

ChiTaRS-3.1—the enhanced chimeric transcripts and RNA-seq database matched with protein–protein interactions

Alessandro Gorohovski, Somnath Tagore, Vikrant Palande, Assaf Malka, Dorith Raviv-Shay and Milana Frenkel-Morgenstern*

Faculty of Medicine in Galilee, Bar-Ilan University, Henrietta Szold 8, Safed 13195, Israel

Received October 02, 2016; Revised October 26, 2016; Editorial Decision October 26, 2016; Accepted October 30, 2016

ABSTRACT

Discovery of chimeric RNAs, which are produced by chromosomal translocations as well as the joining of exons from different genes by trans-splicing, has added a new level of complexity to our study and understanding of the transcriptome. The enhanced ChiTaRS-3.1 database (<http://chitars.md.biu.ac.il>) is designed to make widely accessible a wealth of mined data on chimeric RNAs, with easy-to-use analytical tools built-in. The database comprises 34 922 chimeric transcripts along with 11 714 cancer breakpoints. In this latest version, we have included multiple cross-references to GeneCards, iHop, PubMed, NCBI, Ensembl, OMIM, RefSeq and the Mitelman collection for every entry in the ‘Full Collection’. In addition, for every chimera, we have added a predicted chimeric protein–protein interaction (ChiPPI) network, which allows for easy visualization of protein partners of both parental and fusion proteins for all human chimeras. The database contains a comprehensive annotation for 34 922 chimeric transcripts from eight organisms, and includes the manual annotation of 200 sense-antiSense (SaS) chimeras. The current improvements in the content and functionality to the ChiTaRS database make it a central resource for the study of chimeric transcripts and fusion proteins.

INTRODUCTION

Chimeric RNAs comprise sequences deriving from more than one transcription event. Fusion can occur at either the genomic level as the result of chromosomal rearrangement, or at the RNA level when two different transcripts are combined through a complex trans-splicing process (1–24). While many chimeric transcripts have been shown to be artifacts of *in vitro* reverse transcription reactions (25–32),

recent studies clearly demonstrate that some (mostly cancer chimeric transcripts) are translated into chimeric proteins (11,16,18). Here, we expand our previously published collection of putative chimeric transcripts (ChiTaRS) that includes chimeras whose RNA expression levels have been verified by RNA-sequencing and whose translation into protein products has been shown previously by us, using mass-spectrometry analyses (33,34) by predicted protein–protein interaction networks.

Translation of chimeric transcripts into a fusion protein has been shown to dramatically alter the protein–protein interaction (PPI) networks of the two parental proteins that comprise the fusion. We have built a computational tool for analyzing changes to the PPI networks of chimeric (or ‘fusion’) proteins, called ‘ChiPPI’ (Chimeric PPI), which we have incorporated into the ChiTaRS database, providing a pre-calculated analysis for every human fusion event (<http://chitars.md.biu.ac.il>, see ‘Full Collection’). Using a methodology that treats discrete protein domains as building blocks of interacting proteins, we have catalogued the protein interaction networks for all the chimeric proteins in ChiTaRS. The ChiPPI method (<http://chippi.md.biu.ac.il/>) is unique in that it incorporates the protein domain-domain co-occurrence scores in order to identify interactors of chimeric proteins. Today, the ChiTaRS-3.1 database of ‘Chimeric Transcripts and RNA-Seq data’ is a collection of 34 922 chimeric transcripts identified by Expressed Sequence Tags (ESTs) and mRNAs from the GenBank (35), ChimerDB (26,36), dbCRID (37), TICdb (38) and the Mitelman collection of cancer fusions (39–42) for *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Rattus norvegicus*, *Bos taurus*, *Danio rerio*, *Saccharomyces cerevisiae* and *Sus scrofa* organisms. All the improvements in content, accessibility, usability and functionality (explained below), place ChiTaRS-3.1 as one of the major, up-to-date resources for the study of chromosomal and trans-splicing alterations in cancer.

*To whom correspondence should be addressed. Tel: +972 72 264 4901; Fax: +972 72 264 4901; Email: milana.morgenstern@biu.ac.il

Table 1. The major improvements and data additions in ChiTaRS-3.1 in comparison to ChiTaRS-2.1.

Content	ChiTaRS-2.1	ChiTaRS-3.1	Relevance
The collection of chimeric transcripts	29 500 (total), 20 753 (<i>Homo sapiens</i>), 6226 (<i>Mus musculus</i>), 2151 (<i>D. melanogaster</i>), 4 (<i>Bos taurus</i>), 8 (<i>Rattus norvegicus</i>), 4 (<i>Denio rerio</i>), 5 (<i>S. cerevisiae</i>), 13 (<i>Sus scrofa</i>)	34 922 (total), 24 608 (<i>Homo sapiens</i>), 7457 (<i>Mus musculus</i>), 2740 (<i>D. melanogaster</i>), 6 (<i>Bos taurus</i>), 10 (<i>Rattus norvegicus</i>), 7 (<i>Denio rerio</i>), 5 (<i>S. cerevisiae</i>), 89 (<i>Sus scrofa</i>)	We extended the collection for all eight organisms by ~4500 new entries.
Cancer Breakpoints	1280	11 714 including 69 SaS chimeras (634 FASTA sequences of chimeras)	Bona-fide expression of unique cancer-restricted fusion transcripts extended by more than 10 000 new entries.
Chimeric protein–protein interaction (ChiPPI) networks	No	2081 (validated), 22 527 (predicted)	We added pre-computed ChiPPI networks for every human entry in ‘Full Collection’ and ‘Breakpoints’
Manual annotation of Sense-antiSense (SaS) chimeras	No	200	We have mapped the unique properties of SaS chimeras.
GeneCards, iHop, PubMed, NCBI, Ensembl, OMIM, RefSeq, Mitelman	No	33 124	More than 30 000 links to the extended description for every entry in Full Collection.

IMPROVEMENTS

The major updates and improvements to the content and functionality of ChiTaRS are summarized in Table 1 and Supplementary Table S1. The improvements include: the addition of >4500 chimeric transcripts from eight organisms, and >10 000 cancer breakpoints; prediction of Chimeric protein–protein interaction (ChiPPI) networks, manual annotation of Sense-antiSense (SaS) chimeras, newly added automatic annotation and links to UniProt (43), GeneCards (44), iHop (45), GeneBank (35), Ensembl (46), OMIM (47), RefSeq (48) and the Mitelman collection (39) for every entry in the ‘Full Collection’ (Figure 1, The ChiTaRS-3.1 Interface Screen-shot).

Updated database content

In the current 2016 update, 34 922 chimeric transcripts have been collated from eight organisms (Table 1). We have identified and annotated 11 714 cancer breakpoints from the recent study of Merten *et al* and from the Mitelman collection (39–42). To study all these cancer fusions (see ‘Breakpoints’ collection), we have performed manual confirmation of their veracity using the information from >7700 PubMed articles and >19 000 iHop links (Table 1 and Figure 1). Malignancies with the most frequently found fusions are Adenocarcinoma (6308 fusions, ChiTaRS-3.1), Chronic Leukemia (1140 fusions), Acute Lymphoblastic Leukemia (2078 fusions), and Acute Myeloid Leukemia (AML) (135 fusions) (Supplementary Table S2). ChiTaRS-3.1 consists of 435 chimeric transcripts and their junction sites that have been confirmed by RNA-seq datasets (the Human Body Map dataset analyses from (18)), and 77 chimeras have been confirmed by the mass-spec experiments (18,33,34). Finally, for all the Breakpoints collection, the website tool has been greatly improved to provide a user-friendly interface (see ‘Breakpoints’, and Supplementary Figure S1).

To make the ChiTaRS-3.1 collection the most comprehensive source of chimeric transcripts available, we regularly update the list of chimeras deriving from the GenBank collection of ESTs and mRNAs for *H. sapiens* (UCSC reference genome: GRCh37/hg19), *M. musculus*

(NCBI37/mm9), *D. melanogaster* (BDGP R5/dm3), *R. norvegicus* (RGSC Rnor_6.0/rn6), *B. taurus* (Baylor College of Medicine HGSC Btau_4.6.1/bosTau7), *D. rerio* (Sanger Institute Zv9/danRer7), *S. cerevisiae* (SGD April 2011 sequence/sacCer3) and *Sus scrofa* (Broad/Pig3) (35,49). Over the past two years, we have added additional chimeric transcripts for all eight organisms: *H. sapiens* (22 810), *M. musculus* (7457), *D. melanogaster* (2740), *R. norvegicus* (10), *B. taurus* (6), *D. rerio* (7), *S. cerevisiae* (5), *S. scrofa* (89) (Table 1).

To provide biological context to fusion sequence data, we updated the GenBank (35), RefSeq (48) and Mitelman (39–42) cross-references for all the genes (total 170 797 cross-references). All the UniProt (43) references into ChiTaRS-3.1 have been updated and include now >17 000 unique proteins. Further, we added 29 643 Ensembl (46) cross-references for all gene names (Table 1). Finally, we added links to GeneCards (44), iHop (45), OMIM (47) and PubMed publications for 34 922 entries in the ‘Full Collection’ (Figure 1). Thus, ChiTaRS-3.1 is an easy-to-use resource for the in-depth study of fusion transcripts and proteins on a genome-wide, and multi-species level.

Updated database functionality

The improved user interface of ChiTaRS-3.1 allows for rapid and easy analysis of evolutionary conservation of fusions, literature references and experimental data supporting fusion expression in different organisms (see ‘Compare and Analyze’). We added a separate pop-up window with an extended annotation for every entry in the Full Collection (Figure 1, see a green button of the fusion, “EU216064”), allowing easy cross-reference to other databases (listed above).

Annotation of sense-antisense (SaS) chimeras

The phenomenon of Sense-antiSense (SaS) chimeric transcripts (34) is also covered by the ChiTaRS database in this latest version. While SaSs may result in chimeric protein translation, they also represent potential inhibitors of translation though dsRNA-mediated mechanisms (34) (Supple-

SEARCH DATABASE

You can use special characters (*, & > <) for the search by Keyword, Tissue, Gene Name, UniProtKB, Identity, etc. For example: BCR*, ABL*, AFF1 BCR* & ABL* and also NCBI ID: 611

Items per page: 10

Chimeras of exons from sense and anti-sense strands of the same open-reading frame

Order of column Numbers for GET RESULTS AS TEXT only: 1 2 3 4 5 6 7 8 9

GET RESULTS AS TEXT

Choose parameters to search by: ChiTaRS Full Collection

Dataset updates: ALL

Search filters: Rank: [] Junction Consistence: [] Sense-ANTI sense []

Organisms: Homo Sapiens Mus musculus D.Melanogaster Rattus Norvegicus Bos Taurus Danio Rerio Saccharomyces Cerevisiae Sus Scrofa All organisms

Chimeras of two adjusted genes in the genome

Known Cancer Breakpoints

Chimeric transcripts confirmed by RNA-seq reads mapping the chimeric junction sites

RESULT FOR THE SEARCH: GENE NAME/NCBI ID(S): 'BCR & ABL1'

Total sequences: 58

[1] 2 3 4 5 6 »

Organism	Fusion: Graphical View, Annotation, PPI	Head Gene (1)					Tail Gene (2)					Deviation		Rank	RNAseq and/or Mass-spec evidences	Cancer Breakpoint, PubMed, IHOP reference
		Name ₁	Start ₁	End ₁	Ident ₁ %	Strand ₁	Name ₂	Start ₂	End ₂	Ident ₂ %	Strand ₂	Eloc _(x)	Sloc _(y)			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Homo Sapiens	SplGraphs EU216064 ChiPPI	BCR A0A024R1B H0Y554 P11274 Q12843 Q12844 Q12845 Q12846 Q12847 Q12848 Q12849 Q12850 Q12851 Q12852 Q12853 Q12854 Q12855 Q12856 Q12857 Q12858 Q12859 Q12860 Q12861 Q12862 Q12863 Q12864 Q12865 Q12866 Q12867 Q12868 Q12869 Q12870 Q12871 Q12872 Q12873 Q12874 Q12875 Q12876 Q12877 Q12878 Q12879 Q12880 Q12881 Q12882 Q12883 Q12884 Q12885 Q12886 Q12887 Q12888 Q12889 Q12890 Q12891 Q12892 Q12893 Q12894 Q12895 Q12896 Q12897 Q12898 Q12899 Q12900 Q12901 Q12902 Q12903 Q12904 Q12905 Q12906 Q12907 Q12908 Q12909 Q12910 Q12911 Q12912 Q12913 Q12914 Q12915 Q12916 Q12917 Q12918 Q12919 Q12920 Q12921 Q12922 Q12923 Q12924 Q12925 Q12926 Q12927 Q12928 Q12929 Q12930 Q12931 Q12932 Q12933 Q12934 Q12935 Q12936 Q12937 Q12938 Q12939 Q12940 Q12941 Q12942 Q12943 Q12944 Q12945 Q12946 Q12947 Q12948 Q12949 Q12950 Q12951 Q12952 Q12953 Q12954 Q12955 Q12956 Q12957 Q12958 Q12959 Q12960 Q12961 Q12962 Q12963 Q12964 Q12965 Q12966 Q12967 Q12968 Q12969 Q12970 Q12971 Q12972 Q12973 Q12974 Q12975 Q12976 Q12977 Q12978 Q12979 Q12980 Q12981 Q12982 Q12983 Q12984 Q12985 Q12986 Q12987 Q12988 Q12989 Q12990 Q12991 Q12992 Q12993 Q12994 Q12995 Q12996 Q12997 Q12998 Q12999 Q13000 Q13001 Q13002 Q13003 Q13004 Q13005 Q13006 Q13007 Q13008 Q13009 Q13010 Q13011 Q13012 Q13013 Q13014 Q13015 Q13016 Q13017 Q13018 Q13019 Q13020 Q13021 Q13022 Q13023 Q13024 Q13025 Q13026 Q13027 Q13028 Q13029 Q13030 Q13031 Q13032 Q13033 Q13034 Q13035 Q13036 Q13037 Q13038 Q13039 Q13040 Q13041 Q13042 Q13043 Q13044 Q13045 Q13046 Q13047 Q13048 Q13049 Q13050 Q13051 Q13052 Q13053 Q13054 Q13055 Q13056 Q13057 Q13058 Q13059 Q13060 Q13061 Q13062 Q13063 Q13064 Q13065 Q13066 Q13067 Q13068 Q13069 Q13070 Q13071 Q13072 Q13073 Q13074 Q13075 Q13076 Q13077 Q13078 Q13079 Q13080 Q13081 Q13082 Q13083 Q13084 Q13085 Q13086 Q13087 Q13088 Q13089 Q13090 Q13091 Q13092 Q13093 Q13094 Q13095 Q13096 Q13097 Q13098 Q13099 Q13100 Q13101 Q13102 Q13103 Q13104 Q13105 Q13106 Q13107 Q13108 Q13109 Q13110 Q13111 Q13112 Q13113 Q13114 Q13115 Q13116 Q13117 Q13118 Q13119 Q13120 Q13121 Q13122 Q13123 Q13124 Q13125 Q13126 Q13127 Q13128 Q13129 Q13130 Q13131 Q13132 Q13133 Q13134 Q13135 Q13136 Q13137 Q13138 Q13139 Q13140 Q13141 Q13142 Q13143 Q13144 Q13145 Q13146 Q13147 Q13148 Q13149 Q13150 Q13151 Q13152 Q13153 Q13154 Q13155 Q13156 Q13157 Q13158 Q13159 Q13160 Q13161 Q13162 Q13163 Q13164 Q13165 Q13166 Q13167 Q13168 Q13169 Q13170 Q13171 Q13172 Q13173 Q13174 Q13175 Q13176 Q13177 Q13178 Q13179 Q13180 Q13181 Q13182 Q13183 Q13184 Q13185 Q13186 Q13187 Q13188 Q13189 Q13190 Q13191 Q13192 Q13193 Q13194 Q13195 Q13196 Q13197 Q13198 Q13199 Q13200 Q13201 Q13202 Q13203 Q13204 Q13205 Q13206 Q13207 Q13208 Q13209 Q13210 Q13211 Q13212 Q13213 Q13214 Q13215 Q13216 Q13217 Q13218 Q13219 Q13220 Q13221 Q13222 Q13223 Q13224 Q13225 Q13226 Q13227 Q13228 Q13229 Q13230 Q13231 Q13232 Q13233 Q13234 Q13235 Q13236 Q13237 Q13238 Q13239 Q13240 Q13241 Q13242 Q13243 Q13244 Q13245 Q13246 Q13247 Q13248 Q13249 Q13250 Q13251 Q13252 Q13253 Q13254 Q13255 Q13256 Q13257 Q13258 Q13259 Q13260 Q13261 Q13262 Q13263 Q13264 Q13265 Q13266 Q13267 Q13268 Q13269 Q13270 Q13271 Q13272 Q13273 Q13274 Q13275 Q13276 Q13277 Q13278 Q13279 Q13280 Q13281 Q13282 Q13283 Q13284 Q13285 Q13286 Q13287 Q13288 Q13289 Q13290 Q13291 Q13292 Q13293 Q13294 Q13295 Q13296 Q13297 Q13298 Q13299 Q13300 Q13301 Q13302 Q13303 Q13304 Q13305 Q13306 Q13307 Q13308 Q13309 Q13310 Q13311 Q13312 Q13313 Q13314 Q13315 Q13316 Q13317 Q13318 Q13319 Q13320 Q13321 Q13322 Q13323 Q13324 Q13325 Q13326 Q13327 Q13328 Q13329 Q13330 Q13331 Q13332 Q13333 Q13334 Q13335 Q13336 Q13337 Q13338 Q13339 Q13340 Q13341 Q13342 Q13343 Q13344 Q13345 Q13346 Q13347 Q13348 Q13349 Q13350 Q13351 Q13352 Q13353 Q13354 Q13355 Q13356 Q13357 Q13358 Q13359 Q13360 Q13361 Q13362 Q13363 Q13364 Q13365 Q13366 Q13367 Q13368 Q13369 Q13370 Q13371 Q13372 Q13373 Q13374 Q13375 Q13376 Q13377 Q13378 Q13379 Q13380 Q13381 Q13382 Q13383 Q13384 Q13385 Q13386 Q13387 Q13388 Q13389 Q13390 Q13391 Q13392 Q13393 Q13394 Q13395 Q13396 Q13397 Q13398 Q13399 Q13400 Q13401 Q13402 Q13403 Q13404 Q13405 Q13406 Q13407 Q13408 Q13409 Q13410 Q13411 Q13412 Q13413 Q13414 Q13415 Q13416 Q13417 Q13418 Q13419 Q13420 Q13421 Q13422 Q13423 Q13424 Q13425 Q13426 Q13427 Q13428 Q13429 Q13430 Q13431 Q13432 Q13433 Q13434 Q13435 Q13436 Q13437 Q13438 Q13439 Q13440 Q13441 Q13442 Q13443 Q13444 Q13445 Q13446 Q13447 Q13448 Q13449 Q13450 Q13451 Q13452 Q13453 Q13454 Q13455 Q13456 Q13457 Q13458 Q13459 Q13460 Q13461 Q13462 Q13463 Q13464 Q13465 Q13466 Q13467 Q13468 Q13469 Q13470 Q13471 Q13472 Q13473 Q13474 Q13475 Q13476 Q13477 Q13478 Q13479 Q13480 Q13481 Q13482 Q13483 Q13484 Q13485 Q13486 Q13487 Q13488 Q13489 Q13490 Q13491 Q13492 Q13493 Q13494 Q13495 Q13496 Q13497 Q13498 Q13499 Q13500 Q13501 Q13502 Q13503 Q13504 Q13505 Q13506 Q13507 Q13508 Q13509 Q13510 Q13511 Q13512 Q13513 Q13514 Q13515 Q13516 Q13517 Q13518 Q13519 Q13520 Q13521 Q13522 Q13523 Q13524 Q13525 Q13526 Q13527 Q13528 Q13529 Q13530 Q13531 Q13532 Q13533 Q13534 Q13535 Q13536 Q13537 Q13538 Q13539 Q13540 Q13541 Q13542 Q13543 Q13544 Q13545 Q13546 Q13547 Q13548 Q13549 Q13550 Q13551 Q13552 Q13553 Q13554 Q13555 Q13556 Q13557 Q13558 Q13559 Q13560 Q13561 Q13562 Q13563 Q13564 Q13565 Q13566 Q13567 Q13568 Q13569 Q13570 Q13571 Q13572 Q13573 Q13574 Q13575 Q13576 Q13577 Q13578 Q13579 Q13580 Q13581 Q13582 Q13583 Q13584 Q13585 Q13586 Q13587 Q13588 Q13589 Q13590 Q13591 Q13592 Q13593 Q13594 Q13595 Q13596 Q13597 Q13598 Q13599 Q13600 Q13601 Q13602 Q13603 Q13604 Q13605 Q13606 Q13607 Q13608 Q13609 Q13610 Q13611 Q13612 Q13613 Q13614 Q13615 Q13616 Q13617 Q13618 Q13619 Q13620 Q13621 Q13622 Q13623 Q13624 Q13625 Q13626 Q13627 Q13628 Q13629 Q13630 Q13631 Q13632 Q13633 Q13634 Q13635 Q13636 Q13637 Q13638 Q13639 Q13640 Q13641 Q13642 Q13643 Q13644 Q13645 Q13646 Q13647 Q13648 Q13649 Q13650 Q13651 Q13652 Q13653 Q13654 Q13655 Q13656 Q13657 Q13658 Q13659 Q13660 Q13661 Q13662 Q13663 Q13664 Q13665 Q13666 Q13667 Q13668 Q13669 Q13670 Q13671 Q13672 Q13673 Q13674 Q13675 Q13676 Q13677 Q13678 Q13679 Q13680 Q13681 Q13682 Q13683 Q13684 Q13685 Q13686 Q13687 Q13688 Q13689 Q13690 Q13691 Q13692 Q13693 Q13694 Q13695 Q13696 Q13697 Q13698 Q13699 Q13700 Q13701 Q13702 Q13703 Q13704 Q13705 Q13706 Q13707 Q13708 Q13709 Q13710 Q13711 Q13712 Q13713 Q13714 Q13715 Q13716 Q13717 Q13718 Q13719 Q13720 Q13721 Q13722 Q13723 Q13724 Q13725 Q13726 Q13727 Q13728 Q13729 Q13730 Q13731 Q13732 Q13733 Q13734 Q13735 Q13736 Q13737 Q13738 Q13739 Q13740 Q13741 Q13742 Q13743 Q13744 Q13745 Q13746 Q13747 Q13748 Q13749 Q13750 Q13751 Q13752 Q13753 Q13754 Q13755 Q13756 Q13757 Q13758 Q13759 Q13760 Q13761 Q13762 Q13763 Q13764 Q13765 Q13766 Q13767 Q13768 Q13769 Q13770 Q13771 Q13772 Q13773 Q13774 Q13775 Q13776 Q13777 Q13778 Q13779 Q13780 Q13781 Q13782 Q13783 Q13784 Q13785 Q13786 Q13787 Q13788 Q13789 Q13790 Q13791 Q13792 Q13793 Q13794 Q13795 Q13796 Q13797 Q13798 Q13799 Q13800 Q13801 Q13802 Q13803 Q13804 Q13805 Q13806 Q13807 Q13808 Q13809 Q13810 Q13811 Q13812 Q13813 Q13814 Q13815 Q13816 Q13817 Q13818 Q13819 Q13820 Q13821 Q13822 Q13823 Q13824 Q13825 Q13826 Q13827 Q13828 Q13829 Q13830 Q13831 Q13832 Q13833 Q13834 Q13835 Q13836 Q13837 Q13838 Q13839 Q13840 Q13841 Q13842 Q13843 Q13844 Q13845 Q13846 Q13847 Q13848 Q13849 Q13850 Q13851 Q13852 Q13853 Q13854 Q13855 Q13856 Q13857 Q13858 Q13859 Q13860 Q13861 Q13862 Q13863 Q13864 Q13865 Q13866 Q13867 Q13868 Q13869 Q13870 Q13871 Q13872 Q13873 Q13874 Q13875 Q13876 Q13877 Q13878 Q13879 Q13880 Q13881 Q13882 Q13883 Q13884 Q13885 Q13886 Q13887 Q13888 Q13889 Q13890 Q13891 Q13892 Q13893 Q13894 Q13895 Q13896 Q13897 Q13898 Q13899 Q13900 Q13901 Q13902 Q13903 Q13904 Q13905 Q13906 Q13907 Q13908 Q13909 Q13910 Q13911 Q13912 Q13913 Q13914 Q13915 Q13916 Q13917 Q13918 Q13919 Q13920 Q13921 Q13922 Q13923 Q13924 Q13925 Q13926 Q13927 Q13928 Q13929 Q13930 Q13931 Q13932 Q13933 Q13934 Q13935 Q13936 Q13937 Q13938 Q13939 Q13940 Q13941 Q13942 Q13943 Q13944 Q13945 Q13946 Q13947 Q13948 Q13949 Q13950 Q13951 Q13952 Q13953 Q13954 Q13955 Q13956 Q13957 Q13958 Q13959 Q13960 Q13961 Q13962 Q13963 Q13964 Q13965 Q13966 Q13967 Q13968 Q13969 Q13970 Q13971 Q13972 Q13973 Q13974 Q13975 Q13976 Q13977 Q13978 Q13979 Q13980 Q13981 Q13982 Q13983 Q13984 Q13985 Q13986 Q13987 Q13988 Q13989 Q13990 Q13991 Q13992 Q13993 Q13994 Q13995 Q13996 Q13997 Q13998 Q13999 Q14000	1	1504	100.0	+	ABL1 A0A024R8 A0A0U4E0 A1Z199 A3RL30 A6MF66 A6MF67 A6MF68 A6MF69 A6MF70 A6MF71 A6MF72 A6MF73 A6MF74 A6MF75 A6MF76 A6MF77 A6MF78 A6MF79 A6MF80 A6MF81 A6MF82 A6MF83 A6MF84 A6MF85 A6MF86 A6MF87 A6MF88 A6MF89 A6MF90 A6MF91 A6MF92 A6MF93 A6MF94 A6MF95 A6MF96 A6MF97 A6MF98 A6MF99 A6MF100 A6MF101 A6MF102 A6MF103 A6MF104 A6MF105 A6MF106 A6MF107 A6MF108 A6MF109 A6MF110 A6MF111 A6MF112 A6MF113 A6MF114 A6MF115 A6MF116 A6MF117 A6MF118 A6MF119 A6MF120 A6MF121 A6MF122 A6MF123 A6MF124 A6MF125 A6MF126 A6MF127 A6MF128 A6MF129 A6MF130 A6MF131 A6MF132 A6MF133 A6MF134 A6MF135 A6MF136 A6MF137 A6MF138 A6MF139 A6MF140 A6MF141 A6MF142 A6MF143 A6MF144 A6MF145 A6MF146 A6MF147 A6MF148 A6MF149 A6MF150 A6MF151 A6MF152 A6MF153 A6MF154 A6MF155 A6MF156 A6MF157 A6MF158 A6MF159 A6MF160 A6MF161 A6MF162 A6MF163 A6MF164 A6MF165 A6MF166 A6MF167 A6MF168 A6MF169 A6MF170 A6MF171 A6MF172 A6MF173 A6MF174 A6MF175 A6MF176 A6MF177 A6MF178 A6MF179 A6MF180 A6MF181 A6MF182 A6MF183 A6MF184 A6MF185 A6MF186 A6MF187 A6MF188 A6MF189 A6MF190 A6MF191 A6MF192 A6MF193 A6MF194 A6MF195 A6MF196 A6MF197 A6MF198 A6MF199 A6MF200 A6MF201 A6MF202 A6MF203 A6MF204 A6MF205 A6MF206 A6MF207 A6MF208 A6MF209 A6MF210 A6MF211 A6MF212 A6MF213 A6MF214 A6MF215 A6MF216 A6MF217 A6MF218 A6MF219 A6MF220 A6MF221 A6MF222 A6MF223 A6MF224 A6MF225 A6MF226 A6MF227 A6MF228 A6MF229 A6MF230 A6MF231 A6MF232 A6MF233 A6MF234 A6MF235 A6MF236 A6MF237 A6MF238 A6MF239 A6MF240 A6MF241 A6MF242 A6MF243 A6MF244 A6MF245 A6MF246 A6MF247 A6MF248 A6MF249 A6MF250 A6MF251 A6MF252 A6MF253 A6MF254 A6MF255 A6MF256 A6MF257 A6MF258 A6MF259 A6MF260 A6MF261 A6MF262 A6MF263 A6MF264 A6MF265 A6MF266 A6MF267 A6MF268 A6MF269 A6MF270 A6MF271 A6MF272 A6MF273 A6MF274 A6MF275 A6MF276 A6MF277 A6MF278 A6MF279 A6MF280 A6MF281 A6MF282 A6MF283 A6MF284 A6MF285 A6MF286 A6MF287 A6MF288 A6MF289 A6MF290 A6MF291 A6MF292 A6MF293 A6MF294 A6MF295 A6MF296 A6MF297 A6MF298 A6MF299 A6MF300 A6MF301 A6MF302 A6MF303 A6MF304 A6MF305 A6MF306 A6MF307 A6MF308 A6MF309 A6MF310 A6MF311 A6MF312 A6MF313 A6MF314 A6MF315									

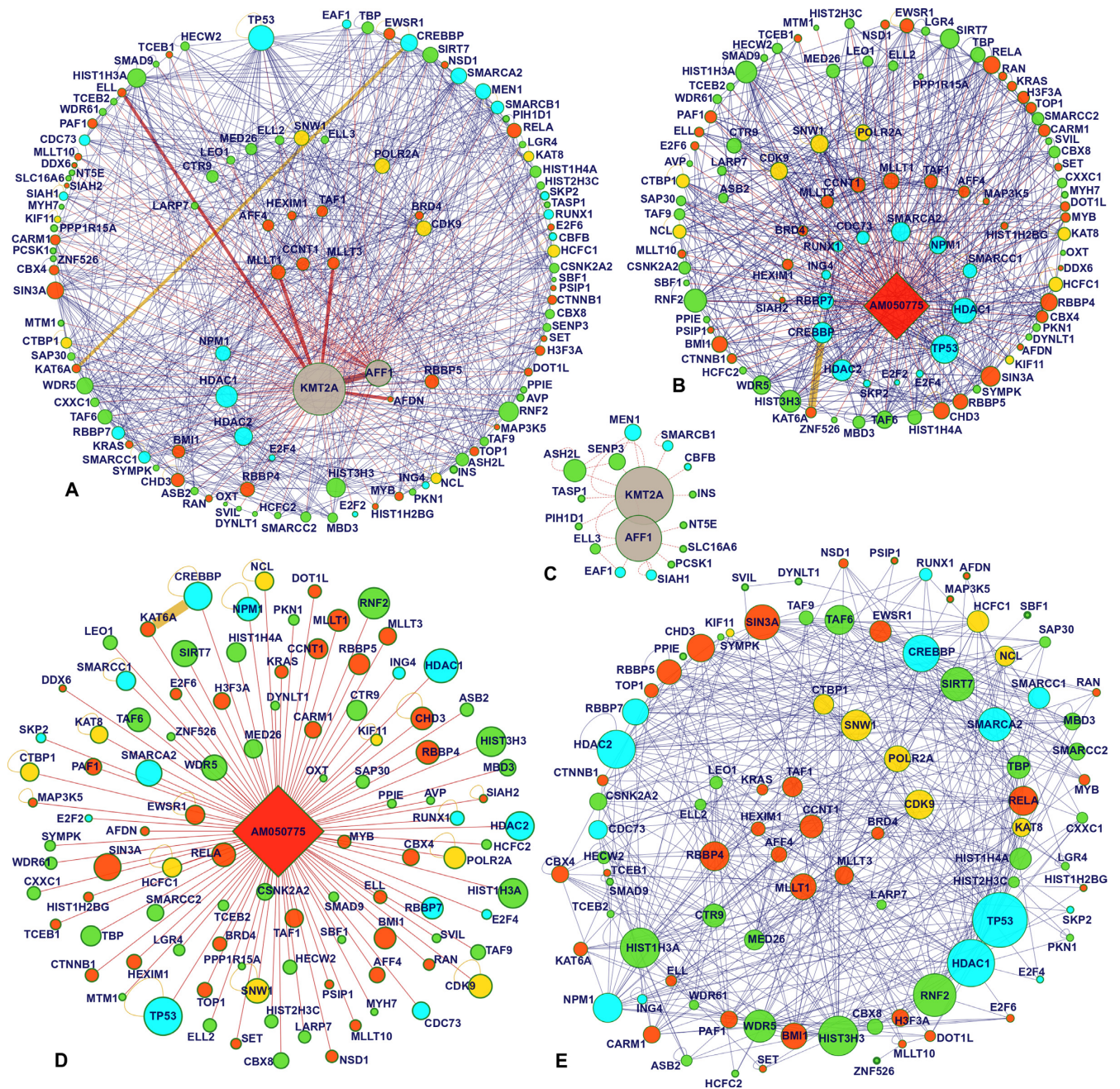


Figure 2. The ChiPPI protein–protein interaction network for the AFF1/KMT2A fusion protein. (A) The initial PPI networks of parental proteins (AFF1 and KMT2A). (B) The ChiPPI network for the AFF1/KMT2A fusion (chimericID: AM050775, ‘Full Collection’ and ‘Breakpoints’). The onco-proteins, parental proteins, potential onco-proteins, tumor suppressors and normal proteins are shown in dark-orange, light-orange, yellow, blue and green colors correspondingly. (C) All the missing interactors are shown on the network. The interactions of KMT2A with the tumor suppressors MEN1, SMARCB1 and CFBF and also the interactions of AFF1 with the tumor suppressors EAF1 and SIAH1 are lost upon the fusion. (D) The network of the ‘affected’ interactions, e.g. those changed upon the fusion event. (E) All the interactions that stay unchanged upon the fusion event.

Chimeric protein–protein interaction (ChiPPI) networks

Next, we asked how fusion protein function can be best assessed using computational methods. A fusion protein typically contains discreet domains from both parental proteins. It has been shown that a fusion event can dramatically alter the protein–protein interaction (PPI) network of the parental proteins. We, therefore, designed a visual means of assessing PPI network perturbations induced by protein

fusion events. To this end, we have assembled a tool for analyzing PPI networks that focuses on individual protein domains as the mediators of PPIs (<http://chippi.md.biu.ac.il/>). Using this tool we have fully pre-computed chimeric protein–protein interactions (ChiPPI) networks for 2081 fusion proteins, and predicted ChiPPI networks for 22 527 human chimeric transcripts. The predicted ChiPPI networks have been produced by the unification of the PPI networks

of two parental proteins of a chimera (see 'Full Collection'). This new feature provides users with the ability to study the protein interaction networks of chimeric proteins for all cancer fusions.

ChiPPI displays PPI networks in a map that gives a broad overview of the consequences of a fusion event from a proteomic perspective. ChiPPI predicts where fusion proteins are likely to lose binding to interactors of the parental proteins. Figure 2 shows the predicted PPI network of AFF1-KMT2A fusion protein. The interactions of KMT2A with the tumor suppressors: MEN1, SMARCB1 and CBFB, as well as the interactions of AFF1 with the tumor suppressors: EAF1 and SIAH1 are lost upon the fusion. Using ChiPPI, we mapped the influence of specific fusion proteins on cellular PPI networks and on essential pathways in cancer development and progression. For example, Supplementary Table S3 shows the mapping of ChiPPI networks for the different ABL1 fusions and their alterations in the 'betweenness centrality', 'clustering coefficient', and scoring for the addition of onco-proteins to, or removal of tumor suppressors from, the PPI network. In general, we find that the PPI networks of fusion proteins often lose tumor suppressor proteins, as well as being enriched in onco-proteins. Thus, ChiPPI is highly suitable for displaying how fusion proteins contribute to the skewing of protein interaction networks as well as of signaling pathways. Particularly, this new feature provides users with the ability to study the protein interaction networks of different cancer fusions (<http://chippi.md.biu.ac.il/>).

CONCLUSIONS AND PERSPECTIVES

The enhanced ChiTaRS-3.1 database is a comprehensive resource dedicated to the study of chimeric alterations at the proteomic, transcriptomic, genomic level in eukaryotes. ChiTaRS and ChiPPI are recently being referenced as sources for publishable data on cancer fusions (e.g. (50)). The updated version 3.1 of the ChiTaRS database provides a vast increase in annotated and verified chimeric transcripts as compared to the previous ChiTaRS releases, and includes a significant extension of specific research-oriented features. ChiTaRS-3.1 provides extensive experimental evidence for chimeras and cancer fusions, which can be effectively applied in the planning of new experiments or for the analysis of large scale RNA-sequencing experiments. International projects like ICGC and TCGA will benefit from this database and on all incremental additions to it, for improving the process of chimera identification and validation. To conclude, the ChiTaRS-3.1 database is designed to advance the field of Cancer Research as well as our understanding of the phenomenon of chimeric transcripts and its evolution in eukaryotes.

AVAILABILITY

The ChiTaRS-3.1 content will be continuously maintained and updated every six months. The database is now publicly accessible at <http://chitars.md.biu.ac.il> and its old version 2.1 is accessible at <http://chitars.bioinfo.cnio.es/>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank MPLabs LTD for the website design and support, and whose efforts make ChiTaRS-3.1 a reality. We thank our users for their consistent support and valuable feedback and our outstanding group for their priceless discussions and suggestions. We would like to thank Prof. Alfonso Valencia and his group at Centro Nacional de Investigaciones Oncológicas (CNIO), Spain, for their valuable suggestions in the overall process of designing the database and writing the manuscript.

FUNDING

PBC (VATAT) (22351, 20027) [to S.T. for Post-Doctoral Fellowship]. Funding for open access charge: VATAT [VaTaT #22351 to M.F.-M. & S.T.].

Conflict of interest statement. None declared.

REFERENCES

- Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigó,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
- Guigó,R., Flicek,P., Abril,J.F., Reymond,A., Lagarde,J., Denoeud,F., Antonarakis,S., Ashburner,M., Bajic,V.B., Birney,E. *et al.* (2006) EGASP: the human ENCODE genome annotation assessment project. *Genome Biol.*, **7**(Suppl 1), 1–31.
- Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
- Griffin,T.J., Gygi,S.P., Ideker,T., Rist,B., Eng,J., Hood,L. and Aebersold,R. (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics*, **1**, 323–333.
- Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E., Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) Characterization of the yeast transcriptome. *Cell*, **88**, 243–251.
- Cirulli,E.T., Singh,A., Shianna,K.V., Ge,D., Smith,J.P., Maia,J.M., Heinzen,E.L., Goedert,J.J., Goldstein,D.B. and (CHAVI), C.f.H.A.VI. (2010) Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol.*, **11**, R57.
- Finta,C. and Zaphiropoulos,P.G. (2002) Intergenic mRNA molecules resulting from trans-splicing. *J. Biol. Chem.*, **277**, 5882–5890.
- Kapranov,P., Drenkow,J., Cheng,J., Long,J., Helt,G., Dike,S. and Gingeras,T.R. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.*, **15**, 987–997.
- Di Segni,G., Gastaldi,S. and Tocchini-Valentini,G.P. (2008) Cis- and trans-splicing of mRNAs mediated by tRNA sequences in eukaryotic cells. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 6864–6869.
- Akiva,P., Toporik,A., Edelman,S., Peretz,Y., Diber,A., Shemesh,R., Novik,A. and Sorek,R. (2006) Transcription-mediated gene fusion in the human genome. *Genome Res.*, **16**, 30–36.
- Parra,G., Reymond,A., Dabbouseh,N., Dermitzakis,E.T., Castelo,R., Thomson,T.M., Antonarakis,S.E. and Guigó,R. (2006) Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.*, **16**, 37–44.
- Romani,A., Guerra,E., Trerotola,M. and Alberti,S. (2003) Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res.*, **31**, e17.

13. Campbell,P.J., Stephens,P.J., Pleasance,E.D., O'Meara,S., Li,H., Santarius,T., Stebbings,L.A., Leroy,C., Edkins,S., Hardy,C. *et al.* (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.*, **40**, 722–729.
14. Ortiz de Mendibil,I, Vizmanos,J.L. and Novo,F.J. (2009) Signatures of selection in fusion transcripts resulting from chromosomal translocations in human cancer. *PLoS One*, **4**, e4805.
15. Li,H., Wang,J., Mor,G. and Sklar,J. (2008) A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science*, **321**, 1357–1361.
16. Li,H., Wang,J., Ma,X. and Sklar,J. (2009) Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell Cycle*, **8**, 218–222.
17. Edgren,H., Murumagi,A., Kangaspeska,S., Nicoric,D., Hongisto,V., Kleivi,K., Rye,I.H., Nyberg,S., Wolf,M., Borresen-Dale,A.L. *et al.* (2011) Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.*, **12**, R6.
18. Frenkel-Morgenstern,M., Lacroix,V., Ezkurdia,I., Levin,Y., Gabashvili,A., Prilusky,J., Del Pozo,A., Tress,M., Johnson,R., Guigo,R. *et al.* (2012) Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.*, **22**, 1231–1242.
19. Frenkel-Morgenstern,M. and Valencia,A. (2012) Novel domain combinations in proteins encoded by chimeric transcripts. *Bioinformatics*, **28**, i67–i74.
20. Asmann,Y.W., Necela,B.M., Kalari,K.R., Hossain,A., Baker,T.R., Carr,J.M., Davis,C., Getz,J.E., Hostetter,G., Li,X. *et al.* (2012) Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res.*, **72**, 1921–1928.
21. Gingeras,T.R. (2009) Implications of chimaeric non-co-linear transcripts. *Nature*, **461**, 206–211.
22. Maher,C.A., Palanisamy,N., Brenner,J.C., Cao,X., Kalyana-Sundaram,S., Luo,S., Khrebtukova,I., Barrette,T.R., Grasso,C., Yu,J. *et al.* (2009) Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 12353–12358.
23. Maher,C.A., Kumar-Sinha,C., Cao,X., Kalyana-Sundaram,S., Han,B., Jing,X., Sam,L., Barrette,T., Palanisamy,N. and Chinnaiyan,A.M. (2009) Transcriptome sequencing to detect gene fusions in cancer. *Nature*, **458**, 97–101.
24. Djebali,S., Lagarde,J., Kapranov,P., Lacroix,V., Borel,C., Mudge,J.M., Howald,C., Foissac,S., Ucla,C., Chrast,J. *et al.* (2012) Evidence for transcript networks composed of chimeric RNAs in human cells. *PLoS One*, **7**, e28213.
25. Prakash,A., Tomazela,D.M., Frewen,B., Maclean,B., Merrihew,G., Peterman,S. and Maccoss,M.J. (2009) Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J. Proteome Res.*, **8**, 2733–2739.
26. Kim,P., Yoon,S., Kim,N., Lee,S., Ko,M., Lee,H., Kang,H. and Kim,J. (2010) ChimerDB 2.0—a knowledgebase for fusion genes updated. *Nucleic Acids Res.*, **38**, D81–D85.
27. Denoed,F., Kapranov,P., Ucla,C., Frankish,A., Castelo,R., Drenkow,J., Lagarde,J., Alioto,T., Manzano,C., Chrast,J. *et al.* (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.*, **17**, 746–759.
28. Houseley,J. and Tollervey,D. (2010) Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS One*, **5**, e12271.
29. McManus,C.J., Duff,M.O., Eipper-Mains,J. and Graveley,B.R. (2010) Global analysis of trans-splicing in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12975–12979.
30. McManus,C.J., Coolon,J.D., Duff,M.O., Eipper-Mains,J., Graveley,B.R. and Wittkopp,P.J. (2010) Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.*, **20**, 816–825.
31. Wu,C.S., Yu,C.Y., Chuang,C.Y., Hsiao,M., Kao,C.F., Kuo,H.C. and Chuang,T.J. (2014) Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency. *Genome Res.*, **24**, 25–36.
32. Yu,C.Y., Liu,H.J., Hung,L.Y., Kuo,H.C. and Chuang,T.J. (2014) Is an observed non-co-linear RNA product spliced in trans, in cis or just *in vitro*? *Nucleic Acids Res.*, **42**, 9410–9423.
33. Frenkel-Morgenstern,M., Gorohovski,A., Lacroix,V., Rogers,M., Ibanez,K., Boulosa,C., Andres Leon,E., Ben-Hur,A. and Valencia,A. (2013) ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. *Nucleic Acids Res.*, **41**, D142–D151.
34. Frenkel-Morgenstern,M., Gorohovski,A., Vucenovic,D., Maestre,L. and Valencia,A. (2015) ChiTaRS 2.1—an improved database of the chimeric transcripts and RNA-seq data with novel sense-antisense chimeric RNA transcripts. *Nucleic Acids Res.*, **43**, D68–D75.
35. Benson,D.A., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2014) GenBank. *Nucleic Acids Res.*, **42**, D32–D37.
36. Kim,N., Kim,P., Nam,S., Shin,S. and Lee,S. (2006) ChimerDB—a knowledgebase for fusion sequences. *Nucleic Acids Res.*, **34**, D21–D24.
37. Kong,F., Zhu,J., Wu,J., Peng,J., Wang,Y., Wang,Q., Fu,S., Yuan,L.L. and Li,T. (2011) dbCRID: a database of chromosomal rearrangements in human diseases. *Nucleic Acids Res.*, **39**, D895–D900.
38. 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.
39. Mertens,F., Johansson,B., Fioretos,T. and Mitelman,F. (2015) The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer*, **15**, 371–381.
40. Mertens,F., Antonescu,C.R. and Mitelman,F. (2016) Gene fusions in soft tissue tumors: recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer*, **55**, 291–310.
41. Mitelman,F., Mertens,F. and Johansson,B. (2005) Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders. *Genes Chromosomes Cancer*, **43**, 350–366.
42. Mitelman,F., Johansson,B. and Mertens,F. (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*, **7**, 233–245.
43. Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, **2011**, bar009.
44. Stelzer,G., Rosen,N., Plaschkes,I., Zimmerman,S., Twik,M., Fishilevich,S., Stein,T.I., Nudel,R., Lieder,I., Mazor,Y. *et al.* (2016) The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics*, **54**, 31–33.
45. Fernández,J.M., Hoffmann,R. and Valencia,A. (2007) iHOP web services. *Nucleic Acids Res.*, **35**, W21–W26.
46. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2011) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.
47. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM.org: online mendelian inheritance in man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
48. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciuffo,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
49. Benson,D.A., Karsch-Mizrachi,I., Clark,K., Lipman,D.J., Ostell,J. and Sayers,E.W. (2012) GenBank. *Nucleic Acids Res.*, **40**, D48–D53.
50. Latysheva,N.S., Oates,M.E., Maddox,L., Flock,T., Gough,J., Buljan,M., Weatheritt,R.J. and Babu,M.M. (2016) Molecular principles of gene fusion mediated rewiring of protein interaction networks in cancer. *Mol. Cell*, **63**, 579–592.