A Novel Pipeline for Identification and Prioritization of Gene Fusions in Patient-derived Xenografts of Metastatic Colorectal Cancer

Paciello Giulia[1], Andrea Acquaviva[1], Consalvo Petti[2], Claudio Isella[2], Enzo Medico[2,3] and Elisa Ficarra[1]

[1]*Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, Turin, Italy*

[2]*Laboratory of Oncogenomics, Institute for Cancer Research at Candiolo (IRCC),
Strada Provinciale 142, Candiolo, Turin, Italy*

[3]*Department of Oncology, Universita' di Torino, Candiolo, Turin, Italy*

Keywords: RNA-sequencing, Metastatic Colorectal Cancer, Patient-derived Mouse Xenografts, Artifacts.

Abstract: Metastatic spread to the liver is a frequent complication of colorectal cancer (CRC), occurring in almost half of the cases, for which personalized treatment strategies are highly desirable. To this aim, it has been proven that patient-derived mouse xenografts (PDX) of liver-metastatic CRC can be used to discover new therapeutic targets and determinants of drug resistance. To identify gene fusions in RNA-Seq data obtained from such PDX samples, we propose a novel pipeline that tackles the following issues: (i) discriminating human from murine RNA, to filter out transcripts contributed by the mouse stroma that supports the PDX; (ii) increasing sensitivity in case of suboptimal RNA-Seq coverage; (iii) prioritizing the detected chimeric transcripts by molecular features of the fusion and by functional relevance of the involved genes; (iv) providing appropriate sequence information for subsequent validation of the identified fusions. The pipeline, built on top of Chimerascan(R.Iyer, 2011) and deFuse(McPherson, 2011) aligner tools, was successfully applied to RNA-Seq data from 11 PDX samples. Among the 299 fusion genes identified by the aforementioned softwares, five were selected since passed all the filtering stages implemented into the proposed pipeline resulting as biologically relevant fusions. Three of them were experimentally confirmed.

1 INTRODUCTION

It is currently known that cancer derives from permanent alterations of the cellular DNA, leading to aberrant growth, invasion of adjacent tissues and metastatic diffusion at distant sites(Hanahan, 2000). Among the various DNA alterations, chromosomal rearrangements leading to gene fusions play a central role in the initial steps of many pathologies such as leukaemias, sarcomas and common epithelial neoplasms, like breast, colorectal and prostate cancer(Aman, 1999). The impact of gene fusions on cellular behavior is due to functional alteration of one or both the genes involved in the chromosomal rearrangements that give rise to chimeric transcripts. Typical consequences of gene fusions, at the RNA and protein level, are strong variation of expression, removal of regulatory domains, forced oligomerization, change of the subcellular location or acquisition of novel binding domains(Edwards, 2010). Gene fusions can therefore have important prognostic and therapeutic implications in the management of malignancies, as shown in recent studies(Mitelman, 2007).

Data produced by Next Generation Sequencing (NGS) technologies are nowadays considered very useful in order to detect genetic abnormalities(Ansorge, 2009)(Mardis, 2008)(Metzker, 2010). In particular, concerning chimeric transcripts identification, the analysis of the so called paired-end RNA-Seq reads can be considered a powerful strategy as shown by Maher and colleagues(Maher, 2009) and confirmed by the subsequent development of numerous tools to perform such activity(McPherson, 2011)(R.Iyer, 2011)(Abate, 2012). Paired-end reads, differently from single-end reads, are obtained by sequencing nucleic acid fragments at both the 5' and 3' ends. When the two sequenced portions of the fragment, called mates, align on different genes, it is likely that the fragment is originated by a chimeric transcript. We applied this strategy to identify relevant gene fusions in colorectal cancer (CRC), one of the most frequent cancers worldwide(Walther, 2000).

Giulia P., Acquaviva A., Petti C., Isella C., Medico E. and Ficarra E..
A Novel Pipeline for Identification and Prioritization of Gene Fusions in Patient-derived Xenografts of Metastatic Colorectal Cancer.
DOI: 10.5220/0004799401420148
In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (BIOINFORMATICS-2014), pages 142-148
ISBN: 978-989-758-012-3
Copyright © 2014 SCITEPRESS (Science and Technology Publications, Lda.)

CRC is frequently complicated by liver metastasis, and features a remarkable heterogeneity in terms of molecular pathogenesis, natural history and response to treatment(Siena, 2009)(Cunningham, 2010). A recent advance in the characterization of such heterogeneity has been brought forward by the propagation of human neoplastic tissue in immunodeficient mice, the so-called patient-derived xenograft (PDX) approach. As proven by Bertotti and colleagues(Bertotti, 2011), PDXs of human metastatic colorectal cancer can be reliably exploited to discover novel determinants of therapeutic response and new oncoprotein targets. PDXs are indeed able to conserve the inter-individual diversity and the genetic heterogeneity typical of the tumors of origin and at the same time to reproduce the disease responses in humans.

In the present work we searched for chimeric transcripts in eleven PDXs of metastatic CRC, by analyzing Illumina RNA-Seq data consisting of 100-base pair (bp) long paired-end reads. RNA was extracted from PDXs at the second passage of propagation in mice. By this stage human stromal cells, not capable of growing in the murine context, are replaced by mouse stromal cells. As a consequence, the extracted RNA is a mixture of human RNA, originated from neoplastic cells, and murine RNA from stromal cells. This required a dedicated step in the RNA-Seq analysis pipeline to distinguish reads originated from RNAs of the two species. The experimental setup is reported in Figure 1.

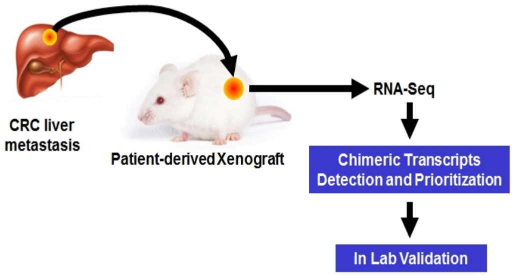
CRC liver metastasis
Patient-derived Xenograft
RNA-Seq
Chimeric Transcripts Detection and Prioritization
In Lab Validation

Figure 1: Experimental Setup. The activities highlighted in blue boxes are respectively that implemented in the pipeline objective of this work and the validation of the obtained results.

In particular rhe proposed research is inspired by the observation that fusion detection tools usually produce in the preliminary steps of their workflow a conspicuous number of putative gene fusions not feasible for in lab validation. To shrink down these candidates lists, gene fusions detection tools implement inside their algorithms different kind of filters generally efficient in discarding false positive fusion transcrpits but at the same time determining poor sensitivity level of the detection. Therefore, as reported By Abate and colleagues (Abate, 2012) and Carrara and colleagues(Carrara, 2013) because the heterogeneity of algorithms and filters applied, the identified fusions usually poorly overlap with a high rate of false positives, but also of false negatives, leading to the need of considering the union set of fusion genes detection tool outputs for in lab validation. The number of fusions to be tested becomes in this way not feasible even considered that the biologically relevant fusions in a sample are usually very few if any(Ozsolak, 2011).

With the aim of reducing the union set of candidates detected by two fusion genes discovery tools (i.e Chimerascan(R.Iyer, 2011) and defuse(McPherson, 2011)) in the aforementioned PDXs samples, we proposed a novel pipeline characterized by the reimplementation of some modules proper of fusion genes detection tools and ad hoc scripts developed to perform different filtering stages.

At the days no tools or algorithms implementing the analysis performed by our pipeline can be identified since all sequencing studies in PDXs such as those of Rossello and colleagues, Conway and colleagues, Valder and colleagues(Rossello, 2013)(Conway, 2012)(Valdes, 2013) circumscribed at the discrimination between murine and human reads don't relying on fusion genes detection beyond their prioritizazion.

The whole workflow is implemented in order to: (i) take account for the murine stroma; (ii) consider the contingent PCR artifacts; (iii) evaluate the role of different kind of reads in the chimeric transcript reliability to maximize sensitivity also at low sequencing coverage; (iv) integrate biological and functional information about the gene fusions. Among the prioritized fusions, three have been at the moment experimentally confirmed in lab with PCR.

2 METHODS

The proposed pipeline is characterized by different activities and filtering stages as shown in Figure 2. In the following all the steps will be detailed.

First Filtering Stage: Gene Fusions Annotation and Selection.

The list of gene fusions detected using Chimerascan(R.Iyer, 2011) and defuse(McPherson, 2011) tools with default run parameters constitute the input of the first step of the proposed workflow on the eleven RNA-Seq samples under examination. The choice of Chimerascan(R.Iyer, 2011) and deFuse(McPherson,

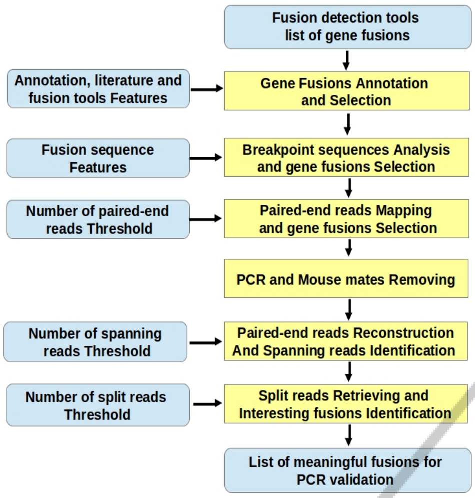
Fusion detection tools list of gene fusions
Annotation, literature and fusion tools Features
Gene Fusions Annotation and Selection
Fusion sequence Features
Breakpoint sequences Analysis and gene fusions Selection
Number of paired-end reads Threshold
Paired-end reads Mapping and gene fusions Selection
PCR and Mouse mates Removing
Number of spanning reads Threshold
Paired-end reads Reconstruction And Spanning reads Identification
Number of split reads Threshold
Split reads Retrieving and Interesting fusions Identification
List of meaningful fusions for PCR validation

Figure 2: Pipeline workflow. Activities are represented in yellow boxes whereas input and output in blue ones.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

2011) tools is dictated by a recent research performed by Carrara and colleagues(Carrara, 2013). Chimerascan(R.Iyer, 2011) and deFuse(McPherson, 2011) tools were indeed proven to achieve good sensitivity levels on real dataset even if generally provide a remarkable number of false positive chimeras. This negative feature has been however properly managed thanks to the following different filtering stages. Furthermore for what is concerning Chimerascan(R.Iyer, 2011) and defuse(McPherson, 2011) run parameters they can be conveniently triggered according to the specific requirements even if we evaluated that from a computational point of view it is more convenient to impose restrictive mapping policies in the following phases of the proposed pipeline. Chimeric transcripts output files, containing all the fusions detected in the samples, are automatically elaborated taking advantage of an annotation tool (http://sourceforge.net/projects/pegasus-fus/) able to perform on the detected fusions, protein domain and functional analysis. The achieved information provide a more detailed overview on each chimeric transcript supplying knowledge about the presence of kinases, the reading frame, the domain conserved or loss, and the breakpoint regions. Of essential importance in this phase are also the data provided by the chimeric transcripts detection tools for what is concerning the reads used to identify a gene fusion. Chimerascan(R.Iyer, 2011) and deFuse(McPherson, 2011) indeed, in order to score the fusions detected in the samples, provide the number of reads used to define the fusion: They distinguish between encompassing reads, in other word paired-end reads with the two mates mapped respectively on the two different partner genes of the chimeric transcript and split mates, representing those mates that harbor the gene fusion breakpoint. Starting from the information collected using the aforementioned annotation tool, the gene fusions detection tools outputs and thanks to biological considerations about the function of the genes involved in the fusion a first step of filtering is performed. Only those fusions characterized by a certain threshold number of split reads that will be discussed in *Results* Section and satisfying also at least another one criteria of those previously listed will be considered for further evaluations.

Second Filtering Stage: Breakpoint Sequences Analysis and Gene Fusions Selection.
Starting from the prioritized list of fusion gene candidates a new stage of filtering is applied. The fusion sequence of each chimeric transcript, provided by defuse(McPherson, 2011) and deduced from the split reads for what is concerning Chimerascan(R.Iyer, 2011) is here analysed. Main objective of this filtering stage consists essentially in the retrieving of those gene fusions sequences that could account for the translation of the chimeric transcript into a functional protein. The presence of a Kozac sequence in the 5' partner gene generally account for an ATG triplet (starting site of the translation process) downstream that allows the beginning of the translation process. In this phase also chimeric transcript sequences characterized by an in frame configuration have been however selected for further analyses since we considered the case in which the ATG triplet is located upstream with respect to the starting point of the sequencing.

Third Filtering Stage: Paired-end Reads Mapping and Gene Fusions Selection.
For the gene fusions selected in the previous step the mapping of the paired-end reads on the fusion sequence is performed taking advantage of Bowtie(Langmead, 2009) tool which parameters were set in order to report for a data read only the best alignment identified. Outputs were later converted using the Samtools(Li, 2009) and the BedTools(Aaron, 2009) utilities to obtain suitable format files for the following phases of the pipeline. Gene fusions not supported by a threshold number of paired-end reads will be not considered in the next steps of the flow.

Fourth Filtering Stage: PCR and Mouse Mates Filtering.
All the mates mapped onto a specific fusion sequence have been analyzed in this phase with the main purposes of removing those deriving from PCR artifacts

and those mapping also on the murine DNA. For what is concerning PCR artifacts, as depicts in Figure 3, if more than a mate is mapped in the same location, only one mate of the group is considered for further analyses (green mates in Figure 3 a).

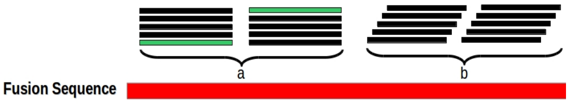

Figure 3: PCR artifacts removal filter. In Subfigure a is reported an example of mates deriving from PCR artifacts, characterized by the same start and end mapping positions on the fusion sequence. Only one read for group (the green ones) is maintained for subsequent analyses. In Subfigure b are reported instead some mates mapped in unambiguous ladder-like pattern on the fusion sequence. All the mates are retained in this case for further examination.

As shown in Figure 4 a only the mates survived to the PCR artifacts removal filter have been mapped using Bowtie(Langmead, 2009) in single-end mode on the mouse genome in order to identify those mates mapped also on this reference. These mates, that probably derives from murine stromal cells sequencing (Figure 4 b), are removed from the list of mates supporting the specific gene fusion (green mates in Figure 4).

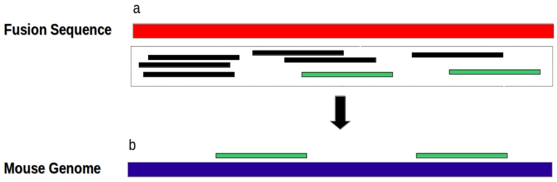

Figure 4: Mouse artifacts removal filter. In Subfigure 4 a are shown the input mates for the murine filter (all the mates not deriving from PCR artifacts) that have been mapped onto the fusion sequence. Mates belonging to such a group that have been mapped also on the murine genome, as shown in Figure 4 b (green mates), are not considered in the following phases of the pipeline.

Fifth Filtering Stage: Paired-end Reads Reconstruction and Spanning Reads Identification.
After the last described stage of filtering each gene fusion results supported by a list of single mates instead of the initial paired-end reads obtained from bowtie run on the fusion sequence. As proven by Maher et colleagues(Maher, 2009) paired-end reads are however more meaningful for gene fusions detection so, starting from the list of single mates supporting a putative chimeric transcript the paired-end read, if present, is reconstructed (green reads in Figure 5). Mates not belonging to a paired-end reads are discarded. Among these reads of great importance are the spanning reads, those reads characterized by the mapping of the mates on the two different partner genes of the chimeric transcript. Only those paired-end reads (spanning reads) having the two mates partially or totally mapped onto the two different partner genes were retrieved for further examination (yellow reads in Figure 5). Gene fusions not supported by a certain number of spanning reads that will be discussed in *Results* Section were discarded.

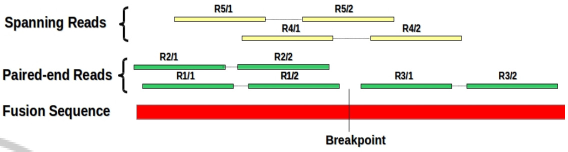

Figure 5: Paired-end and spanning reads representation. In Figure are reported respectively in green and yellow colours the paired-end and the spanning reads.

Sixth Filtering Stage: Split Reads Retrieving and Interesting Fusions Identification.
In this phase, in order to guarantee once again the reliability of both the partner genes involved in the fusion and the breakpoint coordinates on the same partner genes, the search for split reads is performed on the remaining putative chimeric transcripts: Split reads are indeed capable to account for a base-pair resolution of the gene fusion sequence in the breakpoint region. This is the reason for which only those spanning couples, represented with yellow bars in Figure 6, having one or both the mates mapped in the breakpoint region provided by Chimerascan(R.Iyer, 2011) and deFuse(McPherson, 2011) tools are considered. The mates belonging to such a couple are called split mates and are represented in a dashed box in Figure 6. PCR validation will be performed for those gene fusions characterized at least by a threshold number of split mates discussed once a time in *Results* Section.

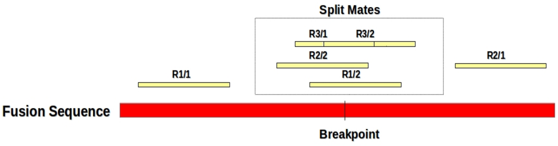

Figure 6: Split mates representation. In Figure are reported some spanning paired-end reads. The mates belonging to such couples that are mapped in the breakpoint region are called split mates and are represented in a dashed box.

3 RESULTS

The run of deFuse(McPherson, 2011) on the eleven metastatic CRC samples detected 132 gene fusions whereas that of Chimerascan(R.Iyer, 2011) 167. The two different outputs didn't show overlapping fusions as was initially supposed considering the re-

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

cent researches before mentioned(Abate, 2012)(Carrara, 2013).These results, not so conspicuous if considered the number of samples under examination and the mean amount of gene fusions usually identified in RNA-Seq samples, are essentially due to the poor coverage of the samples. It is worth noting however that generally higher the coverage of the samples higher will be the number of detected gene fusions with remarkably computational costs spent for the detection. It is at the same time worth noting that not all the 299 fusions can be tested in lab using PCR for known economic and temporal restraints. The proposed pipeline was therefore applied in order to prioritize the identified 299 chimeric transcripts. In the following we refer to the different chimeric transcripts using couples of capital letters (actual gene names cannot be disclosed as the biological results of this research are currently under review).

For what is concerning the first filtering stage a threshold of one split read was imposed: Thirteen gene fusions were here selected because supported by at least one split read and present at least one of the other features described in Subsection *First filtering stage: Gene fusions annotation and selection.* The two thresholds were selected with the intention to be as preservative as possible in evaluating the different candidates. These parameter values can be however tuned in order to satisfy specific needs. Furthemore even if among the initial 299 fusion genes none have been previously detected in cancer samples, all the genes involved in the thireteen fusions are characterized by mutational states related to cancer development and progression (information deriving from literature sources and COSMIC database(Simon, 2010)). Three out of the thirteen chimeras that passed the first filtering stage are moreover characterized by a partner gene found to be fused with other genes in cancer diseases.

For each of these chimeric transcripts, the fusion sequence has been then retrieved and analyzed in order to understand the biological mechanism at the basis of the recombination. The criteria of the second filtering stage reduced the previous list to only eight gene fusions characterized by the presence of a Kozac sequence (or an ATG triplet) at 5'-end or by an in frame configuration. Also frame shifted configurations could however be interesting in case of tumor suppressors 3' partner genes: In the proposed pipeline this scenario has not been considered because no oncogenic suppressor genes were detected among the identified chimeric transcript partner genes.

In Table 1 are reported the results relative to the third and fourth filtering stages. In particular, a threshold of at least one paired-end read was fixed in order to select a fusion for the next phases of the pipeline being as much preservative as possible: Only one gene fusion (i.e gene fusion G-H) was deleted because not supported by paired-end reads as shown in column 3 of Table 1. The differences among the breakpoint sequence lengths (Column 2 of Table 1) can be attributed to the fact that they depend on the number of reads used to define the fusion sequence. So higher the number of reads mapped by the chimeric transcript discovery tool on the supposed breakpoint sequence, higher will be the provided length of the same sequence and the probability of finding with the propose pipeline paired-end reads aligning on the same. The absence of mates removed by the mouse remove filter, shown in Column 5 of Table 1, confirm the fact that effectively the samples were composed exclusively of human tumor cells. On the other and, instead, the PCR in the most of cases caused a remarkably number of artifacts, as it is possible to note from Column 4 of Table 1.

After PCR and mouse mates removal, for each of the remaining seven gene fusions the supporting paired-end reads, if present, are reconstructed. This activity is followed by the identification of the so called paired-end spanning reads if existing. The results are shown in columns 2 and 3 of Table 2.

Two of the previous seven gene fusions have been removed because they are not supported by spanning reads (i.e. gene fusions I-L and M-N). A threshold of one spanning read was indeed imposed in order to consider a gene fusion. The value selected derives, as already largely discussed, from the desire to be very preservative since the previous filtering stages concerning functional and biological properties of fusion genes have been already capable as shown to remove a conspicuous number of not functional chimeric transcrips. It is worth noting however that it is possible to set this parameter according to the specific requirements.

The fourth column of Table 2 reports instead the number of split mates supporting the remaining five chimeric transcripts. In the last filtering stage a threshold of at least one split mate was imposed in order to consider a gene fusion for in lab validation. Even for the split reads the value parameter was selected, as already said in relation to spanning reads threshold, in order to be as conservative as possible. Of the initial 299 gene fusions at the end of the pipeline five were considered priority (i.e gene fusions A-B, C-D, E-F, O-P and Q-R). Three out of five have been actually validated in lab using PCR resulting as true gene fusions. In table 3 is reported a summary of the number of fusion genes obtained after the application of the different filtering stages.

Table 1: Third and Fourth Filtering Stages results.

Gene Fusion	Gene Fusion sequence length	Nr of mapped paired-end Reads	PCR Mates	Mouse Mates
A-B	3668	767	529	0
C-D	127	66	74	0
E-F	168	1	0	0
G-H	139	0	0	0
I-L	599	58	28	0
M-N	311	10	4	0
O-P	1886	32	2	0
Q-R	531	86	58	0

Table 2: Fifth and Sixth Filtering Stages results.

Gene Fusion	Nr of paired-end Reads	Nr of Spanning paired-end reads	Number of Split Mates
A-B	344	2	2
C-D	16	11	12
E-F	1	1	2
I-L	18	0	0
M-N	6	0	0
O-P	30	2	3
Q-R	43	5	8

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 3: Number of chimeric transcripts obtained as output of the different FSs (Filtering Stages).

#Initial fusions	#Fusions I FS	#Fusions II FS	#Fusions III FS	#Fusions IV FS	#Fusions V FS
299	13	8	7	5	5

4 CONCLUSIONS AND FUTURE WORK

Starting from the 299 gene fusions detected by Chimerascan(R.Iyer, 2011) and deFuse(McPherson, 2011) on the eleven metastatic CRC xenopatients RNA-Seq samples, the proposed pipeline was able to progressively reduce the list of the identified chimeric transcripts. Five gene fusions in particular passed all the filtering stages and were considered to be relevant. The in lab validation of three of these gene fusions confirmed the presence of a chimeric transcript product proving that the developed pipeline is able to identify true chimeric transcripts potentially associated to cancer onset and progression. Furthermore all the activities performed within the pipeline have been implemented considering the general features at the basis of a real, productive and biologically relevant gene fusion making this program capable to prioritize gene fusions from PDXs affected also by different diseases. Surely the proposed methodology can be transferred to other RNA-Seq dataset deriving from PDXs even with different pathologies. All the activities performed within the pipeline have been indeed implemented considering the general features at the basis of a real, productive and biologically relevant gene fusion making this program capable to prioritize gene fusions from PDXs affected as said by different diseases. We would like to underline that our aim is just to apply as soon as possible the proposed methodology o other dataset in order to improve the filtering stages with new features and at the same time to make very user friendly the entire tool.

Future works will aim at: (i) validate all the detected fusions using also Real-Time PCR (RT-PCR), (ii) integrate in the proposed pipeline results from other chimeric transcript detection tools, (iii) change the setting of the implemented filters in order to evaluate the filtering stages performances, (iv) investigate the functional role of the identified fusion transcripts, (v) evaluate the occurrence of the detected fusions in public RNA-seq dataset (TCGA).

REFERENCES

Aaron, R. (2009). Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics.

Abate, F. (2012). Bellerophontes: an rna-seq data analysis framework for chimeric transcripts discovery based on accurate fusion model. Bioinformatics.

Aman, P. (1999). Fusion genes in solid tumors. Semin Cancer Biology.

Ansorge, W. (2009). Next-generation dna sequencing techniques. New Biotechnology.

Bertotti, A. (2011). A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies her2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. Cancer Discovery.

Carrara, M. (2013). State-of-the-art fusion-finder algorithms sensitivity and specificity. BioMed Research International.

Conway, T. (2012). Xenome–a tool for classifying reads from xenograft samples. Bioinformatics.

Cunningham, D. (2010). Monoclonal antibodies in the treatment of metastatic colorectal cancer: a review. Colorectal Cancer.

Edwards, P. (2010). Fusion genes and chromosome translocations in the common epithelial cancers. The Journ. of Pathology.

Hanahan, D. (2000). The hallmarks of cancer. Cell.

Langmead, B. (2009). Ultrafast and memory-efficient alignment of short dna sequences to the human genome. Genome Biology.

Li, H. (2009). The sequence alignment/map (sam) format and samtools. Bioinformatics.

Maher, C. (2009). Chimeric transcript discovery by paired-end transcriptome sequencing. PNAS.

Mardis, E. (2008). The impact of next-generation sequencing technology on genetics. Trends in genetics.

McPherson, A. (2011). defuse: An algorithm for gene fusion discovery in tumor rna-seq data. PLOS Computational Biology.

Metzker, M. (2010). Sequencing technologies - the next generation. Nature Reviews Genetics.

Mitelman, F. (2007). The impact of translocations and gene fusions on cancer causation. Nature Reviews.Cancer.

Ozsolak, F. (2011). Rna sequencing: advances, challenges and opportunities. Nat Rev Genet.

R.Iyer (2011). Chimerascan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics.

Rossello, F. (2013). Next-generation sequence analysis of cancer xenograft models. PLoS ONE.

Siena, S. (2009). Biomarkers predicting clinical outcome of epidermal growth factor receptortargeted therapy in metastatic colorectal cancer. JNCI.

Simon, A. (2010). Cosmic: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Research.

Valdes, C. (2013). Characteristics of cross-hybridization and cross-alignment of expression in pseudo-xenograft samples by rna-seq and microarrays. Journal of Clinical Bioinformatics.

Walther, A. (2000). Genetic prognostic and predictive markers in colorectal cancer. Nature Reviews.Cancer.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS