Methods 59 (2013) 187–191

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Detecting and visualizing gene fusions

Jochen Supper*, Claudia Gugenmus, Johannes Wollnik, Tanja Drueke, Matthias Scherf, Alexander Hahn, Korbinian Grote, Nancy Bretschneider, Bernward Klocke, Christian Zinser, Kerstin Cartharius, Martin Seifert

Genomatix Software GmbH, Bayerstr. 85a, 80335 München, Germany

ARTICLE INFO

Article history:
Available online 2 October 2012

ABSTRACT

In recent years, gene fusions have gained significant recognition as biomarkers. They can assist treatment decisions, are seldom found in normal tissue and are detectable through Next-generation sequencing (NGS) of the transcriptome (RNA-seq). To transform the data provided by the sequencer into robust gene fusion detection several analysis steps are needed. Usually the first step is to map the sequenced transcript fragments (RNA-seq) to a reference genome. One standard application of this approach is to estimate expression and detect variants within known genes, e.g. SNPs and indels. In case of gene fusions, however, completely novel gene structures have to be detected. Here, we describe the detection of such gene fusion events based on our comprehensive transcript annotation (ElDorado).

To demonstrate the utility of our approach, we extract gene fusion candidates from eight breast cancer cell lines, which we compare to experimentally verified gene fusions. We discuss several gene fusion events, like BCAS3–BCAS4 that was only detected in the breast cancer cell line MCF7. As supporting evidence we show that gene fusions occur more frequently in copy number enriched regions (CNV analysis). In addition, we present the Transcriptome Viewer (TViewer) a tool that allows to interactively visualize gene fusions. Finally, we support detected gene fusions through literature mining based annotations and network analyses.

In conclusion, we present a platform that allows detecting gene fusions and supporting them through literature knowledge as well as rich visualization capabilities. This enables scientists to better understand molecular processes, biological functions and disease associations, which will ultimately lead to better biomedical knowledge for the development of biomarkers for diagnostics and therapies.

1. Introduction

Gene fusions are found in many cancer types, and have been shown to be prognostic biomarkers in several studies [1–4]. A favorable property of gene fusions as biomarkers is their absence in normal tissue. Thus, targeting fusion transcripts can potentially provide a very specific biomarker. In addition, gene fusions often have a direct functional impact on the molecular processes in the cell. For instance, the well-studied TMPRSS2-ERG gene fusion leads to overexpression of ERG and hence the cancer develops androgen-independence, as seen in many prostate cancers [4].

Through the recent advances in Next-generation sequencing technology it has become possible to screen for known and novel gene fusion events on a genome wide scale. The prerequisite for a robust detection is a paired-end sequencing of the cell's transcriptome. Meanwhile, this sequencing has become a commodity and the bottlenecks in gene fusion detection have shifted towards data analysis and visualization.

Recently, many tools have been published that allow handling specific aspects of the data analysis pipeline needed to detect and validate gene fusions [5–8]. The first data analysis step is mapping the sequenced mRNA fragments to a reference [9,10]. For mapping paired-end RNA-seq data a mapper should have certain capabilities, these are: the ability to process paired-end information, the ability to align fragments that span over multiple exons and the ability to align fragments that originate from separate regions of the genome. The last capability is needed to specify the breakpoints that caused the gene fusion.

After having mapped the reads gene fusion detection tools can be applied. One major concern thereby is to reduce the number of false positive (FP) calls. When sequencing hundreds of millions of reads, a very small percentage of false alignments can lead to a FP gene fusion call. Thus, filtering steps in which suspicious gene fusion events or alignments are removed are paramount. Another important aspect of calling gene fusions is the transcript annotation reference used as basis for their detection. To this end, transcripts from a single resource like RefSeq [11] or Ensembl [12] are often used.

After the gene fusions have been detected they should be compared to known literature and inspected visually. The visual

* Corresponding author. Fax: +49 (0)89 599766 55.
E-mail address: supper@genomatix.de (J. Supper).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.09.013

inspection is an important step to provide a second level of quality control. The expression profiles of the putative gene fusion partners can, for instance, give an indication about the expression of the original genes and the fusion gene, and hence give insight into potential FP calls. Many gene fusion events like TMPRSSS2-ERG are reoccurring in certain cancer subtypes, thus it is helpful to scan the literature for gene fusions and the genes that constitute the fusion. This can be done through dedicated databases that contain gene fusion information, by mining through the known literature or by analyzing networks and gene regulatory effects.

In this work, we present a gene fusion pipeline that covers all aforementioned aspects. We have embedded several levels of quality control (QC), multiple FP filters, gene fusion visualization tools and support of the analysis through annotation and literature knowledge. The annotation used in this work extends over the usage of one transcript resource by generating a comprehensive but non-redundant transcript resource from RefSeq [11], Ensembl [12], GenBank [13] and by orthologous cross mapping. A novel tool presented here is the Transcriptome Viewer (TViewer) that, to the best of our knowledge, is the first tool to directly visualize detected gene fusions on the transcript level.

2. Methods and applications

2.1. NGS datasets

Illumina provided the NGS dataset used in this work within the iDEA challenge 2011 [14]. This dataset consists of eight well-characterized breast cancer cell lines obtained from the American Type Culture Collection (ATCC). These cell lines can be categorized in ER+ (MCF7, T47D, ZR-75-1, BT-474), ER− (MDA-MB-231, MDA-MB-468, BT-20) and non-tumorigenic (MCF10A). Hence, MCF10A is used as control. From this dataset we used the 50 bp paired-end RNA-seq data and low pass genomic sequencing data (50 bps) both sequenced on the Illumina Genome Analyzer. The complete dataset can be obtained from the Gene Expression Omnibus (GSE27003).

2.2. Annotations

As prerequisite for gene fusion calls known transcripts have to be provided. These are usually taken from a single resource like RefSeq or Ensembl. Here we use the ElDorado transcript annotation, a resource that integrates transcript annotations from RefSeq [11], Genbank (full-length cDNAs) [13] and Ensembl [12]. To process these transcripts and cDNAs all respective sequences are mapped to the genome. In a subsequent merging step, all transcripts with identical exon-exon boundaries and no more than 50 bps difference in their 5′- and 3′-ends are merged. In addition, transcripts are inferred in one organism from another with our proprietary transmapping approach. An orthologous transfer of a transcript annotation requires all exons and splice-junctions to match known transcript annotations.

Complementary to the genomic annotations, we have compiled an extensive literature and pathway database to relate detected gene fusions with published literature and biological networks. This database contains expert curated literature annotations from Molecular Connections (NetPro) and Genomatix. These annotations provide semantic information on the relations between genes. In addition, automatically extracted relations between genes, diseases and pathways are collected from NCBI's PubMed database. To understand gene fusions in the context of biological networks and gene regulatory interactions, various canonical pathway resources (e.g. NCI-nature pathway interaction database [15], Biocarta, Reactome [16]) are employed, as well as our transcription factor database MatBase.

2.3. Mapping

The RNA-seq and DNA-seq reads were both aligned to the genome; in addition the RNA-seq dataset was also aligned to the transcriptome. The mapper operates on seeds that are organized in a tree structure. The seed length can range from 8 to 25 bps, 8 being the shortest unique sub word found in the human reference genome and 25 being a length at which most subsequences in the genome are unique. For several reasons we check for multiple seeds in each read. First, parts of the read might not contain a seed. This may be caused by sequencing errors, SNPs, indels or ambiguous regions in the reference. Second, sequenced reads might not fit to the reference in one 'piece', i.e. – chromosomal rearrangements can lead to reads that span from one chromosome to another and RNA-seq reads often span over introns. To cover these cases for a given read multiple seeds are considered as anchor positions. These anchors are then used to align the complete read to the reference, using the Needleman-Wunsch [17] algorithm.

Using multiple seeds allows for the generation of spliced alignments. To do this, two general modes are available. One is the global spliced alignment that allows for breaking reads into two segments with arbitrary distances, or between chromosomes. This mode is used to find chromosomal break points. The second mode is the local spliced alignment, which performs a spliced alignment within a region of 1 million bps, using a subroutine that allows splicing the read multiple times. This mode is used to map mRNA sequences that span over multiple exons. For gene fusion detection we enable both modes.

To incorporate paired-end information into the mapping, the distances between the pairs are matched against the background distribution. If these are not within the expected distance (three standard deviations from the mean), suboptimal alignments are considered. If for one of these suboptimal alignments the mapping is within the expected distance this partner is reported as final alignment. This step is important to reduce the number of false positive gene fusion calls.

2.4. Gene fusion detection

We detect gene fusions of transcripts within and across chromosomes. This detection is based on paired-end reads that are uniquely mapped to transcripts from different loci. The approach consists of four steps. These are: filtering, clustering, inclusion of splice-junction information and scoring. The filtering removes read-pairs if one or both pairs do not uniquely map to the genome, if the pairs align within the expected distance or if any pair is contained in an artificial pileup. All pairs that pass the filters and are mapped to transcripts from different loci are considered for the second step. In the second step, mate-pairs that map to proximal positions and have the same strand orientation are clustered, to generate gene fusion predictions.

For each gene fusion prediction, single-end reads spanning (spliced-alignment) the transcripts are used to determine the exact breakpoint. In the last step, several scores are determined to rank the gene fusion candidates. The first two scores are the number of spliced reads and the number of paired-end reads that support the gene fusion. In addition, a breakpoint score is calculated. This score quantifies the difference between the read coverage upstream and down-stream of the breakpoint. To calculate the breakpoint scores the coverage of the regions 5′ and 3′ of the breakpoint region are compared to the coverage between the breakpoints. The breakpoint score (ranging from 0 to 1) is calculated as:

$(SeqCov_{max} - SeqCov_{min})/SeqCov_{max}$

where $SeqCov_{max}$ is the higher and $SeqCov_{min}$ the lower coverage value.

2.5. Copy number variations (CNVs)

To determine copy number variations a gene centric approach is employed. For each annotated gene locus the number of DNA-seq reads aligning to the locus are counted. From this mapping a normalized copy number value (NE) [18] is calculated for each gene.

2.6. Visualization of gene fusions (TViewer)

The Genomatix Transcriptome Viewer (TViewer – Fig. 1) provides a framework for visually integrating transcript annotations and gene fusions with RNA-seq data. In the standard view each transcript is drawn separately whereas the splicing graph view identical exons and promoters are merged. Exons, promoters and splice junctions are drawn according to their read coverage. The integrated paired-end viewer shows the coverage of sequenced fragments and the distance of the respective paired-end reads along the transcript. Fragments are only considered if both mates uniquely map. This indicates whether the transcript is completely covered with reads and if these are within the expected distance. The gene fusion visualization allows for interactively visualizing not only known transcripts but also novel fusion products. Expression information is also displayed for each exon and splice-junction and can be plotted for the entire fusion transcript at base resolution.

3. Results

Several gene fusions were found that occur in multiple tumorigenic cell lines but not in MCF10A. For instance, the gene fusion EEF1A2-PKN1 has breakpoints that are identically found in the cell lines BT-474 and T47D, and that deviate at most 19 bps in the other cell lines. Another example is the gene fusion SREBF1-NUP210 that occurs in four tumorigenic cell lines. Here, the breakpoints in MCF-7 and BT-20 are identical in the 5′ transcript and are three bps apart in the 3′ transcript. Thus, both fusions candidates produce the same protein, except that one amino acid is deleted in the BT-20 cell line. In the cell line T47D the gene fusion candidate RERG-CBFB was found. RERG is a member of the RAS super family that inhibits cell proliferation, tumor formation and it is estrogen-regulated.

For the MCF-7 cell line 16 gene fusion candidates across chromosomes, 17 gene fusion candidates within chromosomes and 14 read through candidates were detected (Table 1). For 16 of these gene fusions additionally spliced reads are found that span over the fusion breakpoint. As additional information, for each gene involved in the gene fusions the NE value obtained from the CNV analysis is provided (Table 1). These numbers indicate that many fused genes have high copy number values (0.29 is the average copy number value over all MCF-7 genes).

To validate the detected gene fusion candidates they were compared to gene fusions reported recently by Sakarya et al. [7]. In this work, 21 of the gene fusion candidates found here have been experimentally validated. Furthermore, Edgren et al. recently published a paper [5] in which they validated three MCF-7 gene fusions and eleven BT-474 gene fusions. All of their validated MCF-7 gene fusions and nine of the eleven gene fusions they report for BT-474 are detected.

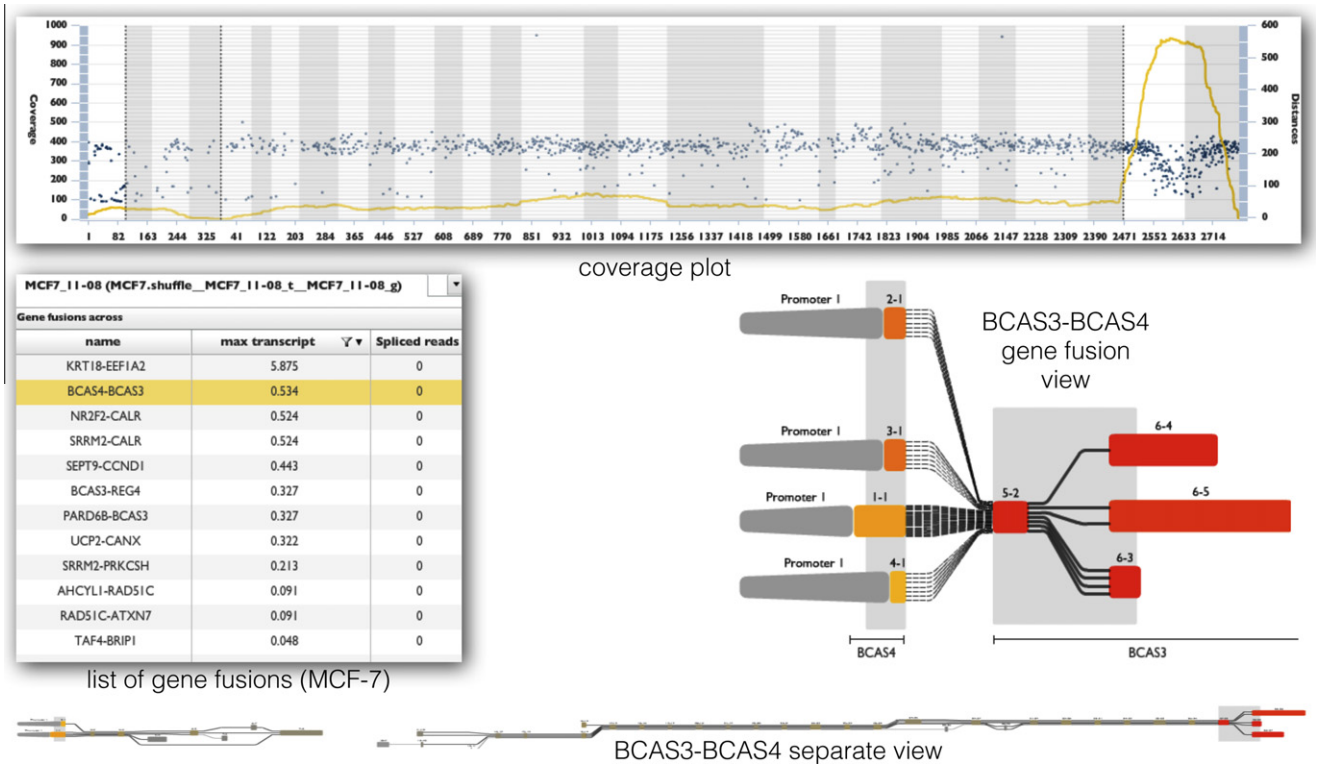

Fig. 1. TViewer visualization of the gene fusion BCAS3–BCAS4 in MCF7. The paired-end view on top shows the coverage and distances of the fused transcripts BCAS4–BCAS3. The shaded area represents skipped exons. Here, we see low coverage on the skipped exons and high coverage on the fashioned exons. In the gene fusion view the dotted lines represent the fusions and the grey areas represent the breakpoint regions. The breakpoint region is the sequence within a transcript between the outmost 5′ and 3′ mates, which support the detected gene fusion.

Table 1
MCF-7 gene fusion candidates within and across chromosomes. Colored fusions have been validated [7]. Given are CNVs of fused genes, read-pairs and spliced-reads spanning over fusions.

5' gene			3 ' gene				
name	**chr**	CNV	name	chr	CNV	read pairs	spliced reads
ARFGEF2	20	0.50	SULF2	20	4.80	145	25
TANC2	17	0.71	CA4	17	5.10	40	14
BCAS4	20	2.09	BCAS3	17	2.56	896	11
RPS6KB1	17	3.63	TMEM49	17	1.13	47	6
SULF2	20	4.80	PRICKLE2	3	3.48	17	3
C16orf45	16	0.31	ABCC1	16	0.36	6	3
SMARCA4	19	0.44	CARM1	19	0.52	21	2
MYO9B	19	0.25	FCHO1	19	0.30	17	2
ATXN7L3	17	0.26	FAM171A2	17	0.36	5	2
BCAS3	17	2.56	ATXN7	3	1.61	2	2
GCN1L1	12	0.44	MSI1	12	0.44	15	1
BCAS4	20	2.09	ZMYND8	20	5.13	8	1
TAF4	20	0.71	BRIP1	17	4.86	8	1
MYH9	22	0.21	EIF3D	22	0.27	6	1
BCAS3	17	2.56	REG4	1	0.65	4	1
SYTL2	11	0.38	PICALM	11	0.29	2	1
RNF4	4	0.31	LOC644006	1	0.42	14	0
ABCA10	17	0.50	PPP4R1L	20	1.06	8	0
GATAD2B	1	0.41	NUP210L	1	0.32	7	0
NAV1	1	0.44	GPR37L1	1	0.93	7	0
C10orf112	10	0.21	PLXDC2	10	0.22	6	0
GGA2	16	0.48	ZFAND5	9	0.21	5	0
HIPK1	1	1.26	DENND2C	1	1.49	5	0
SMAD9	13	0.32	SMAD5	5	0.27	5	0
AHCYL1	1	0.31	RAD51C	17	2.85	4	0
PARD6B	20	4.09	BCAS3	17	2.56	4	0
PDE8A	15	0.46	SCAND2	15	0.54	4	0
PNPLA7	9	0.69	WDR85	9	0.46	4	0
C18orf25	18	0.14	LOC729141	2	0.27	3	0
MTG1	10	0.31	LOC619207	10	0.42	3	0
NCRNA0182	X	0.19	AK124009	X	0.15	3	0
PLCG1	20	1.09	TOP1	20	0.69	3	0
RAD51C	17	2.85	ATXN7	3	1.61	3	0
SHANK2	11	0.52	SHANK3	22	0.23	3	0
SREBF1	17	0.42	NUP210	3	0.64	2	0
USP31	16	0.32	CRYL1	13	0.42	2	0

To analyze the CNV enriched regions, two such regions within the genome of the MCF-7 cell line data were extracted, one on chromosome 17 and one on chromosome 20. These regions have a strong overlap with the enriched regions of the BT-474 cell line. For the selected regions the list of contained genes was extracted. To understand the biological implications of enriching or deleting chromosomal regions, all genes were uploaded into the Genomatix Pathway System (GePS) to perform gene set enrichment analyses and to view the genes in light of regulatory networks. The most significant disease enrichment for these genes was the breast neoplasm's network (p-value: 2.57E-07) and the most significant tissue was breast (p-value: 3.14E-05).

Genes in regions with high copy numbers could be affected in many ways. For instance, genes with high copy numbers could exhibit higher expression. Furthermore, genes that exist in many copies might be more likely to fuse with other genes. The gene fusion candidates have an average CNV of 1.18, where all genes in MCF-7 have an average CNV of 0.34. Although the CNVs of the gene fusion candidates are by no means an indication of a fusion event, it was observed that a high number of fused genes have high CNVs. Owing to this fact many gene fusions can be observed within the breast neoplasm's network. There, 18.75% (6 of 32) of the contained genes are gene fusion candidates. Overall only 0.20% (65 of 32099) of the genes are gene fusion candidates.

4. Conclusion

Many cancer therapies are motivated by the concept that certain genotypes can predict a therapeutic response and provide a prognosis. For instance, the expression of the ESR1 gene is a known biomarker for the endocrine therapy. Microarrays have allowed screening for such expression based biomarkers on a genome wide scale.

The developments in sequencing and analyzing RNA-seq data have extended the capability of genome wide screens to other targets like gene fusions, SNPs, indels and novel isoforms. Here we concentrated on gene fusions and presented a pipeline for robust gene fusion detection. Searching for genetic variants that have been caused by mutations and chromosomal rearrangement can potentially lead to very specific biomarkers, because these variants often only occur in the tumor cells and not in any normal tissue type.

To this end, the presented pipeline enables medical researchers to utilize a robust and comprehensive platform for detecting, visualizing and interpreting gene fusions.

Acknowledgement

We would like to thank Carri-Lyn Mead for providing the dataset within the Illumina iDEA challenge 2011 and Hannes Planatscher for helpful scientific discussions on the Transcriptome Viewer (TViewer).

References

[1] B.G. Barwick, M. Abramovitz, M. Kodani, C.S. Moreno, R. Nam, W. Tang, M. Bouzyk, A. Seth, B. Leyland-Jones, British Journal of Cancer 102 (2010) 570–576. URL: <http://dx.doi.org/10.1038/sj.bjc.6605519>.

[2] D.V. Makarov, S. Loeb, R.H. Getzenberg, A.W. Partin, Annual Review of Medicine 60 (1) (2009) 139–151. URL: <http://dx.doi.org/10.1146/annurev.med.60.042307.110714>.

[3] S.A. Tomlins, D.R. Rhodes, S. Perner, S.M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J.E. Montie, R.B. Shah, K.J. Pienta, M.A. Rubin, A.M. Chinnaiyan, Science 310 (5748) (2005) 644–648. URL: <http://dx.doi.org/10.1126/science.1117679>.

[4] J. Yu, J. Yu, R.-S. Mani, Q. Cao, C.J. Brenner, X. Cao, X. Wang, L. Wu, J. Li, M. Hu, Y. Gong, H. Cheng, B. Laxman, A. Vellaichamy, S. Shankar, Y. Li, S.M. Dhanasekaran, R. Morey, T. Barrette, R.J. Lonigro, S.A. Tomlins, S. Varambally, Z.S. Qin, A.M. Chinnaiyan, Cancer Cell 17 (5) (2010) 443–454. URL: <http://dx.doi.org/10.1016/j.ccr.2010.03.018>.

[5] H. Edgren, A. Murumagi, S. Kangaspeska, D. Nicorici, V. Hongisto, K. Kleivi, I. Rye, S. Nyberg, M. Wolf, A.L.B. Dale, O. Kallioniemi, Genome Biology 12 (1) (2011) R6. URL: <http://dx.doi.org/10.1186/gb-2011-12-1-r6>.

[6] R. Piazza, A. Pirola, R. Spinelli, S. Valletta, S. Redaelli, V. Magistroni, C. Gambacorti-Passerini, Nucleic Acids Research 40 (16) (2012) e123. URL: <http://dx.doi.org/10.1093/nar/gks394>.

[7] O. Sakarya, H. Breu, M. Radovich, Y. Chen, Y.N. Wang, C. Barbacioru, S. Utiramerur, P.P. Whitley, J.P. Brockman, P. Vatta, Z. Zhang, L. Popescu, M.W. Muller, V. Kudlingar, N. Garg, C.-Y. Li, B.S. Kong, J.P. Bodeau, R.C. Nutter, J. Gu, K.S. Bramlett, J.K. Ichikawa, F.C. Hyland, A.S. Siddiqui, PLoS Computational Biology 8 (4) (2012) e1002464. URL: <http://dx.doi.org/10.1371/journal.pcbi.1002464>.

[8] X.-S. Wang, J.R. Prensner, G. Chen, Q. Cao, B. Han, S.M. Dhanasekaran, R. Ponnala, X. Cao, S. Varambally, D.G. Thomas, T.J. Giordano, D.G. Beer, N. Palanisamy, M.A. Sartor, G.S. Omenn, A.M. Chinnaiyan, Nature Biotechnology 27 (11) (2009) 1005–1011. URL: <http://dx.doi.org/10.1038/nbt.1584>.

[9] C. Trapnell, L. Pachter, S.L. Salzberg, Bioinformatics 25 (9) (2009) 1105–1111. URL: <http://dx.doi.org/10.1093/bioinformatics/btp120>.

[10] G. Jean, A. Kahles, V.T. Sreedharan, F. De Bona, G. Rätsch, Current Protocols in Bioinformatics/Editoral Board, Andreas D. Baxevanis, 2010 (Chapter 11: Unit 11.6). URL: <http://dx.doi.org/10.1002/0471250953.bi1106s32>.

[11] K.D. Pruitt, T. Tatusova, G.R. Brown, D.R. Maglott, Nucleic Acids Research 40 (Database issue) (2012) D130–D135. URL: <http://dx.doi.org/10.1093/nar/gkr1079>.

[12] P. Flicek, M.R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A.K. Kahari, D. Keefe, S. Keenan, R. Kinsella, M. Komorowska, G. Koscielny, E. Kulesha, P. Larsson, I. Longden, W. McLaren, M. Muffato, B. Overduin, M. Pignatelli, B. Pritchard, H.S. Riat, G.R.S. Ritchie, M. Ruffier, M. Schuster, D. Sobral, Y.A. Tang, K. Taylor, S. Trevanion, J. Vandrovcova, S. White, M. Wilson, S.P. Wilder, B.L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X.M. Fernandez-Suarez, J. Harrow, J. Herrero, T.J.P. Hubbard, A. Parker, G. Proctor, G. Spudich, J. Vogel, A. Yates, A. Zadissa, S.M.J. Searle, Nucleic Acids Research 40 (D1) (2012) D84–D90. URL: <http://dx.doi.org/10.1093/nar/gkr991>.

[13] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler, Nucleic acids research 36 (Database issue) (2008) D25–D30. URL: <http://dx.doi.org/10.1093/nar/gkm929>.

[14] Z. Sun, Y.W. Asmann, K.R. Kalari, B. Bot, J.E. Eckel-Passow, T.R. Baker, J.M. Carr, I. Khrebtukova, S. Luo, L. Zhang, G.P. Schroth, E.A. Perez, E.A. Thompson, PloS One 6 (2) (2011) e17490. URL: <http://dx.doi.org/10.1371/journal.pone.0017490>.

[15] C.F. Schaefer, K. Anthony, S. Krupa, J. Buchoff, M. Day, T. Hannay, K.H. Buetow, Nucleic Acids Research 37 (Database issue) (2009) D674–D679. URL: <http://dx.doi.org/10.1093/nar/gkn653>.

[16] D. Croft, G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N. Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio, L. Stein, Nucleic Acids Research 39 (Database issue) (2011) D691–D697. URL: <http://dx.doi.org/10.1093/nar/gkq1018>.

[17] M. Waterman, T. Smith, W. Beyer, Advances in Mathematics 20 (3) (1976) 367–387. URL: <http://dx.doi.org/10.1016/0001-8708(76)90202-4>.

[18] M. Sultan, M.H. Schulz, H. Richard, A. Magen, A. Klingenhoff, M. Scherf, M. Seifert, T. Borodina, A. Soldatov, D. Parkhomchuk, D. Schmidt, S. O'Keeffe, S. Haas, M. Vingron, H. Lehrach, M.-L. Yaspo, Science 321 (5891) (2008) 956–960. URL: <http://dx.doi.org/10.1126/science.1160342>.