Gene Data Portal database.

We anticipate to regularly update FARE-CAFE once new versions of the utilized databases are released.

## Summary

FARE-CAFE is the first comprehensive database of extensive annotated oncogenic FG events, including, genomic, regulatory and function elements, which play crucial roles in cancer formation. Overall, this database enables the finding of relevant information (TF, miRNA, domain, DDI and PPI) about cancer-associated FGs at various levels (pre-transcription, post-transcription and post-translation) of regulations and serves as a unique resource for researchers in the field of the cancer biology.

## Supplementary Data

Supplementary data are available at Database Online.

## Acknowledgements

We thank Professor Jim Sheu for his support on providing the use of the Metacore database. Ted Knoy is appreciated for his editorial assistance.

## Funding

Ministry of Science and Technology (MOST) of the Republic of China, Taiwan; Grant of MOST 102-2632-E-468-001-MY3 (to K.-L.N. and J.J.P.T.); Asia University (103-asia-06) and China Medical University (CMU103-ASIA-21) to K.-L.N.; Grant of MOST 102-2632-B-039-001-MY3 (to J.C.); Grant of MOST 104-2221-E-150-039 (to C.-H.H.); National Formosa University (EN103B-1005) to C.-H.H.; China Medical University (CMU103-ASIA-21) to W.-T.H. and H.-Y.C. Funding for open access charge: CMU103-ASIA-21.

*Conflict of interest.* None declared.

## References

- Mitelman,F., Johansson,B. and Mertens,F. (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*, 7, 233–245.
- Rabbitts,T.H. (1994) Chromosomal translocations in human cancer. *Nature*, 372, 143–149.
- Nambiar,M., Kari,V. and Raghavan,S.C. (2008) Chromosomal translocations in cancer. *Biochim. Biophys. Acta*, 1786, 139–152.
- Kong,F., Zhu,J., Wu,J. *et al.* (2011) dbCRID: a database of chromosomal rearrangements in human diseases. *Nucleic Acids Res.*, 39, D895–D900.
- Look,A.T. (1997) Oncogenic transcription factors in human acute leukemias. *Science*, 278, 1059–1064.
- Rowley,J.D. (2001) Chromosomal translocations: dangerous liaisons revisited. *Nat. Rev. Cancer*, 1, 245–250.
- Mitelman,F., Johansson,B. and Mertens,F. (2014) *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*. <http://cgap.nci.nih.gov/Chromosomes/Mitelman>
- Novo,F.J., de Mendibil,I.O. and Vizmanos,J.L. (2007) TICdb: a collection of gene-Mapped translocation breakpoints in cancer. *BMC Genomics*, 8, 33.

9. Kim,D.S., Huh,J.W. and Kim,H.S. (2007) HYBRIDdb: a database of hybrid genes in the human genome. *BMC Genomics*, **8**, 128.
10. Kim,P., Yoon,S., Kim,N. *et al.* (2010) ChimerDB 2.0—a knowledgebase for fusion genes updated. *Nucleic Acids Res.*, **38**, D81–D85.
11. Iida,K., Fukushi,J., Matsumoto,Y. *et al.* (2013) miR-125b develops chemo resistance in Ewing sarcoma/primitive neuroectodermal tumor. *Cancer Cell Int.*, **3**, 21.
12. Xu,C., Fu,H., Gao,L. *et al.* (2014) BCR-ABL/GATA1/miR-138 mini circuitry contributes to the leukemogenesis of chronic myeloid leukemia. *Oncogene*, **33**, 44–54.
13. Finn,R.D., Bateman,A., Jody,C. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
14. Mosca,R., Ceol,A., Stein,A. *et al.* (2014) 3did: a catalog of domain-based interactions of known three-dimensional structures. *Nucleic Acids Res.*, **42**, D374–D379.
15. Chatr-Aryamontri,A., Breitkreutz,B.J., Heinicke,S. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.
16. Portales-Casamar,E., Arenillas,D., Lim,J. *et al.* (2009) The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res.*, **37**, D54–D60.
17. Hsu,S.D., Tseng,Y.T., Shrestha,S. *et al.* (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.*, **42**, D78–D85.
18. Amberger,J., Bocchini,C.A., Scott,A.F. *et al.* (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **37**, D793–D796.
19. Kuhn,R.M., Karolchik,D., Zweig,A.S. *et al.* (2009) The UCSC Genome Browser Database: update-2009. *Nucleic Acids Res.*, **37**, D755–D761.
20. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
21. Boeckmann,B., Bairoch,A., Apweiler,R. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **1**, 365–370.
22. An,O., Pendino,V., D'Antonio,M. *et al.* (2014) NCG 4.0: the network of cancer genes in the era of massive mutational screenings of cancer genomes. *Database.*, 1–10.
23. Wheat,W., Fitzsimmons,D., Lennox,H. *et al.* (1999) The highly conserved beta-hairpin of the paired DNA-binding domain is required for assembly of Pax-Ets ternary complexes. *Mol. Cell Biol.*, **19**, 2231–2241.
24. Hollenhorst,P.C., Chandler,K.J., Poulsen,R.L. *et al.* (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet.*, **5**, 1–12.
25. Mao,S., Frank,R.C., Zhang,J. *et al.* (1999) Functional and physical interactions between AML1 proteins and an ETS protein, MEF: implications for the pathogenesis of t(8;21)-positive leukemias. *Mol. Cell Biol.*, **19**, 3635–3644.
26. Ajore,R., Kumar,P., Dhanda,R.S. *et al.* (2012) The leukemia associated nuclear corepressor ETO homologue genes MTG16 and MTGR1 are regulated differently in hematopoietic cells. *BMC Mol. Biol.*, **13**, 1–16.
27. Wong,N. and Wang,X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**(D1), D146–D152.
28. Betel,D., Wilson,M., Gabow,A. *et al.* (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
29. García,D.M., Baek,D., Shin,C. *et al.* (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other miRNAs. *Nat. Struct. Mol. Biol.*, **18**, 1139–1146.
30. Xie,B., Ding,Q., Han,H. *et al.* (2013) miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*, **29**, 638–644.
31. Shugay,M., Ortiz de Mendivil,I., Vizmanos,J.L. *et al.* (2012) Genomic hallmarks of genes involved in chromosomal translocations in hematological cancer. *PLoS Comput. Biol.*, **8**, e1002797.
32. Mayr,C., Hemann,M.T., Bartel,D.P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*, **315**, 1576–1579.
33. Gomez-Benito,M., Loayza-Puch,F., Oude Vrielink,J. *et al.* (2011) 3' UTR-mediated gene silencing of the mixed lineage leukemia (MLL) gene. *PLoS One*, **6**, e25449.
34. Iwata,S., Mizutani,S., Nakazawa,S. *et al.* (1994) Heterogeneity of the breakpoint in the ABL gene in cases with BCR/ABL transcript lacking ABL exon a2. *Leukemia*, **8**, 1696–1702.
35. Van der Velden,V.H., Beverloo,H.B., Hoogveen,P.G. *et al.* (2007) A novel BCR-ABL fusion transcript (e18a2) in a child with chronic myeloid leukemia. *Leukemia*, **21**, 833–835.
36. Forbes,S.A., Beare,D., Gunasekaran,P. *et al.* (2014) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
37. Yoshihara,K., Wang,Q., Torres-Garcia,W. *et al.* (2014) The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*. doi: 10.1038/onc.2014.406