

RESEARCH

Open Access

HOXD-AS1 is a novel lncRNA encoded in *HOXD* cluster and a marker of neuroblastoma progression revealed via integrative analysis of noncoding transcriptome

Aliaksandr A Yarmishyn^{1†}, Arsen O Batagov^{1†}, Jovina Z Tan¹, Gopinath M Sundaram², Prabha Sampath^{2,3,4}, Vladimir A Kuznetsov^{1,5*}, Igor V Kurochkin^{1*}

From Asia Pacific Bioinformatics Network (APBioNet) Thirteenth International Conference on Bioinformatics (InCoB2014)
Sydney, Australia. 31 July - 2 August 2014

Abstract

Background: Long noncoding RNAs (lncRNAs) constitute a major, but poorly characterized part of human transcriptome. Recent evidence indicates that many lncRNAs are involved in cancer and can be used as predictive and prognostic biomarkers. Significant fraction of lncRNAs is represented on widely used microarray platforms, however they have usually been ignored in cancer studies.

Results: We developed a computational pipeline to annotate lncRNAs on popular Affymetrix U133 microarrays, creating a resource allowing measurement of expression of 1581 lncRNAs. This resource can be utilized to interrogate existing microarray datasets for various lncRNA studies. We found that these lncRNAs fall into three distinct classes according to their statistical distribution by length. Remarkably, these three classes of lncRNAs were co-localized with protein coding genes exhibiting distinct gene ontology groups. This annotation was applied to microarray analysis which identified a 159 lncRNA signature that discriminates between localized and metastatic stages of neuroblastoma. Analysis of an independent patient cohort revealed that this signature differentiates also relapsing from non-relapsing primary tumors. This is the first example of the signature developed via the analysis of expression of lncRNAs solely. One of these lncRNAs, termed HOXD-AS1, is encoded in *HOXD* cluster. HOXD-AS1 is evolutionary conserved among hominids and has all *bona fide* features of a gene. Studying retinoid acid (RA) response of SH-SY5Y cell line, a model of human metastatic neuroblastoma, we found that *HOXD-AS1* is a subject to morphogenic regulation, is activated by PI3K/Akt pathway and itself is involved in control of RA-induced cell differentiation. Knock-down experiments revealed that HOXD-AS1 controls expression levels of clinically significant protein-coding genes involved in angiogenesis and inflammation, the hallmarks of metastatic cancer.

Conclusions: Our findings greatly extend the number of noncoding RNAs functionally implicated in tumor development and patient treatment and highlight their role as potential prognostic biomarkers of neuroblastomas.

* Correspondence: vladimirk@bii.a-star.edu.sg; igork@bii.a-star.edu.sg

† Contributed equally

¹Department of Genome and Gene Expression Data Analysis, Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Matrix, 138671 Singapore

Full list of author information is available at the end of the article

Background

Whereas the coding sequences constitute only 1.5% of human genome, the noncoding constituent of the genome has recently come into light as functionally important. The most well studied part of noncoding transcriptome is represented by microRNAs (miRNAs), which affect expression of protein coding genes by modulating mRNA stability and translation. miRNAs regulate various biological processes and have well established roles in cancer [1]. However, long noncoding RNAs (lncRNAs) represent the largest fraction of transcriptional output in the mammalian genome. Unlike extensively studied protein coding transcripts or miRNAs, the majority of lncRNAs still remain “the dark matter of the genome” with their functionality being debatable. Several lncRNAs have been shown in recent years to be directly involved in oncogenesis [2]. Reprogramming of chromatin state mediated by lncRNA HOTAIR is linked to metastasis and poor prognosis in breast cancer [3]. MALAT-1 has originally been shown to be associated with metastasis in lung carcinomas [4] and regulate invasive potential of tumor cells [5].

Significant fraction of lncRNAs is represented on widely used microarray platforms, however they have traditionally been ignored in cancer studies. Thousands of microarray datasets accumulated over years in public databases provide an immense resource for analysis of lncRNA expression. In this study we set an example of reanalysis of this rich data source by annotating the noncoding probe sets of the popular Affymetrix U133 microarray and applied this annotation to identify lncRNAs associated with aggressive and nonaggressive types of neuroblastoma (NB).

NB is a common pediatric malignancy characterized by tremendous clinical heterogeneity, in which some tumors are extremely aggressive and drug resistant while others can spontaneously differentiate into benign forms and are sensitive to differentiation stimuli [6]. According to International NB Staging System (INSS), based on tumor characteristics and age, NB is classified into 3 localized stages (stages 1, 2a and 2b), advanced locoregional (stage 3) and metastatic (stages 4 and 4s) [7]. About one-half of NB patients have metastatic disease at diagnosis.

One of the most widely used prognostic genomic markers is amplification of *MYCN* locus, which occurs in 20% of NB cases. It is associated with poor survival and is considered a promising therapeutic target in NB [8,9]. All metastatic tumors with amplified *MYCN* gene are aggressive, whereas tumors with non-amplified *MYCN* gene have variable clinical behavior that could be driven by clinical (stage, histology, age at diagnosis etc) and biological factors (mutations, chromosome rearrangements, gene expression profiles).

A large number of clinical genomics and experimental model studies have described protein-coding gene

expression profiles that can differentiate NBs with favorable and unfavorable outcomes [10-14]. The protein-coding gene signatures of NB aggressiveness are characterized by over-expression of genes involved in *MYC* and β -catenin pathways, cell cycle and chromosome segregation, and on the other hand by low expression of genes involved in neural differentiation [15-19]. Unfortunately, reliability and consistency between the gene signatures reported in these and other publications are very poor (V.A. Kuznetsov, unpublished). It suggests high plasticity of gene expression patterns in NB tumors, a large number of correlations between potential biomarkers and high dimensionality of their space. Due to poor understanding of molecular basis of NB and high disease-caused mortality of the NB patients there is an urgent need for discovery of novel molecular drivers of NB progression and its differentiation, as well as high-confidence and reliable clinical biomarkers specifying NB tumors, prognosis of clinical outcome and improvement of NB patients' therapy.

The degree of NB differentiation correlates with stage of NB and its progression. Retinoic acid (RA) is a differentiation agent that is the first choice drug for NB therapy [20]. However, in many cases NB acquires resistance to RA treatment thus leading to relapses. Much effort has been devoted to identify biological subtypes and critical regulators of RA-induced NB differentiation. RA-induced signalling pathway in NB initiates expression of many hundreds of genes including *MYCN* and other proto-oncogenes [21]. It has also been shown to alter drug resistance mechanisms [22]. Identification of new molecular drivers, as well as unbiased, more complete and clinically relevant mechanisms of RA-induced differentiation of NB is still a great challenge.

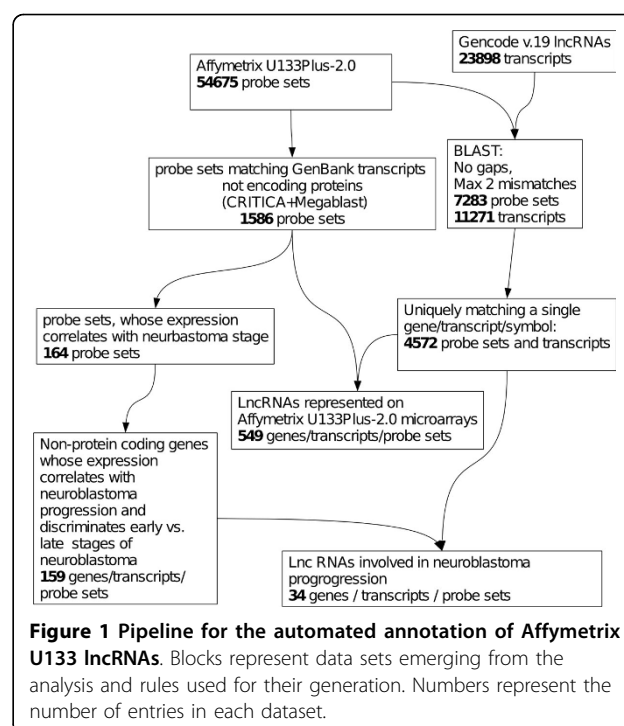
Gain of chromosome arm 17q is one of the most common genetic abnormalities encountered in NB, which is associated with poor prognosis [23]. An example of lncRNA biomarker for NB is ncRAN gene mapped to 17q region [24]. Its over-expression was shown to be associated with aggressiveness of primary NB [24] and confer a set of oncogenic properties, such as acceleration of cell growth and increasing invasiveness [24,25]. One of the classes of highly conserved lncRNAs, known as transcribed ultraconserved regions (T-UCR) has been shown to be implicated in cancer [26]. Indeed, differential expression of T-UCRs was shown to correlate with such oncogenomic parameters as *MYCN* amplification status and be a consequence of DNA copy number changes associated with NB [27]. Another study revealed association of different sets of T-UCRs with poor and long-term survival in the metastatic stage of NB patients [28]. In spite of this evidence for involvement of lncRNAs in NB, systematic and unbiased genome-wide analysis has not been performed.

In this study, we performed systematic genome-wide analysis to identify biomarkers for aggressive stages of NB. For these purposes we took advantage of the extensive coverage of a considerable part of noncoding transcriptome by Affymetrix U133 Plus 2.0 microarray probe sets. These probe sets have traditionally been neglected in most studies aimed to identify tumor biomarkers. Here, we developed a computational pipeline to annotate noncoding probe sets on Affymetrix U133 chips, representing ESTs, non-annotated RNAs, and unknown ncRNAs and identified lncRNA transcripts that differentiate favorable from aggressive stages of NB, as well as relapsing and non-relapsing primary tumors. We found that one of the identified in this study lncRNA transcript, termed HOXD-AS1, is induced by RA, regulated by PI3K/Akt pathway and controls genes implicated in RA-mediated cell differentiation and also angiogenesis and inflammation, the hallmarks of metastasized cancer [29].

Results

Affymetrix U133 series microarrays contain 1,586 probe sets measuring expression of potential lncRNAs, including 549 high-confidence lncRNA transcripts

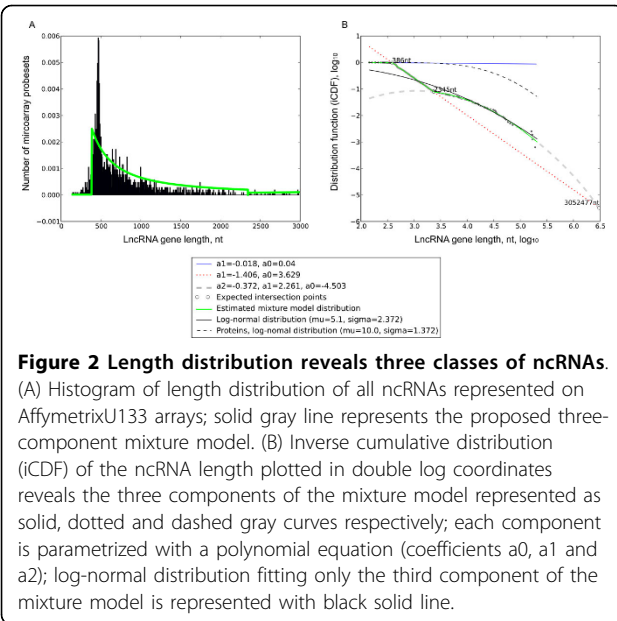
The probe sets of Affymetrix U133 chips were designed to detect not only transcripts of protein-coding genes, but also many transcripts currently annotated as ESTs and non-annotated RNAs(30), some of them can measure expression of unknown lncRNAs. Our goal was to find such probe sets. The 54,675 probe sets of Affymetrix microarrays U133A, U133B and U133Plus-2.0 were subjected to the filtration by quality and uniqueness of transcript matching according to the criteria of APMA database described by us earlier [30] (Figure 1). Following the database annotation, 7,587 probe sets matching unique genomic locations of RNAs not annotated as protein mRNAs were selected. Probe sets overlapping with exons of any known proteins (RefSeq) on the same DNA strand were excluded from the analysis. Thus, 1,586 probe sets were selected (Additional file 1), each of them matched a single transcript. To ensure that the selected transcripts do not encode proteins, we analyzed them with CRITICA software that identifies protein-coding sequences by potential presence of codons in the transcript sequences [31,32]. All 1,586 transcripts passed this filter and, therefore, they were considered as probe sets potentially measuring expression of lncRNA genes. To find which of these genes can be considered as true lncRNA candidates, their length distribution was analyzed. To ensure the reliability of the results, the sequences of individual Affymetrix probe sets were independently scanned against the 23,898 sequences of lncRNA transcripts annotated by Gencode consortium (v.19) using NCBI BLAST tool. Only 4,572 transcripts were found to have uniquely matching probe sets, 549



of which were present among our 1,586 ncRNA candidates. Since there was a limited overlap between the candidates and the Gencode-annotated lncRNAs, we considered the 549 lncRNAs as a high-confidence subset embedded in our data.

Length distribution of the ncRNA candidate transcripts reveals three lncRNA classes

Traditionally, ncRNAs longer than 200 nt are considered lncRNAs. According to this criterion, 1,581 selected ncRNAs were classified as lncRNAs. The distribution of the lncRNAs by length can be described as a mixture of three groups (Figure 2). Two distributions with very long and moderate long sequence lengths were fitted well with power-law like functions [33,34]. The length distribution of relatively shorter 3-rd group of lncRNAs resembles a log-normal distribution whose shape is more similar to the frequency distribution of protein-coding genes (Figure 2B, Additional file 2). The expected lncRNA lengths marking the three ranges correspond to each of the three distributions. They were found from the extrapolation of the components' estimates to their intersection points giving lengths of 386 nt and 2,345 nt. Thus, based on these cutoff estimates, we propose the following groups of lncRNA genes: 1) lncRNAs-1, length from 200 nt to 386 nt, 2) lncRNAs-2, length from 386 to 2,345 nt and 3) lncRNAs-3, length >2,345 nt. These classes comprise approximately 4%, 87% and 9% of 1,581 selected lncRNAs.



Genes encoding lncRNA-2 and lncRNA-3 classes are localized next to transcription factor genes implicated in development

Since the vast majority of the studied lncRNAs have never been functionally classified, we analyzed the functions of their neighboring genes (within ± 10 kb distance). Due to their proximity, they may be co-regulated with the lncRNA genes and thus may suggest biological processes associated with these genomic regions. For 89% (1400/1581) of lncRNAs such neighbors were found. For the classes lncRNA-1, lncRNA-2 and lncRNA-3 this fraction was 89% (62/70), 89% (1218/1372) and 93% (129/138), respectively (Additional files 1 and 3). Gene ontology (GO) analysis of all lncRNAs showed significant enrichment for 13 ontologies, including 'Wnt pathway', 'nucleic acids metabolic processes', 'mitochondria' and 'transcription factors' (Additional file 4). Out of four GO terms associated with lncRNA-1, three ('lysosomal transport', 'hydrolase activity of N-glycosyl compounds', 'phosphodiesterase') were common with all lncRNAs. In contrast, class lncRNA-2 was associated with 10 GO terms common with all lncRNAs. 'Homeobox transcription factors' and 'gut mesoderm development' categories were absent in lncRNA-2 class GO terms and were unique for lncRNA-3 class. Class lncRNA-3 was significantly associated with 15 GO terms, i.e. more than for all lncRNAs. GO terms 'reproduction', 'embryonic development' and 'KRAB box transcription factors' were associated only with lncRNA-3 class. It is remarkable that among the neighbors of lncRNA-3 class genes 'gut mesoderm development' category was represented by 11 genes from 3 *HOX* clusters: *HOXA*, *HOXB*, and *HOXD*. These transcription factors constitute 88% of all 18 lncRNAs annotated with this

GO term. Two other homeobox genes, not belonging to *HOX* were *EMX2* and *MNX1*. Fifteen other transcription factor genes belong to the zinc finger class and the remaining 8 genes to other classes.

159 lncRNAs are associated with NB progression

Data from NB patient tumor samples (GSE12460) were used to make a primary screening of lncRNAs associated with NB. Fourteen tumors in this dataset were MYCN-amplified (MNA), while the rest 50 tumors had no MYCN amplification (non-MNA). 1,581 lncRNAs represented on Affymetrix microarray U133Plus-2.0 were analyzed for their association with the stages of NB progression in the non-MNA group of patients (Figure 3). After comparing expression levels of these genes in early stage (1 and 2) vs. late stage (3 and 4) tumors 223 of them were identified as genes whose expression could discriminate ($P < 0.001$) between the stages. Among them 164 lncRNAs, whose expression levels significantly correlate (Kendall's $|\tau| > 0.25$) with the stage were found (Additional file 5A). Among them 79% (130/164) correlated positively. PAM

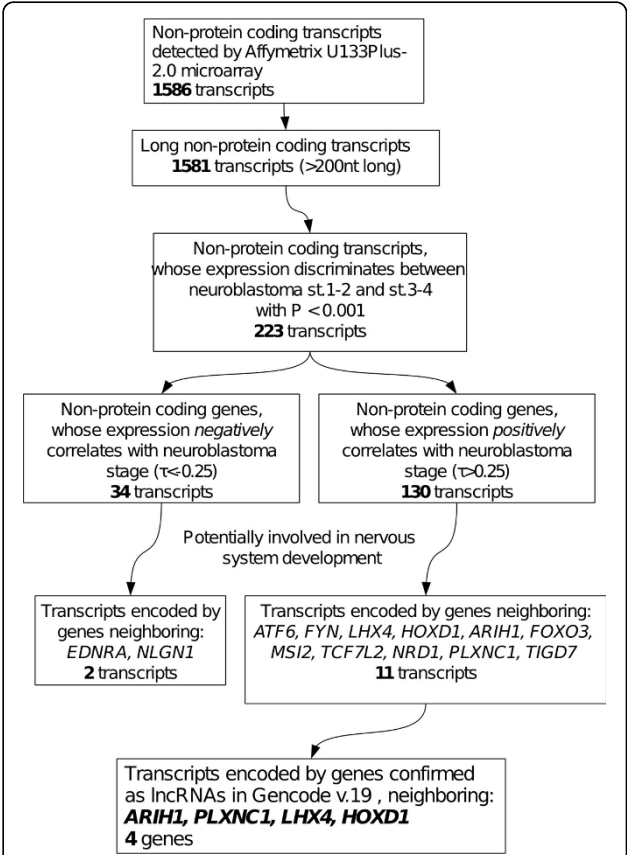


Figure 3 Pipeline for identification of lncRNA biomarkers that discriminate between stages 1-2 and 3-4. Blocks represent data sets emerging from the analysis and rules used for their generation. Numbers represent the number of entries in each dataset.

analysis revealed that 159 of the 164 lncRNAs form a gene signature able to discriminate between the low-aggressive (stages 1, 2) and high-aggressive (stages 3, 4) tumors with 79-88% accuracy (Additional files 5B and 5C). This signature also discriminates between relapsing and not relapsing primary tumors (GSE3446) with 74-80% accuracy (Additional files 5D and 5E). The lncRNAs of this tumor aggressiveness/relapse gene signature are grouped in two clusters by their correlation coefficients (Additional files 5F and 6). Analysis of GO of the genes neighboring the 126 lncRNAs of the main cluster demonstrated a significant enrichment in genes associated with nucleic acid binding and transcription (Additional file 5G).

Importantly, this list of lncRNA candidates included 13 genes (including four high-confidence lncRNAs) whose neighboring protein-coding genes were annotated as involved in embryonic development (Figure 3). Studying them could give a clue for the mechanisms of lncRNA involvement in neuroblastoma progression. Therefore, we tested the expression of each of the 13 lncRNAs in response to the treatment of neuroblastoma cells to RA. RA is a morphogen, which can induce neuronal differentiation in embryo and cell culture and is widely used for therapy of neuroblastoma.

HOXD-AS1 is up-regulated by retinoic acid in SH-SY5Y neuroblastoma cell line

In our experiments, we chose non-MNA neuroblastoma cell line SH-SY5Y, which is a widely used model for RA-induced neuronal differentiation. In our recent work, we performed analysis of differential expression of the 13 lncRNAs upon differentiation of SH-SY5Y cells using custom microarray chip [35]. Among all 13 tested lncRNAs only one lncRNA was upregulated during differentiation [35] was represented among the 164 lncRNA genes, correlating with NB stage, measured in patient tumors by a probe set 239182_at. Using qRT-PCR we measured the expression of these 13 lncRNAs at four time points (0 hr, 6 hr, 1 day and 5 days) post RA induction. The expression of only one transcript AL120749, encoded by gene HOXD-AS1, was progressively up-regulated by RA up to the highest level at the 5th day after RA-induction (Figure 4A). It is located between *HOXD1* and *HOXD3* genes in anti-sense direction to both of them. Noncoding status of HOXD-AS1 was predicted by both Megablast and Critica algorithms (31). Interestingly, expression level of probe set 239182_at correlates with the progression stage of the NB patients ($\tau = 0.32$, $P = 0.001$) (Figure 4B), including non-MNA ($\tau = 0.28$, $P = 0.01$) patients. It also discriminates the tumors at stages 1-2 from the tumors at stages 3-4 ($P < 0.001$), including non-MNA ($P = 0.009$) and MNA ($P = 0.002$) patients (Additional file 7).

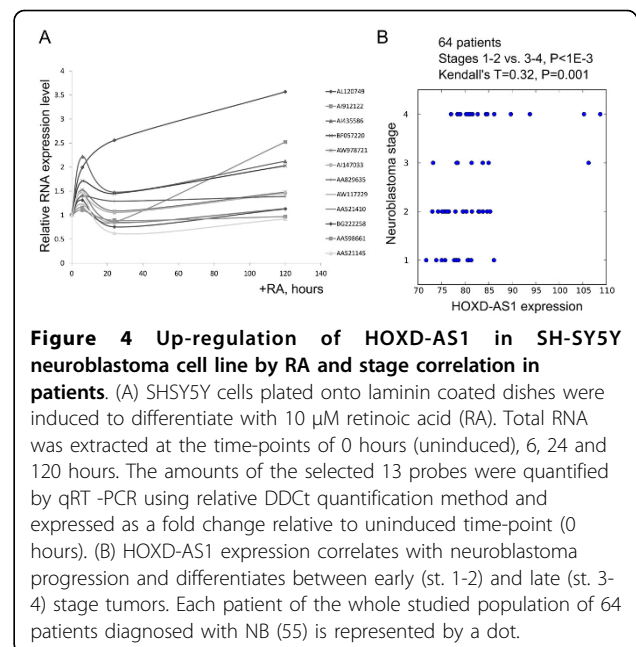
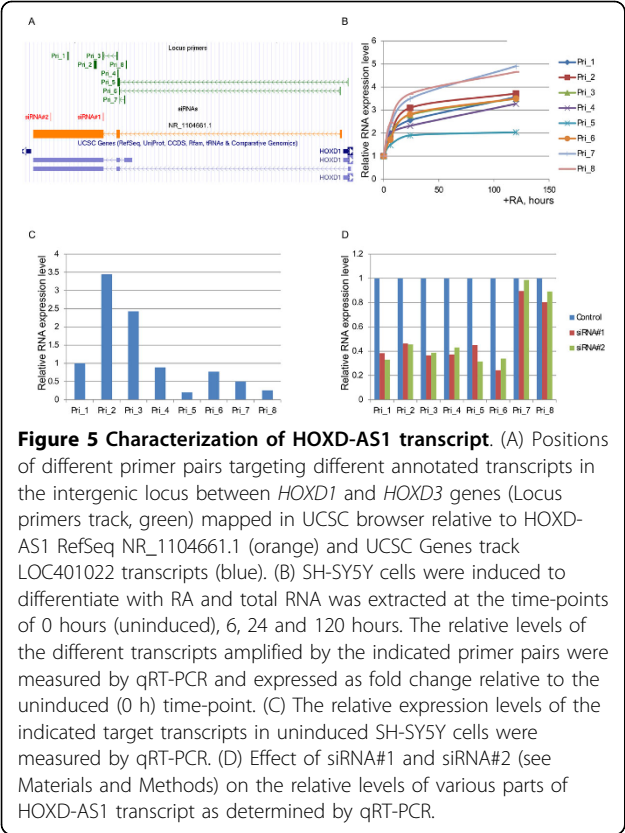


Figure 4 Up-regulation of HOXD-AS1 in SH-SY5Y neuroblastoma cell line by RA and stage correlation in patients. (A) SHSY5Y cells plated onto laminin coated dishes were induced to differentiate with 10 μ M retinoic acid (RA). Total RNA was extracted at the time-points of 0 hours (uninduced), 6, 24 and 120 hours. The amounts of the selected 13 probes were quantified by qRT-PCR using relative DDCT quantification method and expressed as a fold change relative to uninduced time-point (0 hours). (B) HOXD-AS1 expression correlates with neuroblastoma progression and differentiates between early (st. 1-2) and late (st. 3-4) stage tumors. Each patient of the whole studied population of 64 patients diagnosed with NB (55) is represented by a dot.

Characterization of HOXD-AS1 locus and its transcripts

Within the locus between *HOXD1* and *HOXD3* genes several isoforms of HOXD-AS1 have been annotated with two transcription start regions defined by active promoter features, such as CpG islands, H3K27 acetylation and transcription factor ChIP tracks (Additional file 8). The probe set 239182_at, which discriminates the patients, at its 5'-end fits into the boundary of RefSeq transcript NR_1104661.1, which starts at the promoter region 1. In order to define the boundaries of HOXD-AS1 transcript in SH-SY5Y cells, a set of primer pairs targeting different possible exons within this region was designed (Figure 5A, Additional file 8). Treatment of SH-SY5Y cells with RA resulted in induction of all regions under query with a 3-5-fold magnitude for most of them, apart from the primer pair 5 (Pri_5) querying the first two exons of UCSC LOC401022 isoform starting at the promoter region 1, which was induced only 2-fold (Figure 5B). The comparison of the relative levels of different amplicons revealed that the least abundant was UCSC LOC401022 isoform starting at the promoter region 1 (Pri_5), Pri1 - Pri4 querying regions common for all possible transcripts were of the highest abundance and Pri_6 querying exons 1-2 of NR_1104661.1 was of comparably high level (Figure 5C). The knock-down by two siRNAs used in this study as described below (their target positions shown in siRNAs track of Additional file 8) resulted in depletion of regions fitting to the RefSeq NR_1104661.1 (Pri_6) and the UCSC Genes LOC401022 isoform starting at promoter region 1 (Pri_5), but not the UCSC Genes LOC401022 isoform



starting at the promoter region 2 (Pri_7 and Pri_8, Figure 5D). In summary, RefSeq NR_1104661.1 transcript was the most abundant in SH-SY5Y cells and the most responsive to experimental conditions of RA-induced differentiation and siRNA mediated knock-down. Although the probe set 239182_at features the shortened form of the exon 3 as compared to HOXD-AS1 annotated transcripts, the pairs of primers spanning its entire length as in the UCSC geneset ("**Exon 3 primers**" track, Additional file 8) produce comparable relative qRT-PCR signals (Additional file 9 **panel A**) and are depleted by both siRNAs (Additional file 9, **panel B**), implying that the full length of this exon is present in the transcript under query. Indeed, the predicted full length size of NR_1104661.1 transcripts of 3,821 nucleotides is confirmed by Northern blot using the probe targeting exon 3 (Additional file 9 **panel C**).

Evolutionary analysis of HOXD-AS1 gene

A comparison of *HOXD-AS1* gene region against NCBI database of reference genomic sequences with BLAST showed that more than 96% of this gene sequence is evolutionary conserved among hominids (*H. sapiens*, *P. paniscus*, *P. troglodytes* and *P. abelii*). At the same time, only 12% to 52% of *HOXD-AS1* sequence is shared with other primates (*M. mulatta*, *N. leucogenys*, *C. jacchus* and

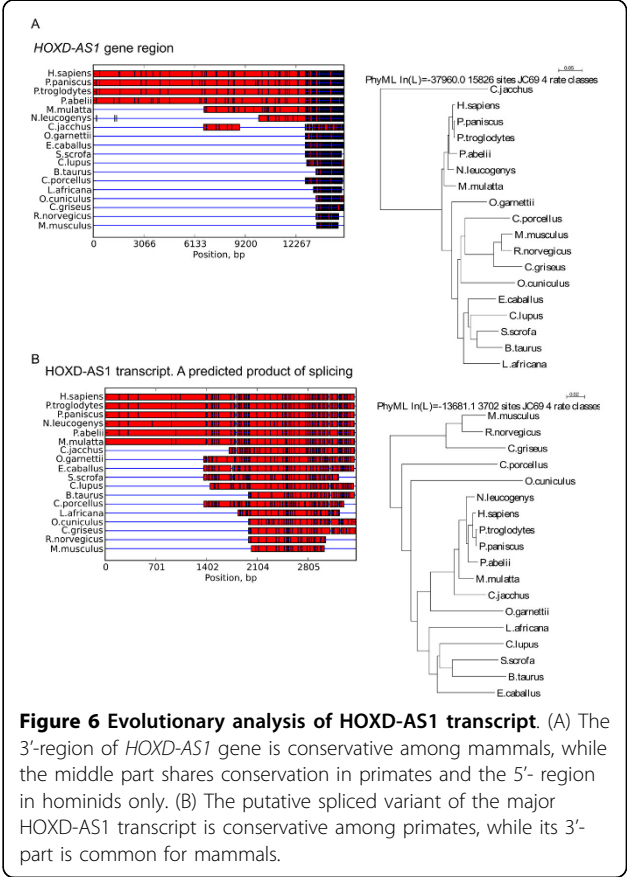
O. garnettii), comparable to 11% of dog (*C. lupus*) and 8-12% of ungulates (*E. caballus*, *S. scrofa* and *B. taurus*) (Figure 6). Thus, a gap between *HOXD-AS1* genomic sequence in apes and other monkeys is apparent.

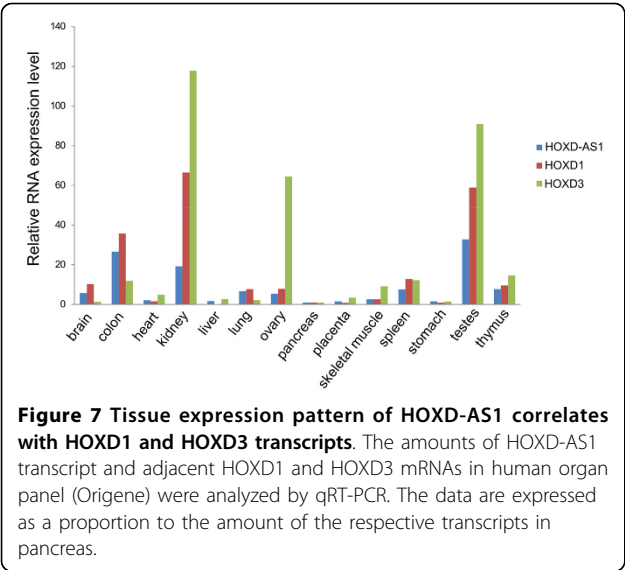
Tissue expression pattern of HOXD-AS1 correlates with HOXD1 and HOXD3 genes expression

We characterized the pattern of expression of HOXD-AS1 in human body by qRT-PCR analysis of human organ cDNA panel. We found that HOXD-AS1 is expressed in multiple tissues with the highest levels of expression in kidney, colon and testes and the lowest in liver, heart, pancreas and stomach (Figure 7). In general, HOXD-AS1 demonstrated highly correlated expression with its neighboring protein coding genes HOXD1 and HOXD3, thus implying that they are subject to the common regulatory mechanism. Notably, in the non-MNA patient tumors HOXD-AS1 expression significantly correlated with HOXD3 ($\tau = 0.25$, $P = 0.007$) (Additional file 10), but not HOXD1.

Differentially expressed genes identified after knock-down of HOXD-AS1 in SH-SY5Y cells

The differential expression of HOXD-AS1 in NB patients and in differentiating NB cell line is an

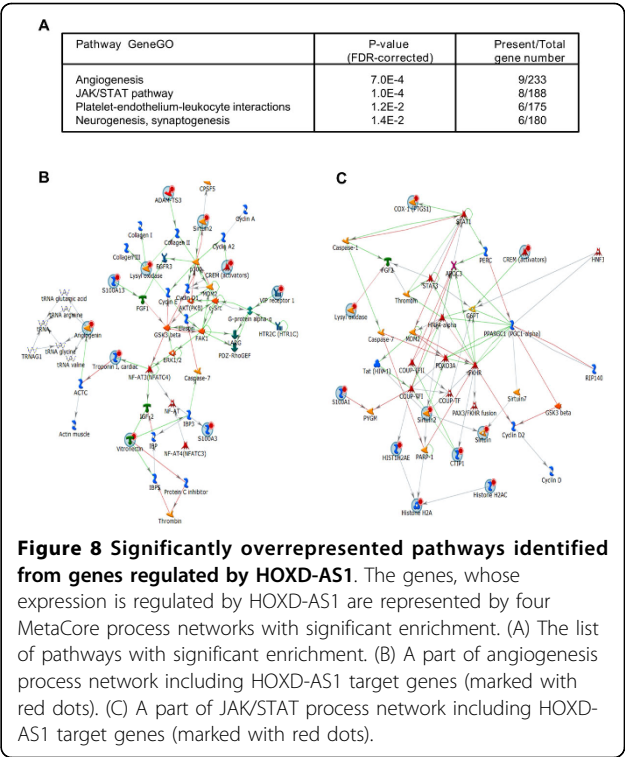




indicator of functionality of this lncRNA. In order to gain more insight into functionality of HOXD-AS1 we sought to investigate its effect on global gene expression profile in the context of differentiation of SH-SY5Y cells. For this purpose we designed two siRNAs targeting different regions of HOXD-AS1. Both of these siRNAs knocked down HOXD-AS1 transcript in the RA-induced SH-SY5Y cells with 70 - 80% efficiency as compared to non-targeting siRNA control (Additional file 11A). The total RNA was hybridized to IlluminaBeadArray (deposited in NCBI GSE40680) and differentially expressed genes common for two siRNAs were identified. In total, we identified 96 differentially regulated genes, among them 43 were up-regulated and 53 were down-regulated (Additional file 12). The pattern of differential expression of a set of genes, most of which were chosen for their involvement in the processes of inflammation and angiogenesis as discussed further, was confirmed by qRT-PCR (Additional file 11B). In order to annotate the functions of differentially expressed genes at the network level we mapped them on existing networks using MetaCore pathway analysis software. We observed significant functional enrichment in pathways regulating angiogenesis and inflammatory JAK/STAT pathway (Figure 8).

Expression of genes controlled by HOXD-AS1 significantly correlates with its expression in clinical neuroblastoma samples

In non-MNA tumors we analyzed pair-wise correlation between expression of HOXD-AS1 and of its targets revealed in the knock-down experiments. Among the genes regulated by HOXD-AS1 in the cell line, 53 were



represented on the Affymetrix chip by 98 probe sets. The expression of 21 of them significantly ($P < 0.05$) correlated with HOXD-AS1 expression in the NB samples (Additional file 12). Sixteen genes correlated positively and 5 negatively. Expression of 20 HOXD-AS1 targets was significantly different in primary tumors of 102 patients with relapse ($P < 0.05$), seven of which at the same time correlated with NB stage (Additional file 12). Importantly, *MAGEA9B* gene expression was also a significant ($P = 0.014$) marker of 5-year survival of neuroblastoma patients (GSE16237). Additionally, *MAGEA9B*, along with two other targets of HOXD-AS1 - *TNF* and *CNR1*, were recently found to be a part of neuroblastoma signature identified in several datasets [36]. Sixteen of HOXD-AS1 target genes were predictive markers of NB relapse (Additional file 13). Remarkably, 12 out of these 16 genes were previously reported to respond to RA treatment (Additional file 14A). Among these genes *SNN*, *TMEM86A*, *VIPR1*, *CREM* were down-regulated and *TSPAN2*, *CNR1*, *CREBL1*, *PTGS1* were up-regulated by RA. Gene signature consisting of four HOXD-AS1 targets, *ADAMTS3*, *AMDMD2*, *ANG*, *ASNA1* could predict NB relapse with 76-80% accuracy (Additional files 14B and 14C). Thus, HOXD-AS1, in response to RA treatment, controls expression of genes involved in NB progression. Some of them can serve as predictive markers of NB relapse and are associated with patient survival.

HOXD-AS1 is evenly distributed between nucleus and cytoplasm

So far most of the characterized lncRNAs that effect gene expression are known to be localized to the nucleus. Therefore, we analyzed the subcellular localization of HOXD-AS1. qRT-PCR analysis of subcellular fractions of SH-SY5Y cells revealed that HOXD-AS1 is equally distributed between the nucleus and the cytoplasm (Additional file 15A). This pattern was largely confirmed by RNA fluorescence *in situ* hybridization. HOXD-AS1-specific probe sets detected individual transcripts both in the nucleus and the cytoplasm (Additional file 15B).

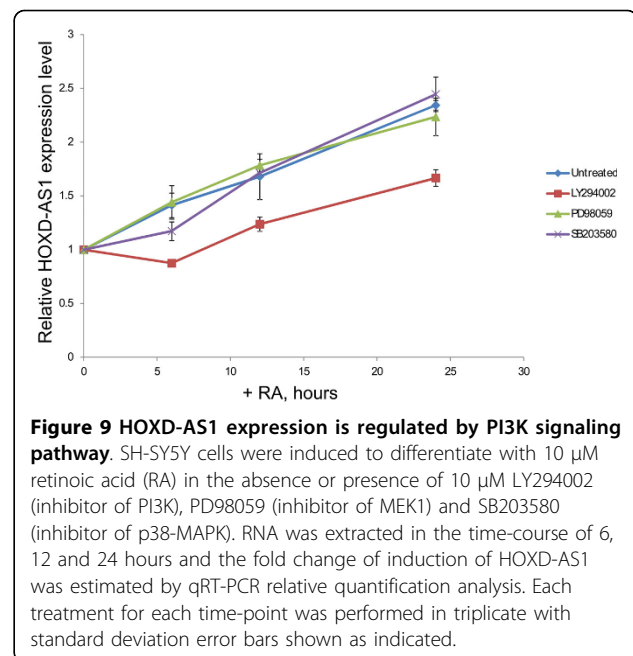
HOXD-AS1 expression is regulated by PI3K/Akt signaling pathway

Induction of differentiation of NB cells by RA is a complex process, requiring involvement of at least two signaling pathways: PI3K/Akt and MAPK/ERK [37]. To characterize regulation of expression of HOXD-AS1 by these signaling pathways we used chemical inhibitors of PI3K (LY294002), MEK1 (PD98059) and p38-MAPK (SB203580). SH-SY5Y cells were induced to differentiate in the presence of these inhibitors in the time course of 6, 12 and 24 hours and change of HOXD-AS1 expression was measured by qRT-PCR. The induction of HOXD-AS1 was significantly reduced in the presence of LY294002, however it was not affected by either PD98059 or SB203580, thus indicating that HOXD-AS1 is regulated by PI3K/Akt pathway, but not by MAPK/ERK (Figure 9).

Discussion

Evidence for involvement of different lncRNAs in oncogenesis and for their potential use as biomarkers of cancer is growing [2,38]. Affymetrix U133 series are among the most widely used microarray platforms with thousands of datasets deposited in the public databases. Given the fact that many lncRNAs were fortuitously represented on this platform these datasets provide a great potential for their reanalysis, with lncRNAs taken into account. Here we annotated the noncoding fraction of Affymetrix probe sets and created a resource for this kind of studies.

Discrimination of coding and non-coding transcripts is not straightforward, and different laboratories are using different methods. Therefore, in our work we used two different methods that discard the possibility of any transcript encoding a protein: Megablast, to eliminate the possible known protein homologs; followed by CRITICA to eliminate the possibility that any of our transcripts may encode a protein, yet to be discovered [31,32]. All 1,586 transcripts studied in our work satisfied both criteria. In addition to that, we applied the criterion of transcript length distribution among proteins



to judge about our dataset as a whole. To maximize the reliability of prediction, we performed an additional analysis and mapped the probe sequences to the lncRNA transcript sequences predicted by Gencode consortium. It allowed us to identify 549 reliable probesets corresponding to lncRNAs out of 1,586 initially predicted. We provided classification of these 1,586 lncRNAs by the length of their transcripts and discriminating three classes among the lncRNA genes longer than 200 nt. Our classification system is vindicated by the fact that the proposed lncRNA classes have distinct gene ontologies.

However, lncRNA prediction strategy utilized in this study does not take into account relative position of candidate lncRNA gene in relation to protein coding gene. It has been a prevalent belief for some time that intronic parts of mRNAs cannot be considered as lncRNAs because they merely represent pre-mRNAs on their way to splicing or spliced-out introns on their way to degradation. The lower expression of the intronic regions than those in exons seemingly was consistent with this notion. However, several recent studies have focused on this important issue and concluded that intronic lncRNAs have a number of features that would be expected from functional, stand-alone RNA species. For example, St Laurent *et al.* have shown that there are thousands of intronic RNAs in the mouse genome changing throughout the inflammation time course that are likely to be functional [39]. The levels of intronic RNAs were independent of those of exons or other introns and some of the introns encoded relatively abundant RNA species. Another study by Tahira *et al.* described

intronic lncRNAs differentially expressed in primary and metastatic pancreatic cancer [40]. The authors suggested that intronic lncRNAs are generated from independent transcriptional units. In support of this notion they have shown enrichment of H3K4me3 marks close to the known boundaries for the set of intronic transcripts, the presence of conserved DNA elements and stable predicted secondary structures in intronic lncRNAs, evidence for the evolutionary selection, and significant enrichment of CAGE tags proximal to known start sites of intronic lncRNAs [40].

Applying the new probe set and ncRNAs annotation system for biomarker discovery in public datasets, we identified a gene expression signature consisting of 159 lncRNAs, which can discriminate NB tumors by their aggressiveness and clinical stages. To our knowledge, this is the first example of a gene signature based on expression of lncRNAs only. Functional studies of lncRNAs are not common and the biological effects of this class of molecules have not been studied well. The primary structures of lncRNAs are much less conservative in comparison with protein structures. Therefore, functional characterization of human lncRNAs is particularly difficult. Characterization of lncRNAs via gene signatures specific to certain diseases reported in the present study can be used as a tool to solve this methodological conundrum. It is important that our syndrome-centered scan for potential candidates was successfully validated and complemented with mechanistic studies on a model cell line. Regarding the clinical potential of gene expression signatures based only on lncRNAs, it is important to note that such signatures can be particularly useful as diagnostic and prognostic tools in combinations with existing protein-based signatures, since lncRNAs and proteins belong to different regulatory realms and thus naturally increase each other's power as patient classifiers.

Among the 159 lncRNAs we selected *HOXD-AS1* for functional characterization as the most potent candidate for playing a role in oncogenesis. The gene encoding *HOXD-AS1* is located in *HOXD* cluster. *HOX* genes are the key developmental regulators and their aberrant expression is often associated with malignancy [41]. We obtained evidence of a two-step evolution of the unspliced *HOXD-AS1* sequence. The first step, characterized with initial two-fold increase of *HOXD-AS1* length, presumably occurred around the time of primates branching from other placental animals between 115 [42] and 80 million years ago [43]. The second step, when the maximal-length *HOXD-AS1* gene in hominids (*H. sapiens*, *P. paniscus*, *P. troglodytes* and *P. abelii*) was established, might have occurred 15-19 million years ago, when hominid diverged from hylobatids (*N. leucogenys*) [44].

Remarkably, *HOXD3* gene, adjacent to *HOXD-AS1* locus, has established role in promoting metastatic potential of lung cancer cells [45]. *HOXD1*, which is also adjacent to *HOXD-AS1*, has been linked with locus rs2072590 associated with an increased risk of ovarian cancer, more significantly for the serous subtype [46]. This locus lies in a non-coding region, ~5 kb distal to *HOXD3* and ~10 kb proximal to *HOXD1* [46]. Using higher resolution map, we found that SNP rs2072590 is localized in *HOXD-AS1* and includes the last intron and last exon regions boundary.

Interestingly, differential expression of *HOXD-AS1* lncRNA was detected by microarray analysis in a number of cancer studies. It was among the 17 genes up-regulated in EGFR-mutated adenocarcinomas, which are characterized by poor post-operative prognosis [47]. In breast cancer, over-expression of breast cancer metastasis suppressor gene (BRMS) leads to down-regulation of *HOXD-AS1* [48], thus further pointing to its potential pro-oncogenic role. In contrast, *HOXD-AS1* was shown to be one of the most drastically down-regulated genes in hepatocellular carcinoma [49]. We also found that *HOXD-AS1* is strongly suppressed in colorectal adenoma discriminating these precancerous lesions from the surround normal mycosa tissue [50]. Here we show that *HOXD-AS1* is highly induced by RA treatment of NB SH-SY5Y cell line. RA causes growth arrest and differentiation of NB cells, therefore it is widely used for therapeutic purpose. The elevation of *HOXD-AS1* levels in metastatic tumors is in a seeming contradiction with its up-regulation during tumor cell differentiation. One possible explanation is that *HOXD-AS1* elevation in metastatic tumors is a compensatory reaction to a loss of its differentiation-promoting function that actually leads to aberrant tumor-promoting effect. This could happen because of the overall loss of homeostasis when inactivation of one pathway leads to a compensatory activation of another pathway.

In this study, we investigate regulation of *HOXD-AS1* expression within a framework of PI3K and MAPK signaling pathways, both of which are necessary for RA mediated differentiation and survival of SH-SY5Y cells [37]. Here we demonstrate that chemical inhibitor of PI3K, but not inhibitors of components of MAPK pathway MEK1 and p38 significantly reduce RA-mediated induction of *HOXD-AS1* (Figure 9). Both PI3K and MAPK pathways are induced in SH-SY5Y cells by ligation of a family of growth factors, known as neurotrophins to their respective receptors belonging to a class of receptor tyrosine kinases. Different types of neurotrophin receptors are known to control differentiation and survival of neuroblastoma cells include TrkA, TrkB, TrkC [51]. Whereas TrkA and TrkC are not regulated, or regulated marginally by RA in SH-SY5Y, TrkB is

efficiently induced by it (Additional file 16 **panel A**). Interestingly, activation of PI3K pathway by ligation of TrkB with its specific ligand BDNF leads to both neuritogenesis and survival. On the other hand, activation of MAPK pathway contributes only to neuritogenesis, but not survival [37]. Indeed, PI3K pathway is known to promote oncogenesis via inhibition of apoptosis as one of the mechanisms [52]. Both activation of PI3K pathway and over-expression of TrkB and BDNF are associated with poor prognosis in NB [13,53,54]. In contrast, patients with increased levels of TrkA and TrkC demonstrate better prognosis [55,56]. Thus our results imply that *HOXD-AS1* may be regulated via PI3K pathway induced by oncogenic BDNF/TrkB axis. This is further supported by our observation that *HOXD-AS1* transcript positively correlates with TrkB and negatively correlates with TrkA and TrkC (Additional file 16 **panel B**).

In agreement with the proposed pro-oncogenic role of *HOXD-AS1*, we observed that its siRNA-mediated knock-down resulted in differential expression of genes implicated in angiogenesis and inflammation, two highly interdependent processes considered as the hallmarks of cancer [29]. Both of these processes are manifested in a crosstalk between tumor cells and their microenvironment that includes endothelial and immune inflammatory cells. In agreement with this, a number of *HOXD-AS1* target genes encode secreted proteins that provide communication between tumor and its microenvironment. Among them are cytokines, such as CX3CL1, CCL20, TNF, GDF15 - key regulators of angiogenesis and lymphangiogenesis ANG and PROX1 - genes involved in remodeling of extracellular matrix LOX and ADAMTS3.

Conclusions

In conclusion, the computational pipeline developed in this study was successfully applied to greatly extend the number of novel NB cancer biomarkers using published Affymetrix U133 chip datasets and to identify candidate lncRNAs functionally implicated in classification of the NB subtypes. We found that one of these lncRNAs, named *HOXD-AS1*, is induced by RA, could be regulated via PI3K/Akt pathway, controls genes involved in RA signaling, angiogenesis and inflammation.

Methods

Cell culture

SH-SY5Y cell line was purchased from American Type Culture Collection (ATCC). SH-SY5Y cells were cultured in Hams F12/ DMEM supplemented with 10% FBS in a humidified incubator at 37°C with 5% CO₂. For cell differentiation experiment SH-SY5Y cells were plated onto laminin coated dishes and on the next day

were induced to differentiate by addition of 10 μM all-trans-retinoic acid (RA, Sigma). For pathway inhibition study SH-SY5Y were induced to differentiate by 10 μM in the absence or presence of 10 μM of LY294002, PD98059 and SB203580 (Sigma).

RNA isolation

Total RNA was purified using Trizol. For microarray experiment RNA was additionally purified using RNeasy Mini Kit (Qiagen) according to manufacturer's instructions.

Northern blot

The probe to detect *HOXD-AS1* was designed to be of 373 nucleotides, mapping to coordinates chr2:177,040,872-177,041,244 in GRCh37/hg19 Human Assembly of UCSC browser. The probe was synthesized (using cDNA derived from SH-SY5Y total RNA as a template) and Northern blot procedures were performed using DIG Northern Starter Kit according to manufacturer's instructions.

Quantitative PCR (qRT-PCR)

Total RNA was used as a template for reverse transcription using QuantiTect Reverse Transcription Kit (Qiagen) using random hexamer primers. The transcript levels were analyzed by qRT-PCR run on Rotor-Gene Q machine using Rotor-Gene SYBR Green PCR Kit (Qiagen). The primers used throughout this study are listed in Additional file 17.

Knock-down experiment using small interfering RNA (siRNA)

siRNA duplexes and control ON-TARGETplus non-targeting pool (Dharmacon) were introduced into SH-SY5Y cells using AmaxaNucleofector System (Lonza AG) according to manufacturer's instructions. Cells were plated onto laminin coated dishes and 24 hours later were induced to differentiate with 10 μM RA. 48 hours later cells were harvested for RNA analysis. The following siRNA target sequences (Dharmacon) were used to knock-down *HOXD-AS1*: GAAAGAAGGACCAAAGTAA, GCACAAAGGAACAAGGAAA.

Microarray analysis

Total RNA was amplified using IlluminaTotalPrep RNA Amplification Kit (Ambion) and amplified cRNA was hybridized to HumanHT-12 v4 BeadChip according to manufacturer's instructions. Briefly, 500 ng of total RNA was converted to double stranded cDNA using T7-oligo(dT) primers followed by IVT reaction to produce biotinylatedcRNA. The chips were scanned on IlluminaBeadArray Reader. The microarray data were deposited to NCBI GSE40680.

Analysis of microarray data from knock-down experiments

Expression values from four technical replicates were analyzed for each of the three biological samples (two siRNA knock-down experiments and one non-matching siRNA control). After RMA normalization, microarray signal distributions were examined. Multiple statistical criteria were used to obtain robust results of differential expression analysis. A differentially expressed probe simultaneously satisfied each of the following criteria: 1) Student's t-test $P(\text{sample vs. control}) < 0.05$, 2) fold change ratio of median sample value to median control value > 1.3 or > 1.4 , 3) the number of sample replicas which signal value is higher than median control value is not less than 3. Based on these criteria 98 differentially expressed probes were selected.

Computational pipeline to define and classify probe sets for lncRNAs on Affymetrix U133- Plus-2.0

Noncoding status of the probe sets was confirmed by assuring that their sequences share no homology with existing protein-coding genes with NCBI BLAST [32]. To further confirm that the selected transcripts were not protein-coding by the base composition statistical criteria, CRITICA software was used as an additional filter [31]. The overlap length distribution of the probe sets and the noncoding transcripts was calculated and parametrized with a 3-component mixture model, a sum of 3 extended power-law distribution functions by least squares method [57]. To generate a high-confidence subset of lncRNA genes used in the study, the all Affymetrix U133-Plus-2.0 probes were matched against Gencode v.19 database of lncRNA transcripts using NCBI BLAST tool with the following options: -F -F -g -F -a 4 -K 1 -e 0.001.

Computational analysis of clinical microarray data

For identification of lncRNAs differentially expressed in neuroblastoma, GSE12460 [58] AffymetrixU133-Plus-2.0 data from GEO database were used. This dataset contains 64 neuroblastoma (stages 1 to 4) tumor samples, in the genomes of 14 of which MYCN gene was amplified. For clinical analysis, lncRNA longer than 200 bp were selected. Mann-Whitney U test with null-hypothesis that the expression of a given lncRNA in the samples of neuroblastoma patients diagnosed at stages I or II is distributed equivalently to the patients diagnosed at stages III or IV and Student's t-test with null hypothesis that the population means are equal were applied with FDR control [59]. For every lncRNA Kendall's tau correlation was calculated across the patients sorted by their diagnosed neuroblastoma stages (I to IV). To study relapsing neuroblastoma tumors 102 patients of GSE3446 dataset [13] was used. To study the effect of

RNA gene expression on 5-years patient survival 44 patients of GSE16237 dataset [60] was used. Mann-Whitney U test for assessing gene differential expression and PAM analysis [61] for gene signature identification were used for these two datasets.

GO analysis

RefSeq genes located in the vicinity of ± 10 kb from each studied lncRNAs were defined as neighboring. For each neighboring gene, the best matching Affymetrix probe set was found, according to APMA annotation (30). For each such probe set, the best matching mRNA was identified. For all such mRNAs found for all neighboring genes Entrez IDs were found according to UniProt annotation. For each Entrez ID, the corresponding GO were found from Panther database (<http://www.pantherdb.org/>). Enrichment was considered significant for ontologies with $P < 0.01$ and enrichment value not less than 2.

Fluorescence in situ hybridization

RNA visualization was performed using QuantiGeneViewRNA technology (Panomics). RNA staining was performed according to manufacturer's protocol. The cells were counterstained with DAPI to label nuclei and images were acquired on Zeiss LSM 5 DUO confocal microscope.

Additional material

Additional file 1: Classification of AffymetrixU133 probe sets representing ncRNAs.

Additional file 2: Length distribution of protein coding RNAs overlapping with Affymetrix U133 probe sets. A) Histogram of length distribution of all protein coding RNAs represented on AffymetrixU133 arrays; green line represents the mixed model based on Gaussian components, red line represents the tail-specific component. B) Inverse cumulative distribution (iCDF) of the ncRNA length plotted in double log coordinates reveals that log-normal distribution (black solid curve) can describe it well and three-component (blue, green and cyan dashed curves respectively) mixture model (yellow solid curve) is unnecessary; each component of the mixture model is parametrized with a polynomial equation (coefficients a_0 , a_1 and a_2).

Additional file 3: Histogram of the ncRNAs (represented by Affymetrix U133 probe sets) length distribution by chromosome. Blue and yellow parts of the bar respectively represent the number of ncRNAs, which genes are with and without neighbouring genes within 10 kb.

Additional file 4: List of ncRNA probe sets neighboring coding genes with their ontologies.

Additional file 5: lncRNAs with expression levels correlating with the disease stage. A. List of 164 lncRNAprobe sets associated with neuroblastoma progression. B. ncRNA gene expression signature for discriminating aggressive (stage III-IV) from not aggressive (stage I-II) neuroblastoma. C. Predictive power of the 159 ncRNA gene expression signature for classification of neuroblastoma tumors into aggressive (stages III-IV) and low-aggressive (stage I-II) cases. D. ncRNA gene expression signature for discriminating primary neuroblastoma tumors of patients with relapse vs. tumors of patients without relapse. E. Predictive power of the 159 ncRNA gene expression signature for classification of

neuroblastoma tumors into aggressive (stages III-IV) and low-aggressive (stage I-II) cases. F. The 159 ncRNAs of the PAM signature for separating high- and low- aggressive forms of neuroblastoma form 2 major clusters. G. 159 lncRNAs are associated with neuroblastoma progression.

Additional file 6: Cluster analysis of Kendall's correlation matrix of 159 ncRNAs. Cluster analysis revealed two major groups of correlating ncRNAs (left panel, red and green) characterized with specific GOs.

Additional file 7: Correlation of HOXD-AS1 expression with neuroblastoma progression. Expression of HOXD-AS1 correlates with neuroblastoma progression and differentiates between early (st. 1-2) and late (st. 3-4) stage tumor belonging to A) non-MNA and B) MNA groups.

Additional file 8: UCSC browser representation of the locus between HoxD1 and HoxD3 genes. Custom tracks depict positions of siRNAs used to knock down HOXD-AS1, pairs of primers targeting different parts of annotated transcripts within this locus (Locus primers and Exon 3 primers), HOXD-AS1 model, the tracks of transcript predictions (UCSC Genes, and Affymetrix U133 Plus probe sets). Two transcription start region defined by CpG Islands, ENCODE H3K27Ac mark and Transcription Factor ChIP-Seq tracks are shown in rectangles.

Additional file 9: The 3' exon of HOXD-AS1 is full length. A set of primers targeting different regions of the entire length of HOXD-AS1 exon 3 was used to measure A) relative expression levels of these regions B) the knock-down efficiency of these regions by 2 indicated siRNAs (Additional file 14). C) Northern blot confirming the size of HOXD-AS1.

Additional file 10: Correlation of HOXD-AS1 and HOXD3 expression in patient tumors. Expression of HOXD-AS1 transcript positively correlates (Kendall's tau) with expression of HoxD3 measured by probe sets 206601_at (A) and 206602_at (B) in non-MNA patient tumors.

Additional file 11: Identification of differentially expressed genes after knock-down of HOXD-AS1 in SH-SY5Y cells. (A) SH-SY5Y cells grown on laminin coated dishes were transfected with non-targeting siRNA pool and two siRNAs targeting different regions of HOXD-AS1 transcript. On the next day transfected cells were induced to differentiate by addition of 10 μ M RA. 48 hours later RNA was extracted and the efficiency of knock-down was estimated by qRT-PCR and expressed as a fold change relative to non-targeting siRNA pool (control). (B) SH-SY5Y cells were transfected with non-targeting siRNA pool and two siRNAs targeting different regions of HOXD-AS1 transcript. On the next day transfected cells were induced to differentiate by addition of 10 μ M RA. 48 hours later RNA was extracted and the relative levels of the selected transcripts were estimated by qRT-PCR.

Additional file 12: List of 96 genes differentially regulated by HOXD-AS1 knock-down.

Additional file 13: Genes regulated by HOXD-AS1 on expression level differ significantly in predicting relapse of neuroblastoma. Expression distribution comparisons.

Additional file 14: HOXD-AS1 target genes involved in NB progression. A. 12 of the 16 relapse-significant gene targets of HOXD-AS1 belong to RA pathway. B. HOXD-AS1 target genes, whose expression forms a signature discriminating between relapsing and not relapsing neuroblastoma tumors C. Predictive power of the 98 HOXD-AS1 target genes expression signature for classification of neuroblastoma tumors into relapsing and not relapsing cases.

Additional file 15: HOXD-AS1 is evenly distributed between nucleus and cytoplasm. Subcellular localization of HOXD-AS1 lncRNA was analyzed by two methods. (A) SH-SY5Y cells were fractionated into cytoplasm and nucleus by centrifugation through sucrose pad (see Materials and Methods). RNA was extracted from nuclear and cytoplasmic fractions and the equal volumes of nuclear and cytoplasmic RNA, corresponding to equivalent numbers of cells were used for the first strand cDNA synthesis. The abundance of HOXD-AS1 transcript in nuclear fraction relative to cytoplasmic was analysed by qRT-PCR. The *bona fide* nuclear transcript MALAT1 and mitochondrial 16S were used to assess the purity of fractionation. (B) Localization of HOXD-AS1 inside SH-SY5Y was directly visualized by hybridization of the specific PanomicsQuantiGeneViewRNA probes followed by confocal microscopy

(see Materials and Methods). HOXD-AS1 transcript was visualized at red channel, cytoplasmic GAPDH transcript was visualized at green channel and nuclei were counter-stained with DAPI (blue).

Additional file 16: Correlation of HOXD-AS1 with TrkA, TrkB and TrkC receptors expression. (A) qRT-PCR demonstrating drastic up-regulation of TrkB, but not TrkA and TrkC in the time-course of differentiation of SH-SY5Y cells. (B) HOXD-AS1 positively correlates with TrkB and negatively correlates with TrkA and TrkC.

Additional file 17: Primer list.

Conflict of interest

The authors declare that they have no competing interests.

Authors' contributions

AAY, AOB, VAK, IVK planned the project. AAY, VAK, IVK designed the experiments. AAY, JZT, GMS, PS, IVK performed experiments. AOB, VAK performed computational analysis. AAY, AOB, VAK, IVK wrote the paper. All authors discussed the data and commented on the manuscript.

Declarations

Publication of this article was funded by Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore. This article has been published as part of *BMC Genomics* Volume 15 Supplement 9, 2014: Thirteenth International Conference on Bioinformatics (InCoB2014): Computational Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/15/S9>.

Authors' details

¹Department of Genome and Gene Expression Data Analysis, Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Matrix, 138671 Singapore. ²Translational Control in Development and Disease Group, Institute of Medical Biology, Agency for Science, Technology and Research (A*STAR), Immunos, 138648 Singapore. ³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597. ⁴Program in Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, 8 College Road, 169857 Singapore. ⁵Division of Software & Information Systems, School of Computer Engineering, Nanyang Technological University, 639798 Singapore.

Published: 8 December 2014

References

1. Palanichamy JK, Rao DS: miRNA dysregulation in cancer: towards a mechanistic understanding. *Front Genet* 2014, **5**:54.
2. Qiu MT, Hu JW, Yin R, Xu L: Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol* 2013, **34**:613-620.
3. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010, **464**:1071-1076.
4. Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Müller-Tidow C: MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003, **22**:8031-8041.
5. Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, Ijiri K, Akimitsu N: MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett* 2010, **584**:4575-4580.
6. Janoueix-Lerosey I, Schleiermacher G, Delattre O: Molecular pathogenesis of peripheral neuroblastic tumors. *Oncogene* 2010, **29**:1566-1579.
7. Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, Castelberry RP, De Bernardi B, Evans AE, Favrot M, Hedborg F, et al: Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *J Clin Oncol* 1993, **11**:1466-1477.

8. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM: **Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage.** *Science* 1984, **224**:1121-1124.
9. Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, Hammond D: **Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas.** *N Engl J Med* 1985, **313**:1111-1116.
10. Berwanger B, Hartmann O, Bergmann E, Bernard S, Nielsen D, Krause M, Kartal A, Flynn D, Wiedemeyer R, Schwab M, Schäfer H, Christiansen H, Eilers M: **Loss of a FYN-regulated differentiation and growth arrest pathway in advanced stage neuroblastoma.** *Cancer Cell* 2002, **2**:377-386.
11. Ohira M, Oba S, Nakamura Y, Isogai E, Kaneko S, Nakagawa A, Hirata T, Kubo H, Goto T, Yamada S, Yoshida Y, Fuchioka M, Ishii S, Nakagawara A: **Expression profiling using a tumor-specific cDNA microarray predicts the prognosis of intermediate risk neuroblastomas.** *Cancer Cell* 2005, **7**:337-350.
12. Schramm A, Schulte JH, Klein-Hitpass L, Havers W, Sieverts H, Berwanger B, Christiansen H, Warnat P, Brors B, Eils J, Eils R, Eggert A: **Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling.** *Oncogene* 2005, **24**:7902-7912.
13. Asgharzadeh S, Pique-Regi R, Sposto R, Wang H, Yang Y, Shimada H, Matthay K, Buckley J, Ortega A, Seeger RC: **Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification.** *J Natl Cancer Inst* 2006, **98**:1193-1203.
14. Oberthuer A, Berthold F, Warnat P, Hero B, Kahlert Y, Spitz R, Ernestus K, König R, Haas S, Eils R, Schwab M, Brors B, Westermann F, Fischer M: **Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification.** *J Clin Oncol* 2006, **24**:5070-5078.
15. Krasnoselsky AL, Whiteford CC, Wei JS, Bilke S, Westermann F, Chen QR, Khan J: **Altered expression of cell cycle genes distinguishes aggressive neuroblastoma.** *Oncogene* 2005, **24**:1533-1541.
16. Fredlund E, Ringnér M, Maris JM, Pålman S: **High Myc pathway activity and low stage of neuronal differentiation associate with poor outcome in neuroblastoma.** *Proc Natl Acad Sci USA* 2008, **105**:14094-14099.
17. Liu X, Mazanek P, Dam V, Wang Q, Zhao H, Guo R, Jagannathan J, Cnaan A, Maris JM, Hogarty MD: **Deregulated Wnt/beta-catenin program in high-risk neuroblastomas without MYCN amplification.** *Oncogene* 2008, **27**:1478-1488.
18. Westermann F, Muth D, Benner A, Bauer T, Henrich KO, Oberthuer A, Brors B, Beissbarth T, Vandesompele J, Pattyn F, Hero B, König R, Fischer M, Schwab M: **Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous regression or malignant progression in neuroblastomas.** *Genome Biol* 2008, **9**:R150.
19. Bourdeaut F, Janoueix-Lerosey I, Lucchesi C, Paris R, Ribeiro A, de Pontual L, Amiel J, Lyonnet S, Pierron G, Michon J, Peuchmaur M, Delattre O: **Cholinergic switch associated with morphological differentiation in neuroblastoma.** *J Pathol* 2009, **219**:463-472.
20. Reynolds CP, Matthay KK, Villablanca JG, Maurer BJ: **Retinoid therapy of high-risk neuroblastoma.** *Cancer Lett* 2003, **197**:185-192.
21. Thiele CJ, Deutsch LA, Israel MA: **The expression of multiple proto-oncogenes is differentially regulated during retinoic acid induced maturation of human neuroblastoma cell lines.** *Oncogene* 1988, **3**:281-288.
22. Bates SE, Mickley LA, Chen YN, Richert N, Rudick J, Biedler JL, Fojo AT: **Expression of a drug resistance gene in human neuroblastoma cell lines: modulation by retinoic acid-induced differentiation.** *Mol Cell Biol* 1989, **9**:4337-4344.
23. Bown N, Cotterill S, Lastowska M, O'Neill S, Pearson AD, Plantaz D, Meddeb M, Danglot G, Brinkschmidt C, Christiansen H, Laureys G, Speleman F, Nicholson J, Bernheim A, Betts DR, Vandesompele J, Van Roy N: **Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma.** *N Engl J Med* 1999, **340**:1954-1961.
24. Yu M, Ohira M, Li Y, Niizuma H, Oo ML, Zhu Y, Ozaki T, Isogai E, Nakamura Y, Koda T, Oba S, Yu B, Nakagawara A: **High expression of ncRAN, a novel non-coding RNA mapped to chromosome 17q25.1, is associated with poor prognosis in neuroblastoma.** *Int J Oncol* 2009, **34**:931-938.
25. Zhu Y, Yu M, Li Z, Kong C, Bi J, Li J, Gao Z, Li Z: **ncRAN, a newly identified long noncoding RNA, enhances human bladder tumor growth, invasion, and survival.** *Urology* 2011, **77**:510.e1-e5.
26. Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu X, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM: **Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas.** *Cancer Cell* 2007, **12**:215-229.
27. Mestdagh P, Fredlund E, Pattyn F, Rihani A, Van Maerken T, Vermeulen J, Kumps C, Menten B, De Preter K, Schramm A, Schulte J, Noguera R, Schleiermacher G, Janoueix-Lerosey I, Laureys G, Powel R, Nittner D, Marine JC, Ringnér M, Speleman F, Vandesompele J: **An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours.** *Oncogene* 2010, **29**:3583-3592.
28. Scaruffi P, Stigliani S, Moretti S, Coco S, De Vecchi C, Valdora F, Garaventa A, Bonassi S, Tonini GP: **Transcribed-Ultra Conserved Region expression is associated with outcome in high-risk neuroblastoma.** *BMC Cancer* 2009, **9**:441.
29. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**:646-674.
30. Orlov YL, Zhou J, Lipovich L, Shahab A, Kuznetsov VA: **Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis.** *In Silico Biol* 2007, **7**:241-260.
31. Badger JH, Olsen GJ: **CRITICA: coding region identification tool invoking comparative analysis.** *Mol Biol Evol* 1999, **16**:512-524.
32. Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, Sunkara S, Furuno M, Bult CJ, Quackenbush J, Kai C, Kawai J, Carninci P, Hayashizaki Y, Pesole G, Mattick JS: **Discrimination of non-protein-coding transcripts from protein-coding mRNA.** *RNA Biol* 2006, **3**:40-48.
33. Kuznetsov VA, Knott GD, Bonner RF: **General statistics of stochastic process of gene expression in eukaryotic cells.** *Genetics* 2002, **161**:1321-1332.
34. Chua ALS, Ivshina AV, Kuznetsov VA: **Pareto-gamma statistic reveals global rescaling in transcriptomes of low and high aggressive breast cancer phenotypes.** *Lecture Notes in Computer Science* 2006, **4146**:49-59.
35. Batagov AO, Yarmishyn AA, Jenjaroenpun P, Tan JZ, Nishida Y, Kurochkin IV: **Role of genomic architecture in the expression dynamics of long noncoding RNAs during differentiation of human neuroblastoma cells.** *BMC Syst Biol* 2013, **7**(Suppl 3):S11.
36. Cornero A, Acquaviva M, Fardin P, Versteeg R, Schramm A, Eva A, Bosco MC, Blengio F, Barzaghi S, Varesio L: **Design of a multi-signature ensemble classifier predicting neuroblastoma patients' outcome.** *BMC Bioinformatics* 2012, **13**(Suppl 4):S13.
37. Encinas M, Iglesias M, Llecha N, Comella JX: **Extracellular-regulated kinases and phosphatidylinositol 3-kinase are involved in brain-derived neurotrophic factor-mediated survival and neurogenesis of the neuroblastoma cell line SH-SY5Y.** *J Neurochem* 1999, **73**:1409-1421.
38. Mattick JS: **The genetic signatures of noncoding RNAs.** *PLoS Genet* 2009, **5**:e1000459.
39. St Laurent G1, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, Urcuqui-Inchima S, Seilheimer B, McCaffrey TA, Kapranov P: **Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells.** *BMC Genomics* 2012, **13**:504.
40. Tahira AC1, Kubrusly MS, Faria MF, Dazzani B, Fonseca RS, Maracaja-Coutinho V, Verjovski-Almeida S, Machado MC, Reis EM: **Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer.** *Mol Cancer* 2011, **10**:141.
41. Shah N, Sukumar S: **The Hox genes and their roles in oncogenesis.** *Nat Rev Cancer* 2010, **10**:361-371.
42. Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W: **Using genomic data to unravel the root of the placental mammal phylogeny.** *Genome Res* 2007, **17**:413-421.
43. Tavaré S, Marshall CR, Will O, Soligo C, Martin RD: **Using the fossil record to estimate the age of the last common ancestor of extant primates.** *Nature* 2002, **416**:726-729.
44. Raau RL, Sterner KN, Novelli CM, Stewart CB, Disotell TR: **Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence.** *J Hum Evol* 2005, **48**:237-257.
45. Miyazaki YJ, Hamada J, Tada M, Furuuchi K, Takahashi Y, Kondo S, Katoh H, Moriuchi T: **HOXD3 enhances motility and invasiveness through the TGF-beta-dependent and -independent pathways in A549 cells.** *Oncogene* 2002, **21**:798-808.
46. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, Widschwendter M, Vierkant RA, Larson MC, Kjaer SK, Birrer MJ,

- Berchuck A, Schildkraut J, Tomlinson I, Kiemeny LA, Cook LS, Gronwald J, Garcia-Closas M, Gore ME, Campbell I, Whittemore AS, Sutphen R, Phelan C, Anton-Culver H, Pearce CL, Lambrechts D, Rossing MA, Chang-Claude J, Moysich KB, Goodman MT, et al: **A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24.** *Nat Genet* 2010, **42**:874-879.
47. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, Mitsudomi T, Takahashi T: **Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors.** *J Clin Oncol* 2006, **24**:1679-1688.
 48. Champine PJ, Michaelson J, Weimer BC, Welch DR, DeWald DB: **Microarray analysis reveals potential mechanisms of BRMS1-mediated metastasis suppression.** *ClinExp Metastasis* 2007, **24**:551-565.
 49. Lee DC, Kang YK, Kim WH, Jang YJ, Kim DJ, Park IY, Sohn BH, Sohn HA, Lee HG, Lim JS, Kim JW, Song EY, Kim DM, Lee MN, Oh GT, Kim SJ, Park KC, Yoo HS, Choi JY, Yeom YI: **Functional and clinical evidence for NDRG2 as a candidate suppressor of liver cancer metastasis.** *Cancer Res* 2008, **68**:4210-4220.
 50. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehauer H, Laczko E, Kuroski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Gomes V, Pastorelli A, Faggiani R, Anti M, Jiricny J, Clevers H, Marra G: **Transcriptome profile of human colorectal adenomas.** *Mol Cancer Res* 2007, **5**:1263-1275.
 51. Thiele CJ, Li Z, McKee AE: **On Trk-the TrkB signal transduction pathway is an increasingly important target in cancer biology.** *Clin Cancer Res* 2009, **15**:5962-5967.
 52. Chang F, Lee JT, Navolanic PM, Steelman LS, Shelton JG, Blalock WL, Franklin RA, McCubrey JA: **Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy.** *Leukemia* 2003, **17**:590-603.
 53. Opel D, Poremba C, Simon T, Debatin KM, Fulda S: **Activation of Akt predicts poor outcome in neuroblastoma.** *Cancer Res* 2007, **67**:735-745.
 54. Nakagawara A, Azar CG, Scavarda NJ, Brodeur GM: **Expression and function of TRK-B and BDNF in human neuroblastomas.** *Mol Cell Biol* 1994, **14**:759-767.
 55. Nakagawara A, Arima-Nakagawara M, Scavarda NJ, Azar CG, Cantor AB, Brodeur GM: **Association between high levels of expression of the TRK gene and favorable outcome in human neuroblastoma.** *N Engl J Med* 1993, **328**:847-854.
 56. Yamashiro DJ, Liu XG, Lee CP, Nakagawara A, Ikegaki N, McGregor LM, Baylin SB, Brodeur GM: **Expression and function of Trk-C in favourable human neuroblastomas.** *Eur J Cancer* 1997, **33**:2054-2057.
 57. Clauset A, Shalizi CR, Newman MEJ: **Power-law distributions in empirical data.** *SIAM Review* 2009, **51**:661-703.
 58. Janoueix-Lerosey I, Lequin D, Brugières L, Ribeiro A, de Pontual L, Combaret V, Raynal V, Puisieux A, Schleiermacher G, Pierron G, Valteau-Couanet D, Frebourg T, Michon J, Lyonnet S, Amiel J, Delattre O: **Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma.** *Nature* 2008, **455**:967-970.
 59. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency.** *Ann Statist* 2001, **29**:1165-1188.
 60. Ohtaki M, Otani K, Hiyama K, Kamei N, Satoh K, Hiyama E: **A robust method for estimating gene expression states using Affymetrix microarray probe level data.** *BMC Bioinformatics* 2010, **11**:183.
 61. Tibshirani R, Hastie T, Narasimhan B, Chu G: **Diagnosis of multiple cancer types by shrunken centroids of gene expression.** *Proc Natl Acad Sci USA* 2002, **99**:6567-6572.

doi:10.1186/1471-2164-15-S9-S7

Cite this article as: Yarmishyn et al.: HOXD-AS1 is a novel lncRNA encoded in HOXD cluster and a marker of neuroblastoma progression revealed via integrative analysis of noncoding transcriptome. *BMC Genomics* 2014 **15**(Suppl 9):S7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

