Journal of Bioinformatics and Computational Biology
© Imperial College Press

# *In silico* search for natural antisense transcripts reveals their differential expression in human tumors.

DMITRY KLIMOV

*Institute of General Genetics RAS,*
*Moscow, 119991, Russia, Gubkin Str. 3*
*dklimov@vigg.ru*

MIKHAIL SKOBLOV

*Medical Genetics Research Center RAMS,*
*Moscow, 115478, Russia, Moskvorechie Str. 1*
*mskoblov@medgen.ru*

ANDREI RYAZANTZEV

*Moscow Institute of Radiotechnics Electronics and Automatics,*
*Moscow, 119454, Russia, Vernadsky avenue, 78*
*kardox@mail.ru*

TATIANA TYAZHELOVA

*National Hematology Research Centre,*
*Moscow, 125167, Russia, Novozykovsky proezd, 4a*
*tatvt@list.ru*

ANCHA BARANOVA

*Center for the Study of Genomics in Liver Diseases, Molecular Biology and Microbiology Department,*
*George Mason University, Fairfax, VA., USA;*
*George Mason-Inova Health System's Translational Research Centers, Falls Church, VA, USA;*
*Institute of General Genetics RAS,*
*Moscow, 119991, Russia, Gubkin Str. 3*
*abaranov@gmu.edu*

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

We created an algorithm that allows high-throughput mapping of sense-antisense (SA) pairs of transcripts. By this method we mapped approximately 32,000 SA pairs of human mRNAs. Collected SA pairs were divided into three groups: SA pairs based on two or more UniGene clusters (17% of all sense-antisense pairs), SA pairs based on ESTs that belong to the same UniGene cluster (42%), and SA pairs formed by UniGene cluster and non-unique unclustered transcripts (41%). To study expression patterns of natural SA pairs we created a software application "Antisense Cluster Filter". We retrieved tissue expression data for all the transcripts forming identified SA pairs, including clustered and unclustered ones. After that, we selected 108 SA pairs represented by transcripts differentially regulated in human tumors. For each of these SA pairs one of the transcripts was expressed only in tumors, another one was expressed both in non-malignant and malignant tissues. Indicated SA pairs may represent a new class of tumor markers. An example of the tumor-specific natural antisense to C3orf4 mRNA is detailed.

*Keywords*: natural antisense transcripts; tumor markers, UniGene clusters

## 1. Introduction

The availability of the complete human genome sequence [1,2] and the accumulation of millions of expressed sequences (mRNAs and ESTs) allowed one to perform large-scale studies of naturally occurring antisense transcription. Indeed, in the past two years, several studies have used publicly available sources of information in attempts to produce comprehensive datasets of sense–antisense pairs [3-7].

Both mRNA expression in a eukaryotic cell and efficiency of its translation into proteins are controlled by multiple regulatory levels subsequent to transcription initiation. As mRNA is a single strand molecule, the expression of a complementary antisense strand may alter the rate of transcription initiation and elongation, mRNA processing and stability, as well as the rate of translation of the template RNA [7]. Functional antisense RNAs have been identified in bacteria[8], but later were shown to be involved in gene regulation and differentiation in several eukaryotic organisms, including mammals[7]. Natural antisense transcripts usually arise via separate transcription initiation on the opposite DNA strand at the same genomic locus as the sense strand. Computational analysis of the data obtained in large-scale sequencing projects has revealed a surprising abundance of antisense transcripts in several eukaryotic genomes. As some antisense transcripts have been shown to regulate gene expression, it is possible that antisense production might be a common mechanism regulating gene expression in eukaryotic cells. Tumors are characterized by general deregulation of transcription initiation revealing itself as “illegitimate transcription” or “mass-production of non-coding RNAs” [9,10]. We hypothesized that this phenomenon could be partially explained by differential regulation of antisense transcripts in human tumors and found 108 sense-antisense (SA) pairs represented by such transcripts. For each of these SA pairs one of the transcripts was expressed only in normal tissues, another one was expressed only in tumors. Indicated SA pairs may represent a new class of tumor markers.

## 2. Materials and methods

Publicly available complete set of human ESTs mapped to human genome was retrieved from GenBank (http://www.ncbi.nlm.nih.gov/entrez/query. fcgi?db =Nucleotide). Information on human EST clusters was retrieved from UniGene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene). Solitary ESTs mapped to human genome as singletons not supported by any UniGene clusters were considered as potential artifacts and were excluded from further studies. Non-solitary ESTs obtained from independent cDNA molecules were considered as true representatives of transcriptome even if they were not supported by existing UniGene clusters. cDNA library descriptions available from CGAP website (CGAP http://cgap.nci.nih.gov/) and other sources[10] were used.

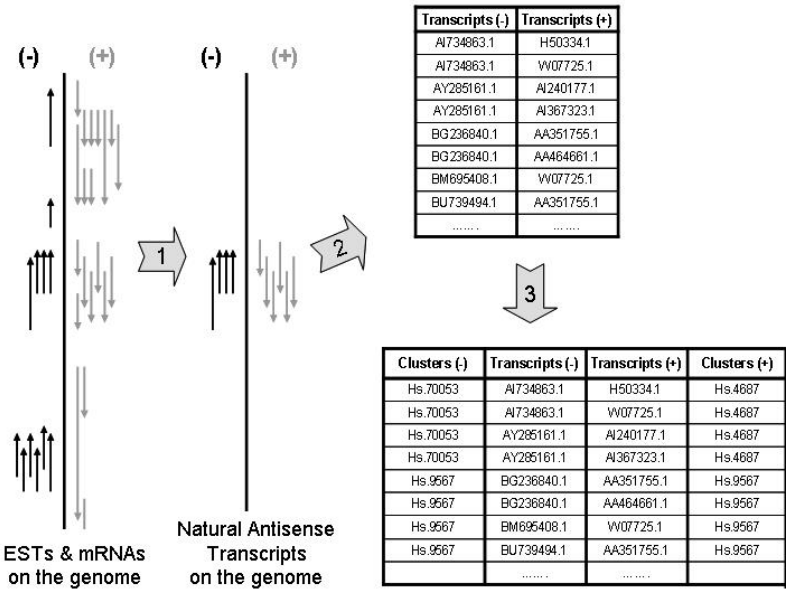
We created an algorithm "Antisense Cluster Filter" implemented in C++. This algorithm allows high-throughput mapping of sense-antisense (SA) pairs of transcripts by 1) retrieving all overlapping pairs that are located on different DNA strands with more than 20 nucleotide overlaps; 2) retrieving all pairs of transcripts that are associated with UniGene clusters; 3) retrieving an intersection of two sets of the above-mentioned sequences. EST clusters containing less than three ESTs were filtered out. Mapping was performed by comparing exact coordinates of the transcript and its orientations on the plus/minus chains of the human genome of the NCBI assembly 35 v. 1 (ftp://ftp.ncbi.nih.gov/genebank/genomes/H_Sapiens). See algorithm schematic in Fig.1.

Natural tumor-specific SA pairs were selected for further analysis according to the following criteria: 1) sense cluster contains more than 10 ESTs; 2) antisense cluster contains more than 10 ESTs; 3) antisense cluster contains no more than 10% of ESTs originating in non-malignant tissues.

## 3. Results

### *2.1. In silico search for natural antisense transcripts in human genome.*

Algorithm "Antisense Searcher" allowed us to retrieve approximately 32,000 sense-antisense human (SA) pairs and map them on human genome (Fig.1). Collected SA pairs were divided into three groups: SA pairs based on two or more UniGene clusters (17%), SA pairs based on ESTs that belong to the same UniGene cluster (42%), and SA pairs formed by UniGene cluster and non-solitary unclustered transcripts (41% of all pairs).




| Transcripts (-) | Transcripts (+) |
|---|---|
| AI734863.1 | H50334.1 |
| AI734863.1 | W07725.1 |
| AY285161.1 | AI240177.1 |
| AY285161.1 | AI367323.1 |
| BG236840.1 | AA351755.1 |
| BG236840.1 | AA464661.1 |
| BM695408.1 | W07725.1 |
| BU739494.1 | AA351755.1 |
| ..... | ..... |

| Clusters (-) | Transcripts (-) | Transcripts (+) | Clusters (+) |
|---|---|---|---|
| Hs.70053 | AI734863.1 | H50334.1 | Hs.4687 |
| Hs.70053 | AI734863.1 | W07725.1 | Hs.4687 |
| Hs.70053 | AY285161.1 | AI240177.1 | Hs.4687 |
| Hs.70053 | AY285161.1 | AI367323.1 | Hs.4687 |
| Hs.9567 | BG236840.1 | AA351755.1 | Hs.9567 |
| Hs.9567 | BG236840.1 | AA464661.1 | Hs.9567 |
| Hs.9567 | BM695408.1 | W07725.1 | Hs.9567 |
| Hs.9567 | BU739494.1 | AA351755.1 | Hs.9567 |
| | ..... | ..... | |

Fig.1. In silico search for natural SA pairs of transcripts in human genome

### *2.2. Expression patterns of natural SA pairs in human genome.*

To study expression patterns of natural SA pairs we created an algorithm “Antisense Cluster Filter” implemented in C++. This software allowed us to retrieve tissue expression data for all the transcripts forming identified SA pairs, including clustered and unclustered ones.

According to tissue expression data, we selected 108 SA pairs represented by transcripts differentially regulated in human tumors. For each of these SA pairs one of the transcripts was expressed only in tumors, another one was expressed both in non-malignant and malignant tissue sources. One of the most prominent examples of tumor-specific SA pairs is described on Fig. 2. Indicated SA pairs may represent a new class of tumor markers.

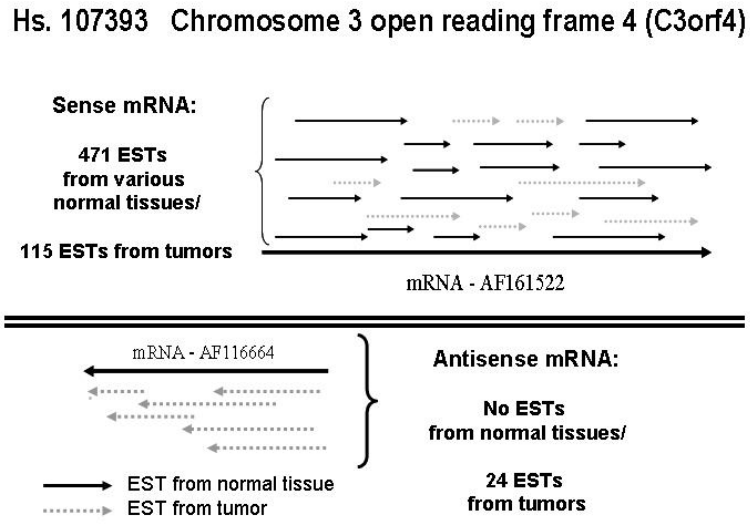



Fig.2. Example of sense-antisense pair when one of pair members in expressed only in tumor cells.

## 4. Discussion

Tumor markers serve as valuable instruments of tumor detection and monitoring, but sensitivity and specificity of tumor marker assays remain regrettably low. Most probably, the future is in tumor detection panels encompassing anywhere from a dozen to hundreds of marker molecules. The level of expression of such markers in each individual tumor most likely is a function of the degree of derangement in the rates of transcription initiation, in alternative splicing events, and in the suppression of the intron retention. An increase in the level of antisense transcription could be one of the facets of tumor-specific expression derangement. Such an increase will reveal itself in novel sense-antisense (SA)

pairs arising in tumor tissue. If the sense molecule possesses tumor suppressor function, indicated antisense molecules may serve as potential oncogenes by downregulation of the cellular level of sense transcript. Expression of such oncogenic antisense transcripts will be subject to positive selection, which will lead to an increase in the percentage of tumor cells expressing antisense within a tumor mass. This, in turn, will increase the chances of antisense transcript detection, and will allow one to profile tumor progression as an increase in antisense transcript levels in subsequent samples taken from the tumor mass. If a sense molecule does not possess tumor suppressive function, the relative level of the antisense molecule expressed *de novo* still could serve as an indicator of tumor progression, as it reflects a combination of per cell and per locus rate of antisense transcription in a given tumor specimen. Based on this logic, we propose that a targeted search for tumor-specific SA pairs might yield a novel set of tumor marker candidates deserving attention both for tumor monitoring and tumor vaccine research. In this study we attempted to select the first set of such candidates.

An example detailed in this paper is a tumor-specific natural antisense to recently characterized human gene C3orf4 located at 3q12.1[11]. This gene is predominantly expressed in cultured oligodendrocytes, the myelinating cells of the central nervous system, but not in astrocytes. In addition to brain, C3orf4 transcripts are found in the heart and, in lesser amounts, in other adult and fetal tissues[11]. Analysis of the corresponding UniGene cluster Hs.107393 supports experimental data. It was previously hypothesized that the C3orf4 gene encodes a membrane protein that could be involved in the differentiation processes at the interface between oligodendrocytes and neurons [11]. In addition to that, we found that expression levels of C3orf4 in tumors may be perturbed by the natural antisense interference. It is tempting to speculate that the loss of C3orf4 function caused by illegitimate antisense RNA production may support the suppression of the cell differentiation and, therefore, may uphold the tumor phenotype.

In our opinion, this and other tumor-specific natural antisense molecules may serve as tumor marker candidates suitable for further experimental studies.

**Acknowledgements**

Authors are extremely grateful to Prof. A. P. Kozlov (Biomedical Center, St. Petersburg, Russia) for scientific discussions leading to important insights, and to Prof. Nick Yankovsky (Moscow, Russia), who inspired the authors to proceed with this study. A.B. was partially covered by NIH 1R15CA113331-01.

## 5. References

1. J.C. Venter et al., “The sequence of the human genome,” Science. 291(5507), 1304-1351 (2001).
2. International Human Genome Sequencing Consortium, “Initial sequencing and analysis of the human genome”, Nature. 409, 860 - 921 (2001).
3. J. Shendure and G. M. Church, “Computational discovery of sense–antisense transcription in the human and mouse genomes,” Genome Biol. 3, 1–14 (2002).

4. B. Lehner, G.Williams, R.D. Campbell and C.M. Sanderson, “Antisense transcripts in the human genome,” Trends Genet. 18(2), 63–65 (2002).
5. R. Yelin, D. Dahary, R. Sorek, E.Y. Levanon, O. Goldstein et al., “Widespread occurrence of antisense transcription in the human genome,”. Nat. Biotechnol. 21(4), 379–386 (2003). .
6. H. Kiyosawa, I. Yamanaka, N. Osato, S. Kondo, Y. Hayashizaki, RIKEN GER Group; GSL Members, “Antisense transcripts with FANTOM2 clone set and their implications for gene regulation,” Genome Res. 13 (6b), 1324–1334 (2003).
7. G. Lavorgna, D. Dahary, B. Lehner, R. Sorek, C.M. Sanderson, G. Casari, “In search of antisense,” Trends Bioche. Sci 29(2), 88-94 (2004).
8. S. Gottesman, “Micros for microbes: non-coding regulatory RNAs in bacteria,” Trends Genet. 21(7):399-404 (2004).
9. A.P. Kozlov, “Gene competition and the possible evolutionary role of tumours,” Med Hypotheses. 46(2):81-4 (1986).
10. A.V. Baranova, A.V. Lobashev, D.V. Ivanov, L.L. Krukovskaya, N.K. Yankovsky, A.P. Kozlov, “In silico screening for tumour-specific expressed sequences in human genome,” FEBS Lett. 508(1):143-8 (2001).
11. N.A. Fayein, B. Stankoff, C. Auffray, MD Devignes, “Characterization of tissue expression and full-length coding sequence of a novel human gene mapping at 3q12.1 and transcribed in oligodendrocytes “. Gene. 289(1-2):119-129 (2002).

**Klimov Dmitry** received his BS degree in applied mathematics and physics from dept. of General & Applied Physics, Moscow Institute of Physics and Technology (MIPT) in 2004. He is now a graduate student at MIPT.

**Skoblov Mikhail** received his Ph.D. degree in Genetics from Institute of General Genetics, Russia, in 2004. From April 2005 he is at Medical Genetics Research Center as a postdoctoral researcher. His research interests are bioinformatics, genome expression, transcription factors and promoters, transcriptional regulation.

**Andrei Ryazantzev** studying since September 2003 in Moscow Institute of Radiotechnics Electronics and Automatics on "Computing Machinery and Systems" faculty by specialty "Computing Machinery and Automated Systems Software Engineering".

**Tatiana Tyazhelova** received her Ph.D degree in Genetics from Institute of General Genetics, Russia, in 2003. From December 2004 she is at National Hematology Research Centre as a postdoctoral researcher. Her research interests are bioinformatics and regulation of gene expression in eukaryotes.

**Ancha Baranova** received her Ph.D degree in Genetics in 1998 and D.Sci degree in Genetics in 2004 from Institute of General Genetics, Russia. From June 2002 she is an Assistant Professor in Molecular and Microbiology Department, George Mason University, Fairfax, Virginia, USA. Her research interests are functional genomics and bioinformatics.